Development of 23 Novel Polymorphic EST-SSR Markers for the Endangered Relict Conifer Metasequoia glyptostroboides

Authors: Jin, Yuqing, Bi, Quanxin, Guan, Wenbin, and Mao, Jian-Feng

Source: Applications in Plant Sciences, 3(9)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1500038

Applications in Plant Sciences 2015 3(9): 1500038

PRIMER NOTE

DEVELOPMENT OF 23 NOVEL POLYMORPHIC EST-SSR MARKERS FOR THE ENDANGERED RELICT CONIFER *METASEQUOIA GLYPTOSTROBOIDES*[1]

YUQING JIN[2], QUANXIN BI[3], WENBIN GUAN[3], AND JIAN-FENG MAO[2,4]

[2]National Engineering Laboratory for Tree Breeding, College of Biological Sciences and Technology, Beijing Forestry University, Beijing 100083, People's Republic of China; and [3]School of Nature Conservation, Beijing Forestry University, Beijing 100083, People's Republic of China

- *Premise of the study: Metasequoia glyptostroboides* is an endangered relict conifer species endemic to China. In this study, expressed sequence tag–simple sequence repeat (EST-SSR) markers were developed using transcriptome mining for future genetic and functional studies.
- *Methods and Results:* We collected 97,565 unigene sequences generated by 454 pyrosequencing. A bioinformatics analysis identified 2087 unique and putative microsatellites, from which 96 novel microsatellite markers were developed. Fifty-three of the 96 primer sets successfully amplified clear fragments of the expected sizes; 23 of those loci were polymorphic. The number of alleles per locus ranged from two to eight, with an average of three, and the observed and expected heterozygosity values ranged from 0 to 1.0 and 0.117 to 0.813, respectively.
- *Conclusions:* These microsatellite loci will enrich the genetic resources to develop functional studies and conservation strategies for this endangered relict species.

Key words: 454 pyrosequencing; Cupressaceae; EST-SSR; *Metasequoia glyptostroboides*; microsatellite; relict plant.

Metasequoia glyptostroboides H. H. Hu & W. C. Cheng (Cupressaceae), the dawn redwood, is a relict conifer of the genus *Metasequoia* H. H. Hu & W. C. Cheng. Its natural population has been found only in highly restricted regions bordering Hubei, Hunan, and Chongqing provinces in China (Farjon, 2001). This tree species is valued for its essential oils and crude extracts, which have great potential for antifungal activity. Moreover, this conifer is widely used as an ornamental tree in eastern China. Given its limited natural population size, *M. glyptostroboides* has been listed as an endangered species by both the Chinese government and the International Union for Conservation of Nature (Li et al., 2005).

Low genetic diversity was previously found in both wild and artificial populations of *M. glyptostroboides* using random amplified polymorphic DNAs (RAPDs) and amplified fragment length polymorphisms (AFLPs) (Chen et al., 2003). Nevertheless, dominant markers, including RAPDs and AFLPs, cannot provide reliable estimates of genetic diversity because they are unable to distinguish heterozygotes from homozygotes (Nybom, 2004). This drawback can be overcome using simple sequence repeats (SSRs). In contrast to RAPDs and AFLPs, SSRs are typically studied separately at each identified locus and can be regarded as codominant markers (Nybom, 2004). Cui et al. (2010) developed 11 polymorphic microsatellite markers with a high polymorphism information content (PIC) by sequencing a microsatellite-enriched library. In recent years, with increasing exploration of expressed sequence tags (ESTs), EST-derived SSRs (EST-SSRs) have emerged as useful tools for estimating functional variation (Andersen and Lübberstedt, 2003; Varshney et al., 2007; Zalapa et al., 2012). Additional genetic markers, especially codominant EST-SSRs, are valuable and will be critical for resolving finer genetic variation patterns and developing functional studies and conservation strategies for this endangered species.

Zhao et al. (2013) generated 1,571,764 high-quality reads (assembled into 97,565 unigenes) from vegetative and female buds by transcriptome sequencing using 454 pyrosequencing technology, which provides a large amount of sequence information for microsatellite mining. In this study, we referenced the transcriptome sequences and developed the first set of EST-SSR markers for *M. glyptostroboides*.

METHODS AND RESULTS

The 97,565 unigene sequences were downloaded from http://www.genome.clemson.edu/node/273 (presented by the transcriptome sequencing project) (Zhao et al., 2013) and used for subsequent microsatellite mining. We first detected the microsatellite sequences from the unigene data set and identified unique microsatellites through an all-against-all BLAST search. Finally, we successfully designed primer pairs for 2087 unique EST-SSR loci. A bioinformatics analysis was implemented for microsatellite development and primer design using QDD version 3.1 (Meglécz et al., 2014). We selected 96 di- or trinucleotide loci with large numbers of repeats for primer synthesis and tested them in eight mature individuals collected from the Beijing Botanical Garden with permission. The sampled trees were introduced separately from natural populations in three counties (Lichuan [108.56°N, 30.18°E], Shizhu [28.28°N, 108.58°E], and Longshan [29.59°N, 106.50°E]) in the 1950s. Voucher specimens

[1]Manuscript received 3 April 2015; revision accepted 4 June 2015.

This work was supported by the Fundamental Research Funds for the Central Universities (no. YX2013-41).

[4]Author for correspondence: jianfeng.mao@bjfu.edu.cn

doi:10.3732/apps.1500038

were deposited in the herbarium at Beijing Forestry University (accession no.: BFU-shuishan201403). Genomic DNA was extracted from the leaves of eight individuals using the cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). Functional annotations were prepared for each EST-SSR locus with polymorphisms detected by querying the Pfam protein families database (Finn et al., 2014).

The 96 selected primer pairs were validated by PCR using the M13-tail technique (Schuelke, 2000) to test for polymorphisms. Three primers were synthesized for each genotyping experiment: a 5′ M13-tailed forward primer, a reverse primer, and a fluorescently labeled M13 primer. M13 primers carrying FAM, HEX, TAMRA, and ROX (BGI, Beijing, China) labels were used separately. Each PCR mixture contained 5 µL of 2× MIX (New Industry, Beijing, China), 10–20 ng of template DNA, 1.6 pmol of reverse primer, 1.6 pmol of single fluorescently labeled M13 primer, and 0.4 pmol of forward primer in a final volume of 10 µL. PCR was performed as follows in a thermal cycler: 94°C for 5 min; 28 cycles of 94°C for 40 s, annealing at 55°C for 40 s, and elongation at 72°C for 45 s; 10 cycles of 94°C for 40 s, annealing at 53°C for 40 s, and elongation at 72°C for 45 s; and a final extension at 72°C for 10 min. The products were analyzed using an ABI 3730 Prism Genetic Analyzer (Applied Biosystems, Foster City, California, USA).

The raw data were analyzed using GeneMarker version 2.62 (SoftGenetics, State College, Pennsylvania, USA). The number of alleles, observed heterozygosity (H_o), and expected heterozygosity (H_e) were calculated using GenAlEx version 6.5 (Peakall and Smouse, 2012). The allelic PIC was also calculated for each locus using CERVUS version 3.0 (Kalinowski et al., 2007).

Of the 96 loci, 53 demonstrated the expected fragment sizes; the remaining loci failed to amplify any product. After capillary electrophoresis scoring, 42 of the 53 loci showed a clear, single peak for each allele, of which 23 (53.49%) were polymorphic and 19 were monomorphic. The EST-SSR loci evaluated did not overlap with putative polymorphic SSRs described in a previous study that detected SSR loci from the same transcriptome data set (Zhao et al., 2013). The primer sequences, repeat motifs, sizes, and functional annotations from Pfam are shown in Table 1. The number of alleles from these 23 polymorphic EST-SSRs varied from two to eight (average: 3). The H_o, H_e, and PIC ranged from 0 to 1.0, 0.117 to 0.813, and 0.110 to 0.789, respectively (Table 2). The mean H_e was 0.551, which is higher than that estimated using RAPDs (Chen et al.,

Table 1. Characteristics of 23 polymorphic EST-SSR loci developed for *Metasequoia glyptostroboides*.[a]

Locus	Primer sequences (5′–3′)	Repeat motif	Allele size (bp)	Functional annotation[b]
BQSS1	F: GTTTCCTAGTTCTGCGCGAC	$(AT)_8$	301	CLP_protease, Ribosomal_S18, MGC-24
	R: TTGCTGCTCATTTCATTCCA			
BQSS2	F: CATTGGCAAATTCACTCCAAT	$(AC)_8$	335	No hit
	R: GACGAAGAGGCATTCCCAT			
BQSS3	F: GCCATTCGTATCTGTTTCCAA	$(AG)_9$	229	KH_1, KH_3, KH_2, KH_4, KH_5, SLS
	R: GCGGTATTTACGAGACGAGC			
BQSS7	F: CTGGCACTCCTGATATGCCT	$(AAG)_8$	154	DEAD, Helicase_C, Helicase_C_2, AAA_22, TATR
	R: GATGACTCGATGCAAGGGTT			
BQSS9	F: AAATCTCTATCGCCCTCCAA	$(AT)_{12}$	187	TauD, DUF1667, Glyoxalase_2
	R: ACTTGGCGAACCTGTAATCC			
BQSS10	F: GGAGGCGTTAGTGGAGGAA	$(AG)_3CG(AGG)_8AG$	140	No hit
	R: ATCATTTGTCCCTCTGCTGC			
BQSS17	F: TCCTCGGTTTCTAGTGCAGG	$(ACC)_8$	117	No hit
	R: CAAGGGAATGTACTCATCTCCA			
BQSS18	F: TCCTCCTCCAAATTATTGCG	$(AAC)_9$	291	RRM_1, RRM_6, RRM_5, PAM2
	R: ATTTCAGAGGCATTTCGGTG			
BQSS21	F: CCAAATACTGGTCGGATGCT	$(AGG)_8$	271	AP2
	R: TGCTCCTCCTCTTCTACCGA			
BQSS26	F: CTTTGGATTGCATACCTGGG	$(AAG)_9$	300	zf-C2H2_6, zf-C2H2_jaz, ThuA, zf-met, Tnp_zf-ribbon_2
	R: GCTTTACAAGTCAGCGGAGG			
BQSS27	F: TTCTGTGACAATGGATGGGA	$(AT)_8$	161	eIF-5a, KOW, EFP_N
	R: GCAAGGCTCTAAAGCTGGAA			
BQSS29	F: CTGGTTCAGGTTTGTGGGTC	$(AT)_8$	290	QH-AmDH_gamma
	R: TTCATTTGAGCTTGGCGG			
BQSS32	F: CACTTCTGCTCTCTTCTCATCG	$(AG)_{15}$	333	FBA_2, Pox_E10
	R: CACAGGAACAGGTAGGAAACG			
BQSS34	F: GGTCAAGAGCATCAGCTTCC	$(AT)_8$	193	VPS28
	R: ATGGTTCGGATTATGGCAAA			
BQSS35	F: AAGATGAATGCAATCCAGGC	$(AG)_9$	154	HEAT_2
	R: ATCAGGCAACTGTTTGGCAT			
BQSS36	F: AATCCTCAACACAAGGATGC	$(AG)_9$	182	GUCT
	R: TTGTACCTTCGTTTGCCAAG			
BQSS38	F: CAAACAAACCAACCAACCAA	$(AT)_9$	176	No hit
	R: CCTGAATTTGATCGAAGATGG			
BQSS45	F: TTGCTGTCATACTGTTGTCGTG	$(AG)_9(AT)_8A$	344	No hit
	R: CTTTGTTGAGTCTGTGCCGA			
BQSS46	F: GTGATGCAAGGTTAGTGCCC	$(AC)_8$	139	AHH
	R: GCATGTGTGTGTTTGAGGGA			
BQSS48	F: TGGGTGAGTTCAAGAAGTTGG	$(AT)_{12}$	277	No hit
	R: GTGTCCATACGCAAGGGTTT			
BQSS49	F: CAAGCCATAGCATAGGCACA	$(AAG)_8$	343	SOBP, Plexin_cytopl, 7tm_7, Herpes_US9
	R: GATGGGTTGTCGAGTGGTTT			
BQSS50	F: CTTCAATCCACTTGTCTTGCC	$(ATGC)_{12}$	128	RP-C_C
	R: GACAAGTAACCGGAACCGAA			
BQSS70	F: CTTTCATTTAGTGCATGGATGG	$(AC)_9$	174	No hit
	R: TTTGCATTGTAATCTTGCGG			

[a] Annealing temperature (T_a) for all loci was 55°C.
[b] Protein family names from function annotation prepared by querying the protein family database, Pfam (Finn et al., 2014).

TABLE 2. Genetic properties of 23 polymorphic EST-SSR loci developed for *Metasequoia glyptostroboides*.[a]

Locus	*A*	H_o	H_e	HWE[b]	PIC
BQSS1	3	0.000	0.594	**	0.511
BQSS2	3	0.625	0.508	ns	0.428
BQSS3	2	0.375	0.305	ns	0.258
BQSS7	2	0.500	0.469	ns	0.359
BQSS9	4	0.750	0.648	ns	0.592
BQSS10	3	0.625	0.539	ns	0.447
BQSS17	2	0.250	0.219	ns	0.195
BQSS18	4	0.750	0.609	ns	0.559
BQSS21	2	0.000	0.219	**	0.195
BQSS26	2	0.375	0.430	ns	0.337
BQSS27	4	0.500	0.719	ns	0.667
BQSS29	2	0.000	0.219	**	0.195
BQSS32	3	0.000	0.531	**	0.468
BQSS34	2	0.125	0.117	ns	0.110
BQSS35	2	0.250	0.219	ns	0.195
BQSS36	3	1.000	0.555	*	0.456
BQSS38	5	0.750	0.656	ns	0.595
BQSS45	4	0.143	0.704	*	0.646
BQSS46	2	0.125	0.117	ns	0.110
BQSS48	8	0.500	0.813	ns	0.789
BQSS49	2	0.375	0.305	ns	0.258
BQSS50	2	0.000	0.469	**	0.359
BQSS70	2	0.625	0.430	ns	0.337

Note: *A* = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity; HWE = Hardy–Weinberg equilibrium; PIC = polymorphism information content.

[a] Sequences were downloaded from the Clemson University Genomics Institute Metasequioa RNA-Seq transcriptome sequencing project (available at http://www.genome.clemson.edu/node/273).

[b] ns = not significant; $*P < 0.05$; $**P < 0.01$; $***P < 0.001$.

2003). Indicators including the percent of polymorphic loci and H_e suggest considerable genetic variability for this endangered conifer. Target sequences for the 23 microsatellite loci are attached as Appendix S1.

CONCLUSIONS

We developed the first set of EST-SSR markers for *M. glyptostroboides*. The 23 polymorphic loci reported here will facilitate finer estimation of the genetic diversity and population structure of this species, as well as the development of functional studies and conservation strategies.

LITERATURE CITED

ANDERSEN, J. R., AND T. LÜBBERSTEDT. 2003. Functional markers in plants. *Trends in Plant Science* 8: 554–560.

CHEN, X. Y., Y. Y. LI, T. Y. WU, X. ZHANG, AND H. P. LU. 2003. Size-class differences in genetic structure of *Metasequoia glyptostroboides* Hu et Cheng (Taxodiaceae) plantations in Shanghai. *Silvae Genetica* 52: 107–109.

CUI, M. Y., S. YU, M. LIU, AND Y. Y. LI. 2010. Isolation and characterization of polymorphic microsatellite markers in *Metasequoia glyptostroboides* (Taxodiaceae). *Conservation Genetics Resources* 2: 19–21.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of leaf tissue. *Phytochemical Bulletin* 19: 11–15.

FARJON, A. 2001. World checklist and bibliography of conifers. Royal Botanic Gardens, Kew, Richmond, Surrey, United Kingdom.

FINN, R. D., A. BATEMAN, J. CLEMENTS, P. COGGILL, R. Y. EBERHARDT, S. R. EDDY, A. HEGER, ET AL. 2014. Pfam: The protein families database. *Nucleic Acids Research* 42: D222–D230.

KALINOWSKI, S. T., M. L. TAPER, AND T. C. MARSHALL. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106.

LI, Y. Y., X. Y. CHEN, X. ZHANG, T. Y. WU, H. P. LU, AND Y. W. CAI. 2005. Genetic differences between wild and artificial populations of *Metasequoia glyptostroboides* Hu et Cheng (Taxodiaceae): Implications for species recovery. *Conservation Biology* 19: 224–231.

MEGLÉCZ, E., N. PECH, A. GILLES, V. DUBUT, P. HINGAMP, A. TRILLES, R. GRENIER, AND J.-F. MARTIN. 2014. QDD version 3.1: A user-friendly computer program for microsatellite selection and primer design revisited: Experimental validation of variables determining genotyping success rate. *Molecular Ecology Resources* 14: 1302–1313.

NYBOM, H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* 13: 1143–1155.

PEAKALL, R., AND P. E. SMOUSE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics (Oxford, England)* 20: 1–3.

SCHUELKE, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

VARSHNEY, R. K., K. CHABANE, P. S. HENDRE, R. K. AGGARWAL, AND A. GRANER. 2007. Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. *Plant Science* 173: 638–649.

ZALAPA, J. E., H. CUEVAS, H. Y. ZHU, S. STEFFAN, D. SENALIK, E. ZELDIN, B. MCCOWN, ET AL. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *American Journal of Botany* 99: 193–208.

ZHAO, Y., S. M. THAMMANNAGOWDA, S. TANG, X. XIA, W. YIN, AND H. LIANG. 2013. An EST dataset for *Metasequoia glyptostroboides* buds: The first EST resource for molecular genomics studies in *Metasequoia*. *Planta* 237: 755–770.