Development and Characterization of EST-SSR Markers in Bombax ceiba (Malvaceae)

Authors: Ju, Miao-Miao, Ma, Huan-Cheng, Xin, Pei-Yao, Zhou, Zhi-Li, and Tian, Bin

Source: Applications in Plant Sciences, 3(4)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1500001

Applications in Plant Sciences 2015 3(4): 1500001

PRIMER NOTE

DEVELOPMENT AND CHARACTERIZATION OF EST-SSR MARKERS IN *BOMBAX CEIBA* (MALVACEAE)[1]

MIAO-MIAO JU[2], HUAN-CHENG MA[2], PEI-YAO XIN[2], ZHI-LI ZHOU[2], AND BIN TIAN[2,3,4]

[2]Key Laboratory of Biodiversity Conservation in Southwest China, State Forestry Administration, Southwest Forestry University, Kunming 650224, People's Republic of China; and [3]Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, People's Republic of China

- *Premise of the study: Bombax ceiba* (Malvaceae), commonly known as silk cotton tree, is a multipurpose tree species of tropical forests. Novel expressed sequence tag–simple sequence repeat (EST-SSR) markers were developed and characterized for the species using transcriptome analysis.
- *Methods and Results:* A total of 33 new EST-SSR markers were developed for *B. ceiba*, of which 13 showed polymorphisms across the 24 individuals from four distant populations tested in the study. The results showed that the number of alleles per polymorphic locus ranged from two to four, and the expected heterozygosity and observed heterozygosity per locus varied from 0.043 to 0.654 and from 0 to 0.609, respectively.
- *Conclusions:* These newly developed EST-SSR markers can be used in phylogeographic and population genetic studies to investigate the origin of *B. ceiba* populations. Furthermore, these EST-SSR markers could also greatly promote the development of molecular breeding studies pertaining to silk cotton tree.

Key words: *Bombax ceiba*; EST-SSR; Malvaceae; transcriptome.

Bombax ceiba L. (Malvaceae), also known as silk cotton tree, is a multipurpose tree species of tropical forests. It provides food, fodder, fiber, fuel, medicine, and many other ecological benefits (Jain et al., 2011), and is an important component of tropical dry deciduous forest ecosystems (Chaudhary and Khadabadi, 2012). Most previous research has focused on biochemistry or morphological characteristics of *B. ceiba* (Baum and Oginuma, 1994; El-Hagrassi et al., 2011); however, few studies have focused on molecular markers, which are important for genetic studies and plant breeding. Until now, only nine genomic simple sequence repeat (SSR) markers for *B. ceiba* have been reported from our group (Tian et al., 2013). Compared to studies of its medicinal properties, studies of the genetics of *B. ceiba* have lagged behind many other economic trees mainly due to a lack of efficient genetic markers (Chaudhary and Khadabadi, 2012; Refaat et al., 2013). To generate high-density genetic markers and improve genetic diversity analysis, more SSR markers need to be developed for this species. Transcriptome sequencing is an efficient method for acquiring expressed sequence tag (EST)–SSR markers (Bouck and Vision, 2007) based on its low cost and high reliability. Therefore, in this paper, we have developed additional polymorphic EST-SSR primers for *B. ceiba* using transcriptome sequencing.

[1]Manuscript received 5 January 2015; revision accepted 21 February 2015.

This study was supported by the Doctoral Scientific Research Foundation of Southwest Forestry University and the National Natural Science Foundation of China (NSFC; 31260050).

[4]Author for correspondence: tianbinlzu@gmail.com

doi:10.3732/apps.1500001

METHODS AND RESULTS

In this study, fresh leaf tissues of three, one-year-old *B. ceiba* seedlings (from Gengma, Yunnan, China) were immediately frozen in liquid nitrogen and stored at −80°C for RNA extraction and transcriptome sequencing. The total RNA of *B. ceiba* was extracted using the cetyltrimethylammonium bromide (CTAB) method (Chang et al., 1993). The RNA quality and quantity were measured using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). Poly-T oligo-attached magnetic beads were used to isolate mRNA after extraction. Fragmentation buffer was added to produce short mRNA fragments. After fragmentation, cDNA was synthesized. The purified cDNA libraries were then amplified by PCR and sequenced by Illumina HiSeq 2000 (Illumina, San Diego, California, USA; sequencing performed by Encode Genomics Bio-Technology Company, Suzhou, Jiangsu Province, China). A total of 136,000,000 raw reads were generated, which were finally turned to 103,344,062 clean reads after removing adapter sequences and low-quality sequences to ensure the precision of acquired reads. Transcriptome de novo assembly was performed to generate a reference genome using Trinity (Grabherr et al., 2011). CD-HIT (Fu et al., 2012) was further used to cluster similar contigs and obtain a high-quality reference genome with nonredundant unigenes. We detected microsatellites using MISA Perl script (MIcroSAtellite identification tool, http://pgrc.ipk-gatersleben.de/misa/) from all unigenes obtained in the study. We screened for SSR motifs containing two to six nucleotides with minimum number of repeats as follows: seven for dinucleotide and five for trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide. Altogether, 71,203 SSR motifs were found, and 42 of them were selected to design primers using Primer3 software (Rozen and Skaletsky, 1999).

Twenty-four individuals of *B. ceiba* representing four distant natural populations (Appendix 1) were used to evaluate the polymorphisms of the target microsatellite loci. A voucher specimen of each population was deposited in the herbarium of Southwest Forestry University (SWFC; Appendix 1). Genomic DNA was extracted from silica-dried leaves using the DNA Extraction Kit (TIANGEN, Beijing, China) following the manufacturer's protocol. PCR amplifications were performed in 25-μL volumes that included 1 μL of genomic DNA, 1 μL of forward primer, 1 μL of reverse primer, 12.5 μL of PCR Master Mix, and 9.5 μL of ddH_2O. The PCR reactions were performed in the S1000 Thermal Cycler (Applied Biosystems, Foster City, California, USA) under the following conditions: initial denaturation was at 94°C for 5 min, followed by 35

TABLE 1. Characteristics of the 33 microsatellite markers developed for *Bombax ceiba*. Loci BC1–13 are polymorphic while loci BC14–33 are monomorphic across the 24 individuals from four distant populations tested in the study.

Locus	Primer sequences (5′–3′)	Repeat motif	Allele size range (bp)	T_a (°C)	Fluorescent dye	GenBank accession no.	BLAST top hit description [organism]	BLAST top hit accesssion no.	*E*-value
BC1	F: TACTCCGAAACTCACGCCTT R: AAAGACGTATCGGTGAAGCG	$(CTT)_7$	270–273	59	6-FAM	KP216639	Nonintrinsic ABC protein 6, putative isoform 2 [*Theobroma cacao*]	XM_007039483.1	2.00E-98
BC2	F: AAAGGTGACAGCTTGTTGGC R: TTTTGTCCATTTGTGCCTCA	$(TA)_{11}$	250–268	60	HEX	KP216640	No hit	—	—
BC3	F: CCTTCGTCCTCGTCTTCATC R: AATGACCCAGATGGCAACTC	$(TTC)_7$	204–207	59	HEX	KP216642	No hit	—	—
BC4	F: CTTCCTTTTCTTGGGAGGCT R: GACCCAGAGGATGAGGATGA	$(TCA)_7$	153–156	59	NED	KP216643	No hit	—	—
BC5	F: ACAAAAATGGCGTCTTCTGG R: GCAGGAGATCCATGGTGATT	$(CAG)_7$	128–134	60	6-FAM	KP216644	No hit	—	—
BC6	F: CCCTTGCCAGATTTCTTTGA R: GGAAAGAGTTGATAGCGGCA	$(TG)_{10}$	149–165	60	6-FAM	KP216645	No hit	—	—
BC7	F: GTGGCATACAGTGCTTCCCT R: GCAGCCTTCGGGTACATATT	$(CA)_{10}$	248–250	60	NED	KP216646	No hit	—	—
BC8	F: CTCCTGTCCCGAACATCAAT R: GGAGCTTCGGAGTTGTCTTG	$(CGA)_8$	156–168	59	NED	KP216647	No hit	—	—
BC9	F: TTTGAAAGGGAGGGTGTTTG R: GAGGAGGAAAAGTTATGTTTTGG	$(GACT)_6$	134–138	57	HEX	KP216648	No hit	—	—
BC10	F: ACCTCCTGCACAGACCATTC R: CATGGGGGAAAATTTTGTTG	$(ACA)_8$	213–216	60	6-FAM	KP216649	No hit	—	—
BC11	F: TGGGAGCTGAGATTTGATCC R: CCCCACTGGATTGATTGATT	$(CAGC)_6$	316–320	60	6-FAM	KP216650	No hit	—	—
BC12	F: TCCATCCAAATTCAGCAACA R: GGTTTGCTGCAAGGAGAGTC	$(CAG)_8$	147–150	60	NED	KP216651	Auxin efflux facilitator isoform 6 [*Theobroma cacao*]	XM_007045067.1	3.00E-37
BC13	F: CACGTGTGGAGAAGCTGAAA R: ATTTTGATGGCTTCCACCTG	$(CTG)_7$	270–281	59	HEX	KP216652	No hit	—	—
BC14	F: GCAACCAAGCTCTCGACTTT R: CTTTAAATCTGGCACGGCAT	$(ATA)_7$	280	60		KP723832	No hit	—	—
BC15	F: CAGTGTGGTGGAAGTTGTGG R: GATCGTGAATGCCAATCTCA	$(GTG)_7$	191	58		KP216641	SET domain protein [*Theobroma cacao*]	XM_007045968.1	3.00E-51
BC16	F: CTTCACATTTCTTCGCCTCC R: TGTCCTTGGCCTTGTAAACC	$(CT)_{10}$	208	57		KP723833	Kinase cdc2 homolog B [*Vitis vinifera*]	XM_002266587.2	2.00E-13
BC17	F: CGAGACTGACCCCGAAGTAG R: AATCGAGCAGGGAGTTTGAA	$(TTA)_7$	277	60		KP723834	No hit	—	—
BC18	F: CCCTCGCTTTCTCTCTTGAA R: TCCCATTCTCGATTTCCTTG	$(TAAT)_6$	196	59		KP723835	Translationally controlled tumor protein homolog [*Vitis vinifera*]	XM_002283806.2	7.00E-23
BC19	F: TTCAAAGAATAACCGTCGGC R: GCTCTCATTCCCTTGACTTCC	$(AAAG)_6$	260	60		KP723836	No hit	—	—
BC20	F: GCCTCTCCCCAAATCTCATT R: AGACCCTCCGGTATCCATTC	$(TG)_{10}$	149	58		KP723837	No hit	—	—
BC21	F: TTTTGAGAGGGCAAGAGAGG R: GCGGCTGAATTTTATGGAAA	$(AT)_{10}$	205	58		KP723838	No hit	—	—
BC22	F: GGTGGTGGAGATGGTGTAGG R: CCGAAACGGAAATAAAGCAA	$(GA)_{10}$	263	60		KP723839	No hit	—	—
BC23	F: TGAAAGGGACCAGAAAATCG R: GCAATCGGAGGAATGAAATC	$(AC)_{10}$	230	57		KP723840	Basic helix-loop-helix DNA-binding superfamily protein, putative isoform 7 [*Theobroma cacao*]	XM_007040192.1	7.00E-33
BC24	F: TAGCGGAAGATTTCTCTGGC R: GTATCGATGGCTGTTGGGAT	$(CA)_{10}$	246	58		KP723841	No hit	—	—
BC25	F: TCCTCCACACTGTTTCCCAT R: ATCACTTTTCTCGGCCCTTT	$(CT)_{11}$	198	58		KP723842	Ubiquitin-like superfamily protein [*Theobroma cacao*]	XM_007045597.1	6.00E-44
BC26	F: CACCATACGAGTCGTGCCTA R: GAAGATCGACGAGTTGGCTC	$(AG)_{10}$	258	60		KP723843	No hit	—	—
BC27	F: GGACAAGGCTCCAAAAACAA R: AGCAGCTAGCATTCCCAGAA	$(TA)_{10}$	226	59		KP723844	Uncharacterized protein [*Theobroma cacao*]	XM_007051709.1	8.00E-18
BC28	F: TATTCGGGGGAACCCTAAAC R: GACCGACTGACAGCACAAAA	$(AG)_{11}$	108	59		KP723845	No hit	—	—
BC29	F: ACAAGCTGTAAGCAAGCCGT R: GGGTATCCTGGGAGGAAGAT	$(GA)_{10}$	124	60		KP723846	No hit	—	—

Table 1. Continued.

Locus	Primer sequences (5′–3′)	Repeat motif	Allele size range (bp)	T_a (°C)	Fluorescent dye	GenBank accession no.	BLAST top hit description [organism]	BLAST top hit accesssion no.	*E*-value
BC30	F: AAGCCCAACTACGACACCAC R: CTCTCACAAAACCTGTGCCA	$(CT)_{10}$	201	59		KP723847	No hit	—	—
BC31	F: CATGTGCTCCTCCACTACCA R: TTCATGACCAAAGCAAGCAG	$(TTG)_8$	148	60		KP723848	No hit	—	—
BC32	F: AAGGTTGCCGCATAACAATC R: TCGTTCACTTTGACAGCCAC	$(CCA)_8$	238	60		KP723849	Caspase, putative [*Theobroma cacao*]	XM_007022689.1	9.00E-72
BC33	F: TTCTCCCTATGGCTGTTTGG R: TATCCTTGAGCCAGCGCC	$(GAGG)_6$	155	57		KP723850	No hit	—	—

Note: T_a = optimal annealing temperature.

Table 2. Genetic diversity statistics for four populations of *Bombax ceiba* based on 13 newly developed polymorphic EST-SSR primers.[a]

	BN (N = 6)			LS (N = 6)			GM (N = 6)			LL (N = 6)			Total	Mean	
Locus	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e
BC1	2	0.667	0.500	2	0.167	0.153	2	0.500	0.375	2	0	0.500	2	0.348	0.423
BC2	2	0	0.444	3	0	0.611	2	0.167	0.153	1	0	0	3	0.043	0.542
BC3	1	0	0	1	0	0	2	0.167	0.153	1	0	0	2	0.043	0.043
BC4	1	0	0	1	0	0	3	0.333	0.403	2	0	0.375	3	0.087	0.198
BC5	1	0	0	1	0	0	2	0.167	0.153	1	0	0	2	0.043	0.043
BC6	1	0	0	4	0.833	0.583	2	0.167	0.486	3	0.250	0.531	4	0.304	0.654
BC7	2	0	0.320	2	0	0.480	2	0	0.278	2	0	0.500	2	0	0.500
BC8	4	0.333	0.514	3	0.500	0.403	4	0.667	0.694	3	0.500	0.406	4	0.478	0.570
BC9	2	0.667	0.444	2	0.667	0.500	2	0.500	0.486	2	0.750	0.469	2	0.609	0.499
BC10	2	0.667	0.444	2	0.167	0.153	2	0.500	0.486	2	0.500	0.375	2	0.435	0.476
BC11	1	0	0	2	0.667	0.444	2	0.167	0.486	2	0.250	0.469	2	0.261	0.423
BC12	2	0.444	0.485	2	0.167	0.153	2	0.167	0.375	1	0	0	2	0.261	0.287
BC13	2	0.486	0.530	2	0.167	0.375	3	0.500	0.403	2	0.250	0.219	3	0.348	0.461

Note: A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity.
[a]Locality and voucher information are provided in Appendix 1.

cycles at 94°C for 30 s, then annealing for 45 s at the optimal temperature for each primer pair (from 58–62°C, see Table 1), and 72°C for 1 min, with a final extension of 10 min at 72°C. To test the utility of the primers, PCR products were detected on 1% agarose gels. Finally, a total of 33 out of the 42 primer pairs were successfully amplified. The other primer pairs gave no product.

Fluorescence-based SSR genotyping was performed using Multiplex-Ready Technology as described by Hayden et al. (2008). The 5′ end primers of EST-SSRs were labeled with three different fluorescent dyes (6-FAM, HEX, and NED; Applied Biosystems) (Table 1). Fluorescently labeled PCR products generated using the protocols mentioned above were diluted 1 : 50 with ddH_2O. Further, 1 μL of the diluted PCR products was added to 12 μL of formamide and 0.1 μL of GeneScan 500 LIZ Size Standard (Applied Biosystems). We denatured samples for 5 min at 95°C and cooled on ice before loading onto an ABI 3730xl Sequence Analyzer (Life Technologies, Carlsbad, California, USA). Allele sizes and number of alleles per locus were called using GeneMarker version 2.4.1 (SoftGenetics, State College, Pennsylvania, USA). The polymorphic SSR loci were analyzed with POPGENE version 32 (Yeh et al., 1999) for the number of alleles per locus, observed heterozygosity, and expected heterozygosity.

A total of 33 EST-SSR markers were developed and characterized, of which 13 loci showed polymorphisms for *B. ceiba* among four populations. The corresponding sequences of these markers are stored in GenBank (Table 1). The number of alleles per locus ranged from two to four, expected heterozygosity per locus varied from 0.043 to 0.654, and observed heterozygosity varied from 0 to 0.609 (Table 2).

CONCLUSIONS

Here we developed and characterized 33 polymorphic EST-SSR markers for *B. ceiba* using transcriptome sequences obtained by an Illumina paired-end sequencing technique, of which 13 markers showed polymorphisms across 24 individuals from four distant populations. These newly developed SSR primers will enable development of phylogeographic and population genetic studies and help investigate the origin of Chinese *B. ceiba* populations. Furthermore, they will be particularly useful for identification of novel genes with traits of interest and markers to assist breeding in silk cotton tree. Additionally, the microsatellite markers reported here provide a valuable tool for forest management and could be tested on other Malvaceae species.

LITERATURE CITED

Baum, D. A., and K. Oginuma. 1994. A review of chromosome numbers in Bombacaceae with new counts for *Adansonia*. *Taxon* 43: 11–20.

Bouck, A., and T. Vision. 2007. The molecular ecologist's guide to expressed sequence tags. *Molecular Ecology* 16: 907–924.

Chang, S. J., J. Puryear, and J. Cairney. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* 11: 113–116.

Chaudhary, P. H., and S. S. Khadabadi. 2012. *Bombax ceiba* Linn.: Pharmacognosy, ethnobotany and phyto-pharmacology. *Pharmacognosy Communications* 2: 2–9.

El-Hagrassi, A. M., M. M. Ali, A. F. Osman, and M. Shaaban. 2011. Phytochemical investigation and biological studies of *Bombax malabaricum* flowers. *Natural Product Research* 25: 141–151.

Fu, L. M., B. F. Niu, Z. W. Zhu, S. T. Wu, and W. Z. Li. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics (Oxford, England)* 28: 3150–3152.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Hayden, M. J., T. M. Nguyen, A. Waterman, G. L. McMichael, and K. J. Chalmers. 2008. Application of multiplex-ready PCR for fluorescence-based SSR genotyping in barley and wheat. *Molecular Breeding* 21: 271–281.

Jain, V., S. K. Verma, S. K. Sharma, and S. S. Katewa. 2011. *Bombax ceiba* Linn: As an umbrella tree species in forests of southern Rajasthan, India. *Research Journal of Environmental Sciences* 5: 722–729.

Refaat, J., S. Y. Desoky, M. A. Ramadan, and M. S. Kamel. 2013. Bombacaceae: A phytochemical review. *Pharmaceutical Biology* 51: 100–130.

Rozen, S., and H. Skaletsky. 1999. Primer3 on the WWW for general users and for biologist programmers. *In* S. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

Tian, B., X. N. Tian, Y. L. Xu, and H. C. Ma. 2013. Developing SSR primers in *Bombax ceiba* (Bombacaceae). *Guihaia* 33: 465–467.

Yeh, F. C., R. C. Yang, and T. Boyle. 1999. POPGENE version 1.32: Microsoft Windows–based freeware for population genetic analysis, quick user guide. Center for International Forestry Research, University of Alberta, Edmonton, Alberta, Canada.

Appendix 1. Locality information for the sampled populations of *Bombax ceiba* used in this study. All voucher specimens are deposited at the herbarium of Southwest Forestry University (SWFC), Kunming, China.

Population code	Location	*N*	Geographic coordinates	Altitude (m)	Voucher no.
BN	Xishuangbanna, Yunnan	6	21°53′N, 100°59′E	570	2010BN
LS	Lushui, Yunnan	6	25°34′N, 98°52′E	1060	2011LS
GM	Gengma, Yunnan	6	23°22′N, 99°38′E	890	2011GM
LL	Longling, Yunnan	6	24°19′N, 99°01′E	750	2011LL

Note: *N* = number of individuals.