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

MIAO-MIAO JU2, HUAN-CHENG MA2, PEI-YAO XIN2, ZHI-LI ZHOU2, AND BIN TIAN3,4

1 Key Laboratory of Biodiversity Conservation in Southwest China, State Forestry Administration, Southwest Forestry University, Kunming 650224, People’s Republic of China; and 2 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.

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 sequence tags 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 Skaltsky, 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 ddH2O. 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

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).

2 Key Laboratory of Biodiversity Conservation in Southwest China, State Forestry Administration, Southwest Forestry University, Kunming 650224, People’s Republic of China.

3 Key Laboratory of Biodiversity and Biogeography, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204, People’s Republic of China.

4 Author for correspondence: tianbinlzu@gmail.com

doi:10.3732/apps.1500001

Applications in Plant Sciences 2015 3(4): 1500001; http://www.bioone.org/loi/apps © 2015 Ju et al. Published by the Botanical Society of America. This work is licensed under a Creative Commons Attribution License (CC-BY-NC-SA).
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<sub>e</sub> (°C) | Fluorescent dye | GenBank accession no. | BLAST top hit description [organism] | BLAST top hit accession no. | E-value |
|-------|--------------------------|-------------|------------------------|----------------|----------------|---------------------|--------------------------------------|-----------------------------|---------|
| BC1   | F: TACTCCGAAAATCCAGGCTTTT | (CTT)<sub>7</sub> | 270–273 | 59 | 6-FAM | KP216639 | Nonintrinsic ABC protein 6, putative isoform 2 [Theobroma cacao] | XM_007039483.1 | 2.00E-98 |
|       | R: AAGAGCTATGGGAAGGAGGCTT | | | | | | | | |
| BC2   | F: AAAGGAGCATCGGTGGTTGCC | (TA)<sub>11</sub> | 250–268 | 60 | HEX | KP216640 | No hit | — | — |
|       | R: TTTTGGCCTATTGTTGCTCA | | | | | | | | |
| BC3   | F: CCTGTCCTGCTGCTTTCACTC | (TTC)<sub>11</sub> | 204–207 | 59 | HEX | KP216642 | No hit | — | — |
|       | R: AATGACCCGAGTGGGACACTC | | | | | | | | |
| BC4   | F: CTGGCTTTTCCTGGGAGGCTT | (TCA)<sub>7</sub> | 153–156 | 59 | NED | KP216643 | No hit | — | — |
|       | R: GCCAGAGGGAGGAGAGGAGGA | | | | | | | | |
| BC5   | F: ACACAAATGTGCTTCTGAGG | (CAG)<sub>7</sub> | 128–134 | 60 | 6-FAM | KP216644 | No hit | — | — |
|       | R: GCAGGAGATCCATGTTGATTT | | | | | | | | |
| BC6   | F: CTTGTGGAGATTTGGTCTGA | (TG)<sub>10</sub> | 149–165 | 60 | 6-FAM | KP216645 | No hit | — | — |
|       | R: GGAAAGTGTTAGACGGCAAGG | | | | | | | | |
| BC7   | F: GTGGAGATACAGCTGCTCTCT | (CA)<sub>10</sub> | 248–250 | 60 | NED | KP216646 | No hit | — | — |
|       | R: GCAGCTCTGGTGATCATATTT | | | | | | | | |
| BC8   | F: CTCCCTGCGGCTACATCAT | (CGA)<sub>8</sub> | 156–168 | 60 | NED | KP216647 | No hit | — | — |
|       | R: GGTTTGCTGCAAGGAGAGTC | | | | | | | | |
| BC9   | F: TTTGAAAGGGAGGGTGTTTG | (GACT)<sub>6</sub> | 134–138 | 57 | HEX | KP216648 | No hit | — | — |
|       | R: GAGGAGGAAAAGTTATGTTTTG | | | | | | | | |
| BC10  | F: ACCTCCTGCACAGACACATT | (ACA)<sub>6</sub> | 213–216 | 60 | 6-FAM | KP216649 | No hit | — | — |
|       | R: CATGGGGGAAAATTTTGTTG | | | | | | | | |
| BC11  | F: NTGGAGTCTGATGGTCTGAC | (CAGC)<sub>6</sub> | 316–320 | 60 | 6-FAM | KP216650 | No hit | — | — |
|       | R: CCCCACTGGATTGATTGATT | | | | | | | | |
| BC12  | F: TTCATTTTCTCTGCGTGAAG | (CAG)<sub>8</sub> | 147–150 | 60 | NED | KP216651 | Auxin efflux facilitator isoform 6 [Theobroma cacao] | XM_007045067.1 | 3.00E-37 |
|       | R: GGTGTTTCTGCTGAGATGGCT | | | | | | | | |
| BC13  | F: CACGTTGGGAGAAGCCTGAA | (CTG)<sub>7</sub> | 270–281 | 59 | HEX | KP216652 | No hit | — | — |
|       | R: ATTTTGTGCTCCACAGCCTT | | | | | | | | |
| BC14  | F: GCCCAACGCTCTCAGCAGATT | (ATA)<sub>7</sub> | 280 | 60 | | KP273832 | No hit | — | — |
|       | R: CTTTTAATCTGAGACAGCAT | | | | | | | | |
| BC15  | F: CAGTGTGGATGATTTTGAGG | (GTG)<sub>7</sub> | 191 | 58 | | KP216641 | SET domain protein [Theobroma cacao] | XM_007045968.1 | 3.00E-51 |
|       | R: GATTTTTTTTTTTTTTTTTT | | | | | | | | |
| BC16  | F: CTGTCAGATTCTGGCCCTCTC | (CT)<sub>10</sub> | 208 | 58 | | KP273833 | kinase cdc2 homolog B [Vitis vinifera] | XM_002266587.2 | 2.00E-23 |
|       | R: TGCTCTTTGCCGTGTTAAACC | | | | | | | | |
| BC17  | F: CGGACGCTGACCCCGAGATTG | (TTA)<sub>7</sub> | 277 | 60 | | KP273834 | No hit | — | — |
|       | R: AATCGCTAGCAGGGATTGAAA | | | | | | | | |
| BC18  | F: CCTGCTTTTTCTCCTGCGAA | (TAAT)<sub>6</sub> | 196 | 59 | | KP273835 | Transcriptionally controlled tumor protein homolog [Vitis vinifera] | XM_002283806.2 | 7.00E-23 |
|       | R: TCCATATTCTGGCTAAGG | | | | | | | | |
| BC19  | F: TTTAGCCAATACCCGTGCCC | (AAAG)<sub>6</sub> | 260 | 60 | | KP273836 | No hit | — | — |
|       | R: GCTCTCCTATCCCTGAGATCC | | | | | | | | |
| BC20  | F: GCTCTCCCTCCAATCTCATT | (TG)<sub>10</sub> | 149 | 58 | | KP273837 | No hit | — | — |
|       | R: AGACCTCCTGGATATCCATTC | | | | | | | | |
| BC21  | F: TTTTTAGGGAGGAGAAGGAGG | (AT)<sub>10</sub> | 205 | 58 | | KP273838 | No hit | — | — |
|       | R: TCTCTCTGATGGTTAAGAA | | | | | | | | |
| BC22  | F: GTGGTGGAGATGGTGTAGG | (GA)<sub>10</sub> | 263 | 60 | | KP273839 | No hit | — | — |
|       | R: CGGACGCTGACCCCGAGATTG | | | | | | | | |
| BC23  | F: TGAAAGGGACGAAAGAATCG | (AC)<sub>10</sub> | 230 | 57 | | KP273840 | Basic helix-loop-helix DNA-binding superfamily protein, putative isoform 7 [Theobroma cacao] | XM_007040193.1 | 7.00E-33 |
|       | R: GCAATTTTTCGAGGGAATG | | | | | | | | |
| BC24  | F: TAGGGGATGCTTCTGCGCC | (CA)<sub>10</sub> | 246 | 58 | | KP273841 | No hit | — | — |
|       | R: GTACGCTATGGCTTGGGAGAT | | | | | | | | |
| BC25  | F: TCTCCGACCATGGTTCCTATT | (CT)<sub>11</sub> | 198 | 58 | | KP273842 | Ubiquitin-like superfamily protein [Theobroma cacao] | XM_007045597.1 | 6.00E-44 |
|       | R: ATCCACTCTTCCCGCTTTTT | | | | | | | | |
| BC26  | F: CACCATGATGCTGCTGCTT | (AG)<sub>10</sub> | 258 | 60 | | KP273843 | No hit | — | — |
|       | R: GAGATGCGAGGCTGGCTTC | | | | | | | | |
| BC27  | F: GCAAGGCTGCTCCTGAGGAA | (TA)<sub>10</sub> | 226 | 59 | | KP273844 | Uncharacterized protein [Theobroma cacao] | XM_007051709.1 | 8.00E-18 |
|       | R: AGCGACTGATCTCCCGAGAA | | | | | | | | |
| BC28  | F: TACCTTGGGGGAGACCTAAC | (AG)<sub>11</sub> | 108 | 59 | | KP273845 | No hit | — | — |
|       | R: GACGAGCTGACAGCCAAA | | | | | | | | |
| BC29  | F: ACAAGCTCTGAAAGCGCCTT | (GA)<sub>10</sub> | 124 | 60 | | KP273846 | No hit | — | — |
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-SSR products were labeled using the protocols mentioned above were diluted 1:50 with ddH₂O. NED; Applied Biosystems) (Table 1). Fluorescently labeled PCR products generated using the protocols mentioned above were diluted 1:50 with ddH₂O. 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

BAILM, 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.

BOUCK, A., AND T. VISION. 2007. The molecular ecologist’s guide to expressed sequence tags. Molecular Ecology 16: 907–924.

BOUCK, A., AND T. VISION. 2007. The molecular ecologist’s guide to expressed sequence tags. Molecular Ecology 16: 907–924.

EL-HAGRASSI, A. M., M. M. ALL, A. F. OSMAN, AND M. SHAABAN. 2011. Phytochemical investigation and biological studies of Bombax mallebaricum flowers. Natural Product Research 25: 141–151.

FU, L. M., B. F. NHU, Z. W. ZHENG, S. T. WU, AND W. Z. LI. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics (Oxford, England) 28: 3150–3152.
Applications in Plant Sciences 2015 3(4): 1500001
doi:10.3732/apps.1500001

Ju et al.—Bombax ceiba microsatellites

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            | Genga, 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.