Applications in Plant Sciences

Applications in Plant Sciences 2016 4(10): 1600067

Primer Note

# Development and characterization of novel EST-SSR markers for *Speranskia tuberculata* (Euphorbiaceae)[1]

Yi Fu[2], Miao-Miao Ju[2], Huan-Cheng Ma[2], Pei-Yao Xin[2], Cheng-Zhong He[2], Dong-Rui Jia[3,5], and Bin Tian[2,4,5]

[2]Key Laboratory of Biodiversity Conservation in Southwest China, State Forestry Administration, Southwest Forestry University, Kunming 650224, People's Republic of China; [3]School of Ecology and Environmental Science, Yunnan University, Kunming 650091, People's Republic of China; and [4]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:* The first set of expressed sequence tag–simple sequence repeat (EST-SSR) markers were developed and characterized for *Speranskia tuberculata* (Euphorbiaceae), a traditional medicinal plant endemic to northern China, to explore the effects of recent habitat fragmentation on the genetic diversity and structure of this species.
- *Methods and Results:* In this study, a total of 18 novel polymorphic microsatellite (EST-SSR) markers were developed for *S. tuberculata* using high-throughput transcriptome sequencing. Analysis of 24 individuals of *S. tuberculata* from four natural populations revealed their robust polymorphic reliability. The number of alleles per locus ranged from two to 11, while the expected and observed heterozygosity per marker varied from 0.187 to 0.827 and 0.042 to 0.917, respectively. Of these markers, 13 showed good amplification results in the closely related species *S. cantonensis.*
- *Conclusions:* These newly generated SSR markers are expected to provide novel tools for genetic studies of *S. tuberculata*, which will contribute to the conservation and sustainable use of the species' wild genetic resources.

**Key words:** Euphorbiaceae; expressed sequence tag–simple sequence repeat (EST-SSR); *Speranskia cantonensis*; *Speranskia tuberculata*; transcriptome sequencing.

*Speranskia* Baill. (Euphorbiaceae) is a small genus endemic to China, comprising three herbaceous perennial species: *S. tuberculata* (Bunge) Baill., *S. cantonensis* (Hance) Pax & K. Hoffm., and *S. yunnanensis* S. M. Hwang (Hwang, 1989). *Speranskia tuberculata* is endemic to northern China and occurs on grassy slopes, grasslands, and thickets. The entire plant is commonly used for Chinese traditional medicine (Mazzio et al., 2014). Although *S. tuberculata* is not listed in the IUCN Red List, it is exhibiting a general decreasing trend or even disappearing completely in many distributional areas because of agricultural intensification and overexploitation of natural population resources. To explore the genetic consequences of recent habitat fragmentation for this medical plant and generate useful information to facilitate the conservation and sustainable use of wild genetic resources, we developed the first set of 18 polymorphic expressed sequence tag–simple sequence repeat (EST-SSR) markers for *S. tuberculata* using high-throughput transcriptome sequencing. We also tested these developed markers in *S. cantonensis*, a closely related species (Hwang, 1989), to identify their cross-species utility.

## METHODS AND RESULTS

Fresh leaves of *S. tuberculata* seedlings were gathered in Beijing (39°59′06″N, 116°02′04″E; voucher specimen accession no. TB2013079, deposited at the Herbarium of Southwest Forestry University [SWFC], Kunming, China) and immediately frozen in liquid nitrogen, and then stored at −80°C. RNA extraction, cDNA library construction, and transcriptome sequencing were conducted following the procedures previously described by Ju et al. (2015). After removing adapter sequences and low-quality sequences, a total of 86,138,489 nonredundant unigenes were assembled from 95,791,418 raw reads. A high-quality reference genome with nonredundant unigenes was then generated by performing de novo transcriptome assembly using Trinity with the parameter of full clean up (Grabherr et al., 2011) and clustering similar contigs using CD-HIT with default parameters (Fu et al., 2012). Furthermore, we used MISA Perl script (MIcroSAtellite identification tool; Thiel et al., 2003) to screen for SSR motifs from all unigenes, and the minimum numbers of repeats were set as seven, five, five, five, and five for di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively. MISA recovered a total of 26,202 SSR motifs, of which 30 were randomly selected for primer design using Primer3 software (Rozen and Skaletsky, 1999). The major parameters for primer pair design were set as follows: primer length of 15–25 bases, PCR product size of 100–400 bp, and annealing temperatures of 55–60°C.

The 30 target EST-SSR markers were initially tested for amplification using DNA from 24 *S. tuberculata* individuals from four natural populations located in different provinces across the distributional range in northern China (populations YA, XZ, YT, and KQ; Appendix 1). Total genomic DNA was extracted from silica gel–dried leaves using the Ezup DNA Extraction Kit (Sangon Biotech, Shanghai, China) following the manufacturer's protocol. PCRs were performed using the S1000 Thermal Cycler (Applied Biosystems, Foster City, California, USA) in a 25-μL total volume with 1 μL (~10 ng) of genomic DNA, 12.5 μL of *Taq* PCR Mix (Sangon Biotech), 9.5 μL of $ddH_2O$, and 1 μL (5 pmol) of each primer. The PCR program consisted of 10 min of initial denaturation at 95°C; followed by 35 cycles of denaturation at 94°C for 45 s, annealing at specific temperature (58–60°C; Table 1) for 1 min, extension at 72°C for 1 min; and a final extension at 72°C for 10 min. All PCR products were run on 1% agarose gels to check for successful amplification. Twenty primer pairs produced clear amplicons of the expected size ranges. Multiplex-Ready PCR technology (Hayden et al., 2008) was then applied for fluorescence-based SSR genotyping. Forward primers for the 20 successfully amplified loci were labeled with three different fluorescent dyes (6-FAM, HEX, and NED; Applied Biosystems; Table 1) and used for

[1]Manuscript received 3 June 2016; revision accepted 21 July 2016.

This study is supported by the National Natural Science Foundation of China (NSFC; 31260050).

[5]Authors for correspondence: tianbinlzu@gmail.com, jia.drui@ynu.edu.cn

doi:10.3732/apps.1600067

 

TABLE 1. Characteristics of the 18 polymorphic microsatellite markers developed for *Speranskia tuberculata*.

| 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 accession no. | *E*-value |
|---|---|---|---|---|---|---|---|---|---|
| 8852 | F: GTGCCTCCATCCGGAAAT<br>R: CAACAGCAGCAAAAACAACAA | $(TGC)_5$ | 361–370 | 60 | 6-FAM | KT285024 | Probable methyltransferase PMT27 [*Ricinus communis*] | XP_002533655.1 | 2E-10 |
| 9441 | F: CCAAAAAGCTAAAACCACTCG<br>R: CTGCTGCTGCTGCTTTTG | $(GCA)_5$ | 225–240 | 59 | 6-FAM | KT285026 | Trihelix transcription factor GTL1 [*Ricinus communis*] | XP_002516129.1 | 0.008 |
| 9832 | F: CTTGCACCTCCAACTCCG<br>R: AGCTTGAGCATGAGCGAGA | $(TCA)_5$ | 192–201 | 60 | 6-FAM | KT285027 | Uncharacterized protein At1g65710 [*Ricinus communis*] | XP_002521410.1 | 0.0000003 |
| 10026 | F: TGCAATTGATTGACATTGTTG<br>R: CACGCGTCCTCAAAGACC | $(TGA)_6$ | 154–172 | 59 | 6-FAM | KT285028 | No hit | | |
| 10117 | F: CCTCAAATCCATCGCCAC<br>R: CGGGGAGTTTCGGAGAAT | $(CAA)_5$ | 203–209 | 60 | NED | KT285029 | Pentatricopeptide repeat-containing protein At3g49240 [*Ricinus communis*] | XP_002516677.1 | 2.00E-05 |
| 10128 | F: TCCAGGGTCGAGATTTGG<br>R: GCAAACCAAGAAAGCCGT | $(TTC)_5$ | 150–159 | 59 | NED | KT285030 | Rho GTPase-activating protein 3-like isoform X3 [*Populus euphratica*] | XP_011039002.1 | 1.00E-23 |
| 10809 | F: GAAGAGCTGAAAAGGCAACCT<br>R: TTCTTTTGCCCCTCAGCTT | $(TGTGG)_6$ | 196–231 | 60 | HEX | KT285032 | 22.7 kDa class IV heat shock protein [*Ricinus communis*] | XP_002521274.1 | 4.00E-06 |
| 10960 | F: GGCATCTCTTCTTATCCTCCC<br>R: CTGAAAGAAAAACGGAGCG | $(TCCTT)_5$ | 211–236 | 59 | NED | KT285033 | Uncharacterized protein LOC8265384 [*Ricinus communis*] | XP_002528323.1 | 0.072 |
| 10997 | F: TCAAGCATGGCAAAGGGT<br>R: TGCATGAACAAAGGTGCC | $(AGA)_7$ | 233–242 | 60 | NED | KT285034 | Probable BOI-related E3 ubiquitin-protein ligase 3 [*Jatropha curcas*] | XP_012081782.1 | 1.00E-32 |
| 16226 | F: TGGCATAAGAGTGCAACCA<br>R: TGATGATGTTGAAAACCTCCA | $(CAT)_7$ | 232–238 | 58 | 6-FAM | KT285035 | No hit | — | — |
| 16859 | F: CACAACACACACACCAACA<br>R: TTTGAAAATTTTGGAAACCCA | $(TAC)_7$ | 160–178 | 59 | HEX | KT285036 | No hit | — | — |
| 22194 | F: CCCTGTTCTGTTGGGTCG<br>R: GAAGAAGCAGTGCTGAGTGC | $(TTTTG)_6$ | 226–241 | 59 | HEX | KT285037 | Amino acid permease 6 [*Ricinus communis*] | XP_002510013.1 | 6.00E-05 |
| 23632 | F: GCGACCAGAGGCAGTGAA<br>R: TCTTCTGCCTCACGCATTT | $(AGTG)_6$ | 224–254 | 60 | HEX | KT285038 | Dynamin-related protein 3A isoform X1 [*Ricinus communis*] | XP_015572520.1 | 2.00E-13 |
| 24490 | F: AAGGGTAAGGGTGCCCAG<br>R: CAAGAGGCGTCATCCACC | $(TGGA)_7$ | 180–208 | 60 | 6-FAM | KT285039 | No hit | — | — |
| 25334 | F: CACAACTCCACCGCATCA<br>R: ACCGCTAGAACTCGCTGC | $(CCT)_9$ | 175–193 | 59 | 6-FAM | KT285040 | No hit | — | — |
| 25439 | F: TCACCGGATTGTTGACGA<br>R: CAGAAACCCCACCTAGAAGAA | $(TGAGGC)_7$ | 147–177 | 59 | HEX | KT285041 | No hit | — | — |
| 26221 | F: ATGGGGACATGATGGTGG<br>R: GCCTTTGTGTTCGTTGAGAGA | $(TGA)_{17}$ | 153–189 | 60 | NED | KT285042 | Transcription factor PIF7 isoform X2 [*Vitis vinifera*] | XP_010663294.1 | 2.00E-07 |
| 26474 | F: TGGCACCATCACCATCAC<br>R: CCCCTCAACTCAATCCATCA | $(AG)_{11}$ | 220–226 | 60 | HEX | KT285043 | Conserved hypothetical protein [*Ricinus communis*] | EEF49157.1 | 3.00E-12 |

*Note*: $T_a$ = annealing temperature.

Table 2. Genetic properties of the 18 novel polymorphic EST-SSR markers developed in four populations of *Speranskia tuberculata*.[a]

| | YA ($N = 6$) | | | XZ ($N = 6$) | | | YT ($N = 6$) | | | KQ ($N = 6$) | | | Total | Mean | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Locus | $A$ | $H_o$ | $H_e$ | $A$ | $H_o$ | $H_e$ | $A$ | $H_o$ | $H_e$ | $A$ | $H_o$ | $H_e$ | $A$ | $H_o$ | $H_e$ |
| 8852 | 4 | 0.500 | 0.708 | 3 | 0.167 | 0.292 | 1 | 0 | 0 | 4 | 0.333 | 0.597 | 4 | 0.250 | 0.626 |
| 9441 | 5 | 0.833 | 0.736 | 1 | 0 | 0 | 2 | 0 | 0.278 | 2 | 0.167 | 0.153 | 5 | 0.250 | 0.386 |
| 9832 | 2 | 0.167 | 0.486 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 2 | 0.042 | 0.187 |
| 10026 | 3 | 0.250 | 0.656 | 3 | 0.333 | 0.486 | 4 | 0.333 | 0.681 | 2 | 0.333 | 0.278 | 7 | 0.318 | 0.705 |
| 10117 | 1 | 0 | 0 | 3 | 0.333 | 0.500 | 2 | 0.167 | 0.153 | 1 | 0 | 0 | 3 | 0.125 | 0.192 |
| 10128 | 1 | 0 | 0 | 4 | 0.333 | 0.597 | 2 | 0 | 0.278 | 1 | 0 | 0 | 4 | 0.087 | 0.271 |
| 10809 | 2 | 0.833 | 0.486 | 4 | 0.500 | 0.597 | 3 | 0.333 | 0.292 | 2 | 0.167 | 0.153 | 5 | 0.458 | 0.556 |
| 10960 | 3 | 0.667 | 0.611 | 3 | 0.500 | 0.569 | 4 | 0.167 | 0.625 | 4 | 0.667 | 0.694 | 6 | 0.500 | 0.752 |
| 10997 | 3 | 0.500 | 0.569 | 2 | 0.167 | 0.375 | 4 | 0.667 | 0.653 | 3 | 0.500 | 0.542 | 4 | 0.458 | 0.588 |
| 16226 | 1 | 0 | 0 | 2 | 0 | 0.278 | 2 | 0.833 | 0.486 | 1 | 0 | 0 | 2 | 0.208 | 0.305 |
| 16859 | 4 | 0.500 | 0.514 | 3 | 0.333 | 0.292 | 3 | 0.667 | 0.569 | 2 | 0.167 | 0.375 | 5 | 0.417 | 0.531 |
| 22194 | 3 | 0.333 | 0.500 | 3 | 0.167 | 0.292 | 1 | 0 | 0 | 4 | 0.667 | 0.708 | 4 | 0.292 | 0.440 |
| 23632 | 5 | 0.667 | 0.681 | 2 | 0.167 | 0.486 | 2 | 0 | 0.320 | 2 | 0.400 | 0.480 | 6 | 0.318 | 0.675 |
| 24490 | 6 | 0.833 | 0.806 | 4 | 0.667 | 0.681 | 3 | 0.500 | 0.500 | 3 | 0.500 | 0.625 | 8 | 0.625 | 0.827 |
| 25334 | 2 | 0.167 | 0.375 | 4 | 0 | 0.667 | 4 | 0.833 | 0.653 | 3 | 0.667 | 0.500 | 7 | 0.417 | 0.806 |
| 25439 | 3 | 0.833 | 0.653 | 4 | 1.00 | 0.708 | 3 | 0.833 | 0.667 | 4 | 1.000 | 0.681 | 6 | 0.917 | 0.748 |
| 26221 | 4 | 0.500 | 0.653 | 5 | 0.333 | 0.750 | 7 | 0.833 | 0.819 | 4 | 0.500 | 0.625 | 11 | 0.542 | 0.826 |
| 26474 | 4 | 0.667 | 0.681 | 4 | 0.200 | 0.700 | 2 | 0 | 0.500 | 4 | 0.333 | 0.597 | 4 | 0.304 | 0.717 |

*Note*: $A$ = number of alleles per locus; $H_e$ = expected heterozygosity; $H_o$ = observed heterozygosity; $N$ = number of individuals sampled.
[a]Locality and voucher information are provided in Appendix 1.

amplifications with the same protocol. The labeled PCR products were analyzed on an ABI 3730 DNA Analyzer with a GeneScan 500 LIZ Size Standard (Applied Biosystems). Allele sizes were called using GeneMarker version 2.6.0 (SoftGenetics, State College, Pennsylvania, USA). Number of alleles per locus ($A$), observed heterozygosity ($H_o$), and expected heterozygosity ($H_e$) were calculated using GenAlEx version 6.2 (Peakall and Smouse, 2006).

Eighteen of the 20 candidate markers showed polymorphisms among the four populations of *S. tuberculata*. The corresponding sequences of these markers were deposited in GenBank (Table 1). The number of alleles per locus ranged from two to 11, $H_e$ ranged from 0.187 to 0.827, and $H_o$ ranged from 0.042 to 0.917 (Table 2).

Cross-species amplification of the 18 newly developed polymorphic markers was tested in 24 *S. cantonensis* individuals from a single population (Ruyuan, Guangdong; Appendix 1), using the same procedures described above. Thirteen loci (72.22%) were successfully amplified in all *S. cantonensis* individuals tested, of which six showed polymorphisms (Table 3).

## CONCLUSIONS

These 18 novel polymorphic SSR markers will be used to evaluate impacts of recent habitat fragmentation on the genetic diversity and structure of *S. tuberculata*, and to develop suitable conservation strategies for the species. Of these SSR markers developed in *S. tuberculata*, 13 were successfully amplified in single population samples of the related species *S. cantonensis*, extending their potential usefulness for future research in the genus *Speranskia* (e.g., comparisons of genetic diversity).

Table 3. Polymorphisms at the 13 successfully cross-amplified EST-SSR markers in single population samples of *Speranskia cantonensis* ($N = 24$).[a]

| Locus | $A$ | $H_o$ | $H_e$ | GenBank accession no.[b] |
|---|---|---|---|---|
| 8852 | 1 | 0 | 0 | KT312943 |
| 9441 | 4 | 0.042 | 0.666 | KT312944 |
| 9832 | 1 | 0 | 0 | KT312945 |
| 10026 | 1 | 0 | 0 | KT312946 |
| 10117 | 2 | 0 | 0.375 | KT312947 |
| 10128 | 1 | 0 | 0 | KT312948 |
| 10960 | 1 | 0 | 0 | KT312949 |
| 10997 | 1 | 0 | 0 | KT312950 |
| 16226 | 1 | 0 | 0 | KT312951 |
| 23632 | 4 | 0 | 0.663 | KT312952 |
| 24490 | 2 | 0 | 0.500 | KT312953 |
| 25334 | 3 | 0.250 | 0.288 | KT312954 |
| 25439 | 3 | 0 | 0.625 | KT312955 |

*Note*: $A$ = number of alleles per locus; $H_e$ = expected heterozygosity; $H_o$ = observed heterozygosity; $N$ = number of individuals sampled.
[a]Locality and voucher information are provided in Appendix 1.
[b]GenBank accession numbers are for the cross-amplified markers in *Speranskia cantonensis*.

## LITERATURE CITED

Fu, L. M., B. F. Niu, Z. W. Zhu, and W. 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.

Hwang, S. M. 1989. A notes on genera *Speranskia* in China (Euphorbiaceae). *Bulletin of Botanical Research* 9: 37–40.

Ju, M. M., H. C. Ma, P. Y. Xin, Z. L. Zhou, and B. Tian. 2015. Development and characterization of EST-SSR markers in *Bombax ceiba* (Malvaceae). *Applications in Plant Sciences* 3: 1500001.

Mazzio, E., R. Badisa, N. Mack, S. Deiab, and K. F. A. Soliman. 2014. High throughput screening of natural products for anti-mitotic effects in MDA-MB-231 human breast carcinoma cells. *Phytotherapy Research* 28: 856–867.

Peakall, R., and P. E. Smouse. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

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.

Thiel, T., W. Michalek, R. K. Varshney, and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 10: 411–422.

Appendix 1. Locality information for the sampled populations of *Speranskia tuberculata* and *S. cantonensis* tested in this study.

| Population | Species | Collection locality | *N* | Geographic coordinates | Altitude (m) | Voucher no.[a] |
|---|---|---|---|---|---|---|
| YA | *S. tuberculata* (Bunge) Baill. | Yan'an, Shaanxi | 6 | 36°35′N, 109°29′E | 1061 | TB2014087 |
| XZ | *S. tuberculata* | Xinzhou, Shanxi | 6 | 39°19′N, 113°34′E | 1160 | TB2014117 |
| YT | *S. tuberculata* | Yantai, Shandong | 6 | 37°17′N, 121°44′E | 120 | TWYT02 |
| KQ | *S. tuberculata* | Chifeng, Inner Mongolia | 6 | 42°57′N, 118°59′E | 631 | TB2013153 |
| RU | *S. cantonensis* (Hance) Pax & K. Hoffm. | Ruyuan, Guangdong | 24 | 24°59′N, 113°08′E | 650 | FL2014098 |

*Note*: *N* = number of individuals.
[a]Voucher specimens deposited at the Herbarium of Southwest Forestry University (SWFC), Kunming, China.