Development of novel EST-SSR markers for *Rhododendron longipedicellatum* (Ericaceae) and cross-amplification in two congeners

Taiqiang Li1*, Xiongfang Liu1*, Zhenghong Li1, Youming Wan1, Xiuxian Liu1, and Hong Ma1,2

**PREMISE OF THE STUDY:** To investigate the genetic background and population characteristics of *Rhododendron longipedicellatum* (Ericaceae), a newly discovered and critically endangered species, expressed sequence tag–simple sequence repeat markers were developed, and transferability was tested in two congeners, *R. molle* and *R. simsii*.

**METHODS AND RESULTS:** Based on the transcriptome sequences of *R. longipedicellatum*, 102 primer sets were designed; 48 primer sets were successfully amplified, with 15 showing polymorphisms in 150 individuals from five extant populations of *R. longipedicellatum*. The number of alleles per locus ranged from four to 18, and the levels of observed and expected heterozygosity for the 15 loci varied from 0.255 to 0.913 and from 0.306 to 0.851, respectively. All 15 loci were found to amplify in *R. molle* and *R. simsii*.

**CONCLUSIONS:** These polymorphic SSR markers can be used in conservation genetic and phylogeographic studies to elucidate the rarity and origin of *R. longipedicellatum*.

**KEY WORDS** conservation genetics; Ericaceae; EST-SSR marker; *Rhododendron longipedicellatum*; transcriptome; transferability.

*Rhododendron* L., renowned for its horticultural and ecological value, is the largest genus in Ericaceae and is one of the most widespread woody plants in the Northern Hemisphere. Its more than 1025 species are distributed throughout Asia, Europe, and North America, and two species extend to eastern Greenland and Queensland, Australia (Chamberlain et al., 1996; Fang et al., 2005; Cai et al., 2016). Wild *Rhododendron* species serve as potential genetic resources for the development of new cultivars, and more than 25,000 *Rhododendron* cultivars have been bred around the world. However, no evergreen rhododendron has a yellow-flowered cultivar (Ureshino et al., 2016). At present, the breeding of flower color in rhododendrons tends to favor pure-colored flowers internationally, especially pure yellow rhododendrons (Lan et al., 2012).

*R. longipedicellatum* Lei Cai & Y. P. Ma (subg. *Rhododendron*, sect. *Vireya*, subsect. *Pseudovireya*) is an unusual evergreen shrub, with brilliantly pure yellow flowers having no blotches or spots. Unlike all other wild *Rhododendron* species, whose flowering times occur between March and June in the Northern Hemisphere, the natural flowering time of *R. longipedicellatum* extends from the last 10-day period of November to the first 10-day period of February. However, *R. longipedicellatum* has a very limited distribution, with only five relict populations found in Malipo County, Yunnan Province, China, and with the largest population comprising about 350 mature plants (Cai et al., 2016). Furthermore, this species is at risk of extinction because of continued disturbance from anthropogenic activities. Therefore, genetic information from *R. longipedicellatum* is urgently needed for current and future conservation activities.

Expressed sequence tag–simple sequence repeat (EST-SSR) markers are increasingly used in population genetic studies because they are codominant, multiallelic, and often highly polymorphic, and they are less susceptible to null alleles and homoplasy than anonymous SSRs are (Ellis and Burke, 2007; Yoichi et al., 2016). However, currently, only 77 EST-SSR markers have been developed in *Rhododendron* (Yoichi et al., 2016; Xing et al., 2017), and only two markers (Rhob_1022 and Rhob_30843, percentage of polymorphic loci = 2.60%; Table 1) are amplified for *R. longipedicellatum*, which is insufficient for unraveling the population dynamics of most species within this genus. Therefore, we developed 15 EST-SSR markers for *R. longipedicellatum* and evaluated their polymorphism and transferability to *R. molle* (Blume) G. Don (an important congeneric species with yellow flowers) and *R. simsii* Planch. (a widespread *Rhododendron* species that occurs with populations of *R. longipedicellatum*). These markers will provide an important genetic resource for rhododendron breeding programs worldwide.
The EST-SSR primers were initially screened for performance with two individuals from each of the five relict R. longipedicellatum wild populations (WBL, WJL, XCW, XL, and ZWL; Appendix 1). Genomic DNA was extracted from silica-dried leaves with a modified cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987). PCR amplifications were performed in a final 10-μL volume, containing 1 μL (10–30 ng) of template DNA, 5 μL of 0.7× Multiplex PCR Master Mix (QIAGEN, Hilden, Germany), 0.5 μL (10 pm) of each primer, and 3 μL of RNase-free water. The PCR thermal profile consisted of an initial denaturation step at 95°C for 5 min; followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 56–59°C for 30 s (Table 2), and elongation at 72°C for 1 min; with a final elongation step at 72°C for 10 min. All PCR products were resolved by electrophoresis in 1% agarose gels to determine whether amplification was successful. Of the 102 primer pairs, 48 (47.1%) target regions were successfully amplified.

PCR fluorescent tagging was performed for further polymorphic screening. The 5’ end of each forward primer for the 48 markers was tagged with one of three fluorescent dyes (FAM, HEX, or ROX [Thermo Fisher Scientific]; Table 2), and multiplex PCR amplifications were performed for the 150 individuals of R. longipedicellatum, representing all extant populations (30 for each population), using the PCR conditions described above. Allele size for the tagged PCR products was obtained using an ABI 3730 sequencer with a GeneScan 500 LIZ Size Standard (Thermo Fisher Scientific) and GeneMapper 4.1 (Thermo Fisher Scientific). Population genetic parameters, including the number of alleles per locus, expected heterozygosity, observed heterozygosity, and deviation from Hardy–Weinberg equilibrium, were analyzed with GENEPOP software (version 3.4; Raymond and Rousset, 1995). Pairwise linkage disequilibrium in each population was tested with FSTAT software (version 2.9.3; Goudet, 1995). The Brookfield Methods

Three mature plants of R. longipedicellatum were collected from population ZWL (voucher specimen accession no. LTQ20160618; Appendix 1) and planted in a greenhouse at the Research Institute of Resources Insects, Chinese Academy of Forestry (Kunming, China). Fresh, tender leaves from the mature plant were gathered 1 y later and mixed in equal proportions for RNA extraction and transcriptome sequencing. Total RNA was extracted with Trizol (Thermo Fisher Scientific, Waltham, Massachusetts, USA) followed by the manufacturer’s instructions. The cDNA library construction and sequencing was performed by staff at the Beijing Genome Institute (Wuhan, China) with a HiSeq 4000 (Illumina, San Diego, California, USA). Altogether, 58.30 Mbp raw reads were obtained and deposited into the National Center for Biotechnology Information (NCBI) sequence read archive (SRA) database (Bioproject ID: SRR6509877). The generated raw reads were filtered to remove reads containing adapters, ambiguous reads (N > 5%), and other low-quality reads (base quality <15% or >20%), and a total of 44.85 Mbp clean reads were obtained and assembled de novo into 94,906 contigs using Trinity software (Grabherr et al., 2011). TGICL software (Perta et al., 2003) was used to cluster similar contigs, which generated 74,092 nonredundant unigenes, with an average length of 938 bp. MISA software (Thiel et al., 2003) was used for SSR motif mining from all unigenes, and the minimum numbers of repeats were set as six, five, four, and four for di-, tri-, tetra-, penta-, and hexanucleotide motifs, respectively. Altogether, 20,304 SSR motifs were found, and 102 of them were selected with at least five tri- and tetranucleotide repeats, 10 dinucleotide repeats, or four penta- and hexanucleotide repeats for primer design in Primer3 software (Rozen and Skaltsky, 1999), with conditions as described by Li et al. (2011).
| Locus | Primer sequences (5′–3′) | Repeat motif | Allele size range (bp) | T<sub>a</sub> (°C) | Fluorescent dye<sup>a</sup> | BLAST top hit description [organism] | BLAST top hit accession no. | E-value | GenBank accession no. |
|-------|--------------------------|--------------|-----------------------|-------------------|-----------------------------|-------------------------------------|-----------------------------|--------|---------------------|
| RL6   | F: GAGCTCTACAAGTTAATATTCCCG<sup>b</sup> R: ATCATCACCACACCTCTACCCG | (AGC)<sub>5</sub> | 135–161 | 58 | FAM<sup>c</sup> | 5-methyltetrahydrofolate-homocysteine methyltransferase reductase mRNA [Monoraphidium neglectum] | XM_014050907.1 | 5.00E−06 | MG585326 |
| RL6   | F: GATGCTCTTCTCTCGGATACC<sup>b</sup> R: GAAGAATATAATCTGCTGGGATACC | (GAQ)<sub>8</sub> | 94–114 | 59 | FAM<sup>d</sup> | No hit | — | — | MG585327 |
| RL16  | F: AGGAGCAAGGTATAAAAGCAGC<sup>b</sup> R: GGGTTCTTTGTTCTCTTCTGACC | (AGG)<sub>6</sub> | 85–95 | 59 | FAM<sup>c</sup> | Uncharacterized LOC100249879 [LOC100249879] [Vitis vinifera] | XM_010648600.1 | 7.00E−11 | MG585328 |
| RL20  | F: CATGTAGGGGCTACTCC<sup>b</sup> R: TGATCCGGGACTTGAATCC ACC | (CCT)<sub>2</sub> | 100–106 | 58 | FAM<sup>c</sup> | CCMP1545 predicted protein [Micromonas pusilla] | XM_003062022.1 | 5.00E−09 | MG585329 |
| RL26  | F: AGATGAACTCCAGTTAAGGG<sup>b</sup> R: CTCTCTCTGTCTTTATAGGT TCTG | (ATG)<sub>8</sub> | 89–105 | 58 | FAM<sup>c</sup> | No hit | — | — | MG585330 |
| RL28  | F: CATGTAGGTTAAAAGGATGG<sup>b</sup> R: TCAGGACTCCAGTTGACATCC | (GTG)<sub>4</sub> | 91–122 | 57 | HEX<sup>c</sup> | Dof zinc finger protein DOF2.4-like (LOC101000103) [Beta vulgaris subsp. vulgaris] | XM_010648746.1 | 4.00E−06 | MG585331 |
| RL37  | F: CCAGTCAGCCGACTCTGCTG<sup>b</sup> R: CTCTCTCTCTCTCTCTCCAGGCC | (CAG)<sub>8</sub> | 90–122 | 58 | FAM<sup>c</sup> | AP2-like ethylene-responsive transcription factor ANT (LOC104591238) [Nelumbo nucifera] | XM_010250039.1 | 2.00E−08 | MG585332 |
| RL54  | F: CGACTCTACTAATAACAGAGG<sup>b</sup> R: CTCCTCTGGAAGAGCTCTAAGG | (GA)<sub>10</sub> | 84–102 | 57 | HEX<sup>c</sup> | Vitellogenin-1 (LOI 04803668) [Tremaya hassleriana] | XM_010527700.1 | 1.00E−15 | MG585333 |
| RL61  | F: GTAGGTCGATGGTAATCATTCTGCTG<sup>b</sup> R: CTCTCTCTCTCTCTCAGGCC | (CT)<sub>10</sub> | 96–138 | 59 | HEX<sup>c</sup> | Uncharacterized LOC101310701 (LOC101310701) [Fragaria vesca subsp. vesca] | XM_002949009.2 | 7.00E−08 | MG585334 |
| RL74  | F: GTCCTGACTTCTGTACACAG<sup>b</sup> R: GTATGAGATCTAGGGCATCGG | (ATC T)<sub>6</sub> | 110–133 | 56 | ROX<sup>d</sup> | No hit | — | — | MG585336 |
| RL89  | F: GCCTGGTCTCTTGATGACA<sup>b</sup> R: CTCTCTCTCTCTCTCTCTCTCAGC | (ATCT)<sub>10</sub> | 102–137 | 57 | ROX<sup>d</sup> | Cone Tongling01–10 microsatellite sequence [Lycoris radiata] | XP665168.1 | 8.00E−08 | MG585337 |
| RL98  | F: CCTCTTTCTCCCTTAATCC<sup>b</sup> R: AGCAGATGTTCTTTCTTCTCCGCC | (AAACA)<sub>4</sub> | 88–103 | 57 | ROX<sup>d</sup> | Uncharacterized LOC106446472 [LOC106446472] [Brassica napus] | XM_013888210.1 | 4.00E−06 | MG585338 |
| RL99  | F: CCCTCTTTCTCCCTTAATCC<sup>b</sup> R: AGCAGATGTTCTTTCTTCTCCGCC | (AAACA)<sub>4</sub> | 85–104 | 57 | ROX<sup>d</sup> | Uncharacterized LOC106446472 [LOC106446472] [Brassica napus] | XM_013888210.1 | 4.00E−06 | MG585339 |
| RL100 | F: CCCTGTGGGAGGTTGTTACC R: CCCATTTAACAATCTCAACACCC | (GTAGG)<sub>2</sub> | 88–112 | 59 | ROX<sup>d</sup> | Oryza sativa Japonica group protein Chromatin remodeling 24 (LOC100646472) transcript variant X2 [Oryza sativa] | XM_015781575.1 | 4.00E−09 | MG585340 |

<sup>a</sup>PCR multiplex sets are indicated as 1, 2, or 3.

<sup>b</sup>Primers are listed in 5′–3′ direction.

<sup>c</sup>Note: T<sub>a</sub> = annealing temperature.

<sup>d</sup>Fluorescent dyes are indicated as FAM, HEX, ROX, and ROX2.
Rhododendron longipedicellatum (Ericaceae) exhibit mono-allelic amplification, and only three loci in the 15 newly developed, polymorphic markers were tested in 30 individual samples from Kunming Botanical Garden, WJL, XCW, XL, and ZWL populations of 6162 Li et al. — Rhododendron longipedicellatum EST-SSRs

ACKNOWLEDGMENTS

This work was supported by the Technology Innovation Talent Project of Yunnan Province (2016HB007).

LITERATURE CITED

Brookfield, J. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology 5: 453–455.

Cai, L., J. Nilsen, Z. L. Dao, and Y. P. Ma. 2016. Rhododendron longipedicellatum (Ericaceae), a new species from southeastern Yunnan, China. Phytotaxa 282: 296–300.

Chamberlain, D. F., R. Hyam, G. Argent, G. Fairweather, and K. S. Walter. 1996. The genus Rhododendron: Its classification and synonymy. Royal Botanic Garden, Edinburgh, United Kingdom.

Dove, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11–15.

Ellis, J. R., and J. M. Burke. 2007. EST-SSRs as a resource for population genetic analyses. Heredity 99: 125–132.

Fang, M. Y., R. Z. Fang, M. Y. He, L. Z. Hu, H. B. Yang, H. N. Qin, and T. L. Min. 2005. Ericaceae. In Z. K. Wu, P. H. Raven, and D. Y. Hong [eds.], Flora of China, vol. 14, 260–455. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Goudet, J. 1995. FSTAT (version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86: 485–486.

Grabherr, M. G., B. J. Haas, M. Sultana, S. Karamycheva, Y. Antonescu, R. Z. Pertea, G. M. Pertea, G. M., X. Q. Huang, F. Liang, V. V. Varshney, and A. Graner. Thiel, T., W. Michalek, R. K. Varshney, and A. Graner. 2003. Exploiting EST data for biologist programmers. Bioinformatics (Oxford, England) 19: 651–652.

Li, L. F., D. X. Yin, N. Song, E. H. Tang, and H. X. Xiao. 2011. Genomic and EST microsatellites for Rhododendron aureum (Ericaceae) and cross-amplification in other congeneric species. American Journal of Botany 98: e250–252.

Li, L. T., X. F. Liu, Z. H. Li, H. Ma, Y. M. Wan, X. X. Liu, and L. Y. Fu. 2018. Study on reproductive biology of Rhododendron longipedicellatum: A newly discovered and special threatened plant surviving in limestone habitat in Southeast Yunnan, China. Frontiers in Plant Science 9: 33.

Li, L. F., D. X. Yin, N. Song, E. H. Tang, and H. X. Xiao. 2011. Genomic and EST microsatellites for Rhododendron aureum (Ericaceae) and cross-amplification in other congeneric species. American Journal of Botany 98: e250–252.

Li, T. Q., X. F. Liu, Z. H. Li, H. Ma, Y. M. Wan, X. X. Liu, and L. Y. Fu. 2018. Study on reproductive biology of Rhododendron longipedicellatum: A newly discovered and special threatened plant surviving in limestone habitat in Southeast Yunnan, China. Frontiers in Plant Science 9: 33.

Pertea, G., X. Q. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y. Lee, et al. 2003. TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. Bioinformatics (Oxford, England) 19: 651–652.

Raymond, M., and F. Roussset. 1995. GENEPOP version 1.2: Population genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248–249.

Rozen, S., and H. Skaltsky. 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 106: 411–422.

Ureshino, K., M. Nakayama, and I. Miyajima. 2016. Contribution made by and characterization of EST-SSR markers for the genus Rhododendron section Brachyalyss (Ericaceae). Plant Species Biology 32: 455–459.

We developed 15 highly informative EST-SSR markers for R. longipedicellatum, which can be used in population genetic diversity, genetic structure, and phylogeographic studies to facilitate development of scientific conservation measures in R. longipedicellatum. The markers may also be valuable for population and evolutionary studies of congeneric species and closely related taxa.
APPENDIX 1. Locality and voucher information for the *Rhododendron* species and populations used in this study.

| Species       | Population code | N  | Collection locality | Geographic coordinates          | Altitude (m) | Voucher no.  |
|---------------|-----------------|----|---------------------|---------------------------------|--------------|--------------|
| *R. longipedicellatum* Lei Cai & Y. P. Ma | WBL 30 | Malipo, Yunnan | 23°09′33.7″N, 104°56′48.7″E | 1312         | MH20150614   |
|               | WJL 30          | Malipo, Yunnan | 23°09′52.4″N, 104°56′34.8″E | 1316         | LTX20161205  |
|               | XCW 30          | Malipo, Yunnan | 23°09′47.3″N, 104°56′45.1″E | 1248         | MH20141124   |
|               | XL 30           | Malipo, Yunnan | 23°10′1.9″N, 104°56′51.1″E | 1183         | DZL3637-1    |
|               | ZWL 30          | Malipo, Yunnan | 23°09′59.8″N, 104°56′22.0″E | 1270         | LTX20160618  |
| *R. molle* (Blume) G. Don | YZC 30 | Kunming, Yunnan | 25°08′24.6″N, 102°44′27.9″E | 1953         | LXF20170322  |
| *R. simsii* Planch. | YSH 30 | Kunming, Yunnan | 25°08′25.0″N, 102°44′31.6″E | 1951         | LXF20170615  |

Note: N = number of individuals sampled.  
*C* Collection localities in China.  
*V* Voucher specimens are deposited at the Herbarium of the Kunming Institute of Botany, Chinese Academy of Sciences (KUN), Kunming, China.