Development and Characterization of EST-SSR Markers for Begonia luzhaiensis (Begoniaceae)

Authors: Tseng, Yu-Hsin, Huang, Han-Yau, Xu, Wei-Bin, Yang, Hsun-An, Liu, Yan, et al.

Source: Applications in Plant Sciences, 5(5)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1700024

Applications in Plant Sciences 2017 5(5): 1700024

PRIMER NOTE

DEVELOPMENT AND CHARACTERIZATION OF EST-SSR MARKERS FOR *BEGONIA LUZHAIENSIS* (BEGONIACEAE)[1]

YU-HSIN TSENG[2], HAN-YAU HUANG[2,3], WEI-BIN XU[4], HSUN-AN YANG[2], YAN LIU[4], CHING-I PENG[2], AND KUO-FANG CHUNG[2,3,5]

[2]Research Museum and Herbarium (HAST), Biodiversity Research Center, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11529, Taiwan; [3]School of Forestry and Resource Conservation, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan; and [4]Guangxi Key Laboratory of Plant Conservation and Restoration Ecology in Karst Terrain, Guangxi Institute of Botany, Guangxi Zhuangzu Autonomous Region, and Chinese Academy of Sciences, Guilin 541006, People's Republic of China

- *Premise of the study:* Microsatellite primers were developed for *Begonia luzhaiensis* (Begoniaceae) to assess genetic diversity and population genetic structure.
- *Methods and Results:* Based on the transcriptome data of *B. luzhaiensis*, 60 primer pairs were selected for initial validation, of which 16 yielded polymorphic microsatellite loci in 57 individuals. The number of alleles observed for these 16 loci ranged from one to nine. The observed and expected heterozygosity ranged from 0.000 to 1.000 and from 0.000 to 0.804 with averages of 0.370 and 0.404, respectively. Five loci could be successfully amplified in *B. leprosa*.
- *Conclusions:* The expressed sequence tag–simple sequence repeat markers are the first specifically developed for *B. luzhaiensis* and the first developed in *Begonia* sect. *Coelocentrum.* These markers will be useful for future studies of the genetic structure and phylogeography of *B. luzhaiensis*.

Key words: *Begonia leprosa*; *Begonia luzhaiensis*; Begoniaceae; expressed sequence tag–simple sequence repeat (EST-SSR); section *Coelocentrum*; transcriptome.

The immense terrain of limestone karsts stretching from southern China to northern Vietnam (Sino-Vietnamese limestone karsts; SVLK) is a renowned biodiversity hotspot. It is noted for extremely high species diversity and endemicity of narrowly distributed calciphilous plants that are increasingly threatened by habitat destruction caused by rapid economic growth in recent decades (Chung et al., 2014). To design effective conservation strategies, a better understanding of the evolutionary mechanisms underlying the rich and distinct SVLK flora is essential. With ca. 70 species distributed exclusively in the SVLK, *Begonia* L. sect. *Coelocentrum* Irmsch. (Begoniaceae) is one of the most characteristic plant groups (Chung et al., 2014; Peng et al., 2014, 2015; Li et al., 2016), presenting an ideal model system for studying limestone plant speciation. Species of the section exhibit great morphological variation in leaf shape, texture, and variegation; they are usually confined to caves and cave-like microhabitats that occur abundantly in the karst areas of southern China (Guangdong, Guangxi, and Yunnan provinces) and northern Vietnam (Chung et al., 2014). Most species of sect. *Coelocentrum* are known from a single or a few localities, with the exception of *B. cavaleriei* H. Lév., *B. leprosa* Hance, and *B. luzhaiensis* T. C. Ku (Gu et al., 2007).

Based on phylogenetic analyses of Asian limestone *Begonia* species, Chung et al. (2014) proposed that the excess of moisture brought by the onset of East Asian monsoons since the late Miocene accelerated rates of karstification of Sino-Vietnamese limestone terrains, triggering widespread allopatric speciation in the SVLK. To test the microevolutionary processes underlying the speciation hypothesis of Chung et al. (2014), we investigate the population genetics and phylogeography of *B. luzhaiensis*. In this article, we report microsatellite markers developed for *B. luzhaiensis*. Expressed sequenced tag–simple sequence repeat (EST-SSR) markers are valuable in tests of cross-transferability, facilitating studies of population genetic diversity in many plant species (e.g., Dikshit et al., 2015; Zhou et al., 2016). Here, we used next-generation transcriptome sequencing to develop a set of microsatellite markers for *B. luzhaiensis.* Additionally, we tested the transferability of these markers for *B. leprosa*, another widespread species of sect. *Coelocentrum*.

METHODS AND RESULTS

To maximize potential loci, total RNAs were extracted from fresh leaves and male floral buds (from *C.-I Peng 18732*, denoted as Pool AC) and fruit (*C.-I Peng 18735*, Pool B) of *B. luzhaiensis*, respectively. RNA extraction was performed using the PureLink RNA Mini Kit (Invitrogen, Carlsbad, California, USA) according to the manufacturer's instructions, and quality and quantity were measured by an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa

[1]Manuscript received 20 March 2017; revision accepted 5 April 2017.

This work was supported by grants from the Biodiversity Research Center, Academia Sinica, Taiwan, and the Ministry of Science and Technology, Taiwan (NSC 101-2621-B-001-003) to C.I.P. and K.F.C.

[5]Author for correspondence: bochung@gate.sinica.edu.tw

doi:10.3732/apps.1700024

TABLE 1. Characteristics of 16 microsatellite loci developed for *Begonia luzhaiensis*.

Locus[a]	Primer sequences (5′–3′)	Repeat motif	Fluorescent label	Allele size range (bp)	Putative function [Organism]	*E*-value	GenBank accession no.
BLZ01	F: TCGGTTGAGCTGCTCTTTTT	$(GA)_{20}$	FAM	188–228	No hit	–	KY659210
	R: CCGCATGAATGCCTCTTATT						
BLZ02	F: AGCTCCCTCTTCCATCTTCC	$(TCT)_{11}$	FAM	164–200	No hit	–	KY659211
	R: CGTGTCCTCCTTCCCTACAA						
BLZ03	F: GACGAACGAAGGGTGACTGT	$(CGAAC)_5$	NED	205–240	No hit	–	KY659212
	R: CGGCTTCAGATTACCAGGAC						
BLZ04	F: TGTCACCGTGGAAGAAGATG	$(AGGGAG)_5$	NED	163–217	Cyclin-L1-1 isoform X1 [*Lupinus angustifolius*]	3.7	KY659213
	R: CTCTCGCTCTCGTTCCTTGT						
BLZ05	F: TGAAGCCAGAGCATGAGATG	$(GATGGA)_5$	PET	209–257	Transcription factor MYB114 isoform X2 [*Eucalyptus grandis*]	0.004	KY659214
	R: ACTGGAACACCCATTTACGC						
BLZ06	F: AGCATCAGAGGAAAGCTTCG	$(TTC)_{12}$	FAM	135–153	Mlo-related protein [*Corchorus capsularis*]	8e-08	KY659215
	R: CCTCCATCCTCAATGGAAAA						
BLZ07	F: GCAAGGAGTTGCAGAGGAAG	$(CAG)_{11}$	VIC	200–248	No hit	–	KY659216
	R: CCAAAACCATTGCCCAGTAG						
BLZ08	F: TTTTGGAGAAGACGACCCTTT	$(GAAAG)_5$	NED	182–192	Hypothetical protein [*Gossypium raimondii*]	0.17	KY659217
	R: AGCGCAAGGCATTCATATTC						
BLZ09	F: GATCTGGCAAGGTTTTGGAA	$(CCCATT)_6$	PET	185–227	Hypothetical protein [*Citrus clementina*]	3e-12	KY659218
	R: GTGGCTGATCTTGCAGCATA						
BLZ10	F: AAAATGGCTAAGACGTGGAAA	$(GAA)_{10}$	VIC	150–210	No hit	–	KY659219
	R: TTTGGTTGCGAGAAGTTGC						
BLZ11	F: ACGTGCGACTCTGGAAAACT	$(AAAT)_5$	FAM	226–242	1,4-dihydroxy-2-naphthoyl-CoA thioesterase 1 [*Malus domestica*]	2e-11	KY659220
	R: GGCTTTCAATTCCTCGAAAA						
BLZ12	F: AAATGGGGAAGAACGGGTTA	$(TTTG)_5$	NED	151–199	Uncharacterized protein [*Vitis vinifera*]	0.46	KY659221
	R: GTATAAGCCCAGCTGCCAAG						
BLZ13	F: TCCAAGTTCAGGACCCAAAA	$(TTCTCC)_6$	PET	186–222	E3 ubiquitin-protein ligase UPL1 isoform X3 [*Cucumis melo*]	9e-13	KY659222
	R: TCTGTGCGTGCATTTCTAGC						
BLZ14	F: ATCCCATGGCTATCAGCATT	$(GACCGA)_5$	VIC	197–227	No hit	–	KY659223
	R: GGTATGGCGAGACCTAGTGC						
BLZ15	F: AGCAACACCCAGACTTCCAC	$(GAGATG)_5$	FAM	210–228	Hypothetical protein [*Gossypium raimondii*]	1e-13	KY659224
	R: CGCCTGAGTCATCGTTTTCT						
BLZ16	F: GACGTGATAAGGCCACAACC	$(TCTCCA)_5$	NED	223–259	No hit	–	KY659225
	R: CTTCTCCTCCACCTTCATCG						

[a] Annealing temperature was 53°C for all loci.

Clara, California, USA). Illumina TruSeq library preparation and sequencing using the Illumina MiSeq system (2 × 250 bp paired-end; Illumina, San Diego, California, USA) were performed by Tri-I Biotech (New Taipei City, Taiwan). Reads were de-multiplexed, quality-trimmed, and assembled using SOAPdenovo2 (Luo et al., 2012). All sequence information has been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA; PRJNA378679). Our assembly generated 40,226 and 28,454 contigs for the Pool AC and Pool B assemblies. The average contig length was 874 and 850 bp for Pool AC and Pool B, respectively. In combined *B. luzhaiensis* assemblies (combined Pool AC and B), 44,368 contigs with an average length of 864 bp were generated.

The program Simple Sequence Repeat Identification Tool (SSRIT; Temnykh et al., 2001; http://archive.gramene.org/db/markers/ssrtool) was used to identify sequences containing at least five di-, tri-, tetra-, penta-, and hexanucleotide microsatellite repeats. In combined *B. luzhaiensis* assemblies, 5602 microsatellite-containing sequences were identified, of which 60 potential loci were selected for primer design using Primer-BLAST (Ye et al., 2012), with the optimum conditions set at a length of 20 bp (18–22 bp), a melting temperature of 60°C (57–63°C), and a product size range of 120–250 bp.

To characterize the degree of polymorphism of each locus, 57 individuals from three populations were genotyped using the 60 newly designed primer pairs (Appendix 1). Total genomic DNA was extracted from silica gel–dried leaves based on protocols outlined in Chung et al. (2014). The PCR reaction was conducted with a final volume of 20 µL containing approximately 30 ng of genomic DNA, 1 µL of 10 µM of each primer, and 10 µL of 2× Master Mix Red (Ampliqon, Odense, Demark). The following PCR conditions were used: an initial denaturation of 94°C for 5 min; 32 cycles of 95°C for 40 s, 53°C for 35 s, and 72°C for 1 min; followed by an extension of 5 min at 72°C. The amplified products were analyzed on an ABI 3500 Genetic Analyzer (Applied Biosystems, Waltham, Massachusetts, USA) with GeneScan 600 LIZ Size Standard (Applied Biosystems). Genotypes were determined using GeneMarker version 3.7 (Holland and Parson, 2011).

Of the 60 primer pairs, 16 loci were polymorphic among the three tested populations (Table 1). The number of alleles per locus, expected heterozygosity, and observed heterozygosity were calculated with GenAlEx 6.503 (Peakall and Smouse, 2012). GENEPOP 4.2 (Raymond and Rousset, 1995) was used to perform exact tests of Hardy–Weinberg equilibrium and linkage disequilibrium. The total number of alleles ranged from one to nine with a mean of 3.379 (Table 2). The observed and expected heterozygosity ranged from 0.000 to 1.000 and from 0.000 to 0.804 with averages of 0.370 and 0.404, respectively. Significant deviations of Hardy–Weinberg equilibrium in terms of heterozygosity deficiency were detected at three loci (BLZ01, BLZ06, BLZ16) in the Luzhai population (LZ, Table 2). Significant linkage disequilibrium was not detected between any pair of loci ($P < 0.001$). The putative functions of SSR-associated sequences were determined by BLASTX against the nonredundant GenBank database.

The interspecific transferability of the 16 markers was evaluated in *B. leprosa*, another widespread species of sect. *Coelocentrum* (Table 3). Five markers were successfully cross-amplified in *B. leprosa*.

Table 2. Genetic characterization of 16 newly developed polymorphic microsatellites of *Begonia luzhaiensis*.[a]

	YS (n = 15)			LZ (n = 20)			XC (n = 22)		
Locus	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e
BLZ01	8	0.667	0.784	9	0.235	0.804*	2	0.048	0.210
BLZ02	4	0.600	0.604	5	0.895	0.733	6	0.773	0.732
BLZ03	4	0.467	0.647	3	0.167	0.531	3	0.167	0.329
BLZ04	5	0.400	0.498	5	1.000	0.722	2	0.455	0.351
BLZ05	3	0.067	0.135	2	0.067	0.064	3	0.190	0.177
BLZ06	4	0.800	0.604	3	0.000	0.304*	2	0.000	0.124
BLZ07	7	0.786	0.661	4	0.667	0.722	3	0.143	0.217
BLZ08	2	0.133	0.444	2	0.556	0.475	2	0.273	0.236
BLZ09	5	0.857	0.712	4	0.778	0.660	4	0.417	0.576
BLZ10	5	0.643	0.663	7	0.737	0.742	2	0.333	0.337
BLZ11	4	0.214	0.258	1	0.000	0.000	1	0.000	0.000
BLZ12	2	0.000	0.124	3	0.500	0.615	2	0.091	0.087
BLZ13	2	0.286	0.245	2	0.050	0.049	2	0.632	0.465
BLZ14	3	0.467	0.380	3	0.684	0.655	4	0.619	0.652
BLZ15	3	0.533	0.518	1	0.000	0.000	2	0.273	0.236
BLZ16	2	0.067	0.064	3	0.000	0.204*	2	0.050	0.049

Note: A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity; n = number of individuals sampled.

[a] Locality and voucher information are provided in Appendix 1.

* Highly significant deviation from Hardy–Weinberg equilibrium ($P < 0.001$).

Table 3. Cross-amplification results for the 16 microsatellites developed for *Begonia luzhaiensis* in seven populations of *B. leprosa* (n = 1).[a]

Locus	LK	QL	TP	FL	DL	LG	LS
BLZ01	—	+	—	—	—	—	—
BLZ02	+	+	—	+	+	+	+
BLZ03	—	—	—	—	—	—	+
BLZ04	+	+	+	+	+	+	+
BLZ05	+	+	—	—	—	+	+
BLZ06	+	+	+	+	+	+	+
BLZ07	—	—	—	—	—	—	+
BLZ08	+	+	+	+	+	+	+
BLZ09	+	+	+	+	+	+	+
BLZ10	+	+	—	+	+	—	+
BLZ11	—	—	—	—	—	—	—
BLZ12	—	+	+	+	+	—	+
BLZ13	+	+	+	+	+	+	+
BLZ14	+	*	*	*	*	+	*
BLZ15	+	*	*	*	*	+	*
BLZ16	+	*	*	*	*	+	*

Note: + = successful amplification; — = failed amplification; * = PCR has not been performed.

[a] Locality and voucher information are provided in Appendix 1.

CONCLUSIONS

The 16 microsatellite markers described here are the first developed for *B. luzhaiensis* and also the first study in *Begonia* sect. *Coelocentrum.* These microsatellites can be applied to the investigation of genetic diversity, population genetic structure, mating system, and gene flow, thus facilitating our understanding of evolutionary mechanisms and species diversification of the limestone flora. Data from such studies will contribute to the conservation and management of *B. luzhaiensis* that is increasingly threatened by habitat destruction.

LITERATURE CITED

Chung, K.-F., W.-C. Leong, R. R. Rubite, R. Repin, R. Kiew, Y. Liu, and C.-I Peng. 2014. Phylogenetic analyses of *Begonia* sect. *Coelocentrum* and allied limestone species of China shed light on the evolution of Sino-Vietnamese karst flora. *Botanical Studies (Taipei, Taiwan)* 55: 1.

Dikshit, H. K., A. Singh, D. Singh, M. S. Aski, P. Prakash, N. Jain, S. Meena, et al. 2015. Genetic diversity in *Lens* species revealed by EST and genomic simple sequence repeat analysis. *PLoS ONE* 10: e0138101.

Gu, C.-Z., C.-I Peng, and N. J. Turland. 2007. Begoniaceae. *In* Z.-Y. Wu, P. H. Raven, and D. Y. Hong [eds.], Flora of China, vol. 13, 153–207. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Holland, M. M., and W. Parson. 2011. GeneMarker® HID: A reliable software tool for the analysis of forensic STR data. *Journal of Forensic Sciences* 56: 29–35.

Li, C., L.-H. Yang, D.-K. Tian, Y. Chen, R.-J. Wu, and N.-F. Fu. 2016. *Begonia leipingensis* (Begoniaceae), a new compound-leaved species with unique petiolule pattern from Guangxi of China. *Phytotaxa* 244: 45–56.

Luo, R.-B., B.-H. Liu, Y.-L. Xie, Z.-Y. Li, W.-H. Huang, J.-Y. Yuan, G.-Z. He, et al. 2012. SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1: https://doi.org/10.1186/2047-217X-1-18.

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)* 28: 2537–2539.

Peng, C.-I, S.-M. Ku, H.-A. Yang, W.-C. Leong, Y. Liu, T. H. Nguyen, Y. Kono, and K.-F. Chung. 2014. Two new species of *Begonia* sect. *Coelocentrum, B. guixiensis* and *B. longa*, from Sino-Vietnamese limestone karsts. *Botanical Studies (Taipei, Taiwan)* 55: 52.

Peng, C.-I, C.-W. Lin, H.-A. Yang, Y. Kono, and H. Q. Nguyen. 2015. Six new species of *Begonia* (Begoniaceae) from limestone areas in Northern Vietnam. *Botanical Studies (Taipei, Taiwan)* 56: 9.

Raymond, M., and F. Rousset. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

Temnykh, S., G. DeClerck, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch. 2001. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. *Genome Research* 11: 1441–1452.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T. L. Madden. 2012. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 13: 134.

Zhou, Q., D. Luo, L.-C. Ma, W.-G. Xie, Y. Wang, Y.-R. Wang, and Z.-P. Liu. 2016. Development and cross-species transferability of EST-SSR markers in Siberian wildrye (*Elymus sibiricus* L.) using Illumina sequencing. *Scientific Reports* 6: 20549.

Appendix 1. Voucher information for *Begonia luzhaiensis* used in this study. All voucher specimens are deposited in the Herbarium of the Biodiversity Research Center (HAST), Academia Sinica, Taipei, Taiwan.

Species	Population code	*n*	Collection locality	Geographic coordinates	Voucher no.
Begonia luzhaiensis T. C. Ku	–	1	Puyi Township, Yanshuo County, Gulin City, Guangxi Province, China	24°43′41.0″N, 110°31′44.0″E	*C.-I Peng 18732*
	–	1	Puyi Township, Yanshuo County, Gulin City, Guangxi Province, China	24°42′34.0″N, 110°32′7.0″E	*C.-I Peng 18735*
	YS	15	County Town, Yanshuo County, Gulin City, Guangxi Province, China	24°46′12.9″N, 110°27′43.4″E	*H.-Y. Huang 008*
	LZ	20	Liuzhou City, Luzhai County, Guangxi Province, China	24°43′37.9″N, 109°39′50.6″E	*H.-Y. Huang 018*
	XC	22	Laibin City, Xingcheng County, Guangxi Province, China	24°03′49.5″N, 108°40′8.9″E	*H.-Y. Huang 102*
B. leprosa Hance	LK	1	Luokeng Town, Qujiang District, Shaoguan City, Guangdong Province, China	24°31′24.4″N, 113°21′19.3″E	*C.-I Peng 23998*
	QL	1	Qinglian Town, Yangshan County, Shaoguan City, Guangdong Province, China	24°27′23.0″N, 112°45′53.3″E	*C.-I Peng 24045*
	TP	1	Taiping Town, Yangshan County, Shaoguan City, Guangdong Province, China	24°10′33.4″N, 112°33′10.8″E	*C.-I Peng 24055*
	FL	1	Fuli Town, Yangshuo County, Gulin City, Guangxi Province, China	24°47′8.2″N, 110°31′22.5″E	*H.-Y. Huang 005*
	DL	1	Donglan County, Hechi City, Guangxi Province, China	24°26′47.1″N, 107°20′34.6″E	*H.-Y. Huang 079*
	LG	1	Longhushan, Longan County, Nanning City, Guangxi Province, China	22°57′10.2″N, 107°38′02.7″E	*H.-Y. Huang 089*
	LS	1	Longhushan, Longan County, Nanning City, Guangxi Province, China	22°57′41.9″N, 107°37′28.2″E	*H.-Y. Huang 092*

Note: *n* = number of individuals sampled.