Development of EST-SSR markers for the Korean endemic species *Chrysosplenium aureobracteatum* (Saxifragaceae)

Jae-Seo SHIN, Bo-Yun KIM[1], Yong-In KIM[2], Jung-Hoon LEE[2] and Young-Dong KIM*

Department of Life Science, Multidisciplinary Genome Institute, Hallym University, Chuncheon 24252, Korea
[1]National Institute of Biological Resources, Incheon 22689, Korea
[2]Multidisciplinary Genome Institute, Hallym University, Chuncheon 24252, Korea
(Received 17 January 2020; Revised 13 March 2020; Accepted 17 March 2020)

ABSTRACT: *Chrysosplenium aureobracteatum* Y. I. Kim & Y. D. Kim (Saxifragaceae) is a recently described endemic species growing in the central part of the Korean peninsula. It requires constant monitoring for conservation due to its limited distributions. There is also a need for molecular markers for proper assessments of the genetic differentiation of *C. aureobracteatum* from species morphologically similar to it. In this study, we developed microsatellite markers that can be used to evaluate the genetic diversity of this species, representing fundamental data with which to conserve the natural populations of the species. A total of 17 expressed sequence tag–simple sequence repeat (EST-SSR) markers were developed by the Illumina pair-end sequencing of the transcriptomes of *C. aureobracteatum*. These markers were successfully applied to populations of *C. aureobracteatum* and to its most closely related species, *C. barbatum*, revealing high polymorphism in both species. The EST-SSR markers developed in this study were proven to be useful not only to monitor the population genetic structure of *C. aureobracteatum* for conservation purposes but also to study the genetic delimitation of the species from species closely related to it.

Keywords: population genetics, endemic species, cross-amplification, conservation

Chrysosplenium aureobracteatum Y. I. Kim & Y. D. Kim (Saxifragaceae) is a recently described species only growing in the central part of the Korean peninsula (Kim and Kim, 2015). The recognition of the species as a separate taxon has long been hampered due to the lack of prominent diagnostic traits during the flowering period, causing many researchers to misidentify it as either *C. barbatum* Nakai, *C. fulvum* A. Terracc., or *C. sphaerospermum* Maxim. (Kim, 2014). Kim and Kim (2015) noted the shape of the leaves on the sterile stems which developed after the flowering period as the most reliable diagnostic character for the above species, which are morphologically very similar to each other (Hara, 1957).

C. aureobracteatum requires constant monitoring in terms of conservation due to its limited distribution only in northeastern Gyeonggi-do and northwestern Gangwon-do in South Korea. Genetic diversity data of populations is used as a key component for the establishment of conservation plans for rare endemic species (Solórzano et al., 2016; Jung et al., 2018). Although the species is well segregated from taxa closely resembling it in the internal transcribed spacer phylogeny (Kim, 2014), population genetic data would be important for understanding the evolutionary history and for clarifying the taxonomic delimitation of these species. Expressed sequence tag–simple sequence repeat (EST-SSR) markers developed from the RNAs of the expressed genes have relatively conservative characteristics (Ellis and Burke, 2007; Hu et al., 2011), making them useful not only for population genetic studies but also for investigations of the evolutionary relationships among closely related species (Balbino et al., 2019). In this study, EST-SSR markers based on the transcriptome data of *C. aureobracteatum* were developed to evaluate the genetic diversity of this newly described species.

*Author for correspondence: ydkim@hallym.ac.kr

Table 1. Locality and voucher information for the *Chrysosplenium aureobracteatum* and *C. barbatum* populations examined in this study. Voucher specimens were deposited in the HHU.

Taxon	Locality	No.	Voucher No.
C. aureobracteatum	Hwacheon, Gangwon-do (Gweongdeoksan Mt.)	20	KYI_2014358
C. barbatum	Wonju, Gangwon-do (Chiaksan Mt.)	20	KYI_2014271

No., number of individuals sampled.

Materials and Methods

RNA was extracted from the fresh leaves of an individual of *Chrysosplenium aureobracteatum* plant sampled from Gwangdeoksan Mt. (Voucher No. KYI_2014358) (Table 1) using a RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). The extracted RNA was used for a cDNA synthesis using a TruSeq cDNA Library Prep Kit, version 2 (Illumina, San Diego, CA, USA). The synthesized cDNA library was used for Illumina Hiseq 100 bp paired-end de novo transcriptome sequencing. Raw reads acquired by RNA sequencing were subjected to quality checking and trimming using Trimmomatic 0.32 (Bolger et al., 2014). EST-SSR motifs were obtained from the trimmed data using the Trinity (Hass et al., 2013) and the MIcroSAtellite (MISA) identification tools (Thiel et al., 2003). Primer sets were constructed to amplify the selected EST-SSR motifs using the Primer3 program (Rozen and Skaletsky, 1999). EST-SSR candidate loci (markers) were selected based on the number of SSR motif repetitions (5–8 times), the PCR product size (100–300 bp), the coverage depth (20–110), and the primer annealing temperature (55–60°C).

The candidate EST-SSR markers developed from *C. aureobracteatum* were applied to its most closely related species, *C. barbatum*, to confirm and obtain markers exhibiting polymorphism in both species. We examined 20 individuals of *Chrysosplenium aureobracteatum* and *C. barbatum* for the marker evaluation and cross-specific amplification test. The voucher specimens were deposited in the HHU (Herbarium of Hallym University) in Korea (Table 1).

PCR for the selected loci was conducted using a mixture with a total volume of 25 μL consisting of 10× Ex Taq buffer at 2.5 μL (TaKaRa Bio Inc., Otsu, Japan), 2.5 mM dNTP at 2 μL, 10 pmol forward and reverse primer labeled with fluorescent dye (FAM, HEX, NED) each at 0.5 μL, template DNA (5–10 ng) at 2 μL, Ex Taq DNA polymerase at 0.12 μL and SIGMA free water using a GeneAmp PCR System 9700 thermocycler (Applied Biosystems Inc., San Diego, CA, USA). The PCR process was performed under the following conditions: pre-denaturation (98°C, 5 min), denaturation (95°C, 1 min), annealing (59°C, 1d min), and extension (72°C, 1 min

30 sec). After going through the above reaction 35 times, the final extension was processed (72°C, 10 min). The PCR products were electrophoresed at 110 V in 1.2% agarose gel for 90 min, stained with ethidium bromide, and confirmed using a UV illuminator. Alleles were identified with an ABI 3730XL (automated sequencer) equipped with the GeneScan 500 LIZ Size Standard (Applied Biosystems). Genotyping was performed using GeneMapper 3.7 (Applied Biosystems), and peaks were scored manually by visual inspections. The PCR products with sizes deviating from the expected size ranges (as estimated by the Primer3 program) were eliminated from the genotyping process. Finally, the resulting genotypes were encoded to calculate the number of alleles (A), the observed heterozygosity (H_o), and the expected heterozygosity (H_e) using the GenAlEx 6.5 (Peakall and Smouse, 2012).

Results and Discussion

As a result of RNA sequencing, raw data of 8,486,210,570 bp were generated (GenBank BioProject number: PRJNA606349). The raw data were processed to produce 7,818,963,245 bp of trimmed data, with this data then utilized to produce a total of 70,753,963 bp of transcriptome data. A total of 15,547 SSR motifs were identified from the trimmed data. The identified SSR motifs consisted of 40% mono-nucleotides, 36% di-nucleotides, 23% tri-nucleotides, and 1% tetra-, penta-, and hexa-nucleotides (Table 2).

From the identified SSR motifs, 96 candidate markers were selected in consideration of the number of repetitions, the

Table 2. Types and ratios of SSR motifs identified in *Chrysosplenium aureobracteatum*.

Unit size	No. of SSRs (%)
Mononucleotide	6,260 (40.27)
Dinucleotide	5,531 (35.58)
Trinucleotide	3,593 (23.11)
Tetranucleotide	133 (0.86)
Pentanucleotide	18 (0.12)
Hexanucleotide	12 (0.08)

Table 3. Information on 17 EST-SSR loci selected from *Chrysosplenium aureobracteatum*.

Locus	Primer sequence (5'-3')	Repeat motif	Allele size range (bp)	GenBank accession No.
CA003	F:TCTGGTGCCAAATCTCACAA R:CGATCTCTGGTCCGAGGTC	$(AG)_7$	191–231	MN961272
CA005	F:AGAAGCAGAAGCAGGATCGG R:CTCGAATCAATAATTGAAGATCTGA	$(TC)_7$	209–221	MN961273
CA006	F:CTCATTGCAGCTCAAAACCCA R:CAGTAATCTCTGTCTCTCTTTCCA	$(AG)_8$	213–239	MN961274
CA008	F:TACCAGAGAATGCCGTGAGC R:CGTCTTTCACCAGCCAAAACA	$(TC)_7$	210–218	MN961275
CA010	F:TGCTCTTTGCTCTCACCAGG R:ATAATCCCGTCCGTCAAGGG	$(GA)_7$	225–233	MN961276
CA017	F:TCCGTCGAGCTTTTGGTCAA R:TGGTCCCTTCAATCCCCAAC	$(AT)_8$	244–256	MN961277
CA052	F:CTCTCTCATACCTTCCCGGC R:CCTGACATCATCCCAGCGAA	$(TC)_7$	278–296	MN961278
CA056	F:AGAAGACACTTGTTGCCTGA R:TTTGCCCTCGCCAATTGTTG	$(AAG)_5$	215–221	MN961279
CA060	F:ACAAGGCTGCAACCCAAAAC R:TCGAGAGGCCGATCGTAGTA	$(CCG)_5$	270–294	MN961280
CA063	F:TGGTGCTTGGAGATAACCCA R:GAGGCTTGGCAATGCAAACA	$(TTG)_6$	201–216	MN961281
CA066	F:TGACGGGTCCAAATCGTCAA R:GAGAGGAGGTGGTGGTGTTG	$(CCA)_6$	211–229	MN961282
CA073	F:TCCTCATCCACAGCAAGCTT R:AGTGCGAGTGAAAAGAGCGA	$(CTT)_5$	266–275	MN961283
CA076	F:TCGTTGTCGCTACTGCTACC R:CCACCTGAGCAGTCACAAGA	$(TTC)_5$	209–230	MN961284
CA080	F:GCGAATAACGAAACGGACCG R:ACAGACAAGCAGCGGTAAGA	$(GAT)_5$	345–372	MN961285
CA091	F:ATCTCACCGTCGATTCAGCC R:CCGGAGTTCGAACATCACCT	$(GAA)_5$	207–222	MN961286
CA092	F:TGGTCTCTCTAATCCGTGCC R:CATGCACAATCGGAAGCGTT	$(TTC)_5$	257–272	MN961287
CA104	F:TCCAGTTCCAAAGACCACCC R:CAGCCCGCTAGTCAAAATGC	$(AGC)_5$	197–209	MN961288

coverage depth, the annealing temperature, and the PCR product size. Excluding nonspecific bands with sizes deviating from the expected sizes, a total of 17 EST-SSR loci (markers) showing polymorphisms for *Chrysosplenium aureobracteatum* and *C. barbatum* were finally developed (Table 3).

As a result of applying the 17 EST-SSR markers on a population (with 20 individuals) of *C. aureobracteatum*, polymorphisms were detected from 11 markers, and a total of 33 alleles were identified (2 to 7 alleles depending on the polymorphic markers). The observed heterozygosity (H_o) and

Table 4. Genetic diversity in *Chrysosplenium aureobracteatum* and *C. barbatum* based on 17 developed microsatellite markers (loci).

Locus	*C. aureobracteatum* ($n = 20$)			*C. barbatum* ($n = 20$)		
	A	Ho	*He*[a]	A	Ho	*He*[a]
CA_003	2	0.350	0.289	5	0.250	0.721***
CA_005	1	0.000	0.000	5	0.550	0.725
CA_006	1	0.000	0.000	5	0.650	0.779**
CA_008	2	0.350	0.349	-	-	-
CA_010	1	0.000	0.000	3	0.250	0.224
CA_017	4	0.750	0.543	-	-	-
CA_052	1	0.000	0.000	4	0.800	0.674*
CA_056	2	0.000	0.095***	1	0.000	0.000
CA_060	2	0.150	0.139	4	0.400	0.586**
CA_063	3	0.600	0.621	3	0.500	0.454*
CA_066	1	0.000	0.000	3	0.700	0.564
CA_073	3	0.550	0.484	1	0.000	0.000
CA_076	2	0.050	0.049	3	0.850	0.559*
CA_080	2	0.500	0.495	2	0.750	0.469**
CA_091	3	0.150	0.224	2	0.550	0.489
CA_092	2	0.250	0.219	3	0.400	0.339
CA_104	1	0.000	0.000	3	0.700	0.611

n, number of individuals; *A,* number of alleles; *Ho,* observed heterozygosity; *He,* expected heterozygosity; - unamplified.
[a]Significant deviation from Hardy-Weinberg equilibrium (*$p < 0.05$, **$p < 0.01$, ***$p < 0.0001$).

the expected heterozygosity (H_e) ranged from 0.000 to 0.750 and 0.000 to 0.543, respectively.

Cross-specific amplification for a population (with 20 individuals) of *C. barbatum* confirmed that 13 markers were polymorphic out of 15 amplifiable markers, resulting in a total of 47 alleles (2 to 8 alleles depending on the polymorphic markers). The observed heterozygosity (H_o) and the expected heterozygosity (H_e) ranged 0.000 to 0.850 and 0.000 to 0.779, respectively (Table 4).

It is important to note that the six monomorphic markers in *C. aureobracteatum* (CA_005, CA_006, CA_010, CA_052, CA_066, CA_104) were highly polymorphic in *C. barbatum*. In contrast, the monomorphic markers for *C. barbatum* (CA_056, CA_073) exhibited polymorphism in *C. aureobracteatum* (Table 4). These results suggest that the EST-SSR markers developed for *C. aureobracteatum* are useful not only to examine the population genetic structure but also for species delimitation among related taxa.

ORCID: Jae-Seo SHIN https://orcid.org/0000-0001-9632-1930; Bo-Yun KIM https://orcid.org/0000-0002-1269-4068; Yong-In KIM https://orcid.org/0000-0001-7482-7228; Jung-Hoon LEE https://orcid.org/0000-0001-6760-7913; Young-Dong KIM https://orcid.org/0000-0001-5904-4727

Acknowledgments

Conflict of Interest

Young-Dong Kim, a contributing editor of the Korean Journal of Plant Taxonomy, was not involved in the editorial evaluation or decision to publish this article. All remaining authors have declared no conflicts of interest.

Literature Cited

Balbino, E., G. Martins, S. Morais and C. Almeida. 2019. Genome survey and development of 18 microsatellite markers to assess genetic diversity in *Spondias tuberosa* Arruda Câmara (Anacardiaceae) and cross-amplification in congeneric species. Molecular Biology Reports 46: 3511–3517.

Bolger, A. M., M. Lohse and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114–2120.

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

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, D. Eccles, B. Li, M. Lieber, M. D. MacManes, M. Ott, J. Orvis, N. Pochet, F. Strozzi, N. Weeks, R. Westerman, T. William, C. N. Dewey, R. Henschel, R. D. LeDuc, N. Friedman and A. Regev. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8: 1494–1512.

Hara, H. 1957. Synopsis of genus *Chrysosplenium* L. Journal of the Faculty of Science, University of Tokyo. Section III: Botany 7: 1–90.

Hu, J., L. Wang and J. Li. 2011. Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. Biologia Plantarum 55: 577–580.

Jung, E.-K., D.-H. Kang, K.-O. Yoo, M. Kwak, Y.-D. Kim and B.-Y. Kim. 2018. Isolation and characterization of EST-SSR markers for *Astilboides tabularis* (Saxifragaceae), endangered species in Korea. Korean Journal of Plant Taxonomy 48: 195–200.

Kim, Y. I. 2014. Systematic studies of *Chrysosplenium* L. series *Pilosa* Maxim. (Saxifragaceae). PhD dissertation, Hallym University, Chuncheon, Korea. Pp. 89–91. (in Korean)

Kim, Y.-I. and Y.-D. Kim. 2015. *Chrysosplenium aureobracteatum* (Saxifragaceae), a new species from South Korea. Novon 23: 432–436.

Peakall, R and P. E. Smouse. 2012. GenALEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics 28: 2537–2539.

Rozen, S. and H. Skaletsky. 1999. Primer 3 on the WWW for general users and for biologist programmers. Methods in Molecular Biology 132: 365–386.

Solórzano, S., S. Arias and P. Dávila. 2016. Genetics and conservation of plant species of extremely narrow geographic range. Diversity 8: 31.

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 Applied Genetics 106: 411–422.