Development and characterization of 30 microsatellite loci for *Plagiorhegma dubium* (Berberidaceae)

Bo-Yun Kim¹, Soo-Rang Lee¹, Yong-In Kim², Dae-Hyun Kang¹, and Young-Dong Kim¹

PREMISE OF THE STUDY: *Plagiorhegma dubium* (Berberidaceae) has been listed as an endangered species in Korea due to extensive collection and destruction of natural habitats. In this study, 30 microsatellite loci, including 25 polymorphic loci, were developed for *P. dubium* for use in population-level genetic analyses.

METHODS AND RESULTS: We carried out transcriptome sequencing and isolated a total of 30 expressed sequence tag–simple sequence repeat markers from *P. dubium* using Illumina Hiseq high-throughput sequencing. To test utility of the developed markers, we genotyped 60 individuals from three populations and estimated the number of alleles and levels of observed and expected heterozygosity. Expected heterozygosity levels ranged from 0.000 to 0.594, 0.000 to 1.000, and 0.000 to 0.744 in the three populations, respectively.

CONCLUSIONS: These transcriptome-derived simple sequence repeat markers are highly polymorphic and can be widely used in characterization of the endangered *P. dubium*. Population genetic studies with these markers will provide valuable insights for conservation by unraveling evolutionary patterns of *P. dubium*.

KEYWORDS: Berberidaceae; EST-SSR marker; genetic diversity; medicinal plant; microsatellites; *Plagiorhegma dubium*.

The genus *Plagiorhegma* Maxim. (Berberidaceae) is composed of only one species, *P. dubium* Maxim., which has a narrow distribution in northeastern Asia and Siberia (Jeong and Sivanesan, 2016). *Plagiorhegma dubium* has been listed as an endangered species by the Ministry of the Environment of South Korea due to its rarity and specific habitat preference (Kim, 2006; Ghimire and Heo, 2012). The plants have a long history of medicinal uses: roots have long been used as folk medicine for stomachache, and the phytochemicals extracted from the plants have shown remedial effects on elevated cholesterol levels (Kong et al., 2004). *Plagiorhegma dubium* is also appreciated as an ornamental plant for commercial use due to its showy flowers and heart-shaped leaves (Huang, 1995). Despite growing interest in the species in both the pharmaceutical industry and horticulture, there is a complete lack of knowledge on the *P. dubium* genome. No applicable microsatellite markers have been developed for genetic studies and genetic analyses at the population level, which can provide insights for conservation and management plans of rare and threatened species (Ottewell et al., 2016). Simple sequence repeats (SSRs), also known as microsatellites, are the most frequently used molecular markers due to their abundance, codominant mode of inheritance, and multiallelic and highly polymorphic nature (Lopez et al., 2015). However, the development of microsatellite markers is an expensive and time-consuming processes (Squirrell et al., 2003). Expressed sequence tag (EST)–derived SSRs can overcome some of the drawbacks of older methods, for example by decreasing processing time (Zhou et al., 2016). Here, we developed and characterized 30 EST-SSR markers for the rare and threatened *P. dubium* using transcriptome sequencing with the Illumina paired-end sequencing platform. We evaluated the performance of these markers using 60 individuals representing three populations of *P. dubium*.

METHODS AND RESULTS

Transcriptome sequencing

To prepare a cDNA library, total RNA of *P. dubium* was extracted from a fresh leaf of a single sample from Korea (voucher no. NIBRVP0000556155; Appendix 1). RNA was extracted using the RNeasy Kit version 2.2 (Illumina, San Diego, California, USA) following the manufacturer’s instructions, and was used for TruSeq cDNA library preparation. The RNA libraries were sequenced on the Illumina Hiseq 2000 platform, producing 150-bp paired-end reads.
| Locus  | Primer sequence (5′–3′)                                                                 | Repeat motif | Allele size range (bp) | Fluorescent dye | $T_c$ (°C) | GenBank accession no. | Putative function       | E-value |
|--------|----------------------------------------------------------------------------------------|--------------|------------------------|-----------------|------------|----------------------|-------------------------|---------|
| JD_03a | F: CATGCCACCTCCTACCCAATCCTCA R: GCCAGTGGACCTGCGGACACAGTCAA                              | (TG)$_a$     | 247                    | HEX             | 60         | Pr032816664          | Not found               | –       |
| JD_05  | F: TCGGCAAGATTGAGACACCTT R: GGTTGAGTTGCTGACATAT                                          | (TG)$_a$     | 273–277                | FAM             | 60         | Pr032816654          | Not found               | –       |
| JD_11  | F: AGACTTTGGGAGTTCAACCAACC R: TCTCCCCCTTTGGGCTGATT                                     | (TA)$_a$     | 196–198                | FAM             | 60         | Pr032816640          | Not found               | –       |
| JD_14a | F: TTGGGTTGGCAAGGCACTG    R: GCTTTGGAAGGCCTAGATT                                      | (TA)$_a$     | 248                    | FAM             | 60         | Pr032816662          | Hypothetical protein PRUPE_ppa020282mg [Prunus persica] | 4E-130  |
| JD_17b | F: GTGTAGCCATCCTCAAGGCA R: GTGTTAAGGTTGTAAGGCA                                       | (GT)$_a$     | 184–186                | FAM             | 55         | Pr032816636          | Predicted: transcription elongation factor B polypeptide 2-like isoform X2 [Citrus sinensis] | 1E-52   |
| JD_19  | F: TGATTGCCACCAAGCTTCAAGA R: GCTTTGACACTGTCGCCAGG                                      | (GT)$_a$     | 262–266                | FAM             | 60         | Pr032816656          | Hypothetical protein MTR_Bg085190 [Medicago truncatula] | 1E-56   |
| JD_24b | F: GCATGATGCTGCTCTGTGTG R: GGTTGAGTTGAGGAAGG                                      | (GA)$_a$     | 222–226                | HEX             | 60         | Pr032816650          | Mitogen-activated protein kinase kinase 5 [Petroselinum crispum] | 2E-159  |
| JD_27b | F: TCTGCAAAATGGGCTGTCG    R: GCATTTGCACTTGCTGAGTCG                                     | (CT)$_a$     | 165–170                | FAM             | 60         | Pr032816655          | Unnamed protein product [Vitis vinifera] | 0.0     |
| JD_29  | F: ACTTGCTAGCTAGGCGTTG    R: TCAAGACTCAACCTGCTTC                                    | (CT)$_a$     | 176–180                | FAM             | 60         | Pr032816660          | Predicted: E3 ubiquitin-protein ligase UFL3-like isoform 1 [Vitis vinifera] | 0.0     |
| JD_30b | F: ATCTGCTGTTGCTGCTGTCT    R: AGACGACATCGTTTACCAGGCA                                      | (CA)$_a$     | 198                    | FAM             | 55         | Pr032816651          | R2R3-MYB transcription factor MYB9 [Epimedium sagittatum] | 3E-126  |
| JD_33  | F: CTGCCCCCTGATGCTCGCG    R: ACAGTTGGCTAGATGCTGTT                                      | (CA)$_a$     | 204–212                | FAM             | 60         | Pr032816663          | Not found               | –       |
| JD_35b | F: AGTGCAGCCTAGAAGCTGTG    R: TTTAGCAGCTGCCTGTG                                      | (AT)$_a$     | 266–268                | FAM             | 55         | Pr032816657          | Not found               | –       |
| JD_36b | F: TGCTGCTCCACTACCTGGTG    R: TCAAGACTCCAGGCTGCTT                                    | (AT)$_a$     | 191–193                | FAM             | 60         | Pr032816658          | Predicted: E3 ubiquitin-protein ligase At1g63170 [Prunus mume] | 6E-142  |
| JD_38  | F: GCAATGACCATCGACATCGCC    R: ACTGCGCTGACAAGCTGAA                                      | (AT)$_a$     | 182–184                | FAM             | 60         | Pr032816661          | Hypothetical protein JCGZ_15866 [Jatropha curcas] | 8E-125  |
| JD_42b | F: CTGTACAAAGACTCCCGCGCTC    R: CGAGGTTGAATCTGTGGCCTC                                     | (AG)$_a$     | 142–144                | HEX             | 55         | Pr032816645          | Hypothetical protein PRUPE_ppa006575mg [Prunus persica] | 0.0     |
| JD_43b | F: TTGTGCGCCATGCTGCTGG    R: AACCGTGGAAAGCACATGAA                                      | (AG)$_a$     | 480–488                | FAM             | 55         | Pr032816652          | Predicted: uncharacterized membrane protein At3g27390-like [Solanum tuberosum] | 2E-121  |
| JD_50  | F: GCCCCAATCTACCTCTTGAC    R: ACAGCGAAAAACCTCCAACCTACCTACCTACCTACA                              | (TTT)$_a$    | 275–278                | HEX             | 55         | Pr032816659          | Unnamed protein product [Vitis vinifera] | 9E-194  |
| JD_52b | F: AGTGCATAACAGAAGCTGTTG    R: TCAGCACTCACTACACTG                                     | (TTT)$_a$    | 218–220                | FAM             | 60         | Pr032816637          | Predicted: uncharacterized protein LOC100261915 [Vitis vinifera] | 9E-158  |
| JD_54b | F: ACCGCAACAATCCGGCAAAAAAC    R: ACCTTCTATCCGGCTATTCACCTAC                                   | (TGG)$_a$    | 183                    | FAM             | 60         | Pr032816639          | Not found               | –       |
| JD_66  | F: CCCCGCCCGCCCGCCACCTG    R: CCAAGTATCAGAGGCGCACA                                      | (GCA)$_a$    | 215–224                | HEX             | 60         | Pr032816644          | Unnamed protein product [Vitis vinifera] | 4E-108  |
| JD_68  | F: TCCCCACCTGCTCTCTCTC    R: AGCTTCTTTGCGCCAGTCA                                       | (GAG)$_a$    | 282–294                | FAM             | 60         | Pr032816647          | Hypothetical protein POPTR_0010s13850g [Populus trichocarpa] | 4E-101  |
| JD_71a | F: GCAGCAAACAATCTCGAACTCAA    R: GTCCTGTGCAAAACTTGGCA                                      | (GAA)$_a$    | 266                    | FAM             | 60         | Pr032816646          | Predicted: transcriptional regulator ATRX homolog [Prunus mume] | 1E-28   |
| JD_72b | F: GCCGACTGTTCCTGAAGTGCA    R: ACATAGCCTGACCGGACCAA                                      | (CTT)$_a$    | 259–262                | FAM             | 60         | Pr032816642          | Not found               | –       |

(Continues)
reads. All raw reads were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (Bioproject ID PRJNA472226). Trimmomatic 0.32 (Bolger et al., 2014) was used to remove adapters and low-quality reads with the following parameters: seed mismatch of 2, palindrome clip threshold of 30, simple clip threshold of 10, a sliding window size of 4, with an average quality of 20 and a minimum sequence length of 50 bases. After trimming, reads were assembled into 76,725 unigenes. 

Development of microsatellite markers based on ESTs of *P. dubium*

Microsatellites from the unigenes were detected using MISA version 1.0.0 (Thiel et al., 2003) with the default parameters. The criteria for identifying SSRs were as follows: the minimum number of nucleotide repeats was six for mono-, di-, tri-, tetra- penta-, and hexanucleotides. MISA identified 11,458 SSRs, of which 96 primer pairs were selected for further testing based on (1) region containing at least five repetitions of di- or trinucleotide motifs; (2) PCR product size of 140–500 bp; (3) length ranging from 12 to 24 nucleotides; (4) annealing temperature 55–60°C; and (5) minimum GC content 50%. Primer pairs 

| Locus | Primer sequence (5’–3’) | Repeat motif | Allele size range (bp) | Fluorescent dye | GenBank accession no. | Putative function [Organism] | E-value |
|-------|-------------------------|--------------|------------------------|----------------|----------------------|----------------------------|--------|
| JD_77 | F: CTCGGCTTACCTGGAGGCCTACGTCGTA R: CTGTTCGGGTTATGCACGGA | (CGG),6 | 180–186 | FAM | 60 | Pr032816641 | Predicted: zinc finger A2O and AN1 domain-containing protein 4 isoform 1 [Vitis vinifera] | 2E-59 |
| JD_78 | F: TGGATGAGCTTGGCACCACCAT R: GTCAGGTTCAAATGGTCACTCA | (CGG),6 | 200–203 | FAM | 55 | Pr032816653 | Unnamed protein product [Vitis vinifera] | 1E-168 |
| JD_79 | F: AGGCGGTCAGAAGTAGGGGTCA GTCAAGCCCCGTC | (CCT),6 | 220–223 | HEX | 60 | Pr032816638 | Decarboxylating-like 6-phosphogluconate dehydrogenase [Medicago truncatula] | 0.0 |
| JD_81 | F: CCTGCTTTTCCAAACTTGCR CCAAGGCGAGGCGATAGT | (CCA),6 | 223–229 | HEX | 55 | Pr032816649 | RNA-binding CRS1/YhbY domain-containing protein, putative [Theobroma cacao] | Not found |
| JD_82 | F: GCCCTGGAGGTTTTGAGAGCT R: GCTCCGGTCGATGAAATAG | (CAT),6 | 254–257 | FAM | 55 | Pr032816643 | Not found |
| JD_83 | F: AAGGACACGCGACGATGAA R: GGGGATAGGGGTTGGGAAA | (CAG),6 | 216–219 | FAM | 60 | Pr032816648 | GATA transcription factor 9 [Morus notabilis] | 5E-61 |
| JD_93 | F: TCACTCGACCGCCCTCATTT R: CCCCCGGGCTCATTAGT | (ACA),6 | 234–247 | HEX | 55 | Pr032816635 | Hypothetical protein AMTR_s00029p00127080 [Amborella trichopoda] | 3E-25 |

Note: *T* = annealing temperature.

⁎Monomorphic loci.

⁎Fixed heterozygotes.

The utility of the selected 96 microsatellite markers was evaluated in a GeneAmp PCR System 9700 thermocycler (Applied Biosystems, Carlsbad, California, USA) using the following protocol: initial denaturation at 98°C for 5 min; followed by 30 cycles of denaturation at 95°C for 1 min, annealing at locus-specific annealing temperature (Table 1) for 1 min, and extension at 72°C for 1.5 min; and a final extension step at 72°C for 10 min. Of the 96 candidate markers, 34 markers amplified and were suitable for further testing. Finally, 30 markers were selected and included in the following analysis after excluding four markers with a low amplification rate of 80% or less. The PCR products were labeled with fluorescent dyes (HEX and FAM) and run on an ABI 3730XL automated sequencer (Applied Biosystems) using the GeneScan 500 LIZ Size Standard (Applied Biosystems). Genotyping was manually determined with GeneMapper 3.7 (Applied Biosystems). Identified microsatellite markers (Table 1) producing clear and polymorphic bands were subsequently used for genetic diversity assessments.

Genetic parameters

Fresh leaves from 60 *P. dubium* individuals were sampled from three populations from Korea, Japan, and China. The voucher specimens were deposited in the National Institute of Biological Resources Herbarium (KB), Incheon, Republic of Korea (Appendix 1). The precise locations of the sites have been withheld to prevent illegal collection. The level of polymorphism at each locus was assessed by calculating the number of alleles per marker (*A*), observed heterozygosity, and expected heterozygosity using GenAIEx 6.5 (Peakall and Smouse, 2012). Deviation from Hardy–Weinberg equilibrium was estimated with Arlequin 3.5 (Excoffier and Lischer, 2010).

Functional annotations for these 30 markers were compared against the NCBI nonredundant (NR) protein database with BLASTX (E-value 1 × 10⁻4). A total of 30 markers were successfully amplified, of which polymorphism was detected in 25 (Table 1).
Because 11 markers (17, 24, 27, 35, 36, 42, 43, 52, 72, 79, and 83) were fixed for heterozygotes in all 60 samples, we present genetic parameters only for the remaining 14 markers (Table 2).

The number of alleles per locus was estimated from 14 polymorphic EST-SSR markers among three populations of *P. dubium* (*A* = 1–4, average = 2.143; Table 2). Levels of observed heterozygosity for each locus ranged from 0.000 to 1.000. Levels of expected heterozygosity ranged from 0.000 to 0.594, 0.000 to 1.000, and 0.000 to 0.744 in the three sampled populations (Table 2). No significant linkage disequilibrium was found in all pairs of 30 loci after Bonferroni correction (*α* = 0.05), whereas some markers revealed significant deviation from Hardy–Weinberg equilibrium (Table 2).

**CONCLUSIONS**

We developed and amplified a set of novel microsatellite markers for *P. dubium*. These markers will be used for constructing an in situ and ex situ conservation strategy of the species by estimating the level of genetic diversity and population structure in wild populations. Assessment of its genetic variation could contribute to infer-ences of the past evolutionary history of *P. dubium*. Furthermore, the 11,458 SSR loci identified, additional markers could be developed to address research needs.

**ACKNOWLEDGMENTS**

This research was supported by the grant “The Genetic Diversity Study of Indigenous Vascular Plant Species, 2nd year” (NIBR2 01603203) funded by the National Institute of Biological Resources, Republic of Korea.

**DATA ACCESSIBILITY**

Raw reads were submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (Bioproject ID PRJNA472226). Sequence information for the developed primers has been deposited to NCBI; GenBank accession numbers are provided in Table 1.

**LITERATURE CITED**

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

Excoffier, L., and H. Lischer. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564–567.

Ghimire, B., and K. Heo. 2012. Embryology of *Jeffersonia dubia* Baker et S. Moore (Berberidaceae) and comparison with allied genera. *Korean Journal of Plant Taxonomy* 42: 260–266.

Haas, B. J., A. Papanicolaou, M. Yassour, M. Grabherr, P. D. Blood, J. Bowden, M. B. Couger, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* 8: 1494–1512.

Huang, M. 1995. New ornamental crops in Asia. *Acta Horticulturae* 397: 43–57.

Jeong, B. R., and I. Sivanesan. 2016. Micropropagation, berberine content and antitumor activity of *Jeffersonia dubia* (Maxim.) Benth et Hook. *Plant Cell Tissue and Organ Culture* 124: 453–458.

Kim, Y. S. 2006. Conservation of plant diversity in Korea. *Landscape and Ecological Engineering* 2: 163–170.

Kong, W., J. Wei, P. Abidi, M. Lin, S. Inaba, C. Li, Y. Wang, et al. 2004. Berberine is a novel cholesterol-lowering drug working through a unique mechanism distinct from statins. *Nature Medicine* 10: 1344–1351.

Lopez, L., R. Barreiro, M. Fischer, and M. A. Koch. 2015. Mining microsatellite markers from public expressed sequence tags databases for the study of threatened plants. *BMC Genomics* 16: 781.

Otowell, K. M., D. C. Bickerton, M. Byrne, and A. J. Lowe. 2016. Bridging the gap: A genetic assessment framework for population-level threatened plant conservation prioritization and decision-making. *Diversity and Distributions* 22: 174–188.

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. 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.

---

**TABLE 2.** Genetic diversity in three *Plagiorhegma dubium* populations based on 14 newly developed polymorphic microsatellite markers.a

| Locus | Korea (n = 20) | China (n = 20) | Russia (n = 20) |
|-------|---------------|---------------|----------------|
|       | A            | H_o           | H_e           | A          | H_o | H_e     | A        | H_o | H_e     |
| JD_05 | 3            | 1.000         | 0.584**       | 2          | 1.000 | 0.500** | 3        | 1.000 | 0.525** |
| JD_11 | 2            | 0.684         | 0.450*        | 2          | 0.100 | 0.095    | 2        | 0.950 | 0.499** |
| JD_19 | 3            | 1.000         | 0.594**       | 3          | 0.950 | 0.559**  | 3        | 1.000 | 0.605** |
| JD_27 | 2            | 0.400         | 0.320         | 1          | 0.000 | 0.000    | 2        | 0.050 | 0.139** |
|JD_33  | 2            | 1.000         | 0.500**       | 2          | 1.000 | 1.000**  | 3        | 1.000 | 0.623** |
| JD_38 | 2            | 0.250         | 0.289         | 2          | 0.850 | 0.499**  | 2        | 0.400 | 0.455  |
| JD_50 | 2            | 1.000         | 0.500**       | 2          | 0.400 | 0.375    | 2        | 0.313 | 0.264  |
| JD_66 | 1            | 0.000         | 0.000         | 1          | 0.000 | 0.000    | 3        | 0.200 | 0.580** |
| JD_68 | 2            | 1.000         | 0.500**       | 4          | 1.000 | 0.590    | 4        | 1.000 | 0.744** |
| JD_77 | 2            | 0.050         | 0.469**       | 1          | 0.000 | 0.000    | 3        | 0.100 | 0.185** |
| JD_78 | 2            | 0.263         | 0.411         | 2          | 0.200 | 0.180    | 1        | 0.000 | 0.000  |
| JD_81 | 2            | 0.050         | 0.494         | 1          | 0.000 | 0.000    | 2        | 0.050 | 0.139** |
| JD_82 | 2            | 0.250         | 0.219         | 2          | 1.000 | 0.500**  | 2        | 0.950 | 0.499** |
| JD_93 | 2            | 0.950         | 0.499**       | 2          | 0.950 | 0.499**  | 2        | 1.000 | 0.500** |
| Mean  | 2.071         | 0.564         | 0.384         | 1.929      | 0.532 | 0.307    | 2.429    | 0.572 | 0.411  |

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

aLocality and voucher information are provided in Appendix 1.

bSignificant deviation from Hardy–Weinberg equilibrium after correction for multiple tests (*P* < 0.05 and **P* < 0.01).
Squirrel, J. P. M. Hollingsworth, M. Woodhead, J. Russell, A. J. Lowe, M. Gibby, and W. Powell. 2003. How much effort is required to isolate nuclear microsatellites from plants? Molecular Ecology 12: 1339–1348.

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.

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

APPENDIX 1. Locality and voucher information for *Plagiorhegma dubium* populations sampled in this study.a,b

| Population | Locality          | N  | Voucher no.          |
|------------|-------------------|----|----------------------|
| Korea      | Uiseong, Gyeongbuk| 20 | NIBRP0000556155      |
| China      | Yanbian, Jilin    | 20 | NIBRP0000601483      |
| Russia     | Vladivostok, Primorskiy | 20 | NIBRP0000556157      |

Note: N = number of individuals.

*a Voucher specimens were deposited in the Herbarium of the National Institute of Biological Resources (KB), Incheon, Republic of Korea.

*b Precise locations of the collection sites have been withheld to prevent illegal collection.