Development and characterization of microsatellite markers for *Rhododendron purdomii* (Ericaceae) using next-generation sequencing

Ningning Zhang¹, Mengyun Qin², Shixin Zhu³, Ziyang Huang³, Hao Dong², Yang Yang³, Lili Yang¹ and Yang Lu², ⁴*¹Academy of Chinese Medical Sciences, Henan University of Chinese Medicine, Zhengzhou 450046, China ²School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China ³School of Life Sciences, Zhengzhou University, Zhengzhou 450001, China ⁴Funiu Mountain Forest Ecosystem Observation and Research Station, Zhengzhou University, Zhengzhou 450001, China

(Received 8 May 2021, accepted 13 September 2021; J-STAGE Advance published date: 15 December 2021)

*Rhododendron purdomii* (Ericaceae) is an endangered ornamental species endemic to the Qinling Mountains of China. Due to the impact of climate change and human disturbance, *R. purdomii* is threatened by habitat loss, and conservation of this species is urgently needed. In this study, we developed and characterized 13 novel microsatellite markers for *R. purdomii* based on next-generation sequencing data. For the 13 microsatellite markers in three *R. purdomii* populations, the number of alleles ranged from two to 12, the number of effective alleles was from 1.000 to 8.892, Shannon’s information index was from 0.000 to 2.320, and the observed and expected heterozygosity were from 0.000 to 1.000 and from 0.000 to 0.888, respectively. Private alleles were found in all three populations. Moderate differentiation between population pairs was indicated by pairwise $F_{ST}$ values. The microsatellite markers developed in this study will provide opportunities for examining the genetic diversity and population structure of *R. purdomii* and contribute to the effective conservation of this species.

**Key words:** microsatellites, *Rhododendron purdomii*, RAD-seq, Ericaceae, population genetics

*Rhododendron* is the largest genus of the family Ericaceae, containing more than 1,000 species worldwide (Ma et al., 2014). This genus includes many important ornamental plants. There are approximately 571 *Rhododendron* species in China, of which 409 are endemic (Fang et al., 2005). *Rhododendron* has been suggested as a good model system for evolutionary, ecological and horticultural studies (Ma et al., 2014), and there are many related studies on species located in hotspot areas (e.g., Hengduan Mountains) in South China (Yu et al., 2017; Shrestha and Wang, 2018). However, little is known about the *Rhododendron* species that are restricted to the central region of China (Zhao et al., 2012).

*Rhododendron purdomii* Rehder & E. H. Wilson is an evergreen shrub or small tree species endemic to the Qinling Mountains of China, occurring in Henan, Shaanxi and Gansu provinces (Fig. 1). It was originally collected on Taibai Mountain of Shaanxi province by William Purdom in 1910, and is therefore named in honor of the collector (Fang et al., 2005). It is an important component of the montane ecosystem at 1,800–3,500 m altitude, and plays a vital role in erosion control and climate regulation (Yu et al., 2017). In addition, *R. purdomii* is a horticulturally valuable plant with colorful flowers. However, this species has been subject to some level of habitat disturbance due to human destruction and climate change (Ma et al., 2014; Yu et al., 2017). Based on a recent evaluation, *R. purdomii* is listed as vulnerable to extinction in the Red List of China Higher Plants (Qin et al., 2017). Therefore, understanding its genetic diversity and structural patterns is urgently required to effectively...
monitor and conserve this species.

Previous studies on \textit{R. purdomii} are mainly about ornamental horticulture and resource development, such as evaluation of morphological traits (Si et al., 2012) and investigation of germplasm resources (Zhao et al., 2013). Despite its great value for horticulture research, the genetic diversity and population structure of \textit{R. purdomii} have not been clarified, the only exception being that some populations in Shaanxi were assessed based on amplified fragment length polymorphism markers (Zhao et al., 2012). Due to the characteristics of codominant inheritance, high polymorphism and wide distribution in the genome, microsatellites are proven to be useful in population genetic studies (Feng et al., 2014; Qin et al., 2021). In this study, we identified microsatellite markers for \textit{R. purdomii} based on restriction site-associated DNA sequencing (RAD-seq) data, and assessed the genetic diversity and differentiation of three \textit{R. purdomii} populations using these novel markers.

To develop the microsatellite markers and evaluate their polymorphism, samples from three natural populations of \textit{R. purdomii} were collected from Laojun Mountain (LJS, \(n = 11\)), Longyuwan (LYW, \(n = 15\)) and Laojieling (LJL, \(n = 17\)) in Henan province, China. The locality information of the sampled populations is detailed in Supplementary Table S1. Fresh and healthy leaves of the investigated individuals were collected, and the individuals sampled within one site were at least 10 m apart. Permissions were obtained from the local nature reserve for collecting plant materials. Leaves were dried using silica gel and stored in plastic bags until DNA extraction, and voucher specimens were deposited in the Herbarium of Zhengzhou University.

Genomic DNA of all the individuals investigated was extracted from dried leaf tissues using a modified CTAB method (Doyle and Doyle, 1987). An RAD library of one \textit{R. purdomii} individual was constructed using the EcoRI (5’-GAATTC-3’) enzyme following Miller et al. (2007) and Baird et al. (2008). The library was sequenced at Novogene (Beijing, China) using the Illumina HiSeq 2000 platform with 150-bp paired-end reads. After filtering low-quality reads, \textit{de novo} assembly of the clean reads was performed using Velvet (Zerbino and Birney, 2008) with default parameters. Perfect microsatellite motifs with a repeat unit of 2–6 bp and a minimum number of four repeats were detected using MISA (Beier et al., 2017) with default settings. The program Primer3 (Untergasser et al., 2012) was used to design microsatellite primers; the length of primers ranged from 20 to 28 bp, and the annealing temperature was 60–65 °C.

Microsatellite loci were validated through three steps as follows. First, we randomly selected three individuals from three different populations of \textit{R. purdomii} to test the success of amplification for the designed primer pairs. Next, the forward primers of microsatellites successfully amplified in the three individuals were labeled with fluorescent dye, and fluorescent PCR products of six individuals selected from three different populations were analyzed for polymorphism. Third, microsatellites showing the expected size range on agarose gels, as well as clear peaks and polymorphism during capillary electrophoresis in six individuals, were further amplified in all the investigated individuals of \textit{R. purdomii}. The PCR reaction mixture volume was 50 μl, consisting of distilled water (22 μl), 2 × PCR Mixture (25 μl, Beibei Biotech, Henan, China), 10 μM forward and reverse primers (1 μl for each primer) and genomic DNA (1 μl). The PCR reaction conditions were: an initial denaturation at 94 °C for 4 min; 35 cycles at 94 °C for 30 s, annealing temperature for 30 s and 72 °C for 30 s; a final extension at 72 °C for 7 min. The success of amplification was determined by electrophoresis on a 1% agarose gel. The fluorescent PCR products were analyzed through capillary electrophoresis on an ABI 3730XL DNA Sequencer (Applied
Rhododendron purdomii microsatellites

Biosystems, Foster City, CA, USA) at the Sangon Biotech Corporation (Shanghai, China), and the genotypes were obtained using GeneMarker (SoftGenetics, State College, PA, USA).

For population genetics analyses, the number of alleles ($N_A$), number of effective alleles ($N_E$), Shannon’s information index ($I$), number of private alleles ($N_{PA}$), observed heterozygosity ($H_O$) and expected heterozygosity ($H_E$) were estimated using GenAlEx (Peakall and Smouse, 2006). Pairwise $F_{ST}$ values and number of migrants ($N_m$) were also calculated using GenAlEx. Polymorphism information content (PIC) was calculated using CERVUS (Kalinowski et al., 2007). Tests for Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium between microsatellite loci were conducted using GENEPOP (Rousset, 2008). Frequencies of null alleles were estimated by Micro-Checker (van Oosterhout et al., 2004).

For the RAD-seq data of $R. purdomii$, a total of 16,152,280 paired-end reads were generated after quality filtering. After de novo assembly of the clean reads, 301,199 contigs with a mean size of 353 bp and a mean GC content of 40.17% were obtained. The software MISA detected 6,853 perfect microsatellite loci in the assembled contigs, 6,714 of which were suitable for primer design. Finally, of the 60 primer pairs selected for the initial tests, 31 could be amplified with clear bands. In these 31 microsatellite loci, 13 loci presenting reproducible amplification, clear peaks and rich polymorphism were selected for $R. purdomii$.

Of the 13 polymorphic microsatellite loci, five were dinucleotide repeats, three were trinucleotides, three were tetranucleotides and two were hexanucleotides (Table 1). All 13 sequences were submitted to GenBank (MW736532–MW736544). The number of alleles per locus ranged from four (P24, P60) to 18 (P23), $I$ ranged from 0.000 to 2.320, the $H_O$ ranged from 0.000 to 1.000 and the $H_E$ ranged from 0.000 to 0.888 (Table 2). Of all the loci, locus P23 had the highest NPA, while loci P57, P58 and P60 contained no private alleles. Private alleles were found in all three populations and the LYW population possessed the highest number of private alleles (12). Null alleles were found in the locus P23 in three populations, P28 in the LJS population, P33 in the LYW population, P57 and P58 in the LJL population, and P59

| Locus | Primer sequences (5′–3′) | Repeat motif | Size range (bp) | $N_A$ | PIC | $T_a$ (°C) | GenBank accession no. |
|-------|---------------------------|--------------|----------------|-------|-----|-----------|-----------------------|
| P3    | F:AAGAATGCTGAAAAATGCTTCTCA R:ACCATCGGCTTTTATGTCTTCC | (TC)$_{16}$ | 133–167 | 13 | 0.813 | 52 | MW736532 |
| P16   | F:GTTCCTAAGATCAGCAGGCTCTCTGG R:AGAACAATAACAGGACATAAGCG | (TC)$_{15}$ | 97–133 | 15 | 0.910 | 53 | MW736533 |
| P23   | F:TTGGGTTGTCGAAAAATACAGGAGGATTACG | (AG)$_{16}$ | 135–207 | 18 | 0.891 | 52 | MW736534 |
| P24   | F:GTTCCTAGAAAAAGATGCTTCTACA R:CGCTACGCTGAGTAGATCCTA | (TTT)$_{15}$ | 118–127 | 4 | 0.460 | 54 | MW736535 |
| P28   | F:GCTATTACCTCTCATATGCGACCCG R:CTGGACAGAGGAAATATGGACCA | (CT)$_{15}$ | 113–157 | 13 | 0.802 | 52 | MW736536 |
| P33   | F:GTAATACGCTGCTATGCTCCCA R:TCAATTGAAATTTCTGAAACACAC | (CT)$_{15}$ | 121–157 | 15 | 0.898 | 52 | MW736537 |
| P41   | F:ATGAAAAGGAGAGAGGATGAGA R:AAAGTCAACACAGGATCTCCCA | (AGA)$_{8}$ | 100–127 | 7 | 0.687 | 51 | MW736538 |
| P42   | F:GACGCCAACAATAACTAAACACATT R:ATAGATTGAGAAACCACTCGGACA | (TAA)$_{8}$ | 101–188 | 13 | 0.789 | 50 | MW736539 |
| P46   | F:GCTTTCTCCATATAGTATTTAAGGTT R:CATACGAGTACCTACGTTGATT | (TTTA)$_{5}$ | 136–156 | 6 | 0.625 | 54 | MW736540 |
| P57   | F:TGTTGTCCTATTTATTCTCCCAATT R:ACCGTACGACGACCAAGATTTC | (TGT)$_{5}$ | 121–141 | 13 | 0.802 | 52 | MW736541 |
| P58   | F:GTTGACACATCGACAAGGTCTTCCTT R:TCTTTTGCTCTCTTTCATTTAT | (CGGG)$_{5}$ | 131–179 | 6 | 0.624 | 54 | MW736542 |
| P59   | F:GTGTTGCTGAGCAACAAAGATGATG R:ACCAAAAATATACACGTGGG | (TTTT)$_{5}$ | 125–149 | 5 | 0.572 | 52 | MW736543 |
| P60   | F:GTAATAGGGTTGATGGGGAAG R:ATAATGCAAGTGAACGTAGGCA | (TTT)$_{5}$ | 145–157 | 4 | 0.596 | 52 | MW736544 |

$N_A$, number of alleles; PIC, polymorphism information content; $T_a$, annealing temperature.

Table 1. Characterization of microsatellite markers for Rhododendron purdomii
in the LJS and LYW populations. After a Bonferroni correction for multiple comparisons, no significant deviations from HWE or linkage disequilibrium were found for any loci. Three populations showed moderate differentiation as indicated by pairwise $F_{ST}$ values (Table 3), among which the LJS and LJL populations indicated the highest level of differentiation (0.058) while the LJS and LYW populations showed the lowest (0.041). $N_{m}$ values between population pairs ranged from 4.091 to 5.902.

In this study, we developed and characterized 13 microsatellite markers for *Rhododendron purdomii* based on next-generation sequencing data and successfully validated them in three *Rhododendron purdomii* populations. These markers will be useful for future population genetic and evolutionary history studies of *Rhododendron purdomii*, which are necessary for the efficient management and conservation of this species.

We thank local staff at the Funiu Mountain Nature Reserve for their assistance during the fieldwork. This work was funded by the National Natural Science Foundation of China (Grant Nos. 31800551, 81903747) and supported by the Supercomputing Center in Zhengzhou University.

---

**Table 2. Genetic diversity parameters of 13 polymorphic microsatellite loci in three populations of *Rhododendron purdomii***

| Locus | LJS (n = 11) | LYW (n = 15) | LJL (n = 17) |
|-------|-------------|-------------|-------------|
|       | $N_A$ | $N_E$ | $I$ | $H_O$ | $H_E$ | NPA | Null present | $N_A$ | $N_E$ | $I$ | $H_O$ | $H_E$ | NPA | Null present | $N_A$ | $N_E$ | $I$ | $H_O$ | $H_E$ | NPA | Null present |
| P3    | 9     | 6.205 | 1.981 | 1.000 | 0.839 | 2   | no          | 4     | 2.018 | 0.936 | 0.467 | 0.504 | 0   | no          | 10    | 5.780 | 1.994 | 0.824 | 0.827 | 3   | no          |
| P16   | 11    | 8.345 | 2.248 | 0.909 | 0.580 | 2   | no          | 9     | 7.895 | 2.119 | 0.933 | 0.873 | 0   | no          | 9     | 6.721 | 2.035 | 0.941 | 0.851 | 2   | no          |
| P23   | 11    | 7.118 | 2.177 | 0.636 | 0.860 | 2   | no          | 10    | 7.500 | 2.147 | 0.600 | 0.867 | 2   | yes         | 12    | 8.892 | 2.320 | 0.706 | 0.886 | 3   | yes         |
| P24   | 2     | 1.198 | 0.305 | 0.182 | 0.165 | 0   | no          | 2     | 1.385 | 0.451 | 0.333 | 0.278 | 0   | no          | 3     | 2.117 | 0.805 | 0.706 | 0.528 | 1   | no          |
| P28   | 6     | 2.916 | 1.397 | 0.455 | 0.657 | 0   | no          | 9     | 5.844 | 1.948 | 0.667 | 0.829 | 2   | no          | 8     | 4.699 | 1.752 | 0.882 | 0.778 | 0   | no          |
| P33   | 9     | 6.205 | 1.981 | 0.636 | 0.839 | 0   | no          | 11    | 7.627 | 2.199 | 0.600 | 0.869 | 4   | yes         | 8     | 5.207 | 1.833 | 0.941 | 0.808 | 0   | no          |
| P41   | 5     | 3.408 | 1.384 | 1.000 | 0.707 | 0   | no          | 6     | 3.516 | 1.434 | 1.000 | 0.716 | 1   | no          | 6     | 3.753 | 1.470 | 1.000 | 0.734 | 1   | no          |
| P42   | 4     | 2.444 | 1.053 | 0.455 | 0.591 | 0   | no          | 6     | 3.214 | 1.372 | 0.533 | 0.689 | 1   | no          | 7     | 4.188 | 1.596 | 0.706 | 0.761 | 0   | no          |
| P46   | 4     | 2.814 | 1.162 | 0.909 | 0.645 | 1   | no          | 5     | 3.147 | 1.320 | 0.933 | 0.682 | 2   | no          | 3     | 2.206 | 0.869 | 0.588 | 0.547 | 0   | no          |
| P57   | 5     | 2.916 | 1.277 | 0.727 | 0.657 | 0   | no          | 3     | 1.613 | 0.683 | 0.333 | 0.380 | 0   | no          | 3     | 2.847 | 1.073 | 0.412 | 0.649 | 0   | yes         |
| P58   | 3     | 1.322 | 0.485 | 0.273 | 0.244 | 0   | no          | 1     | 1.000 | 0.000 | 0.000 | 0.000 | 0   | no          | 3     | 1.197 | 0.355 | 0.059 | 0.164 | 0   | yes         |
| P59   | 5     | 4.172 | 1.520 | 0.364 | 0.760 | 1   | yes         | 4     | 3.285 | 1.282 | 0.400 | 0.696 | 0   | yes         | 4     | 2.460 | 1.114 | 0.471 | 0.593 | 0   | no          |
| P60   | 3     | 2.142 | 0.837 | 0.636 | 0.533 | 0   | no          | 4     | 2.778 | 1.139 | 0.600 | 0.640 | 0   | no          | 3     | 2.042 | 0.829 | 0.412 | 0.510 | 0   | no          |

$n$, number of sampled individuals; $N_A$, number of alleles; $N_E$, number of effective alleles; $I$, Shannon's information index; $H_O$, observed heterozygosity; $H_E$, expected heterozygosity; NPA, number of private alleles.

---

**REFERENCES**

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., and Johnson, E. A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, e3376.

Beier, S., Thiel, T., Münch, T., Schlö, U., and Mascher, M. (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33, 2583–2585.

Doyle, J. J., and Doyle, J. L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.

Fang, M., Fang, R., He, M., Hu, L., Yang, H., and Chamberlain, D. F. (2005) *Rhododendron*. In Flora of China 14. (eds.: Wu C., Raven, P. H., and Hong, D.), pp. 260–455. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, USA.

Feng, X., Wang, Y., and Gong, X. (2014) Genetic diversity, genetic structure and demographic history of *Cycas simplicipinna* (Cycadaceae) assessed by DNA sequences and SSR markers. BMC Plant Biol. 14, 187–203.

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16, 1099–1106.

Ma, Y., Nielsen, J., Chamberlain, D. F., Li, X., and Sun, W. (2014) The conservation of *Rhododendrons* is of greater urgency than has been previously acknowledged in China. Biodivers. Conserv. 23, 3149–3154.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., and Johnson, E. A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17, 240–248.

Peakall, R., and Smouse, P. E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6, 288–295.

Qin, H., Yang, Y., Dong, S., He, Q., Jia, Y., Zhao, L., Yu, S., Liu, H., Liu, B., Yan, Y., et al. (2017) Threatened species list of China’s higher plants. Biodiv. Sci. 25, 696–744.
Qin, M., Zhang, N., Zhu, S., Yue, C., Huang, J., Dong, H., and Lu, Y. (2021) Development of polymorphic microsatellite markers for the Tertiary relict tree species *Taiwania cryptomerioides* (Cupressaceae) in East Asia. Mol. Biol. Rep. 48, 3031–3036.

Rousset, F. (2008) GENEPOP’007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol. Ecol. Resour. 8, 103–106.

Shrestha, N., and Wang, Z. (2018) Selecting priority areas for systematic conservation of Chinese *Rhododendron*: hotspot versus complementarity approaches. Biodivers. Conserv. 27, 3759–3775.

Si, G., Zhang, Y., Zhao, B., and Xu, H. (2012) Phenotypic variation of natural populations in *Rhododendron purdomii* in Qinling Mountains. Acta Bot. Boreal.-Occident. Sin. 32, 1560–1566 (in Chinese).

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res. 40, e115.

van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol. Ecol. Notes 4, 535–538.

Yu, F., Skidmore, A. K., Wang, T., Huang, J., Ma, K., and Groen, T. A. (2017) *Rhododendron* diversity patterns and priority conservation areas in China. Divers. Distrib. 23, 1143–1156.

Zerbino, D. R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829.

Zhao, B., Yin, Z.-f., Xu, M., and Wang, Q.-c. (2012) AFLP analysis of genetic variation in wild populations of five *Rhododendron* species in Qinling Mountain in China. Biochem. Syst. Ecol. 45, 198–205.

Zhao, B., Zhang, G., Si, G., Gu, X., and Zhang, Y. (2013) Investigation on *Rhododendron* germplasm in Qinling mountains. Journal of Northwest Forestry University 28, 104–109 (in Chinese).
Supplementary Table S1. Locality and voucher information of *Rhododendron purdomii*

| Population code | n  | Locality                      | Latitude (°N) | Longitude (°E) | Elevation (m) | Voucher accession no. |
|-----------------|----|--------------------------------|---------------|----------------|---------------|-----------------------|
| LJS             | 11 | Laojun Mountain, Luoyang, Henan | 33°43′5.66″N  | 111°38′23.59″E | 2170          | LY2020051601          |
| LYW             | 15 | Longyuwan, Luoyang, Henan      | 33°39′37.41″N  | 111°47′13.31″E | 2120          | LY2020051701          |
| LJJ             | 17 | Laojieling, Nanyang, Henan     | 33°39′51.48″N  | 111°46′22.66″E | 1800          | LY2020082601          |

n, number of sampled individuals.