Development and Characterization of Microsatellite Markers for *Rhododendron Purdomii* Using Next-Generation Sequencing

Ningning Zhang
Academy of Chinese Medical Sciences, Henan University of Chinese Medicine, Zhengzhou 450046, China
Mengyun Qin
School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China
Shixin Zhu
School of Life Sciences, Zhengzhou University, Zhengzhou 450001, China
Ziyang Huang
School of Life Sciences, Zhengzhou University, Zhengzhou 450001, China
Hao Dong
School of Agricultural Sciences, Zhengzhou University , Zhengzhou 450001, China
Yang Yang
School of Life Science, Zhengzhou University, Zhengzhou 450001, China
Lili Yang
Academy of Chinese Medical Sciences, Henan University of Chinese Medicine, Zhengzhou 450046, China
Yang Lu (luyang@zzu.edu.cn)
School of Agricultural Sciences, Zhengzhou University, Zhengzhou 450001, China https://orcid.org/0000-0002-2564-1993

Short Report

Keywords: Microsatellite, Rhododendron purdomii, RAD-seq, Cross-species amplification, Population genetics

Posted Date: April 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-449799/v1

Abstract

Rhododendron purdomii, an endangered ornamental species endemic to the Qinling Mountains, is an important component of montane ecosystem in central China. Due to the impact of climate change and human disturbance, management and conservation of this species are in urgent needs. In this study, we developed 13 novel microsatellite markers for *R. purdomii* based on next-generation sequencing data, and tested these markers' utility in congeneric species *R. concinnum*. For the 13 microsatellite markers in three *R. purdomii* populations, number of alleles ranged from two to 12, 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. Cross-species amplification for *R. concinnum* indicated eight microsatellite loci were successfully amplified and polymorphic. 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.

Introduction

Rhododendron is the largest genus of the family Ericaceae, containing more than 1000 species worldwide [1]. This genus has many important ornamental plants. There are approximately 571 *Rhododendron* species in China, of which 409 species are endemic [2]. *Rhododendron* has been suggested as a good model system for evolutionary, ecological, and horticultural studies [1], and there are many related studies on species located in the hotspot areas (e.g. Hengduan Mountains) in South China [3, 4]. However, little is known about the *Rhododendron* species restricted to the central region of China [5].

Rhododendron purdomii Rehder & E. H. Wilson is an evergreen shrub or small tree species endemic to the Qinling Mountains of China, occurred in Henan, Shaanxi and Gansu provinces (Fig. 1). *R. purdomii* was originally collected on Taibai Mountain of Shaanxi province by William Purdom in 1910, so it is named in honour of the collector [2]. 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 [4]. In addition, this species is a horticulturally significant plant with colorful flowers (Fig. 2). Based on our recent field survey in Henan, we found that local people collected and transplanted wild individuals of *R. purdomii* together, in order to build *Rhododendron* garden to attract tourists. Moreover, this species has been subject to some level of habitat disturbance due to human destruction and climate change [1, 4]. Based on recent evaluation, *R. purdomii* is listed as vulnerable to extinction in the Red List of China Higher Plants [6]. Therefore, understanding the genetic diversity and structural patterns of *R. purdomii* is urgently required to effectively monitor and conserve this species.

Previous studies on *R. purdomii* are mainly about ornamental horticulture and resource development, such as evaluation on morphological traits [7], and investigation of germplasm resources [8]. Despite its great value for horticulture research, genetic diversity and population structure of *R. purdomii* are not clear, with the exception that only some populations in Shaanxi were assessed based on amplified fragment length polymorphism (AFLP) [5]. Due to the characteristics of codominant inheritance, high polymorphism and wide distribution in genome, microsatellites are proven to be useful in population genetic studies [9, 10]. In this study, we developed microsatellite loci for *R. purdomii* based on restriction-site associated DNA sequencing (RAD-seq), and cross-species amplification was conducted for these newly developed markers in *R. concinnum*.

Materials And Methods

Sample collection and DNA extraction

To develop the microsatellite markers and evaluate the polymorphism of the markers, three natural populations of *R. purdomii* were collected from Laojun Mountain (LJS, n = 11), Longyuwan (LYW, n = 15) and Laojieling (LJL, n = 17) in Henan province, China, respectively. Furthermore, to validate the selected microsatellite primers in other *Rhododendron* species, we sampled two populations of *R. concinnum* from Yao Mountain (CYS, n = 10) and Laojun Mountain (CLJ, n = 10) in Henan, respectively. The locality information of the sampled populations was detailed in Table S1. Fresh and healthy leaves of the investigated individuals were collected and sampled individuals were at least 10 m apart within one site. Permissions were obtained from the local nature reserve for collecting plant materials. Leaves were dried by silica gel and stored in plastic bags until DNA extraction, and voucher specimens were deposited in the Herbarium of Zhengzhou University (ZZU).

RAD sequencing and microsatellite mining

Genomic DNA of all the investigated individuals was extracted from dried leaf tissues using a modified CTAB method [11]. A restriction-site associated DNA (RAD) library of one *R. purdomii* individual was constructed using the *EcoR*I (5'-GAATTC-3') enzyme following Miller et al. [12] and Baird et al. [13]. 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, *de novo* assembling of the clean reads was performed by Velvet [14] with default parameters. Microsatellite motifs with a repeat unit of 2-6 bp and a minimum number of four repeats were detected using MISA [15] with default settings. The program Primer3 [16] was used to design microsatellite primers, the length of primers ranged from 20 to 28bp, the annealing temperature was 60-65℃.

Validation of microsatellite loci and cross-species amplification

The validation of microsatellite loci was performed through three steps as follows. Firstly, we randomly selected three individuals from three different populations of *R. purdomii* to test the success of amplification for the designed primer pairs. Then, the forward primers of microsatellites successfully amplified in three individuals were labeled with fluorescent dye, and fluorescent PCR products of the selected six individuals from three different populations were analyzed for polymorphism. Thirdly, microsatellites showing expected size range on agarose gels, clear peaks and polymorphism during capillary electrophoresis in six individuals were further amplified in all the investigated individuals of *R. purdomii*. In the above processes, PCR reaction mixture used was 50 μL in volume, 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), 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, 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 Biosystems, Foster City, CA, USA) at the Sangon Biotech Corporation (Shanghai, China), and the genotypes were obtained using GeneMarker (Soft Genetics).

Samples of *R. concinnum* were used to evaluate the transferability of the developed microsatellite markers in congeneric species. The screening of primer pairs suitable for *R. concinnum* was consistent with three steps described above, excepting that two individuals from two different populations were used for amplification validation and four from two populations respectively were for polymorphism test.

Population statistics

For population genetics analyses, number of alleles (N_A), number of effective alleles (N_E), Shannon's information index (I), observed heterozygosity (H_O) and expected heterozygosity (H_E) were estimated using GenAlEx [17]. Polymorphism information content (PIC) was calculated by CERVUS [18]. Test for Hardy-Weinberg equilibrium and linkage disequilibrium between microsatellite loci was conducted using GENEPOP [19]. Frequencies of null alleles were estimated by Micro-Checker [20].

Results And Discussion

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* assembling of the clean reads, 301,199 contigs with the mean size of 353 bp and the mean GC content of 40.17% were obtained. The software MISA detected 6,853 microsatellite loci in the assembled contigs, and 6,714 of which were suitable for primer design. Of the 60 primer pairs selected for the initial tests, 31 of them can 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* finally. In the amplification for *R. concinnum*, nine of the 13 loci were amplified successfully, of which one locus showed monomorphism and eight polymorphic loci were detected.

Of the 13 polymorphic microsatellite loci, five were dinucleotide repeats, three were trinucleotides, three were tetranuclotides and two were hexanucleotides (Table 1). All the 13 sequences were submitted to GenBank (MW736532 - MW736544). The number of alleles per locus ranged from four (P24, P60) to 18 (P23) over all investigated *R. purdomii* individuals, while polymorphism information content values were from 0.240 (P58) to 0.915 (P23). In three investigated populations of *R. purdomii*, the number of alleles ranged from two to 12, number of effective alleles ranged from 1.000 to 8.892, Shannon's information index ranged from 0.000 to 2.320, observed heterozygosity ranged from 0.000 to 1.000, and expected heterozygosity ranged from 0.000 to 0.888 (Table 2). Significant deviations from Hardy-Weinberg equilibrium in terms of heterozygosity deficiency were found in three loci of LJS population, four of LYW, and four of LJL population, respectively. Null alleles were found in the locus P23 in three populations, P28 in LJS population, P33 in LYW population, P57 and P58 in LJL population, and P59 in LJS and LYW population. Deviations from Hardy-Weinberg equilibrium of some loci might be related to the presence of null alleles. Linkage disequilibrium was detected in one pair loci (P16 & P60) in LJL population. The linkage disequilibrium might indicate physical proximity of the loci on the chromosome or some evolutionary processes, such as selection, introgression and genetic drift, and those loci should be used with caution in different analyses [21, 22].

In two populations of *R. concinnum*, the number of alleles ranged from three to 10, number of effective alleles ranged from 2.198 to 7.692, Shannon's information index ranged from 0.856 to 2.155, observed heterozygosity ranged from 0.500 to 1.000, and expected heterozygosity ranged from 0.545 to 0.870 (Table 3). Two loci in CYS population and one locus in CLJ population deviated from Hardy-Weinberg equilibrium. No evidence of null alleles was identified. Linkage disequilibrium was found in loci P28 and P57 in CYS population.

Declarations

Acknowledgements This work is funded by the National Natural Science Foundation of China (Grant No. 31800551, 81903747). This work is supported by Supercomputing Center in Zhengzhou University (Zhengzhou). We thank local staff at the Funiu Mountain Nature Reserve for their assistance during the fieldwork.

Conflicts of interest: The authors declare that they have no conflict of interest.

Ethics approval: Not applicable.

Consent to participate: Not applicable.

Consent for publication: Not applicable.

Availability of data and material: Not applicable.

Code availability: Not applicable.

Authors' contributions: Not applicable.

References

1. Ma YP, Nielsen J, Chamberlain DF, Li XY, Sun WB (2014) The conservation of *Rhododendrons* is of greater urgency than has been previously acknowledged in China. Biodivers Conserv 23:3149–3154. https://doi.org/10.1007/s10531-014-0764-9
2. Fang MY, Fang RC, He MY et al (2005) Rhododendron. In: Wu CY, Raven PH, Hong DY (eds) Flora of China, vol 14. Science Press& Missouri Botanical Garden Press, Beijing & St. Louis, pp 260–455
3. Shrestha N 1, Wang ZH (2018) Selecting priority areas for systematic conservation of Chinese *Rhododendron*: hotspot versus complementarity approaches. Biodivers Conserv 27:3759–3775. https://doi.org/10.1007/s10531-018-1625-8
4. Yu FY, Skidmore AK, Wang TJ et al (2017) *Rhododendron* diversity patterns and priority conservation areas in China. Divers Distrib 23:1143–1156. https://doi.org/10.1111/ddi.12607
5. Zhao B, Yin ZF, Xu M, Wang QC (2012) AFLP analysis of genetic variation in wild populations of five *Rhododendron* species in Qinling Mountain in China. Biochem Syst Ecol 45:198–205. http://dx.doi.org/10.1016/j.bse.2012.07.033
6. Qin HN, Yang Y, Dong SY et al (2017) Threatened species list of China's higher plants. Biodiv Sci 25:696–744. https://doi.org/10.17520/biods.2017144
7. Si GC, Zhang YL, Zhao B, Xu H (2012) Phenotypic variation of natural populations in *Rhododendron purdomii* in Qinling Mountains. Acta Bot Boreal-Occident Sin 32:1560–1566
8. Zhao B, Zhang G, Si GC, Gu X, Zhang YL (2013) Investigation on *Rhododendron* germplasm in Qinling mountains. Journal of Northwest Forestry University 28:104–109. https://doi.org/10.3969/j.issn.1001-7461.2013.01.20
9. Feng X, Wang Y, 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. https://doi.org/10.1186/1471-2229-14-187
10. Qin MY, Zhang NN, Zhu SX et al (2021) Development of polymorphic microsatellite markers for the Tertiary relict tree species *Taiwania cryptomerioides* (Cupressaceae) in East Asia. Mol Biol Rep. https://doi.org/10.1007/s11033-021-06287-z
11. Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15
12. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res 17:240–248. https://doi.org/10.1101/gr.5681207
13. Baird NA, Etter PD, Atwood TS et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD Markers. PLoS ONE 3:e3376. https://doi.org/10.1371/journal.pone.0003376
14. Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107
15. Beier S, Thiel T, Muench T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33:2583–2585. https://doi.org/10.1093/bioinformatics/btx198
16. Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3-new capabilities and interfaces. Nucleic Acids Res 40:e15. https://doi.org/10.1093/nar/gks596
17. Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295. https://doi.org/10.1111/j.1471-8286.2005.01155.x
18. Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x
19. Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. Mol Ecol Resour 8:103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x
20. Van OC, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes 4:535–538. https://doi.org/10.1111/j.1471-8286.2004.00684.x
21. Hull JM, Sage GK, Sonsthagen SA, Gravley MC, Talbot SL (2020) Isolation and characterization of microsatellite loci in merlins (*Falco columbarius*) and cross-species amplification in gyrfalcons (*F. rusticolus*) and peregrine falcons (*F. peregrinus*). Mol Biol Rep 47:8377–8383. https://doi.org/10.1007/s11033-020-05842-4
22. Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits, vol 1. Sinauer, Sunderland, MA

Tables

Table 1 Characterization of the microsatellite markers for *Rhododendron purdomii*

Locus	Primer sequences (5'–3')	Repeat motif	Size range (bp)	N_A	PIC	*T*a (°C)	GenBank accession no.
P3	F:AAGAATGCTGAAAATGTCTTCCA R:ACCATCTGGCTTCTTTAGTTTCC	$(TC)_{16}$	133-167	13	0.813	52	MW736532
P16	F:GTTCTAAGATCCAAGCCTTCTGG R:AGAACAAATCAAGGACATAAGCG	$(TC)_{15}$	97-133	15	0.910	53	MW736533
P23	F:TTGGGTGTCAAAATAAAACCAAG R:GAAGCCAGGAACAGGATAGAATC	$(AG)_{16}$	135-207	18	0.915	52	MW736534
P24	F:GTCCTAGGAAAAAGATGCCTTACA R:CGCTACGAGTGGATTCAGCTA	$(TTC)_7$	118-127	4	0.460	54	MW736535
P28	F:GCTATTTACCTTCTATTGCACGC R:TCTGGACAGAGAGAATATGGACC	$(CT)_{15}$	113-157	13	0.802	53	MW736536
P33	F:GTAATACGATGCTATCGCTCCAC R:TCAATTGAATTTCGAACACACAC	$(CT)_{15}$	121-157	15	0.898	52	MW736537
P41	F:ATGAAATTGAGAGGAACGATTGA R:AAGTCAATCCACAGAGATTCCAA	$(AGA)_8$	100-127	7	0.687	51	MW736538
P42	F:GCACGCAAACTAATCAAAACATT R:ATAGATTGAAAAACCATCGGACA	$(TAA)_8$	101-188	13	0.789	50	MW736539
P46	F:GGGCTTCCAATAGATTTAAGGGT R:CCATACGAGACCTTACCCTGATT	$(TTTA)_5$	136-156	6	0.625	54	MW736540
P57	F:TGGGTCCTACTTATTCCCAATTT R:AACGTACGACGACCAAGATTTC	$(TGTA)_5$	121-141	5	0.647	53	MW736541
P58	F:GGTACACATCGACAAGCTCTCTT R:TCCTTTCGGCTCTCCTTACTTAT	$(CGGGAG)_5$	131-179	6	0.240	54	MW736542
P59	F:TGTTTCGGACAACAAAGAGTATG R:ACCAAACTAAAATACAACTGCGG	$(TTTTAT)_6$	125-149	5	0.572	52	MW736543
P60	F:GTAATAGGGTTGGTATGGGGAAG R:ATAATCGAAATGAACGTAAGCCA	$(TTCT)_5$	145-157	4	0.596	52	MW736544

N_A number of alleles, PIC polymorphism information content, *T*a annealing temperature.

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

	LJS (n = 11)						LYW (n = 15)						LJL(n =17)				
Locus	N_A	N_E	I	H_O	H_E	Null present	N_A	N_E	I	H_O	H_E	Null present	N_A	N_E	I	H_O	H_E
P3	9	6.205	1.981	1.000	0.839	no	4	2.018	0.936	0.467	0.504	no	10	5.780	1.994	0.824	0.82
P16	11	8.345	2.248	0.909	0.880	no	9	7.895	2.119	0.933	0.873	no	9	6.721	2.035	0.941	0.85
P23	11	7.118	2.177	0.636*	0.860	yes	10	7.500	2.147	0.600*	0.867	yes	12	8.892	2.320	0.706*	0.88
P24	2	1.198	0.305	0.182	0.165	no	2	1.385	0.451	0.333	0.278	no	3	2.117	0.805	0.706	0.52
P28	6	2.916	1.397	0.455	0.657	yes	9	5.844	1.948	0.667	0.829	no	8	4.699	1.752	0.882	0.78
P33	9	6.205	1.981	0.636	0.839	no	11	7.627	2.199	0.600*	0.869	yes	8	5.207	1.833	0.941	0.80
P41	5	3.408	1.384	1.000	0.707	no	6	3.516	1.434	1.000	0.716	no	6	3.753	1.470	1.000	0.73
P42	4	2.444	1.053	0.455	0.591	no	6	3.214	1.372	0.533	0.689	no	7	4.188	1.596	0.706	0.76
P46	4	2.814	1.162	0.909*	0.645	no	5	3.147	1.320	0.933*	0.682	no	3	2.206	0.869	0.588	0.54
P57	5	2.916	1.277	0.727	0.657	no	3	1.613	0.683	0.333	0.380	no	3	2.847	1.073	0.412	0.64
P58	3	1.322	0.485	0.273	0.244	no	1	1.000	0.000	0.000	0.000	no	3	1.197	0.355	0.059*	0.16
P59	5	4.172	1.520	0.364*	0.760	yes	4	3.285	1.282	0.400*	0.696	yes	4	2.460	1.114	0.471*	0.59
P60	3	2.142	0.837	0.636	0.533	no	4	2.778	1.139	0.600	0.640	no	3	2.042	0.829	0.412*	0.51

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;

*significant deviation from Hardy–Weinberg equilibrium at $P < 0.05$.

Table 3 Genetic diversity parameters of eight polymorphic microsatellite loci in two populations of *Rhododendron concinnum*

	CYS (n = 10)						CLJ (n = 10)					
Locus	N_A	N_E	I	H_O	H_E	Null present	N_A	N_E	I	H_O	H_E	Null present
P16	7	5.714	1.822	1.000	0.825	no	10	6.452	2.085	1.000	0.845	no
P24	8	3.922	1.704	1.000	0.745	no	7	4.167	1.639	1.000	0.760	no
P28	9	5.405	1.917	1.000*	0.815	no	9	7.143	2.068	1.000	0.860	no
P33	7	4.348	1.696	0.900	0.770	no	9	7.407	2.085	1.000	0.865	no
P41	5	3.704	1.432	1.000	0.730	no	3	2.667	1.040	0.900	0.625	no
P42	10	7.692	2.155	1.000	0.870	no	9	5.000	1.873	0.900	0.800	no
P57	5	3.279	1.344	1.000	0.695	no	4	2.410	1.106	0.700	0.585	no
P60	5	3.774	1.458	0.500*	0.735	no	3	2.198	0.856	0.900*	0.545	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;

*significant deviation from Hardy–Weinberg equilibrium at $P < 0.05$.

Figures

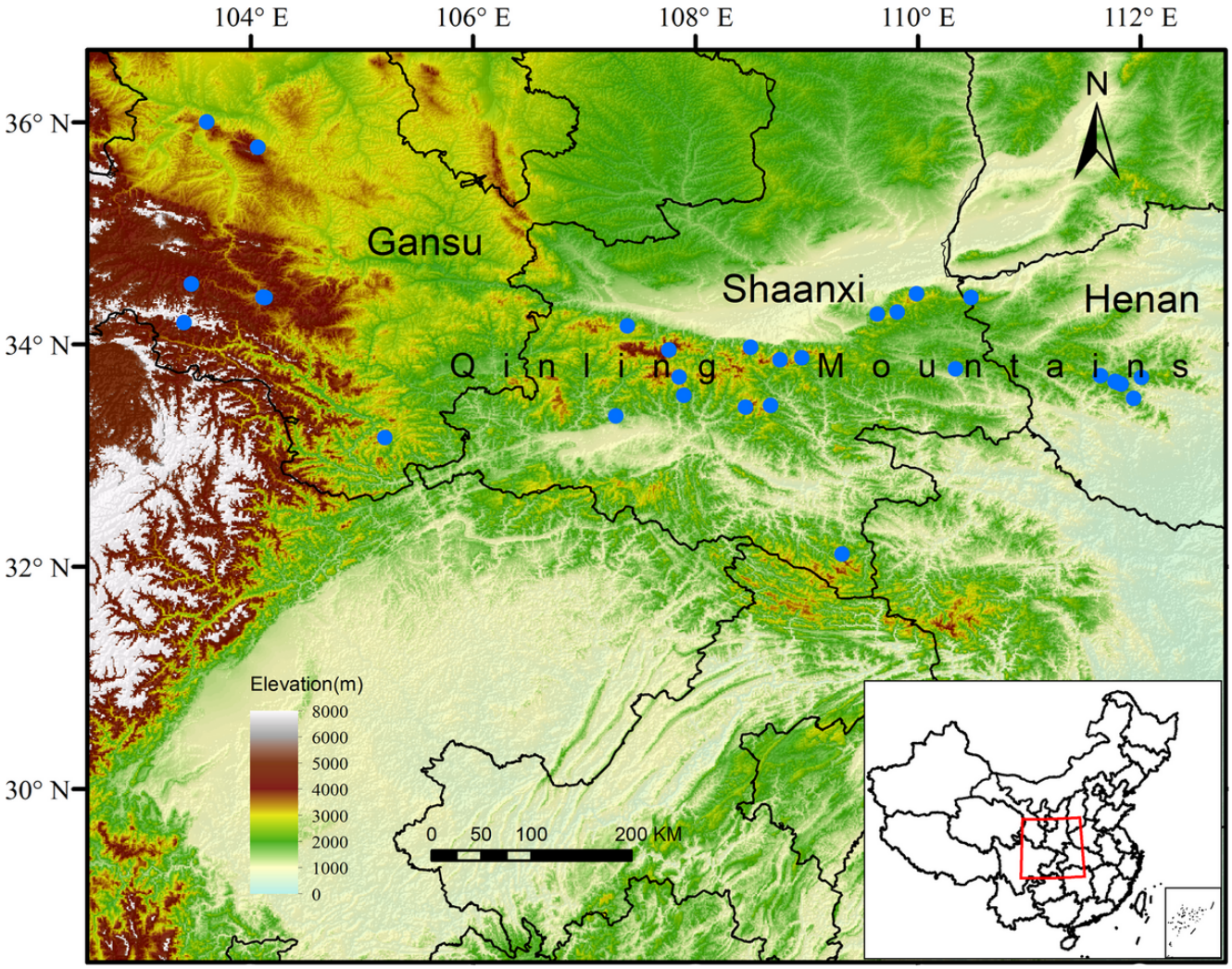

Figure 1

Geographic distribution of Rhododendron purdomii in China. The locality information was based on specimen records Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.

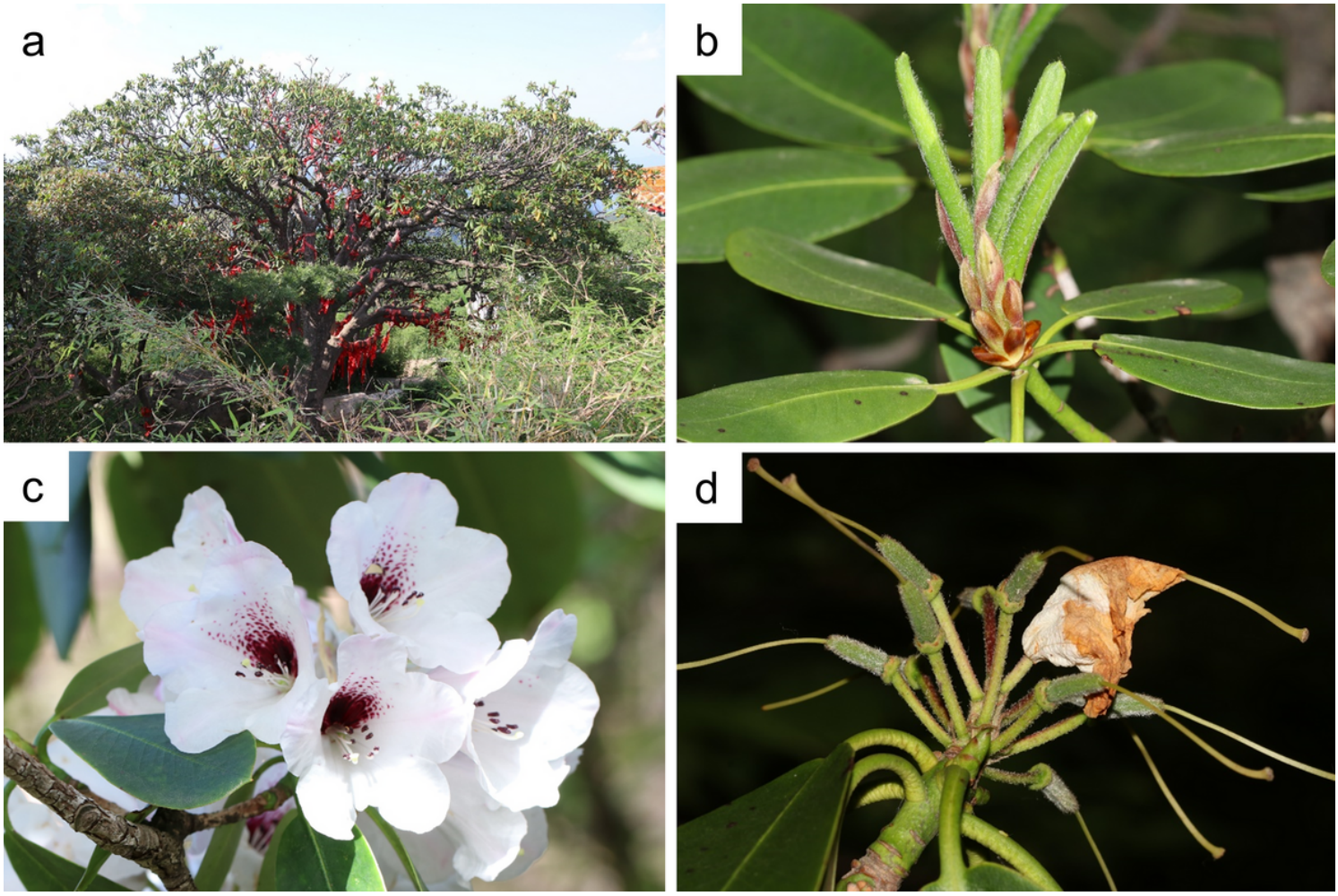

Figure 2

Photographic images of Rhododendron purdomii. a A typical individual of R. purdomii. b Leaves and leaf buds. c R. purdomii inflorescence. d Capsule of R. purdomii

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.docx