Development of polymorphic simple sequence repeat markers in *Huperzia serrata* (Lycopodiaceae)

Bin Guo1,*, Jing-yu Ren1,*, Mei-na He1,*, Kai Yao1, Tian-shu Wang1, Li-qing Wang1, Xin Liu1, Wei He1, Yan-ping Fu1, De-li Wang2, and Ya-hui Wei1,3

Manuscript received 27 February 2019; revision accepted 4 April 2019.

1Key Laboratory of Resource Biology and Biotechnology in Western China, Department of Life Science, Northwest University, Xi’an 710069, People’s Republic of China
2Hainan Branch Institute of Medicinal Plant Development, Peking Union Medical College and Chinese Academy of Medical Sciences, Haikou, Hainan 570311, People’s Republic of China
3Author for correspondence: weiyahui@nwu.edu.cn

These authors contributed equally to this work and should be considered co-first authors.

Citation: Guo, B., J.-Y. Ren, M.-N. He, K. Yao, T.-S. Wang, L.-Q. Wang, X. Liu, W. He, Y.-P. Fu, D.-L. Wang, and Y.-H. Wei. 2019. Development of polymorphic simple sequence repeat markers in *H. serrata* (Lycopodiaceae). *Applications in Plant Sciences* 7(7): e11273.

doi:10.1002/apps3.11273

**PREMISE:** The natural population size of *Huperzia serrata* (Lycopodiaceae) has dramatically decreased and the species has become endangered due to overexploitation. Here, we developed simple sequence repeat (SSR) markers for *H. serrata* to survey both its genetic diversity and population structure.

**METHODS AND RESULTS:** Based on 177 individuals, 120 SSR primer pairs were developed and optimized from five regions of the *H. serrata* transcriptomic data. Of these primer pairs, 20 were successfully amplified and 10 showed obvious polymorphism. These polymorphic loci were investigated to study the genetic diversity of *H. serrata*. Two to 11 alleles per locus were identified, the level of observed heterozygosity ranged from 0.00 to 1.00, and the level of expected heterozygosity ranged from 0.19 to 0.79. All loci were successfully amplified in *H. crispa*, *H. sutchuiana*, and *H. selago*.

**CONCLUSIONS:** The 10 polymorphic primer pairs developed here will be valuable for studies of the endangered *H. serrata* and other related species.

**KEY WORDS** *Huperzia serrata*; Lycopodiaceae; microsatellite primers; transcriptome.

*Huperzia serrata* (Thunb.) Trevis., also named qiancengta in China, is a member of the Lycopodiaceae (Almeida, 2016). More than 90% of *H. serrata* species are distributed east of the Hengduan Mountains and south of the Tsining Mountains–Huai River line, in an area with a subtropical monsoon climate. *Huperzia serrata* is a valuable medicinal plant because it contains the alkaloid huperzine A, which has been shown to effectively attenuate cognitive deficits and has been used for the treatment of Alzheimer’s disease (Lei et al., 2015). The rapidly growing demand for natural *H. serrata* and the unrestricted and continuous harvesting has led to the rapid extinction of wild resources. Therefore, many scientists have begun to conduct research on *H. serrata*, including the study of its cultivation and reproduction physiology. However, only a small number of markers (3459 expressed sequence tags) are currently available in the National Center for Biotechnology Information (NCBI), and the current research on *H. serrata* is still limited, especially with regard to molecular information.

Simple sequence repeat (SSR) markers have been widely used as effective genetic markers for plant breeding and genetic applications (Sharma et al., 2009), and have been applied to genetic diversity analysis, genetic map construction, molecular breeding, and germplasm conservation (Kumar et al., 2015). Luo et al. (2010) discovered thousands of SSR loci in *H. serrata* and selected 10 SSR sequences, including several candidate gene-encoding enzymes involved in bioactive compound biosyntheses, for further detection and verification. However, no optimum SSR loci for the study of genetic diversity and population structure in *H. serrata* have been reported. In this study, microsatellite markers were developed based on the *H. serrata* transcriptome, which will help to investigate the reproductive characteristics of *H. serrata*, evaluate its evolutionary potential, and develop a reasonable strategy for its protection, development, and utilization.

**METHODS AND RESULTS**

In this study, 177 *H. serrata* individuals were collected from the Chinese localities of Luan, Enshi, Jizhou, Hanzhong, and Jiping (Appendix 1). Total RNA was extracted from 100 mg of fresh leaves using TRIZol following the instructions of the manufacturer (TIANGEN, Beijing, China). To eliminate potential DNA contamination, we used DNase to purify total RNA following the manufacturer instructions (QIAGEN, Hilden, Germany). RNA purity and concentration were determined by NanoDrop Spectrophotometer (Qubit2.0, Agilent 2100; Shimadzu, Kyoto, Japan). mRNA was isolated using magnetic oligo (dT) beads, and then cut into short fragments using NEBNext Poly(A) mRNA Magnetic Isolation Module according to the manufacturer’s instructions (New England
### TABLE 1. Characteristics of 20 SSR markers developed for Huperzia serrata.

| Locus           | Primer sequences (5′–3′) | Repeat motif       | Allele size range (bp) | A | PIC | Putative function [Organism] | GenBank accession no. |
|-----------------|--------------------------|--------------------|------------------------|---|-----|-----------------------------|----------------------|
| c52026.graph_c0 | F: TCAAAAACCAACACTCCAACA | (AAGG)<sub>n</sub> | 138–152                | 5.6 | 6   | 0.476 Light-harvesting complex [Selaginella moellendorffii] | MH298194 |
|                 | R: TCTTCCTGCAACACCTGATA  |                    |                        |    |     |                             |                      |
| c63431.graph_c0 | F: TGCTTCTTTCTGTCTCTCTCTT | (GCT)<sub>n</sub> | 192–195                | 5.5 | 2   | 0.375 Unknown [Picea sitchensis] | MH298199 |
|                 | R: GCTGAAAGCAGGACAGCACAG |                    |                        |    |     |                             |                      |
| c51797.graph_c0 | F: CTTGTGGGAAAGGGAATAA   | (TC)<sub>n</sub>   | 190–208                | 5.8 | 9   | 0.730 Unknown [Picea sitchensis] | MH298193 |
|                 | R: GCTGAAACACAAACACAGAT  |                    |                        |    |     |                             |                      |
| c59934.graph_c0 | F: CTTGGATTGGAAAGGGAATAA | (CATC)<sub>n</sub> | 256–266                | 6.1 | 7   | 0.601 Predicted protein [Physcomitrella patens] | MH298200 |
|                 | R: AAAACGTTGCAACAGGAGG   |                    |                        |    |     |                             |                      |
| c52211.graph_c0 | F: GCACCTCTTTATCTGGGCAGG | (AG)<sub>n</sub>   | 244–256                | 5.6 | 2   | 0.313 Hypothetical protein SELMODRAFT_17213, partial [Selaginella moellendorffii] | MH298197 |
|                 | R: GTGAAGAGGGACAGGCAGAG  |                    |                        |    |     |                             |                      |
| c65171.graph_c0 | F: ATCAACGCTGCGAGCCACCTAC | (CGATCG)<sub>n</sub> | 233–257                | 5.2 | 11  | 0.723 Predicted protein [Physcomitrella patens] | MH298192 |
|                 | R: GACCGGGGTCATGAGGAGA   |                    |                        |    |     |                             |                      |
| c50318.graph_c0 | F: CCTTTATACGAGTGCGCACC  | (GT)<sub>n</sub>   | 252–270                | 6.9 | 8   | 0.618 Phosphoribosylaminomimidazole-succinocarboxamide synthase, chloroplastic isoform X1 [Musa acuminata subsp. malaccensis] | MH298198 |
|                 | R: GATACGGGCAAGGACAGCA   |                    |                        |    |     |                             |                      |
| c52257.graph_c0 | F: GCATGATAAACACTTCCGTTG | (GA)<sub>n</sub>   | 231–241                | 5.2 | 6   | 0.648 Predicted protein [Physcomitrella patens] | MH298196 |
|                 | R: GACCGGGGAAAGGCAATAGAT |                    |                        |    |     |                             |                      |
| c66382.graph_c0 | F: GCTGGCTCATGGTGCTGCTCC | (GA)<sub>n</sub>   | 231–241                | 5.2 | 6   | 0.648 Predicted protein [Physcomitrella patens] | MH298196 |
|                 | R: GATTCGCTTACCTGGGCC   |                    |                        |    |     |                             |                      |
| c60778.graph_c0 | F: GACCACTGAGAGAATGCGCA  | (GA)<sub>n</sub>   | 177–207                | 5.3 | 6   | 0.544 Hypothetical protein AMTR_s00031p00115090 [Amborella trichopoda] | MH920532 |
|                 | R: GAGTTTGCTGCTCTGGCG    |                    |                        |    |     |                             |                      |
| c38689.graph_c0 | F: GGGATCTGTCTATGAGTTACG | (GA)<sub>n</sub>   | 256–266                | 5.8 | 6   | 0.565 Unnamed protein product [Coffea canephora] | MH920819 |
|                 | R: GAATTCCTGCTGACGGGAGG  |                    |                        |    |     |                             |                      |
| c56357.graph_c0 | F: CTCTCTCTGTGACGCCTTTCT | (GA)<sub>n</sub>   | 244–256                | 5.6 | 2   | 0.313 Hypothetical protein SELMODRAFT_17213, partial [Selaginella moellendorffii] | MH298200 |
|                 | R: GTGAAGAGGGACAGGCAGAG  |                    |                        |    |     |                             |                      |
| c59441.graph_c0 | F: ATGGGAAATGACATTGATTTC | (GA)<sub>n</sub>   | 231–241                | 5.2 | 6   | 0.648 Predicted protein [Physcomitrella patens] | MH298196 |
|                 | R: GATACGGGCAAGGACAGCA   |                    |                        |    |     |                             |                      |
| c60018.graph_c0 | F: GCACAACTGGCAAACACAAA  | (GA)<sub>n</sub>   | 226–241                | 5.9 | 6   | 0.565 Unnamed protein product [Coffea canephora] | MH920819 |
|                 | R: AATACATCCAGCCGCAAGAAA|                    |                        |    |     |                             |                      |
| c61426.graph_c0 | F: AGGAAAGGGAAGGATTTTGGG | (CCAT)<sub>n</sub> | 240–260                | 6.0 | 2   | 0.565 Unnamed protein product [Coffea canephora] | MH920819 |
|                 | R: CAACCTCTCTGCTCCTCAAA  |                    |                        |    |     |                             |                      |
| c62215.graph_c0 | F: GTGCTATGCTCCAGCTGCTCC | (GCC)<sub>n</sub>  | 246–260                | 6.0 | 1   | 0.600 Hypothetical protein SELMODRAFT_17213, partial [Selaginella moellendorffii] | MH298200 |
|                 | R: TCAAGGCAACGGCTCTACTC  |                    |                        |    |     |                             |                      |
| c62350.graph_c0 | F: ACAATCGAGCTTTTGGCCTT | (GGAG)<sub>n</sub> | 214–240                | 6.0 | 1   | 0.600 Hypothetical protein SELMODRAFT_17213, partial [Selaginella moellendorffii] | MH298200 |
|                 | R: ATGGGACCTGTGCTGCGA    |                    |                        |    |     |                             |                      |
| c62412.graph_c0 | F: CTGCGAGGTTCTACCCCGCTT | (AGGC)<sub>n</sub> | 177–200                | 6.0 | 1   | 0.600 Hypothetical protein F775_13731 [Aegilops tauschii] | MH920537 |
|                 | R: CTTGAAACAGGGTGCCGCGAG |                    |                        |    |     |                             |                      |
| c63947.graph_c0 | F: CTGCACGGCAACGGGAAAATA | (CGT)<sub>n</sub>  | 244–255                | 6.0 | 1   | 0.600 Hypothetical protein JCGZ_15140 [Jatropha curcas] | MH920538 |
|                 | R: GATAGGGAATGGTGGCCAGAC |                    |                        |    |     |                             |                      |
| c64437.graph_c0 | F: CATATGCTCTCTCTCTTGC  | (GA)<sub>n</sub>   | 212–237                | 6.0 | 1   | 0.600 Hypothetical protein JCGZ_15140 [Jatropha curcas] | MH920538 |
|                 | R: ACGGAAATGACCTGTTGTTTT |                    |                        |    |     |                             |                      |

Note: A = number of alleles; PIC = polymorphism information content; T<sub>a</sub> = annealing temperature.
Biolabs, Ipswich, Massachusetts, USA). First-strand cDNA synthesis used random hexamer primers, buffer, dNTPs, and RNase H, and second-strand cDNA was synthesized by supernumerary DNA polymerase I. The total high-quality RNA was used to construct the cDNA library. Then, the cDNA library of *H. serrata* was sequenced based on synthesis by sequencing (SBS) technology using the Illumina HiSeq2500 Sequencing platform (Illumina, San Diego, California, USA). After trimming the sequencing linker and primer sequences, only 2064 loci could be used for the design and validation of primer pairs. To investigate the genetic diversity of *H. serrata*, annotated unigenes were used to identify SSRs. The identification and localization of SSR markers are valuable for the study of wild *H. serrata* and assessed their transferability in related species. The sequencing data were shown to be polymorphic (Table 3).

To analyze the genetic diversity of *H. serrata*, annotated unigenes were used to identify SSRs. The identification and localization of SSRs were performed using the MicroSatellite Identification Tool (MISA; Thiel et al., 2003). A total of 4395 SSR loci were found by MISA in 3685 unigenes (24.7%), which was higher than previously reported for *H. serrata* (Luo et al., 2010) and for bryophytes such as *Physcomitrella patens* (Hedw.) Bruch & Schimp. (Kobayashi and Morita, 2005). A relatively high frequency of repeats with di- and trinucleotides was detected in *H. serrata* (6.3%) (Kobayashi and Morita, 2005). A relatively high frequency of repeats with di- and trinucleotides was detected in *H. serrata* (6.3%) (Kobayashi and Morita, 2005).

Due to short flanking sequences of the SSR loci or inappropriate sequences, only 2064 loci could be used for the design and validation of primer pairs. To investigate the genetic diversity of *H. serrata*, 120 SSR makers were randomly selected and synthesized. DNA amplification was performed with Ex Taq (TaKaRa Biotechnology Co., Beijing, China) following the manufacturer's script and unigenes. Furthermore, reads were divided into 25-bp (k-mer) segments using Trinity software (Grabherr et al., 2011). The final assembly was composed of 111,251 unigenes and had an N50 size of 997 bp.

The 20 primer pairs were used to evaluate the polymorphism information in five populations of *H. serrata*, identifying a total of 72 alleles. The polymorphism information content (PIC) value for SSR markers ranged from 0.313 (primer c52211) to 0.730 (primer c51797) with an average of 0.568; this value was higher than that found in other plants (e.g., *Sesamum indicum* L. [Cho et al., 2011]). The level of expected heterozygosity of the genetic diversity ranged between 0.06 and 0.79 and the level of observed heterozygosity ranged from 0.00 to 1.00 (Table 2). Eight SSR markers had a high value for SSR markers ranging from 0.313 (primer c52211) to 0.730 (primer c51797) with an average of 0.568; this value was higher than that found in other plants (e.g., *Sesamum indicum* L. [Cho et al., 2011]). The level of expected heterozygosity of the genetic diversity ranged between 0.06 and 0.79 and the level of observed heterozygosity ranged from 0.00 to 1.00 (Table 2). Eight SSR markers had a high value for SSR markers ranging from 0.313 (primer c52211) to 0.730 (primer c51797) with an average of 0.568; this value was higher than that found in other plants (e.g., *Sesamum indicum* L. [Cho et al., 2011]). The level of expected heterozygosity of the genetic diversity ranged between 0.06 and 0.79 and the level of observed heterozygosity ranged from 0.00 to 1.00 (Table 2).

### Table 3. Cross-amplification (showing allele size range in base pairs) of the 10 microsatellites developed for *Huperzia serrata* in *H. crispata*, *H. sutchueniana*, and *H. selago.*

| Locus          | *H. crispata*  (N = 5) | *H. sutchueniana*  (N = 5) | *H. selago*  (N = 5) |
|---------------|------------------------|---------------------------|----------------------|
|                | H                | H                | H                   |
| c52211.graph_c0 | 247–254             | 248–254             | 254–255              |
| c5026.graph_c0  | 137–154             | 138–151             | 176–184              |
| c63411.graph_c0 | 193–197             | 188–197             | 193–197              |
| c51797.graph_c0 | 194–207             | 197–200             | 193–209              |
| c59934.graph_c0 | 257–261             | 257–265             | 257–270              |
| c65171.graph_c0 | 234–246             | 234–235             | 234–239              |
| c50318.graph_c0 | 248–278             | 250–279             | 248–264              |
| c52257.graph_c0 | 230–235             | 231–239             | 231–241              |
| c66382.graph_c0 | 180–184             | 176–184             | 176–184              |
| c60778.graph_c0 | 185–188             | 187–205             | 187–225              |

*Note: N = number of individuals sampled.*

10 polymorphic primer pairs were used to evaluate the polymorphism information in five populations of *H. serrata*, identifying a total of 72 alleles. The polymorphism information content (PIC) value for SSR markers ranged from 0.313 (primer c52211) to 0.730 (primer c51797) with an average of 0.568; this value was higher than that found in other plants (e.g., *Sesamum indicum* L. [Cho et al., 2011]). The level of expected heterozygosity of the genetic diversity ranged between 0.06 and 0.79 and the level of observed heterozygosity ranged from 0.00 to 1.00 (Table 2). Eight SSR markers had a high value for SSR markers ranging from 0.313 (primer c52211) to 0.730 (primer c51797) with an average of 0.568; this value was higher than that found in other plants (e.g., *Sesamum indicum* L. [Cho et al., 2011]). The level of expected heterozygosity of the genetic diversity ranged between 0.06 and 0.79 and the level of observed heterozygosity ranged from 0.00 to 1.00 (Table 2). Eight SSR markers had a high value for SSR markers ranging from 0.313 (primer c52211) to 0.730 (primer c51797) with an average of 0.568; this value was higher than that found in other plants (e.g., *Sesamum indicum* L. [Cho et al., 2011]). The level of expected heterozygosity of the genetic diversity ranged between 0.06 and 0.79 and the level of observed heterozygosity ranged from 0.00 to 1.00 (Table 2).

Cross-species amplification of 10 microsatellite primers was tested on DNA extracts in three related species: *H. crispata* (Ching) Ching, *H. sutchueniana* (Herter) Ching, and *H. selago* (L.) Bernh. Ten loci were successfully amplified in three related species and were shown to be polymorphic (Table 3).

### CONCLUSIONS

This study successfully developed 10 polymorphic primers from *H. serrata* and assessed their transferability in related species. The selected polymorphic microsatellites are valuable for the study of wild *H. serrata* resources with regard to its genetic diversity, population structure, and evolution.
ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China (no. 31702159, 81303159, 31572665), the Key Research and Development Plan Project of Shaanxi Province (2018ZDXXM-SF-016), the Natural Science Foundation Research Project of Shaanxi Province (grant no. 2016JM3002, 2018JQ3029), the Research Program of Key Laboratory of Shaanxi Education Department (18JS111), and the Bureau of Science and Technology in Shaanxi Province, Xi’an City, Beilin District (GX1702).

DATA ACCESSIBILITY

The raw sequence data reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession number: SRR8402085). Sequence information of the developed primers has been deposited in NCBI’s GenBank, and accession numbers are provided in Table 1.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Summary of di- and trinucleotide repeats in Huperzia serrata.

APPENDIX 1. Voucher and location information for Huperzia species used in this study.

| Species                        | Population code | N   | Voucher no. | Location | Geographical coordinates | Elevation (m) |
|-------------------------------|-----------------|-----|-------------|----------|--------------------------|---------------|
| Huperzia serrata (Thunb.) Trevis. | AL              | 19  | NWUHS1001   | Luan, Anhui, China | 31°28’N, 116°12’E | 150            |
|                               | HE              | 55  | NWUHS1002   | Enshi, Hubei, China | 30°5’N, 109°11’E | 880            |
|                               | HJ              | 40  | NWUHS1003   | Jizhou, Hubei, China | 30°08’N, 112°04’E | 330            |
|                               | SH              | 33  | NWUHS1004   | Hanzhong, Shaanxi, China | 32°30’N, 107°09’E | 840            |
|                               | YJ              | 30  | NWUHS1005   | Jinping, Yunnan, China | 22°54’N, 103°19’E | 1420           |
| Huperzia crispata (Ching) Ching | HC              | 5   | NWUHC1001   | Jizhou, Hubei, China | 30°08’N, 112°04’E | 330            |
| Huperzia selago (L.) Bernh.    | PS              | 5   | NWUHS3001   | Longyan, Fujian, China | 24°23’N, 115°51’E | 460            |
| Huperzia sutchueniana (Herter) Ching | HS          | 5   | NWUHS2001   | Shizhu, Chongqing, China | 27°29’N, 108°39’E | 1500           |

Note: N = number of individuals analyzed.

*The samples were stored in the Key Laboratory of Resource Biology and Biotechnology in Western China, Department of Life Science, Northwest University.

APPENDIX S1. Summary of di- and trinucleotide repeats in Huperzia serrata.

LITERATURE CITED

Almeida, T. E. 2016. A community-derived classification for extant lycophytes and ferns. Journal of Systematics and Evolution 54: 563–603.

Cho, Y. I., J. H. Park, C. W. Lee, W. H. Ra, J. W. Chung, J. R. Lee, K. H. Ma, et al. 2011. Evaluation of the genetic diversity and population structure of sesame (Sesamum indicum L.) using microsatellite markers. Genes and Genomics 33: 187–195.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644.

Kobayashi, J. I., and H. Morita. 2005. The Lycopodium alkaloids. In G. A. Cordell [ed.]. The alkaloids: Chemistry and biology, vol. 61, 1–57. Academic Press, Salt Lake City, Utah, USA.

Kumar, B., U. Kumar, and H. K. Yadav. 2015. Identification of EST–SSRs and molecular diversity analysis in Mentha piperita. Crop Journal 3: 335–342.

Lei, Y., L. Yang, C. Y. Ye, M. Y. Qin, H. Y. Yang, H. L. Jiang, X. C. Tang, and H. Y. Zhang. 2015. Involvement of intracellular and mitochondrial Aβ in the ameliorative effects of Huperzine A against oligomeric Aβ42-induced injury in primary rat neurons. PLoS ONE 10(5): e0128366.

Luo, H., L. Ying, S. Chao, Q. Wu, J. Song, Y. Sun, A. Steinmetz, and S. Chen. 2010. Comparison of 454-ESTs from Huperzia serrata and Phlegmariurus carinatus reveals putative genes involved in lycopodium alkaloid biosynthesis and developmental regulation. BMC Plant Biology 10: 209.

Sharma, R. K., P. Bhardwaj, R. Negi, T. Mohapatra, and P. S. Ahuja. 2009. Identification, characterization and utilization of unigene derived microsatellite markers in tea (Camellia sinensis L.). BMC Plant Biology 9: 53.

Thiel, T., W. Michalek, R. 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.