Applications in Plant Sciences

PRIMER NOTE

Development of 16 single-copy nuclear gene markers for *Oreocharis auricula*, a perennial herb in China

Zhixia Sun[1,2], Chao Feng[1], Lihua Yang[1], and Hanghui Kong[1,3]

Manuscript received 18 August 2019; revision accepted 25 September 2019.

[1] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, People's Republic of China

[2] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, People's Republic of China

[3] Author for correspondence: konghh@scbg.ac.cn

Citation: Sun, Z., C. Feng, L. Yang, and H. Kong. 2019. Development of 16 single-copy nuclear gene markers for *Oreocharis auricula*, a perennial herb in China. *Applications in Plant Sciences* 7(11): e11304.

doi:10.1002/aps3.11304

PREMISE: We developed single-copy nuclear DNA markers for a perennial herb, *Oreocharis auricula* (Gesneriaceae), to help infer the evolutionary history of the genus *Oreocharis* in subtropical China.

METHODS AND RESULTS: We screened 834 putative single-copy nuclear genes from transcriptomes of 11 species of *Primulina* using the Illumina HiSeq 2000 platform. Based on the screening results, we developed 16 primer sets for accessions of *O. auricula* representing three wild populations. The number of alleles per locus ranged from three to 34. The levels of expected and observed heterozygosity varied from 0.000 to 0.372 and 0.000 to 0.650, respectively. The markers were successfully cross-amplified in the related species *O. magnidens*, *O. speciosa*, *O. maximowiczii*, and *O. henryana*.

CONCLUSIONS: The newly developed markers will facilitate further studies on genetic diversity and phylogeographic structure throughout the distributional range of *O. auricula*. Additionally, these markers may be useful for other related species in Gesneriaceae.

KEY WORDS genetic diversity; Gesneriaceae; Illumina; *Oreocharis auricula*; transcriptome.

Oreocharis auricula (S. Moore) C. B. Clarke is a perennial herb in the Gesneriaceae family. It primarily grows on shaded and damp rocks by streams, in valleys, or on forested slopes, and is distributed across a broad elevational range of 200–1800 m mostly within subtropical China (Wang et al., 1998). The widespread geographic distribution of *O. auricula* makes it a suitable model to investigate phylogeographic structure and contribute to a better understanding of the evolutionary history of plant species in subtropical China.

Oreocharis auricula is widely regarded as a valuable source of natural essential oil, and irritating volatile oil compounds can be extracted from its stems and leaves (Zhu et al., 2004). Traditionally, this species has been used in folk herbal medicine (known as yanbaicai or yantongcao) in China to detoxify and attain hemostasis, especially in treating bruises, acne, swelling, and pain (Feng, 2015). Therefore, it is important to ensure the sustainable utilization of the wild germplasm resources of *O. auricula*.

Here, we developed 16 primer sets based on 834 putative single-copy nuclear genes for *O. auricula* to establish tools for investigating its evolutionary and phylogeographic history, as well as to advance molecular studies of the genus and the Gesneriaceae family. We screened these single-copy nuclear gene markers from transcriptomes of 11 *Primulina* Hance species using the Illumina HiSeq 2000 platform (Ai et al., 2015). We successfully used these primers to assess polymorphism among three wild populations of *O. auricula*, and we cross-amplified the primers in *O. henryana* Oliv., *O. magnidens* Chun ex K. Y. Pan, *O. maximowiczii* C. B. Clarke, and *O. speciosa* (Hemsl.) Mich. Möller & W. H. Chen. Consequently, we propose that the 16 primer sets for the single-copy nuclear genes can be used in further investigations of genetic diversity and population structure of *O. auricula* and in related species within the genus and family.

METHODS AND RESULTS

We collected fresh leaves during the early growth stages of samples representing 60 individuals from three populations of *O. auricula* in subtropical China (20 from Anfu, Jiangxi; 20 from Hefeng, Hubei; and 20 from Wuyi Mountain, Fujian). Additionally, we sampled 10 individuals of *O. henryana*, three individuals of *O. speciosa*, two individuals of *O. magnidens*, and six individuals of *O. maximowiczii* (Appendix 1). We dried all samples in silica gel and used the dried material for genomic DNA extraction following a modified 3% cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1987). We screened single-copy nuclear genes from transcriptomes of 11 species of *Primulina* sequenced using RNA-Seq in our group's

TABLE 1. Characteristics of 16 single-copy nuclear gene markers developed in *Oreocharis auricula*.

Locus[a]	Primer sequences (5′–3′)	Allele size (bp)	BLASTN top hit description	*E*-value	GenBank accession no.[b]
47FR	F: CCAGAGCACCAACCTTAC R: AATCCCACCTTGAGATAGTA	194	Protein PAM68 [*Sesamum indicum*]	2e-17	MK710715–MK710734 MK710870–MK71088 MK711167–MK711186 MK981648–MK981658
68FR	F: ACGTGCAGGAAATGTACAG R: GTTTATCGTTGGCACTGGC	476	Ubiquinone biosynthesis monooxygenase Coq6 [*Vigna unguiculata*]	9e-71	MK709894–MK709913 MK710049–MK710067 MK710330–MK710347 MK981659–MK981669
72FR	F: CCTTCTGACCGTAGGAAAG R: AAAGCCCACAATACTTGAGC	427	RAP domain-containing protein [*Sesamum indicum*]	2e-136	MK706534–MK706553 MK706689–MK706708 MK706975–MK706993 MK981670–MK981680
79FR	F: TACAAGGGGGCGTTTGATG R: TAGCCGAAAAGCATGCACTG	389	Protein MITOFERRINLIKE 1 [*Sesamum indicum*]	1e-119	MK981701–MK981751 MN225011–MN225050
80FR	F: AGCATCAGGTTTCCGGATC R: GCTGCATATCCATTAAGC	845	Peptide methionine sulfoxide reductase A5 [*Sesamum indicum*]	2e-31	MN225051–MN225131
92FR	F: AGAACAAGCTCTCCCTCTCC R: ACCACCTCCTTAAAATAGCC	600	Pentatricopeptide repeat-containing protein At1g74850 [*Sesamum indicum*]	1e-179	MN225132–MN225212
96FR	F: GCTAAACAGGGATGGAAACC R: TTCTGAAGTTTCACCACATC	795	Protein CLP1 homolog [*Sesamum indicum*]	3e-48	MN225213–MN225293
97FR	F: TCGTCGACCTGAAACAAACC R: ATGCATCCACGAAATCTTCAG	697	Uncharacterized protein 97FR [*Primulina bicolor*]	0	MN225294–MN225374
100FR	F: CGATCATAACCGAAGAACC R: TGCCCATTTCTTCACTACC	502	Protein AUXIN RESPONSE 4 [*Sesamum indicum*]	1e-153	MK716403–MK716422 MK716560–MK716579 MK716845–MK716864 MK981752–MK981762
107FR	F: TTGCCTTGGTTTAGACTGAC R: TTACAACCGAATCCTTCATC	716	Protein ROOT PRIMORDIUM DEFECTIVE 1 [*Sesamum indicum*]	4e-167	MN225375–MN225455
109FR	F: GGTATGTCTCAGGGCATCAC R: CATCCCACACATTTCACA	438	Rhomboid-like protein 14, mitochondrial [*Sesamum indicum*]	4e-63	MN225456–MN225536
111FR	F: TCGATTTCTCCGTCAGCG R: CATTCACCTCAATTTGCC	743	Shewanella-like protein phosphatase 2 [*Sesamum indicum*]	0	MK708235–MK708254 MK708391–MK708410 MK708676–MK708695 MK981763–MK981773
112FR	F: CTCAAAGTGGATTGAGAAGCC R: CAGCCAAACTTCTCCACAC	570	RINT1-like protein MAG2 [*Sesamum indicum*]	5e-170	MK981792–MK98184 MN225537–MN225576
115FR	F: TCCAGCAGAATTGGCATTAG R: ACTCTGGAGCAAGATATCC	340	C-type lectin receptor-like tyrosine-protein kinase At1g52310 [*Sesamum indicum*]	1e-117	MK711565–MK711584 MK711723–MK711742 MK712005–MK712024 MK981843–MK981853
117FR	F: ATGAGATTTGGGATAAGCCAGC R: AGGTTCTACCATTTCCAGAAGC	584	Rhomboid protein 2 gene [*Primulina suichuanensis*]	0	MK709250–MK709290 MK709419–MK709437 MK709706–MK709725 MK981854–MK981864
165FR	F: TCTTCGCTGCCTTCTCTCTC R: TCTCCATACTCGTTCACTTCCTC	644	Uncharacterized protein 165FR [*Primulina lungzhouensis*]	0	MK707382–MK707401 MK707539–MK707558 MK707826–MK707845 MK981865–MK981875

[a]An annealing temperature of 55°C was used for all loci.
[b]Multiple samples were sequenced for each locus.

previous study (Ai et al., 2015). The previously obtained data set comprised 834 putative single-copy nuclear genes with alignment lengths from 129 to 2847 bp. Compared to organellar genes, nuclear genes are useful for phylogenetic and phylogeographic inference because of their rapid evolutionary rates, and single-copy genes circumvent analytical complications that arise from paralogy (Sang, 2002; Mort and Crawford, 2004).

The procedure for detecting the putative single-copy nuclear genes is detailed in Ai et al. (2015). Based on the 834 putative single-copy nuclear genes in the transcriptomes of *Primulina* species, we randomly selected 74 of them and initially designed 115 PCR primers in Primer3 version 2.3.4 (Rozen and Skaletsky, 1999). The primers were designed using the following criteria: length of PCR product of 300–900 bp, located within exon sequences, primer melting temperatures between 50–55°C. Preliminarily, we screened 115 PCR primers for amplification using eight randomly selected samples of *O. auricula*. After optimization, we selected 16 primer pairs that showed amplification of a single clear band, were polymorphic, and cross-amplified in the related species to estimate genetic diversity in all 60 individuals of *O. auricula*. We performed

TABLE 2. Genetic characterization of 16 single-copy nuclear gene markers across three populations of *Oreocharis auricula*.[a]

	Anfu (*N* = 20)				Hefeng (*N* = 20)				Wuyi Mountain (*N* = 20)			
Locus	*A*	H_e	H_o	HWE[b]	*A*	H_e	H_o	HWE[b]	*A*	H_e	H_o	HWE[b]
47FR	8	0.045	0.050	0.096 ns	8	0.063	0.125	0.000***	8	0.235	0.331	0.146 ns
68FR	5	0.089	0.125	0.189 ns	5	0.000	0.000	NA	5	0.372	0.650	0.003**
72FR	7	0.132	0.168	0.075 ns	7	0.097	0.171	0.069 ns	7	0.090	0.121	0.061 ns
79FR	14	0.145	0.168	0.474 ns	14	0.272	0.429	0.063*	14	0.040	0.045	0.101 ns
80FR	23	0.166	0.237	0.116 ns	23	0.114	0.209	0.027***	23	0.132	0.233	0.026 ns
92FR	3	0.258	0.333	0.413 ns	3	0.167	0.333	0.000***	3	0.000	0.000	NA
96FR	8	0.040	0.050	0.033 ns	8	0.036	0.044	0.043 ns	8	0.339	0.644	0.065 ns
97FR	8	0.057	0.069	0.123 ns	8	0.063	0.125	0.000***	8	0.337	0.581	0.035 ns
100FR	8	0.174	0.241	0.056 ns	8	0.000	0.000	NA	8	0.194	0.338	0.065 ns
107FR	6	0.164	0.208	0.122 ns	6	0.066	0.075	0.191 ns	6	0.253	0.433	0.312 ns
109FR	3	0.125	0.167	0.045 ns	3	0.000	0.000	NA	3	0.313	0.550	0.010*
111FR	11	0.182	0.268	0.104 ns	11	0.048	0.093	0.085 ns	11	0.100	0.118	0.169 ns
112FR	6	0.146	0.250	0.023 ns	6	0.271	0.417	0.068 ns	6	0.102	0.146	0.091 ns
115FR	25	0.238	0.340	0.296 ns	25	0.202	0.282	0.325 ns	25	0.211	0.296	0.187 ns
117FR	34	0.186	0.237	0.357 ns	34	0.185	0.231	0.366 ns	34	0.153	0.194	0.389 ns
165FR	6	0.148	0.188	0.101 ns	6	0.158	0.258	0.000***	6	0.123	0.213	0.052 ns

Note: *A* = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity; HWE = Hardy–Weinberg equilibrium; *N* = number of individuals sampled.
[a]Voucher and locality information are provided in Appendix 1.
[b]Deviations from HWE were not statistically significant (ns) and statistically significant (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$); NA = not applicable (i.e., monomorphic locus).

PCR amplifications in 25-μL volumes consisting of 12.5 μL 2× *Taq* PCR Master Mix (Tsingke Biotech, Guangzhou, China), 1 μL of each primer, 1 μL of template DNA (20–40 ng), and 9.5 μL of ddH_2O. For all primers, the PCR thermocycling procedure comprised enzyme activation of 3 min at 94°C; followed by 38 cycles of 30 s at 94°C, 30 s at 55°C, and 90 s at 72°C; ending with a final extension of 10 min at 72°C. All PCR products were checked by 1% agarose gel electrophoresis, which we ran for 15 min at 120 V. The PCR products were sequenced by Tsingke Biotech. Using the same PCR conditions, we cross-amplified the genes within *O. henryana*, *O. magnidens*, *O. maximowiczii*, and *O. speciosa* (Appendix 1). All 16 markers showed successful cross-amplification in all four species.

The PCR primer pairs, characteristics of the 16 newly developed single-copy nuclear DNA markers, and GenBank accession numbers of the resulting sequences are presented in Table 1. From the sequences that we obtained, the alleles of the heterozygotes were determined by using PHASE version 2.2.1 with default settings (phase threshold = 90%, iteration steps = 100, thinning interval = 1, burn-in steps = 100) (Stephens et al., 2001). The input and output formats for PHASE were converted by using SeqPhase version 2.1.1 (Flot, 2010), and the phased sequences were converted from FASTA file format to GENEPOP format by using PGDSpider version 2.1.1.5 (Lischer and Excoffier, 2012). We used GENEPOP version 4.2 (Rousset, 2008) to calculate the probability of deviation from Hardy–Weinberg equilibrium, and we calculated basic genetic variation (e.g., the number of alleles per locus, expected heterozygosity, and observed heterozygosity) in GenAlEx version 6.5 (Peakall and Smouse, 2012).

The number of alleles per locus ranged from three to 34. The levels of expected and observed heterozygosity varied from 0.000 to 0.372 and 0.000 to 0.650, respectively (Table 2). We measured the genetic diversity of all samples derived from the pairwise number of site differences, including nucleotide diversity (π), haplotype diversity (h),

TABLE 3. Genetic diversity statistics for the 16 newly developed single-copy nuclear gene markers developed for *Oreocharis auricula*, as measured from *O. auricula*, *O. magnidens*, *O. speciosa*, *O. maximowiczii*, and *O. henryana*.

Locus	*n*	Sequence length (bp)	*S*	*k*	*h*	$\pi \times 10^{-3}$	$\theta_w \times 10^{-3}$	*D*	Fu's F_s
47FR	162	194	45	3.908	0.900	20.46	41.61	−1.84739	−8.12
68FR	156	476	55	3.285	0.767	7.47	22.23	−2.11234	−3.603
72FR	162	427	48	6.622	0.919	15.51	19.85	−0.93128	−1.535
79FR	162	389	44	6.571	0.934	16.89	19.98	−0.62866	−6.138
80FR	162	845	150	21.300	0.954	32.37	40.26	−0.83273	2.270
92FR	162	600	37	4.693	0.782	7.82	10.89	−0.89054	1.825
96FR	162	795	70	14.732	0.849	30.25	25.39	0.30314	14.758
97FR	162	697	108	10.221	0.909	14.66	27.37	−1.57337	−1.58
100FR	162	502	74	12.085	0.896	24.27	26.25	−0.51704	0.756
107FR	162	716	81	8.115	0.910	11.41	20.12	−1.41621	−3.672
109FR	162	438	64	7.307	0.840	16.76	25.93	−1.21078	−0.456
111FR	162	743	73	7.775	0.915	10.49	17.40	−1.42475	−4.717
112FR	162	570	87	8.691	0.931	15.25	26.96	−1.44096	−2.071
115FR	162	340	53	12.634	0.979	37.27	27.61	0.91681	−26.412
117FR	158	584	74	14.281	0.993	24.45	22.48	−0.09434	−33.238
165FR	162	644	79	8.639	0.932	13.48	21.77	−1.39153	−8.094

Note: *D* = Tajima's D; *h* = haplotype diversity; *k* = average number of pairwise nucleotide differences; *n* = number of sequences for analysis (consists of four sequences of *O. magnidens*, six sequences of *O. speciosa*, 12 sequences of *O. maximowiczii*, 20 sequences of *O. henryana*, and 120 sequences of *O. auricula*); π = average number of pairwise nucleotide differences per site calculated based on all sites; *S* = number of polymorphic (segregating) sites; θ_w = Watterson estimator per site from *S*.

Watterson estimator per site from S (θ_w), Tajima's D, and Fu's F_s (Table 3) using DnaSP version 5.10 (Librado and Rozas, 2009). We found that the average h was 0.901 (0.767–0.993), π was 18.68×10^{-3} (7.47×10^{-3} to 37.27×10^{-3}), θ_w was 24.76×10^{-3} (10.89×10^{-3} to 41.61×10^{-3}), Tajima's D was −0.94 (−2.11–0.92), and Fu's F_s was −5.00 (−33.24–14.76). The Tajima's D values were not significant, but most populations showed negative values for Tajima's D and Fu's F_s.

CONCLUSIONS

We successfully developed new primers for and amplified 16 single-copy nuclear DNA markers in three populations of *O. auricula*. We determined alleles of the heterozygous sequences based on degenerate bases in PHASE without cloning. These newly developed genetic resources will be useful in further analyses of genetic diversity and population structure at a broad geographic scale to elucidate the demographic history of *O. auricula* and to inform its sustainable use and conservation. Moreover, successful cross-amplification of these single-copy nuclear genes with close congeners of *O. auricula* suggests potential applications for species-level phylogenetic inference and related evolutionary analyses.

ACKNOWLEDGMENTS

The authors thank Dr. A. J. Harris (Oberlin College, Oberlin, Ohio, USA) for her critical reading of the manuscript and English editing. This study was supported by grants from the National Natural Science Foundation of China (31970231, 31900178).

DATA AVAILABILITY

Raw sequence reads in Ai et al. (2015) were deposited to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA; accession no. *Primulina eburnea*: SRR1184287, *P. fimbrisepala*: SRR1184436, *P. swinglei*: SRR1184437, *P. pteropoda*: SRR1184438, *P. huaijiensis*: SRR1198999, *P. lobulata*: SRR1199001, *P. sinensis*: SRR1199003, *P. lutea*: SRR1199004, *P. heterotricha*: SRR1199005, *P. tabacum*: SRR1199007, *P. villosissima*: SRR1199008). Assembled single-copy gene sequences have been deposited in the Dryad repository (https://doi.org/10.5061/dryad.h738s; Ai et al., 2014). Sequence information for the developed primers in this study has been deposited to NCBI; GenBank accession numbers are provided in Table 1.

LITERATURE CITED

Ai, B., Y. Gao, X. Zhang, J. Tao, M. Kang, and H. Huang. 2014. Data from: Comparative transcriptome resources of eleven *Primulina* species, a group of "stone plants" from a biodiversity hotspot. Dryad Repository. https://doi.org/10.5061/dryad.h738s.

Ai, B., Y. Gao, X. Zhang, J. Tao, M. Kang, and H. Huang. 2015. Comparative transcriptome resources of eleven *Primulina* species, a group of "stone plants" from a biodiversity hot spot. *Molecular Ecology Resources* 15: 619–632.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Feng, C. Y. 2015. A phylogenetic study on medicinal plant *Oreocharis* and its related species. Master's dissertation, Henan Agricultural University, Zhengzhou, Henan, China (in Chinese).

Flot, J.-F. 2010. SeqPhase: A web tool for interconverting PHASE input/output files and FASTA sequence alignments. *Molecular Ecology Resources* 10: 162–166.

Librado, P., and J. Rozas. 2009. DnaSP v5: A sofware for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451.

Lischer, H. E. L., and L. Excoffier. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28: 298–299.

Mort, M. E., and D. J. Crawford. 2004. The continuing search: Low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon* 53: 257–261.

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.

Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* 8: 103–106.

Rozen, S., and H. Skaletsky. 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.

Sang, T. 2002. Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology* 37: 121–147.

Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* 68: 978–989.

Wang, W. T., K. Y. Pan, and Z. Y. Li. 1998. Gesneriaceae. *In* Z. Y. Wu and P. H. Raven [eds.], Flora of China, Vol. 18. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.

Zhu, S. Y., J. Hou, X. Z. Zheng, Y. D. Yu, and G. L. Zou. 2004. Chemical composition of essential oil of *Oreocharis auricula*. *Amino Acids and Biotic Resources* 26: 8–9.

APPENDIX 1. Voucher and location information for populations of *Oreocharis auricula* and four congeners used for cross-amplification in this study.

Species	Voucher specimen accession no.[a]	Collection locality	Latitude (°N)	Longitude (°E)	*N*
Oreocharis auricula (S. Moore) C. B. Clarke	YLH470	Anfu, Jiangxi	27.0481	114.3694	20
O. auricula	YLH527	Hefeng, Hubei	30.0614	110.0675	20
O. auricula	FJWYS	Wuyishan, Fujian	27.6890	117.6511	20
O. magnidens Chun ex K. Y. Pan	YLH471	Anfu, Jiangxi	27.5223	114.2480	2
O. speciosa (Hemsl.) Mich. Möller & W. H. Chen	YLH501	Huaihua, Hunan	27.7400	110.5786	3
O. maximowiczii C. B. Clarke	YLH454	Yong'an, Fujian	26.0239	117.4169	6
O. henryana Oliv.	YLH570	Wanzhou, Chongqing	30.5153	108.2839	10

Note: N = number of individuals.

[a]One voucher was collected for each population used; all voucher specimens were deposited in the South China Botanical Garden Herbarium (IBSC), Guangzhou, Guangdong, China. All vouchers were collected by Lihua Yang and Hanghui Kong.