

DEVELOPMENT OF MICROSATELLITE MARKERS USING ILLUMINA MISEQ SEQUENCING TO CHARACTERIZE *EPHEDRA GERARDIANA* (EPHEDRACEAE)¹

Ji DE^{2,3}, WEIDONG ZHU³, TIANMENG LIU³, ZHE WANG², AND YANG ZHONG^{2,3,4}

²School of Life Sciences, Fudan University, Shanghai 200438, People's Republic of China; and ³Science Faculty, Tibet University, Lhasa 850000, Tibet Autonomous Region, People's Republic of China

- **Premise of the study:** *Ephedra gerardiana* (Ephedraceae), occurring in the Himalayan ranges, is an important plant species used in Tibetan medicine. Due to the lack of molecular markers to characterize genetic diversity, knowledge for conservation and uses of *E. gerardiana* resources is limited; we therefore developed microsatellite markers for use in this species.
- **Methods and Results:** Using Illumina MiSeq sequencing technology, we developed 29 polymorphic microsatellite loci suitable for *E. gerardiana*, of which 15 loci also showed polymorphisms in two related *Ephedra* species, *E. saxatilis* and *E. monosperma*. The average number of effective alleles per locus ranged from two to six. The observed and expected heterozygosity ranged from 0.23 to 0.83 and 0.44 to 0.86, respectively, in *E. gerardiana* populations.
- **Conclusions:** The developed 29 microsatellite markers are effective for the study of genetic structure and genetic diversity of *E. gerardiana*, and 15 of these markers are suitable for related *Ephedra* species.

Key words: conservation; *Ephedra gerardiana*; Ephedraceae; genetic diversity; next-generation sequencing; simple sequence repeat markers.

The genus *Ephedra* L. (Ephedraceae), also referred to as ma huang in Chinese, contains species that are sources of important Chinese traditional and Tibetan medicines (Konno et al., 1985). *Ephedra gerardiana* Wall. ex C. A. Mey. is distributed at altitudes above 3900 m in the Himalayan ranges (Editorial Committee of Chinese Flora, 1978). Because it is both a drought-resistant plant species (Shen, 1995) and an important Tibetan medicine (Pandey, 2006), *E. gerardiana* is useful for the study of the adaptive evolution and maintenance of genetic diversity in *Ephedra*. Relatively high genetic differentiation and variation are found for plants distributed on the Qinghai–Tibet Plateau due to its great geographical variability, in addition to frequent natural hybridization and polyploidization events (Wen et al., 2014). Chloroplast fingerprints revealed a high level of genetic differentiation as reflected by high levels of genetic diversity (F_{ST}) among populations of *Ephedra* species on the Qinghai–Tibet Plateau, such as in *E. gerardiana* (0.98) and *E. saxatilis* (Stapf) Royle ex Florin (0.86) (Qin et al., 2013).

Microsatellite (also referred to as simple sequence repeat [SSR]) markers are widely used in studies of plant population genetics

and genetic diversity because of their high levels of polymorphism, stability, and codominance. However, traditional methods to develop microsatellite markers are time-consuming and complex. High-throughput and low-cost next-generation sequencing has accelerated the identification of large numbers of microsatellite markers (Rico et al., 2013). In this paper, we report the development of microsatellite markers for *Ephedra* species collected from the Qinghai–Tibet Plateau using Illumina MiSeq genome sequencing technology.

METHODS AND RESULTS

We collected a total of 106 individuals representing three *E. gerardiana* populations, two *E. saxatilis* populations, and one *E. monosperma* C. A. Mey. population from various locations on the Qinghai–Tibet Plateau. Vouchers of the sampled population materials were deposited in the Herbarium of Tibet University, Lhasa, China. The materials and their locality information are listed in Appendix 1.

The high-quality DNA sample for Illumina MiSeq sequencing (Illumina, San Diego, California, USA) was isolated from a randomly selected individual in an *E. gerardiana* population (Eg_QM, Appendix 1) using a DNA extraction kit (Tiangen Biotech, Beijing, China). DNA samples of all other *Ephedra* individuals used for the validation of the identified microsatellite markers were extracted using the cetyltrimethylammonium bromide (CTAB) method following Qin et al. (2013). The quality of extracted DNA samples was monitored on 1% agarose gels. The purity and concentration of the DNA samples were determined using the NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

For the development of microsatellite markers, an Illumina paired-end library was constructed using the TruSeq DNA Sample Prep Kit (Illumina) following the manufacturer's instructions and sequenced using the Illumina MiSeq platform at Majorbio Bio-Pharm Technology Co. Ltd. (Shanghai, China) to generate 250-bp paired-end reads. A total of 3,306,253,832 high-quality reads, consisting

¹ Manuscript received 31 August 2016; revision accepted 28 December 2016.

This work was supported by the National Natural Science Foundation of China (91131901, 31300201), the Specimen Platform of China (Teaching Specimen sub-platform), and the Program for Changjiang Scholars and Innovative Research Team (PSCIRT). The authors thank Min Xu for identifying the *Ephedra* samples and Bao-Rong Lu for his critical comments on this manuscript.

⁴ Author for correspondence: yangzhong@fudan.edu.cn

doi:10.3732/apps.1600104

TABLE 1. Characteristics of 29 microsatellite primer pairs developed in *Ephedra gerardiana*.

Locus name	Primer sequences (5'–3')	Repeat motif	Allele size range (bp)	T _a (°C)	Fluorescent dye	GenBank accession no.
EgSSR01*	F: TGATGATTCATGAAGTATGATTGTG R: CCAATGGCTTTCGGATAACT	(AT) ₂₀	140–172	56	JOE	KR073011
EgSSR02*	F: CCCAATGCTTAAAGAGAGGC R: TTCCAAGCCAAATCATCCTC	(GA) ₁₇	210–246	59	FAM	KR073012
EgSSR03*	F: CATGGCCGAATAATGCTTC R: TTATTGCCCCACCTAAACA	(CT) ₁₃	220–244	59	JOE	KR073013
EgSSR04*	F: CCAAATGTACCCAAAGCAAGA R: TGAGTGAACGCACAAGGAG	(AAT) ₈	264–288	59	FAM	KR073014
EgSSR05*	F: GCCTCCACTCATGAATTACCA R: TGCAAATGAAGCAAGGACTTC	(ATG) ₈	146–164	56	JOE	KR073015
EgSSR06*	F: CTCCCTCCATGTTTTAGCCG R: TGTTTTGAGGAACTTGTGAGA	(ATT) ₇	222–243	54	JOE	KR073016
EgSSR07*	F: TGTGTTTGGAAATCACACTTTT R: GACCAGGTGGAAGTCTCTGAA	(ATT) ₇	140–240	58	FAM	KR073017
EgSSR08*	F: GATGGAATAATGTGGCTGC R: CCAATCCATAAAGGAACTCACA	(AT) ₁₆	188–200	58	JOE	KR073018
EgSSR09*	F: CAACAGGTTCCTTCCAAGA R: TTCAAGCAACAAGGTCCAAA	(AG) ₁₆	212–248	60	JOE	KR073019
EgSSR10*	F: TGAAGAGACATGAACAATGGG R: AGAAGGCCCAAGACCTTT	(TA) ₁₄	176–214	60	FAM	KR073020
EgSSR16	F: TCACATGACAGACAAATGAAGA R: CTTGTACGGAAAAGCCCTGA	(GA) ₁₂	285–295	58	FAM	KX077619
EgSSR17	F: TCATGCCATAGTGTAGTTGGC R: CCTCCTCAAACCAATTTCCA	(AT) ₁₇	256–268	58	FAM	KX077620
EgSSR18	F: AACTCATGATGGGATGTGACC R: GGTTTTGAGATCTTACGAAATGAC	(AT) ₁₆	158–214	60	JOE	KX077621
EgSSR19	F: TCTCACCAATATGACTTTTACCGA R: TCACTATGGGCAGTTGGTCA	(AT) ₁₅	280–290	60	FAM	KX077622
EgSSR21	F: TTCCATCAATCTCAAGGGC R: TGCCAGACTGCTATGTTGCT	(AT) ₁₅	172–180	60	JOE	KX077624
EgSSR23*	F: TTGAGGAAGAGCGCCTTTA R: AAACACAACAACAAGAAAACTCC	(AT) ₁₄	180–196	60	JOE	KX077626
EgSSR24	F: TGTAGCCGACCCCAATAAT R: ATAAACGCATCTTGCATCCC	(TTG) ₇	241–250	60	FAM	KX077627
EgSSR29	F: TGTTCTCGGTCTCACACTCG R: TTTAGTGCCACCACCATG	(ATC) ₇	232–238	59	FAM	KX077632
EgSSR31	F: TTTAGGAGTAGCCTAAGATTTACAG R: GCATGTGGTTATTCATTTGTTCA	(AAT) ₅	150–169	60	JOE	KX077634
EgSSR33*	F: TGGCTGAGTGATGCAGATTC R: ATTCCACAGATGGCATTTC	(TG) ₇ (TA) ₇	237–253	60	FAM	KX077636
EgSSR35*	F: GCATGTCTATGGACTGCCAA R: TTGTAAGAGCATGCGGTG	(CAAC) ₅	209–228	60	JOE	KX077638
EgSSR37	F: TGACCCTTACAATGCCCCAAA R: TCAGGTTTGTCTCTTGCCC	(AAAT) ₅	236–262	60	FAM	KX077640
EgSSR39	F: CCAAGTCAGAAAAGACAGAAACA R: TATGCATCAGTGGCATCCTC	(AC) ₁₁	174–192	59	JOE	KX077642
EgSSR40*	F: CCTGGAAACTTCCATCTCCA R: TCCAGAAAATGCGGGTAAG	(AAT) ₉	204–219	60	JOE	KX077643
EgSSR41	F: TGTGAACACCAATAAGGGG R: GGATGTTATGAACAATGTATGCAA	(AT) ₁₄	246–256	58	JOE	KX077644
EgSSR42	F: TGGCCTAAAATAGTGAAGTCGG R: TCCTATGGGATTGCTCTTTTT	(TA) ₉ (GA) ₈	240–254	57	FAM	KX077645
EgSSR43	F: GTTAAAATTTGCGGGGAGT R: ATGGGTCTATGCTTTTGGCT	(AT) ₉ (TAA) ₉	189–214	60	JOE	KX077646
EgSSR44*	F: CACTTTACAAGGAAAGGCCG R: GAGGATTTATCCACGCAAGG	(AAT) ₇	112–216	60	JOE	KX077647
EgSSR45	F: TCAGGTGGACAAATCCATGA R: TTGCTTGAGAAAGTGCCAAA	(CCAAAG) ₅	198–216	60	JOE	KX077648

Note: T_a = annealing temperature.

*These 15 loci were screened in all samples of *E. gerardiana* and in two related species (*E. saxatilis* and *E. monosperma*) in the genus.

of 5,492,116 bp (Q20 = 95.22%), were generated for the *E. gerardiana* samples (GenBank BioProject number: PRJNA354648). Raw reads were trimmed using SeqPrep (https://github.com/jstjohn/SeqPrep) and Sickle software (Joshi and Fass, 2011). Those reads with quality scores of <20 and lengths <20 bp were removed, leaving only the clean paired-end sequences. The high-quality filtered reads were assembled into 71,461 contigs by the GS De Novo Assembler version 2.8 (Roche Applied Science, Mannheim, Baden-Württemberg, Germany)

with a length of 46,547,889 bp, N50 scaffold size of 1478 bp, N90 scaffold size of 1070 bp, and 32.78% GC content. Consequently, 3442 microsatellite sequences were identified using the MicroSatellite Identification Tool (MISA; Thiel et al., 2003). Frequencies of each repeat type were classified as follows: di-, tri-, tetra-, penta-, and hexanucleotide repeats of 34.1% (1176 loci), 14.12% (486 loci), 0.67% (23 loci), 0.12% (4 loci), and 0.11% (2 loci), respectively. Of these, 2655 loci were designed for primer pairs using Primer3 software (Rozen

TABLE 2. Allelic diversity of *Ephedra gerardiana* populations based on 15 microsatellite loci.^a

Locus	Total (n = 59)			Eg_Qm (n = 20)			Eg_Chd (n = 19)			Eg_Ngz (n = 20)		
	A _e	H _o	uH _e	A _e	H _o	uH _e	A _e	H _o	uH _e	A _e	H _o	uH _e
EgSSR01	2	0.689	0.572	2	0.684	0.575	3	0.632	0.642	2	0.750	0.499
EgSSR02	6	0.250	0.859	5	0.400	0.828	6	0.000	0.842	9	0.350	0.908
EgSSR03	5	0.471	0.797	4	0.300	0.765	4	0.263	0.767	6	0.850	0.858
EgSSR04	3	0.466	0.644	2	0.211	0.508	3	0.632	0.629	4	0.556	0.795
EgSSR05	3	0.575	0.655	3	0.600	0.701	2	0.526	0.539	3	0.600	0.724
EgSSR06	3	0.829	0.641	3	0.737	0.656	2	1.000	0.608	3	0.750	0.659
EgSSR07	2	0.233	0.438	1	0.000	0.097	2	0.000	0.569	3	0.700	0.647
EgSSR08	2	0.722	0.488	2	0.750	0.501	2	0.667	0.457	2	0.750	0.506
EgSSR09	5	0.314	0.808	4	0.176	0.758	5	0.211	0.805	6	0.556	0.862
EgSSR10	6	0.484	0.768	7	0.421	0.886	2	0.421	0.522	8	0.611	0.895
EgSSR23	3	0.516	0.654	3	0.176	0.622	2	0.538	0.575	4	0.833	0.763
EgSSR33	4	0.816	0.774	4	0.647	0.765	5	0.800	0.839	3	1.000	0.717
EgSSR35	3	0.827	0.683	4	0.750	0.738	3	0.842	0.647	3	0.889	0.665
EgSSR40	3	0.642	0.646	3	0.643	0.690	2	0.615	0.443	5	0.667	0.805
EgSSR44	3	0.617	0.569	1	0.263	0.235	3	0.882	0.697	4	0.706	0.775

Note: A_e = number of effective alleles; H_o = observed heterozygosity; uH_e = Nei's unbiased expected heterozygosity (Nei, 1978).

^aLocality and voucher information are provided in Appendix 1.

and Skaletsky, 1999), and 135 loci (mostly dinucleotide or pure nucleotide repeats) were determined suitable to use. Primer pairs were then synthesized at the 135 loci by Sangon Biotech Co. Ltd. (Shanghai, China), and we tested whether these primer pairs could produce PCR products in eight randomly selected individuals from two *E. gerardiana* populations (Eg_Chd and Eg_Qm; Appendix 1).

PCR amplifications were performed in a 10-μL volume with 10 ng of template DNA, 5 μL of 2× Taq PCR MasterMix (Tiangen Biotech), 0.4 μM of forward primers, 0.4 μM of reverse primers, and 3.7 μL of ddH₂O following the protocol by Xu et al. (2014), except for the annealing temperatures as indicated in Table 1 for 30 s. PCR products were visualized on a 6% polyacrylamide gel electrophoresis (PAGE) gel with a 10-bp DNA ladder marker. Consequently, 45 primer pairs could produce PCR products with clear bands on PAGE, and only 29 of these primer pairs produced polymorphic bands among the eight *E. gerardiana* individuals. To confirm the reproducibility of polymorphisms of the 29 primer pairs in the eight *E. gerardiana* individuals, PCR amplifications were repeated with fluorescent dye-labeled (Jiang et al., 2012) forward primers and the PCR products were analyzed using the ABI 3710XL DNA analyzer (Applied Biosystems, Foster City, California, USA), using GeneScan 500 LIZ (Applied Biosystems) as the internal size standard. The peaks of the loci were read by Peak Scanner Software version 1.0 (Applied Biosystems). The PCR reaction was performed in a 10-μL reaction solution containing 10 ng of template DNA, 3.8 μL ddH₂O, 5 μL 2× Taq PCR MasterMix (Tiangen Biotech), 0.4 μM reverse

primers, 0.04 μM M13-tailed forward primers, and 0.36 μM M13 label (5'-CAC-GACGTTGTAAAACGAC-3') following the program: initial denaturation at 94°C for 3 min; followed by 12 cycles of denaturation at 94°C for 30 s, annealing for 35 s at the temperatures described in Table 1, and extension at 72°C for 35 s; followed by another 24 cycles again at these conditions; and a final extension at 72°C for 7 min. The results showed perfect reproducibility of polymorphisms of the 29 primer pairs in the eight *E. gerardiana* individuals.

Fifteen randomly selected loci from the 29 confirmed microsatellite markers were used to test their polymorphisms and transferability in a larger sample set of *E. gerardiana* (59 individuals in three populations) and two related *Ephedra* species (47 individuals in *E. saxatilis* and *E. monosperma*) (Appendix 1). The data matrices of the scored bands from all *Ephedra* species based on the 15 primer pairs were subject to analysis for genetic diversity estimates. Genetic parameters included the number of effective alleles per locus, observed heterozygosity, and Nei's unbiased expected heterozygosity (Nei, 1978). All the analyses were conducted using GenAlEx software version 6.0 (Peakall and Smouse, 2012). Our results showed moderate to high polymorphisms of the 15 primer pairs among *E. gerardiana* populations, with effective alleles ranging from two to six per locus and observed and expected heterozygosity ranging from 0.23–0.83 and 0.44–0.86, respectively (Table 2). Results further indicated high polymorphisms of the 15 loci in the populations of *E. saxatilis* and *E. monosperma* (Table 3). All the results suggest that these 15 microsatellite loci are suitable for

TABLE 3. Allelic diversity of populations of *Ephedra saxatilis* and *E. monosperma* based on 15 microsatellite loci developed in *E. gerardiana*.^a

Locus	<i>Ephedra saxatilis</i>						<i>Ephedra monosperma</i>		
	Es_Tr (n = 20)			Es_Chd (n = 19)			Em_Cn (n = 8)		
	A _e	H _o	uH _e	A _e	H _o	uH _e	A _e	H _o	uH _e
EgSSR01	3	0.895	0.716	2	0.778	0.606	3	0.625	0.725
EgSSR02	3	0.200	0.677	3	0.789	0.679	2	0.800	0.644
EgSSR03	6	1.000	0.863	5	0.842	0.824	4	0.125	0.792
EgSSR04	2	0.200	0.354	2	0.684	0.491	3	0.800	0.778
EgSSR05	2	0.350	0.550	1	0.368	0.309	1	1.000	0.000
EgSSR06	2	0.450	0.396	2	1.000	0.539	3	0.333	0.697
EgSSR07	1	0.000	0.097	2	0.053	0.366	3	0.625	0.725
EgSSR08	2	0.450	0.442	2	0.316	0.548	2	1.000	0.533
EgSSR09	4	0.250	0.737	2	0.316	0.508	2	0.125	0.592
EgSSR10	4	0.474	0.767	5	0.895	0.839	3	1.000	0.758
EgSSR23	2	0.050	0.529	2	0.357	0.542	4	0.750	0.783
EgSSR33	2	0.267	0.570	3	1.000	0.742	2	0.750	0.542
EgSSR35	3	0.600	0.633	2	0.294	0.483	1	0.000	0.000
EgSSR40	2	0.556	0.527	3	0.278	0.648	2	0.500	0.433
EgSSR44	2	0.450	0.376	2	0.000	0.398	1	0.375	0.325

Note: A_e = number of effective alleles; H_o = observed heterozygosity; uH_e = Nei's unbiased expected heterozygosity (Nei, 1978).

^aLocality and voucher information are provided in Appendix 1.

use across species of the genus *Ephedra* in addition to the 14 loci that are suitable for use in *E. gerardiana* (Table 1).

CONCLUSIONS

We developed and confirmed polymorphic microsatellite markers at 29 loci based on *E. gerardiana* populations, using the Illumina MiSeq sequencing method. Fifteen of these microsatellite loci were highly transferable to two related species in the genus, *E. saxatilis* and *E. monosperma*. These microsatellite markers will be useful for the characterization of genetic diversity and analysis of genetic structure for *Ephedra* species.

LITERATURE CITED

- EDITORIAL COMMITTEE OF CHINESE FLORA. 1978. Flora Popularis Reipublicae Sinicae, vol. 7, 481. Science Press, Beijing, China (in Chinese).
- JIANG, Z. X., H. XIA, B. BASSO, AND B. R. LU. 2012. Introgression from cultivated rice influences genetic differentiation of weedy rice populations at a local spatial scale. *Theoretical and Applied Genetics* 124: 309–322.
- JOSHI, N. A., AND J. N. FASS. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [software]. Website <https://github.com/najoshi/sickle> [accessed 18 January 2017].
- KONNO, C., T. MIZUNO, AND H. HIKINO. 1985. Isolation and hypoglycemic activity of ephedrins A, B, C, D and E, glycans of *Ephedra distachya* herbs. *Planta Medica* 51: 162–163.
- NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
- PANDEY, M. R. 2006. Use of medicinal plants in traditional Tibetan therapy system in upper Mustang, Nepal. *Our Nature* 4: 69–82.
- PEAKALL, R., AND P. E. SMOUSE. 2012. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics (Oxford, England)* 28: 2537–2539.
- QIN, A. L., M. M. WANG, Y. Z. CUN, F. S. YANG, S. S. WANG, J. H. RAN, AND X. Q. WANG. 2013. Phylogeographic evidence for a link of species divergence of *Ephedra* in the Qinghai-Tibetan Plateau and adjacent regions to the Miocene Asian aridification. *PLoS ONE* 8: e56243.
- RICO, C., E. NORMANDEAU, A. M. DION-CÔTÉ, M. I. RICO, G. CÔTÉ, AND L. BERNATCHEZ. 2013. Combining next-generation sequencing and on-line databases for microsatellite development in non-model organisms. *Scientific Reports* 3: 3376.
- 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.
- SHEN, G. 1995. Distribution and evolution of the genus *Ephedra* in China. *Acta Botanica Yunnanica* 17: 15–20.
- 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.
- WEN, J., J. Q. ZHANG, Z. L. NIE, Y. ZHONG, AND H. SUN. 2014. Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Frontiers in Genetics* 5: doi:10.3389/fgene.2014.00004.
- XU, L., L. ZENG, B. LIAO, AND Y. ZHONG. 2014. Microsatellite markers for a mangrove species, *Cerbera manghas*, from South China. *Conservation Genetics Resources* 6: 45–48.

APPENDIX 1. Geographic and voucher information of *Ephedra* populations used in this study.

Species	Population name	Population size	Locality information ^a	GPS coordinates	Voucher ID ^b
<i>E. gerardiana</i> Wall. ex C. A. Mey.	Eg_Qm	20	Qomolangma	28°9.891'N, 86°50.571'E	Eg-Qm01–20
<i>E. gerardiana</i>	Eg_Ch	19	Chamdo	30°19.147'N, 97°15.028'E	Eg-Chd01–19
<i>E. gerardiana</i>	Eg_Ngz	20	Nagarze	28°53.556'N, 090°17.224'E	Eg-Ngz01–20
<i>E. saxatilis</i> (Stapf) Royle ex Florin	Es_Ch	19	Chamdo	30°07.647'N, 097°17.248'E	Eg-Chd01–19
	Es_Tr	20	Tingri	28°28.817'N, 087°33.856'E	Es-Tr01–20
<i>E. monosperma</i> C. A. Mey.	Em_Cn	8	Cuona	28°12.559'N, 86°49.368'E	Em-Cn01–08

^aCollection localities are on the Qinghai–Tibet Plateau.

^bVouchers were deposited in the Herbarium of Tibet University, Lhasa, China.