Development of novel EST-SSR markers for *Ephedra sinica* (Ephedraceae) by transcriptome database mining

Si-Qian Jiao1, Yan-Qiang Sun1, Dong-Xu Zhang2, Qiong Gao1, Yuqing Jin1, Hui Liu1, Yongpeng Ma3, Yong Yang4, Ilga Porth5,6, and Jian-Feng Mao1,7

PREMISE OF THE STUDY: *Ephedra sinica* (Ephedraceae) is a gymnosperm shrub with a wide distribution across Central and Eastern Asia. It is widely cultivated as a medicinal plant, but its wild populations are monitored to determine whether protection is needed.

METHODS AND RESULTS: Thirty-six microsatellite markers, including 11 polymorphic markers, were developed from *E. distachya* RNA-seq data deposited in the National Center for Biotechnology Information dbEST database. Among 100 genotyped *E. sinica* individuals originating from five different population groups, the allele number ranged from three to 22 per locus. Levels of observed and expected heterozygosity ranged from 0 to 0.866 (average 0.176) and 0 to 0.876 (average 0.491), respectively. Allelic polymorphism information content ranged from 0.000 to 0.847 (average 0.333). Cross-species amplifications were successfully conducted with two related *Ephedra* species for all 11 di- or trinucleotide simple sequence repeats.

CONCLUSIONS: This study provides the first set of microsatellite markers for genetic monitoring and surveying of this medicinal plant.

KEYWORDS: *Ephedra sinica*; Ephedraceae; expressed sequence tag–simple sequence repeat (EST-SSR) marker; genetic diversity; gymnosperm; medicinal plant.

*Ephedra sinica* Stapf (also known as Chinese ephedra or ma huang; Ephedraceae), a gymnosperm shrub, is distributed across southern Siberia, Mongolia, and China, and is found in arid areas and highlands, occurring on slopes, dry river beds, sandy places, or fields in mountainous areas (Lin et al., 2002). The species is reported as dominant in some areas, but little is known about its entire population size. *Ephedra sinica* has been used in Chinese herbal medicine for thousands of years (Fabricant and Farnsworth, 2001). The stems of most members in the genus *Ephedra* L. contain the alkaloid ephedrine, which is used for treatment of asthma and other respiratory ailments (Liu, 1989; Nam et al., 2003). Recently, *E. sinica* has become extensively exploited in a large market developed for nutritional supplements and stimulants involving this plant. *Ephedra sinica* is recorded on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (Bell and Bachman, 2011). The IUCN lists the species as Least Concern; however, wild populations still need to be monitored to determine whether protection is required, as a species of Least Concern may still be critically endangered within a particular region where numbers are very small or declining.

Recently, 29 polymorphic microsatellite loci were developed for a distantly related species, *E. gerardiana* Wall. ex C. A. Mey., by mining the whole-genome-skimming data from Illumina MiSeq sequencing (De et al., 2017). However, no DNA markers have been developed for *E. sinica*, limiting our ability to monitor its population dynamics and employ conservation genetic measures. The present study developed a crucial set of di- or trinucleotide microsatellite markers by mining an *E. distachya* expressed sequence tag (EST)–derived database. The EST–simple sequence repeat (SSR) markers developed here will enrich the genetic marker set for *Ephedra* species.
METHODS AND RESULTS

A total of 4981 ESTs generated from mRNA sequencing of *E. distachya* were retrieved from the National Center for Biotechnology Information (NCBI)Expressed Sequence Tags database (dbEST) (accessed by searching with “*Ephedra* AND *Ephedra distachya*[porgn:__txid3389]”). Microsatellites with a minimum repeat number of five were detected for 324 ESTs with a minimum length of 200 bp. We obtained 203 unique EST-SSR loci by an all-against-all BLAST analysis and successfully designed primers for 171 unique EST-SSR loci. All bioinformatic operations were performed using the microsatellite detection and development pipeline QDD version 3.1 (Meglécz et al., 2014). Finally, we selected 88 di- or trinucleotide loci with at least five repeats for further evaluation.

We sampled five populations (100 individuals total) of *E. sinica* in Datong, Shanxi Province, China (Appendix 1). Voucher specimens were deposited in the Herbarium of Beijing Forestry University (BJFC). In order to test for successful amplification of the 88 EST-SSR loci selected, we conducted PCR analysis using eight individual plants of *E. sinica*. These eight individuals were collected in the Beijing Botanical Garden, Chinese Academy of Sciences. The genomic DNA was extracted from dried leaves using the cetyltrimethylammonium bromide (CTAB) protocol (Doyle and Doyle, 1987). An M13 tail (FAM, HEX, TAMRA, ROX) was attached to primer (4 pM), 4 μL of mixed forward and reverse primers, and 2 μL (20 ng) of DNA. The following PCR conditions were used: 94°C incubation for 5 min; 25 cycles at 94°C for 40 s, 55°C for 40 s, and 72°C for 1 min; and a final extension at 72°C for 10 min.

Among the 88 identified di- or trinucleotide loci, 38 displayed the expected size bands. After final capillary electrophoresis analysis on an ABI 3730 sequencer (Applied Biosystems, Waltham, Massachusetts, USA), SSR alleles were called with GenMarker version 2.20 (SoftGenetics, State College, Pennsylvania, USA). Of these 38 loci, 36 showed clear, single peaks for each allele as essential for confident scoring, and 11 of these loci were polymorphic among the initially screened eight individuals. Characteristics of the 25 pairs of monomorphic microsatellite loci developed for *E. sinica* are shown in Appendix 2. The 11 polymorphic primer pairs were subsequently used to screen five *E. sinica* populations (with sample sizes n = 20 per population) and two additional populations originating from *E. likiangensis* Florin (n = 20) and *E. equisetina* Bunge (n = 6) (Appendix 1). Table 1 shows the primer sequences, repeat motifs, amplification sizes, GenBank accession number of the target sequences, and functional annotations determined with the protein family database, Pfam (Finn et al., 2014). We employed GenALEx version 6.5 (Peakall and Smouse, 2012) to calculate genetic diversity parameters. The allelic polymorphism information content (PIC) was calculated using CERVUS 3.0 (Kalinowski et al., 2007). Allele numbers ranged from three to 22, with an average of 11.55 alleles per locus. Levels of observed and expected heterozygosity ranged from 0 to 0.842 (average 0.176) and 0 to 0.883 (average 0.491), respectively. In addition, PIC values ranged from 0 to 0.847 (average 0.333). The genetic parameters calculated for the 11 polymorphic EST-SSR loci are detailed in Table 2. The target sequences for all microsatellite loci are provided in Appendices S1 and S2.

Furthermore, we conducted cross-species amplification of the 11 polymorphic primer pairs on two related species: *E. likiangensis* from Yulong, Yunnan Province, and *E. equisetina* from Datong, Shanxi Province, China (Appendix 1). All 11 primer pairs successfully amplified *E. likiangensis*, except for locus E-20, which produced

| Locus | Primer sequences (5′–3′) | Repeat motif | Allele size range (bp) | Fluorescent dye | Function annotationa | GenBank accession no. |
|-------|-------------------------|--------------|-----------------------|----------------|----------------------|----------------------|
| E-2   | F: GAGAGAGAAGTCATATGG   | (AGG)₅       | 192–231               | TAMRA          | Peroxidase           | JG722437             |
|       | R: CCATCTCTGCTTCTTCTGC  |             |                       | FAM            |                      |                      |
|       |                         | (ATC)₅       |                       |                |                      |                      |
| E-18  | F: AGTCAAGGAAACACGCGAC  | (ATC)₅       | 153–228               | TAMRA          | Dev_Cell_Death       | JG719586             |
|       | R: CTCTGGAAAGAGACTGGTA  |             |                       | FAM            |                      |                      |
| E-20  | F: GATATGATGGAAGACAAGG  | (AAC)₃       | 164–170               | HEX            | DUF260, Oxidored₉₁   | JG721857             |
|       | R: ATCCACACCTCGATACCC   |             |                       |                |                      |                      |
| E-33  | F: TTGATGATGCTTGAGCGGC  | (ATC)₃       | 186–246               | ROX            | MGS, AICARFT_IMPCHas | JG720119             |
|       | R: AGTGGCAAGAATGTGGGT   |             |                       |                |                      |                      |
| E-35  | F: GGTGTTTCGATGCGATTCA  | (AGG)₅       | 182–188               | TAMRA          | CK₂_R beta           | JG720356             |
|       | R: ATCTTTGATCTCTTCTCTG  |             |                       |                |                      |                      |
| E-49  | F: CCTTGGGGCGGTTTACAG   | (AGG)₅       | 175–253               | TAMRA          | MIT                  | JG721444             |
|       | R: CGCAAGATCTGAAATCACCT |             |                       |                |                      |                      |
| E-58  | F: GCTCTGGACGAAAGACCGA  | (ATC)₃       | 149–200               | TAMRA          | U-box, zf-RING₉₁, DOPA, dioxygen | JG722187 |
|       | R: GGGTGGAACCTTGAAGCTCTT|             |                       | FAM            |                      |                      |
| E-59  | F: GGATGCTACGGGAGAAGGAG | (AGG)₅       | 174–226               | TAMRA          | Ycl                  | JG722338             |
|       | R: AGGCCCATGCTATCATC    |             |                       |                |                      |                      |
| E-62  | F: TGGCTGGCTGTGCTGTG    | (ATC)₃       | 173–254               | TAMRA          | No hit               | JG722724             |
|       | R: TTGGCTGGCTGTGCTGTG   |             |                       |                |                      |                      |
| E-71  | F: AAAGCGTCGCAAGACAGGT  | (CAA)₅                                          | TAMRA          | AAA_assoc           | JG723111             |
|       | R: TCCCTCTCTCTCTTCACTCA |             |                       |                |                      |                      |
| E-83  | F: GTCTATGCTATGCTACACCAGC | (ATC)₅       | 255–264               | TAMRA          | Pkinase, Pkinase₉₁, Kdo, APH, RIO₁, YbrL-PhoP_reg | JG719186 |
|       | R: GCCACCTCTACGTGCTTCC   | (TTC)₅       |                       | FAM            |                      |                      |

aPfam annotation refers to the protein functional annotation.
monomorphic bands in the species (Table 3). For *E. equisetina*, nine out of the 11 primers tested were polymorphic, and two loci failed to amplify. The interspecific amplification profile may be partially related to the phylogenetic relationships between species, as the relationship between *E. equisetina* and *E. sinica* is more distant (Ickert-Bond and Wojciechowski, 2004). In terms of polymorphisms, except for primers at the *E*-49 locus, the remaining primer pairs showed moderate polymorphism in *E. equisetina*, possibly due to the small sample size.

### CONCLUSIONS

The EST-SSR polymorphic markers developed in this study will be potentially useful for studies of population structure and genetic diversity in *E. sinica* conservation genetics. These new markers will also be applicable for *E. likiangensis* and *E. equisetina* and can enrich the number of DNA markers available for Ephedra.

### ACKNOWLEDGMENTS

The authors thank Dr. X.-R. Wang and Dr. X.-Y. Kang for their valuable suggestions. This study was supported by grants from the Fundamental Research Funds for the Central Universities (no. YX2013-412018BLCB08).

### DATA ACCESSIBILITY

Expressed sequence tags used for primer development were downloaded from the National Center for Biotechnology Information (NCBI) Expressed Sequence Tags database (dbEST). GenBank accession numbers for target sequences of both polymorphic and monomorphic SSR loci are provided in Table 1 and Appendix 2.

### SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

### APPENDIX S1. Monomorphic microsatellite target sequences from microsatellite marker development in Ephedra sinica

### APPENDIX S2. Polymorphic microsatellite target sequences from microsatellite marker development in Ephedra sinica

### LITERATURE CITED

Bell, A., and S. Bachman. 2011. The IUCN Red List of Threatened Species 2011: Ephedra sinica. Website https://doi.org/10.2305/iucn.uk.2011-2.rlts.t201702a9168958.en [accessed 14 September 2018].
De, J., W. Zhu, T. Liu, Z. Wang, and Y. Zhong. 2017. Development of microsatellite markers using Illumina MiSeq sequencing to characterize Ephedra gerardiana (Ephedraceae). Applications in Plant Sciences 5: 1600104.

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

Fabricant, D. S., and N. R. Farnsworth. 2001. The value of plants used in traditional medicine for drug discovery. Environmental Health Perspectives 109: 69.

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, et al. 2014. Pfam: The protein families database. Nucleic Acids Research 42: D222–D230.

Ickert-Bond, S. M., and M. F. Wojciechowski. 2004. Phylogenetic relationships in Ephedra (Gnetales): Evidence from nuclear and chloroplast DNA sequence data. Systematic Botany 29: 834–849.

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

Lin, W., L. Li, and K. Chen. 2002. Advances in Ephedra sinica research. Primary Journal of Chinese Materia Medica 16: 48–50.

Liu, X. X. 1989. Pharmacological action and its clinical application of Ephedra sinica. Journal of Integrated Traditional and Western Medicine 9: 255–256.

Meglécz, E., N. Pech, A. Gilles, V. Dubut, P. Hingamp, A. Trilles, R. Grenier, and J.-F. Martin. 2014. QDD version 3.1: A user-friendly computer program for microsatellite selection and primer design revisited: Experimental validation of variables determining genotyping success rate. Molecular Ecology Resources 14: 1302–1313.

Nam, N.-H., C.-W. Lee, D.-H. Hong, H.-M. Kim, K.-H. Bae, and B.-Z. Ahn. 2003. Antiinvasive, antiangiogenic and antitumour activity of Ephedra sinica extract. Phytotherapy Research 17: 70–76.

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.

APPENDIX 1. Voucher and locality information for Ephedra species used in this study.

| Species                  | Population code | Voucher specimen accession no. | Collection locality | Geographic coordinates | n  |
|--------------------------|-----------------|--------------------------------|---------------------|------------------------|----|
| Ephedra likiangensis Florin | —               | ELF201807281b                  | Baishui River, Jade Dragon Snow Mountain, Lijiang County, Yunnan Province | 27.13205°N, 100.248755°E | 20 |
| Ephedra equisetina Bunge  | —               | —                              | Datong, Shanxi Province | 39.95878°N, 113.776324°E | 6  |
| Ephedra sinica Stapf      | MH-1            | EEB201807301c                  | Fan Yao village, Yanggao County, Datong, Shanxi Province | 40.28975°N, 113.648139°E | 20 |
| Ephedra sinica MH-2       | ESS201806232c   | Fan Yao village, Yanggao County, Datong, Shanxi Province | 39.95890°N, 113.76347°E | 20 |
| Ephedra sinica MH-3       | ESS201806233c   | Yang Lao Wa village, Xubao township, Yuncheng District, Datong, Shanxi Province | 40.85174°N, 113.852189°E | 20 |
| Ephedra sinica MH-4       | ESS201806234c   | Longhun Mountain, Kang Yao village, Yuncheng District, Datong, Shanxi Province | 40.26208°N, 113.622244°E | 20 |
| Ephedra sinica MH-5       | ESS201806255c   | Bai Deng Mountain, Pingcheng District, Datong, Shanxi Province | 40.12804°N, 113.372931°E | 20 |

Note: n = number of individuals sampled

*All voucher specimens are deposited in the Herbarium of Beijing Forestry University (BJFC), Beijing, China.

Collector Yong-peng Ma.

Collector Dong-xu Zhang.

APPENDIX 2. Characteristics of 25 pairs of monomorphic microsatellite loci developed for Ephedra sinica.

| Locus | Primer sequences (5’–3’) | Repeat motif | Allele size (bp) | Fluorescent dye | GenBank accession no. |
|-------|--------------------------|--------------|-----------------|----------------|-----------------------|
| E-1   | F: CGCACTACGCCTCCTTT     | (CT)5         | 151             | FAM            | JG721273              |
|       | R: GGCTGGTTCTCCCATTT     |              |                 |                |                       |
| E-6   | F: CAGTCAGGTCTCTGCGCTC  | (CA)9         | 200             | TAMRA          | JG723006              |
|       | R: TCGAACCGTTGATAGAGAGC  |              |                 |                |                       |
| E-12  | F: TAGCTTGTGGCTATGCGCTC | (TAG)5        | 144             | HEX            | JG719000              |
|       | R: ACCCTCTCCTCTATTGTTG   |              |                 |                |                       |
| E-13  | F: ATCCACATTGGCCCGACAACAA| (CAT)5        | 151             | ROX            | JG719115              |
|       | R: CCTCTTGTCTAGCACGCTCT  |              |                 |                |                       |
| E-19  | F: GAGGCAAGAGCAGAAGATGC  | (GCA)6        | 194             | FAM            | JG720107              |
|       | R: TTTGAGGCTGAGCTAGTG    |              |                 |                |                       |
| E-21  | F: TTTCAGGTGGTGCTGAGAGC | (AG)6         | 244             | TAMRA          | JG719754              |
|       | R: ACTCTCTCTGCCTCCTCTCC  |              |                 |                |                       |
| E-22  | F: AGGCTGACGACCAATCTC    | (GGC)5        | 230             | HEX            | JG723316              |
|       | R: GTGAGGCGGTAAGTGGAGGA  |              |                 |                |                       |
| E-23  | F: TAAACGACGGTTCTCCTCCA  | (TG)5         | 173             | ROX            | JG719685              |

(Continues)
### APPENDIX 2. (Continued)

| Locus | Primer sequences (5′–3′) | Repeat motif | Allele size (bp) | Fluorescent dye | GenBank accession no. |
|-------|--------------------------|--------------|------------------|-----------------|----------------------|
| E-25  | R: TCAAAGTCGTCGAGGAGGAG  | (GGCACA)$_5$ | 186              | FAM             | JG719706             |
|       | F: GAAACCGGCCAGACAGCACGAC |             |                  |                 |                      |
| E-26  | R: GATTTCAGATCCATTATGCG   | (TT)$_9$    | 115              | TAMRA           | JG719755             |
|       | F: TGTTCCTCTCTCTGCGATT    |             |                  |                 |                      |
| E-30  | R: TCCTTGGAGACGTCGAGAGCGT | (CT)$_9$    | 186              | HEX             | JG720051             |
|       | F: GGAACGGACAGTTGGAGGAG   |             |                  |                 |                      |
| E-36  | R: ATGAGACCGAGACAGACAG    | (TT)$_9$    | 243              | ROX             | JG720438             |
|       | F: TGGATCGAGACGAGACAG     |             |                  |                 |                      |
| E-38  | R: TGTCCTGGTTCATCCCCTTC   | (AG)$_5$, (CAC)$_5$ | 228              | FAM             | JG720528             |
|       | F: CTCTCAAAAAATTCCCCCCACTC |             |                  |                 |                      |
| E-39  | R: AGACGAAAGGCAGACAGAG    | (AGG)$_5$, (GCAAGG)$_5$ | 249              | TAMRA           | JG720562             |
|       | F: TCCTTGGTTCATCCCCTTC    |             |                  |                 |                      |
| E-41  | R: TGGGGAGGAGCAGAGCAGCAGC | (AG)$_5$    | 214              | HEX             | JG720763             |
|       | F: TACGTTGAGCAGACAGAGCAGC |             |                  |                 |                      |
| E-46  | R: GTCAAGGACAGAGAGAGGAG   | (ATG)$_5$   | 144              | ROX             | JG721163             |
|       | F: TTCTGCGGAGTTCTCTGCGG   |             |                  |                 |                      |
| E-47  | R: AACGGACAGAGAGAGAGAG    | (TCA)$_5$   | 222              | FAM             | JG721187             |
|       | F: ATTCAGGCGGAGGAGGAGGAG  |             |                  |                 |                      |
| E-54  | R: TCTCTGTCCCTTTCTTAAGTTT | (TG)$_5$    | 165              | TAMRA           | JG721879             |
|       | F: TCGGATTACCACTCACCACTAC |             |                  |                 |                      |
| E-55  | R: AGGCTTTTCTCCAGTGTC     | (GCA)$_9$   | 253              | HEX             | JG721940             |
|       | F: GAGCAAGGCTTCTTTCGACGTA |             |                  |                 |                      |
| E-60  | R: CTTGGAGTTGGCATGAGAGGC  | (GAG)$_5$, (TGG)$_5$, (TT)$_5$ | 167              | ROX             | JG722642             |
|       | F: CGGTGCTTCACTACAGATGAGC |             |                  |                 |                      |
| E-61  | R: GGCTGATCTCTACATGAGCAC  | (TA)$_10$   | 249              | FAM             | JG722646             |
|       | F: GTTCTGAGGCGAGCAAGCAG    |             |                  |                 |                      |
| E-65  | R: GGTGATCTAGGCGAGCCAG    | (AG)$_5$    | 223              | TAMRA           | JG723017             |
|       | F: TGGTTGCTTCTTCTCTGCTTG  |             |                  |                 |                      |
| E-74  | R: GGATCGTGGGGCGAGAAAAG   | (TAT)$_5$   | 193              | HEX             | JG723206             |
|       | F: CTTGCTTCTTCTCAGTTGAG   |             |                  |                 |                      |
| E-84  | R: TCAACTCTCTACATTCTACACAGC | (TC)$_5$, (TA)$_4$ | 183              | ROX             | JG719254             |
|       | F: GAAACGGACACAGGTAAGAG   |             |                  |                 |                      |
| E-88  | R: TGACCAAGCTTACAGACAGAAG | (ACA)$_6$   | 166              | TAMRA           | JG719585             |