CHARACTERIZATION AND DEVELOPMENT OF EST-DERIVED SSR MARKERS IN *SINOWILSONIA HENRYI* (HAMAMELIDACEAE)

Zuo-Zhou Li², Hua Tian², and Jin-Ju Zhang²,³,⁴

²Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, People’s Republic of China; and ³School of Life Sciences, Central China Normal University, Wuhan 430079, People’s Republic of China

• **Premise of the study:** Polymorphic microsatellite markers were developed to reveal the genetic diversity of extant populations and the mating system of *Sinowilsonia henryi* (Hamamelidaceae).

• **Methods and Results:** In this study, nuclear simple sequence repeat (SSR) markers were developed using the Illumina high-throughput sequencing technique (RNA-Seq). The de novo–assembled transcriptome generated a total of 64,694 unique sequences with an average length of 601 bp. A total of 2941 microsatellite loci were detected. Of the 121 tested loci, 13 loci were polymorphic and eight were monomorphic among 72 individuals representing three natural populations of the species. The number of alleles per locus ranged from one to four, and the observed and expected heterozygosity at population level were 0.00–1.00 and 0.10–0.66, respectively.

• **Conclusions:** The developed expressed sequence tag (EST)–SSRs will be useful for studying genetic diversity of *S. henryi* as well as assessing the mating system among *Sinowilsonia* species.

**Key words:** Hamamelidaceae; microsatellite; RNA-Seq; *Sinowilsonia henryi*.

The tree genus *Sinowilsonia* Hemsl. is a member of the Hamamelidaceae family and comprises only one species, *S. henryi* Hemsl. This species is narrowly distributed in the mountains of central China at an elevation of 600–1400 m (Zhang et al., 2003). Currently, the natural habitats of this species are severely deteriorated and fragmented, with population sizes ranging from as few as five individuals to approximately 50 flowering plants (Zhou et al., 2014). Thus, *S. henryi* has been listed as an endangered plant species in the China Plant Red Data Book (Fu and Jin, 1992).

Knowledge of genetic diversity and genetic structure of extant populations is essential to the formulation of effective conservation and management strategies for threatened species (Frankham et al., 2002). Due to their codominance, hyper-variability, and reliable scorability, microsatellite markers have been widely used in population genetic studies (Selkoe and Toonen, 2006). However, microsatellite markers for *S. henryi* are currently not available. High-throughput RNA sequencing (RNA-Seq) is one of the most useful next-generation sequencing techniques for identifying microsatellites. In the current study, we developed and characterized 21 expressed sequence tag–simple sequence repeat (EST-SSR) markers for *S. henryi* using RNA-Seq.

**METHODS AND RESULTS**

Total RNAs were isolated from young leaves using a cetyltrimethylammonium bromide (CTAB) procedure (Chang et al., 1993). The poly(A)⁺ RNA (mRNA) was puriﬁed with the RNA Clean-up Kit (Invitrogen, Carlsbad, California, USA) according to the manufacturer’s instructions. The puriﬁed RNA was subsequently fragmented into small pieces (200 bp) by the fragmentation buffer. Then, the cleaved RNA fragments were used for ﬁrst-strand cDNA synthesis using reverse transcriptase (Invitrogen) with random hexamer primers. Subsequently, second-strand cDNA was synthesized using RNase H and DNA polymerase I (Tiangen, Beijing, China). Illumina paired-end sequencing adapters were then ligated to the ends of the 3′-adenylated cDNA fragments. The cDNA library was sequenced by Shanghai Haiyu Biotechnology Co. Ltd. on the Illumina HiSeq 2000 instrument (Illumina, San Diego, California, USA). Before assembly, raw reads were filtered to remove those containing adapter or low-quality reads (more than 20% of nucleotides with Q-value ≤ 10) and reads containing poly N (>10% ambiguous base calls). Transcriptome assembly was performed using the Trinity package (version 2013-02-25) with the default parameters (Grabherr et al., 2011).

A total of 28.7 million 300-bp, clean, paired-end reads were obtained. All clean reads are available from the National Center for Biotechnology Information (NCBI) Short Read Archive (SRA) database (Bioproject accession no. PRJNA394173). De novo assembly of clean reads resulted in 64,694 unique sequences with an average length of 601 bp and an N50 length of 999 bp. The MicroSatellite identification tool (MISA; Thiel et al., 2003) was used to screen for the presence of microsatellites. The parameters used to identify microsatellites were seven repeats for di-, five for tri- and tetra-, four for penta-, and three for hexanucleotide repeats. Subsequently, SSR primers were designed with minimum GC content of 40% and an expected product size ranging from 100 to 280 bp using Primer3 (Rozen and Skaltsky, 1999).

A total of 8892 SSRs containing repeats from di- to pentanucleotides were identified from 64,694 unique sequences. Dinucleotides were the most abundant repeat type (5232), followed by trinucleotides (2198), hexanucleotides (1035), pentanucleotides (259), and tetranucleotides (168). The dinucleotide repeat (AG/CT), (3646) was followed by (AT/AT), (1192), (AC/CT), (384), and (CG/CG), (11). Among the trinucleotide repeat motifs, the most frequent SSR motif was (CT)n (3646) was followed by (AT/AT)n (1192), (AC/GT)n (384), and (CG/CG)n (11).
PCR primers of SSR loci were used for validation of amplification and polymorphism; of these, 13 revealed microsatellite polymorphism. To the best of our knowledge, this is the first study to develop microsatellites for *S. henryi*. These EST-derived SSRs could provide valuable tools for studying genetic diversity and assessing the mating system among *Sinowilsonia* species. In addition, because EST-derived SSRs may be associated with functional genes, the remaining untested 2820 SSRs and 21 loci developed in the current study may be useful for examining adaptive variation using genome scan methods.

### LITERATURE CITED

CHANG, S., J. PURVEY, AND J. CARNEY. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* 11: 113–116.

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.

FRANKHAM, R., J. D. BALLOU, AND D. A. BRISCOE. 2002. Introduction to conservation genetics. Cambridge University Press, Cambridge, United Kingdom.

FU, L., AND J. M. JIN. 1992. China plant red data book–rare and endangered plants. Science Press, Beijing, China.

GRABHERR, M. G., B. J. HAAS, M. YASSOUR, J. Z. LEVIN, D. A. THOMPSON, I. AMIT, X. ADICONIS, ET AL. 2011. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* 29: 644–652.

MARSHALL, T. C., J. SLATE, L. E. B. KRUK, AND J. M. PEMBERTON. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7: 639–655.

RAYMOND, M., AND F. ROUSSET. 1995. GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86: 248–249.

ROZEN, S., AND H. SKALETSKY. 1999. Primer3 on the WWW for general users and for biologist programmers. In S. MISER and S. A. KRAWZ [eds.], Methods in molecular biology, vol. 132: Bioinformatics: Methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

SEIKLE, K. A., AND R. J. TOOMIN. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters* 9: 615–629.

THIEL, T., W. MICHALEK, R. K. VARSHEM, AND A. GRANEK. 2003. Exploiting EST databases for the development and characterization of gene-derived microsatellites. *Applications in Plant Sciences* 2017 5(11): 1700080 doi:10.3732/apps.1700080

### TABLE 1. Frequency of repeat motifs in nonredundant *Sinowilsonia henryi* ESTs.

| SSR motifs          | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | >12  | Total |
|---------------------|------|------|------|------|------|------|------|------|------|------|------|-------|
| AC/GT               | 106  | 86   | 71   | 64   | 54   | 3    | 0    | 0    | 0    | 0    | 0    | 384   |
| AG/CT               | 717  | 802  | 1268 | 742  | 111  | 6    | 0    | 0    | 0    | 0    | 0    | 3646  |
| AT/TA               | 283  | 275  | 325  | 248  | 59   | 1    | 1    | 1192 |
| CG/GC              | 5    | 4    | 1    | 0    | 0    | 0    | 0    | 11   |
| AAC/GTT            | 59   | 17   | 13   | 6    | 0    | 0    | 0    | 95   |
| AGG/CTT            | 300  | 200  | 164  | 3    | 0    | 0    | 0    | 667  |
| AAT/ATT            | 147  | 103  | 61   | 3    | 0    | 0    | 0    | 314  |
| ACG/GGT            | 111  | 47   | 33   | 3    | 0    | 0    | 0    | 194  |
| ACG/CGT            | 35   | 14   | 10   | 4    | 0    | 0    | 1    | 63   |
| ACT/AGT            | 19   | 9    | 2    | 1    | 0    | 0    | 0    | 31   |
| AGC/CTG            | 139  | 94   | 64   | 4    | 0    | 0    | 0    | 301  |
| AGCC/CTT           | 81   | 44   | 32   | 7    | 0    | 0    | 0    | 164  |
| ATC/ATG            | 136  | 57   | 54   | 5    | 0    | 0    | 0    | 252  |
| CCG/CGG            | 73   | 31   | 9    | 4    | 0    | 0    | 0    | 117  |
| Tetra              | 146  | 22   | 0    | 0    | 0    | 0    | 0    | 168  |
| Penta              | 252  | 7    | 0    | 0    | 0    | 0    | 0    | 259  |
| Hexa               | 844  | 191  | 0    | 0    | 0    | 0    | 0    | 1035 |

Note: — = number of repeats not calculated.
### Table 2. Characterization of 21 EST-SSR primers developed in *Sinowilsonia henryi*.

| Locus | Primer sequences (5′–3′) | Repeat motif | Allele size range (bp) | $T_a$ (°C) | BLASTX top hit description | E-value | GenBank accession no. |
|-------|--------------------------|--------------|------------------------|-----------|-----------------------------|---------|----------------------|
| SH01  | F: TTTGACCCCGAAACACACC  | (CAT)$_2$    | 209–213                | 58        | —                           | —       | MF503975             |
|       | R: TGATACGGCTCAAGTCTCC   |              |                        |           |                             |         |                      |
| SH02  | F: CATCACCTTCTGCTGAAAGC  | (TC)$_10$    | 223–231                | 58        | Hypothetical protein         | 3E-40   | MF503976             |
|       | R: ACCCGGAGGATATATGACGC  |              |                        |           | EUGRSLZ_G03166              |         |                      |
| SH03  | F: CCACTCTGGTCTCTGCTTC  | (ATC)$_7$    | 210–213                | 58        | Phospho-N-acetyluramoyl-     | 2E-176  | MF503977             |
|       | R: CCTGACGTTAAAGGAAACGC  |              |                        |           | pentapeptide-transferase    |         |                      |
| SH04  | F: GTAGTCGGAGGCTTTTGGG   | (TTC)$_7$    | 258–275                | 60        | NUDIX domain-containing      | 3E-165  | MF503978             |
|       | R: GTCTTCGAGAACCTGAAAGG  |              |                        |           | protein                      |         |                      |
| SH05  | F: TATGCTAGTGGTGTTCTGT   | (GCA)$_7$    | 195–202                | 58        | —                           | —       | MF503979             |
|       | R: TAGCTCTGCGGCTCATAC    |              |                        |           |                             |         |                      |
| SH06  | F: ATGGAGGGCTTTTAGCTCGG  | (GCC)$_7$    | 148–158                | 58        | —                           | —       | MF503980             |
|       | R: TGGCTTCCCCTCTCTTCTTT  |              |                        |           |                             |         |                      |
| SH07  | F: TGACATGGAGGGTGGTGGG   | (ATG)$_7$    | 183–186                | 58        | —                           | —       | MF503981             |
|       | R: TACACTCTTCTATGGCTCTT  |              |                        |           |                             |         |                      |
| SH08  | F: GAAGCTGAGTTGGTTACCGG  | (GTT)$_8$    | 214–225                | 58        | —                           | —       | MF503982             |
|       | R: CTTCGGGCCCTATAGTGGGT  |              |                        |           |                             |         |                      |
| SH09  | F: GGGGTGTTGCTCACGTCTTT  | (ACC)$_7$    | 232–240                | 58        | CBL-interacting protein      | 0       | MF503983             |
|       | R: CCACAGTGTGTTGAGAGG    |              |                        |           | kinase 07                   |         |                      |
| SH10  | F: AACCAACAGGGCTGCTCTTT  | (AGC)$_7$    | 225–239                | 59        | Pre-mRNA-splicing factor     | 0       | MF503984             |
|       | R: CGGCTGCAGATAAGTTGGA   |              |                        |           | SYF1                         |         |                      |
| SH11  | F: GGATGGCCTATCGGCTTTG   | (TC)$_10$    | 209–215                | 58        | Transmembrane protein 230   | 1E-61   | MF503985             |
|       | R: AGCAAATTTGGCAGCTGGAG  |              |                        |           |                             |         |                      |
| SH12  | F: GGATCCACAGTGTGCTAGAG  | (TC)$_10$    | 154–156                | 58        | —                           | —       | MF503986             |
|       | R: ACTCTCGGGGCTCATCTCT   |              |                        |           |                             |         |                      |
| SH13  | F: AAGGACGAGAGTGAAGG    | (GCC)$_2$    | 265–268                | 56        | —                           | —       | MF503987             |
|       | R: CCCAATTCCCTCGAGAGT    |              |                        |           |                             |         |                      |
| SH14  | F: TCACCATCATACACACCTC  | (TTG)$_7$    | 175                    | 56        | ABC transporter G family     | 0       | MF501055             |
|       | R: AGGTCTGATGGTTACAGCT   |              |                        |           | member 5-like                |         |                      |
| SH15  | F: AGCAAGAGGAGCACACACTCT | (AAG)$_7$    | 200                    | 58        | —                           | —       | MF501056             |
|       | R: TGCTGCTTTTACTCTTCTCT  |              |                        |           |                             |         |                      |
| SH16  | F: CCAAGAGCCCCACACACTA  | (GCT)$_7$    | 256                    | 56        | —                           | —       | MF501057             |
|       | R: AGACGTCTGAGTTCTTCTGT  |              |                        |           |                             |         |                      |
| SH17  | F: TGGCTTCCACACTCCTCAA  | (ACA)$_8$    | 250                    | 56        | —                           | —       | MF501058             |
|       | R: GGTGGGGAAGGAGAGAGGAGG |              |                        |           |                             |         |                      |
| SH18  | F: ACCCGGCATACATGTGACA  | (CTG)$_7$    | 165                    | 56        | DExH-box ATP-dependent       | 3E-35   | MF501059             |
|       | R: GGGCCGTCATACCTGGCTCT  |              |                        |           | RNA helicase                 |         |                      |
| SH19  | F: GAGCAACACACCAATCCACA  | (GAG)$_7$    | 200                    | 58        | —                           | —       | MF501052             |
|       | R: GCTGCCATGTTGAGAAACACA |              |                        |           |                             |         |                      |
| SH20  | F: GGAGGCTGCTAGGCTACA  | (CT)$_10$    | 200                    | 56        | NADP-dependent malic enzyme  | 0       | MF501051             |
|       | R: AGAGGGAGAGGTCACACA   |              |                        |           |                             |         |                      |
| SH21  | F: CCAATCCTCCGCCGCAAATAG | (GGT)$_7$    | 275                    | 58        | Receptor-like protein 1,     | 2E-35   | MF501052             |
|       | R: GCTCAATTTGGCTACTCTTGGGAG |              |                        |           | putative isof orm 2          |         |                      |

Note: $T_a$ = annealing temperature.

SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics* 106: 411–422.

Zhang, Z., H. Zhang, and P. K. Endress. 2003. *Flora of China*, vol. 9. Science Press, Beijing, China, and Missouri Botanical Garden Press, St. Louis, Missouri, USA.
Table 3. Genetic diversity of 13 SSR loci in three populations of *Sinowilsonia henryi*.a

| Locus | SNJ (N = 15) | FS (N = 25) | WD (N = 32) |
|-------|-------------|-------------|-------------|
|       | A | $H_e$ | $H_o$ | A | $H_e$ | $H_o$ | A | $H_e$ | $H_o$ |
| SH01  | 2 | 0.18 | 0.13 | 2 | 0.30 | 0.28 | 3 | 0.47 | 0.41 |
| SH02  | 3 | 0.56 | 0.88 | 3 | 0.52 | 0.24 | 3 | 0.59 | 0.63 |
| SH03  | 2 | 0.50 | 0.50 | 1 | 0.00 | 0.00*** | 1 | 0.00 | 0.00*** |
| SH04  | 2 | 0.31 | 0.38 | 3 | 0.63 | 0.32 | 3 | 0.48 | 0.38 |
| SH05  | 3 | 0.60 | 1.00 | 3 | 0.25 | 0.28 | 3 | 0.42 | 0.41 |
| SH06  | 3 | 0.51 | 0.75 | 3 | 0.37 | 0.32 | 3 | 0.41 | 0.38 |
| SH07  | 1 | 0.00 | 0.00 | 2 | 0.42 | 0.20 | 1 | 0.00 | 0.00*** |
| SH08  | 2 | 0.22 | 0.25 | 3 | 0.62 | 0.72 | 2 | 0.48 | 0.38 |
| SH09  | 1 | 0.00 | 0.00*** | 2 | 0.08 | 0.08 | 1 | 0.00 | 0.00*** |
| SH10  | 2 | 0.38 | 0.25 | 3 | 0.63 | 0.56 | 3 | 0.63 | 0.59 |
| SH11  | 3 | 0.53 | 0.75 | 3 | 0.66 | 0.80 | 4 | 0.36 | 0.38 |
| SH12  | 2 | 0.50 | 0.75 | 2 | 0.34 | 0.36 | 2 | 0.49 | 0.47 |
| SH13  | 2 | 0.22 | 0.25 | 2 | 0.39 | 0.36 | 2 | 0.44 | 0.34 |
| Average | 2.15 | 0.35 | 0.45 | 2.46 | 0.40 | 0.35 | 2.31 | 0.37 | 0.44 |

Note: $A$ = number of alleles; $H_e$ = expected heterozygosity; $H_o$ = observed heterozygosity; $N$ = number of individuals sampled.

*** Denotes significant departure from Hardy–Weinberg equilibrium after Bonferroni correction ($P < 0.0006$).

Appendix 1. List of vouchers of *Sinowilsonia henryi* used in this study.

| Population code | N | Location | Voucher no. | Geographic coordinates | Altitude (m) |
|-----------------|---|----------|-------------|------------------------|--------------|
| SNJ             | 15 | Shennongjia Mountain, Hubei Province | Q. G. Ye 1102 | 31°30’09”N, 110°24’03”E | 1405 |
| FS              | 25 | Yangchashan, Fang County, Hubei Province | Q. G. Ye 1108 | 31°53’01”N, 110°27’53”E | 1201 |
| WD              | 32 | Wudang Mountain, Hubei Province | Q. G. Ye 1109 | 32°40’59”N, 111°01’01”E | 1035 |

Note: $N$ = number of individuals sampled.

*All vouchers are deposited at the Wuhan Botanical Garden Herbarium (HIB), Wuhan, Hubei Province, China.