Characterization and Development of EST-Derived SSR Markers in Sinowilsonia henryi (Hamamelidaceae)

Authors: Li, Zuo-Zhou, Tian, Hua, and Zhang, Jin-Ju

Source: Applications in Plant Sciences, 5(11)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1700080

Applications in Plant Sciences 2017 5(11): 1700080

PRIMER NOTE

CHARACTERIZATION AND DEVELOPMENT OF EST-DERIVED SSR MARKERS IN *SINOWILSONIA HENRYI* (HAMAMELIDACEAE)[1]

ZUO-ZHOU LI[2], HUA TIAN[2], AND JIN-JU ZHANG[2,3,4]

[2]Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, People's Republic of China; and [3]School 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, hypervariability, 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.

[1]Manuscript received 31 July 2017; revision accepted 5 September 2017.

The authors thank Xiao-Peng Li, Qi-Gang Ye, and Ping Tang for their assistance and advice. This work was supported by the Natural Scientific Foundation of China (grant no. 31400476).

[4]Author for correspondence: jinjvzhang@163.com

doi:10.3732/apps.1700080

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 Skaletsky, 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 $(AG/CT)_n$ (3646) was followed by $(AT/AT)_n$ (1192), $(AC/GT)_n$ (384), and $(CG/CG)_n$ (11). Among the trinucleotide repeat motifs, the most frequent SSR motif was

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

	No. of repeats											
SSR motifs	3	4	5	6	7	8	9	10	11	12	>12	Total
AC/GT	—	—	—	—	106	86	71	64	54	3	0	384
AG/CT	—	—	—	—	717	802	1268	742	111	6	0	3646
AT/TA	—	—	—	—	283	275	325	248	59	1	1	1192
CG/GC	—	—	—	—	5	4	1	1	0	0	0	11
AAC/GTT	—	—	59	17	13	6	0	0	0	0	0	95
AAG/CTT	—	—	300	200	164	3	0	0	0	0	0	667
AAT/ATT	—	—	147	103	61	3	0	0	0	0	0	314
ACC/GGT	—	—	111	47	33	3	0	0	0	0	0	194
ACG/CGT	—	—	35	14	10	4	0	0	1	0	0	63
ACT/AGT	—	—	19	9	2	1	0	0	0	0	0	31
AGC/CTG	—	—	139	94	64	4	0	0	0	0	0	301
AGG/CCT	—	—	81	44	32	7	0	0	0	0	0	164
ATC/ATG	—	—	136	57	54	5	0	0	0	0	0	252
CCG/CGG	—	—	73	31	9	4	0	0	0	0	0	117
Tetra-	—	—	146	22	0	0	0	0	0	0	0	168
Penta-	—	252	7	0	0	0	0	0	0	0	0	259
Hexa-	844	191	0	0	0	0	0	0	0	0	0	1035

Note: — = number of repeats not calculated.

AAG/CTT (667), followed by AAT/ATT (314), AGC/CTG (301), and ATC/ATG (252) (Table 1). Of the 8892 identified SSRs, 2941 (33%) were suitable for designing locus-specific primers (Appendix S1).

SSR loci with a minimum of 10 repeats for dinucleotides and seven for trinucleotides were selected for amplification. A total of 121 primer pairs were selected and used for further characterization. Eight individuals of *S. henryi* from Wuhan Botanical Garden, China, were collected to initially assess microsatellite polymorphism. Genomic DNA was isolated using the CTAB method (Doyle and Doyle, 1987). PCR reactions were performed in a 10-μL reaction mixture (final volume) containing approximately 50 ng of genomic DNA, 0.2 μM each of forward and reverse primer, 10 mM Tris-HCl (pH 8.4), 50 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, and 1 unit *Taq* polymerase (Fermentas, Vilnius, Lithuania). The PCR cycling program included 5 min of initial denaturation at 94°C; followed by 35 cycles of 50 s at 94°C, 50 s at 56–60°C depending on the primer pair (Table 2), and 1 min at 72°C; followed by a final 10-min extension step at 72°C. The PCR products were separated on a 6% polyacrylamide denaturing gel of high resolution with silver stain. A 25-bp marker ladder (Promega Corporation, Madison, Wisconsin, USA) was used to identify the alleles.

Of the 121 primer pairs tested, 21 successfully amplified the target fragments (Table 1); of these, 13 loci were polymorphic (SH01–SH13), while eight were detected as monomorphic (SH14–SH21; Table 2). The level of genetic variability was estimated by genotyping 72 individuals of *S. henryi* from three wild populations (Appendix 1). For each locus, the number of alleles (*A*), observed heterozygosity (H_o), and expected heterozygosity (H_e) were estimated using the program GENEPOP version 3.4 (Raymond and Rousset, 1995). Null alleles were detected at three loci (SH03, SH04, and SH07) using the program CERVUS 2.0 (Marshall et al., 1998). In the SNJ population, *A* ranged from one to three, H_e ranged from 0 to 0.60, and H_o ranged from 0 to 1.00. In the FS population, *A* ranged from one to three, H_e ranged from 0 to 0.66, and H_o ranged from 0 to 0.80. In the WD population, *A* ranged from one to four, H_e ranged from 0 to 0.63, and H_o ranged from 0 to 0.63. Three loci deviated from Hardy–Weinberg equilibrium after correction for multiple tests (Table 3). The observed departures from Hardy–Weinberg equilibrium may be due to null alleles. Significant linkage disequilibrium was observed in 10 pairs of loci before correction for multiple tests ($P < 0.05$). However, no loci were observed to be in linkage disequilibrium after correction for multiple tests ($P < 0.0006$). The sequences containing microsatellites were BLASTed against the NCBI nonredundant protein database using BLASTX with a threshold of *E*-value < 2.00E-5. Ten loci showed significant similarities to known proteins in the NCBI nonredundant protein database (Table 2).

CONCLUSIONS

In the current study, a total of 2941 primer pairs were successfully designed based on transcriptome sequences. In total, 121 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. PURYEAR, AND J. CAIRNEY. 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. K., 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. KRUUK, 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. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics: Methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

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

THIEL, T., W. MICHALEK, R. K. VARSHNEY, AND A. GRANER. 2003. Exploiting EST databases for the development and characterization of gene-derived

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: TTTGACCCCGAAACAACAGC R: TGATACCGCTCAAGTCTCCC	$(CAT)_7$	209–213	58	—		MF503975
SH02	F: CATCACCTTCTGCTGGAACG R: ACCCCGGAGCATATATCAGC	$(TC)_{10}$	223–231	58	Hypothetical protein EUGRSUZ_G03166	3E-40	MF503976
SH03	F: CCACCTCGTTCTTCTCGTCT R: CCTGACGGTAAAGAGAAACGC	$(ATC)_7$	210–213	58	Phospho-*N*-acetylmuramoyl-pentapeptide-transferase	2E-176	MF503977
SH04	F: GAGTCGGAGTCCCATTTTGC R: GTCTTCGAACATGAGGCGTC	$(TTC)_7$	258–275	60	NUDIX domain-containing protein	3E-165	MF503978
SH05	F: TATGCTAGTGGTGGTGCTGT R: TAGTCGTCGGGCTCATCATC	$(GCA)_7$	195–202	58	—		MF503979
SH06	F: ATTGAAGGCGTTTGGATCCG R: TGGCTTCCCTCTCGTCTTTT	$(GCC)_7$	148–158	58	—		MF503980
SH07	F: TGACATGGAGGGTAGTGTGG R: TCACCTCTTCCATTGCCTTCT	$(ATG)_7$	183–186	58	—		MF503981
SH08	F: GAAGCTGGAGTTTGTGACGG R: CTTCGGGGCCTATAGTTGGT	$(GTT)_8$	214–225	58	—		MF503982
SH09	F: GGGGTGTTGTCCATTGATACAG R: CCAGCAGTTGAAGTTCAGGAG	$(ACC)_7$	232–240	58	CBL-interacting protein kinase 07	0	MF503983
SH10	F: AACCAAATCAGGCTCGCTTT R: CCGCTGCCAGATGAAATTGA	$(AGC)_7$	225–239	59	Pre-mRNA-splicing factor SYF1	0	MF503984
SH11	F: GGATTGCCATCATGCTGTTG R: AGCAAATTTGGCCACTGGAG	$(TC)_{10}$	209–215	58	Transmembrane protein 230	1E-61	MF503985
SH12	F: GGCATCCACAGTGTTGCTAG R: ACTTCTGGGGCCATTTCCTT	$(TC)_{10}$	154–156	58	—		MF503986
SH13	F: AAGGACGAGGATGAATGGGG R: CCCAATTCCCCTCGAGAAGT	$(GCG)_7$	265–268	56	—		MF503987
SH14	F: TCACCATCATCACCACCTTCA R: AGGCTCATGGGTTTACAGCT	$(TTG)_7$	175	56	ABC transporter G family member 5-like	0	MF510515
SH15	F: AGCAAGAGGACCAACACTCT R: TGCTGCTTTTACTTCCCCTC	$(AAG)_7$	200	58	—		MF510516
SH16	F: CCAAGAGACCCCACCAACTA R: AGACGTTGCCTCAGTCTTGT	$(GCT)_7$	256	56	—		MF510517
SH17	F: TGGCTTCCAACCTCCTCAAA R: GGTGGGGTGGAAGAAGAGAG	$(ACA)_8$	250	56	—		MF510518
SH18	F: ACCCGCGATCATACTGACAA R: GGTCCGTCATCACTTCTCCT	$(CTG)_7$	165	56	DExH-box ATP-dependent RNA helicase	3E-35	MF510519
SH19	F: GAGCAAACCCACAATCCAGA R: GCTGCCATGGTGAAGAAACA	$(GAG)_7$	200	58	—		MF510520
SH20	F: GGGTGGGGAGAATAGGGAAG R: AGAGGGAGAGAGGGTCACAA	$(CT)_{10}$	200	56	NADP-dependent malic enzyme	0	MF510521
SH21	F: CCATATCCGCCGCCAATAAG R: GCTCAATTTGCTACCTTTGAAG	$(GGT)_7$	275	58	Receptor-like protein 1, putative isoform 2	2E-35	MF510522

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.

ZHOU, T. H., K. X. WU, Z. Q. QIAN, G. F. ZHAO, Z. L. LIU, AND S. LI. 2014. Genetic diversity of the threatened Chinese endemic plant, *Sinowilsonia henryi* Hemsl. (Hamamelidaceae), revealed by inter-simple sequence repeat (ISSR) markers. *Biochemical Systematics and Ecology* 56: 171–177.

TABLE 3. Genetic diversity of 13 SSR loci in three populations of *Sinowilsonia henryi*.[a]

	SNJ (N = 15)			FS (N = 25)			WD (N = 32)		
Locus	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	2	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.
[a]Locality and voucher information are provided in Appendix 1.
*** 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.[a]	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.
[a]All vouchers are deposited at the Wuhan Botanical Garden Herbarium (HIB), Wuhan, Hubei Province, China.