Development and characterization of SSR markers for *Sanguinaria canadensis* based on genome skimming

Renyu Liao1,*, Yuxin Luo1,*, Gulbar Yisilam1, Ruisen Lu1, Yuguo Wang2, and Pan Li1,3

**PREMISE:** Polymorphic nuclear simple sequence repeat (nSSR) markers were developed for *Sanguinaria canadensis* (Papaveraceae), a spring ephemeral native to eastern North America.

**METHODS AND RESULTS:** Based on the genome skimming data of *S. canadensis*, a total of 240 nSSR primer pairs were designed for 80 loci from the assembled nuclear contigs. Of these primer pairs, 19 were selected for initial validation in four populations (80 individuals). All 19 loci produced heterologous amplification. The numbers of alleles per locus ranged from one to 21; the levels of observed and expected heterozygosity per locus ranged from 0.000 to 1.000 and from 0.000 to 0.847, respectively. Transferability of the loci was tested in the related species *Eomecon chionantha*.

**CONCLUSIONS:** The developed nSSR markers revealed polymorphism in the four studied populations and may contribute to investigations of the genetic diversity of *S. canadensis* and *E. chionantha*.

**KEY WORDS:** eastern North America; *Eomecon*; microsatellites; Papaveraceae; *Sanguinaria canadensis*.

*Sanguinaria canadensis* L. (bloodroot; Papaveraceae) is the only species of *Sanguinaria* L., a genus endemic to eastern North America. Bloodroot is a polymorphic species in regard to the morphology of its petals and stems, which vary greatly among individuals. Despite this polymorphism, most botanists consider the different phenotypes to represent variation within a single species; the variation is clearly continuous, exhibiting intermediate forms, and having no obvious correlation with either geography or habitat (Kiger, 1997). Bloodroot is a spring ephemeral herb commonly occurring in temperate deciduous forests, ranging from southern Ontario to eastern Texas in the west, and from northern Florida to New England in the east. Spring ephemerals exhibit a common strategy strongly associated with temperate deciduous forests that allows understory herbs to take advantage of the high levels of sunlight in spring reaching the forest floor prior to formation of a canopy by woody plants (Archibold, 1995). Understanding the population distribution and genetic structure of *S. canadensis* may not only shed light on the evolutionary history of the species, but also provide insights into the formation and evolution of the North American temperate deciduous forest biome.

Simple sequence repeats (SSRs) have been shown to be highly useful genetic markers for assessing genetic diversity and characterizing population genetic structure (Emanuelli et al., 2013; Lind and Gailing, 2013; Chen et al., 2017; Liu et al., 2018). However, there have been no SSR markers developed for *S. canadensis*. In this study, we developed 19 polymorphic nuclear SSR (nSSR) markers for *S. canadensis* using genome skimming and applied these markers to characterize the genetic diversity and population structure of four natural populations. Furthermore, we tested their cross-amplification in the related species *Eomecon chionantha* Hance. Our results suggest that these nSSRs markers will be valuable for future studies on population genetics and phylogeography of *S. canadensis* across its entire geographic range.

**METHODS AND RESULTS**

Two individuals of *S. canadensis* were selected for genome skimming analysis. Fresh leaves of these individuals (XGC1 and XGC2; Appendix 1) were sampled from the field and dried with silica gel. Total genomic DNA was extracted using Plant DNAzol Reagent (Thermo Fisher Scientific, Waltham, Massachusetts, USA) following the manufacturer’s protocol. The high-molecular-weight DNA was sheared using a Covaris LE220 Focused-ultrasonicator (Covaris, Woburn, Massachusetts, USA), then the library was prepared using a NEBNext DNA Library Prep Master Mix Set for Illumina (New England Biolabs, Ipswich, Massachusetts, USA), followed by a size selection of 300–350 bp using the Agencourt AMPure PRIMER NOTE
XP system (Beckman Coulter, Shanghai, China). Finally, sequencing was conducted using the Illumina HiSeq 2500 (Illumina, San Diego, California, USA) at the Beijing Genomics Institute (Shenzhen, China), with 150-bp paired-end sequencing. The raw reads were filtered and then assembled de novo into contigs using the CLC Genome Workbench version 4.0.6 (QIAGEN, Hilden, Germany), following Liu et al. (2017). The raw reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (BioProject ID: PRJNA512069 for XGC1, PRJNA512066 for XGC2).

Contigs mapped to the plastome of Coreanomecon hyloconoides Nakai (NC_031446) and the mitochondria of Nelumbo nucifera Gaertn. (NC_030753) were removed, so that only nuclear contigs remained. BLAST 2.2.26 (Altschul et al., 1990) was used for remapping. The remaining nuclear contigs were analyzed using CandiSSR (Xia et al., 2016) to identify candidate polymorphic nSSRs by comparing the contigs between the two S. canadensis individuals. For CandiSSR analysis, default parameters were used except that flanking sequence length = 50. For each SSR locus, primers were automatically designed in the pipeline via the Primer3 package (Koressaar and Remm, 2007), which generated three primer pairs for each of the 80 loci. The 240 SSR primer pairs (Table 1, Appendix S1) were then evaluated by OLIGO version 6.67 (Molecular Biology Insights, Cascade, Colorado, USA); primers that had ΔG values <4.5 kcal/mol and no annealing troubles were selected.

In total, 19 primer pairs were selected to test in 80 individuals from four natural populations (20 individuals each) of S. canadensis (Appendix 1; Tables 1, 2). DNA was extracted from the leaf tissues as described above. Two rounds of PCR amplifications were performed using a Tsingke PCR kit (Tsingke Biotech Company, Beijing, China). For the first round, a mixture of 15 μL containing approximately 5 ng DNA, 7.5 μL 2 × PCR Master Mix, 5 pM forward primer (synthesized with an 18-bp M13 tail [5′-TGTAAAACGACGGCCAGT-3′] at the 5′ end), and 5 pM of reverse primer. The PCR thermal profile involved an initial denaturation at 95°C for 5 min; followed by 35 cycles of denaturation at 95°C for 30 s, annealing at optimal temperature (Table 1) for 30 s, and extension at 72°C for 30 s; and a final extension of 72°C for 10 min. For the second round, a final volume of 30 μL contained approximately 3 μL of the first round's product.

### Table 1. Characteristics of 19 microsatellite loci developed for Sanguinaria canadensis.

| Locus | Primer sequences (5′–3′) | Repeat motif | Allele size range (bp) | T_a (°C) | A | GenBank accession no. |
|-------|--------------------------|--------------|------------------------|---------|---|----------------------|
| SSR04 | F: TGCAATTCACACCCGTTGAC | (AAT)6       | 142–183                | 59.7    | 8 | MK369726             |
|       | R: GTCGGAACCCTGCAAGTTGA  | (A)6         | 134–146                | 55.8    | 4 | MK369728             |
| SSR05 | F: TGCTGAGGCTTACACCGCTCA | (AC)6        | 118–134                | 55.8    | 9 | MK369735             |
|       | R: GCTTCGCCCAATTCGTCCTA  | (AG)6        | 135–137                | 55.8    | 2 | MK369732             |
| SSR11 | F: TCCCCTGACAAGTACACCGCT | (AG)6        | 178–197                | 57.1    | 3 | MK369724             |
|       | R: CACATCGATTACACCACTCGCT | (AG)6       | 157–163                | 59.7    | 3 | MK369727             |
| SSR21 | F: TGCTGGGATTCTCTGACATGC | (CAG)6       | 154–163                | 55.8    | 3 | MK369719             |
| SSR44 | F: ATCCTGCCCACTTGGAGCTG  | (CT)5        | 181–239                | 55.8    | 21| MK369725             |
| SSR51 | F: ACCGTCCTACCTATGCTCCTCC | (CTC)5      | 155–170                | 58.4    | 5 | MK369729             |
| SSR56 | F: ACAATTGCGACCTGAAAGGAG | (GA)5       | 130–162                | 58.4    | 13| MK369731             |
| SSR59 | F: GTCTTCGGTATTGCCATCCTC | (GA)6       | 154–183                | 55.8    | 3 | MK369723             |
| SSR64 | F: ACGCGGAGAGAACCACTTCT | (CAG)6       | 165–174                | 61.9    | 4 | MK369720             |
| SSR67 | F: CAGAGGTCCTGGGAGATTGTC | (GTG)5      | 131–143                | 58.4    | 4 | MK369736             |
| SSR68 | F: TGCTGGGCAGTTCTGCTCTC | (GTTT)5     | 189                    | 60.9    | 1 | MK369733             |
| SSR80 | F: GGTCTGGCTGAAAAGGCTGCTG | (TC)5       | 123–131                | 57.1    | 5 | MK369730             |
| SSR90 | F: CAGCACTGCTGCTGGATGATG | (TCT)5      | 109–121                | 53.8    | 5 | MK369718             |
| SSR92 | F: GACGACACGAGAATGACGACG | (TCT)5      | 166–184                | 59.7    | 7 | MK369722             |
| SSR93 | F: TTCCACAGCTGGCTCTCTCTC | (TCT)5      | 110–128                | 54.7    | 5 | MK369721             |
| SSR94 | F: ACTAGTCCTCTGCTCTGCTG  | (TGA)6       | 143–173                | 62.6    | 7 | MK369734             |

Note: A = total number of alleles per locus detected in the four populations; T_a = optimized annealing temperature.

http://www.wileyonlinelibrary.com/journal/AppsPlantSci © 2019 Liao et al.
15 μL of 2× PCR Master Mix, 100 μM of fluorescent-labeled (FAM, ROX, HEX, or TAMRA) universal M13 primer, and 10 pM of reverse primer. The PCR reactions were performed with an initial denaturation at 94°C for 2 min; followed by 38 cycles of 94°C for 60 s, 56°C for 45 s, 72°C for 1 min; and a final 10-min extension step at 72°C. The first round of PCR added an adapter to the fragments, which were then fluorescent-labeled in the second round of PCR. The final amplification products of four different loci (with four different fluorescent labels) for each plant sample were pooled and genotyped in an ABI 3730xl DNA Analyzer (Applied Biosystems, Foster City, California, USA) with GeneScan 500 LIZ Size Standard (Applied Biosystems) as the internal reference.

Allele identification was performed using GeneMarker version 2.2.0 (SoftGenetics, State College, Pennsylvania, USA) with default parameters. The presence of null alleles and their bias on genetic diversity were evaluated based on the expectation maximization method implemented in FreeNA (Chapuis and Estoup, 2006). FSTAT version 2.9.3 (Goudet, 1995) was used to test for Hardy–Weinberg equilibrium, and GENEPOP version 4.0.7 (Rousset, 2008) was used to test for linkage disequilibrium. The number of alleles, levels of observed and expected heterozygosity, and polymorphism information content were calculated using CERVUS version 3.0.3 (Kalinowski et al., 2007).

In total, we detected 112 alleles among 19 loci of 80 individuals. The number of alleles at each locus ranged from one to 21, and PIC values ranged from 0.000 to 0.808 (Table 2), suggesting moderate to high levels of polymorphism (Botstein et al., 1980). Levels of observed and expected heterozygosity for each locus varied from 0.000 to 1.000 and 0.000 to 0.847, respectively (Table 2). These data suggest the presence of abundant genetic diversity in the species. Five loci (SSR4, SSR18, SSR64, SSR80, and SSR90) were shown to have null alleles (Table 2; Chapuis and Estoup, 2006). Loci SSR18, SSR80, and SSR90 revealed significant deviation from Hardy–Weinberg equilibrium in populations NC and VA, population WI, and population MN (P > 0.05; Table 2), respectively. Significant linkage disequilibrium was only detected between loci SSR18 and SSR92. These data may indicate non-random mating, and possibly inbreeding, of the plants in these populations.

| Locus | WI (n = 20) | MN (n = 20) | NC (n = 20) | VA (n = 20) |
|-------|------------|------------|------------|------------|
|       | A          | H<sub>e</sub> | H<sub>s</sub> | PIC        | A          | H<sub>e</sub> | H<sub>s</sub> | PIC        | A          | H<sub>e</sub> | H<sub>s</sub> | PIC        |
| SSR4  | 1          | 0.000      | 0.000      | 0.000      | 3          | 0.000      | 0.000      | 0.000      | 6          | 0.000      | 0.000      | 0.000      |
| SSR5  | 1          | 0.500      | 0.500      | 0.500      | 3          | 0.500      | 0.500      | 0.500      | 6          | 0.500      | 0.500      | 0.500      |
| SSR7  | 1          | 0.000      | 0.000      | 0.000      | 3          | 0.000      | 0.000      | 0.000      | 6          | 0.000      | 0.000      | 0.000      |
| SSR8  | 1          | 0.000      | 0.000      | 0.000      | 3          | 0.000      | 0.000      | 0.000      | 6          | 0.000      | 0.000      | 0.000      |
| SSR9  | 1          | 0.000      | 0.000      | 0.000      | 3          | 0.000      | 0.000      | 0.000      | 6          | 0.000      | 0.000      | 0.000      |

Table 2. The genetic parameters (per locus) of 19 microsatellite loci in four populations of *Sanguinaria canadensis*.

**Note:** A<sub>n</sub> = number of alleles per locus detected in each population; H<sub>e</sub> = expected heterozygosity; H<sub>s</sub> = observed heterozygosity; n = number of individuals sampled; PIC = polymorphism information content.

**Voucher and locality information are provided in Appendix 1.**

**Significant deviations (P < 0.05) from Hardy–Weinberg equilibrium detected in the population.**

**Estimated null allele frequency r > 0.05.**

**TABLE 3.** Fragment sizes detected in cross-amplification tests of 19 microsatellite markers developed for *Sanguinaria canadensis* in one population (n = 20) of *Eomecon chionantha*.

| Locus | Allele size range (bp) |
|-------|------------------------|
| SSR4  | 141–171                |
| SSR5  | 134–146                |
| SSR6  | 118–120                |
| SSR7  | 125–151                |
| SSR8  | 181–197                |
| SSR9  | 157–163                |
| SSR10 | 151–157                |
| SSR11 | 161–167                |
| SSR12 | 144–150                |
| SSR13 | 110–134                |
| SSR14 | 156–162                |
| SSR15 | 128–137                |
| SSR16 | 181                    |
| SSR17 | 117–125                |
| SSR18 | 109–130                |
| SSR19 | 169–184                |
| SSR20 | 110–128                |
| SSR21 | 143–170                |

**Voucher and locality information are provided in Appendix 1.**
To verify their potential for heterologous amplification, all 19 SSRs were tested in one population (20 individuals) of *E. chionantha* (Appendix 1), the sister species of *S. canadensis* (Wang et al., 2009). The procedures of PCR amplification and genotyping were the same as described above. All 19 loci amplified in *E. chionantha*, with polymorphism detected in all loci except SSR68 (Table 3), suggesting their potential utility in future studies.

**CONCLUSIONS**

In this study, 19 highly polymorphic nSSR markers were successfully developed for *S. canadensis*. These markers can be applied to investigate population genetics and phylogeography of *S. canadensis* and its close relatives.

**ACKNOWLEDGMENTS**

The authors thank Dr. Quyun (Jenny) Xiang, Dr. Ken Cameron, and Dr. Thomas Wentworth for helpful discussion and suggestions. The project was supported by the National Natural Science Foundation of China (31500184), the Natural Science Foundation of Zhejiang Province (LY19C030007), and the Dimensions of Biodiversity program (grant no. 31461123001).

**DATA AVAILABILITY**

Raw sequencing reads were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (BioProject ID: PRJNA512069 for XGC1, PRJNA512066 for XGC2). Sequence information for the developed primers has been deposited in NCBI's GenBank, and accession numbers are provided in Table 1.

**SUPPORTING INFORMATION**

Additional Supporting Information may be found online in the supporting information tab for this article.

**APPENDIX S1.** The remaining 221 primers developed through the pipeline for *Sanguinaria canadensis*.

**LITERATURE CITED**

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Archibold, O. W. 1995. Ecology of world vegetation. Chapman and Hall, London, United Kingdom.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32: 314–331.

Chapuis, M. P., and A. Estoup. 2006. Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* 24: 621–631.

Chen, C., R. Lu, S. Zhu, I. Tamaki, and Y. Qiu. 2017. Population structure and historical demography of *Dipterocarpus digeriana* (Sapindaceae), an extremely narrow palaeoendemic plant from China: Implications for conservation in a biodiversity hot spot. *Heredity* 119: 95–106.

Emanuelli, E., S. Lorenzi, L. Grzeskowiak, V. Catalano, M. Stefanini, M. Troggio, S. Myles, et al. 2013. Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC Plant Biology* 13: 59.

Goudet, J. 1995. FSTAT (version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* 86: 485–486.

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.

Kiger, R. W. 1997. Papaveraceae. In *Flora of North America* Editorial Committee [eds.], *Flora of North America North of Mexico*, Vol. 3, 300–339. Oxford University Press, New York, New York, USA.

Koressaar, T., and M. Remm. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23: 1289–1291.

Lind, J. F., and O. Gailing. 2013. Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E.J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetics & Genomes* 9: 707–722.

Liu, L., R. Li, J. R. Worth, X. Li, P. Li, K. M. Cameron, and C. Fu. 2017. The complete chloroplast genome of Chinese bayberry (*Morella rubra*, Myrtaceae): Implications for understanding the evolution of Fagales. *Frontiers in Plant Science* 8: 968.

Liu, L., Y. Wang, P. He, P. Li, J. Lee, D. E. Soltis, and C. Fu. 2018. Chloroplast genome analyses and genomic resource development for epilithic sister genera *Oreostrope* and *Mukdenia* (Saxifragaceae), using genome skimming data. *BMC Genomics* 19: 235.

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

Wang, W., A. Lu, Y. Ren, M. E. Endress, and Z. Chen. 2009. Phylogeny and classification of Ranunculaceae: Evidence from four molecular loci and morphological data. *Perspectives in Plant Ecology, Evolution and Systematics* 11: 81–110.

Xia, E., Q. Yao, H. Zhang, J. Jiang, L. Zhang, and L. Gao. 2016. CandiSSR: An efficient pipeline used for identifying candidate polymorphic SSRS based on multiple assembled sequences. *Frontiers in Plant Science* 6: 1171.

**APPENDIX 1.** Voucher information for species used in this study.

| Species                  | Population code | Voucher no.* | Locality               | Geographic coordinates | Altitude (m) | n  |
|--------------------------|-----------------|--------------|------------------------|------------------------|--------------|---|
| *Sanguinaria canadensis* L. | XGC1            | Pan Li, LP162238 | Winslow, Arkansas, USA  | 35.78180°N, 94.21827°W | 350          |    |
|                          | XGC2            | Pan Li, LP162495 | Marriana, Florida, USA  | 30.79820°N, 85.22835°W | 21           |    |
|                          | WI              | Pan Li, LP185609 | Lyndon Station, Wisconsin, USA | 43.67250°N, 90.03027°W | 380          | 20 |
|                          | MN              | Pan Li, LP185619 | Rochester, Minnesota, USA | 44.03513°N, 92.43248°W | 320          | 20 |
|                          | NC              | Pan Li, LP185735 | Micaville, North Carolina, USA | 35.81235°N, 82.14401°W | 1136         | 20 |
|                          | VA              | Pan Li, LP185767 | McLean, Virginia, USA | 38.96440°N, 77.15205°W | 68           | 20 |
| *Eomecon chionantha Hance* | CQ              | Pan Li, LP185255 | Nanchuan, Chongqing, China | 29.11748°N, 107.33582°E | 1289         | 20 |

*Note: n = number of individuals per population.*

*Vouchers were deposited in the Herbarium of Zhejiang University (HZU), Hangzhou, Zhejiang, China.*