Development and Characterization of Microsatellite Primers in Geranium carolinianum (Geraniaceae) with 454 Sequencing

Authors: Shirk, Rebecca Y., Glenn, Travis C., Chang, Shu-Mei, and Hamrick, J. L.

Source: Applications in Plant Sciences, 1(8)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1300006

Applications in Plant Sciences 2013 1(8): 1300006

PRIMER NOTE

DEVELOPMENT AND CHARACTERIZATION OF MICROSATELLITE PRIMERS IN *GERANIUM CAROLINIANUM* (GERANIACEAE) WITH 454 SEQUENCING[1]

REBECCA Y. SHIRK[2,3,5], TRAVIS C. GLENN[4], SHU-MEI CHANG[2], AND J. L. HAMRICK[2]

[2]Department of Plant Biology, University of Georgia, Athens, Georgia 30602 USA; [3]Department of Botany, University of Wisconsin–Madison, Madison, Wisconsin 53706 USA; and [4]Department of Environmental Health Science, University of Georgia, Athens, Georgia 30602 USA

- *Premise of the study:* Microsatellite primers were developed for *Geranium carolinanum*, a North American winter annual herb, for use in population genetic analyses.
- *Methods and Results:* Genomic DNA enriched for repeat-containing fragments was sequenced on the Roche 454 Titanium platform, resulting in 470 primer pairs developed from 1115 microsatellite-containing sequences. A subset of 37 primer pairs was screened for polymorphism across three native and three invasive populations. We identified four monomorphic and eight polymorphic loci. Polymorphic loci contained between two and seven alleles per locus, and mean within-population expected heterozygosity ranged from 0.100 to 0.290. Within populations, observed heterozygosity for individual loci ranged from zero to 0.857, and expected heterozygosity ranged from 0.046 to 0.559.
- *Conclusions:* These microsatellite markers will be useful for future studies of genetic diversity, structure, and mating systems across the geographic range of *G. carolinianum*, and may be transferable to other closely related species.

Key words: 454 sequencing; biological invasions; Geraniaceae; *Geranium carolinianum*.

Geranium carolinianum L. (Geraniaceae) is a weedy winter annual herb native to North America and naturalized in East Asia, South America, and the Caribbean (Aedo, 2000). In China, it is naturalized along roadsides and in agricultural fields, and is considered an invasive species with minor environmental impacts (Liu et al., 2006). The origin and colonization history of these introduced populations is unknown. Although it is not closely related to the horticultural geranium (genus *Pelargonium* L'Hér. ex Aiton), genetic resources developed for *G. carolinianum* are potentially transferable to other *Geranium* L. species, which includes annuals and perennials that are both native and introduced in North America (Aedo, 2000). Here, we present 12 microsatellite loci developed using targeted enrichment and next-generation sequencing.

METHODS AND RESULTS

DNA was extracted from leaf tissue using a QIAGEN Plant Mini Kit (QIAGEN, Valencia, California, USA) and enriched for microsatellites using three probe mixes [Mix 2 = $(AG)_{12}$, $(TG)_{12}$, $(AAC)_6$, $(AAG)_8$, $(AAT)_{12}$, $(ACT)_{12}$, $(ATC)_8$; Mix 3 = $(AAAC)_6$, $(AAAG)_6$, $(AATC)_6$, $(AATG)_6$, $(ACAG)_6$, $(ACCT)_6$, $(ACTC)_6$, $(ACTG)_6$; Mix 4 = $(AAAT)_8$, $(AACT)_8$, $(AAGT)_8$, $(ACAT)_8$, $(AGAT)_8$] as in Glenn and Schable (2005). DNA from one individual was digested separately with *Rsa*I (New England Biolabs, Ipswich, Massachusetts, USA) and *Alu*I (New England Biolabs). Digests were pooled, a single adenosine was added with exo-Klenow fragment (New England Biolabs), and fragments were ligated to CAG-SimpleXT-13 adapters (upper oligo = GTTTCAGTCGGGCGTCATCA CCAGCACCGGAACAT; lower oligo = /5Phos/TGTTCCGGTGCTGG). Linker-ligated DNA was pooled, denatured, and hybridized to biotinylated microsatellite oligonucleotide mixes, which were subsequently captured on magnetic streptavidin beads (Dynabeads M-280; Invitrogen, Carlsbad, California, USA), and remaining DNA was washed away. The enriched DNA was eluted and amplified in PCR using Phos_Pig_CAG (/5Phos/GTTTCAGTCGGGCGTCATCA). PCR reactions were run in a 25 µL total volume with 2 µL eluted DNA, 0.2 µL Phusion *Taq* polymerase (5 units/µL; New England Biolabs), 1× Phusion reaction buffer (New England Biolabs), 0.025 mg/mL bovine serum albumin (BSA; New England Biolabs), 2.0 mM $MgCl_2$, 0.15 mM each dNTP (Sigma-Aldrich, St. Louis, Missouri, USA), and 0.5 µM Phos_Pig_CAG. Cycling conditions were 2 min at 95°C; 25 cycles of 20 s at 95°C, 20 s at 60°C, and 1.5 min at 72°C; and a final 30-min extension at 72°C. Multiple libraries with additional different SimpleX sequences were then pooled and ligated into a 454 RL MID-tagged library, pooled with additional 454 RL MID-tagged libraries, and sequenced on half a run of the Roche 454 FLX with Titanium chemistry using the manufacturer's protocols.

Following demultiplexing, 1115 sequences with microsatellites were identified from 5586 sequence reads using MSATCOMMANDER version 0.8.1 (Faircloth, 2008). All microsatellite-containing sequences are provided in Appendix S1. MSATCOMMANDER also designs primers for microsatellite-containing sequences using Primer3 (Rozen and Skaletsky, 2000). Sequences were discarded if Primer3 could not design acceptable primers. All primers (excluding those tested as described below, a total of 433 pairs) are reported in Appendix S2. To avoid selecting primers within highly repetitive regions of the genome, sequences were assembled into contigs by automatic assembly using default parameters in Sequencher version 4.7 (Gene Codes Corporation, Ann

[1]Manuscript received 15 January 2013; revision accepted 25 February 2013.

The authors thank B. Faircloth for demultiplexing the 454 reads and revisions to MSATCOMMANDER, and an anonymous reviewer who provided helpful comments. This material is based upon work supported by the National Science Foundation's Partnerships for International Research and Education (PIRE) program under grant no. OISE 0730218.

[5]Author for correspondence: rshirk@wisc.edu

doi:10.3732/apps.1300006

Arbor, Michigan, USA). Microsatellite-containing sequences that were at high frequency (i.e., members of contigs with >10 sequences) were not considered for screening.

Of the remaining sequences, we selected 37 primer pairs for testing. The majority of primers tested were based on unique sequences in the data set and amplified tri- or tetranucleotide repeats. Forward primers were tagged on the 5′ end with the universal "CAG" sequence (CAGTCGGGCGTCATCA) to allow fluorescent labeling of PCR product using a 3-primer protocol as detailed below. Additionally, a 5′GTTT "pigtail" was added to some reverse primers to ensure consistency in amplicon size (Brownstein et al., 1996). These 37 primer pairs were screened using DNA template from 12 samples across *G. carolinianum*'s native range. All primer pairs were initially tested using a "touchdown 58" PCR protocol. Cycling conditions were 3 min at 95°C; nine touchdown cycles of 30 s at 95°C, 30 s at 65–58°C (starting at 65°C and decreasing one degree each cycle), and 45 s at 72°C; 30 cycles of 30 s at 95°C, 30 s at 50°C, and 45 s at 72°C; and a final extension of 20 min at 72°C. Primer pairs that did not amplify consistently at touchdown 58 were tested with a "touchdown 50" protocol, which was identical to touchdown 58 except the annealing temperatures for the touchdown cycles decreased from 58°C to 50°C, followed by 30 cycles at an annealing temperature of 50°C. All PCR reactions were carried out in a 12.5 μL total volume with 10–200 ng DNA template, 0.1875 units JumpStart *Taq* DNA polymerase (Sigma-Aldrich), 1× JumpStart reaction buffer without magnesium (Sigma-Aldrich), 1× (0.1 mg/mL) BSA, 2 mM $MgCl_2$, 0.15 mM each dNTP, 0.025 μM forward (CAG-tagged) primer, 0.25 μM reverse (untagged) primer, and 0.25 μM universal CAG primer labeled with a 5′ FAM, HEX, or NED fluorescent dye. Primer pairs with consistent amplification at either touchdown 58 or touchdown 50 were screened for polymorphism by fragment analysis. PCR products from the initial 12 DNA samples were sized on an ABI 3730xl capillary sequencer (Applied Biosystems, Carlsbad, California, USA) using a ROX-labeled internal size standard (GGF500R; Georgia Genomics Facility, Athens, Georgia, USA). Chromatograms were analyzed using GeneMapper version 3.7 (Applied Biosystems).

Of the 37 primer pairs tested, we identified eight polymorphic and four monomorphic loci with high-quality amplification (Table 1). Primers that performed poorly in the screening are listed in Appendix S3. Monomorphic loci were tested further against either 48 or 96 individual DNA samples from across the native and invasive ranges, but no polymorphism was found. Polymorphic loci were used to assess genetic variation in 103 individual DNA samples from three native populations in the United States and three invasive populations in eastern China (Table 2). For GC7 and GC35, only a subset of DNA samples were available for one of the populations from Georgia, USA, and the population from Henan, China, so sample sizes from other populations were increased to reach a total sample size of roughly $n = 100$ per locus. Voucher specimens from the two Georgia and three China populations have been deposited at the University of Georgia Herbarium (GA). We were unable to obtain a voucher for the North Carolina population.

We calculated the number of alleles, observed heterozygosity (H_o), and expected heterozygosity (H_e) for each polymorphic locus and population in GenAlEx version 6.2 (Peakall and Smouse, 2006). Results are given in Table 2. Across all samples, number of alleles per polymorphic locus ranged from two to seven, and five loci were monomorphic in at least one population. Mean population H_e (averaged across loci) ranged from 0.100 to 0.290. Within populations, H_o for individual loci ranged from zero to 0.857, and H_e ranged from 0.046 to 0.559. We did not test for Hardy–Weinberg equilibrium because this is a colonizing, mixed-mating species.

CONCLUSIONS

We report 12 microsatellite loci in *G. carolinianum*, eight of which are shown to be variable in native and invasive populations. Forward primer sequences are labeled with a 5′ universal sequence to allow fluorescent tagging of amplicons during PCR, and reverse primer sequences (with the exception of GC1-R, GC6-R, GC7-R, and GC10-R) are labeled with a 5′ GTTT pigtail sequence for consistent genotyping. Although not tested, addition of a 5′ pigtail to these primers should facilitate accurate genotyping. These microsatellite loci will be useful for studies of genetic variation, gene flow, and mating system in *G. carolinianum* and may be transferable to related species.

TABLE 1. Characteristics of microsatellite primers developed in *Geranium carolinianum*.

Locus	Primer sequences (5′–3′)	Repeat motif	Fragment size (bp)	T_a (°C)	Dye[a]	GenBank accession no.
GC1[b]	F: CAGTCGGGCGTCATCATTGTGAGCTTGCTCTTGCC R: AAGGCATCCCAACAGAGGG	$(CTT)_4$	203	65–58	FAM	KC433595
GC6[b]	F: CAGTCGGGCGTCATCACGAGTTGCAGCTACCAAGC R: TGGAGCCTCTATTGCACCC	$(CTT)_4$	240	65–58	HEX	KC433596
GC7	F: CAGTCGGGCGTCATCATCTCGCTCATCATCACTCTCC R: GAACGAAGCAATCCGCTGG	$(CTT)_{18}$	217	58–50	NED	KC429676
GC10	F: CAGTCGGGCGTCATCAGACCTGGAGGTAAGTCCCTG R: GGAGCTCGGCTACTCTTCC	$(CTT)_{16}$	195	58–50	HEX	JX075892
GC29	F: CAGTCGGGCGTCATCAACTGCGCTTGTAGAAATCTG R: GTTTGTGATTTTGACGGTGGGC	$(AGG)_7$	154	58–50	HEX	JX075893
GC31	F: CAGTCGGGCGTCATCAGTGGTTGGGGTGTGTGAAC R: GTTTCGAAAGAACCGAACCGGAC	$(GAGT)_4$	244	65–58	FAM	JX075894
GC33[b]	F: CAGTCGGGCGTCATCACCGGGAATGGCTAGTACG R: GTTTGGATGCCTAAGCTGTCCAAG	$(GTTT)_5$	188	58–50	FAM	KC433597
GC35	F: CAGTCGGGCGTCATCACTCTCTTCTCTCGGCCACC R: GTTTCGAACGAGGGGCATTTTCG	$(ATT)_9$	223	58–50	HEX	KC429677
GC36[b]	F: CAGTCGGGCGTCATCACTTGCTCTGGTCAGTCTTGG R: GTTTGGGAGGATAGGGAATCTGCTG	$(CTT)_{14}$	268	58–50	HEX	KC433598
GC38	F: CAGTCGGGCGTCATCAGCTAGGATCAGCAGTCCCG R: GTTTGCTCAATGTCTCGCAGG	$(GCCT)_5$	150	58–50	FAM	JX075895
GC39	F: CAGTCGGGCGTCATCAGCTCGTGAGTTCATTATGTTTGC R: GTTTCAATCCAGCCACCTTTCGC	$(GTTT)_6$	242	58–50	NED	JX075896
GC47	F: CAGTCGGGCGTCATCATGGAGTCTCTCGCAACAC R: GTTTCAACTCAGGCTCTGCTCC	$(ACAT)_6$	204	58–50	FAM	JX075897

Note: T_a = range of annealing temperatures used in touchdown PCR.

[a] Fluorescent dye used for fragment analysis.

[b] Monomorphic locus.

TABLE 2. Results of primer screening of eight polymorphic microsatellite loci in three native (United States) and three invasive (China) populations of *Geranium carolinianum*.

	Georgia, USA (34.08°N, 84.65°W; $n = 18$)			Georgia, USA (34.82°N, 85.24°W; $n = 18$)			North Carolina, USA (35.85°N, 80.13°W; $n = 19$)			Hunan, China (26.89°N, 112.62°E; $n = 10$)			Jiangxi, China (29.74°N, 116.02°E; $n = 21$)			Henan, China (32.13°N, 114.20°E; $n = 17$)			All samples ($n = 103$)
Locus	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A	H_o	H_e	A
GC7[a]	4	0.056	0.650	1	0.000	0.000	3	0.043	0.559	1	0.000	0.000	2	0.000	0.111	1	0.000	0.000	7
GC10	2	0.056	0.153	3	0.333	0.323	3	0.368	0.492	2	0.300	0.375	2	0.048	0.046	2	0.059	0.057	6
GC29	2	0.056	0.153	2	0.000	0.105	4	0.211	0.320	4	0.500	0.480	2	0.857	0.490	3	0.765	0.524	4
GC31	2	0.111	0.475	2	0.167	0.239	3	0.053	0.148	2	0.000	0.180	2	0.476	0.363	2	0.529	0.438	3
GC35[b]	2	0.000	0.091	1	0.000	0.000	1	0.000	0.000	1	0.000	0.000	1	0.000	0.000	3	0.091	0.244	4
GC38	2	0.278	0.239	1	0.000	0.000	1	0.000	0.000	2	0.100	0.095	1	0.000	0.000	1	0.000	0.000	3
GC39	2	0.111	0.278	1	0.000	0.000	1	0.000	0.000	2	0.100	0.095	2	0.143	0.210	2	0.059	0.057	2
GC47	2	0.000	0.278	2	0.000	0.105	2	0.000	0.100	1	0.000	0.000	1	0.000	0.000	2	0.059	0.057	4

Note: A = number of alleles; H_e = expected heterozygosity; H_o = observed heterozygosity; n = sample size.

[a] Population sample sizes for locus GC7 are as follows: $n = 18$, 8, 23, 18, 17, and 14; total $n = 98$.

[b] Population sample sizes for locus GC35 are as follows: $n = 21$, 5, 22, 19, 22, and 11; total $n = 100$.

LITERATURE CITED

AEDO, C. 2000. The genus *Geranium* L. (Geraniaceae) in North America. I. Annual species. *Anales del Jardin Botanico de Madrid* 51: 39–82.

BROWNSTEIN, M. J., J. D. CARPTEN, AND J. R. SMITH. 1996. Modulation of non-templated nucleotide addition by Taq DNA polymerase: Primer modifications that facilitate genotyping. *BioTechniques* 20: 1004–1006, 1008–1010.

FAIRCLOTH, B. 2008. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources* 8: 92–94.

GLENN, T. C., AND N. A. SCHABLE. 2005. Isolating microsatellite DNA loci. *In* E. A. Zimmer and E. H. Roalson [eds.], Methods in enzymology, vol. 395, Molecular evolution: Producing the biochemical data, Part B, 202–222. Academic Press, San Diego, California, USA.

LIU, J., M. DONG, S. L. MIAO, Z. Y. LI, M. H. SONG, AND R. Q. WANG. 2006. Invasive alien plants in China: Role of clonality and geographical origin. *Biological Invasions* 8: 1461–1470.

PEAKALL, R., AND P. E. SMOUSE. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

ROZEN, S., AND H. J. SKALETSKY. 2000. 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.