Applications in Plant Sciences 2015 3(8): 1500051

PRIMER NOTE

CHARACTERIZATION OF 13 MICROSATELLITE MARKERS FOR *CALOCHORTUS GUNNISONII* (LILIACEAE) FROM ILLUMINA MISEQ SEQUENCING[1]

RYAN S. FULLER[2,3], SETH FRIETZE[2], AND MITCHELL E. MCGLAUGHLIN[2]

[2]School of Biological Sciences, University of Northern Colorado, 501 20th Street, Greeley, Colorado 80639 USA

- *Premise of the study:* Microsatellite primers were designed for *Calochortus gunnisonii* (Liliaceae), a montane lily species of the central and southern Rocky Mountains, using next-generation DNA sequencing. The markers will be used to investigate population structure, genetic diversity, and demographic history.
- *Methods and Results:* Thirteen polymorphic microsatellite loci were isolated from *C. gunnisonii* using Illumina MiSeq next-generation DNA sequencing and bioinformatic screening. The mean number of alleles per locus ranged from 4.15 to 5.92 (avg. = 4.97). Observed and expected heterozygosity ranged from 0.077 to 0.871 and 0.213 to 0.782, respectively. The primers were also tested for cross-species amplification value with *C. flexuosus*, *C. nuttallii*, *C. kennedyi* var. *kennedyi*, and *C. subalpinus*.
- *Conclusions:* These primers will be useful for genetic and evolutionary studies across *C. gunnisonii*'s range within the southern and central Rocky Mountains. Furthermore, these markers have proven valuable for cross-species amplifications within *Calochortus*.

Key words: *Calochortus gunnisonii*; Illumina; Liliaceae; microsatellites; MiSeq.

Calochortus Pursh (Liliaceae) is a large genus of bulbous geophytes (ca. 70 spp.) originating in California ~7 million years ago (Patterson and Givnish, 2003). Its range includes a center of diversity in California that spreads north to British Columbia, east to the Dakotas, and south to Guatemala (Ownbey, 1940; Patterson and Givnish, 2003; Henss et al., 2013). The genus also occurs in a wide range of habitats including grasslands, deserts, vernal pools, woodland meadows, springs, montane woodlands, and forest understories, with most taxa occupying narrow geographic ranges (Ownbey, 1940; Patterson and Givnish, 2003; Fiedler and Zebell, 2012; Henss et al., 2013).

Gunnison's mariposa lily, *C. gunnisonii* S. Watson, is a North American endemic populating portions of the central and southern Rocky Mountains. *Calochortus gunnisonii* has a broad distribution encompassing northeastern Arizona, northern New Mexico, much of Colorado, eastern Utah, large portions of Wyoming, southern Montana, and western South Dakota at elevations of 1200–3300 m (Fiedler and Zebell, 2012). This species achieves some of the highest elevations for the genus in the southern portion of the Rocky Mountains. Disjunctions in the northern portions of *C. gunnisonii*'s range exist across mountain ranges of the Big Horn Mountains, Black Hills, Absaroka Range, Sierra Madre, Medicine Bow, and Laramie Range. A population genetic study of *C. gunnisonii* across multiple, disjunct populations in the central and southern Rocky Mountains is currently being conducted. However, previous genetic studies of members within the genus are limited to amplified fragment length polymorphisms (Henss et al., 2013) and chloroplast sequence comparisons (Patterson and Givnish, 2003). Here, we report the characterization of 13 microsatellite loci that will be used to investigate the role of glacial oscillatory demographic changes in shaping genetic structure of *C. gunnisonii* across multiple montane disjunctions of the central and southern Rocky Mountains.

METHODS AND RESULTS

Next-generation sequencing was used to acquire a large quantity of genomic sequence data in search of microsatellite repeats. Genomic DNA (gDNA) was extracted from leaf tissue using a modified cetyltrimethylammonium bromide (CTAB) protocol (Friar, 2005). Two individual DNA samples collected from two separate geographic populations (Hell Canyon Road, South Dakota: 43.724753°N, 103.854112°W; La Prele Reservoir, Wyoming: 42.706796°N, 105.578531°W [Appendix 1]) were pooled and sent to the Center for Genome Research and Biocomputing at Oregon State University. A total of ~400 ng of gDNA was used for library preparation and Illumina MiSeq sequencing (Illumina, San Diego, California, USA). A single, 300-bp paired-end MiSeq sequencing run resulted in 18,332,564 reads with an average length of 301 bases. Raw sequence reads were filtered, reformatted, and trimmed using the default commands of the Trimmomatic v.0.32 program (Bolger et al., 2014). Trimmomatic yielded 12,080,556 high-quality contigs (6,040,278 forward and reverse). Contigs were de novo assembled using Trinity (v. 07-04-2014) (Grabherr et al., 2011) producing 486,538 contigs, with an average size of 516 bp (N50 = 583). MSATCOMMANDER v.1.0.8 (Faircloth, 2008) found a total of 4118 perfect microsatellite repeats from the assembled contigs: 85 hexanucleotide with at least four repeat units, 585 pentanucleotide with at least four repeat units, 481 tetranucleotide with at least five repeat units, and 2967 trinucleotide with at least five repeat units. Dinucleotide repeats were excluded to avoid stutter and subsequent scoring problems in later analyses. The inset version of Primer3 (Rozen and Skaletsky, 2000) in MSATCOMMANDER designed 1153 primer pairs using default parameters (except that GC clamp = yes and repeat motif was ≥4 for penta- and hexanucleotide repeats). The PCR product size was set to 120–400 bp. One primer of each pair was designed with a common tag at the 5′ end following the procedure of Boutin-Ganache et al. (2001). Two common

[1]Manuscript received 30 April 2015; revision accepted 29 May 2015.

The authors thank the Colorado Federation of Garden Clubs and National Garden Club for funding this research. The University of Northern Colorado Field Botany course, BIO 329, collected *Calochortus gunnisonii* tissue samples from central Colorado populations. David Hardy assisted in DNA extractions.

[3]Author for correspondence: full6186@bears.unco.edu

doi:10.3732/apps.1500051

TABLE 1. Characteristics of 13 microsatellite loci designed for *Calochortus gunnisonii*.

Locus	Primer sequences (5′–3′)[a]	5′ Tag[b]	Repeat motif	Fluorescent dye	Allele size range (bp)	T_a (°C)	$MgCl_2$/$MgSO_4$[c]	GenBank accession no.
CAGU_14	F: TTGTCAAGTGGGCAAGTGTC	M13	$(ACACC)_4$	FAM	246–266	62.4	$MgCl_2$	KR139841
	*R: ATACAACGCACCGCATACC							
CAGU_15	*F: ATCCTCACTGCCCACTCAC	M13	$(AGAGG)_4$	VIC	385–415	62.4	$MgCl_2$	KR139842
	R: GTGCAGCAGATCCACATTCC							
CAGU_22	F: CACATGGTGTTGATGCAGGG	CAGT	$(ATCC)_5$	PET	237–257	62.4	$MgCl_2$	KR139843
	*R: TTGTATCGTGCAACAGTCCC							
CAGU_31	F: CCACCCAAGAAGAGCTAAAGG	CAGT	$(ACATC)_4$	PET	310–340	62.4	$MgCl_2$	KR139844
	*R: TTCACCGTCTCCAACAC							
CAGU_35	F: TAATACCCGTGAACTCCGGC	M13	$(ATGC)_5$	VIC	263–283	62.4	$MgCl_2$	KR139845
	*R: TGCAACCGAGTAGTGGACC							
CAGU_36	F: CCCTCGATCTCAGCCGATTG	M13	$(ATCG)_5$	FAM	378–398	62.4	$MgCl_2$	KR139846
	*R: TAAGTTAGGTACGGAGCAGGC							
CAGU_39	F: TTACACCAAGCTCCGCAG	CAGT	$(ACTG)_6$	FAM	335–374	62.4	$MgCl_2$	KR139847
	R: GTCTTGCTCATGTTGCTCCC							
CAGU_42	*F: TCGTGTTCAAGTGCTACCATCG	M13	$(AAGAC)_5$	PET	380–415	66.9	$MgCl_2$	KR139848
	R: CTTGCTGGAATCACATCACCC							
CAGU_45	*F: TCTAGTAGAAACCAGGGTGCC	M13	$(ACCG)_5$	FAM	190–214	62.4	$MgCl_2$	KR139849
	R: ACCCTAGCTCTTGTTCGGTG							
CAGU_46	F: GCACTCGATTTCAGAGCTGC	T7	$(ACTAT)_4$	PET	164–184	59.8	$MgSO_4$	KR139850
	*R: TTCAAGGATGGATGGCATCG							
CAGU_47	F: TCCCTCAGTCCAACGTC	CAGT	$(AGCTCC)_4$	VIC	345–381	61.1	$MgCl_2$	KR139851
	*R: GAACCTTCCTTTGTCCGCAC							
CAGU_48	F: TGCACCCTTAGAGACCATGG	CAGT	$(ACAGAT)_4$	FAM	406–448	61.1	$MgSO_4$	KR139852
	*R: GTCAAAGCATCCCGTCCTC							
CAGU_50	*F: TAGGGAGGGCTTCAGGAAC	CAGT	$(ACAGT)_5$	PET	364–409	56.9	$MgCl_2$	KR139853
	R: TAGGTGTCGGCAGGTCTAATG							

Note: T_a = PCR annealing temperature.

[a]The primer tagged with either M13R, CAGT, or T7 is indicated by an asterisk.

[b]The 5′ tag used for incorporation of the fluorescent tag: M13R (AGGAAACAGCTATGACCAT), T7 (GCTAGTTATTGCTCAGCGG), or CAGT (ACAGTCGGGCGTCATCA).

[c]$MgCl_2$ or $MgSO_4$ used in PCR reactions.

Table 2. Allelic diversity at 13 variable microsatellite loci for three populations of *Calochortus gunnisonii*.

	Dixon				Sand Lake				Shavano Camp			
Locus	*N*	*A*	H_o	H_e[a]	*N*	*A*	H_o	H_e[a]	*N*	*A*	H_o	H_e[a]
CAGU_14	31	4	0.469	0.469	31	3	0.233	0.376*	33	4	0.559	0.493
CAGU_15	31	8	0.563	0.607	32	4	0.419	0.348	34	3	0.571	0.519
CAGU_22	31	5	0.125	0.252*	32	5	0.387	0.356	33	5	0.441	0.440
CAGU_31	31	6	0.531	0.582*	32	5	0.645	0.674	34	5	0.457	0.602
CAGU_35	30	6	0.419	0.601	29	5	0.241	0.405*	28	6	0.393	0.695*
CAGU_36	30	5	0.387	0.455*	18	3	0.118	0.213	34	4	0.294	0.407*
CAGU_39	30	6	0.452	0.655	32	5	0.806	0.691	34	6	0.600	0.666*
CAGU_42	31	7	0.250	0.750*	30	4	0.448	0.542	25	5	0.077	0.643*
CAGU_45	31	7	0.344	0.616*	32	4	0.548	0.575	33	5	0.500	0.599*
CAGU_46	28	4	0.517	0.633	32	3	0.516	0.533	33	4	0.618	0.478
CAGU_47	30	5	0.387	0.363	31	4	0.400	0.575	33	3	0.118	0.452*
CAGU_48	30	8	0.871	0.763	28	5	0.692	0.648	34	6	0.686	0.782
CAGU_50	29	6	0.633	0.554	30	4	0.586	0.691	34	7	0.714	0.738
Mean	30.23	5.92	0.458	0.562	29.92	4.15	0.465	0.510	32.46	4.85	0.464	0.578

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

[a] Statistical significance associated with departure from Hardy–Weinberg equilibrium (HWE) is indicated with an asterisk (*$P \leq 0.05$).

tags, M13R (GGAAACAGCTATGACCAT) and CAGT (CAGTCGGGCGTCATCA), and one custom tag, T7 (GCTAGTTATTGCTCAGCGG), were used. Fifty loci with optimum primer conditions were screened for amplification success and within-population variability. Microsatellite loci were amplified in 12-μL reaction volumes that included 2.5 mM of tagged primer, 25 mM of nontagged primer, 25 mM of fluorescently labeled tag (6-FAM, PET, or VIC; Applied Biosystems, Wilmington, Delaware, USA), 5× Colorless GoTaq Flexi Buffer (Promega Corporation, Madison, Wisconsin, USA), 2.5 mM each dNTP, 0.02 units GoTaq Flexi DNA Polymerase (Promega Corporation), 0.1 μL 100× bovine serum albumin, 25 mM $MgCl_2$ or 100 mM $MgSO_4$ (Table 1), ~20–50 ng of gDNA, and brought to volume with purified water. Thermal cycler conditions were 94°C for 3 min; 35 cycles of 94°C for 30 s, 56–63°C for 30 s, and 72°C for 30 s; followed by an extension for 10 min at 72°C; and a final 30-min extension at 72°C. PCR products were mixed with 1× SYBR Green (LONZA, Rockland, Maine) and visualized on a 1% agarose gel. Successful PCR products were diluted with water and mixed with Hi-Di formamide and GeneScan 500 LIZ Size Standard (Applied Biosystems) before electrophoresis on an Applied Biosystems 3730 Genetic Analyzer at Arizona State University. Fragments were sized using the microsatellite plug-in of Geneious v.8.0.4 (created by Biomatters, http://www.geneious.com/).

Of the 50 initial screenings, 22 failed to yield reliable PCR product via gel electrophoresis. Twenty-eight loci yielded consistent products and were amplified. Preliminary sizing of the products revealed 13 primer pairs that consistently yielded amplicons within the target size range (120–400 bp) (Table 1). Three sample populations of *C. gunnisonii* were selected to evaluate variability in the isolated loci (Table 2). One population represents central Colorado (Shavano Campground; $n = 34$), the second represents a low-elevation population in northern Colorado (Dixon Reservoir; $n = 31$), and the third is a population from south-central Wyoming near the northern edge of the southern Rockies (Sand Lake; $n = 32$).

All 13 microsatellite loci were variable and polymorphic among sampled populations. Population statistics and deviations from Hardy–Weinberg equilibrium (HWE) were calculated using GenAlEx v.6.5 (Peakall and Smouse, 2012). The mean number of alleles per locus ranged from 4.15 to 5.92, with an average of 4.97 (Table 2). The observed and expected heterozygosity ranged from 0.077 to 0.871 and 0.213 to 0.782, respectively. Significant deviations from HWE were observed for individual loci, but no consistent pattern across multiple populations was observed, suggesting the deviations are due to population processes (Table 2). Observed and expected heterozygosities are high when examined across all populations, indicating the value of these markers for future studies.

All 13 loci were screened for amplification success in four additional *Calochortus* species spread throughout the *Calochortus* phylogeny (Patterson and Givnish, 2003). Cross-amplification success was high in the closely related *C. nuttallii* Torr. & A. Gray (11), moderate in *C. kennedyi* Porter var. *kennedyi* (8), and low in *C. flexuosus* S. Watson (7) and *C. subalpinus* Piper (4) (Appendix 2).

CONCLUSIONS

We identified 13 *C. gunnisonii* microsatellite loci that are variable and informative. These markers will be used to investigate the population genetic structure and levels of genetic variability of *C. gunnisonii* in the central and southern Rocky Mountains. Intra- and intermontane patterns of gene flow and divergence will be inferred within *C. gunnisonii*. Cross-species amplification was high in a closely related taxon, *C. nuttallii*, and decreased in more divergent sampled taxa.

LITERATURE CITED

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114–2120.

Boutin-Ganache, I., M. Raposo, M. Raymond, and C. F. Deschepper. 2001. M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *BioTechniques* 31: 24–28.

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

Fiedler, P. L., and R. K. Zebell. 2012. *Calochortus*: Flora of North America online. Available at website http://www.efloras.org/florataxon.aspx?flora_id=1taxon_id=105173 [accessed 15 January 2014].

Friar, E. A. 2005. Isolation of DNA from plants with large amounts of secondary metabolites. *Methods in Enzymology* 395: 1–12.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nature Biotechnology* 29: 644–652.

Henss, J. M., J. R. Moeller, T. J. Theim, and T. J. Givnish. 2013. Spatial scales of genetic structure and gene flow in *Calochortus albus* (Liliaceae). *Ecology and Evolution* 3: 1461–1470.

Ownbey, M. 1940. A monograph of the genus *Calochortus*. *Annals of the Missouri Botanical Garden* 27: 371–560.

Patterson, T. B., and T. J. Givnish. 2003. Geographic cohesion, chromosomal evolution, parallel adaptive radiations, and consequent floral adaptations in *Calochortus* (Calochortaceae): Evidence from a cpDNA sequence phylogeny. *New Phytologist* 161: 253–264.

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.

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.

Appendix 1. Voucher information for *Calochortus* species used in this study.

Species	Collector information[a]	Voucher specimen accession no.[b]	Collection locality[c]	Geographic coordinates	*N*
C. gunnisonii[d]	MEM 608	GREE 23640	Dixon Reservoir; Larimer County, CO	40.54135329°N, 105.1343894°W	31
C. gunnisonii[d]	RSF 47	GREE 23643	Sand Lake; Carbon County, WY	41.33943317°N, 106.1658852°W	32
C. gunnisonii[d]	MEM 669	GREE 23639	Shavano Campground; Chaffee County, CO	38.586457°N, 106.215402°W	34
C. gunnisonii[d]	RSF 36	GREE 23642	Hell Canyon Road; Custer County, SD	43.72462205°N, 103.8544076°W	1
C. gunnisonii[d]	MEM 621	GREE 23644	La Prele Reservoir; Converse County, WY	42.70679583°N, 105.5785312°W	1
C. nuttallii[d]	MEM 620	GREE 23646	Hell Canyon Road; Custer County, SD	43.727663°N, 103.847306°W	20
C. nuttallii[d]	MEM 631	GREE 23647	Shirley Basin; Carbon County, WY	42.341401°N, 106.411052°W	15
C. flexuosus[d]	MEM 644	GREE 23645	Big Gypsum Valley; San Miguel County, CO	38.068645°N, 108.791213°W	10
C. kennedyi var. *kennedyi*[e]	Lehto, Hensel & Pinkava 11039	GREE 20646	Oracle; Pima County, AZ	NA	1
C. subalpinus[e]	L. E. Detling 3581	GREE 20658	Gold Hill; Lane County, OR	44.1687015°N, 122.4452742°W	1

Note: *N* = number of individuals; NA = not available.
[a] MEM = Mitchell E. McGlaughlin; RSF = Ryan S. Fuller.
[b] Vouchers deposited at the University of Northern Colorado Herbarium (GREE), Greeley, Colorado, USA.
[c] Locality, county, and state in USA (AZ = Arizona; CO = Colorado; SD = South Dakota; WY = Wyoming).
[d] Fresh tissue used in DNA extractions, microsatellite population statistics, and cross-species amplifications.
[e] Herbarium specimen tissue used in cross-species amplifications.

Appendix 2. Cross-species amplification information for 13 microsatellite loci developed for *Calochortus gunnisonii* with four related *Calochortus* species.

Locus	*C. flexuosus*	*C. nuttallii*	*C. kennedyi* var. *kennedyi*	*C. subalpinus*
CAGU_14	—	—	—	—
CAGU_15	—	+	+	+
CAGU_22	+	+	+	—
CAGU_31	—	+	+	—
CAGU_35	+	+	+	+
CAGU_36	+	+	+	+
CAGU_39	+	+	+	—
CAGU_42	+	+	+	+
CAGU_45	—	+	—	—
CAGU_46	—	+	—	—
CAGU_47	—	—	+	—
CAGU_48	+	+	—	—
CAGU_50	+	+	—	—

Note: + = amplification product fell within the product size range for the locus; — = amplification product was absent or did not fall within the product size range expected for the locus.