Short title: *AJB* Primer Notes & Protocols – 13 Microsatellites in *Solanum rostratum*

Thirteen microsatellites developed by SSR-enriched pyrosequencing for *Solanum rostratum* (Solanaceae) and related species[1]

Mario Vallejo-Marín[2], Lislie Solis-Montero, Cecile F. E. Bacles and Olivier Lepais

Biological and Environmental Sciences, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, United Kingdom

Email addresses: MVM: mario.vallejo@stir.ac.uk

LSM: lislie.solismontero@stir.ac.uk

CFEB: bacles@hotmail.com

OL: lepais@gmail.com

Number of words: 1463. Abstract (115), Introduction (176), Methods and Results (790), Conclusions (85), and Literature cited (297)

[1]Manuscript received_____; revision accepted_____.

[2]Author for correspondence: mario.vallejo@stir.ac.uk

The authors thank B. Igic, G. Lye, C. Domínguez, J. Fornoni and R. Pérez for support during field work. This work was partially funded by a Horizon Ph.D. studentship from the University of Stirling to LSM, a Leverhulme Trust (Early Career Fellowship ECF/2010/0166) to OL, and by a Royal Society of London Research Grant (RG2010R1) and a Scottish Plant Health Licence (PH/38/2010) to MVM.

ABSTRACT

- *Premise of the study:* Microsatellite markers were developed using second-generation sequencing in *Solanum rostratum* as a tool to study the reproductive biology and genetic structure of this invasive species.
- *Methods and Results:* Thirteen microsatellites were successfully discovered and amplified in a single multiplexed PCR. All loci showed genetic variation in *S. rostratum*. Cross –amplification in five closely related taxa was successful for a subset of loci.
- *Conclusions:* The set of 13 microsatellite markers developed here provides a time and cost effective genetic tool to study the reproductive biology of *S. rostratum*. The demonstrated transferability of the PCR multiplex to related taxa also highlights its usefulness for evolutionary studies across *Solanum* Section *Androceras*.

Key words: invasive species; population genetics; reproductive biology; *Solanum rostratum; Solanum fructu-tecto; Solanum heterodoxum; Solanum grayi var. grayi; Solanum grayi var. grandiflorum; Solanum lumholtzianum; Solanum* Section *Androceras*.

INTRODUCTION

Solanum rostratum Dunal (Solanaceae) is a diploid, annual, self-compatible herb with weakly zygomorphic bee-pollinated nectarless yellow flowers (Whalen, 1979). It forms part of a clade of 12 species of Section *Androceras*, a group that has been used as a model to investigate the relationship between flower form, and reproductive isolation and mating patterns (e.g. Whalen, 1979, Vallejo-Marín et al., 2009). The native range of *S. rostratum* extends from Central Mexico to the U.S.A. (Whalen, 1979). However, it is now found in China, Russia, Australia, and Europe (Whalen, 1979; Lin and Tan, 2007; Vallejo-Marín, unpublished). The limited availability of genetic markers in *S. rostratum* currently thwarts studies on the reproductive biology and genetic structure of both native and invasive populations.

In this study, we describe 13 new microsatellite markers for *S. rostratum*, in order to enable further studies on its phylogeography and reproductive biology. We used second-generation

sequencing and bioinformatic tools to optimize a single microsatellite PCR multiplex (Guichoux et al., in press) for cost and time effective amplification of these markers in *S. rostratum* and related taxa.

METHODS AND RESULTS

Seven *S. rostratum* individuals were sampled from two populations (Tehuacán, 18.480° N, 97.411° W; Mexico City, 19.313° N, 99.178° W; Mexico; Appendix 1) Genomic DNA was isolated from silica-dried leaf tissue with Qiagen DNeasy Plant Mini kit and sent to Genoscreen (Lille, France) for microsatellite-enriched library preparation and sequencing by 454 GS FLX Titanium (Roche Applied Science) according to Malausa et al. (in press). Briefly, the pooled sample of seven individuals was subject to genomic DNA fragmentation, ligated to standard adapters, and enriched with eight microsatellite probes (TG, TC, AAC, AAG, AGG, ACG, ACAT, ACTC). The enriched DNA was then amplified using adapter-specific primers as described in Malausa et al. (in press). The resulting library was tagged with a specific multiplex identifier (MID) tag sequence, and pooled together with eight other samples in a quarter of a 454 GS FLX Titanium run for sequencing. The resulting 33,491 reads (average length = 254±107 bp; mean±SD) were analyzed with QDD v1.3 (Meglécz et al., 2010) to design microsatellite primers using selection criteria detailed in Lepais and Bacles (in press). These criteria were chosen to optimize potential for single PCR multiplexing of the designed primers, and included limiting the length of the expected PCR product to between 90 and 400 bp, optimal primer length of 24bp (range 21-30bp), optimal annealing temperature of 63°C (range 60-66°C), and 50% GC content (range 40-60%). 557 microsatellites were identified from which 355 had designed primers (Appendix S1).

Two screenings of 24 primer pairs were performed following the selection strategy of Lepais and Bacles (in press). In brief, microsatellite loci containing dinucleotide (AG and AC) and trinucleotide (AAC, AAG and AGG) repeat motifs were categorized in one of six expected PCR product size classes and ranked based on the number of motif repeats. In the first screening, a selection of 24 primer pairs representing all six size classes was chosen for testing in simplex PCR format on a panel of 19 *S. rostratum* individuals. Based on the results of this first screening, a new set of 24 primer

pairs was then selected to try to obtain successfully amplifying loci across all size classes, and screened in the same 19 individuals. Simplex PCR cycles consisted of a denaturing step of 5 min at 94°C followed by 30 cycles of 94°C for 30 s, 58°C for 45 s and 72°C for 45 s, and then eight cycles of 94°C for 30 s, 53°C for 45 s and 72°C for 45 s, and a final elongation step of 10 min at 72°C (Lepais and Bacles in press) Fragment analysis was performed on an ABI 3730xl capillary sequencer (Applied Biosystems) at DNA Sequencing & Services (Dundee, UK) and subsequently analyzed using STRand (VGL, UC Davis, CA). Out of 48 tested primer pairs, 29 successfully amplified, and 15 were polymorphic with repeatable profiles (Appendix S1).

Thirteen loci were found to be compatible for simultaneous PCR multiplexing using Multiplex Manager (Holleley and Geerts, 2009) and were evaluated using a panel of 38 *S. rostratum* individuals from two populations (Teotihuacán, 19.683° N, 98.858° W; Plan de Fierro, 18.333° N, 97.572° W; Mexico; Appendix 1). In addition, marker transferability and multiplex applicability were tested on 2 individuals from each of five taxa in *Solanum* Sect. *Androceras*: *S. fructu-tecto* Cav., *S. heterodoxum* Dunal, *S. grayi* var. *grandiflorum* Whalen, *S. grayi* var. *grayi* Whalen, and *S. lumholtzianum* Bartlett (Appendix 1). The multiplex PCR reaction was done using 1X Qiagen Type-it Microsatellite PCR Kit, various concentrations (Table 1) of each of the 13 fluorescent forward primers labeled with one of 6-FAM (Eurofins MWG Operon), VIC, PET or NED (Applied Biosystems) dyes and reverse primer and approximately 5 ng of template DNA. PCR cycles consisted of a denaturing step of 5 min at 95˚C, followed by 30 cycles of 95˚C for 30 s, 58˚C for 180s and 72˚C for 30s and a final elongation step of 30 min at 60˚C. Products were analyzed in an ABI3730xl capillary sequencer. Fluorescence profiles were analyzed using STRand and exported to MsatAllele (Alberto, 2009) in R version 2.12.0 (R Development Core Team, 2010) to determine suitable allele bin range.

All thirteen loci were polymorphic in at least one population with 2 to 13 alleles detected (Table 2), and showed moderate genetic diversity with expected heterozygosity ranging from 0.00 to 0.86 (Table 2).

All loci amplified in *S. fructu-tecto*; Sr21, Sr06 and Sr02 failed to amplify in S. *heterodoxum*; Sr21 and Sr06 did not amplify in *S. grayi* var. *grayi*; Sr21, Sr06 and Sr02 did not amplify in *S. grayi*

var. *grandiflorum*; and Sr21, Sr06 and Sr26 failed to amplify in *S. lumholtzianum*. Importantly, loci that amplified in these taxa, did so within the expected size range thus demonstrating the transferability of the multiplex protocol.

CONCLUSIONS

Second-generation sequencing and novel bioinformatic approaches are very effective tools to isolate microsatellite markers in non-model organisms. This allows discovery of numerous microsatellites that can be combined in one or few PCR reactions, reducing both time and cost of genotyping (Lepais and Bacles, in press). Here we developed a set of 13 polymorphic microsatellite markers for *S. rostratum* that can be amplified in a single multiplexed PCR, and demonstrated its potential use in related taxa, thus enabling future investigation of numerous ecological and evolutionary questions.

LITERATURE CITED

ALBERTO, F. 2009. MsatAllele_1.0: An R package to visualize the binning of microsatellite alleles. *Journal of Heredity* 100: 394-397.

GUICHOUX, E., L. LAGACHE, S. WAGNER, P. CHAUMEIL, P. LÉGER, O. LEPAIS, C. LEPOITTEVIN, T. MALAUSA, E. REVARDEL, F. SALIN, AND R. J. PETIT. 2011. Current trends in microsatellite genotyping. *Molecular Ecology Resources* in press.

HOLLELEY, C. E., AND P. G. GEERTS. 2009. Multiplex Manager 1.0: a cross-platform computer program that plans and optimizes multiplex PCR. *Biotechniques* 46: 511-517

LEPAIS, O., AND C. F. E. BACLES. 2011. Comparison of random and SSR-enriched shotgun pyrosequencing for microsatellite discovery and single multiplex PCR optimisation in *Acacia harpophylla* F. Muell. Ex Benth. *Molecular Ecology Resources* in press. DOI: 10.1111/j.1755-0998.2011.03002.x

LIN, Y. AND D. TAN. 2007. The potential and exotic invasive plant: *Solanum rostratum*. *Acta Phytotaxonomica Sinica* 45: 675-685.

MALAUSA, T., A. GILLES, E. MEGLÉCZ, H. BLANQUART, S. DUTHOY, C. COSTEDOAT, V. DUBUT, N. PECH, P. CASTAGNONE-SERENO, C DÉLYE, N. FEAU, P. FREY, P. GAUTHIER, T. GUILLEMAUD, L. HAZARD, V. LE CORRE, B. LUNG-ESCARMANT, P. MALÉ, S. FERREIRA, AND J.-F. MARTIN. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. *Molecular Ecology Resources* in press.

MEGLÉCZ, E., C. COSTEDOAT, V. DUBUT, A. GILLES, T. MALAUSA, N. PECH, AND J.-F. MARTIN. 2010. QDD: a user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics* 26: 403-404.

R DEVELOPMENT CORE TEAM. 2010. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org [accessed 02 February 2011].

VALLEJO-MARÍN, M., J. S. MANSON, J. D. THOMSON, AND S. C. H. BARRETT. 2009. Division of labour within flowers: heteranthery, a floral strategy to reconcile contrasting pollen fates. *Journal of Evolutionary Biology* 22: 828:839.

WHALEN, M. 1979. Taxonomy of *Solanum* section *Androceras*. *Gentes Herbarum* 11:359-426.

Tables

Table 1. Characteristics of 13 microsatellite primers developed in *Solanum rostratum* and optimized to co-amplify in a single multiplex PCR.

Locus	Repeat type	EMBL accession	Primer sequences (5' -> 3')		Dye	[Primer] (nM)	Size range (bp)
Sr09	$(ac)_8$	FR846150	F:	TCACTTTGAGACCCCTAACACCTC	FAM	170	204-214
			R:	TAAGAGGAACAGGAAGAAGAGGGC			
Sr18	$(ca)_6$	FR846159	F:	AATCACCCACCTACTGTGACGTTT	FAM	170	292-310
			R:	ATCCAGTGCTTGTGTTGATAGGCT			
Sr30	$(tc)_8$	FR846171	F:	ATGCTCCCCATTTTCCATTTTC	FAM	120	109-117
			R:	ATCTGCTGAGAAGTTGAATTTCCG			
Sr33	$(gt)_6$	FR846174	F:	ATACTTCATTTGTTGCAGGAGCTG	FAM	340	141-167
			R:	CAAAAGCTAAAACCCAAGACAGGA			
Sr06	$(ag)_8$	FR846147	F:	ATGAGGACCCAGTTGAGTTTCTTG	VIC	340	190-206
			R:	CTTTAAATTCCTCCCATCCAGCTC			
Sr22	$(aac)_6$	FR846163	F:	CTAACAATTTCTCCAACAACCTTGG	VIC	170	346-358
			R:	CCAAAACTTTCACCAGAAAACTCAC			
Sr26	$(ct)_9$	FR846167	F:	GCTATTTCCCCTACTCCGGTTCTT	VIC	120	107-141
			R:	GTAGGTGCCCAAATATTGATCCAG			
Sr05	$(tc)_9$	FR846146	F:	CTGAATGTTGTAATTGGGTGTCCA	NED	340	173-199
			R:	ACAAGAACCGAAAACGAAGAACAG			
Sr21	$(aac)_8$	FR846162	F:	GGTCGATTGCCTCTATCTACTGTTG	NED	200	370-378
			R:	TGGTAGTGGTAAGGTCTGCGTACA			
Sr31	$(tc)_7$	FR846172	F:	AACTCAGCCATAGTTCCAGACACC	NED	170	96-112
			R:	AGAGGTGCTGGAGTTGAGAAAAGA			
Sr38	$(gaa)_6$	FR846179	F:	GATCTCAAAGAAGGGTCTCCCCTA	NED	170	256-260
			R:	AGTGCAGAAAATGAAGTGCTCTGG			
Sr02	$(ct)_{13}$	FR846143	F:	GGAATAGAGGGAGTTATACAGAAT ACACGA	PET	200	96-164
			R:	GGCGAGACCAGTTCTTGTCATATT			
Sr12	$(tc)_7$	FR846153	F:	GGTTAGGCCCAAACGTTGAAATAA	PET	170	217-223
			R:	ACCAGAGATGGATCAAACTTCAGC			

Notes: Shown for each primer pair are the repeated motif type, the accession number at the European Molecular Biology Laboratory—Nucleotide Sequence Database (EMBL-Bank; www.ebi.ac.uk/embl/), the forward and the reverse primer sequence, the fluorescent dye added to the 5’ end of the forward primer, the final primer concentration ([Primer]) in the PCR mixture (nM) and the allele size range (bp).

157 **Table 2.** Results of initial loci screening in two populations of *Solanum rostratum*. N = Number of
158 genotyped individuals, N_a = number of alleles; H_e = expected heterozygosity. Population 1 = Teotihuacán,
159 Estado de México; Population 2 = Plan de Fierro, Puebla.

	Population 1 (N=15)		Population 2 (N=23)		Total
Loci	N_a	H_e	N_a	H_e	N_a
Sr09	2	0.238	3	0.343	4
Sr18	2	0.186	6	0.783	6
Sr30	3	0.476	3	0.573	5
Sr33	4	0.612	4	0.489	5
Sr06	4	0.667	5	0.612	6
Sr22	4	0.352	3	0.606	4
Sr26	4	0.531	5	0.501	6
Sr05	8	0.852	6	0.754	12
Sr21	2	0.457	3	0.625	3
Sr31	3	0.440	6	0.792	8
Sr38	1	0.00	2	0.417	2
Sr02	7	0.660	9	0.862	13
Sr12	3	0.676	5	0.543	5

160

161

Figure Legend

Fig. 1: Example of a typical electropherogram profile obtained for one individual with the multiplex PCR genotyping protocol presented here (**a**), and diagram showing the allele size range and fluorescent dyes of each of the 13 loci (**b**). In (**a**), down-turned triangles indicate alleles at each locus; fragments sizes (bp) of the 500 LIZ size standards are indicated by numbers above each corresponding peaks. In (**b)**, dark rectangles represent the observed allele range in 34 *S. rostratum* individuals; light rectangles represent an arbitrary potential allele range used during the multiplex design to avoid overlap of loci with the same fluorescent dye.

Appendix 1

Appendix 1. Voucher information for taxa used in this study. All vouchers deposited at the University of Stirling.

Species—Country and Locality, Accession number
***Solanum rostratum* Dunal** — Mexico, Tehuacán, Puebla, 08s104.
***Solanum rostratum* Dunal** — Mexico, Mexico City, Distrito Federal, 10s110.
***Solanum rostratum* Dunal** — Mexico, Plan de Fierro, Puebla, TP-8.
***Solanum rostratum* Dunal** — Mexico, Teotihuacán, Estado de México, TEM-19.
***Solanum fructu-tecto* Cav.** — Mexico, Atitalaquia, Hidalgo, AH-9
***Solanum heterodoxum* Dunal** — Mexico, Fresnillo, Zacatecas, FZ-24
***Solanum grayi var. grandiflorum* Whalen** — Mexico, Los Zapotes, Sinaloa, 07s197
***Solanum grayi var. grayi* Whalen** — Mexico, Los Álamos, Sonora, 07s189
***Solanum lumholtzianum* Bartlett** — Mexico, El Progreso, Sinaloa, 07s41

Online Supplementary Material

Figure S1. Histogram of read lengths obtained from the 454 GS FLX Titanium sequencing for *Solanum rostratum.*

Appendix S1. List and detailed characteristics of the 355 microsatellites with designed primers identified by the bioinformatics analysis with annotations indicating the criteria used to select primers pairs to screen and results of the initial screening