Development of microsatellite markers for the threatened species *Coleocephalocereus purpureus* (Cactaceae) using next-generation sequencing

Daphne Amaral Fraga¹, Anderson Figueiredo de Carvalho¹, Ricardo Souza Santana²,
Marlon Câmara Machado¹, Gustavo Augusto Lacorte¹.

1. Molecular Biology Lab, Department of Science and Languages, Federal Institute of
Minas Gerais – Bambuí Campus, Bambuí, Minas Gerais State, Brazil;

2. Nativa Serviços Ambientais Ltda, Curvelo, Minas Gerais State, Brazil;

Corresponding author: Gustavo Augusto Lacorte; e-mail: gustavo.lacorte@ifmg.edu.br;

ORCID iD: 0000-0002-6866-348X; Phone number: +55 37 3431-5600.

Abstract

Ten microsatellite loci were developed and validated for the endangered cactus species
Coleocephalocereus purpureus. The markers were obtained from sequences generated by
whole genome shotgun sequencing approaches. A testing group of 36 specimens of the
main grouping were genotyped and all described markers presented suitable outcomes to
population genetic studies, showing polymorphic status for *C. purpureus* testing group
with clean and reproducible amplification. No evidence for scoring errors, null alleles or
linkage disequilibrium was detected. Number of alleles per locus ranged from 3 to 6 and
expected heterozygosity ranged from 0.78 to 0.99. These new microsatellite loci are
suitable to be used in future diversity and structure population studies of *C. purpureus*.

Keywords: Microsatellites; Population genetics; Cactus; Endangered species.

Author information

Daphne Amaral Fraga¹

e-mail: daphnefraga@hotmail.com

Anderson Figueiredo de Carvalho¹

e-mail: anderson_fbi45@outlook.com

Ricardo Souza Santana²

e-mail: souzasantana@hotmail.com

ORCID ID: 0000-0002-1107-3040

Marlon Câmara Machado¹

e-mail: marloncmachado@gmail.com

Corresponding author: Gustavo Augusto Lacorte¹

e-mail: gustavo.lacorte@ifmg.edu.br

ORCID iD: 0000-0002-6866-348X

Introduction

Coleocephalocereus purpureus (Buining & Brederoo) F. Ritter is a rare columnar cactus species, reaching up to 90 cm high, with grey wool and golden yellow to brown bristles, that forms a lateral cephalium up to 50 cm long near the stem's tip with pink melocactus-like flowers (Figure 1) [1]. As a typical columnar cactus, *C. purpureus* specimens provide nectar, pollen and fruits to a broad spectrum of bat, bird, and insect species, playing a crucial support ecological role of a fragile arid ecosystem [2].

The species was ranked by IUCN as “critically endangered” due its high endemism, whose occurrence of populations (or even of a single population, the population boundaries is unknown) is limited to only four small habitat patches restricted to inland cliffs and mountain peaks placed in a small area of Caatinga biome from Southeast Brazil (Figure 2), as well due the presence of quarrying activities of the rock where the species grows and other minor human impacts such as collection, deforestation, introduction of invasive species, fire and cattle trampling [3].

Due to the high risk of extinction of the species imposed by quarrying activities in its home range, a conservation unit (named “RPPN Pasmado”) has recently been established to protect one of the main native populations of *C. purpureus* as well as to harbour rescued individuals from areas where quarrying activities has been authorized and all vegetation has been removed [4]. Since the translocation programs were implemented, the RPPN Pasmado Conservation Unit became the largest grouping of *C. purpureus* constituted by native individuals as well translocated individuals from other populational groups that became extinct due to quarrying activity. In addition, as the other three habitat patches harbouring *C. purpureus* individuals are located in unprotected areas of high value for

quarrying activities, the RPPN Pasmado Conservation Unit has become the unique expectance of preservation of this species.

Strategies for genetic variability retention of the remaining populations is essential for the persistence of endangered species since low levels of genetic variation decrease population resilience and are associated to fitness decreasing by inbreeding depression [5]. Some genetic studies on cacti species using microsatellite molecular markers have provided insights into genetic diversity, demographic history, population structure and connectivity, which could offer critical guidance for conservation and management strategies [6-8].

The traditional methods for microsatellite isolation were expensive and time-consuming task, drawbacks that often reduced the use microsatellite markers in genetic studies of natural populations [9]. Recently, Next-generation sequencing (NGS), specifically whole genome shotgun sequencing approaches, has become a viable alternative for microsatellite markers isolation since NGS methods comprise faster and less costly approaches besides to generating thousands of markers per procedure [10]. Here, we use whole genome shotgun sequencing approach to develop a set of microsatellite markers for the endemic and endangered species *Coleocephalocereus purpureus* for application in population genetic studies, as a way to guide conservation and management strategies of *C. purpureus* species.

Material and methods

One *C. purpureus* specimen collected from the RPPN Pasmado was selected for whole genome shotgun sequencing. Total genomic DNA was extracted from a fresh stem slice using NucleoSpin® Plant II DNA extraction kit (Macherey-Nagel, Düren, Germany) and quantified using a Qubit 2.0 device (Life Technologies, Carlsbad, CA, USA). Genomic

libraries was constructed using Illumina TruSeq kit (Illumina Inc., San Diego, CA) and were run in a Illumina NextSeq Sequencer (Illumina Inc., San Diego, CA) from BPI Sequencing Facility company (Botucatu, SP. Brazil). The microsatellite motifs were identified from sequences obtained by NGS with MSATCOMMANDER software [11], indicating to the program to search for di, tri, tetra and pentanucleotide microsatellites motifs. Sequencing reads including microsatellite motifs with following specifications were selected to primer design using PRIMER3 [12]: (1) perfect motif microsatellite ≥ 5 tandems repeats; (2) melting temperature around 54°C and with a maximum of 2°C difference between melting temperature for each paired primers; (3) PCR product between 100 and 300 pb length; (4) GC content > 40%; (5) no complementarity between the primers.

A testing group of 36 *C. purpureus* specimens from RPPN Pasmado Conservation Unit were selected and a small stem slice from each individual were collected to DNA extraction using NucleoSpin® Plant II DNA extraction kit (Macherey-Nagel, Düren, Germany). Based on the five defined criteria for primer design, Forty-four primer pair were recovered from PRIMER3 results and subsequently manufactured with 6-FAM™ fluorophore labelling in order to perform PCR optimization and polymorphism tests. The amplification of testing group DNA was performed using a Mastercycler gradient thermocycler (Eppendorf, Germany) and a 25 µL volume reaction containing 50 ng/µL of DNA, 5 µL of PCR Buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl₂), 0.4 µM of each primer; 0.2 mM dNTPs, 2 units of Taq DNA Polymerase and 6 µL of MiliQ H₂O. The amplification program included an initial denaturation at 95 °C for 4 min, 30 cycles at 95°C (45 s)/54°C (45 s)/72°C (45 s) and final extension at 72°C for 7 min. Amplification products were visualized in ABI 3730 Genetic Analyzer (Applied

Biossystem®) using GeneScan™ 500 LIZ® Size Standard v2.0 (Applied Biosystem®).

The genotypes were obtained with Peak Scanner v1.0 software (Applied Biosystems®).

The genetic diversity was evaluated estimating the number of alleles per locus (N_A),

number effective of alleles per locus (N_E), expected and observed heterozygosity (H_E and

H_O , respectively) using the program ARLEQUIN V 3.5.1.2 [13]. In addition, deviations

from Hardy–Weinberg equilibrium (HWE) and linkage disequilibrium were tested also

using ARLEQUIN V 3.5.1.2 [13]. Frequencies of null alleles were estimated using

FREENA [14]. DNA sequences of microsatellite loci described here were deposited at

GenBank Database with accession numbers MN200436 to MN200445.

Specimen collection for this study was authorized by Brazilian Environmental Agency

(ICMBio) licence number 58257-1. The access to *C. purpureus* genetic material

performed in this study was properly registered in official database of Brazilian genetic

patrimony – SISGEN – with access number A8AFD21.

Results and discussion

This is the first time that microsatellite markers for *C. purpureus* were isolated and the

NGS approach used in this study generated thousands of sequence reads containing

microsatellite motifs (Figure 3). In total, it was generated 17,895,688 reads, of which

14769 sequences containing microsatellite motifs were identified. However, most of

sequences containing microsatellite motifs were not suitable to be considered a potential

microsatellite molecular marker because they did not present perfect (or even almost

perfect) microsatellite motifs, or more than 5 tandems repeats, or sequence length > 100

bp. So, only about two thousand of sequence reads really could be considered potentially

amplifiable microsatellite loci.

Of all the almost 2,000 potential candidate loci, we selected only those that were accepted by the parameters implemented in the PRIMER3 primer design program, resulting in a list of 44 microsatellite loci. It is important to highlight that in this study we included melting temperature parameter in PRIMER3 software run just to simplify the process of PCR optimization. Thus, more microsatellite loci could have been characterized if the annealing temperature parameter were relaxed or not pre-established. Our results of recovering potentially amplifiable loci, using the whole genome shotgun technique, showed results similar to those performed with the same technique and purpose but for different plant species [15-18]. Our results support the robustness and applicability of NGS approaches for microsatellite loci characterization.

Of the forty-four microsatellite loci selected for the initial tests, ten of them presented suitable outcomes to population genetic studies, showing polymorphic status for *C. purpureus* testing group with clean and reproducible amplification (Table 1). The microsatellite loci presented different patterns of genetic diversity on testing group of 36 *C. purpureus* specimens designed for this study, which the number of alleles per locus ranged from 3 (Cpur-1) to 6 (Cpur-10), the observed heterozygosity (H_o) among loci ranged from 0.22 (Cpur-8) to 0.92 (Cpur-3) and expected heterozygosity (H_E) ranged from 0.78 (Cpur-8) to 0.99 (Cpur-3). Six loci presented HWE deviation attributed to deficit of heterozygotes, which is expected to small populations [19]. No evidence of scoring errors were detected in the genotyping by stuttering or due to large allele dropout as well no evidence of null alleles was verified. No significant linkage disequilibrium was detected. These markers information could be useful for describing the genetic diversity and population structure of the remnant *C. purpureus* populations. They will also be used for understanding patterns of gene flow and potential connectivity among the populations. All this information is crucial for the design of management plans for this threatened

species aiming the retention of the residual genetic diversity present in the remnant populations, maximizing the resilience and preventing the effects of inbreeding depression.

Acknowledgements

The authors thank the Program for Technological Development in Tools for Health-PDTIS FIOCRUZ for use of its facilities and Nativa Serviços Ambientais Ltda for logistic support during collections of biological material.

References

1. Taylor NP, Zappi DC (2004). Cacti of eastern Brazil. Royal Botanic Gardens, Kew, 2004.
2. Felger, R. S., & Moser, M. B. (1985). People of the Desert and Sea: Ethnobotany of the Seri Indians. xv+ 438 pp. Tucson: University of Arizona Press.
3. Taylor NP, Machado M, Braun P (2013) *Coleocephalocereus purpureus*. The IUCN Red List of Threatened Species 2013: e.T40913A2943088. <http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T40913A2943088.en>.
4. IEF-MG (2010) Portaria IEF n° 149, de 02 de agosto de 2010. <http://www.siam.mg.gov.br/sla/action/consultaPublicacoes.do>. Accessed 29 October 2019
5. Frankham R (2005). Genetics and extinction. Biological conservation 126:131-140. <https://doi.org/10.1016/j.biocon.2005.05.002>
6. Solórzano S, Dávila P (2015) Identification of conservation units of *Mammillaria crucigera* (Cactaceae): perspectives for the conservation of rare species. Plant Ecology & Diversity 8:559-569. <https://doi.org/10.1111/j.1755-0998.2008.02422.x>
7. Maya-García R, Arizaga S, Cuevas-Reyes P, Peñaloza-Ramírez JM, Ramírez VR, Oyama K (2017) Landscape genetics reveals inbreeding and genetic bottlenecks in the extremely rare short-globose cacti *Mammillaria pectinifera* (Cactaceae) as a result of habitat fragmentation. Plant diversity 39:13-19. <https://doi.org/10.1016/j.pld.2016.09.005>
8. Khan G, Ribeiro PM, Bonatelli IA, Perez MF, Franco FF, Moraes EM (2018) Weak grouping structure and no genetic erosion in *Pilosocereus aureispinus*: A

- microendemic and threatened cactus species from eastern Brazil. PloS One 13:e0195475. <https://doi.org/10.1371/journal.pone.0195475>
9. Vieira MLC, Santini L, Diniz AL, Munhoz, CDF (2016) Microsatellite markers: what they mean and why they are so useful. Genetics and molecular biology 39:312-328. <http://dx.doi.org/10.1590/1678-4685-GMB-2016-0027>
10. Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99:193–208. <https://doi.org/10.3732/ajb.1100394>.
11. Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol Resour 8:92–94. <https://doi.org/10.1111/j.1471-8286.2007.01884.x>
12. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic acids research 40:e115-e115. <https://doi.org/10.1093/nar/gks596>.
13. Excoffier L, Lischer H (2010) Arlequin suite V.3.5: a new series of programs to perform grouping genetics analyses under Linux and Windows. Mol Ecol Resour 10(3):564–567. <https://doi.org/10.1111/j.1755-0998.2010.02847.x>
14. Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of grouping differentiation. Mol Biol Evol 24:621–631. <https://doi.org/10.1093/molbev/msl191>.
15. Morillo, E., Buitron, J., Limongi, R., Vignes, H., & Argout, X. (2016). Characterization of microsatellites identified by next-generation sequencing in the Neotropical tree *Handroanthus billbergii* (Bignoniaceae). Applications in plant sciences, 4(5), 1500135. <https://doi.org/10.3732/apps.1500135>
16. Aranguren-Díaz, Y. C., Varani, A. M., Michael, T. P., & Miranda, V. F. (2018). Development of microsatellite markers for the carnivorous plant *Genlisea aurea* (Lentibulariaceae) using genomics data of NGS. Molecular biology reports, 45(1), 57-61. <https://doi.org/10.1007/s11033-017-4140-1>
17. Bastías, A., Correa, F., Rojas, P., Almada, R., Munoz, C., & Sagredo, B. (2016). Identification and characterization of microsatellite loci in maqui (*Aristotelia chilensis* [Molina] Stunz) using next-generation sequencing (NGS). PloS one, 11(7), e0159825. <https://doi.org/10.1371/journal.pone.0159825>
18. Zhang, X., Zhou, Y., Li, Y. L., & Liu, J. X. (2019). Development of microsatellite markers for the seagrass *Zostera japonica* using next-generation sequencing. Molecular biology reports, 46(1), 1335-1341. <https://doi.org/10.1007/s11033-018-4491-2>

19. Young A, Boyle T, Brown T. (1996) The grouping genetic consequences of habitat fragmentation for plants. Trends in ecology & evolution 11:413-418.
[https://doi.org/10.1016/0169-5347\(96\)10045-8](https://doi.org/10.1016/0169-5347(96)10045-8).

Figure 1. Images of *Coleocephalocereus purpureus*. A, a typical individual of *C. purpureus* with several columnar stems. B, a typical cephalium near the stem's tip. C, pink melocactus-like flowers.

Figure 2. *C. purpureus* distribution. Natural color Google Earth image showing the location of the known occurrence areas of *C. purpureus* remnant individuals. Unprotected areas is represented by red dots and the protected and managed area by a yellow dot. Jequitinhonha and Itinga rivers are depicted in blue. Map of eastern Brazil showing the area studied within the red square (on the left).

Figure 3: Number of sequence reads containing potentially microsatellite loci (on the top of each bar is the exactly number of reads).

337 Table 1. Summary data for 10 microsatellite loci developed for *C. purpureus*

Locus	Primer sequence (5' - 3')	Repeat motif	Size (bp)	N _A	N _E	H _O	H _E	HWE	FN
Cpur-1	F: CGAACAGCAGATCATCGACT	(CGCCG) ₅ *	236-256	3	2,83	0.608	0.905	S	0
	R: ACCTGGGCGAACTGACGAT								0
Cpur-2	F: AACAGCAACTGCAACAATCG	(ATGT) ₈	228-240	4	3,40	0.405	0.829	S	0
	R: CCCTTGTTCTTCTGTAAACCC								0
Cpur-3	F: GCACCACCAATCCTTCCAT	(AG) ₁₃	206-214	5	4,85	0.921	0.994	NS	0
	R: TTCGTTCTAGTTCGCCACAG								0
Cpur-4	F: AAAGGAGATGTGCACTCACG	(AG) ₁₄	157-167	4	3,17	0.730	0.952	NS	0
	R: ATTAGCCGTACCGTCAGAGT								0
Cpur-5	F: AACAGCCTAAAGAAGCCCAG	(AG) ₁₅	152-162	5	3,11	0.605	0.925	S	0
	R: ATTCTGTGCGCTATGGTGC								0
Cpur-6	F: GTAACGGTTGGGAATGGGAA	(GAT) ₁₂	230-236	3	2,98	0.833	0.972	NS	0
	R: AAGTCCTCCCTCAGTTCCAA								0
Cpur-7	F: CAAACACAGCAACAGCACA	(GA) ₁₀	236-246	4	2,15	0.875	0.991	NS	0
	R: TTTGGACACTGTTGGACTCG								0
Cpur-8	F: CCCTGAACCAACCTGTGTT	(TC) ₁₈	230-240	6	4,51	0.229	0.785	S	0
	R: GGGTTCAGGTCTTTGGACAT								0
Cpur-9	F: GTGATGACCTCGTCTTTCGA	(TTA) ₁₂	213-221	4	2,78	0.467	0.849	S	0
	R: TGGGTATTAAAGGCTCACGA								0
Cpur-10	F: GTGCTGATGAGTGTGTGACA	(ATGA) ₉	245-257	4	2,97	0.676	0.895	S	0
	R: AATGAATGCACCGACCTCAC								0

338 N_A = number of alleles, N_E = number effective of alleles; H_O and H_E = observed and expected heterozygosity
339 (respectively), HWE = deviation from Hardy-Weinberg Equilibrium (S significant and NS not significant), FN = null
340 allele frequency.*Maximum number of motif repeats.

341

342