Molecular Biology Reports (2019) 46:4593–4597
https://doi.org/10.1007/s11033-019-04819-2

SHORT COMMUNICATION

Development of ten microsatellite markers for *Alibertia edulis* (Rubiaceae), a Brazilian savanna tree species

Liliane G. Dantas[1,3] · Lucas Alencar[1,4] · Bruno Huettel[2] · Andrea Pedrosa-Harand[1]

Received: 15 December 2018 / Accepted: 12 April 2019 / Published online: 30 April 2019

Abstract

Ten microsatellite markers were developed using next-generation sequencing data for *Alibertia edulis* (Rubiaceae), a widely distributed species typical of Cerrado (Brazilian savanna) vegetation. The markers were polymorphic in the two populations analyzed. The numbers of alleles, and observed (H_O) and expected (H_E) heterozygosities per polymorphic locus ranged from 2 to 11, 0.091 to 1.0, and 0.100 to 0.937 respectively. The SSR loci demonstrated moderate to high polymorphism values in both populations analyzed, with PIC values ranging from 0.26 to 0.91, and total allele numbers ranging from three to 16. The inbreeding coefficient values were generally higher in the Piauí population (ranging from −0.593 to 0.762) than in the Mato Grosso population (ranging from −1 to 0.575). The differences observed between those disjunct populations suggest they harbor different alleles, which has implications for Cerrado conservation strategies. Those loci will be useful for population studies of *A. edulis*.

Keywords NGS · SSR loci · Cerrado · Population genetics · Rubiaceae

Introduction

Alibertia edulis (Rich.) A.Rich. ex DC. (Rubiaceae), commonly known as black guava or small cerrado quince, is a tree species with unisexual white flowers and succulent fruits that can grow to eight meters in height [1, 2]. It is a widely distributed species in the Cerrado (Brazilian savanna) [3] and therefore an interesting taxon for phylogeographic analyses, especially of less-studied or marginal savanna populations in the Amazon region and northeastern Brazil.

Brazilian Cerrado vegetation covers approximately 2 million km^2, with a wide variety of climates and soils [4, 5] that contribute to its high floristic diversity, richness, and levels of endemism [6]. More than half of its area has been converted to agriculture and livestock in the past 35 years, however—three times more area than lost in the Amazon forest during the same period [7]. Accordingly, the Cerrado was classified by Mittermeir et al. [8] as one of the world's 34 priority hotspots for conservation.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11033-019-04819-2) contains supplementary material, which is available to authorized users.

✉ Liliane G. Dantas
liliane.dantas@univasf.edu.br

Lucas Alencar
alencar.lucasc@gmail.com

Bruno Huettel
huettel@mpipz.mpg.de

Andrea Pedrosa-Harand
andrea.pedrosaharand@pq.cnpq.br

[1] Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Centre of Biosciences, Federal University of Pernambuco, R. Prof. Moraes Rego, s/n, CDU, Recife, PE 50670-420, Brazil

[2] Max Planck Genome Centre Cologne, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, 50829 Cologne, Germany

[3] Present Address: Natural Sciences Collegiate, São Francisco Valley Federal University, Av. Tomaz Guimarães, S/N Santos Dumont, Senhor do Bonfim, BA 48970-000, Brazil

[4] Present Address: Laboratório de Ecologia Vegetal Aplicada (LEVA), Department of Botany, Centre of Biosciences, Federal University of Pernambuco, R. Prof. Moraes Rego, s/n, CDU, Recife, PE 50670-420, Brazil

Increasing numbers of molecular markers have been developed to better understand relationships between plant populations, including Cerrado species [e.g., 9, 10]. Microsatellites (also called simple sequence repeats—SSRs) are commonly employed markers for population studies as they are co-dominant and demonstrate high levels of intraspecific polymorphism and random distributions within the genome [11]. A partial knowledge of a species' genome is usually necessary to develop loci-specific primers and access SSR polymorphism [12]. The availability of more accessible next-generation genomic sequencing techniques (NGS) has enabled the efficient development of molecular markers (including SSRs) that can be used in population analyses [e.g., 10, 13]. No molecular markers, however, have yet been developed for the tree species *A. edulis*.

The present work therefore sought to develop SSR markers for *A. edulis* that would be useful for population analysis at several scales, including examinations of the phylogeographic relationships between different Cerrado areas and the identification of priority areas for conservation.

Materials and methods

DNA extraction and library construction

Genomic DNA of an *A. edulis* specimen from Paraopeba, Minas Gerais (19.26°S–44.40°W), was extracted from silica-dried leaves using a conventional CTAB extraction protocol [14]. A 5 µg DNA aliquot was ultrasonically fragmented (COVARIS S2 system) and genomic libraries prepared for pyrosequencing, following the manufacturer's instructions (Roche Diagnostics, Mannheim, Germany). The libraries were quantified by fluorometry (Qubit dsDNA HS Assay Kit, Invitrogen, USA) and shotgun-sequenced on a Roche 454 GS-FLX with the Titanium Sequencing Kit XLR70 (Roche Diagnostics). The sequence reads generated were aligned using CodonCode Aligner 4.0.4 DEMO software (CodonCode Corp., Dedham, Mass., USA).

SSR loci searching

MsatCommander 1.0.2-alpha [15] was used to search for sequence contigs and singletons for SSRs. The minimum thresholds adopted were: di- and trinucleotides with at least six repeat units, tetranucleotides with at least five repeat units, pentanucleotides with at least four repeat units, and hexanucleotides with at least three repeat units. Compound SSRs were also considered. A GenBank search was conducted for all SSR sequences to determine whether the sequences were mitochondrial, chloroplastidial, or associated with highly repetitive genomic sequences (e.g., retrotransposons) of other sequenced species.

Primer design and SSR loci screening

Primer design was performed using the Primer3 tool, implemented in MsatCommander; and the best primer pairs were selected according to the following parameters: product size range (150–300 bp), Tm (50–65 °C), and the absence of dimers or hairpins. Primer pairs were tested in five randomly chosen individuals from five populations across the geographic distribution of *A. edulis* in the Cerrado, whose total genomic DNA was extracted from silica-dried leaves, following a CTAB protocol [16]. PCR was carried out in 10 µL volumes using an Eppendorf Mastercycler (Eppendorf, Hamburg, Germany) or an Applied Biosystems Veriti thermocycler (Applied Biosystems, USA). Each assay contained approximately 10 ng of genomic DNA, 1 × PCR reaction buffer (Invitrogen), 0.5 µg of Bovine Serum Albumin (BSA), 2 mM of $MgCl_2$, 0.25 mM of each dNTP, 0.5 mM of each primer, and 0.3 U of *Taq* DNA polymerase (Invitrogen). Alternatively, a Trehalose-BSA-Tween-20 (TBT) solution was added to a final concentration of 1 × instead of BSA, as described by Samarakoon et al. [17]. PCR conditions included an initial denaturation at 94 °C for 5 min, followed by 30 cycles of 45 s at 94 °C, 1 min at 53–65 °C, 30 s at 72 °C, and a final extension at 72 °C for 15 min. PCR products were initially analyzed on 3% agarose gels in 1 × TAE buffer and stained either with ethidium bromide or GelRED™ (Biotium) and visualized under ultraviolet light. Thirteen of the 50 primer pairs yielded distinct, single bands of the expected size range on agarose; 23 did not show amplification, and fourteen showed either unspecific amplifications or were outside the expected size range. The 13 primer pairs were used to genotype a set of 20 individuals: ten from a population located at Chapada dos Guimarães in Mato Grosso State (15.46°S × 55.75°W), referred to here after as "MT", and ten from a population located at Uruçuí in Piauí State (7.56°S × 44.44°W), referred to hereafter as "PI". Those two areas are approximately 1500 km from one another, and considered central and marginal Cerrado populations respectively. Sampling respected a minimum distance of 10 meters between individuals within each population. PCR fragments were separated on 6% denaturing polyacrylamide gels in 1 × TBE buffer and stained with silver nitrate [18]. Fragment sizes were scored using a 10 bp ladder (Invitrogen).

SSR loci characterization

The numbers of alleles (Na), observed and expected heterozygosity (H_O and H_E), Hardy–Weinberg equilibrium, and linkage disequilibrium were determined using

ARLEQUIN v 3.5.1.3 software [19]. Wright's inbreeding coefficient (F_{IS}) was calculated using GENEPOP [20]. The polymorphism information content (PIC) was estimated using CERVUS v. 3.0.7 software [21], and the presence of null alleles was tested with Micro-checker v. 2.2.3 software, based on the "allelic signature" given by deficiencies or excesses of particular genotypes [22]. Identity (P_{ID} and P_{IDsib}) and parentage exclusion (P_E1, P_E2, and P_E3) probabilities, as well as the numbers of private alleles (Np) and number of migrants (Nm) were calculated using GenAlEx software [23].

Results and discussion

A total of 52,865 sequence reads were generated and aligned, resulting in 4697 contigs at least 300 bp long. Of those, 278 contained SSRs: 28 compound and 250 simple SSR. Among the contigs with simple SSRs, 57.1% were mononucleotide, 30.9% were dinucleotides, 4% trinucleotides, 2.4% tetranucleotides, 4% pentanucleotides, and 1.6% hexanucleotides. Di-, tri-, and tetranucleotides were also selected from singleton sequences (Table 1 and Online Resource 1). Twelve contigs and 38 singleton regions were selected for primer design, totaling 50 pairs of SSR primers (Table 1 and Online Resource 1).

Of the 13 loci that showed single bands within the expected size range on agarose, ten presented clearly scorable alleles and were polymorphic in polyacrylamide gels. Among the polymorphic loci, from two to nine alleles per locus were observed within the MT population, and from two to 11 alleles within the PI population (Table 2). Although possibly biased by the low numbers of individuals per population, we observed high numbers of private alleles unique to each population: out of the 74 alleles found in both populations, 19 were private to the MT population (26%), while 25 were private to the PI population (34%), so that 60% of all alleles were private to one or the other population (Table 2). The uniqueness of alleles within the two groups presumably reflects their phylogeographic isolation.

The loci described here have PIC values ranging between 0.26 and 0.91 per locus, with a relatively high mean value (0.64) (Table 2). The probability that two unrelated individuals would share the same genotype (PI), and the probability that two full-sibs would have identical genotypes (PI_{sib}), are presented in Online Resource 2. PI and PI_{sib} values across all 10 loci ranged from $2.2\times E^{-2}$ (locus AeSSR*C9* in the *AEPI* population) to $8.2\times E^{-1}$ (locus AeSSR*s26* in the *AEMT* population), and from $9.1\times E^{-1}$ (locus AeSSR*s26* in the *AEMT*

Table 1 Characteristics of ten polymorphic microsatellite loci and primer pairs developed for *Alibertia edulis*

Locus	Primer sequences (5′–3′)	Repeat motif	Size (bp)	T_a (°C)	GenBank accession number
AE_SSR_C2	F:GATCACGGCCACATAGAATACC R:CTCCGGTAATTGCTCTTTCAAT	$(ACCAA)_4$	240	59	KR075121
AE_SSR_C9	F:TGTTGAAGGACCACAAGAACAT R:AATTAACTCCACTTCCCATTGC	$(GA)_{12}$	225	65	KR075128
AE_SSR_S1	F:CACGAAGTGGATTGAGGCAC R:CGGTAGTTTGTTCTTCCGGC	$(TA)_{11}$	249	59	KR075082
AE_SSR_S3	F:TCACTACTAGGTAGGCAAATGAAG R:TCCCAGAAACTGATTCAAACCC	$(CA)_{12}$	338	58	KR075084
AE_SSR_S6	F:GCGTGAGTGGTGCAAATTC R:CACCCACTCCACAAGGAAG	$(CA)_{11}$	224	58	KR075087
AE_SSR_S8	F:CTGTCTCTTGGCGAGTCTTC R:ATGGCACGTCTCCATCCAC	$(GA)_{12}$	179	58	KR075089
AE_SSR_S9	F:CTCTGCCTACCCTCCTTGG R:TGCAACAAGTTGGCTTCGG	$(AGG)_{11}$	384	59	KR075090
AE_SSR_S26	F:ATCTTTCTTGCTTTACCACCA R:GAGGGAGAGTGCGGTTAC	(CCT)5	215	56	KR075107
AE_SSR_S30	F:AAAAGAATCCAACCAAGAAAG R:CAAGTAATAGCAGCATCCAAT	(ACG)5	195	53	KR075111
AE_SSR_S33	F:TTACTCTTCTATTGGCATTGG R:TGGAGATCAAAAAGACTGCTA	(TTC)6	168	55	KR075114

T_a annealing temperature as calculated by Primer3 software

Table 2 Population genetic parameters determined in two *Alibertia edulis* populations (Mato Grosso—MT and Piaui—PI) for ten polymorphic microsatellites loci

Locus	AEMT						AEPI						mNa	tNa	Nm	PIC	Null alleles
	N	Na	H_O	H_E	F_{IS}	*PA*	N	Na	H_O	H_E	F_{IS}	*PA*					
AeSSRC2	9	2	**1**	0.529	−1	0	10	4	**1**	0.647	−0.593	2	3	4	15.93	0.48	No
AeSSRC9	10	8	0.7	0.863	0.198	2	10	11	0.8	0.937	0.153	5	9.5	13	7.77	0.87	No
AeSSRS1	10	2	0.182	0.416	0.575	1	10	2	**0.091**	0.368	0.762	1	2	3	2.92	0.37	No
AeSSRS3	10	6	0.6	0.837	0.294	0	10	6	**0.4**	0.837	0.536	0	6	6	21.02	0.75	No
AeSSRS6	7	6	**0.429**	0.791	0.478	3	10	5	**0.3**	0.726	0.6	2	5.5	8	1.62	0.79	Yes
AeSSRS8	10	9	0.9	0.9	0	8	9	8	0.889	0.863	−0.032	7	8.5	16	2.62	0.91	No
AeSSRS9	10	6	0.8	0.784	−0.021	2	10	7	**0.8**	0.842	0.053	3	6	9	2.80	0.84	No
AeSSRS26	10	2	0.1	0.1	0	1	10	4	0.5	0.437	−0.154	3	3	5	3.92	0.26	No
AeSSRS30	10	5	1	0.711	−0.44	0	10	6	1	0.721	−0.417	1	5.5	6	45.33	0.65	No
AeSSRS33	8	3	0.25	0.425	0.429	2	10	2	0.3	0.521	0.437	1	2.5	4	1.59	0.46	No

Loci out of Hardy–Weinberg equilibrium are written in bold

N number of analyzed individuals, *Na* number of observed alleles, *Ho* observed heterozygosity, H_E expected heterozygosity, F_{IS} Wright's inbreeding coefficient, *PA* private alleles, *mNa* mean number of observed alleles in both populations, *tNa* number of total observed alleles, *Nm* number of migrants, *PIC* polymorphism information content

Table 3 PI, PIsibs and probabilities of exclusion (P1, P2 and P3) for increasing locus combinations for ten polymorphic microsatellites loci

Population	N	PI	Pisibs	P1	P2	P3
AEMT	18	1.4E−08	7.7E−04	1.0E+00	9.7E−01	1.0E+00
AEPI	17	1.5E−09	3.8E−04	1.0E+00	9.8E−01	1.0E+00

N number of analyzed individuals, *PI* probability of identity, *PIsibs* probability of identity when related individuals are included in the sample, *P1* probability of exclusion for one putative parent when the other parent's genotype is known. *P2* probability of exclusion for one putative parent when the genotype of the other parent is missing, *P3* probability of excluding a putative parent pair

population) to $3.1 \times E^{-1}$ (locus AeSSR*C9* in the *AEPI* population) respectively. When considering all ten described loci together, however, the PI and PIsibs values were very low, as can be seen in Table 3. The parentage exclusion probabilities (P1, P2 and P3) for each locus, and for increasing locus combinations, are shown in Online Resource 2 and Table 3 respectively. Parentage exclusion values were very high when considering all ten loci together (ranging from 1.0 to 0.97), indicating those loci as useful tools for population genetic analysis in *A. edulis* when analyzed together.

Observed heterozygosity ranged from 0.09 (locus AeSSRS1 in the PI population) to 1.00 (loci AeSSRC2 and AeSSRS30 in both populations), while the expected heterozygosity ranged from 0.10 (locus AeSSRS26 in the MT population) to 0.94 (locus AeSSRC9 in the PI population). Heterozygote excesses or mainly deficiencies were observed for a few loci, resulting in significant deviation from the Hardy–Weinberg equilibrium (p values < 0.05) for the PI population in two loci and for both populations in three loci, as well as the presence of null alleles in one locus. The inbreeding coefficient values were, in these cases, higher in the PI than in the MT population (Table 2), suggesting that reproduction within those populations was not random. Linkage disequilibrium was also estimated for both populations (Table 3).

Previous studies developed and used SSR markers for several other Cerrado plant species, mostly analyzing central Cerrado populations, as part of increasing efforts to understand and conserve that threatened biome [e.g., 9, 10]. The Cerrado biome has experienced expansion-retraction dynamics related to climatic oscillations that contributed to the isolation and diversification of its populations [5], especially those located in northern and northeastern Brazil (such as disjunct populations in the states of Pará and Piauí) [24]. Because of their geographic isolation, those areas would be expected to show higher inbreeding coefficient values than those located in core Cerrado sites. *A. edulis* is also found in gallery forests, which connected and expanded during the Plio-Pleistocene period [25]. Therefore, the higher F_{IS} values observed in the Piauí population might be related to a long period of isolation from other Cerrado populations, most likely since the Pleistocene. Alternatively, those values could be related to selection due to the non-neutrality of certain loci, or to population substructuring [13]. Subsequent

studies using the markers developed here in larger numbers of populations could help clarify the genetic composition of central and more isolated *A. edulis* populations.

The polymorphic markers developed in the present study represent informative tools for phylogeographic and population analysis in *Alibertia edulis,* and potentially for other related Rubiaceae species, and will contribute to a better understanding of the biogeographic history of the Cerrado and, consequently, its conservation.

Acknowledgements We thank Ana Paula Moraes, Felipe Amorim and the Conservation Area managers for their support in the field, and Ana Benko-Iseppon and João Pacífico Bezerra Neto for their assistance with the NGS data analysis. We also thank the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq—fellowship 140699/2010-4), the Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE APQ-0862-2.05/10), and the Fundação Grupo Boticário de Proteção à Natureza (Grant No. 0920_20112) for financial support. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001. The authors declare that they have no conflicts of interest.

References

1. Silva FAB, Pereira LAR, Silveira CES (2008) Micropropagation of Alibertia edulis Rich. Braz Arch Biol Technol 51:1103–1114
2. Silva Júnior MC, Pereira BAS (2009) +100 árvores do cerrado matas de galeria. Rede de Sementes do Cerrado, Brasília
3. Ratter JA, Bridgewater S, Ribeiro JF (2003) Analysis of the floristic composition of the Brazilian cerrado vegetation III: comparison of the woody vegetation of 376 areas. Edinburgh J Botany 60:57–109
4. Furley PA, Ratter JA (1988) Soil resources and plant communities of the central Brazilian cerrado and their development. J Biogeogr 15:97–108
5. Ratter JA, Ribeiro JF, Bridgewater S (1997) The Brazilian cerrado vegetation and threats to its biodiversity. Ann Bot 80:223–230
6. Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403:853–858
7. Klink CA, Machado RB (2005) Conservation of the Brazilian cerrado. Conserv Biol 19:707–713
8. Mittermeier RA, Gil RP, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, Lamourex J, Fonseca GAB (2005) Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions. University of Chicago Press, Boston
9. Ritter LMO, Bajay MM, Monteiro M, Souza RGVC, Moreno MA, Kageyama PY (2012) Development of microsatellite markers for Qualea grandiflora (Vochysiaceae), a typical species of the Brazilian cerrado. Am J Bot 99:97–98
10. Camacho LMD, Schatzer CAF, Alves-Pereira A, Zucchi MI, Carvalho MAM, Gaspar M (2017) Development, characterization and cross-amplification of microsatellite markers for *Chrysolaena obovata*, an important Asteraceae from Brazilian Cerrado. J Genet. https://doi.org/10.1007/s12041-017-0812-9
11. Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1–16
12. Squirrell J, Hollingsworth PM, Woodhead M, Russel J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:1339–1348
13. Wöhrmann T, Pinangé DSB, Krapp F, Benko-Iseppon A-M, Huettel B, Weising K (2013) Development of 15 nuclear microsatellite markers in the genus Dyckia (Pitcairnioideae; Bromeliaceae) using 454 pyrosequencing. Conserv Genet Resour 5:81–84
14. Weising K, Nybom H, Wolff K, Kahl G (2005) DNA fingerprinting in plants: principle, methods and applications. Taylor & Francis Group, Florida
15. Faircloth B (2008) Msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol Ecol 8:92–94
16. Ferreira ME, Grattapaglia D (1996) Introdução ao Uso de Marcadores Moleculares em Análise Genética. EMBRAPA/CENARGEN, Brasília
17. Samarakoon T, Wang SY, Alford MH (2013) Enhancing PCR amplification of DNA from recalcitrant plant specimens using a trehalose-based additive. APPS 1:1200236
18. Creste S, Tulmann NA, Figueira A (2001) Detection of simple sequence repeat polymorphisms in denaturing polyacrilamide sequencing gels by silver staining. Plant Mol Biol Rep 19:299–306
19. Excoffier L, Laval G, Schneider S (2005) Arlequin v. 3.0: An integrated software package for population genetics data analysis. Evol Bioinform 1:47–50
20. Raymond M, Rousset F (1995) GENEPOP (v. 1.2): population genetics software for exact tests and ecumenicism. J Hered 86:248–249
21. Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106. https://doi.org/10.1111/j.1365-294X.2007.03089.x
22. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Res. https://doi.org/10.1111/j.1471-8286.2004.00684.x
23. Peakall R, Smouse PE (2006) GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295
24. Machado RB, Ramos Neto MB, Pereira PGP, Caldas EF, Gonçalves DA, Santos NS, Tabor K, Steininger M (2004) Estimativas da perda da área do cerrado brasileiro. Relatório técnico não publicado. Conservação Internacional, Brasília DF
25. Batalha-Filho H, Fjeldsa J, Fabre P-H, Miyaki CY (2013) Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. J Ornithol 154:41–50