SHORT COMMUNICATION

First genomic microsatellite markers developed for *Platonia insignis* (Clusiaceae), a Brazilian fruit tree

Luany Caroline Ribeiro Paraense[1] · Dayane Nascimento Pena[2] · Sylvain Henri Darnet[3] ·
Simone de Miranda Rodrigues[4] · Ilmarina Campos de Menezes[4] · Elisa Ferreira Moura[4]

Abstract

Platonia insignis is a fruit tree native of Brazil with allogamous and asexual reproduction. The production of fruits is mainly obtained by exploitation of natural populations and the impact of genetic structuring on plant production may be evaluated. For this purpose, codominant and multiallelic markers such as microsatellite are the most suitable, but they need to be developed for this species. Thus, the aim of this work was to develop and validate microsatellite markers for *P. insignis*. We used Roche 454 GS FLX sequencing platform of a single *P. insignis* genotype and 1702 microsatellite sequences were identified. Based on some pre-requisites, we could develop 50 primer pairs to be tested. Twenty-two primer pairs successfully amplified fragments and they were tested in 31 genotypes of *P. insignis* that belong to a germplasm bank and were sampled in the northeast of Pará State, Brazil. Thirteen primers were polymorphic and the number of alleles per loci varied from 5 (PI18 and PI27) to 2 (PI08, PI25, PI31, PI33 and PI 37). Expected heterozygosity (H_E) varied from 0.74 (PI27) to 0.12 (PI31) and observed heterozygosity (H_O) varied from 1.00 (PI25) to 0.00 (PI08, PI31, PI33 and PI37). Principal coordinates could separate the genotypes of *P. insignis* in clusters and we can conclude that the primers can estimate the genetic diversity of *P. insignis* populations.

Keywords Simple sequence repeats · Genetic diversity · Oligonucleotides · Next generation sequencing

Introduction

'Bacurizeiro' (*Platonia insignis* Mart.) is a fruit tree native of Brazil. Its fruits have a special flavor and are rich in lipid content and total dietary fiber [9]. Besides the value of its fruit, the tree can also be exploited for wood. It belongs to the family Clusiaceae, subfamily Clusioideae and it is the only member of the genus *Platonia*. *P. insignis* has allogamous reproduction associated with sporophytic self-incompatibility, and parrots were identified as seed and pollen dispersers [5]. Besides, its roots can expand and grow abundantly in sandy soils and, in recently degraded areas, more than 40,000 ramets generated from the roots can be found per hectare [1].

'Bacuri' pulp has been gaining more attention from gastronomy chefs due to its flavor and the pulp is being purchased at higher prices on local markets [3]. The fruits are commonly collected from natural population areas, since there are few planted areas of *P. insignis* to be commercially exploited and the few agronomical technologies available for the species management are still not very adopted. The self-incompatibility reproduction associated with asexual propagation may contribute to low yield of fruits in natural regenerated areas and molecular studies can help the management of natural populations as well as the genetic breeding of the species. Microsatellite or simple sequence repeats (SSR) can be an ideal tool, due to their wide distribution in the genome, codominancy and high level of polymorphism [4], helping in the identification of clones and in genetic population studies. There are no reports of microsatellite

✉ Elisa Ferreira Moura
elisa.moura@embrapa.br

1 Programa de Pós-graduação em Biotecnologia aplicada à Agropecuária, Universidade Federal Rural da Amazônia, Belém, Pará 66077-530, Brazil

2 Programa de Pós-graduação em Agronomia, Universidade Federal Rural da Amazônia, Belém, Pará 66077-530, Brazil

3 Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém, Pará 66077-830, Brazil

4 Laboratório de Genética Molecular, Embrapa Amazônia Oriental, Belém, Pará 66095-903, Brazil

primers for *P. insignis*. Thus, the aim of this study was to develop microsatellite markers of *P. insignis*.

Material and methods

Total genomic DNA was extracted from 31 accessions of *P. insignis* from a germplasm bank using a cetyltrimethylammonium bromide (CTAB) protocol described by Pontes et al. [8]. The accession "Fernando" was used to be sequenced by the Roche 454 GS FLX platform. Genomic libraries and sequencing were performed according to the manufacturer's protocol. Before search for microsatellite regions, the filtering process of raw sequencing reads was performed. The MIcroSAtellite identification tool—MISA program (https://pgrc.ipk-gatersleben.de/misa/) was used to detect microsatellite repetitions, considering perfect microsatellites with one to six nucleotides. Primers were designed using Primer3 program [10] considering di to pentanucleotides with a minimum of five repetitions per motif; temperature inferior to 2 °C between the forward and reverse primers; CG percentage of 40 to 60% and amplification product between 100 and 280 bp. After primers were designed, it was selected the ones that amplify dinucleotide repetitions, and sequences A/T were eliminated due to lower amplification stability. This selection was possible due to the high number of primer sequences generated.

Fifty primers were synthetized and were tested on two samples of *P. insignis*. Polymerase chain reactions were prepared for a final volume of 20 μL containing 10 ng of genomic DNA, 100 μM of each dNTP, 10 μM of each primer, 1.0 mg mL^{-1} BSA (bovine serum albumin), in 1 × PCR buffer (50 mM KCl; 10 mM Tris–HCl, pH 8.8, 0.1% Triton X-100, 1.0 mM MgCl$_2$) and 1.0 U Taq DNA polymerase (Invitrogen, São Paulo, Brazil). Reactions were carried out in an Amplitherm TX96 thermocycler (Applied Biosystems, Foster City, CA, USA), with the following program: an initial cycle of 95 °C for 4 min; followed by 30 cycles of 94 °C for 30 s; 60 s at appropriate annealing temperature for each primer (Table 1), and extension at 72 °C, for 60 s. After the 30 cycles, there was a final extension at 72 °C, for 10 min. Amplification of PCR products were checked on 3% agarose gels. Twenty-two (Table 1) out of the fifty pair of primers tested amplified fragments successfully on the two samples of *P. insignis* and were selected for polymorphism tests on 31 accessions from a germplasm bank of *P. insignis*, sampled in municipalities of Pará State, Brazil. Each accession was represented by one clone.

PCR were run under the same conditions described above and the amplification products were visualized on vertical electrophoresis (Omniphor, HMEDI15, England), using 7% polyacrilamide gel, using 10-bp and 100-bp DNA ladder (Invitrogen, Carlsbad, California, USA) as a reference. Gels were revealed with silver nitrate and scanned for image analyses. Suitable primers were those that exhibited fragments of the expected size, clear banding patterns, and no more than two alleles per individual for each locus.

For each locus, the observed number of alleles (*A*), effective number of alleles (N$_e$), observed heterozygosity (H$_O$), expected heterozygosity (H$_E$) and fixation index (*f*), as well as test for deviation from Hardy–Weinberg equilibrium (HWE) were calculated using the GenAlEx 6.501 program [7]. The probability of genetic identity and the probability of exclusion were obtained according to the method suggested by Paetkau and Strobeck [6] and calculated with GenAlEx. The PAST program [2] was used for computing principal coordinate analysis (PCoA).

Results and discussion

The sequencing with 454-Roche FLX platform generated a total of 85,909 reads, 32,487,452 base pairs and reads were 378.16 bp length in average. Seventy-five percent of the reads were 400 to 500 bp length. Sequence library data of this study were deposited to the Sequence Read Archive of the National Center for Biotechnology Information (NCBI; BioProject ID: SUB4665203). After filtering process, microsatellite sequences were searched in 36,739 reads with 14,225,957 base pairs and 1702 microsatellite regions were identified, representing 3.95% of total base pairs sequenced. When considering di- to hexanucleotide tandem repeats, di- were the most frequent repeats (66.5%), followed by tri- (26%), tetra- (6.1%), penta- (0.83%) and hexa- (0.35%). Among di- repeats, AG/CT was the most frequent (42.74% of di-repeats), followed by AT/AT (40.52%).

Fifty primers were constructed, and tests with two *P. insignis* DNA samples detected amplification in 22 pair of primers (Table 1). The 22 microsatellite loci were PCR amplified with 31 accessions that belong to a germplasm bank and were sampled in different sites of Northeastern Brazilian Amazon. Thirteen loci amplified polymorphic bands (Table 2). The total number of alleles per loci varied from five (PI18 and PI27) to two (PI08, PI25, PI31, PI33 and PI37) and the effective number of alleles ranged from 3.84 (PI27) to 1.14 (PI31). The expected heterozygosity (H$_E$) ranged from 0.74 (PI27) to 0.12 (PI31), whereas for observed heterozygosity, PI25 detected only heterozygotes (Ho = 1.00) and loci PI08, PI31, PI33 and PI37 detected only homozygotes (H$_O$ = 0.00). This may be due to the presence of null alleles or due to the characteristic of asexual propagation of *P. insignis* [1]. Fixation index was 0.29 on average, and values ranged from 1.00 (PI08, PI31, PI33 and PI37) to 0.00 (PI25). Significant deviation from Hardy–Weinberg equilibrium (P < 0.001) was detected for 11 loci (Table 2), except for PI01 and PI07. This can be explained by the fact

Table 1 Characteristics of 22 microsatellite loci developed for the fruit tree *Platonia insignis* Mart.

Primers	Repeat motifs	Primer sequences (5′–3′)	Size (bp)	(Ta) °C	GenBank accession no.
PI01	$(AG)_{13}$	F: TGCACAATTCTTGAGGCGT R: ACATGGACGGTCATGCGT	240–280	54	KY911456
PI03	$(AG)_6$	F: CGGCCAGTAGAAGACGACC R: CACATTCGCCAGGCATGC	252	60	KY911458
PI04	$(AG)_8$	F: ACACAGTGACCAACCAAACA R: GGCTTCACTGGCATCCACT	270–300	53	KY911459
PI05	$(CA)_6$	F: CCCGGTCCAGAGTGCTTC R: TGTGGTAACTGCATGGGGC	245	57	KY911460
PI06	$(CA)_9$	F: AGGTCATATGCCTTGTCCAGA R: ACTTGAATGGTGGAAGTGACA	162–173	53	KY911461
PI07	$(CT)_{13}$	F: CAAACCAGCCATCAAAGAGGG R: TCCCGAGGAAAACGTAAACA	168	60	KY923231
PI08	$(CT)_9$	F: ACAATGAGGGGACAGGGGA R: GCCAACTAACCAATCGGCG	235	60	KY923232
PI13	$(TG)_{11}$	F: GGTAGGAGTGGCTCAAACCA R: GCATGCCATACATCAGAGTCT	208	59	KY923237
PI14	$(TG)_6$	F: TGTTGGTCTTGGTGGGAAACA R: AGCACTCTGTTGGGCTGAC	255	53	KY923238
PI15	$(CT)_6$	F: CCTCTCCCTCAGGGTCCA R: GGTTGCCATGGATGAGGCT	245	57	KY923239
PI17	$(AG)_7$	F: ACGGGTGTTACAACAGTGGT R: AGCACTGCACTCTCTTCTGAG	235	53	KY923241
PI18	$(GA)_8$	F:CTGAACCAACTGCTAAAGGAGG R: AAGGGTAATGTCTCTTCCACCC	120	59	KY923242
PI25	$(AG)_6$	F: GGTTCGTGACCCTGAACCA R:TCGTGCTCCTAAACCACCA	175–220	55	MF002446
PI27	$(CA)_6$	F: ACACACACAACAAGGGAGCA R: GTGTCCTGCAAACACGAACA	165–170	57	MF002448
PI31	$(CA)_7$	F:GGATCGAATCCCACCATTGC R:GCT TTGGTCATGGCAGCC	165	59	MF002452
PI32	$(GA)_{23}$	F: GTGGTTGATAGGGCCACCC R: TCCACATGCACAACCCAACT	170	60	MF002453
PI33	$(CA)_7$	F: CGCTTTCCCCTTGCTCGT R: TGGGGATGTGTGGGAGAGA	255	60	MF002454
PI37	$(TC)_7$	F: TCCAAGTACATGCAAGGTGT R: TGGGAGCTTAGAAGGGTGGA	260	57	MF002458
PI40	$(TC)_7$	F: CCATGCCATTTAGCCAAGCA R: AGAGATGAGCGAAGGGGAGA	175	60	MF002461
PI47	$(GA)_9$	F: ACATGGCCCCTCGTACCT R: ACGATTTGCCACCTGCCA	280	60	MF038810
PI48	$(TG)_7$	F:CTGAACCAACTGCTAAAGGAGG R: AAGGGTAATGTCTCTTCCACCC	270	57	MF038809
PI49	$(TC)_7$	F: CCTCACTGGTTAGGTCACCG R: GGGAGAGAGCAATCTTTGAGC	240–250	59	MF038808

that the set of primers were tested in a group of genotypes sampled in different places, which could be not exchanging alleles, due to geographic distance, thus justifying the deviation from Hardy–Weinberg equilibrium. The probability that two individuals share the same genetic profile by chance with this set of thirteen primers and the allelic frequencies obtained was 2.1×10^{-7}.

The first two main coordinates on the analysis of principal coordinates explained 50.87% of the total genetic variation and could separate the accessions in five groups (Fig. 1).

Thus, it can be noticed that the set of primers developed for *P. insignis* could detect genetic variation and could form clusters of different genotypes. It could be expected that the application of these primers in a group of genotypes sampled

Table 2 Genetic variation of the 13 polymorphic microsatellite loci in 31 accessions of *Platonia insignis* Mart. sampled in Brazilian Amazonia

Locus	A	A_E	H_O	H_E	f	*HWE*
PI01	4	2.99	0.58	0.67	0.13	0.715[ns]
PI04	3	1.54	0.13	0.35	0.63	0.000***
PI06	3	2.47	0.71	0.60	− 0.19	0.015*
PI07	4	1.44	0.23	0.30	0.26	0.434[ns]
PI08	2	1.84	0.00	0.46	1.00	0.000***
PI17	4	1.77	0.32	0.44	0.26	0.000***
PI18	5	2.34	0.87	0.57	− 0.52	0.048*
PI25	2	2.00	1.00	0.50	− 1.00	0.000***
PI27	5	3.84	0.48	0.74	0.35	0.003**
PI31	2	1.14	0.00	0.12	1.00	0.000***
PI33	2	1.95	0.00	0.49	1.00	0.000***
PI37	2	2.00	0.00	0.50	1.00	0.000***
PI49	4	0.72	0.77	0.72	− 0.07	0.007**
Average	3.23 ± 0.32	2.22 ± 0.22	0.39 ± 0.10	0.50 ± 0.05	0.30 ± 0.18	

A number of alleles, A_E effective number of alleles, H_E expected heterozygosity, *Ho* observed heterozygosity, f index fixation, *HWE* test for deviation from Hardy Weinberg equilibrium

[a]Deviation from Hardy–Weinberg equilibrium at *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ns = not significant
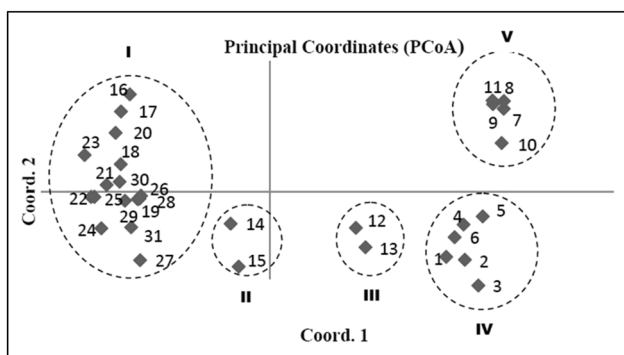
Fig. 1 Principal coordinate analysis based on the genotyping of 31 genotypes of *Platonia insignis* Mart. sampled on Pará State, Brazil with 13 microsatellite primers developed for the species

in geographically more distant places could increase the genetic variation detected.

Conclusions

From genomic next generation sequencing data, we searched and characterized microsatellite regions in an individual of *P. insignis* and we validated thirteen sets of primers using genotypes sampled in different places of northeastern Amazonia. This resulted in thirteen polymorphic microsatellite markers for *P. insignis*, which can be used for plant population studies and assist genetic breeding efforts.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest directly or indirectly and informed consent to publish this study and that the manuscript complies with the ethical standards of the journal.

References

1. Ferreira MSG, Medina G (2004) Bacuri: gold in the market, delicious in the mouth. In: Alexiades M, Shanley P (eds) Productos forestales, medios de subsistencia y conservácion: estudios de caso sobre sistemas de manejo de productos forestales no moderables. CIFOR, Bogor, pp 5–8
2. Hammer O, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron 4:1–9
3. Homma AKO, de Menezes AJE, de Carvalho JEU, de Matos GB (2018) Manejo e plantio de bacurizeiros (*Platonia insignis* Mart.): a experiência no manejo e domesticação de um recurso da biodiversidade amazônica. Inc Soc 12:48–57
4. Kalia RK, Raí MK, Kalia S, Rohtas S, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. Euphytica 177:309–334
5. Maués MM, Venturieri GC (1996) Ecologia da polinização do bacurizeiro (*Platonia insignis* Mart.) Clusiaceae. Boletim de pesquisa 170, Embrapa Amazônia Oriental, Belém, Pará. 24p.
6. Paetkau D, Strobeck C (1994) Microsatellite analysis of genetic variation in black bear populations. Mol Ecol 3:489–495
7. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28:2537–2539

8. Pontes LCG, Moura EF, Moura MF, Rodrigues SM, Oliveira MSP, Carvalho JEU, Therrier J (2017) Molecular characterization of progenies of bacurizeiro (*Platonia insignis*) from Marajó Island, northeastern Amazon. Acta Amaz 47:293–300. https://doi.org/10.1590/1809-4392201701302

9. Rogez H, Buxant R, Mignolet E, Souza JNS, Silva EM, Larondelle Y (2004) Chemical composition of the pulp of three typical Amazonian fruits: araça-boi (*Eugenia stipitata*), bacuri (*Platonia insignis*) and cupuaçu (*Theobroma grandiflorum*). Eur Food Res Technol 218:380–384

10. Rosen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol 132:365–386