Isolation and characterization of microsatellite markers for *Bertholletia excelsa* (Lecythidaceae) population genetic analysis

P.S. Sujii[1], P.W. Inglis[2], A.Y. Ciampi[2], V.N. Solferini[1] and V.C.R. Azevedo[2]

[1]Departamento de Genética, Evolução e Bioagentes, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brasil
[2]Laboratório de Genética Vegetal, Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brasil

Corresponding author: V.C.R. Azevedo
E-mail: vania.azevedo@embrapa.br

ABSTRACT. Seven polymorphic microsatellite markers were developed and validated for *Bertholletia excelsa* (Brazil nut tree) population genetic studies. This species is a widespread monotypic Amazonian tree with high non-timber economic value. Unfortunately, Brazil nut production is currently less than 25% of historical production levels, because of extensive deforestation. All pairs of primers produced clearly interpretable and polymorphic bands. No linkage disequilibrium was observed in an analysis of 46 individuals from one population, three to seven alleles per locus were observed; the expected heterozygosity ranged from 0.378 to 0.978, with significant heterozygote excess for four loci. An analysis of individuals from two populations showed private alleles at all loci. These primer pairs will be useful for population studies, especially for comparing samples from different parts of the Amazon forest.

Key words: Brazil nut; SSR; Plant genetics; Amazon forest

INTRODUCTION

The Brazil nut tree (*Bertholletia excelsa* Humb.; Bonpl.) is a widespread monotypic species in upland forest areas in the Amazon. Its reproductive characteristics and longevity make this species a good model to understand the genetic diversity of Amazon Forest species. Brazil nuts are important non-timber products, and in some over-exploited areas, a lack of juvenile trees and seedlings has been observed (Peres et al., 2003).

Genetic diversity across a species' distribution throughout the Amazon Forest is an understudied issue, especially in plants. Quantifying and understanding its population genetic structure, as well as studies of gene flow and mating systems, provide important information for the development of conservation strategies and species management. Such information can also be helpful in identifying effects of habitat fragmentation (Erickson et al., 2004; Azevedo, 2007; Vandergast et al., 2007).

The use of genetics as a tool for conservation requires information about allelic distribution, heterozygosities, fixation indices and population structure. The need for information on the populations' recent history and the scarcity of information for non-model species, as is the case for most tropical trees, make microsatellite markers valuable tools (Schlötterer, 2004). Microsatellites are co-dominant markers, with a high degree of polymorphism, found at high frequency, and are well distributed throughout higher plant genomes (Tóth et al., 2000). The building of an enriched genomic library is a method capable of rapidly developing plenty of these markers (Ostrander et al., 1992).

In the present study, we aimed to develop microsatellite primers for *B. excelsa* and validate them for use in population genetic studies.

MATERIAL AND METHODS

Samples of *B. excelsa* were collected from two different areas in Brazil. One of them was the Chico Mendes Extractive Reserve (Resex Chico Mendes), in Acre State, where six individuals were collected for genomic library construction and polymorphism tests. In another site, the Iratapuru River Sustainable Development Reserve (RDS do Rio Iratapuru), in Amapá State, a sample of 46 individuals was collected in a continuous forest area for the characterization and validation of the markers. DNA was obtained from vascular cambium tissue using a 2% CTAB protocol (Doyle and Doyle, 1987), which was adapted.

A dinucleotide-enriched genomic library was constructed using the protocol developed by Rafalski et al. (1996) and adapted by Buso et al. (2003). A DNA digestion test was performed with the restriction enzymes Sau3A (↓GATC), *Mse*I (↓TAA) and Tsp 509 (↓AATT), where the largest number of fragments from 200 to 800 bp was obtained with *Mse*I. Digested fragments were purified and recovered from a 2% agarose gel using the QIAquick Gel Extraction kit (Qiagen; Hilden, Germany) and were linked to specific adaptors for the *Mse*I restriction site. Fragments with adapters were linked to a complex containing poly-AG/TC oligonucleotide fragments, biotin and magnetic beads (Dynabeads - Streptavidin; Boehringer; Mannheim, Germany) and were recovered and amplified by polymerase chain reaction (PCR). Amplified fragment purification was performed using the QIAquick kit (Qiagen). Fragments containing SSR sequences were ligated into the pGEM-T Easy Vector (Promega; Fitchburg, Wisconsin, USA) and transformed into *Escherichia coli* strain XL1-Blue by electroporation.

Positive clones were selected based on lack of β-galactosidase gene activity. Sequencing reaction templates were generated by PCR amplification DNA fragments with M13 forward and reverse primers, complementary to vector sequences. Templates were generated from 192 positive clones and were purified with exonuclease I and shrimp alkaline phosphatase (SAP). Sequencing reactions were performed with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems). PCR products were purified following purification by precipitation with ethanol/EDTA, using the manufacturer protocol. Sequencing was performed using the 96-capillary ABI 3700 automatic DNA analyzer.

Sequences with di or tri-nucleotide repeats were selected for primer design with Primer3 web-based software (Rozen and Staletsky, 2000). Each forward primer in all primer pairs designed was labeled with green (HEX) or blue (6-FAM) fluorochrome. Ideal PCR conditions were determined using DNA from four *B. excelsa* trees sampled from natural populations. PCR reaction mixtures contained 3-5 ng template DNA, 0.2 μmol each primer, 1 U Taq DNA polymerase, 250 μmol each dNTP, 0.40 μg BSA and 1X reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$) in a total volume of 13 μL. Cycling conditions were: 94°C for 5 min (one cycle); then 94°C for 1 min, 50° to 58°C (according to the primer annealing temperature) for 1 min and 72°C for 1 min (30 cycles); and 72°C for 7 min (one cycle). Amplified fragments were evaluated on a 3.5% agarose gel stained with ethidium bromide.

Polymorphism analysis was performed using DNA from six individuals from Acre (Resex Chico Mendes) and six from Amapá (RDS do Rio Iratapuru), to detect loci with intra- and inter-population polymorphism. DNA amplified fragments were separated and analyzed using a 96-capillary ABI3700 Automatic Analyzer and GeneScan and Genotyper software (Applied Biosystems). SSR loci were selected according to mean number of alleles per locus and presence of private alleles, by direct observation of genotypes. Primer pairs that amplified regions with at least three alleles or one private allele, with clearly identifiable bands, were selected for population analysis.

A population analysis was performed with the sample from Amapá. Genetic diversity was characterized by estimates of allele frequencies, mean number of alleles per locus, and observed and expected heterozygosities under Hardy-Weinberg equilibrium (HWE). The inbreeding coefficient *f* was calculated with the F_{IS} estimator (Weir and Cockerham, 1984), using FSTAT (Goudet, 1995). Confidence intervals (95%CI) of *f* estimates were obtained using 10,000 bootstrap replicates, and linkage disequilibrium was tested using GDA software (Lewis and Zaykin, 2001). The presence of null alleles (*NA*) was tested with FreeNA software (Chapuis and Estoup, 2007).

RESULTS AND DISCUSSION

Of the 171 DNA fragments sequenced, it was possible to design 16 primer pairs complementary to SSR flanking region sequences. Seven primer pairs produced clearly interpretable and polymorphic bands and were used in the population analysis (Table 1). All seven loci showed private alleles in at least one of the Acre or Amapá subpopulations.

Genetic diversity estimates of samples from Amapá are presented in Table 2. Among the individuals sampled, we found a high heterozygote proportion, and many alleles were only in heterozygotes. The mean number of alleles was lower than that usually observed in population analysis with SSR markers (Collevatti et al., 2001; Bittencourt and Sebbenn, 2009; Bizoux et al., 2009). Observed and expected heterozygosities were similar between subpopulations.

Table 1. Characterization of seven microsatellite loci developed for *Bertholletia excelsa*, using 12 individuals from Amapá and Acre.

Locus	GenBank	Repeat motif	Primer pair sequence (5'-3')	Allele range (bp)	PA	N_A
Bet 01	EU179316	$(GA)_6$	F: TTTAACTGATGAAAGGCGGACT R: TACGCAGAACAGACTCGCTAAA	222-238	2	4
Bet 05	JX964793	$(TC)_{14}$	F: TAATCTCACAACAAATAACG R: CTAGCTTGATCCTAGAGAAA	102-110	1	5
Bet 06	JX964794	$(TC)_7$	F: CTCTAGGATCAAGCTAGCCCAA R: AGGTTATGCTCCAAATAGCAGG	145-149	1	3
Bet 12	JX964795	$(TC)_{11}$	F: ATAAGGACCGCCCATCATC R: ATAGCGAGAGCAACCTTTGAAC	111-119	2	6
Bet 14	JX964796	$(AG)_{15}$	F: GTGTACTTCTCTGGTTGGGGC R: CCCGAGTTCATTACCCAAACT	97-129	8	6
Bet 15	JX964797	$(GA)_{19}(AGA)_{13}$	F: ACTGCCATCACCAGCATGTAG R: GTCCCTTGTGGTCTCTCACAAT	186-256	8	7
Bet 16	JX964798	$(AG)_9$	F: TCTTCAAACACTCAAAGGGACA R: TGTCTATAAATAGGGGCCTCCC	124-136	4	4
Mean over all loci						5

PA = number of private alleles; N_A = mean number of alleles.

Fixation index (*f*) was positive for the locus Bet 01, not significant for Bet 05 and negative for the other five loci. Estimates of fixation index indicated no inbreeding, with most values less than or equal to zero. No linkage disequilibrium was observed. Moderate and low frequencies of null alleles were found at loci Bet 01 and Bet 05, respectively (Table 2).

Table 2. Genetic diversity of samples from Amapá in each locus, with confidence intervals (95%CI) to fixation index estimates (*f*) and null allele frequencies (*NA*).

	N_A	H_E	H_O	f (95%CI)	NA
Bet 01	4	0.591	0.378	0.364 (0.114, 0.575)	0.13
Bet 05	5	0.643	0.674	-0.049 (-0.304, 0.163)	0.02
Bet 06	3	0.385	0.477	-0.242 (-0.356, -0.153)	0
Bet 12	6	0.833	0.978	-0.176 (-0.232, -0.128)	0
Bet 14	6	0.699	0.867	-0.243 (-0.368, -0.117)	0
Bet 15	7	0.760	0.911	-0.201 (-0.337, -0.077)	0
Bet 16	4	0.624	0.711	-0.141 (-0.318, 0.021)	0
Overall	5	0.648	0.714	-0.103 (-0.176, -0.051)	

N_A = mean number of alleles; H_E = expected heterozygosity; H_O = observed heterozygosity.

The primer pairs designed produced clearly interpretable and polymorphic bands. A small number of alleles were found per locus, but all loci showed private alleles in samples from Acre and Amapá. No linkage disequilibrium was observed. This study provides useful information for population genetic studies of a very important tree species widely distributed in the Amazon Forest.

ACKNOWLEDGMENTS

REFERENCES

Azevedo VCR (2007). Desenvolvimento e Aplicações de Microssatélites, Análise de cpDNA e Modelagem Computacional para Estudos da Estrutura e Dinâmica Genética de Maçaranduba - *Manilkara huberi* (Ducke) Chev. Sapotaceae. Doctoral thesis, UnB, Brasília.

Bittencourt JVM and Sebbenn AM (2009). Genetic effects of forest fragmentation in high-density *Araucaria angustifolia* populations in Southern Brazil. *Tree Genet. Genomes* 5: 573-582.

Bizoux JP, Dainou K, Bourland N, Hardy OJ, et al. (2009). Spatial genetic structure in *Milicia excelsa* (Moraceae) indicates extensive gene dispersal in a low-density wind-pollinated tropical tree. *Mol. Ecol.* 18: 4398-4408.

Buso GSC, Ciampi AY, Moretzsohn MC and Souza ZP (2003). Protocolo para Desenvolvimento de Marcadores Microssatélites. Circular Técnica 20, Embrapa Cenargen, Brasília.

Chapuis MP and Estoup A (2007). Microsatellite null alleles and estimation of population differentiation. *Mol. Biol. Evol.* 24: 621-631.

Collevatti RG, Grattapaglia D and Hay JD (2001). Population genetic structure of the endangered tropical tree species *Caryocar brasiliense*, based on variability at microsatellite loci. *Mol. Ecol.* 10: 349-356.

Doyle JJ and Doyle JL (1987). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Erickson DL, Hamrick JL and Kochert GD (2004). Ecological determinants of genetic diversity in an expanding population of the shrub *Myrica cerifera*. *Mol. Ecol.* 13: 1655-1664.

Goudet J (1995). FSTAT (vers. 1.2): a computer program to calculate F-statistics. *J. Hered.* 86: 485-486.

Lewis PO and Zaykin D (2001). Genetic Data Analysis: Computer Program for the Analysis of Allelic Data. Available at [http://lewis.eeb.uconn.edu/lewishome/software.html]. Accessed January 20, 2012.

Ostrander EA, Jong PM, Rine J and Duyk G (1992). Construction of small-insert genomic DNA libraries highly enriched for microsatellite repeat sequences. *Proc. Natl. Acad. Sci. U. S. A.* 89: 3419-3423.

Peres CA, Baider C, Zuidema PA, Wadt LH, et al. (2003). Demographic threats to the sustainability of Brazil nut exploitation. *Science* 302: 2112-2114.

Rafalski JA, Morgante M, Powell W and Vogel JM (1996). Generating and Using DNA Markers in Plants. Academic Press, New York.

Rozen S and Skaletsky H J (2000). Primer3 for General Users and for Biologist Programmers. Humana Press, Totowa.

Schlötterer C (2004). The evolution of molecular markers - just a matter of fashion? *Nat. Rev. Genet.* 5: 63-69.

Tóth G, Gaspari Z and Jurka J (2000). Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* 10: 967-981.

Vandergast AG, Bohonak AJ, Weissman DB and Fisher RN (2007). Understanding the genetic effects of recent habitat fragmentation in the context of evolutionary history: phylogeography and landscape genetics of a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: *Stenopelmatus*). *Mol. Ecol.* 16: 977-992.

Weir B and Cockerham CC (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.