Biochemical Systematics and Ecology 59 (2015) 168–173

Contents lists available at ScienceDirect

Biochemical Systematics and Ecology

journal homepage: www.elsevier.com/locate/biochemsyseco

Development and use of novel microsatellite markers from double-enriched genomic libraries in Guatemalan *Jatropha curcas*

Ramon da Silva Raposo [a, 1], Fábio Barros Britto [b, 1], Isis Gomes de Brito Souza [c], Adriana Márcia Ferreira de Carvalho [c], Adilson Kenji Kobayashi [d], Bruno Galvêas Laviola [d], Paul Bentzen [e], Fábio Mendonça Diniz [c, *]

[a] *Universidade de Fortaleza, Núcleo de Biologia Experimental (NUBEX), Fortaleza, CE, 60.811-905, Brazil*
[b] *Universidade Federal do Piauí, Campus Professora Cinobelina Elvas, Caixa Postal 58, Bom Jesus, PI, 64.900-000, Brazil*
[c] *Embrapa Mid-North, CP: 01, Teresina, PI, 64.006-220, Brazil*
[d] *Embrapa Agroenergy, PqEB, Brasilia, DF, 70.770-901, Brazil*
[e] *Department of Biology, Dalhousie University, Halifax, NS, B3H 4J1, Canada*

ARTICLE INFO

Article history:
Received 14 November 2014
Accepted 24 January 2015
Available online 14 February 2015

Keywords:
Short tandem repeats
Energy crop
Physical nut
Genetic studies
Polymorphism

1. Introduction

Jatropha curcas is a perennial bush-tree from the Euphorbiaceae family and possibly native to Central America, Mexico or neighboring parts of the continent. The genus *Jatropha*, which contains approximately 170 known species, is found in almost all tropical and subtropical countries of the world (Fairless, 2007). At present, the species *J. curcas,* rich in oil content, has been seriously considered as a biofuel alternative. One of the major limitations for successful cultivation of *J. curcas*, however, is the selection of planting material due to its narrow genetic base (Ranade et al., 2008).

The development of molecular tools for monitoring and selecting natural populations for breeding and conservation programs has become of paramount importance with the increasing interest in *J. curcas* as a crop for oil seed production. The use of molecular markers to gather genetic information for the estimation of intra-specific diversity, the selection of

* Corresponding author. Tel.: +55 86 3198 0593; fax: +55 86 3198 0500.
E-mail address: fabio.diniz@embrapa.br (F.M. Diniz).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.bse.2015.01.019

prospective plant materials and the identification of useful traits in marked-assisted selection can be helpful in efficient breeding strategies (Kumar et al., 2011).

Among the different classes of markers microsatellite DNA, or Simple Sequence Repeats (SSRs), has proved to be of particular interest due to its high levels of inter- and intra-specific polymorphism. Microsatellite loci are present in nuclear and organellar DNA and consist of short stretches of tandemly-repeated DNA with 2–5 base pairs in length. The genetic variation is present in the number of repeats of certain microsatellite loci, a result of slippage of the DNA polymerase during replication (Tautz, 1986). Microsatellites have proven to be extremely useful as nuclear DNA tags and their efficiency is reflected in the fact that polymorphism can be even detected between closely-related individuals (Chen and Cheng, 2013). These multi-allelic markers are easy to score by polymerase chain reaction (PCR)/gel electrophoresis, and allelic variants of a microsatellite locus are codominant and show Mendelian inheritance (Jarne and Lagoda, 1996). So far, only dinucleotide microsatellites are available in the literature for *J. curcas*. However, tri and tetranucleotide repeat microsatellites have the advantage of giving less pronounced "stutter" bands in PCR, more reproducibility, and consequently easier to genotype. Therefore, these types of microsatellite markers may have higher applicability than dinucleotides in population genetic analysis (Ellegren, 2004).

Herein, we describe the isolation and amplification of 23 novel microsatellite loci from *J. curcas*, using a new double-enrichment hybridization procedure, and capture a snapshot of the genetic diversity of Guatemalan accessions referred to in the literature as a putative center of origin in Central America.

2. Materials and methods

2.1. Sample collection, tissue preparation and DNA extraction

Fresh young plant leaves of *J. curcas* were collected from the Biocombustibles de Guatemala S.A. (Guatemala City, Guatemala), stored into a plastic bag with silica gel to dehydrate and kept on ice. Approximately 20 mg of silica gel-dried young leaves were grounded in 2.0 mL microtubes containing ceramic grinding beads (CK28, BIOAMERICA) with two 16 s pulses at 5200 rpm with 10 s intervals on a Precellys®24 Tissue Homogenizer (BERTIN TECHNOLOGIES). High molecular weight DNA was extracted from these using the *DNeasy Plant Mini Kit* (QIAGEN) according to manufacturer's instructions. Extracts were checked for quality and quantity on a 1% agarose gel electrophoresis with ethidum bromide. DNA concentration was measured using NanoDrop 2000 Spectrophotometer (Thermo Scientific) as well. DNA was maintained at −20 °C until further analyses.

2.2. Genomic library construction and SSR isolation

Two genomic libraries enriched for tetranucleotide microsatellites were prepared as outlined in Glenn and Schable (2005) and Diniz et al. (2007) with modifications. In both libraries DNA (≈200 ng/μl) was digested separately with *Hin*cII, *Rsa*I, *Bst*UI and *Hae*III (NEW ENGLAND BIOLABS) in the presence of bovine serum albumin (BSA) overnight. Digests were ligated to double stranded SNX linkers (SNX-For: 5′-CTAAGGCCTTGCTAGCAGAAGC-3′ and SNX-Rev: 5′-pGCTTCTGCTAGCAAGGCCTTAGAAAA-3′) with T4 ligase (NEW ENGLAND BIOLABS) in the presence of *Xmn*I.

Linker-ligated inserts were amplified by a symmetric PCR with the SNX-For linker as primer. Cycle number was optimized to minimize over-amplification of products. PCR conditions were as follows: [1 × 95 °C for 5 min, 20 × (95 °C for 45 s, 62 °C for 1 min, and 72 °C for 2 min), 1 × 72 °C for 30 min]. Amplifications were carried out in 50-μL reaction volume containing 20–100 ng DNA, 1 × Thermopol buffer (+1.5 mM MgCl), 50 μM each dNTP, 0.5 U *Taq* DNA polymerase (Thermopol, NEW ENGLAND BIOLABS), 0.3–0.5 μM of SNX-For on an MJ Research DNA Engine Tetrad PTC-225 thermalcycler.

PCR products (linker-ligated DNA) were recovered using Qiaquick PCR Purification columns. They were enriched for repeats by subtractive hybridization with biotinylated tetranucleotide probes [i.e., Mix A: 2.5 μM $(ACTC)_6$; 2.5 μM $(ACTG)_6$; 2.5 μM $(ACCT)_6$ 2.5 μM $(ACAG)_6$ and Mix B: 2.5 μM $(AATG)_6$; 2.5 μM $(AAAC)_6$; 2.5 μM $(AATC)_6$; 2.5 μM $(AAAG)_6$] (Operon Technologies, Alameda, Calif.) bound to magnetic beads (DYNAL BIOTECH Inc., Lake Success, NY).

Enriched-DNA recovered from the beads was amplified with SNX-For linker to generate double stranded DNA. PCR conditions were similar to those described above, except the number of cycles, which was 10in this case. Amplified-enriched DNA from this hybridization/enrichment step was cleaned with Qiaquick purification columns with a final elution of 50 μL. Then, a second round of hybridization was performed using a small fraction of the post-enrichment amplified inserts, using the same probe mixtures and hybridization conditions described in this section. This double-enrichment procedure was performed to increase the chance of recovering SSR-containing sequences within DNA fragments (Diniz et al., 2007).

Amplified-enriched DNA from the double-enrichment procedure was cleaned with Purelink™ PCR purification kit (INVITROGEN) and ligated into Qiagen pDrive™ Vector (QIAGEN PCR CLONING KIT). The cloning vector was transformed into NEB 5-alpha competent *Escherichia coli* (DH5″ derivative; NEB) and plated on imMedia Amp Blue agar (Invitrogen), and then grown overnight at 37 °C. After the cloning procedure, positive colonies were identified in a white (positive) and blue (negative) screen. The positive colonies were transferred to Luria–Bertani (LB) broth with ampicillin (100 μg/mL) and allowed to grow for 15 hours with constant shaking (225 rpm). Then, 1.0 μL of LB broth containing positive bacteria was used in PCR amplifications of target sequences, using M13 forward (−20) and reverse (−40) primers. Plasmid DNA of appropriate length (500–1000 bp inserts) was prepared using a Qiaprep spin miniprep kit (QIAGEN) and cycle-sequenced [96 °C for 3 min, 40 × (96 °C for 20s, 52 °C for 20s and 60 °C for 4 min] in one direction with a universal primer (T7/M13F) using the BigDye

Terminator v3.1 Cycle Sequencing Kit (APPLIED BIOSYSTEMS). Electrophoresis of sequencing products was performed on an ABI 3130 automated DNA analyzer (APPLIED BIOSYSTEMS). Confirmed positives were further sequenced for the opposite strand using primers SP6/M13R.

After sequencing, repeat regions were initially identified using the software TANDEM REPEAT OCCURRENCE LOCATOR (Castelo et al., 2002). Primer pairs for microsatellite loci were designed on the unique flanking regions of each microsatellite locus using PRIMER 3 (Rozen and Skaletsky, 2000).

2.3. SSR primers validation and characterization of loci

A snapshot of the genetic diversity of Guatemalan *J. curcas* was estimated by the level of polymorphism using a panel of 50 individual accessions. Microsatellites were amplified using the following PCR conditions: (I) 1 × 95 °C for 10 min, 30 × (30 s at 95 °C, 30 s at Ta°C (Table 1), 45 s at 72 °C) and 10 min at 72 °C. To optimize the PCR reaction for some primers, an alternate touchdown (TDW) PCR was used based on the following profile (II): 95 °C for 10 min, 10 × (94 °C for 1 min, Ta°C for 30s with

Table 1
Characteristics of microsatellite loci developed for *Jatropha curcas*.[a]

Locus	Primers sequence (5′-3′)	Core Motif	Ta (°C)	R_S/C_S (bp)	*A*	A_R	H_O	H_E	F_{IS}	Accession number
JcSSR-01	F: GCTCGAAACTCCATTCACTTG R: CACCAGCGTCATTCCATTC	$(AAAC)_6(CGAA)_{12}$	55	265	1	–	–	–	–	JX069297
JcSSR-02	F: AATGGAATGACGCTGGTG R: CGGAGATTGTGTTGCTACTTC	$(AAAC)_{14}$	53	139	1	–	–	–	–	JX069298
JcSSR-03	F: CGAAGTAGCAACACAATCT R: CACGTAAGCTCTCTCACA	$(TG)_{26}$	48	106	1	–	–	–	–	JX069299
JcSSR-04	F: TTTCAGTTTTCCCTCACACG R: AGGGTGTTTGGCTTCTGCG	$(TC)_{14}$	55/60§	88-108/100	6	5.60	0.720	0.648	−0.112	JX069300
JcSSR-05	F: AAGAGAGAACTGAGACGGTTTG R: GATGACGAAGCGTTACTGAAC	$(TTGT)_{10}$	50	132-140/136	3	3.00	0.680	0.537	−0.270*	JX069301
JcSSR-06	F: CAGCTACCGCCCAAGCAC R: CGTCACTCCCGCTAAGTC	$(TTTG)_9$	64	199-227/227	6	6.00	0.775	0.725	−0.070	JX069302
JcSSR-07	F: TCATCAAGCTAACAAGCCACT R: TCAAGTTTGGACTTGGAAGG	$(AAT)_9$	60	182-203/194	7	6.61	0.646	0.601	−0.075	JX069303
JcSSR-08	F: CTTTTGAAGATTTGCCTGTG R: CAATCCCTCCATGTTTCTCC	$(TG)_{34}$	54	113	1	–	–	–	–	JX069304
JcSSR-09	F: TGTGTTTCTGTGTCCAATGTG R: TCGAGAGTATGTGTCTGTTTATGG	$(GT)_{34}$	54	161	1	–	–	–	–	JX069305
JcSSR-10	F: TGAAAGTAAGAAAGGAAG R: TCTCTCTATTTAGTTATCTG	$(AAAG)_{10}(AAAG)_6$	45	225-249/229	6	5.99	0.646	0.734	0.121	JX069306
JcSSR-11	F: CAGACACACTCACATTCCATC R: CCATCACATCTTCCTTTTCTCC	$(TC)_{21}$	54	198	1	–	–	–	–	JX069307
JcSSR-12	F: CGTGATCCGGGTGTATTATC R: AATCAATTCAGCGAACCAAC	$(TG)_{21}$	53	188	1	–	–	–	–	JX069308
JcSSR-13	F: ACATAATGTGGTCAGGTTAT R: AAAGAAGAAAGAAAGAAGG	$(TTTC)_{13}$	47	147	1	–	–	–	–	JX069309
JcSSR-14	F: CTTCCTTTCTCTTTCTCC R: AGATCGCGTGACTGCAC	$(TC)_{16}(TTC)_3$	50	121	1	–	–	–	–	JX069310
JcSSR-15	F: CTCATATTGATGCACGGTAG R: TTCTCCATTCCTTTCTTCC	$(AAAG)_5(GAAA)_{10}$	50	172	1	–	–	–	–	JX069311
JcSSR-16	F: GGAAGAAAGGAATGGAGA R: CTTCTTTTCTTTCTCTTTC	$(AAGA)_7$	50	139	1	–	–	–	–	JX069312
JcSSR-17	F: ACTCGCCAACTTCATAATCG R: TGAAATCCAAGGGAATGG	$(TCA)_6$	58	169-181/172	5	4.81	0.571	0.606	0.057	JX069313
JcSSR-18	F: TGCTGAGGGACAAGTCAAC R: TTCTTCTACACACTCATAGAGACC	$(GA)_{36}(AC)_{13}(AC)_{14}$	53	306	1	–	–	–	–	JX069314
JcSSR-19	F: GGCTTGTGTGGGTGGTATC R: AATAATGCTATCGCTGTGACC	$(AAAC)_{14}$	54	133	1	–	–	–	–	JX069315
JcSSR-20	F: GGATAACTCATTCGTGTCTTGC R: GACGCTACGTGATTTCGAG	$(AACA)_7(AACAAACG)_7$ $(AACAAACGAACAAACA)_4(AACA)_{14}$	54	388	1	–	–	–	–	JX069316
JcSSR-21	F: AGTGGATGAGAACAATGCAC R: GAAAGCAACGGTAGTATCTGG	$(ATC)_6$	58	126-135/129	4	3.99	0.640	0.632	−0.012	JX069317
JcSSR-22	F: GCAATCACATAACCTCCAAGA R: AGCCTTTTCCTGTTGTTTCAG	$(AG)_{14}$	58	271-285/271	7	7.00	0.792	0.743	−0.066	JX069318
JcSSR-23	F: TGTGAAGTGGCTCAAACG R: CAGTAGGAGAAGACAGAGACC	$(CTTC)_6$	53	250	1	–	–	–	–	JX069319

[a] 50 Guatemalan samples genotyped; Ta, annealing temperature; S_R, Observed distribution of alleles at each locus; C_S, Size of the PCR product amplified from the clone used to develop each locus; H_O, Observed heterozygosity; H_E, Expected heterozygosity; F_{IS}, Inbreeding coefficient; *: deviation from Hardy–Weinberg equilibrium; §, touchdown PCR.

annealing temperature decreasing 1 °C per cycle and 72 °C for 30s), and 25 × (94 °C for 1 min, Ta°C for 45s, 72 °C for 45s) and finally 72 °C for 1 min. Amplifications were carried out in 20-μL reaction volume containing 20–100 ng DNA, 1.5–2.5 mM MgCl, 50 μM each dNTP, 0.5 U Taq DNA polymerase (RBC), 0.3–0.5 μM of primer and 1 × PCR buffer (10 mM Tris–HCl, pH 8.3; 50 mM KCl) in a VERITI thermalcycler (APPLIED BIOSYSTEMS). Microsatellite markers were screened by silver nitrate detection on

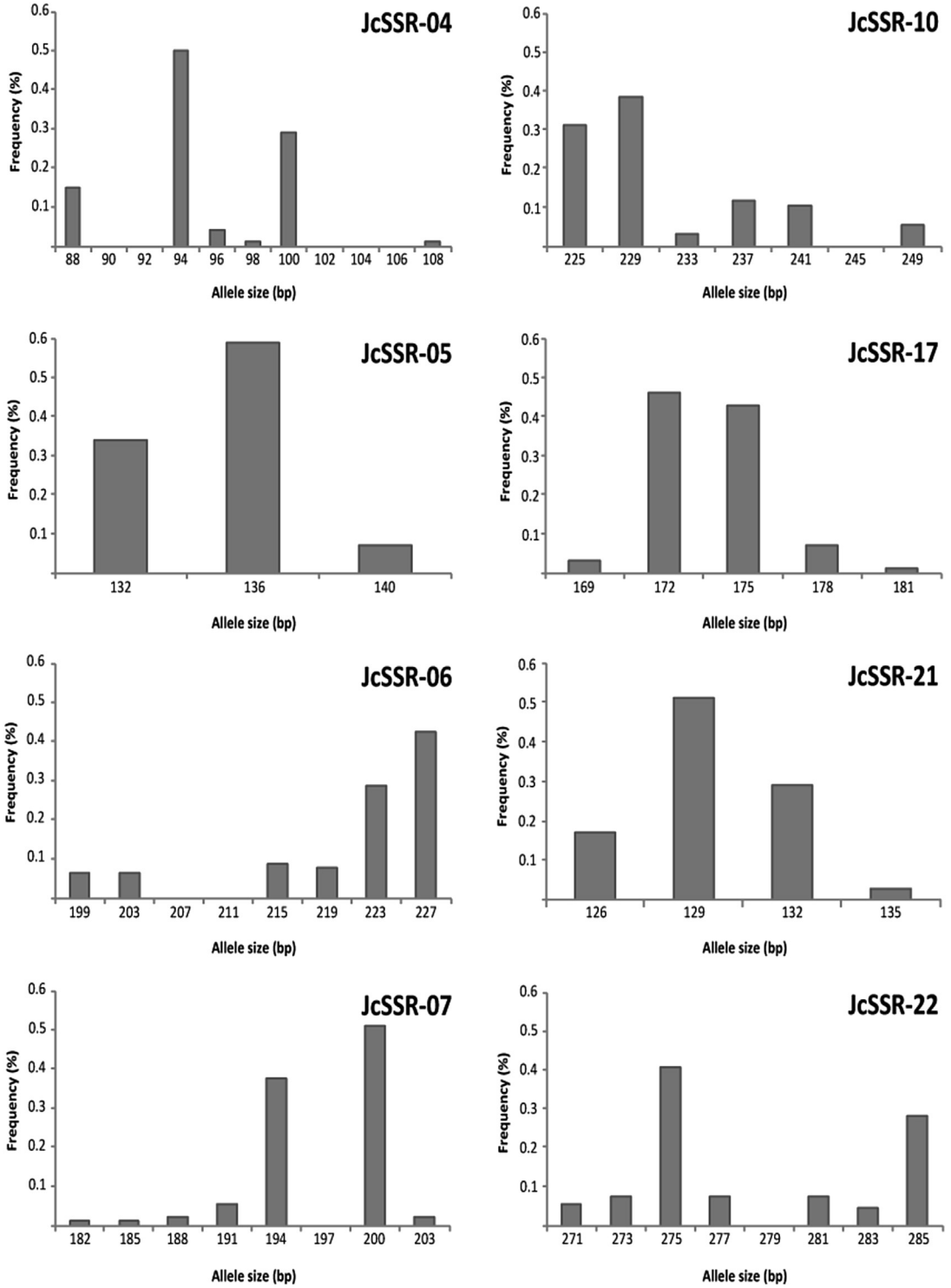

Fig. 1. Distribution of allele frequencies at the eight polymorphic microsatellite loci in *Jatropha curcas* populations.

denatured 6% polyacrylamide gels. Allele sizes were initially determined against a 10 bp DNA ladder (INVITROGEN) and by comparison with the expected size of the cloned fragment and then scored manually. If necessary, two or more runs were performed to verify the allele typing, by re-ordering the samples.

Observed and expected heterozygosities and tests for departure from Hardy–Weinberg expectations (HWE) were performed using GENEPOP (Raymond and Rousset, 1995). The occurrence of linkage disequilibrium between loci and allelic richness for each locus was calculated using FSTAT 2.9.3 (Goudet, 1995). Significance levels were adjusted using sequential Bonferroni corrections. MICRO-CHECKER 2.2.3 (van Oosterhout et al., 2004) was used to check for null alleles.

3. Results and discussions

All tested restriction enzymes (RE) generated fragments between 300 and 1500 bp. However, only *Rsa*I was selected for further experiments because of the intense and homogeneous track shown on the agarose gel electrophoresis profile. The selection of RE for microsatellite development is more related to the features presented by the genome than the enzyme itself. Therefore, the occurrence of enzyme recognition sites (5′-GT↓AC-3′ for *Rsa*I) on the studied genome will determine at which frequency such enzyme will digest the *J. curcas* DNA and, consequently, determine the restriction pattern for a specific enzyme (Lynn et al., 1980).

After cloning and transformation of amplified-enriched DNA from the double-enrichment procedure, 69 positive clones were identified by 'white/blue plaque selection'. PCR amplification of target sequences and subsequent electrophoresis revealed 67 colonies (97% of the clones) containing fragments larger than 500 bp. The high frequency of positive clones having DNA fragments longer than 500 bp increases the chance of detecting sequences with repeat arrays and longer microsatellite flanking regions for efficient primer design (Diniz et al., 2007). The overall success rate of the enrichment protocol employed here was probably due to the presence of large numbers of repeats within the *J. curcas* genome, the effectiveness of using the hybridization capture (biotin/streptavidin) enrichment technique (Kijas et al., 1994), and the addition of a second round of subtractive hybridization with biotinylated probes (Hakki et al., 2002).

Several sequences containing repeat regions were found to be unsuitable for primer design because they either had too short or no available flanking regions. SSRs that *had low GC content* in their flanking region were also excluded. A total of 29 microsatellite primer pairs were designed and after optimization, 23 *loci* (79.3%) had yielded repeatable and scorable PCR products, and 6 failed to amplify. Regions flanking microsatellites with a small number of repeats were not considered for primer designing, because of their likelihood of having very low levels of variability or being fixed (Amos, 1999). Array motifs and GenBank accession numbers are listed in Table 1; most arrays were pure tetranucleotide repeats. Of 23 scorable *loci*, 10 (43.48%) contained tetranucleotide microsatellites of which 7 (70.0%) were pure. Eight *loci* (34.78%) were dinucleotides, three (13.04%) were pure trinucleotides, and two microsatellites (8.69%) were a composition of different core repeat motifs.

Optimization of PCR reactions revealed that optimal annealing temperatures varied according to microsatellite loci (Table 1), but the concentrations of $MgCl_2$ and primer pairs were fixed at 1.5 mM and 0.3 μM, respectively. This would greatly facilitate the designing of a multiplex PCR in order to reduce the number of PCR reactions needed for genotyping a population and consequently the overall cost of the study.

Genotyping of *J. curcas* accessions from Guatemala revealed 15 monomorphic loci. Eight loci were selected for further use based upon polymorphism and allele sizes (Table 1).

The average number of alleles per locus and allelic richness were 5.5 ± 1.4 and 5.4 ± 1.4, respectively, expected heterozygosity was 0.653 ± 0.074 and observed heterozygosity was 0.684 ± 0.074. The allelic sizes ranged from 88 to 285 bp. Distributions of allele frequencies in *J. curcas* populations are depicted in Fig. 1. Analysis with Micro-Checker indicated few null alleles for locus JcSSR-05 in Guatemalan populations, which may be a possible cause of its deviation from Hardy–Weinberg equilibrium, even after Bonferroni correction for multiple comparisons. No loci showed significant linkage disequilibrium after Bonferroni correction.

The eight polymorphic microsatellite loci described in this study are expected to be useful markers for estimating fine scale population processes of the species. The remaining 15 monomorphic loci may also become useful as a new source of data for plant species-level phylogenetics and for population genetics. Current research has indicated that their flanking regions may be used as a new tool for genetic studies (Nazareno and dos Reis, 2011; Chatrou et al., 2009). On the other hand, it is still expected that the reported monomorphic loci may become polymorphic when additional accessions are sampled. Ongoing research is involved in applying these markers for genetic structure in cultivated and natural populations of *J. curcas* throughout its range.

Acknowledgments

This study was supported by the Brazilian Government, the Brazilian Ministry of Science and Technology (MCT) through a FINEP grant (BRJATROPHA Project), The Bank of Northeastern Brazil (BNB/ETENE/FUNDECI) and the Brazilian Agricultural Research Corporation (Embrapa). R.S. Raposo is a doctorate candidate from the Rede Nordeste de Biotecnologia (RENORBIO). I.G.B. Souza and F.B. Britto were supported by CNPq (research scholarship no. 384697/2011-9) and the FAPEPI/CNPq-DCR Fellowship Program, respectively. We thank two anonymous referees and the editor for their remarks and suggestions.

References

Amos, W., 1999. A comparative approach to the study of microsatellite evolution. In: Goldstein, B., Schlotterer, C. (Eds.), Microsatellites: Evolution and Applications. Oxford University Press, Oxford, pp. 67–79.
Castelo, A., Martins, W., Gao, G., 2002. Bioinformatics 18, 634–636.
Chatrou, L.W., Escribano, M.P., Viruel, M.A., Maas, J.W., Richardson, J.E., Hormaza, J.I., 2009. Mol. Phylogenet. Evol. 53 (3), 726–733.
Chen, W., Cheng, Q., 2013. Bio. Sys. Ecol. 47, 111–115.
Diniz, F.M., Iyengar, A., Lima, P.S.C., Maclean, N., Bentzen, P., 2007. Gen. Mol. Biol. 30, 380–384.
Ellegren, H., 2004. Nat. Rev. Gen. 5, 435–445.
Fairless, D., 2007. Nature 449, 652–655.
Glenn, T.C., Schable, N.A., 2005. Methods Enzym. 395, 202–222.
Goudet, J., 1995. J. Hered. 86, 485–486.
Hakki, E., Peev, C., Karagyozov, L., Akkaya, M., 2002. Bulg. J. Plant Physiol. 28 (1–2), 3–10.
Jarne, P., Lagoda, P.J.L., 1996. Trends Ecol. Evol. 11, 424–429.
Kijas, J.M.H., Fowler, J.C.S., Garbett, C.A., Thomas, M.R., 1994. BioTechniques 16, 656–662.
Kumar, S., Kumaria, S., Sharma, S.K., Rao, S.R., Tandon, P., 2011. Biomass. Bioenerg. 35 (7), 3063–3070.
Lynn, S.P., Cohen, L.K., Kaplan, S., Gardner, J.F., 1980. J. Bacteriol. 142, 380–383.
Nazareno, A.G., dos Reis, M.S., 2011. Am. J. Bot. 98 (10), e265–e267.
Ranade, S.A., Srivastava, A.P., Rana, T.S., Srivastava, J., Tuli, R., 2008. Biomass. Bioenerg. 32, 533–540.
Raymond, M., Rousset, F., 1995. J. Hered. 86, 248–249.
Rozen, S., Skaletsky, H.J., 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz, S., Misener, S. (Eds.), Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ, pp. 365–386.
Tautz, D., 1986. Nature 322, 652–656.
van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P., 2004. Mol. Ecol. Notes 4, 535–538.