DOI: 10.1007/s10535-012-0134-y BIOLOGIA PLANTARUM 56 (4): 789-792, 2012

BRIEF COMMUNICATION

New polymorphic microsatellite loci for *Theobroma cacao*: isolation and characterization of microsatellites from enriched genomic libraries

E.S.L. SANTOS[1,2], C.B.M. CERQUEIRA-SILVA[1,2], G.M. MORI[2], D. AHNERT[3], R.X. CORRÊA[3] and A.P. SOUZA[2,4]*

Departamento de Estudos Básicos e Instrumentais, Universidade Estadual do Sudoeste da Bahia, Rodovia BR415, Km03, s/n, Itapetinga, 45700-000, Bahia, Brazil[1].
Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas, Av. Cândido Rondon, 400, CP 6010, Campinas, 13083-875, São Paulo, Brazil[2].
Centro de Biotecnologia e Genética, Departamento de Ciências Biológicas, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna Km 16, Ilhéus, 45662-900, Bahia, Brazil[3].
Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, CP 6109, Campinas, 13083-970, São Paulo, Brazil[4]

Abstract

Seventeen polymorphic microsatellite markers were isolated from enriched genomic libraries for *Theobroma cacao*, providing additional tools for studying the genetic diversity and map saturation of this species. These markers were characterized in 32 accessions of the *T. cacao* germplasm collection from the Centro de Pesquisas do Cacau. The number of alleles at each locus varied from 2 to 8, with an average of 4.41 alleles per locus. The polymorphism information content varied from 0.060 to 0.695, with an average of 0.333. The markers characterized in this study will be employed in map saturation studies and diversity assessments of cacao genotypes.

Additional key words: cacao, genetic diversity, molecular markers, PCR.

Theobroma cacao L. (2n=20) is a neotropical plant that is native to the northern regions of South America and has been domesticated in Central America (Motamayor *et al.* 2002). Among the 22 species of the *Theobroma* genus cacao is the only species cultivated on a large scale for chocolate production (Luz and Silva 2001). International genetic breeding programs for cacao have been developed as a means of obtaining cultivars with improved productivity, increased resistance and good industrial quality (Guiltinan *et al.* 2008). As a useful tool for genetic studies, molecular markers, especially microsatellite markers, have been developed for different species, such as *Jatropha curcas* (Phumichai *et al.* 2011) and *Aegiphila sellowiana* (Ruas *et al*. 2011). Specifically in cacao, some simple sequence repeat (SSR) markers have already been developed (Lanaud *et al.* 1999, Pugh *et al.* 2004, Araújo *et al.* 2007, Lima *et al.* 2008) and have been used for diversity studies and breeding programs (Aikpokpodion *et al.* 2009, Lanaud *et al.* 2009, Yamada *et al.* 2009). To study the genetic diversity and genetic map saturation of *T. cacao*, we isolated and characterized a new set of polymorphic microsatellites from *T. cacao*. These markers may also be used for future genetic studies of this species.

To identify and characterize microsatellites, two microsatellite-enriched libraries for *T. cacao* were developed following the protocol described by Billotte *et al.* (1999) from two phenotypically and genetically

---

Received 14 July 2011, *accepted* 20 December 2011.
Abbreviations: CTAB - cetyltrimethylammonium bromide; HWE - Hardy-Weinberg equilibrium; LD - linkage disequilibrium; PCR - polymerase chain reaction; PIC - polymorphism information content; SSR - simple sequence repeat; TBE - Tris/borate/EDTA
Acknowledgments: The authors are grateful to Conselho de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, PROCAD-NF2008) for financial support, to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for providing graduate fellowships to G.M. Mori and E.S.L. Santos, respectively, and a research fellowship (from CNPq) to A.P. Souza.
* Corresponding author: fax: (+55) 19 35211089, e-mail: anete@unicamp.br

distinct clones: TSH 1188 (Upper Amazon rainforest of Trinidad) and GU 261 (Lower Amazon rainforest of French Guiana). These clones were sampled from the germplasm collection at the Centro de Pesquisas do Cacau/Comissão Executiva do Plano da Lavoura Cacaueira (CEPEC/CEPLAC) in Itabuna, Bahia, Brazil. Genomic DNA was isolated from fresh leaves from each clone using a cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle 1990) with modifications specific for cacao (Faleiro *et al.* 2002). The DNA samples (5 µg) were digested with *Afa*I and ligated to the double-stranded adapters 5′-CTCTTGCTTACGCGTGGACTA-3′ and 5′-TAGTCCACGCGTAAGCAAGAGCACA-3′. The enrichment was performed using a hybridization-based capture with $(GT)_8$ and $(CT)_8$ biotin-linked probes and magnetic-coated streptavidin beads (streptavidin *Magne-Sphere* paramagnetic particles, *Promega*, Madison, WI, USA). Selected fragments were then cloned into a *pGEM-T Easy* (*Promega*) plasmid vector and transformed into *Escherichia coli* XL1-Blue competent cells (*Stratagene*, La Jolla, CA, USA). Recombinant colonies were selected by blue/white screening. Clones were randomly selected and were double-sequenced using the T7 and SP6 primers and the *v3.1 Big Dye* terminator kit (*Applied Biosystems*, Foster City, CA, USA) on an *ABI PRISM 377* automated DNA sequencer (*Applied Biosystems*). Every obtained sequence was aligned, edited and eliminated if redundant using *SeqMan* software (*DNASTAR*, Madison, WI, USA). *MICROSAT* software (A.M. Risterucci, CIRAD, personal communication) was used to eliminate adapters and restriction sites from the sequences. The simple sequence repeat identification tool (*SSRIT*) (Temnykh *et al.* 2001) was used to identify the following microsatellites: di- and trinucleotides with four or more repeats; and tetra- and pentanucleotides with three or more repeats.

Primer pairs were designed using the *PrimerSelect* (*DNASTAR*) and *Primer3Plus* (Untergasser *et al.* 2007) programs. The polymerase chain reaction (PCR) was conducted in a final volume of 0.015 cm^3 containing 15 ng of template DNA, 1× PCR buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl), 1.5 mM $MgCl_2$, 0.2 µM of

Table 1. Characteristics of the forward (F) and reverse (R) primers, repeat motif, and amplification conditions (AC) of 17 microsatellite loci from *Theobroma cacao* L. * - TD65-55 indicates touchdown PCR with temperatures ranging from 65 to 55 °C.

Locus	GenBank acc.	Primer sequence (5´-3´)		Repeat motif	AC*
mTc-UNICAMP01	JF812968	F:	TCATGCAAAGCAAAGTGAAG	$(AG)_{11}$	TD65-55
		R:	ACGGGAAACTCATCATTACA		
mTc-UNICAMP02	JF812969	F:	GGTCCTCCAAGCTGAGTAACA	$(TC)_{15}$	TD65-55
		R:	CTCCCTATTTGCATCGCATT		
mTc-UNICAMP03	JF812970	F:	GCGATGCCAAAAGTTGTGTA	$(AG)_{16}$	TD65-55
		R:	AGAGAAGGGATGGGTGTGTG		
mTc-UNICAMP04	JF812971	F:	GCAGCGAGAGACAAAGATA	$(TC)_{21}$	60 ºC
		R:	TTTGACTGAAATGGTGGTAA		
mTc-UNICAMP05	JF812972	F:	AGCTGTTTATGATTCACATCC	$(TA)_7(CA)_6$	60 ºC
		R:	GAAGCAGCAATTGTAACCAC		
mTc-UNICAMP06	JF812973	F:	AACTCTGGCATTTCATTTGG	$(TC)_{16}TGT(CA)_{10}$	60 ºC
		R:	GTGATGGATGCAGGTGGAAT		
mTc-UNICAMP07	JF812974	F:	CATGCCTGAGCTATGAACA	$(GA)_{11}...(GA)_5$	TD65-55
		R:	TCAGATATGGGTAAAGGAAGAG		
mTc-UNICAMP08	JF812975	F:	GGGAAACTGGGCATCACTTA	$(AC)_5AG(AT)_9$	60 ºC
		R:	TTAATGGGTTGAGGAGAGCAT		
mTc-UNICAMP09	JF812976	F:	TTCGGCAGTTCGATCTATGA	$(TAAAG)_7$	TD65-55
		R:	ATCCACCGTAAGCCTTTCCT		
mTc-UNICAMP10	JF812977	F:	AACTGTCGCATACGCTCATC	$(AG)_9$	60 ºC
		R:	TCAAAAGTGAAGCCCACCAT		
mTc-UNICAMP11	JF812978	F:	GCTTGTTCCTCGACACACAT	$(CA)_6$	60 ºC
		R:	TGGAAGCTAGTTAGGAATCACA		
mTc-UNICAMP12	JF812979	F:	GCCATCTCGCTATTCATGTG	$(AG)_{16}$	60 ºC
		R:	GTCATTGCTAACGTAACCAGAA		
mTc-UNICAMP13	JF812980	F:	ACTTGTCGCTTGCTTCCTGT	$(AT)_6$	TD65-55
		R:	TTCCTCAACCTCTCCGAATG		
mTc-UNICAMP14	JF812981	F:	ACATGATGCCTGAGCTGTTG	$(AG)_{11}$	TD65-55
		R:	CTCCAATTTTGACCCCAGAA		
mTc-UNICAMP15	JF812982	F:	TGCTTTGAGGCACTTGTCC	$(TG)_{12}$	TD65-55
		R:	TGGCATTCATTGAGAGGTGA		
mTc-UNICAMP16	JF812983	F:	CAGGAAGGATACTTCTTAAAGG	$(CT)_{12}(CA)_9...(CT)_4$	TD65-55
		R:	AGTAGAGTCGAGTGGCTTGA		
mTc-UNICAMP17	JF812984	F:	CTGCACAGCTTCATGGACTC	$(CA)_9$	TD65-55
		R:	TGATGATCAGGTGGTTTCTCA		

Table 2. Characteristics of the product size range in base pairs, number of effectively evaluated genotypes (N), number of alleles (N_A), observed (Ho) and expected (H_E) heterozygosity, polymorphism information content (PIC) and *P* value of Hardy-Weinberg equilibrium (HWE) for 17 microsatellite loci from *Theobroma cacao* (* - departs significantly from HWE at $P < 0.003$ after the Bonferroni correction,† - significant values for 'null allele' occurrence after the Bonferroni correction).

Locus	Size range	N	N_A	Ho	H_E	PIC	*P* value
mTc-UNICAMP01	190-208	29	7	0.269	0.286	0.273	0.194
mTc-UNICAMP02	260-290	31	8	0.381	0.746	0.695	0.000*†
mTc-UNICAMP03	282-291	31	5	0.433	0.719	0.661	0.000*
mTc-UNICAMP04	162-186	30	7	0.786	0.742	0.695	0.024
mTc-UNICAMP05	233-242	29	4	0.269	0.544	0.454	0.003†
mTc-UNICAMP06	140-170	32	4	0.107	0.123	0.120	0.032
mTc-UNICAMP07	265-312	28	4	0.240	0.225	0.211	1
mTc-UNICAMP08	278-286	32	5	0.581	0.673	0.605	0.044
mTc-UNICAMP09	148-190	31	6	0.032	0.667	0.596	0.000*†
mTc-UNICAMP10	230-236	32	3	0.094	0.091	0.088	1
mTc-UNICAMP11	290-296	32	2	0.226	0.204	0.180	1
mTc-UNICAMP12	240-290	32	6	0.226	0.369	0.347	0.008†
mTc-UNICAMP13	140-144	30	2	0	0.066	0.064	0.017†
mTc-UNICAMP14	254-258	28	3	0.174	0.166	0.156	1
mTc-UNICAMP15	292-302	32	2	0.067	0.064	0.060	1
mTc-UNICAMP16	244-254	31	4	0.207	0.255	0.240	0.153
mTc-UNICAMP17	296-300	29	3	0.174	0.236	0.213	0.311

each dNTP, 0.5 μM of each primer and 1 U of *Taq* DNA polymerase (*Invitrogen*, Carlsbad, CA, USA). All primers were first evaluated in a PCR amplification protocol with a 60 °C annealing temperature as follows: 94 °C for 5 min; 34 cycles of 94 °C for 1 min, 60 ºC for 1 min, 72 °C for 1 min; and a final extension at 72 °C for 10 min. Later, if necessary, touchdown PCR was performed. The touchdown protocol consisted of 5 steps as follows: *a*) initial denaturation (94 °C for 5 min); *b*) 10 cycles of denaturation (94 °C for 1 min), annealing (65 - 1 °C per cycle for 1 min) and extension (72 °C for 1 min); *c*) step *b* repeated; *d*) 14 cycles of denaturation (94 °C for 1 min), annealing (55 °C for 1 min) and extension (72 °C for 1 min); *e*) a final extension (72 °C for 10 min). The amplification products were visualized in 0.5× TBE/3 % agarose gels stained with ethidium bromide prior to vertical electrophoresis in 1× TBE/6 % denaturing polyacrylamide silver nitrate–stained gels. Product sizes were determined by comparison with a 10 bp DNA ladder (*Invitrogen*). Statistical analyses were performed with *GENEPOP* on the web (Raymond and Rousset 1995), *PIC Calculator* (Kemp 2002) and *MICRO-CHECKER* (Oosterhout *et al.* 2004).

In all, 192 positive colonies were sequenced and 163 microsatellites were obtained. Dinucleotide motifs were the most abundant, followed by tri-, tetra- and penta-nucleotide motifs (with proportions of approximately 94, 3, 2 and 1 %, respectively). Thirty-five primer pairs were designed such that the product size ranged between 100 and 300 bp for accurate genotyping using vertical electrophoresis in 1× TBE/6 % denaturing polyacrylamide gels. Twenty-five of these primer pairs generated consistent patterns of amplification, with fragments matching size expectations based on the sequence information, and these pairs were used for further characterization. For polymorphism analysis, 32 DNA samples of *T. cacao* were extracted and analyzed from the germplasm collection of *CEPEC*/*CEPLAC*. To sample greater genetic variation, we selected genotypes collected from the Amazon in addition to Bahia cacao farms. Of these 25 primer pairs evaluated, 17 tagged polymorphic loci whose numbers of alleles varied (Table 1). To measure the frequency of identity between the polymorphic loci presented here and those described in the literature, we used nucleotide *BLAST* (Altschul *et al.* 1990) (using the megablast algorithm). We observed identity (varying from 95 to 97 %) to four sequences (loci: mTc-UNICAMP03, 04, 07, 09). However, these primers had annealing sites different from those described previously. Of these four loci, three (mTc-UNICAMP03, 04 and 07) were similar to microsatellite loci published by Pugh *et al.* (2004) and are included here to evaluate statistical parameters such as polymorphism information content (PIC) and Hardy-Weinberg equilibrium (HWE). No previous paper has described locus mTc-UNICAMP09. The mean number of alleles was 4.41 (range: 2 to 8 alleles), and the PIC varied from 0.060 to 0.695. Genotypic frequencies of the 17 polymorphic microsatellite markers were tested for HWE and linkage disequilibrium (LD) using the Bonferroni correction. The analysis revealed significant deviations from HWE at three microsatellite loci (Table 1), possibly due to the occurrence of null alleles (as observed for mTc-UNICAMP02 and 09) or LD (as observed for mTc-UNICAMP02, 03 and 09). Significant LD was observed among 19 of 136 possible pairs of markers (mTc-UNICAMP01 and 04, 05, 11, 12; mTc-UNICAMP03 and 07, 11; mTc-UNICAMP05 and 14; mTc-UNICAMP12 and 03, 04, 05, 11; mTc-UNICAMP16 and 01, 02, 03, 06, 07, 09, 11, 12).

The probability of 'stuttering', 'large allele dropout' and 'null alleles' was also evaluated. No microsatellite loci showed a significant probability of 'stuttering' or 'large allele dropout'. Five loci (mTc-UNICAMP02, 05, 09, 12, 13) showed significant values for 'null alleles' according to the Bonferroni correction ($P < 0.003$). Eight other microsatellite markers were monomorphic for the evaluated population.

Microsatellite markers characterized in this study may be used in future genetic map saturation and genetic diversity studies of cacao genotypes.

References

Aikpokpodion, P.O., Motamayor, J.C., Adetimirin, V.O., Adu-Ampomah, Y., Ingelbrecht, I., Eskes, A.B., Schnell, R.J., Kolesnikova-Allen, M.: Genetic diversity assessment of sub-samples of cacao, *Theobroma cacao* L. collections in West Africa using simple sequence repeats marker. - Tree Genet. Genom. **5**: 699-711, 2009.

Altschul, S., Gish, W., Miller, W., Myers, E., Lipman, D.: Basic local alignment search tool. - J. mol. Biol. **215**: 403-410, 1990.

Araújo, I.S., Intorne, A.C., Pereira, M.G., Lopes, U.V., De Souza Filho, G.A.: Development and characterization of novel tetra-, tri- and di-nucleotide microsatellite markers in cacao (*Theobroma cacao* L.). - Mol. Breed. **20**: 73-81, 2007.

Billotte, N., Lagoda, P.J.R., Risterucci, A.M., Baurens, F.C.: Microsatellite-enriched libraries: applied methodology for the development of SSR markers in tropical crops. - Fruits **54**: 277-288, 1999.

Doyle, J.J., Doyle, J.L.: Isolation of plant DNA from fresh tissue. - Focus **12**: 13-15, 1990.

Faleiro, F.G., Santos, I.S., Bahia, R.C.S., Santos, R.F., Yamada, M.M., Anhert, D.: Otimização da extração e amplificação de DNA de *Theobroma cacao* L. visando obtenção de marcadores RAPD. [Optimization of DNA extraction and amplification of *Theobroma cacao* L. aiming obtainment of RAPD markers.] - Agrotrópica **14**: 31-34, 2002. [In Port.]

Guiltinan, M.J., Verica, J., Zhang, D., Figueira, A.: Genomics of *Theobroma cacao,* "the food of the Gods". - In: Moore, P.H., Ming, R. (ed.): Genomics of Tropical Crop Plants. Pp. 145-170. Springer, New York 2008.

Kemp, S.: PIC Calculator. Extra. http://www.genomics.liv.ac.uk/animal/Pic1.html. 2002 <Accessed 08 march 2011>

Lanaud, C., Fouet, O., Clément, D., Boccara, M., Risterucci, A.M., Surujdeo-Maharaj, S., Legavre T., Argout, X.: A meta-QTL analysis of disease resistance traits of *Theobroma cacao* L. - Mol. Breed. **24**: 361-374, 2009.

Lanaud, C., Risterucci, A.M., Pieretti, I., Falque, M., Bouet, A., Lagoda, P.J.L.: Isolation and characterization of microsatellites in *Theobroma cacao* L. - Mol. Ecol. **8**: 2141-2143, 1999.

Lima, L.S., Gramacho, K.P., Gesteira, A.S., Lopes, U.V., Gaiotto, F.A., Zaidan, H.A., Pires, J.L., Cascardo, J.C.M., Micheli, F.: Characterization of microsatellites from cacao-*Moniliophthora perniciosa* interaction expressed sequence tags. - Mol. Breeding **22**: 315-318, 2008.

Luz, E.D.M.N., Silva, S.D.V.M.: Podridão-parda dos frutos, cancro e outras doenças causadas por *Phytophthora* no cacaueiro. [Black pod, cancer and other diseases caused by *Phytophthora* in cocoa.] - In: Luz, E.D.M.N., Santos, A.F., Matsuoka, K., Bezerra, J.L. (ed.): Doenças Causadas por *Phytophthora* no Brasil. [Diseases Caused by *Phytophthora* in Brazil.] Pp. 175-265. Rural, Campinas 2001. [In Port.]

Motamayor, J.C., Risterucci, A.M., Lopez, P.A., Ortiz, C.F., Moreno, A., Lanaud, C.: Cacao domestication I: the origin of the cacao cultivated by the Mayas. - Heredity **89**: 380-386, 2002.

Oosterhout, C.V., Hutchinson, W.F., Wills, D.P.M., Shipley, P.: *MICRO-CHECKER*: software for identifying and correcting genotyping errors in microsatellite data. - Mol. Ecol. Notes **4**: 535-538, 2004.

Phumichai, C., Phumichai, T., Kongsiri, N., Wongkaew, A., Sripichit, P., Kaveeta, R.: Isolation of 55 microsatellite markers for *Jatropha curcas* and its closely related species. - Biol. Plant. **55**: 387-390, 2011.

Pugh, T., Fouet, O., Risterucci, A.M., Brottier, P., Abouladze, M., Deletrez, C., Courtois, B., Clement, D., Larmande, P., N'Goran, J.A., Lanaud, C.: A new cacao linkage map based on codominant markers: development and integration of 201 new microsatellite markers. - Theor. appl. Genet. **108**: 1151-1161, 2004.

Raymond, M., Rousset, F.: *Genepop* (version 1.2): population genetics software for exact tests and ecumenicism. - J. Hered. **86**: 248-249, 1995.

Ruas, E.A., Damasceno, J.O., Conson, A.R.O., Costa, B.F., Rodrigues, L.A., Reck, M., Santos Vieira, A.O., Ruas, C.F., Medri, C., Ruas, P.M. Isolation and characterization of eleven polymorphic microsatellite loci in *Aegiphila sellowiana* and their transferability. - Biol. Plant. **55**: 396-399, 2011.

Temnykh, S., De Clerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., McCouch, S.: Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. - Genome Res. **11**: 1441-1452, 2001.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., Leunissen, J.A.M.: Primer3Plus, an enhanced web interface to Primer3. - Nucl. Acids Res. **35**: 71-74, 2007.

Yamada, M.M., Faleiro, F.G., Flores, A.B., Lopes, U.V., Pires, J.L., Corrêa, R.X., Santos, R.F.: Microsatellite diversity and heterozygosity of parents of a cocoa breeding population. - Crop Breed. appl. Biotechnol. **9**: 17-22, 2009.