Mol Breeding (2007) 20:73–81
DOI 10.1007/s11032-006-9057-7

SHORT COMMUNICATION

Development and characterization of novel tetra-, tri- and di-nucleotide microsatellite markers in cacao (*Theobroma cacao* L.)

Ioná S. Araújo · Aline C. Intorne · Messias G. Pereira · Uilson V. Lopes · Gonçalo A. de Souza Filho

Received: 7 August 2006 / Accepted: 27 September 2006 / Published online: 13 April 2007

Abstract The cacao plant, *Theobroma cacao* L., produces white seeds (beans) that form the major ingredient of processed chocolate. A great deal of research effort has been expended to the development of new genetically modified cacao plants with improved productivity and resistance and beans of good industrial quality. The availability of suitable genetic markers is an important aspect of the efficient selection and breeding of this perennial species. We describe the development of 123 microsatellite loci of cacao. An optimized protocol was used to construct and screen a microsatellite-enriched genomic library from which we isolated 64 di-nucleotide, 45 tri-nucleotide and 14 tetra-nucleotide microsatellite loci. The primers were tested on samples from five different *T. cacao* accessions, one accession from *T. grandiflorum* and one accession from *Herranea sp*. Among the 123 loci, 54 were polymorphic, 61 were monomorphic and eight did not present an amplification product. These new markers will be useful in future studies by increasing the accuracy of genotypic assessments in diverse cocoa tree populations as well as in other species of the *Theobroma* genus.

Keywords Microsatellite · Molecular markers · Primers · *Theobroma*

I. S. Araújo · M. G. Pereira
Laboratório de Melhoramento Genético Vegetal, Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, RJ 28013-600, Brazil

A. C. Intorne · G. A. de S. Filho (✉)
Laboratório de Biotecnologia, Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, RJ 28013-600, Brazil
e-mail: goncalos@uenf.br

U. V. Lopes
Centro de Pesquisas do Cacau (CEPEC)/CEPLAC, Caixa Postal 7, Itabuna, BA 45600-000, Brazil

Theobroma cacao L. is unique among the 22 species of its genus in that it is commercially exploited on a large scale for the production of cacao beans. Cocoa butter and powder, which are extracted from the fermented and dried cocoa beans, are the main ingredients used for the commercial manufacture of chocolate. Improved cocoa cropping requires the development of genetic materials that have a higher productivity and an increased resistance as well as cacao beans of good industrial quality (Figueira and Cascardo 2001). To obtain such characteristics, plant geneticists have made abundant use of molecular markers, especially microsatellites. Microsatellites are powerful genetic markers due to several characteristics, including their abundance in eukaryotic genomes, high levels of polymorphism, Mendelian inheritance, co-dominance and locus-specificity (Merdinoglu et al. 2005). In cocoa, microsatellites have been applied in studies of DNA fingerprint, genetic diversity, variety characterization and genetic

mapping (Charters and Wilkinson 2000; Faleiro et al. 2004; Pugh et al. 2004; Saunders et al. 2004).

In 1999, Lanaud et al. developed the first group of simple sequence repeat (SSR) markers for *T. cacao*. More recently, Pugh et al. (2004) developed 387 new SSR markers for this species. However, all of these SSR loci were isolated using dinucleotide probes during the screening of the genomic library. Consequently, SSR loci consisting of repeats of tri- and tetra-nucleotides remain to be searched in the *T. cacao* genome. The aim of the present study was to develop a new group of SSR markers, including tri- and tetra-nucleotide repeats. To this end, we have attempted to construct 13 genomic libraries, enriched for different SSR sequences, and subsequently to use these libraries to identify and characterize the new microsatellites.

Genomic DNA was extracted from leaves of the *T. cacao* (Scavina-6 clone) following the method of Faleiro et al. (2002). The DNA (1.0 µg) was first digested with the *Alu*I and *Hae*III restriction enzymes (New England Biolabs. Beverly, Mass.) at 37°C for 4 h. The DNA fragments, ranging in size from 250 to 750 bp, were then electrophoresed on a 0.8% agarose gel, excised from the gel and purified with the Wizard SV Gel kit (Promega, Madison, Wis.). The product was linked to the ''*Hae* adapter'' (formed by the association of the oligonucleotides *Hae*1: 5′-CCATCCGCGGCTAG-CAGCATAAAA-3′ and *Hae*2: 5′-AT-GCTGCTAGCCGCGGATGG-3′) in the presence of T4 DNA ligase (Promega). Following ligation, 1-µl aliquots were amplified by PCR in the presence of the *Hae*2 oligonucleotide. The amplified samples were used to construct 13 different SSR-enriched libraries using hybridization capture method. After melting at 94°C for 5 min, each sample was incubated with one of 13 distinct biotin-labeled SSR oligonucleotides: $(AC)_{10}$, $(AT)_{12}$, $(GA)_{13}$, $(GC)_8$, $(GT)_{13}$, $(TC)_{10}$, $(CAA)_7$, $(TAT)_7$, $(AAT)_7$, $(ATT)_7$, $(GATA)_5$, $(ACAG)_5$ or $(GACA)_5$. Fragments containing SSR regions were captured with streptavidin-conjugated magnetic particles (Streptavidin MagneSphere Paramagnetic Particles; Promega). After elution, the enriched fragments were amplified by PCR with the primer *Hae*2. The products were cloned in the pBluescriptKS plasmid (Stratagene, La Jolla, Calif.). Among the 1536 candidate clones, 556 were picked and submitted to fluorescent DNA sequencing, of which 222 were found to be positive for the presence of microsatellites. Among these sequences, 123 were selected for primer design with the PRIMER 3 software (http://frodo.wi.mit.edu/cgi-bin/promer3primer3.bin).

Among the SSRs selected, 64 consisted of sequences containing di-nucleotides, 45 contained tri-nucleotides and 14 contained tetra-nucleotides. The number of replications varied from four (AG, AT, CG, CT, GA, TC, TG, AGG, CAA, CAG, CTG, TGC, TTG and GACA) to 46 (AG). The CAA trinucleotide was the most abundant repetition found in the *T. cacao* L. genome. None of the 123 new regions was homologous to cocoa SSR sequences previously deposited in the GenBank sequence database.

The developed primers were used in PCR reactions with DNA samples from five cocoa accessions (EEG29, EET 397, CCN 10, RB 39 and CAB 169), one accession of *T. grandiflorum* L. and one accession of *Herranea sp/* (a genus genetically related to *Theobroma*) obtained from the Germplasm Collection of the Cocoa Research Center ICEPEC), Ilhéus, Bahia, Brazil. The reactions were performed in a 20-µl final volume containing 10 ng DNA, 10 m*M* Tris-HCl (pH 8.3), 50 m*M* KCl, 2 m*M* $MgCl_2$, 100 μM each desoxyribonucleotide (dATP, dTTP, dGTP and dCTP), 0.2 μM of each primer (forward and reverse) and 1.0 U *Taq* DNA polymerase (Promega). The amplifications were carried out in a Mastercycler gradient cycler (Eppendorf) according to the following program: 94°C for 4 min, followed by 32 cycles at 94°C for 30 s, T_a of each primer (see Table 1) for 1 min, 72°C for 1 min, and a final extension at 72°C for 7 m. Following amplification, the samples were electrophoresed on a high-resolution agarose gel (4%) (Sigma, St. Louis, Mo.), and the allele sizes were estimated by comparison with a 10-bp standard molecular weight (Sigma). The results revealed that 54 SSR were polymorphic, 61 (49%) were monomorphic and eight (8%) did not produce any amplification product. Of the 115 microsatellites that had produced amplification product, 76 were positive for *T. grandiflorum* and *Herranea* sp. and 16 were exclusive for cacao. These new markers will be useful in breeding programs for cocoa involving marker-assisted selection and diversity genetics studies and have already been used for genome mapping.

Table 1 Genetic characterization of microsatellite loci isolated from *Theobroma cacao* L.

Locus	Repeat motif	Primer sequence (5′→ 3′)[a]	T_a(°C)[b]	Allele size (bp)	Polymorphic[c]
UENF/CEPLAC 04	$(GA)_{28}$	(F) TGTTATTTTACCGGCAGT	45	390	No
		®) TCTTGTTAAGAATCAT			
UENF/CEPLAC 05	$(GT)_{14}$	(F) ACCTGCTGTATGTCT	45	170	No
		(R) TGCTAGCCGCGGA			
UENF/CEPLAC 06	$(GT)_4$	(F) TCACTATAGGGCGAATTGGA	55	ND	ND
		(R) ACTTGTGGGGATTTATGGTA			
UENF/CEPLAC 08	$(AT)_4$	(F) AGCATATAGTACAATTACAT	40	310	No
		(R) AACTTCGCTCAT			
UENF/CEPLAC 09	$(GA)_4$	(F) ACACTTGGACTTTGT	40	210	No
		(R) ACCCGATCTCTGTCACA			
UENF/CEPLAC 10	$(GAA)_5$	(F) TGCAGCGGTGGCGGTGGA	50	100	Yes
		(R) TCGAGCCTTCTTCTTCT			
UENF/CEPLAC 11	$(AG)_{12}$	(F) TGCCGCCAAAGAGAGAACTA	47	180	Yes
		(R) TGATCGGTCCTTCCT			
UENF/CEPLAC 12	$(GA)_8$	(F) AGGGCTTTTGGGTTACGAT	54	180	No
		(R) TTTCTTTACTCCGACGTCGA			
UENF/CEPLAC 13	$(AG)_{14}$	(F) TCGGCAGGAGAAGAGTGAGA	50	150	Yes
		(R) AACCCAAGCACCAGAT			
UENF/CEPLAC 14	$(TC)_{18}$	(F) ACACCTTACATGCGTCCACA	54	280	Yes
		(R) AATTTAGAAGACACCCTGGAT			
UENF/CEPLAC 15	$(AG)_{19}$	(F) ATCCTTGCTTGCTTCTCCT	54	180	Yes
		(R) TCACATTCCAACTGGCGAT			
UENF/CEPLAC 16	$(AG)_6$	(F) TTAACGCTGCTGTTGATA	50	180	Yes
		(R) ACCGCCTTTCTTCACTCCTT			
UENF/CEPLAC 17	$(GA)_4$	(F) TCGATCAAAGAACGGTGAA	50	210	No
		(R) AATCGGTGGCAAACCTA			
UENF/CEPLAC 18	$(AT)_6(GT)_4$	(F) TTTGCCACACATTTGAAT	50	220	Yes
		(R) ACTCCACCTTCATCA			
UENF/CEPLAC 19	$(AG)_{20}$	(F) AGATGGGGAAGAGACATA	51	200	Yes
		(R) ACAAAAGAAACATGGGAGCA			
UENF/CEPLAC 20	$(AG)_{21}$	(F) TCCTGCCATTAGAGCGTGT	45	120	Yes
		(R) AAGAGCGGTTTCACT			
UENF/CEPLAC 21	$(AT)_6$	(F) TGCGCACTCTCTACTTGCA	54	350	No
		(R) ATGATTACGCCAAGCTCGA			
UENF/CEPLAC 23	$(AT)_4$	(F) ATATCATTACACAATA	40	ND	ND
		(R) ACCGGTGAGTGAACTTGCAT			
UENF/CEPLAC 24	$(TG)_4$	(F) TGCCAATTGAGTTGTTGTGA	53	110	Yes
		(R) AGTGGAGATGTGGGGTCT			
UENF/CEPLAC 26	$(AG)_{46}$	(F) TAATTGGCATACCTTTGT	40	290	No
		(R) ACACGAGCGATAT			
UENF/CEPLAC 27	$(CT)_{11}$	(F) ATATGTGATCTGTGTGT	45	130	Yes
		(R) AGGGAGCGAAGAGAGAG			
UENF/CEPLAC 28	$(GA)_{22}$	(F) TGAGACGGATATGGA	43	200	Yes
		(R) AGACTCTCTAATATATG			

Table 1 continued

Locus	Repeat motif	Primer sequence (5′→ 3′)[a]	T_a(°C)[b]	Allele size (bp)	Polymorphic[c]
UENF/CEPLAC 30	$(AG)_7$	(F) TTTTGGATTTCCCCAGT	48	400	No
		(R) ACCGGCAGCAACCCA			
UENF/CEPLAC 31	$(AG)_4$	(F) TTGTTATCCTTTCCCA	44	190	No
		(R) TAAACTCCAGATCTCA			
UENF/CEPLAC 32	$(AC)_6(N)_{15}(CA)_8$	(F) AGCCAAAACATATCCAACACA	44	320	Yes
		(R) ATGGCTATTGAAGGAT			
UENF/CEPLAC 33	$(GACA)_4$	(F) ATAGAGGAAGCGCAGGTGA	55	210	Yes
		(R) ATGATTACGCCAAGCTCGAA			
UENF/CEPLAC 34	$(CAGA)_{23}$	(F) TAAAAGAGCATACA	40	310	No
		(R) ATAAGTTAGATATGTT			
UENF/CEPLAC 35	$(TCTG)_{15}$	(F) TCTGTTTAATCTGTAATT	42	290	No
		(R) TAAATCACAGGGGGCATGTT			
UENF/CEPLAC 36	$(AGAC)_{13}$	(F) ATACTAGTGAACCGCT	46	80	Yes
		(R) ATGATTACGCCAAGCTCGA			
UENF/CEPLAC 37	$(AG)_{14}(N)_6(AG)_6$	(F) AGACCGATCGACGACCA	40	220	No
		(R) TTTGTATATACACAGA			
UENF/CEPLAC 39	$(ATCT)_{15}$	(F) TTGTGAAAATTTGCCCCT	49	300	No
		(R) TGGAGAAGAACGTCTGAGCA			
UENF/CEPLAC 40	$(GATA)_{11}$	(F) TTGTGAAAATTTGCCCCT	49	150	No
		(R) TGGAGAAGAACGTCTGA			
UENF/CEPLAC 41	$(GC)_8(AC)_{10}$	(F) ATCACCTCCCTCTCAAATGCT	55	90	No
		(R) ATGATTACGCCAAGCTCGAA			
UENF/CEPLAC 42	$(AT)_5$	(F) TTGCCAAATGGTTTTATGTTGT	52	200	Yes
		(R) ACGGGCACTATCATCATCCA			
UENF/CEPLAC 43	$(GT)_4$	(F) TGCATGATGAATGAT	40	200	No
		(R) ACACGGGCACTATCA			
UENF/CEPLAC 44	$(TG)_{13}$	(F) TAGGGGAATCATCCCATCA	54	210	Yes
		(R) ACATGGACAACCCGAAAGT			
UENF/CEPLAC 45	$(AC)_{14}$	(F) ATGGCTCAACAGAGGTCTT	51	200	Yes
		(R) AAGCTCGAAATTCACCCT			
UENF/CEPLAC 46	$(TA)_6$	(F) TCCCTACTCGGCCACTAAT	55	ND	ND
		(R) TGGGAACAAGCCCTTGATAA			
UENF/CEPLAC 47	$(GA)_4$	(F) ACATGGGTGGACT	40	290	Yes
		(R) TGGTTTTAGCAACT			
UENF/CEPLAC 48	$(TATG)_{11}(TATC)_{21}$	(F)TTGCTAAAACCAAAGACCTGT	54	350	No
		(R) ATTACGCCAAGCTCGAAACT			
UENF/CEPLAC 49	$(GTT)_4$	(F) TTGGCGTAATCATGGTCATA	40	200	Yes
		(R) TCATTTAGTTTTCTTA			
UENF/CEPLAC 51	$(TACA)_8$	(F) ATACCTGTTACCTGTGTGAAA	52	300	No
		(R) TGGGTGGACTCCAAGAA			
UENF/CEPLAC 52	$(GATA)_{12}$	(F) ATAATGTCTCTAGGTCCTCCTGA	52	200	Yes
		(R) ATGACCATGATTACGCCAA			
UENF/CEPLAC 53	$(TC)_4$	(F) TCCCGTTTTTGTCTTTTCCT	53	ND	ND
		(R) ACGGAAGGCGAAGTGATTA			

Table 1 continued

Locus	Repeat motif	Primer sequence (5′→ 3′)[a]	T_a(°C)[b]	Allele size (bp)	Polymorphic[c]
UENF/CEPLAC 54	$(AG)_{26}$	(F) ATTTCCTTGTTATTTTA	40	300	Yes
		(R) ACTCACTCTCCGCTGTTCCT			
UENF/CEPLAC 55	$(AT)_7(AC)_4$	(F) TCTGTTTGTGTGTTTCT	45	300	Yes
		(R) AGGGGCACAAGGGGGT			
UENF/CEPLAC 56	$(CT)_{26}$	(F) TGGGTGAATGAACTAAAT	40	ND	ND
		(R) ACAGAAAAGAGATAT			
UENF/CEPLAC 57	$(TG)_7(N)_{12}(TG)_4$	(F) TCCATCATTCCCACAT	42	100	No
		(R) AAACCCGGGGAAAA			
UENF/CEPLAC 58	$(CA)_{15}$	(F) AGTAACTGGATTGTATGTA	44	190	No
		(R) AGAAAATCCAACCAGA			
UENF/CEPLAC 59	$(AG)_{20}$	(F) ATGTAATTTTGTAGGTAT	41	300	Yes
		(R) TTTGATTTGAGAGAGA			
UENF/CEPLAC 61	$(CA)_5$	(F) ACTCTGCACATGCACTCA	47	300	No
		(R) AATTTCACACAGGAAACA			
UENF/CEPLAC 62	$(CAA)_5$	(F) TATTGCAGACTCCTCT	47	200	No
		(R) TCAGTGAAAGGTGATCTTTTT			
UENF/CEPLAC 63	$(CAA)_{17}$	(F) AAGAACAGCAACAACAACAA	40	280	No
		(R) AATTCATATCCTCCAA			
UENF/CEPLAC 64	$(CAG)_4(N)_{33}(CAA)_6$	(F) TATTGCACACTCCT	40	180	No
		(R) TACTGCTGTGGCTGTGA			
UENF/CEPLAC 65	$(CAA)_5$	(F) AACAGCAGCAGCAGCAAA	53	130	No
		(R) TGATGGTGGAGGTGGAGA			
UENF/CEPLAC 66	$(CAA)_{20}$	(F) ACGACGACAACAACAAGA	50	220	Yes
		(R) ATGGGCGGAATCCATAT			
UENF/CEPLAC 67	$(CAA)_{10}$	(F) ACAGCAACAACGCCGACTA	50	220	Yes
		(R) ATCCTCCAAAGGCAGA			
UENF/CEPLAC 68	$(AGC)_7(AAC)_5$	(F) AGAGGCATGCTCAGCAGTA	40	400	No
		(R) AATTCATATCCTCCA			
UENF/CEPLAC 69	$(CAA)_{10}$	(F) ACGACGACAACAACAAGA	51	210	Yes
		(R) AGGCTGAGGTGGTGCTACA			
UENF/CEPLAC 70	$(CAA)_9$	(F) ACGACGACAACAACAAGA	41	300	Yes
		(R) ATGGAACTGCTTAATA			
UENF/CEPLAC 71	$(CAA)_{12}$	(F) ACAGCAGCAGCAACAACAA	52	250	Yes
		(R) TGACCATGATAAACGCCAA			
UENF/CEPLAC 72	$(CCA)_7$	(F) ACAACAACAAGAACAGCA	41	220	Yes
		(R) AATTCATATCCTCCAA			
UENF/CEPLAC 73	$(CT)_4$	(F) ACGGGTCCTTTTCTT	45	100	Yes
		(R) TCGGCTAAGTAGACGG			
UENF/CEPLAC 75	$(TTG)_5$	(F) ATATCCTCCAAAGGCA	46	160	Yes
		(R) AACAGCACTGCTCTGAACAA			
UENF/CEPLAC 76	$(TTG)_{18}$	(F) AGCATTGCCGTCGTCA	46	170	Yes
		(R) ATGGCTAAGAACAGCA			
UENF/CEPLAC 78	$(CAA)_7$	(F) AGCCACAGCAGCAT	40	300	Yes
		(R) ATGAGATACGCCAA			

Table 1 continued

Locus	Repeat motif	Primer sequence (5′→ 3′)[a]	T_a(°C)[b]	Allele size (bp)	Polymorphic[c]
UENF/CEPLAC 79	$(TTG)_{12}$	(F) ATGGCTGCTGCTGCA	50	180	Yes
		(R) ACAGCAATTGCAGGGACA			
UENF/CEPLAC 80	$(GTT)_5$	(F) TGCTGTGGCTGTGACTGT	53	150	No
		(R) TGGCAGACTCCTCTAAGT			
UENF/CEPLAC 81	$(CAA)_9(N)_9(CAA)_7$	(F) AACAGCAACAACGCCGACGA	41	210	Yes
		(R) AATTCATATCCTCCAA			
UENF/CEPLAC 82	$(CAA)_7$	(F) ACAGCACTGAGCCCAAA	50	200	No
		(R) ACGCCAAGTTGGAAATTAA			
UENF/CEPLAC 83	$(CTG)_4$	(F) TCACGTGAAGAACTGGCTGT	49	180	Yes
		(R) AATGAAGTGCCGTCAAAT			
UENF/CEPLAC 84	$(CAA)_5$	(F) TGAAACATTCTCATCAACAACAA	53	300	No
		(R) ACCGTCAAAGGGAGGGAA			
UENF/CEPLAC 85	$(TTG)_{11}$	(F) TCTTGTCCATCACTTT	44	100	Yes
		(R) TGCCACCAACACAAACAA			
UENF/CEPLAC 86	$(TG)_{10}$	(F) TAGTCGCGGAGATCAGAA	49	250	No
		(R) TCGAAACACACCCTCA			
UENF/CEPLAC 87	$(TG)_{14}$	(F) AATATCAAAGAAACAT	40	190	Yes
		(R) AACAAACAAAGCAGAT			
UENF/CEPLAC 88	$(GACA)_{13}$	(F) ATAGAGGAAGCGCAGGTGA	46	130	No
		(R) ATGATTACGCCAAGCT			
UENF/CEPLAC 89	$(TG)_{11}$	(F) AGCTTGGAGGCAGACAT	52	300	No
		(R) ACAGCTCATGAGCCCGA			
UENF/CEPLAC 90	$(TTG)_4$	(F) AGACATACAGGGATTCAA	40	180	No
		(R) TCAGAAAGGGAACA			
UENF/CEPLAC 91	$(AC)_{14}(AT)_6$	(F) TTAACAATGCCAACATGCA	50	250	No
		(R) ATGACCATGATTACGCCAA			
UENF/CEPLAC 92	$(TG)_{15}$	(F) TAGCCTCCACCAAAAGCAT	40	ND	ND
		(R) AGGGTTCAAGA			
UENF/CEPLAC 94	$(TG)_{16}$	(F) TGGGAAATAACTTAGA	40	100	Yes
		(R) AGGAGAACTAAA			
UENF/CEPLAC 95	$(CAA)_5$	(F) AGAGAAAACATGGACAAA	47	190	No
		(R) AGGGGAGATTGCTGAA			
UENF/CEPLAC 96	$(AGG)_4$	(F) TCAAGAGGCATAGCAGGA	42	ND	ND
		(R) ATGGCTTTTTCCCTT			
UENF/CEPLAC 97	$(CAA)_5$	(F) TTCAGTGGCAGTCACAGGT	51	210	Yes
		(R) TGAAGAAGATGCTGCTGT			
UENF/CEPLAC 99	$(AG)_{37}$	(F) AAAATTTTTGTTTCCA	40	530	No
		(R) AAGAGAGGCCCCTGAGA			
UENF/CEPLAC 100	$(CG)_4$	(F) GGCGGTCTGATTTT	45	480	No
		(R) ATCGGTGGGATAGACGAT			
UENF/CEPLAC 101	$(TG)_{10}$	(F) ACTGACCAGGGTGGACTA	47	350	No
		(R) TAGAGGTCGACGGTA			
UENF/CEPLAC 102	$(CAA)_5$	(F) AACAGCAGCAGCA	40	120	Yes
		(R) AGCATTGCCGTCGTCA			

Table 1 continued

Locus	Repeat motif	Primer sequence (5′→ 3′)[a]	T_a(°C)[b]	Allele size (bp)	Polymorphic[c]
UENF/CEPLAC 103	$(CAA)_6$	(F) ATGCTAATGACAAGA	40	200	No
		(R) ACAGCTATGACCATGA			
UENF/CEPLAC 104	$(GACA)_6$	(F) TCATCACCATCACCCTGACT	54	500	No
		(R) AGGTCGACGGTATCGATAA			
UENF/CEPLAC 105	$(AG)_4$	(F) TCATGGTCATACCTGTT	48	480	No
		(R) AGGCCTACTTTCTCTGGGT			
UENF/CEPLAC 106	$(TC)_4$	(F) TCTCCCTTCTCTCTAAAA	49	520	No
		(R) AGCGATTCGCCGAGTA			
UENF/CEPLAC 107	$(GT)_4$	(F) TGTATAATGATTGCAA	39	170	Yes
		(R) ATTCCCCCACATCA			
UENF/CEPLAC 108	$(CA)_{10}$	(F) AATCACAACACACCACACAGA	53	160	No
		(R) AATTTCACACCAGGAGACAT			
UENF/CEPLAC 109	$(AT)_5$	(F) ATTCATGCACTATT	40	570	No
		(R) TGTCGCCATCATCA			
UENF/CEPLAC 110	$(CAA)_4(N)_{12}(CAA)_5$	(F) ACGACAACAACAAGAA	44	520	Yes
		(R) TGGTGCTACAGCAGA			
UENF/CEPLAC 112	$(GCT)_7(GTT)_5$	(F) TGCAGTTGTGGCT	40	500	Yes
		(R) AGCAGCAGCAGCAAA			
UENF/CEPLAC 113	$(CAA)_{11}$	(F) AGCTGTTCTCCCACA	47	200	No
		(R) AGCAAAGTCGTCGGTCT			
UENF/CEPLAC 114	$(CAA)_{21}$	(F) AACAATAACAGCAGCAGCA	51	ND	ND
		(R) TGGACGGGAATCGATAAA			
UENF/CEPLAC 115	$(CAA)_{18}$	(F) TAAGAACAGCAACAACGCCGACGA	40	150	Yes
		(R) ATTCGATGTGACA			
UENF/CEPLAC 117	$(TTG)_6$	(F) TCTTGTCCATCACT	40	100	No
		(R) ACCAATGCTACCAA			
UENF/CEPLAC 118	$(CAGA)_{18}$	(F) TGGGTCTGCATAGAGGA	41	210	No
		(R) ATTAACCCTCACTAAA			
UENF/CEPLAC 121	$(CAA)_{10}(N)_9(CAA)_7$	(F) ACGACGACAACAACAAGA	49	220	Yes
		(R) ATCCTCCAAAGGCAGA			
UENF/CEPLAC 122	$(CAA)_{11}$	(F) TCCACCTCCAGACACCAAT	55	170	Yes
		(R) ACAGGAAACAGCTATGACCA			
UENF/CEPLAC 123	$(TTG)_4$	(F) TCTACATACGGAACACGCT	41	160	No
		(R) AATGCTAATGACAAGA			
UENF/CEPLAC 124	$(ACAG)_{22}$	(F) TGGCGCCTGACATACA	40	200	No
		(R) TAGACAGACAGACA			
UENF/CEPLAC 125	$(TC)_{19}$	(F) ACCTTTCTCACACTA	42	190	Yes
		(R) TCCCCACTACAAGTAGA			
UENF/CEPLAC 126	$(TG)_{12}$	(F) ATGTTAGCTTCCACCAAA	40	480	No
		(R) TGTGTCCCCAGTT			
UENF/CEPLAC 127	$(AT)_6$	(F) TCCTCGGCTTATGCACA	40	280	No
		(R) ATGGATTTTCAACA			
UENF/CEPLAC 128	$(CAT)_5(N)_{11}(CAT)_5$	(F) AACTAACATGGAAACGCAT	43	210	No
		(R) ATCGATAAGCTTGATAT			

Table 1 continued

Locus	Repeat motif	Primer sequence (5′→ 3′)[a]	T_a(°C)[b]	Allele size (bp)	Polymorphic[c]
UENF/CEPLAC 129	$(AT)_4(N)_5(CA)_{19}$	(F) TCCAATCAAGATAAGCA	42	200	Yws
		(R) TAGGGAGTGCACAT			
UENF/CEPLAC 131	$(CAT)_5$	(F) AGCGAAATACAAGGCA	46	270	Yes
		(R) ATGATGATGGGATGGTGT			
UENF/CEPLAC 132	$(AC)_{17}$	(F) ACATTAACCAACCAACCA	40	200	Yes
		(R) AAGAATTAGCTTTGT			
UENF/CEPLAC 133	$(TTG)_{11}$	(F) AGTCTTCCACGCAAGTTTTGA	51	190	No
		(R) AGCAGCCACAGTCACA			
UENF/CEPLAC 134	$(TTG)_6$	(F) AAAGGCAGAGGCTGA	47	180	No
		(R) ACAACAATGTCAACAGCAGT			
UENF/CEPLAC 135	$(CAA)_4$	(F) TGCTCTGAACAACACTGCT	54	150	No
		(R) AGGCTGAGGTGGTGCTACA			
UENF/CEPLAC 136	$(GT)_{20}$	(F) TTTTCTGTGCTGAAGAGAGGTCCT	55	500	No
		(R) AGGGTCTAGAGCTTACCATT			
UENF/CEPLAC 137	$(AC)_{18}$	(F) ACGCTACTCCCCCAA	40	140	Yes
		(R) AGGAGGATAGAAA			
UENF/CEPLAC 138	$(GT)_{13}$	(F) TTCAATGGTGTGCAAT	42	390	No
		(R) ATTCGCCGCAATATT			
UENF/CEPLAC 140	$(AG)_{31}$	(F) ACATTCTCTCTCACTA	44	320	No
		(R) AGTTCAAGGGTGTGATTT			
UENF/CEPLAC 141	$(CAGA)_{17}$	(F) TGGGTCTGCATAGAGGA	41	360	No
		(R) TCGATAAGCTTGATAT			
UENF/CEPLAC 142	$(GCT)_6$	(F) TGTTGATGAAGAAG	40	330	No
		(R) ACAGCACCAGCAGCAAA			
UENF/CEPLAC 144	$(CAA)_9$	(F) TGAGAATCATAGAGAGCCAT	49	200	Yes
		(R) TAGTGGTGGTGGCATA			

[a] F, Forward; R, Reverse

[b] T_a, Annealing temperature

[c] The polymorphic date were based in five cocoa accessions, one of *Theobroma grandflorum* and one of *Herranea sp.* ND, Not determined

Acknowledgements The authors wish to thank the funding from the International Cocoa Organization (ICCO) and the Common Fund for Commodities (CFC) through project CFC/ICCO/BIOMOL/FUNPAB and the collaboration of Valéria Marques, Verônica S. Lima and Vanessa F. Leite.

References

Charters YM, Wilkinson MJ (2000) The use of self-pollinated progenies as ''in-groups'' for the genetic characterization of cocoa germoplasm. Theor Appl Genet 100:160–166

Faleiro FG, Pires JL, Monteiro WR, Lopes UV, Yamada MM, Piedra AG, Moura AD, Gardini EA, Marques JRB, Gramacho KP, Faleiro ASG, Santos MCM (2004) Variability in cacao accessions from the Brazilian, Ecuadorian and Peruvian Amazons based on molecular markers. Crop Breed Appl Biotechnol 4:227–233

Faleiro FG, Araújo IS, Bahia RCS, Santos RF, Yamada MM, Ahnert D (2002) Otimização da extração e amplificação de DNA de *Theobroma cacao* L. visando a obtenção de marcadores RAPD. Agrotropica 14:31–34

Figueira AVO, Cascardo JCM (2001) Marcadores moleculares no melhoramento. In: Dias LAS (ed) Melhoramento Genético do Cacaueiro, 1st edn. Funape, Viçosa, pp 385–438

Lanaud C, Risterucci AM, Pieretti I, Falque M, Bouet A, Lagoda PJL (1999) Isolation and characterization of microsattelites in *Theobroma cacao* L. Mol Ecol 8:2141–2143

Merdinoglu D, Butterlin G, Bevilacqua L, Chiquet V, Adam-Blondon AF, Decroocq S (2005) Development and characterization of a large set of microsatellite markers in

grapevine (*Vitis vinifera* L.) suitable for multiplex PCR. Mol Breed 15:349–366

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

Saunders JA, Mischke S, Leamy EA, Hemeida AA (2004) Selection of international molecular standards for DNA fingerprint of *Theobroma cacao*. Theor Appl Genet 110:41–47