GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Research Note

New microsatellite markers for bananas (*Musa* spp)

E.P. Amorim[1], P.H. Silva[2], C.F. Ferreira[1], V.B.O. Amorim[1], V.J. Santos[3], A.D. Vilarinhos[1], C.M.R. Santos[4], M.T. Souza Júnior[5] and R.N.G. Miller[6]

[1]Embrapa Mandioca e Fruticultura, Cruz das Almas, BA, Brasil
[2]Universidade Federal de Lavras, Lavras, MG, Brasil
[3]Universidade Federal do Recôncavo da Bahia, Cruz das Almas, BA, Brasil
[4]Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, Brasil
[5]Embrapa Agroenergia, Brasília, DF, Brasil
[6]Departamento de Biologia Celular, Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, DF, Brasil

Corresponding author: E.P. Amorim
E-mail: edson@cnpmf.embrapa.br

Genet. Mol. Res. 11 (2): 1093-1098 (2012)
Received September 12, 2011
Accepted January 16, 2012
Published April 27, 2012
DOI http://dx.doi.org/10.4238/2012.April.27.8

ABSTRACT. Thirty-four microsatellite markers (SSRs) were identified in EST and BAC clones from *Musa acuminata burmannicoides* var. Calcutta 4 and validated in 22 *Musa* genotypes from the Banana Germplasm Bank of Embrapa-CNPMF, which includes wild and improved diploids. The number of alleles per locus ranged from 2 to 14. The markers were considered highly informative based on their polymorphism information content values; more than 50% were above 0.5. These SSRs will be useful for banana breeding programs, for studies of genetic diversity, germplasm characterization and selection, development of saturated genetic linkage maps, and marker assisted selection.

Key words: Genetic breeding; Microsatellites; Primer validation; *Musa* spp

INTRODUCTION

Bananas (*Musa* spp) are grown in more than 120 tropical and subtropical countries, mainly by small farmers. In 2010, the total harvested area and fruit production were approximately 4.7 million hectares and 102 million tons, respectively (FAO, 2012). Such numbers indicate the importance of bananas as a strong commodity, playing key economic and social roles in many countries producing bananas worldwide.

Banana breeding programs are currently focused on the introgression of diverse traits that range from disease resistance/tolerance to yield and fruit quality (Heslop-Harrison and Shawarzacher, 2007). Molecular genetic studies are of fundamental importance for increasing our knowledge base and resources for accelerated genetic improvement of the crop, by allowing the analysis of genetic diversity, construction of genetic linkage maps, and quantitative trait locus mapping of markers for alleles linked to resistance to biotic and abiotic stresses of particular importance. The availability of reliable and robust molecular markers such as simple sequence repeat (SSRs) might contribute to the acceleration of *Musa* breeding programs.

Although many *Musa* SSRs have been developed thus far by using a number of different strategies (Jarret et al., 1994; Crouch et al., 1997; Lagoda et al., 1998; Creste et al., 2003, 2006; Wang et al., 2008; Miller et al., 2010), the number of SSR loci available for genetic analysis is still limited, with many more validated and polymorphic SSRs required (Creste et al., 2004; Wang et al., 2009).

MATERIAL AND METHODS

A total of 52 SSR primer pairs were developed in this study from *Musa acuminata* expressed sequence tag and bacterial artificial chromosome clone consensus sequence data available in the DATA *Musa* database (http://bioinformatics.cenargen.embrapa.br/musagene/tiki-index.php?page=DATAMUSA) (Togawa et al., 2010). For this, sequences were analyzed using an in-house bioinformatics pipeline that detects microsatellite loci on the basis of a PERL script that scans for imperfect repeat SSRs using the scan-for-matches program (Dsouza et al., 1997). A maximum of 5 primer sets per SSR locus were designed using Primer3 (Rozen and Skaletsky, 2000). Optimization of polymerase chain reaction (PCR) amplification conditions was performed using the Fast PCR professional 5.2 software (http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm).

Microsatellites were then optimized and validated using *M. acuminata* accessions ranging from wild to improved diploids. For this, plantlets of 11 wild banana diploids (Burmanica, Calcutta-4, Microcarpa, Berlin, Lidi, Khai Nai On, Nyarma Yik, Jari Buaya, Raja Uter, Sowmuk, and Tjau Lagada) and 11 improved diploids (017041-01, SH32-63, 042079-06, 013018-01, 003023-03, 013004-06, 001016-01, 091079-03, F2P2, 003115-Plant-1, and 15-Plant-2) from the *Musa* germplasm bank maintained at Embrapa Cassava and Fruits Research Centre (www.cnpmf.embrapa.br) were selected. These 22 diploids were chosen on the basis of agronomical characteristics and disease resistance.

Genomic DNA was extracted from young leaves of the wild and improved banana diploids by using the cetyltrimethylammonium bromide (CTAB) method proposed by Doyle

and Doyle (1990). PCRs contained 30 ng DNA, 1.5 mM $MgCl_2$, 100 µM each dNTPs, 0.2 µM each primer, and 0.75 U *Taq* polymerase in 10X buffer (Biosystems) in a final reaction volume of 15 µL. Amplifications were carried out using a BIORAD MyCycler-Thermal Cycler with the following amplification program: 94°C for 3 min, followed by 30 cycles of 94°C for 40 s, 55°C for 40 s (annealing temperature modified according to the primer used), and 72°C for 1 min. A final polymerase extension step at 72°C for 4 min was also included, followed by a final incubation at 10°C. PCR optimization included identification of the optimal $MgCl_2$ concentrations and annealing temperatures by using 3 different protocols. Amplification products were electrophoresed on 6% denaturing polyacrylamide gels and silver stained according to the protocol described by Creste et al. (2001). Allele size was estimated using a 50-bp ladder molecular size standard (Invitrogen).

RESULTS AND DISCUSSION

SSR markers combine several features of an ideal genetic marker, owing to their abundance and widespread dispersal in genomes, hypervariability, co-dominant nature, accessibility, reliability, and ease of interpretation (Gaudeul et al., 2004). Therefore, assessment of *Musa* genetic variability solely by using SSR markers continues to be widely used (e.g., Amorim et al., 2008; Jesus et al., 2009).

Of the 52 primer pairs tested in this study, 1 was monomorphic (data not shown), 3 did not yield any PCR product, and 14 required adjustment in PCR conditions. Thirty-four SSR primer combinations produced polymorphic DNA amplification products and were validated as being useful markers for genetic studies in *Musa* spp (Table 1).

The number of alleles per primer varied from 2 to 14, with an average number of 5.73. Markers CNPMF-1, -9, -15, -21, -24, -25, and -56 presented the lowest number of alleles (2) and marker CNMPF-45 yielded the highest number of alleles (14). This CNPMF-series of SSR markers was classified according to their repeat motifs. The microsatellites were compared with the NCBI public database (http://www.ncbi.nih.gov/BLAST/) using the BLASTx-Swissprot option. A cut-off E-value of $<10^{-5}$ was used to define similar orthologs and sequences that did not meet this requirement, which were annotated as unknown. The Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium were accessed by the exact test, and the polymorphism information content (PIC) was calculated using the POWERMARKER software package (Liu and Muse, 2005). The expected heterozygosity (H_E) and observed heterozygosity (H_O) were determined using the GENETIX software (Version 4.03) (Belkhir et al., 1999). H_E varied among loci from 0.0868 to 0.8988, with an average of 0.6526, and H_O ranged from 0.0909 to 0.9474. Thirty-eight percent of the loci showed a significant deviation from the HWE. The mean PIC values ranged from 0.083 to 0.8946, with ~70% showing a PIC value above 0.5000.

These highly informative SSR markers might contribute to the acceleration of banana breeding programs worldwide, improving studies of genetic diversity and leading to the development of saturated genetic linkage maps. This would ultimately enable marker-assisted selection and improve selection efficiency.

Table 1. Primer sequences identified in seven EST libraries (leaves with temperature stress and infected with *Mycosphaerella fijiensis*: early and late periods, male flower, roots and green peel) and 5 Calcutta 4 BAC consensus datasets (Vilarinhos et al. 2003).

Name	Genbank accession No.	Locus name	Primer sequence (5'-3')	Repeat motif	N_A	bp	Ta	H_E	H_O	PIC	LD (P value)	HWE (5% Prob.)	BLASTx-Swissprot	E-value
CNPMF-1	MaAES_ALL0897	MabCNPMFEst-1	F: TGATGCATTGGATGATCTCG R: AAAACACACCAACTCCATCCC	$(TGG)_7$	2	336	55	0.0868	0.0909	0.0830	1.0000	1.0000	Pectinesterase inhibitor	1e-05
CNPMF-4	MaAES_ALL1578	MabCNPMFEst-4	F: TCTGTACGTCTGCCTTGCAC R: TGCAGGAGGTGGATCCATAG	$(GT)_7$	4	166	54	0.6927	0.8095	0.6305	0.0010	0.6200	Unknown	-
CNPMF-5	MaAES_ALL0329	MabCNPMFEst-5	F: CCCTGACAGATCCTTTGTGG R: GGAGACTTCCACCTTTTCCG	$(GAA)_8$	9	205	54	0.8316	0.5909	0.8143	0.0080	0.0000*	Unknown	-
CNPMF-8	MaAES_ALL0656	MabCNPMFEst-8	F: ATCGAGGAATTTGGGAGAGG R: ATCCACAATCCGATCAGCTC	$(GAA)_6$	3	166	55	0.5248	0.363	0.4660	0.3950	0.0000*	Unknown	-
CNPMF-9	MaAES_ALL0697	MabCNPMFEst-9	F: CCTTCATCATCATCACGGC R: ACCACGACCTCCTCCTCTTC	$(GAA)_6$	2	244	54	0.4965	0.7500	0.4080	0.5390	0.3150	Unknown	-
CNPMF-10	MaAES_ALL1591	MabCNPMFEst-10	F: CACATCACACGCTCTGCTTC R: TTTTTCGGCTGATCCAATTC	$(CCA)_7$	6	288	54	0.6322	0.7727	0.5371	0.0770	0.7750	Unkown	-
CNPMF-12	MaAES_ALL1148	MabCNPMFEst-12	F: CAAAGTTTGAAAGGGAGGGG R: CTCGGACCACTAGCTTCCTG	$(CCA)_6$	3	189	58	0.4845	0.6364	0.4362	0.0500	0.2320	psa2 preprotein	3e-22
CNPMF-13	MaAES_ALL1541	MabCNPMFEst-13	F: GGGATGGCGCACTTCTTC R: AATCCGGGTTGTAAGGAACC	$(CAA)_6$	5	331	54	0.6725	0.8636	0.6222	0.1980	0.0020*	Unkwon	-
CNPMF-14	MaAES_ALL0052	MabCNPMFEst-14	F: CATCGAGGATGCACATCAAG R: CCAAAAGAGCCACGATTCAG	$(CT)_{13}$	9	236	54	0.8213	0.8636	0.8000	0.0060	0.0630	Ferredoxin-dependent glutamate synthase	8e-57
CNPMF-15	MaAES_ALL0088	MabCNPMFEst-15	F: TGCTGCGATTCTACTCATCG R: TGCCCTCCTTGTCCTGTATC	$(CT)_6$	2	302	56	0.2604	0.3077	0.2265	0.9280	1.000	Ubiquitin	2e-101
CNPMF-16	MaAES_ALL0125	MabCNPMFEst-16	F: TGTGTGACTACTCCCGGTTTC R: GTCTGCTGCTCTATCCCGAG	$(CT)_5$	3	216	56	0.5800	0.2667	0.4887	0.0490	0.0080*	Unkown	-
CNPMF-19	MaAES_ALL0167	MabCNPMFEst-19	F: GTGTTCGAGAGCTTTCAGCC R: AGAACAATCAAGCCAGCAGC	$(CT)_8$	7	196	54	0.6622	0.6364	0.6247	0.0580	0.0040*	Oxygen-evolving enhancer protein 1	8e-40
CNPMF-20	MaAES_ALL0191	MabCNPMFEst-20	F: CCTCGCACATCAACCCTTAC R: CATGATCACCATTTCCTCCC	$(CT)_7$	8	223	56	0.8199	0.9474	0.7973	0.3890	0.2280	Heat shock protein binding	3e-16
CNPMF-21	MaAES_ALL0303	MabCNPMFEst-21	F: TGAACTCTTGCTACCCCAGC R: TTAGTGGCTTCTGTCCCAGG	$(CT)_6$	2	249	55	0.5455	0.4545	0.1948	1.0000	1.0000	Probable alanine aminotransferase	1e-31
CNPMF-24	MaAES_ALL0772	MabCNPMFEst-24	F: AGGGCACAAAGCGCTCAAG R: CAATGAACGCATCACAGTCC	$(CT)_9$	2	239	56	0.4959	0.2727	0.3750	0.6730	0.0400*	GTP binding/ phospholipase activator	2e-99
CNPMF-25	MaAES_ALL0808	MabCNPMFEst-25	F: CGGAGGATGTTGTTCTCGTC R: CACGGGCTGTATTTGGTAGG	$(CT)_{12}$	2	314	56	0.5111	0.5333	0.2688	0.3130	0.4880	Genomic DNA, chromosome	1e-13
CNPMF-26	MaAES_ALL0849	MabCNPMFEst-26	F: TGGAGATGAAGAAGATCGCC R: TCATCAAGTGCGTTGCATTC	$(CT)_7$	6	348	55	0.7376	0.5909	0.7339	0.4270	0.0410*	Heat shock cognate 70-kDa protein	5e-139
CNPMF-29	MaAES_ALL1151	MabCNPMFEst-29	F: TCGTCGTAGCATTTGCCTTC R: ACGCACGAGTAAGTTGTCCC	$(CT)_{13}$	7	184	58	0.7190	0.6818	0.6732	0.3520	0.1930	Unkown	-

Continued on next page

Table 1. Continued.

Name	Genbank accession No.	Locus name	Primer sequence (5'-3')	Repeat motif	N_A	bp	Ta	H_E	H_O	PIC	LD (P value)	HWE (5% Prob.)	BLASTx-Swissprot	E-value
CNPMF-30	MaAES_ALL1241	MabCNPMFEst-30	F: TTCCTTTCTCTCCTCCCACC R: TAGGGTTTTCAGATCGACGC	$(CT)_9$	8	188	54	0.7608	0.8571	0.7444	0.6340	0.8110	Unknown	-
CNPMF-31	MaAES_ALL1342	MabCNPMFEst-31	F: AGCGGAAGAGGGTAGAGAGC R: ATCTTCTGCTGGTTCATGGC	$(CT)_{10}$	5	276	56	0.6178	0.2000	0.5422	0.5660	0.0000*	Unkown	-
CNPMF-32	MaAES_ALL1387	MabCNPMFEst-32	F: AGGCTTCGACCACAAACTCC R: AGCGTTCTCGTTCCAATCAC	$(CT)_9$	6	158	56	0.6105	0.5909	0.5813	0.6430	0.1780	NOI-like protein	8e-13
CNPMF-37	MA4_BAC0 08L021_110000	MabCNPMFBac-37	F: GAGCCGTGGCTGTCACTAAG R: TATACTCTCGATCACCGGGC	$(GAA)_6$	5	305	54	0.5850	0.6000	0.5441	0.3670	0.2420	TT12 (transparent/testa12); antiporter/solute: hydrogen antiporter	1e-28
CNPMF-38	MICRO_CT_ 0020A43_10000	MabCNPMFBac-38	F: TCGCAAGAATCTCACCTTCC R: TGGTCTTCAGGTTCCGTTTC	$(CT)_{13}$	10	237	56	0.7362	0.5500	0.7023	0.0290	0.0000*	ABC transporter A family	2e-54
CNPMF-39	BAC111B014 _130000	MabCNPMFBac-39	F: TCAGCAGACAATGCAAGAGC R: GCAGTCCAACTGGCCTTATG	$(CT)_{13}$	5	342	55	0.5568	0.6818	0.4900	0.7410	0.0110*	Unknown	-
CNPMF-41	MICRO_CT_0058 FROM_110000	MabCNPMFBac-41	F: GCTGCTCTCGCTTGTTATCC R: GCTGCTCTCGCTTGTTATCC	$(CT)_{24}$	6	349	56	0.7722	0.7692	0.7386	0.2650	0.4140	SNL2 (SIN3-like 2)	4e-49
CNPMF-43	MICRO_AG_0002 -10-A39_10000	MabCNPMFBac-43	F: AAACCCTCCACCAACACCTC R: GTTTGGTGCTCATTGCTGTG	$(AG)_{12}$	6	289	55	0.7663	0.6154	0.7354		0.0000*	Alcohol dehydrogenase	1e-47
CNPMF-45	MICRO_AG_0008 A40_50000	MabCNPMFBac-45	F: GTCAATTCCTAGCGAGAGCG R: AAATTGAGCTCCACAGGGTG	$(AG)_9$	14	312	55	0.8988	0.9000	0.8946	0.2440	0.4090	TT2 DNA binding/ transcription factor	3e-12
CNPMF-49	MICRO_AG_0013 -6--20-A41_60000	MabCNPMFBac-49	F: AGAAGGGCTGGAAAGAGAGC R: AATCCTGCAATTATGGCTGC	$(GA)_{19}$	5	275	56	0.6844	0.8667	0.6282	0.4450	0.6810	DNA binding/ transcription factor	9e-18
CNPMF-52	MICRO_AG_0017 A43_20000	MabCNPMFBac-52	F: GCGTCAGGTTGTCATTTCG R: CGGCATATCTATCCACCACC	$(AG)_8$	10	278	56	0.8088	0.8000	0.7924	0.0050	0.1290	Hypersensitive-induced response AtHIR1	3e-52
CNPMF-53	MICRO_AG_0018 -9-A44_3000	MabCNPMFBac-53	F: GGAACACAAACACGATGCAG R: TTTGCACTTTGTTCAGGCAG	$(AG)_{11}$	7	231	56	0.7789	0.7979	0.7525	0.4890	0.0130*	Unknown	-
CNPMF-56	MICRO_GCT_ 0001-6--12--10-A40_60	MabCNPMFBac-56	F: AACATGCAGAGGGAGTGGTC R: ATTGCTGATGGAGATGGAGG	$(GCT)_6$	2	309	56	0.6094	0.7895	0.2024	0.0030	1.000	Protein Brevis radix-like 2; OsBRXL2	6e-40
CNPMF-57	MICRO_AG_ 0032.SEQ	MabCNPMFBac-57	F: TATCAAGCCTAATCGGCCAC R: TGCATCAAAAATTTCCAGCTC	$(AG)_7$	8	312	52	0.8079	0.5445	0.7857	1.0000	0.0000*	Soluble starch synthase 3, chloroplastic/amyloplastic	2e-21
CNPMF-60	MICRO_AG 0044.SEQ0	MabCNPMFBac-60	F: TGAAATCTGAACCCTGGTGG R: ACGCACACACACACACAATG	$(GA)_{18}$	8	243	56	0.7895	0.8947	0.7615	0.8390	0.7200	NAC1; transcription factor	2e-42
CNPMF-61	MICRO_AG_ 0044.SEQ01	MabCNPMFBac-61	F: TGATGCTCTTAACCCTTGCC R: CGGTCCGATTCAATATCGTC	$(GA)_{18}$	8	261	56	0.8306	0.8182	0.8080	0.8820	0.2120	NAC1; transcription factor	2e-42

Name, Genbank accession No., locus name, primer sequence (5'-3'), repeat motif, N_A = number of alleles; bp = base pair original cloned allele; Ta = optimum annealing temperature; H_E = expected heterozygosity; H_O = observed heterozygosity; PIC = polymorphism information content; LD = linkage disequilibrium; HWE = Hardy-Weinberg equilibrium, BLASTx-Swissprot and E-value for 34 *Musa* spp SSR loci validated in 22 banana accessions ranging from wild to improved diploids. *Significant departure from HWE.

ACKNOWLEDGMENTS

Research partially supported by the CNPq - the National Council for Scientific and Technological Development (Project #506165/2004-3) and Embrapa (Macroprograma 2 - Banana Genetic Breeding).

REFERENCES

Amorim EP, Reis RV, Santos-Serejao JA, Amorim VBO, et al. (2008). Variabilidade genética estimada entre diplóides de banana por meio de marcadores microssatélites. *Pesq. Agropec. Bras.* 43: 1045-1052.

Belkhir K, Boesa P, Goudet J, Chikhi L, et al. (1999). Genetix Lociciel Sous Windows TM pour la Genetique dés Populations Laboratoire Genome. CNRS UPR, Université de Montpellier II, Montpellier.

Creste S, Tulmann-Neto A and Figueira A (2001). Detection of single sequence repeat polymorphism in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol. Biol. Rep.* 19: 299-306.

Creste S, Tulmann-Neto A, Silva SO and Figueira A (2003). Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. *Euphytica* 132: 259-268.

Creste S, Tulmann-Neto A, Vencovsky R, Silva SO, et al. (2004). Genetic diversity of *Musa* diploid and triploid accessions from the Brazililan banana breeding program estimated by microsatellite markers. *Genet. Resour. Crop Evol.* 51: 723-733.

Creste S, Benatti TR, Orsi MR and Risterucci A-M (2006). Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. *Mol. Ecol. Notes* 6: 303-306.

Crouch JH, Crouch HK, Ortiz R and Jarret RL (1997). Microsatellite markers for molecular breeding of *Musa*. *InfoMusa* 6: 5-6.

Doyle JJ and Doyle JL (1990). Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Dsouza M, Larsen N and Overbeek R (1997). Searching for patterns in genomic data. *Trends Genet.* 13: 497-498.

FAO (2012). FAOSTAT. Available at [http://faostat.fao.org]. Accessed April 12, 2012.

Gaudeul M, Till-Bottraud I, Barjon F and Manel S (2004). Genetic diversity and differentiation in *Eryngium alpinum* L. (Apiaceae): comparison of AFLP and microsatellite markers. *Heredity* 92: 508-518.

Heslop-Harrison JS and Shawarzacher T (2007). Domestication, genomics and the future for banana. *Ann. Bot.* 100: 1073-1084.

Jarret RL, Bhat KV, Cregan P, Ortiz P, et al. (1994). Isolation of microsatellite DNA markers in *Musa*. *InfoMusa* 3: 3-4.

Jesus ON, Ferreira CF, Silva SO, Câmara TR, et al. (2009). Characterization of recommended banana cultivars using morphological and molecular descriptors. *Crop Breed. Appl. Biotechnol.* 9: 164-173.

Lagoda PJ, Noyer JL, Dambier D, Baurens FC, et al. (1998). Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Mol. Ecol.* 7: 659-663.

Liu K and Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21: 2128-2129.

Miller RN, Passos MA, Menezes NN, Souza MT Jr, et al. (2010). Characterization of novel microsatellite markers in *Musa acuminata* subsp. *burmannicoides*, var. Calcutta 4. *BMC Res. Notes* 3: 148.

Rozen S and Skaletsky H (2000). Primer3 on the www for general users and for biologist programmers. *Methods Mol. Biol.* 132: 365-386.

Vilarinhos AD, Piffanelli P, Lagoda P, Thibivilliers S, et al. (2003). Construction and characterization of a bacterial artificial chromosome library of banana (*Musa acuminata* Colla). *Theor. Appl. Genet.* 106: 1102-1106.

Wang JY, Chen YY, Liu WL and Wu YT (2008). Development and application of EST-derived SSR markers for bananas (*Musa nana* Lour.). *Yi Chuan* 30: 933-940.

Wang JY, Zheng LS, Huang BZ, Liu WL, et al. (2009). Development, characterization, and variability analysis of microsatellites from a commercial cultivar of *Musa acuminata*. *Genet. Resour. Crop Evol.* 57: 553-563.

