View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Wageningen University & Research Publications

Identification and Validation of EST-Derived Molecular Markers, TRAP and VNTRs, for Banana Research

S.A.L. Garcia[1, 4], R. Talebi[1], C.F. Ferreira[2], B.I. Vroh[3], L.V. Paiva[4], G.H.J. Kema[1] and M.T. Souza Jr.[1, 5]
[1]Plant Research International, 6708 PB, Wageningen, The Netherlands
[2]Embrapa Cassava & Tropical Fruits, Cruz das Almas, 44380-000, Bahia, Brazil
[3]IITA-Nigeria Ibadan, PMB 5320, Ibadan, Oyo State, Nigeria
[4]Universidade Federal de Lavras, Caixa Postal 3037, Lavras-MG, Brazil
[5]Embrapa LABEX Europe, 6708 PB, Wageningen, The Netherlands

Keywords: EST sequence, molecular markers VNTR and TRAP, *Musa*

Abstract

The advent of high-throughput sequencing technology has generated abundant information on DNA sequences for the genomes of many plant species. Expressed Sequence Tags (ESTs), which are unique DNA sequences derived from a cDNA library and therefore representing genes transcribed in specific tissues or at some stage of development, are one type of DNA sequences highly available today for many important crop species. Molecular markers are used for bridging DNA sequence information with particular phenotypes and are useful tools for genotyping germplasm collections and also for tagging genes involved in desirable agronomic traits. In this sense, there is always a strong demand for suitable marker techniques to better utilise existing sequence information. A transcriptome database from banana (*Musa* spp.), DATAMusa, containing 42,724 ESTs from 11 different cDNA libraries and encompassing approximately 24 Mb of DNA sequence, was used in this study for the design of primers to PCR-amplify two types of EST-derived molecular markers, Variable Nucleotide Tandem Repeat (VNTR) and Target Region Amplification Polymorphism (TRAP). These primers were then validated against a panel of 14 diploid *Musa* genotypes and produced 32 (VNTR) and 119 (TRAP) alleles. Used separately or together, both types of markers were able to discriminate *Musa* genotypes from different genome background (A or B genomes). The TRAP alleles identified were derived from only one EST, while the VNTR alleles were derived from 12 unigenes. Based on the results of this study, EST-derived markers can be an important source of polymorphism to be used in genetic diversity and gene discovery studies in banana.

INTRODUCTION

Banana (*Musa* spp.) belongs to the *Musaceae* family and has a genome of about 600 Mb distributed over 11 chromosomes. The wide range of edible bananas, including cooking types such as plantains and a wide range of dessert types, was derived mainly from intraspecific crosses of several subspecies of *Musa acuminata* (A genome) as well as from hybrids from this species with *Musa balbisiana* (B genome) (Simmonds and Shepherd, 1995).

DNA markers are found in abundance and are not influenced by the environment or plant developmental stages, making them ideal for genetic relationship studies. The development and application of technologies based on molecular markers provides powerful tools that are adequate enough to reveal polymorphisms at the DNA sequence level and are often robust to detect genetic variability within populations and between individuals (Kresovich et al., 1995; Reddy et al., 2002; Simmons et al., 2007). Many studies on genetic diversity of *Musa* have been conducted by applying different molecular markers including isozymes (Bhat et al., 1992), RFLP (Jarret et al., 1992), RAPD (Guimarães et al., 2009), AFLP (Wong et al., 2001) and SSR or VNTR (Kaemmer et al., 1997; Crouch et al., 1999; Ning et al., 2007).

Variable nucleotide tandem repeats (VNTRs) are multiallelic loci which consist of

repeated core sequences (>6 nucleotides), also known as minisatellites, tandemly repeated and flanked by segments of non-repetitive sequence, allowing the VNTR blocks to be extracted with restriction enzymes and analyzed by RFLP, or amplified by PCR and their size determined by gel electrophoresis. Like all genetic loci, VNTRs are also subject to point mutation, but slipped-strand mispairing at meiosis or during DNA replication may also occur, changing the number of repeat units. This process can generate a large number of alleles at a single minisatellite locus, each differing by one or more copies of the same repeat unit (Lai and Sun, 2003). Over the past decade, there has been a tremendous increase in the accessibility of DNA sequence data from a wide variety of taxa, including a wealth of expressed sequence tags (ESTs) that are typically unedited, automatically processed, single-pass sequences produced from cDNAs. Moreover, it has been shown that EST-based SSR or VNTR markers can be rapidly and inexpensively developed from existing EST databases (Holton et al., 2002; Thiel et al., 2003). Thus, the use of such databases for marker development appears to be a promising alternative to the development of traditional "anonymous" SSRs and VNTRs following standard methods.

Target region amplification polymorphism (TRAP) is a new type of molecular marker based on a simple and rapid PCR-based marker system. It was developed using EST information and a bio-informatics approach to generate polymorphic markers around targeted candidate gene sequences. TRAPs are amplified by one fixed primer designed from a known EST and by a second primer of arbitrary sequence with either an AT- or GC-rich core to anneal with an intron or exon, respectively. The arbitrary primers are 5'end-labelled with an infrared dye (IR-700 or IR-800) for self-detection of the amplified fragments with the Global DNA Sequencer (Li-Cor Biosciences) (Hu and Vick, 2003). The TRAP technique was initially developed with sunflower (*Helianthus annus* L.) DNA samples (Hu et al., 2003), but has also proved reliable in other plant species such as lettuce (*Lactuca sativa* L.), sugarcane (*Saccharum* spp. L.) and common bean (*Phaseolus vulgaris* L.) (Hu et al., 2005; Alwala et al., 2006; Miklas et al., 2006).

A banana transcriptome database, DATAMusa (Souza Jr. et al., 2005), containing 42,724 ESTs from 11 different cDNA libraries and encompassing approximately 24 Mb of DNA sequence, was used in this study for the design of primers for PCR amplification of VNTR and TRAP markers. We assessed the feasibility of these markers in a panel of 14 diploid banana varieties, representing both the A and B genomes of *Musa* spp. This is the first study to show the possibility of using EST sequences for generating VNTR markers and the application of the TRAP technique in *Musa* genotypes.

MATERIALS AND METHODS

Plant Material and DNA Isolation

A panel of 14 diploid accessions from the *Musa* germplasm collection at the International Institute of Tropical Agriculture (IITA) was used in this study; representing both the A and the B genome (Table 1). Two grams of the cigar leaf were ground to a powder in liquid nitrogen and used to extract DNA using an SDS method (Dellaporta et al., 1983). Due to the high content of polyphenols in banana leaves, the following modifications were introduced: extraction buffer was supplemented with 2% PVP, and 4 µL of beta-mercaptoethanol (Sigma_Aldrich, St. Louis, USA) was added to each sample following addition of extraction buffer. The DNA was precipitated by the isopropanol procedure.

VNTR Identification, Primer Design and PCR Amplification

Sequence data obtained from the DATAMusa database were submitted to analysis in a bio-informatics pipeline for VNTR detection. This pipeline consisted of a primer development module which splices genomic sequences into 1500-bp size fragments with a 100-bp overlap and then detects VNTRs using a Perl script which allows imperfect repeat VNTR detection using the scan-for-matches programme (Dsouza et al., 1997). Finally, a maximum of five primer sets were generated using Primer3 software (Rozen

and Skaletsky, 2000) for specific PCR amplification of the VNTR. Additional optimisation of the primer pairs was performed using the Fast PCR Professional 5.2 software (http://www.biocenter.helsinki.fi/bi/programs/fastpcr.htm). All primers were synthesised by Sigma-Aldrich (U.S.A). Each PCR reaction was performed in a 25-µl total volume, containing 50 ng of template genomic DNA, 2 mM $MgCl_2$, 600 µM dNTPs, 5 µM of each primer and 0.4 U of Taq-DNA polymerase (Roche, Mannhein, Germany). Temperature cycling was conducted with the following program: 94°C for 2 min, 13 cycles of 94°C for 30 s, 66°C for 30 s (-1°C per cycle) and 72°C for 30 s, followed by 28 cycles of 94°C for 30 s, 53°C for 30 s and 72°C for 30 s and a final elongation period of 7 min at 72°C. PCR products were separated by electrophoresis using 3.0% agarose gels containing 0.3 µg/ml ethidium bromide in 0.5 x TBE buffer at 120 V, for approximately 5 h. Amplification products were visualised using a UV transilluminator and photographed using an Eagle Eye II still video system.

TRAP Primer Design and PCR Amplification

TRAP marker analysis was conducted as described by Hu and Vick (2003). A single fixed primer was used in the PCR reactions in combination with two arbitrary primers, each labelled with IR-700 or IR-800. Three fixed primers were designed against the EST MUC4FL1018_H03 selected from the DATAMusa database: *Musa*.1 sequence forward (5'-gggaggccttgatctcgt-3'), *Musa*.2 sequence forward (5'-gctggccagatggtgaac-3') and *Musa*.3 sequence forward (5'-gaggaggaccgcagatga-3'). The fixed primers were selected by using the program "Primer 3" (Rozen and Skaletsky, 2000), with the following parameters: primer optimum, maximum and minimum Tm at 53, 55 and 50°C, respectively. For the development of the arbitrary primers, the general principles of PCR primer design were upheld so as to avoid self-complementarities and improper GC content (40-60%). Three principals were considered in the construction of each arbitrary primer: 1) the selective nucleotides were 3 to 4 nucleotides at the 3'end, 2) the "core" consisted of 4 to 6 nucleotides with AT- or GC-rich regions, and 3) a filler sequence making up the 5'end. In addition, the arbitrary primers were 3'end-labelled with IR-700 or IR-800 for self-detection of the amplified fragment with the Li-Cor Global DNA Sequencer (Li-Cor Bioscienses, Lincoln, NE) (Hu and Vick, 2003). A total of four arbitrary primers were used in this study: T03 sequence reverse (5'-cgtagcgcgtcaattatg-3'), 5'end-labeled with IR-700; T04 sequence reverse (5'-cgtagtgatcgaattctg-3'), 5'end-labeled with IR-700; T13 sequence reverse (5'-gcgcgatgataaattatc-3'), 5'end-labeled with IR-800; and T14 sequence reverse (5'-gtcgtacgtagaattcct-3'), 5'end-labeled with IR-800. For the Li-Cor Genotyper system, PCR was conducted with a final reaction volume of 10 µl in 96-well microtiter plates using the PTC-200 Peltier Thermal Cycle (MJ Research, Waltham, MA) with the following components: 50 ng of DNA sample, 1.0 µl of 10X reaction buffer, 1.0 µl of 25 mM $MgCl_2$, 0.1 µl of 10 mg/ml BSA, 0.4 µl of 5 mM dNTP, 2 pmol each of IR-700 and IR-800 dye-labelled arbitrary primers, 10 pmol of the fixed primer and 0.5 units of Taq DNA polimerase (SuperTaq). Temperature cycling was conducted with the following program: 94°C for 2 min, 1 cycle of 94°C for 45 s, 35°C for 45 s and 72°C for 1 min, followed by 35 cycles of 94°C for 45 s, 35°C for 45 s and 72°C for 1 min, plus an extra elongation period of 7 min at 72°C. Afterwards, the reaction products were mixed with an equal volume (10 µl) of formamide-loading buffer (98% formamide, 10 mM EDTA pH8.0 and 0.1% Bromophenol Blue). The total mixture was carefully vortexed and heated for 5 min at 94°C in the denaturation hotblock and then quickly cooled in ice. A 0.5-µl aliquot was loaded onto a 6.5% denaturing polyacrylamide sequencing gel in a Li-Cor Global DNA Genotyper. The sequencing gel was prepared using protocols recommended by the manufacturer (Li-Cor Biosciences). Electrophoresis was conducted at 1500 V for 3.5 h and images were collected using SAGA software (Li-Cor Bioscienses).

Scoring and Analysing the Amplified Fragments

The polymorphic fragments generated by the VNTR and TRAP markers were visually scored from the printed images. The scoring codes were (1) for present, (0) for absent and (9) for missing data. Jaccard's coefficient was used to calculate pairwise genetic dissimilarity matrices using the software GENES (Cruz, 2001). The genetic dissimilarity matrices were then used to construct dendrograms via UPGMA (Unweighted Pair Group Method with Arithmetic mean) employing the STATISTICA software (Statistica for Windows, 2002). In order to evaluate the reliability of the clusters formed, the dataset was also submitted to bootstrapping with 300, 1000 and 100 simulations for the VNTR, TRAP and VNTR/TRAP data, respectively.

RESULTS AND DISCUSSION

Microsatellite markers, developed from genomic libraries, can belong to either the transcribed region or the nontranscribed region of a genome, and rarely there is information available regarding their functions. Their frequency is higher in transcribed regions, especially in the untranslated portions and particularly common in the 5'-untranslated region (UTR) and, to a lesser extent, in the 3' UTR (Morgante et al., 2002). The development of SSRs from the transcribed region has become a very important tool for genetic relationship studies. However, it has been suggested that the frequency of SSR-containing sequences in plant-derived EST databases is typically on the order of 2-5% (Kantety et al., 2002). EST-SSRs are likely to be useful across a much broader taxonomic range than the anonymous SSR because the former come exclusively from transcribed regions of the genome (Pashley et al., 2006). EST-SSR markers can contribute to 'direct allele selection', if they are shown to be completely associated or even responsible for a targeted trait (Sorrells and Wilson, 1997). Putative functions for a significant proportion of EST-SSR markers have been reported (Gao et al., 2004; Han, 2004).

A total of 13 EST-VNTR loci were characterised using a panel of 14 accessions, and nine (69%) generated amplicons that were polymorphic among these accessions (Table 2). Two loci (15%), *Musa*_Cachaco_O324TF and MUC4PE1021_A03_.b_017, detected more than three amplicons in the same genotype, and other two (15%), *Musa*_Cachaco_N189TFB and MACVLIMFES026C14C8, did not generate polymorphic amplicons. Other studies in *Musa* showed that 21% (Ning et al., 2007) and 95% (Creste et al., 2006) of the SSR loci selected from genomic libraries were polymorphic. In addition, several studies in other crop plants have shown that EST-SSR loci were less polymorphic compared to genomic SSRs because of greater DNA sequence conservation in transcribed regions (Scott, 2000; Gao, 2003). The average number of alleles amplified per VNTR primer was 3.55, ranging from two to six, with a total of 32 alleles identified. Some studies using SSR from genomic library have obtained higher values, such as 4.4 (Creste et al., 2006), 7.5 (Amorim et al., 2009) and 9.2 (Ning et al., 2007).

Polymorphism Information Content (PIC), a measure of allelic diversity at a locus, was estimated for each polymorphic EST-VNTR loci detected (Table 2). In comparison to genomic SSRs, SSR-ESTs revealed low polymorphic information content value in germplasm characterisation and genetic diversity studies (Scott, 2000; Eujayl, 2001; Thiel et al., 2003). However, in the present work, the PIC values ranged from 0.43 (*Musa*_Cachaco_ET53TF) to 0.93 (*Musa*_Cachaco_FN22TF), and the mean PIC value estimated across all the polymorphic VNTR loci was 0.76. This result is similar to other studies that used genomic SSRs (Creste et al., 2006; Amorim et al., 2009).

The number of alleles per genotype ranged from 7 to 13 for BB accessions, and from 10 to 14 for AA accessions. 'Borneo' (*M. acuminata* ssp. *microcarpa*) presented the highest number of alleles (14); 'Singapuri' (*M. balbisiana*) the lowest (7). The *Musa*_Cachaco_FN22TF locus presented the highest number of alleles (6), while the loci *Musa*_Cachaco_EB55TF and *Musa*_Cachaco_ET53TF, presented the lowest number (2).

The dendrogram presented in Fig. 1 indicates two main clusters, which correspond exactly to the genome designation of the AA and BB accessions, showing that VNTR

markers were able to separate the A genome from the B genome. The bootstrap values were low due to the small number of analysed bands (32), but were sufficient to provide valuable information regarding the constitution of these genomes.

The same accessions were used in the validation of TRAP markers. From 12 TRAP PCR reactions that employed three fixed primers in combination with four random primers, all primer combinations (100%) generated polymorphic bands and a total of 119 alleles (Table 3). This means that the EST MUC4FL1018_H03 sequence alone provided 119 different markers identified in 14 *Musa* accessions. The average number of alleles amplified per primer combination was 9.91, ranging from 5 to 16. The best primer combination was *Musa* SBP.1 x T03, providing the highest PIC value (0.83). The number of alleles per genotype ranged from 42 to 74 for BB accessions, and from 33 to 53 for AA accessions. 'Butohan' (*M. balbisiana*) presented the highest number of alleles (74); 'Calcutta 4' (*M. acuminata* ssp. *burmannicoides*) the lowest (33).

The PIC value was also analysed and ranged from 0.66 to 0.83, with the mean PIC value estimated across all the polymorphic TRAP loci being 0.77. In Figure 2, a fraction of the IR-700 and IR-800 images generated by the Li-Cor system is shown. The excellent quality of the pictures provided by the Li-Cor DNA Sequencer makes the scoring of the amplified fragments faster, easier and more reliable than VNTR. Moreover, the TRAP fragments can be detected using other amplified fragment detection systems, such as the SRAP technique (Quiros, 2001).

Approximately 3.5-12.5% of the amplified fragments, generated by the TRAP technique, were polymorphic. This percentage was influenced by intensity of the amplified fragments that could be classified in strong, intermediate or weak. In order to provide more reliable results, only fragments with strong and intermediate band intensities were analysed.

The dendrogram presented in Figure 3 indicates two main cluster groups with nodes supported by bootstrap values of 99.8 and 100%. These correspond exactly to the genome designation of the AA and BB accessions, showing that TRAP markers were also able to separate the A genome from the B genome. The bootstrap values were high due to the large number of analysed bands.

Several studies have applied the TRAP technique to different purposes. In one, which used this technique for detecting quantitative trait loci in wheat, a total of 357 markers and 24 markers were obtained per reaction using 15 fixed primers and four arbitrary primers, respectively. In addition, this study revealed that the TRAP markers were highly efficient for genetic mapping in wheat (Liu et al., 2005). In another study, this technique was used to genotype a panel of 53 lettuce cultivars, using ten fixed and four arbitrary primers. A total of 388 TRAP markers were encountered, which showed the power of this technology to fingerprint lettuce (Hu et al., 2005). An additional study used this technique for developing chromosome-specific TRAP markers in wheat. The combination of ten fixed primers and two arbitrary primers generated 307 markers, and the authors considered that the chromosome-specific markers developed provided an identity for each chromosome, facilitating the molecular and genetic characterisation of the individual chromosome, including genetic mapping and gene identification (Li et al., 2007). Miklas et al. (2006) used the TRAP technique for mapping and tagging disease resistance traits in common bean. Seventy lines were genotyped using seven fixed primers and 11 arbitrary primers, and 85 TRAP markers were generated. This study revealed that the TRAP markers had great potential for mapping regions of common bean, linked to disease resistance, and as observed in earlier studies, the TRAP technique detected numerous polymorphic markers that were reproducible and heritable as either dominant or codominant markers. The TRAP results presented in this work are in agreement with the results cited above, demonstrating that the TRAP technique is applicable to diverse applications in *Musa* research. Additionally, the technique proved to be a quick, reliable and efficient way to examine genetic variability between *Musa* genotypes.

Finally, the dendrogram presented in Figure 4 represents the data from combined analysis of VNTR/TRAP markers. Two main cluster groups were formed, with 99 and

100% bootstrap support. These groups also correspond exactly to the genome designation of the AA and BB accessions. The Cophenetic correlation value (Mantel, 1967), calculated for the VNTR and TRAP markers together, was r=0.57, at 5% probability. This value indicates greater reliability of our results, showing that both matrices were consistent in being able to differentiate between AA and BB accessions.

The dendrograms shown in Figures 1 and 3 illustrate a clear distinction between the A and B genomes, although subgroups varied for both. Probably, such differences could be due to the template sequences used for primer design. VNTR primers were designed using EST sequences from several cDNA libraries of different banana accessions, while TRAP-fixed primers were designed using only one EST sequence from only one genotype. Nevertheless, the main focus of this work was not to discuss the similarity or differences among these accessions, but to show that both techniques (VNTR and TRAP) can be very useful for identification of polymorphism to be used with different purposes in banana research.

There are some advantages for using expressed gene sequences as compared to anonymous sequences for identification and use of genetic markers (Thiel et al., 2003). First, if an EST marker is associated to a trait of interest, it may be possible that this could be the gene affecting the trait directly. Therefore, EST-derived markers can provide opportunities for gene discovery and enhance the role of genetic markers by assaying variation in transcribed and known function genes. Second, EST-derived markers are likely to be more highly conserved and therefore may be more transferable between species than anonymous sequence-derived markers. Third, ESTs that share homology with candidate genes can be specifically targeted for genetic mapping and can be useful for aligning genome linkage across distantly related species for comparative analysis (Holton et al., 2002).

Literature Cited

Alwala, S., Suman, A., Arro, J.A., Veremis, J.C., Kimbeng, C.A. 2006. Target region amplification polymorphism (TRAP) for assessing genetic diversity in sugarcane germplasm collections. Crop Science 46:448-455.

Amorim, E.P.L., Ledo, C.A.D., Amorim, V.B.D., Reis, R.V. dos, Santos-Serejo J.A. and Silva, S.D.E. 2009. Agronomic and molecular characterization of diploid improved banana genotypes. Revista Brasileira de Fruticultura 31:154-161.

Bhat, K.V., Bhat, S.R. and Chandel, K.P.S. 1992. Survey of isozyme polymorphisms for clonal identification in *Musa*. HortScience 67:501-507.

Creste, S., Benatti, R.T., Orsi, M.R., Risterucci, A.M. and Figueira, A. 2006. Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. Molecular Ecology Notes 6:303-306.

Crouch, H.K., Jarret, R.L., Tenkouano, A. and Ortiz, R. 1999. VNTR-based diversity analysis of 2x and 4x full-sib *Musa* hybrids. Electronic Journal of Biotechnology 2:130-139.

Cruz, C.D. 2001. Programa Genes: versão Windows. Viçosa, MG: UFV. Software.

Dellaporta, S.L., Wood, J. and Hicks, J.B. 1983. A plant DNA minipreparation: version II. Plant Molecular Biology Report 1:19-21.

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

Eujayl, I. 2001. Assessment of genotypic variation among cultivated durum wheat based on EST-SSRs and genomic SSRs. Euphytica 119:39-43.

Gao, L.F. 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches. Molecular Breeding 12:245-261.

Gao, L.F., Jing, R.L., Huo, N.X., Li, Y., Li, X.P., Zhou, R.H., Chang, X.P., Tang, J.F., Ma, Z.Y. and Jia, J.Z. 2004. One hundred and one new microsatellite loci derived from ESTs (EST-SSRs) in bread wheat. Theoretical and Applied Genetics 108:1392-1400.

Guimarães, N.C.C., Torga, P.P., Resende, E.C., Paiva, E. and Paiva, L.V. 2009.

Identificação de variantes somaclonais em bananeiras 'Prata Anã', utilizando técnicas moleculares e citogenéticas. Ciência e Agrotecnologia 33:448-454.
Han, Z.G., Guo, W.Z., Song, X.L., Zhang, T.Z. 2004. Genetic mapping of EST-derived microsatellites from the diploid *Gossypium arboreum* in allotetraploid cotton. Molecular Genetics Genomics 272:308-327.
Holton, T.A., Christopher, J.T., McClure, L., Harker, N. and Henry, R.J. 2002. Identification and mapping of polymorphic SSR markers from expressed gene sequences of barley and wheat. Molecular Breeding 9:63-71.
Hu, J.G. and Vick, B.A. 2003. Target region amplification polymorphism: a novel marker technique for plant genotyping. Plant Molecular Biology Reporter 21:289-294.
Hu, J.G., Seiler, G.S., Jan, C.C. and Vick, B.A. 2003. Assessing Genetic Variability among Sixteen Perennial *Helianthus* Species Using PCR-based TRAP Markers. Fargo: Agricultural Research Service Northern Crop Science Laboratory.
Hu, J.G., Ochoa, O.E., Truco, M.J. and Vick, B.A. 2005. Application of the TRAP technique to lettuce (*Lactuca sativa* L.) genotyping. Euphytica 144:225-235.
Jarret, R.L., Gawel, N., Whittemore, A. and Sharrock, S. 1992. RFLP-based phylogeny of *Musa* species in Papua New Guinea. Theoretical and Applied Genetics 84:579-584.
Kaemmer, D., Fischer, D., Jarret, R.L., Baurens, F.C., Grapin, A., Dambier, D., Noyer, J.L., Lanaud, C., Kahl, G. and Lagoda, P.J.L. 1997. Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. Euphytica 96:49-63.
Kresovich, S., Szewc-Mcfadden, A.K., Blick, S. and Mcferson, J.R. 1995. Abundance and characterization of simple sequence repeats (SSRs) isolated from a size fractionated genomic library of *Brassica napus* L. (rapeseed). Theoretical and Applied Genetics 106:411-422.
Kantety, R.V., Rotal, M.L., Matthews, D.E. and Sorrells, M.E. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. Plant Molecular Biology 48:501-510.
Lai, Y. and Sun, F. 2003. The relationship between microsatellites slippage rate and the number of repeats units. Molecular Biology and Evolution 20:2123-2131.
Li, G., Klindworth, D.L., Shireen, F., Cai, X., Hu, J. and Xu, S.S. 2007. Molecular characterization and chromosome-specific TRAP-marker development for Langdon durum D-genome disomic substitution lines. Genome 49:1545-1554.
Liu, Z.H., Anderson, J.A., Hu, J., Friesen, T.L., Rasmussen, J.B. and Faris, J.D. 2005. A wheat intervarietal genetic linkage map based on microsatellite and target region amplified polymorphism markers and its utility for detecting quantitative trait loci. Theoretical and Applied Genetics 111:782-794.
Mantel, N. 1967. The detection of disease clustering and generalized regression approach. Cancer Research 27:209-220.
Miklas, P.N., Hu, J., Grunwald, N.J. and Larsen, K.M. 2006. Potential application of TRAP (Targeted Region Amplified Polymorphism) markers for mapping and tagging disease resistance traits in common bean. Crop Science 46:910-916.
Morgante, M., Hanafey, M., Powel, W. 2002. Microsatellites are preferentially associated with no-repetitive DNA in plant genomes. Nature Genetics. 30:194-200.
Ning, S.P., Xu, L.B., Lu, Y., Huang, B.Z. and Ge, X.J. 2007. Genome composition and genetic diversity of *Musa* germplasm from China revealed by PCR-RFLP and SSR markers. Scientia Horticulturae 114:281-288.
Pashley, C.H., Ellis, J.R., Mccaulley, D.E., Burke, J.M. 2006. EST databases as a source of molecular markers: lessons from helianthus. Journal of Heredity.
Quiros, G.L.C.F. 2001. Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. Theoretical and Applied Genetics 103:455-461.
Reddy, P.M., Saral, N. and Siddiq, E.A. 2002. Inter simple sequence repeat (ISSR) polymorphism and its application in plant breeding. Euphytica 128:9-17.
Rozen, S. and Skaletsky, H. 2000. Primer3 on the WWW for general users and for biologist programmers. p.365-386. In: S.A. Krawetz and S. Misener (eds.),

Bioinformatics Methods and Protocols: Methods in Molecular Biology
Scott, K.D. 2000. Analysis of SSRs derived from grape ESTs. Theoretical and Applied Genetics 100:723-726.
Simmons, M.P., Zhang, L.B., Webb, C.T. and Müller, K. 2007. A penalty of using anonymous dominant markers (AFLPs, ISSRs, and RAPDs) for phylogenetic inference. Molecular Phylogenetics and Evolution 42:528-542.
Simmonds, N.W. and Shepherd, K. 1995. The taxonomy and origins of the cultivated bananas. Journal of the Linnean Society of London (Botany) 55:302-312.
Sorrells, M.E. and Wilson, W.A. 1997. Direct classification and selection of superior alleles for crop improvement. Crop Science 37:691-697.
Souza Junior, M.T., Santos, C.R., Martins, N.F., Silva, F.R., Togawa, R., Cassiano, L.A. P., Alemida, E.R.P., Coelho, M.C.F., Caetano, A.R., Ciampi, A.Y., Piffanelli, P. and Miller, R.N.G. 2005. Transcriptoma de *Musa acuminata* no DataMusa. Boletim de Pesquisa e Desenvolvimento. Embrapa Recursos Genéticos e Biotecnologia 109:1-21.
Statistica for Windows. 2002. Computer program manual. Version 6. Tulsa: Statsoft, 1 CD-ROM.
Thiel, T., Michalek, W, Varshney, R.K. and Graner, A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics 106:411-422.
Wong, C., Kiew, R., Loh, J.P., Gan, L.H., Set, O., Lee, S.K., Lum, S. and Gan, Y.Y. 2001. Genetic diversity of the wild banana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. Annals of Botany 88:1017-1025.

Tables

Table 1. Panel of *Musa* diploid accessions, representing both the A and B genome.

Accession name	Genome group/species	Subgroup/subspecies	Plant status
Pisang Lilin	AA		landrace
Pisang Mas	AA	Sucrier	landrace
Borneo	*Musa acuminata*	ssp. *microcarpa*	wild
Calcutta 4	*Musa acuminata*	ssp. *burmannicoides*	wild
Madang	*Musa acuminata*	ssp. *banksii*	wild
Selangor	*Musa acuminata*	ssp. *malaccensis*	wild
Zebrina	*Musa acuminata*	ssp. *zebrina*	wild
Butohan	*Musa balbisiana*		wild
Etikehel	*Musa balbisiana*		wild
Los Banos	*Musa balbisiana*		wild
Montpellier	*Musa balbisiana*		wild
Singapuri	*Musa balbisiana*		wild
Tani	*Musa balbisiana*		wild
Truncata BS-252-A	*Musa acuminata*	ssp. *truncata*	wild

Table 2. EST-VNTR Locus, primer sequence forward and reverse, number of alleles and Polymorphic Information Content (PIC).

EST-VNTR Locus	Repeated sequence	Sequence (5' to 3') F/R	N° of alleles	PIC
*Musa*_Cachaco_DH15TF	$(GTGTT)_4$	CCTTCGGTAGTTTGATCAGC AACAATTAGATAAGGAACTTG	3	0.69
*Musa*_Cachaco_EB55TF	$(GAACC)_4$	CAACAGGGCAAGCTGGAGGTACAG ACTGTGAAACAATGTGAGATGGGTC	2	0.7
*Musa*_Cachaco_ET53TF	$(GCAGT)_5$	GGAGGAGAGGATACATGGAA CTTGAATGATCAGCAAGCCAG	2	0.43
*Musa*_Cachaco_FN22TF	$(ATGCC)_7$	CAACAGGGCAAGCTGGAGGTACAG ACTGTGAAACAATGTGAGATGGGTC	6	0.93
*Musa*_Cachaco_0421TF	$(CGAGG)_4$	TGGAAGAGCGAGGGAGGAAGA GCCCTGGAATCGGAGTCGTGGTC	4	0.79
MUC4LH1008_F03_.b_027	$(CACTG)_4$	TGTCTATGGGAATTGAGGATGAG GTACCGATCTCAAAATTGTCAGG	5	0.89
MACVLIMFLS013C_B12_.b	$(TTTTTA)_4$	TTGGCAGGATTAAGCACAGTTTCC GAAGACGGATAATGCCTTTTGC	3	0.88
MACVLIMFLS013D_E05_b	$(GAAAA)_4$	CCTTATTCTGTCGGGAAATCCA GTTCACAAGTGTCCCATGCAACAA	3	0.83
Root_01	$(CCGACG)_4$	ACACAAATATCCTTCCCGTCC TTCTTGGTAACGGTGAGATCG	4	0.75
*Musa*_Cachaco_N189TFB	$(AGAGG)_4$	TAGCAGTGGAAGGGGCATGCA TTATGCTCACAACACACACAC	0	0
*Musa*_Cachaco_O324TF	$(AGGGTT)_4$	TCACCGCCATGGACGTCGTCTA CATCAGTGACATTGCTCATCTC	>3*	0
MACVLIMFES026C14C8	$(AGGGAG)_4$	AACCCCCCACCTTAAAAACATC ATCCCTCCAGCGTCAATCACCG	0	0
MUC4PE1021_A03_.b_017	$(GTTTTT)_5$	CCTTATTCTGTCGGGAAATCCA GTTCACAAGTGTCCCATGCAACAA	>3*	0
Total alleles			32	

*More than two bands in diploid varieties.

Table 3. Characteristics of the amplicons obtained with 12 TRAP primer pairs used to analyse the genetic diversity of *Musa* accessions.

Primer combination	N° of bands	N° of polymorphic bands	% polymorphism	PIC
Musa.1 x T03	128	9	7	0.83
Musa.2 x T03	157	14	8.91	0.82
Musa.3 x T03	144	13	9	0.78
Musa.1 x T04	150	13	8.66	0.76
Musa.2 x T04	128	16	12.5	0.79
Musa.3 x T04	147	5	3.4	0.82
Musa.1 x T13	142	5	3.52	0.66
Musa.2 x T13	188	9	4.78	0.76
Musa.3 x T13	173	6	3.46	0.73
Musa.1 x T14	100	10	10	0.83
Musa.2 x T14	172	12	7	0.71
Musa.3 x T14	163	7	4.29	0.82
Total	1792	119	6.64	9.31

Figures

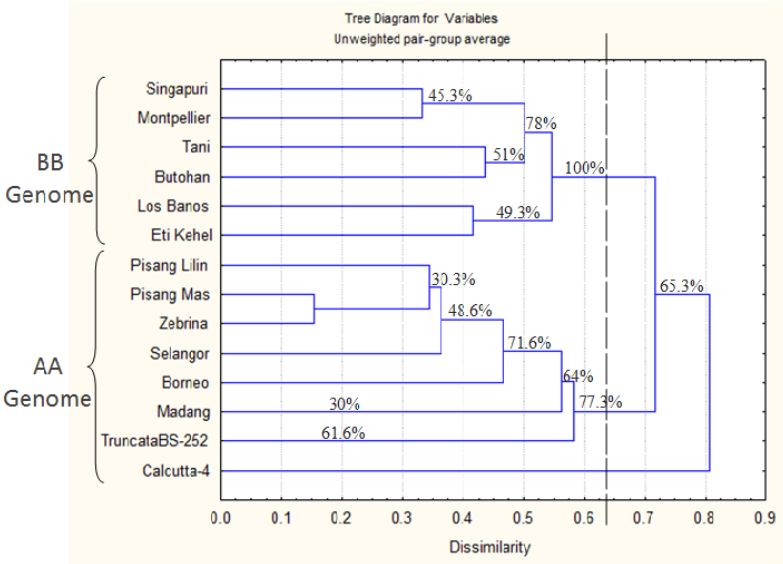

Fig. 1. Dendrogram showing dissimilarity between *Musa* accessions based on analysis with VNTR markers.

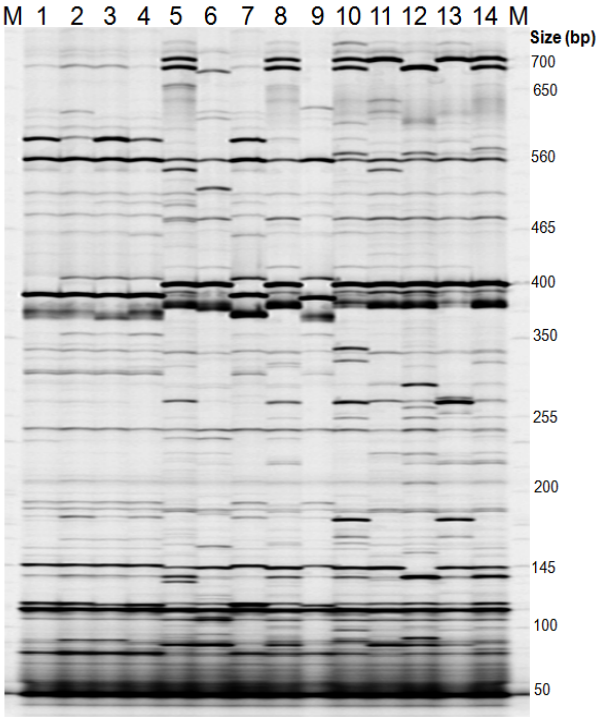

Fig. 2. A portion of Target Region Amplification Polymorphism (TRAP) profile of DNA from 14 banana accessions using the fixed primer *Musa* SBP.1 and the arbitrary primer T04. Lanes are: 1 - Singapuri, 2 - Los Banos, 3 - Tani, 4 - Butohan, 5 - Pisang Lillin, 6 - Calcutta 4, 7 - Etikehel, 8 - Truncata BS-252-A, 9 - Montpellier, 10 - Borneo, 11 - Madang, 12 - Pisang Mas, 13 - Selangor, 14 - Zebrina. *M* = 1 kb molecular DNA marker.

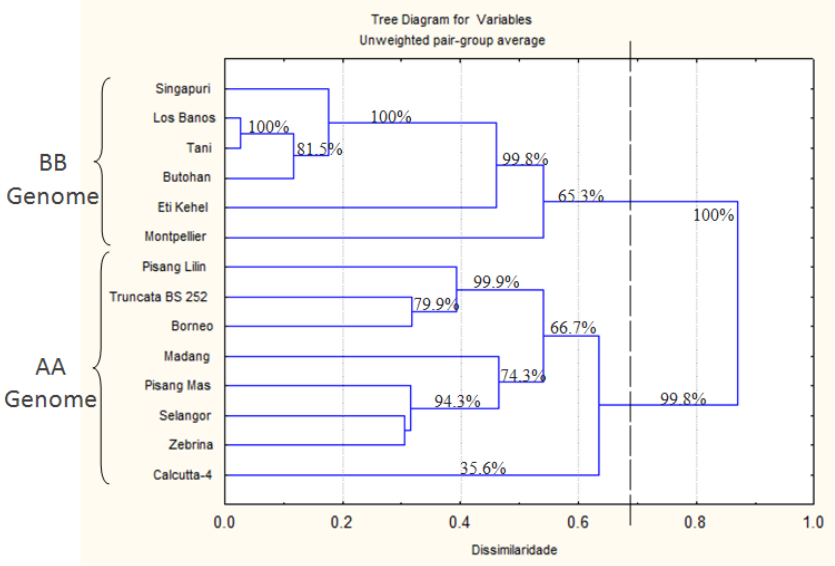

Fig. 3. Dendrogram showing dissimilarity between *Musa* accessions based on analysis with TRAP markers.

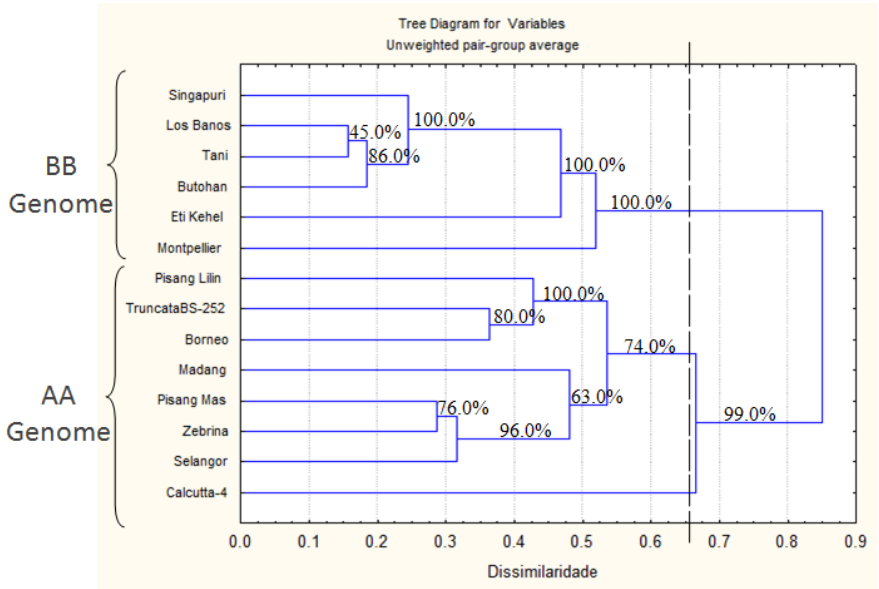

Fig. 4. Dendrogram showing dissimilarity between *Musa* accessions based on a joint analysis using VNTR and TRAP markers.