

# Genetic diversity between improved banana diploids using canonical variables and the Ward-MLM method

Valquíria Martins Pereira<sup>(1)</sup>, Cristine Vanz Borges<sup>(2)</sup>, Livia Pinto Brandão<sup>(2)</sup>, Larissa Santos Oliveira<sup>(1)</sup>, Cintia Paula Feitosa Souza<sup>(2)</sup>, Zalmar Santana Gonçalves<sup>(2)</sup>, Sebastião de Oliveira e Silva<sup>(2)</sup>, Janay Almeida dos Santos-Serejo<sup>(3)</sup>, Claudia Fortes Ferreira<sup>(3)</sup>, Edson Perito Amorim<sup>(3)</sup> and Carlos Alberto da Silva Ledo<sup>(3)</sup>

<sup>(1)</sup>Universidade Federal de Lavras, Departamento de Agricultura, Caixa Postal 3037, CEP 37200-000 Lavras, MG, Brazil. E-mail: vaumarpe@hotmail.com, larizinha.so@gmail.com <sup>(2)</sup>Universidade do Recôncavo da Bahia, CEP 44380-000 Cruz das Almas, BA, Brazil. E-mail: crisvanz@bol.com.br, liviapintobrandao@yahoo.com.br, cintiapaula\_2006@hotmail.com, zalmarufb@hotmail.com, ssilva3000@gmail.com <sup>(3)</sup>Embrapa Mandioca e Fruticultura, Rua Embrapa, s/nº, CEP 44380-000 Cruz das Almas, BA, Brazil. E-mail: janay@cnpmf.embrapa.br, claudiaf@cnpmf.embrapa.br, edson@cnpmf.embrapa.br, ledo@cnpmf.embrapa.br

**Abstract** – The objective of this work was to estimate the genetic diversity of improved banana diploids using data from quantitative analysis and from simple sequence repeats (SSR) marker, simultaneously. The experiment was carried out with 33 diploids, in an augmented block design with 30 regular treatments and three common ones. Eighteen agronomic characteristics and 20 SSR primers were used. The agronomic characteristics and the SSR were analyzed simultaneously by the Ward-MLM, cluster, and IML procedures. The Ward clustering method considered the combined matrix obtained by the Gower algorithm. The Ward-MLM procedure identified three ideal groups (G1, G2, and G3) based on pseudo-F and pseudo-t<sup>2</sup> statistics. The dendrogram showed relative similarity between the G1 genotypes, justified by genealogy. In G2, 'Calcutta 4' appears in 62% of the genealogies. Similar behavior was observed in G3, in which the 028003-01 diploid is the male parent of the 086079-10 and 042079-06 genotypes. The method with canonical variables had greater discriminatory power than Ward-MLM. Although reduced, the genetic variability available is sufficient to be used in the development of new hybrids.

**Index terms:** *Musa*, hybrids, microsatellites, morphological characteristics.

## Diversidade genética entre diploides melhorados de bananeira por meio de variáveis canônicas e do método Ward-MLM

**Resumo** – O objetivo deste trabalho foi estimar a diversidade genética entre diploides melhorados de bananeira por meio de dados quantitativos e de marcadores de sequências simples repetidas (SSR), simultaneamente. O experimento foi conduzido com 33 diploides, em blocos aumentados com 30 tratamentos regulares e três comuns. Dezoito características agrônômicas e 20 iniciadores SSR foram usados. Os dados agrônômicos e de SSR foram analisados, simultaneamente, via os procedimentos Ward-MLM, de agrupamento e IML. O método de agrupamento de Ward considerou matriz híbrida obtida pelo algoritmo de Gower. O procedimento Ward-MLM identificou três grupos (G1, G2 e G3) baseados nas estatísticas de pseudo-F and pseudo-t<sup>2</sup>. O dendrograma mostrou similaridade relativa entre os genótipos do G1, justificada pela genealogia. No G2, 'Calcutta 4' aparece em 62% das genealogias. Comportamento similar foi observado no grupo G3, em que o diploide 028003-01 é o parental masculino dos genótipos 086079-10 e 042079-06. O método com uso de variáveis canônicas teve maior poder discriminatório que o Ward-MLM. Embora reduzida, a variabilidade genética disponível é suficiente para ser usada no desenvolvimento de novos híbridos.

**Termos para indexação:** *Musa*, híbridos, microssatélites, características morfológicas.

### Introduction

Bananas are derived from the wild species *Musa acuminata* Colla (AA) and *Musa balbisiana* Colla (BB). The fruit is produced by smallholders, and millions of people in developing countries use it as staple food. According to the Food and Agriculture Organization of the United Nations (2012), the total area harvested in

2009 was approximately five million hectares, which produced 95 million tons.

Banana genetic breeding programs use diploids as the basis for developing new commercial hybrids, since triploids present sterility as a major constraint (Amorim et al., 2011). In addition, diploids have traits of agronomic importance, such as disease resistance and high fruit yield (Lessa et al., 2010). The agronomic

characterization of these diploids – including the estimation of genetic variability using molecular markers – is important when it comes to choosing the progenitors for the crosses between divergent genotypes, aiming to explore heterosis and to develop new, improved diploids, which can be used in crosses with triploid and tetraploid genotypes in order to develop new commercial banana hybrids (Amorim et al., 2008).

Simple sequence repeats (SSR) or microsatellite markers have been widely used for characterizing species, especially due to their co-dominant nature, repeatability, and easy data interpretation (Creste et al., 2003). Genetic diversity is usually estimated considering, separately, quantitative data – such as plant height and pseudostem girth – and qualitative data, including anthocyanin content, leaf position, pulp color, and molecular marker data (Cabral et al., 2010). Moreover, strategies to rank genotypes considering combined data have also been proposed using clustering methods, such as the Ward method and the unweighted pair group method with arithmetic mean (UPGMA) (Gonçalves et al., 2009).

The modified location model (MLM) procedure, proposed by Franco et al. (1998), is another interesting strategy for quantifying genetic variability using quantitative and qualitative variables simultaneously. The MLM has two stages. In the first one, the Ward clustering method is used to define the groups using the Gower dissimilarity matrix. In the second one, the average of the quantitative variable vector is estimated by the MLM, regardless of the value of the qualitative variables. This procedure has been used for a variety of crops, such as common beans (Cabral et al., 2010), tomatoes (Gonçalves et al., 2009), and bananas (Pestana et al., 2011).

The objective of this work was to estimate the genetic diversity of improved banana diploids using quantitative and simple sequence repeats (SSR) marker data simultaneously.

## Materials and Methods

Thirty-two improved diploids, developed by the banana genetic breeding program at Embrapa Mandioca e Fruticultura, and the SH3263 diploid, developed by the Fundación Hondureña de Investigación Agrícola, were used (Table 1).

The experiment was carried out in Cruz das Almas, BA, Brazil (12°40'19"S and 39°06'22"W, 220 m above sea level). The climate of the region is humid tropical, Aw to Am, according to Köppen, with an annual average temperature of 24.5°C, relative humidity of 80%, and average annual precipitation of 1,249.7 mm (Agritempo, 2008). Federer's augmented block experimental design was used (Federer, 1956) with 31 regular treatments – diploids 1 to 28, 32 and 33 repeated once in each block with replicates only in the plots – and three common treatments – diploids 29 to 31, considered as controls, repeated in the five blocks. Each plot consisted of six plants with spacing of 2.5x2.5 m.

Evaluations were made in the first production cycle and considered the following 18 agronomic traits: plant height (PH, cm); pseudostem girth (PSG, cm); number of tillers during flowering (NTF); number of leaves during flowering (NLF); number of days from bunch emission to harvest (NBH); presence of pollen grains (PPG), based on a scale in which 1 represents the absence of pollen grains, 2 a small amount of pollen grains, 3 the average amount of pollen grains, and 4 the abundance of pollen grains; number of hands per bunch (NHB); number of fruits per hand (NFH); yellow Sigatoka at emergence of flowering (SEF) and at harvest (SH), using the scale proposed by Stover (1972); number of leaves at harvest (NLH); weight of second hand (WSH, kg); length and fruit diameter (LF and FD, respectively, cm); length and diameter of pedicel (LP and DP, respectively, mm); presence of seeds (PS), evaluated with the following scale: 1 is the absence of seeds, 2 represents 1 to 10 seeds, 3 represents 11 to 20 seeds, and 4 is more than 21 seeds; and length and diameter of stalk (LS and DS, respectively, cm).

Twenty pairs of SSR primers were used for molecular characterization, with five microsatellite markers from the Ma series (Crouch et al., 1998), 12 from the AGMI series developed by Lagoda et al. (1998), and three from the MaOCEN series (Creste et al., 2006) (Table 2). DNA was extracted from young leaves using the CTAB method (Doyle & Doyle, 1990). Amplification reactions were done in a final volume of 13  $\mu\text{L}^{-1}$ , containing the following reagents: KCl 50  $\text{mmol L}^{-1}$ , Tris-HCl 10  $\text{mmol L}^{-1}$  (pH 8.3),  $\text{MgCl}_2$  2.5  $\text{mmol L}^{-1}$ , 100  $\mu\text{mol L}^{-1}$  of each of the dNTPs (dATP, dTTP, dGTP, and dCTP), 0.2  $\mu\text{mol L}^{-1}$  of each primer, 50 ng of genomic DNA, and one unit of Taq DNA polymerase.

Amplifications were performed in a Perkin Elmer 9700 thermal cycler (Perkin Elmer do Brasil Ltda., São Paulo, SP, Brazil) with the following program: an initial cycle of 3 min at 94°C, followed by 40 s at 94°C, specific annealing temperature for each primer during 40 s, 1 min at 72°C, followed by 30 cycles of 40 s at 94°C, 40 s at 45°C and 60 s at 72°C, and final

temperature of 8°C. The fragments were separated in 3% Ultrapure Agarose-1000 gels (Invitrogen, Carlsbad, CA, USA) under standard conditions, stained with ethidium bromide, visualized under ultraviolet light, and photo-documented using the UVITEC equipment.

The intrablock procedure was adopted for agronomic data analysis (Duarte, 2000). Intrablock analysis was

**Table 1.** Banana diploid hybrids used in the study.

Code	Improved diploids <sup>(1)</sup>	Year obtained	Genealogy
1	091087-01*	2002	001016-01 (Borneo x Guyod) x 003038-01 (Calcutta 4 x Heva)
2	086079-10	2001	003037-02 (Calcutta 4 x Galeo) x 028003 (Tuugia x Calcutta 4)
3	028003-01	1995	Tuugia x Calcutta 4
4	013019-01	1995	Malaccensis x Tjau Lagada
5	TH-0301	1995	Terrinha x Calcutta 4
6	086094-20	2003	03037-02 (Calcutta 4 x Galeo) x SH3263
7	073041-01	1999	Khai x 003004-01 (Calcutta 4 x Madang)
8	091079-03	2001	01016-01 (Borneo x Guyod) x 028003 (Tuugia x Calcutta 4)
9	003023-03	1995	Calcutta 4 x S/No. 2
10	058054-03	1998	003005-01 (Calcutta 4 x Pahang) x 001004-01 (Borneo x Madang)
11	042049-04	1999	M53 x M48
12	091094-04	2002	001016-01 (Borneo x Guyod) x SH3263
13	086079-12	2002	003037-02 (Calcutta 4 x Galeo) x 028003-01 (Tuugia x Calcutta 4)
14	017041-01	1995	Jari Buaya x 003004-01 (Calcutta 4 x Madang)
15	003037-02	1995	Calcutta 4 x Galeo
16	042079-13	1999	M53 x 028003-01 (Tuugia x Calcutta 4)
17	013004-06	1995	Malaccensis x Madang
18	042085-02	2002	M53 x 015003-01 (Madu x Calcutta 4)
19	042023-06	1995	M53 x No. 2 cultivar not named
20	003004-02	1995	Calcutta 4 x Madang
21	086094-15	2003	03037-02 (Calcutta 4 x Galeo) x SH3263
22	041054-08	1998	003004-01 (Calcutta 4 x Madang) x 001004-01 (Borneo x Madang)
23	042052-03	1997	M53 x Kumburgh
24	086094-05	2000	03037-02 (Calcutta 4 x Galeo) x SH3263
25	042052-04	1997	M53 x Kumburgh
26	042015-02	1997	M53 x Madu
27	042049-05	1995	M53 x M48
28	089087-01	2002	013018-01 (Malaccensis x Sinwobogi) x 003038-01 (Calcutta 4 x Heva)
29	042079-06 <sup>(2)</sup>	1997	M53 x 028003-01 (Tuugia x Calcutta 4)
30	001016-01 <sup>(2)</sup>	1995	Borneo x Guyod
31	013018-01 <sup>(2)</sup>	1995	Malaccensis x Sinwobogi
32	SH3263	-	FHIA <sup>(3)</sup> hybrid
33	013004-04	1995	Malaccensis x Madang

<sup>(1)</sup>The first three numbers refer to the female parent, the following three to the male parent, and the last two to the number of the selection. <sup>(2)</sup>Controls.

<sup>(3)</sup>FHIA, Fundación Hondureña de Investigación Agrícola.

obtained admitting a fixed model by considering the effects as constants, except for the experimental error. The analysis was carried out with the SAS software package using the procedure for general linear models (Proc GLM). The average of the treatments was adjusted by the minimum squares using the lsmeans SAS module. Standard computational procedures were carried out as proposed by Duarte (2000).

The quantitative (agronomic) and qualitative (molecular marker) data were analyzed simultaneously using the Ward-MLM procedure (Franco et al., 1998) and the cluster and interactive matrix programming (IML) command of the SAS program for cluster formation. The Ward cluster method considered the combined data matrix obtained by the Gower algorithm.

The correlation between the agronomic variables and the canonical variables was obtained graphically using the Candisc command in the SAS software.

In order to define the ideal number of groups, the procedure indicated for the MLM model, based on pseudo-F and pseudo-t<sup>2</sup> statistics, was used. Taking into account the definition of the optimal number of groups, a hierarchical classification was obtained by the Ward method, which makes available the initial

value needed to program the final step of the MLM (Franco et al., 1998).

## Results and Discussion

The SH3263 (Cod. 32) and 013004-04 (Cod. 33) diploids were not evaluated due to their bad development in the field, mainly caused by the attack of the banana weevil (*Cosmopolites sordidus*).

Significant differences were detected between diploids and controls for plant height, days from emission to harvest, and number of fruits, which are key characteristics for the selection of improved diploids (Table 3 and 4). Genotypes significantly affected the agronomic traits, except for fruit diameter. This result indicates the existence of genetic variability between the diploids and justifies the use of molecular markers and multivariate methodologies to quantify it.

The analysis using microsatellites showed 133 alleles (102 polymorphic), with an average of 6.65 alleles per primer. The highest number of alleles was identified for the AGMI33/34 and Ma1/27 primers (ten alleles), and the lowest for the AGMI35/36, Ma3/103, and MaOCEN13R/13F primers (four alleles) (Table 2).

**Table 2.** Microsatellite primers used in the study.

Name	Forward and reverse (5'-3') sequences	No. of alleles	PIC	Reference
AGMI103/104	acagaatcgctaaccctaactctca/cccttgcgtgccctaa	8	0.97	Lagoda et al. 1998
AGMI187/188	gcaactttggcagcatttt/tgatggactcatgttacactat	6	0.97	Lagoda et al. 1998
AGMI25/26	ttaaagtggttagcattagg/ttgatgtcacaatgggtgtcc	7	0.53	Lagoda et al. 1998
AGMI93/94	aacaactaggatggtaatgtgtgaa/gatctgaggatggttctgttgagtg	6	0.74	Lagoda et al. 1998
AGMI101/102	tcagttgacaaccccacaca/tgggaagaaaataagaataga	5	0.83	Lagoda et al. 1998
AGMI33/34	agtttcaccgattggtcat/taacaaggactaatcatgggt	10	0.76	Lagoda et al. 1998
AGMI105/108	tcccaaccctgcaaccact/atgacctgtgacacatcttt	5	0.66	Lagoda et al. 1998
AGMI35/36	tgaccacgagaaaagaagc/ctctccatagcctgactgc	4	0.87	Lagoda et al. 1998
AGMI125/126	tcccataagttaatcctcagtt/ctccatcccccaagtcataaag	8	0.74	Lagoda et al. 1998
AGMI129/130	ggaggcccaacataggaagaggaat/cataaacgacagtagaaatagcaac	6	0.57	Lagoda et al. 1998
AGMI95/96	acttattccccgcactcaa/actctgcccatcttcatcc	8	0.83	Lagoda et al. 1998
AGMI99/100	atttctctttcataccttta/taatgagacgctatggagcac	9	0.66	Lagoda et al. 1998
Ma2/7	tgaatcccaagttgtgcaaga/caactctgtccctcaactca	6	0.87	Crouch et al. 1998
Ma1/17	aggcggggaatcggtaga/gcggggagacagatggagt	7	0.86	Crouch et al. 1998
Ma1/24	gagccattaaagctgaaca/ccgacagtcaacatacaata	5	0.74	Crouch et al. 1998
Ma1/27	tgaatcccaagttgtgcaaga/caaacactgtccccatctc	10	0.80	Crouch et al. 1998
Ma3/103	tcgcctcttttagctctg/tgtggaggactgagattg	4	0.57	Crouch et al. 1998
MaOCEN1F/1R	tctcaggaaggcaacaatc/ggaccaagggaagaaacc	9	0.80	Creste et al. 2006
MaOCEN3F/3R	ggaggaatggaggtcaaca/ttcggatagggaggagg	6	0.64	Creste et al. 2006
MaOCEN13R/13F	gctgctattttgccttgggtg/cttgatctgggattctgg	4	0.84	Creste et al. 2006
Total		133	-	
Average		6.65	0.76	

**Table 3.** Analysis of variance for the agronomic characteristics<sup>(1)</sup> of the 31 improved banana diploids in the first production cycle.

Source of variation	df	Mean square								
		PH	PSG	NTF	NLF	NLH	PFEH	SL	SD	NFB
Blocks	4	146.65	1.32	0.43	3.82	0.25	71.86	28.90	0.64	103.41
Treatments (diploids vs. control)	1	2,364.60**	1.58**	0.59 <sup>ns</sup>	2.34 <sup>ns</sup>	0.05 <sup>ns</sup>	1,380.71*	62.16*	5.45**	750.78*
Genotypes within treatments	29	1,618.70**	11.88**	1.18*	6.88*	3.00*	645.30*	94.69**	3.57**	975.48**
Diploids	27	1,694.51**	12.46**	0.99*	6.01*	2.63*	625.82*	80.48**	3.81**	768.72**
Controls	2	595.19*	4.08*	3.72**	18.53**	7.98**	908.26*	286.45**	0.29 <sup>ns</sup>	3,766.67**
Error	8	126.41	0.69	0.35	1.33	0.65	135.35	9.79	0.26	69.01
Coefficient of variation (%)		5.80	5.79	14.88	9.21	27.05	8.30	9.27	11.49	8.47
General average		194.00	14.00	4.00	13.00	3.00	140.00	34.00	4.00	98.00
Regulars										
Average		188.08	14.54	4.08	12.39	3.02	135.84	32.78	4.75	94.87
Maximum		320.97	23.76	6.30	18.45	6.49	179.25	51.89	13.12	206.66
Minimum		133.04	9.56	2.25	7.45	0.36	49.76	12.53	2.99	51.17
Commons										
Average		203.68	14.14	3.83	12.88	2.94	147.76	35.31	4.00	103.66
Maximum		216.11	15.17	4.82	14.99	3.78	159.66	43.84	4.17	130.27
Minimum		195.69	13.51	3.20	11.22	1.49	133.13	29.41	3.73	75.45

<sup>(1)</sup>PH, plant height (cm); PSG, pseudostem girth (cm); NTF, number of tillers during flowering; NLF, number of leaves during flowering; NLH, number of leaves during harvest; PFEH, period between flower emission and harvest (days); SL, stalk length (cm); SD, stalk diameter (cm); NFB, number of fruits per bunch. <sup>ns</sup>Not significant. \*\* and \*Significant at 1 and 5% probability by the F test, respectively.

**Table 4.** Analysis of variance for the agronomic characteristics<sup>(1)</sup> of the 31 improved banana diploids in the first production cycle.

Source of variation	df	Mean square								
		NHB	WSH	NFSH	PL	PD	FL	FD	PS	PPG
Blocks	4	0.10	0.02	0.89	0.01	0.01	2.27	1.15	0.04	0.06
Treatments (diploids vs. controls)	1	0.12 <sup>ns</sup>	0.00 <sup>ns</sup>	14.17**	0.64**	0.03 <sup>ns</sup>	3.64 <sup>ns</sup>	1.15 <sup>ns</sup>	9.71**	0.00 <sup>ns</sup>
Genotypes within treatments	29	2.57**	0.17**	4.25**	0.70**	0.07*	9.05**	0.26 <sup>ns</sup>	1.87**	0.24*
Diploids	27	1.60**	0.16**	4.14**	0.73**	0.07*	8.47**	0.27 <sup>ns</sup>	1.27**	0.18*
Controls	2	15.68**	0.29**	5.74**	0.40*	0.00 <sup>ns</sup>	16.78**	0.21 <sup>ns</sup>	9.92**	1.06**
Error	8	0.13	0.01	0.47	0.04	0.02	1.60	1.35	0.04	0.06
Coefficient of variation (%)		5.78	18.67	4.38	18.15	17.92	12.71	54.49	8.62	6.93
General average		6.00	0.57	16.00	1.22	0.81	10.00	2.00	3.00	4.00
Regulars										
Average		6.39	0.57	15.23	1.32	0.83	9.74	2.00	2.19	3.72
Maximum		10.64	2.04	19.73	5.21	2.02	18.99	3.27	4.19	4.07
Minimum		4.29	0.12	12.22	0.48	0.41	6.53	0.44	0.85	3.07
Commons										
Average		6.50	0.58	16.44	1.06	0.78	10.35	2.35	3.19	3.73
Maximum		8.42	0.79	17.67	1.26	0.81	12.23	2.56	4.00	4.00
Minimum		4.92	0.31	15.74	0.73	0.73	8.57	2.14	1.56	3.20

<sup>(1)</sup>NHB, number of hands per bunch; WSH, weight of second hand (kg); NFSH, number of fruits of second hand; PL, pedicel length (mm); PD, pedicel diameter (mm); FL, fruit length (cm); FD, fruit diameter (cm); PS, presence of seeds; PPG, presence of pollen grains. <sup>ns</sup>Not significant. \*\* and \*Significant at 1 and 5% probability by the F test, respectively.

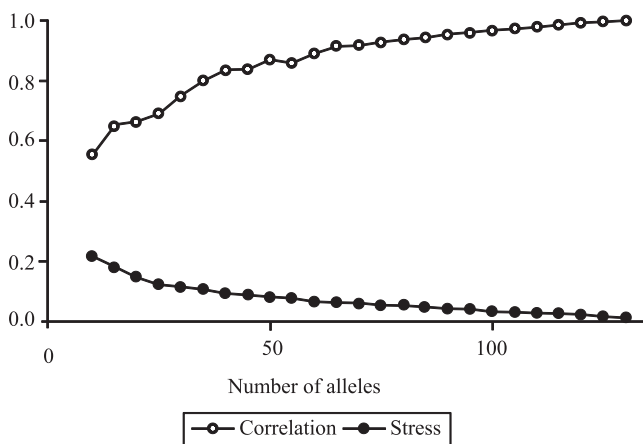
The average number of alleles per primer was similar to the ones reported in other molecular studies with bananas (Kaemmer et al., 1997; Grapin et al., 1998; Creste et al., 2003, 2004, 2006; Ning et al., 2007; Amorim et al., 2009; Mattos et al., 2010).

Polymorphic information content (PIC) varied from 0.53, for primer AGMI25/26, to 0.97, for the AGMI103/104 and AGMI187/188 primers, with an average of 0.76. Resampling analysis indicated that 85 alleles were sufficient for a precise estimation of the genetic diversity between the 31 banana diploid hybrids (Figure 1).

The correlation between the matrix with all the 133 alleles and the matrix with 85 alleles was 0.94, with sum of the deviation squares (SDq) of 0.52 and stress value (S) of 0.048. According to Kruskal (1964), an S value  $\leq 0.05$  indicates excellent precision.

In the literature, the number of SSR markers and alleles used to characterize banana genotypes is inferior to the ones used in the present study. Ning et al. (2007) considered the use of ten SSR markers sufficient to genotype 50 banana accessions from different origins, finding 92 alleles. Creste et al. (2004) used nine SSR markers to genotype 49 banana diploids, using 115 alleles, while Creste et al. (2003) genotyped 35 banana cultivars using 11 SSR markers, with 67 alleles. Mattos et al. (2010) identified 94 alleles using 13 SSRs to genotype 26 banana accessions.

The estimate of genetic variability between genotypes is a critical point for cluster and genetic diversity analyses between and within populations (Kosman &



**Figure 1.** Resampling analysis for a precise estimation of the genetic variability between the 31 improved banana diploids developed by Embrapa Mandioca e Fruticultura in the first production cycle.

Leonard, 2005). The Ward-MLM procedure, initially proposed by Franco et al. (1998), has been used in the combined analysis of multicategorical, quantitative, and molecular data. Therefore, the estimate of the correlation between the dissimilarity matrices obtained by the Gower distance (quantitative data) and the binary data (SSR) was carried out using the simple coincidence index. The correlation between the two matrices was  $r = -0.15$ . For Souza & Sorrells (1991), the low association between morphological and molecular data can be attributed to the partial and insufficient representation of the genome when morphological data is used. This low correlation can be explained by the absence of association between the loci that control the morphological characters studied and the alleles identified by the SSR markers. The average genetic dissimilarity between the diploids was 0.49, varying from 0.33, between the 042052-04 and 042052-03 diploids, to 0.59, between the 017041-01 and 091087-01 genotypes.

The dendrogram based on the combined analysis by the Ward-MLM procedure is shown in Figure 2. The cophenetic correlation value was high ( $r = 0.71$ ,  $p < 0.0001$ , 10,000 permutations) and adequate, since values of  $r \geq 0.56$  are considered ideal, reflecting good agreement with the genetic similarity values (Vaz Patto et al., 2004).

The Ward-MLM procedure indicated that the ideal number of groups was three, based on pseudo-F and pseudo- $t^2$  statistics, considering the Gower algorithm. Group 1, with seven diploids; group 2, with 21; and group 3, with three.

The dendrogram shows relative similarity between the group 1 genotypes, which is justified by genealogy, since the 001016-01 (Borneo x Guyod) genotype is the female parent of the 091087-01 and 091079-03 hybrids. The same occurred with the 013019-01 and 01318-01 diploids, which have the wild *Malaccensis* diploid as one of the ancestors.

Group 2 was formed by 21 diploids, which was an already expected result since the 'Calcutta 4' diploid is present in 62% of these genealogies as the male or female parent (Table 1 and Figure 2). This diploid is widely used in many banana breeding programs as a source of alleles for resistance to black Sigatoka, caused by *Mycosphaerella fijiensis* Morelet (Amorim et al., 2008). Another diploid, the M53, had a appearing frequency greater than 33% in the genealogies.

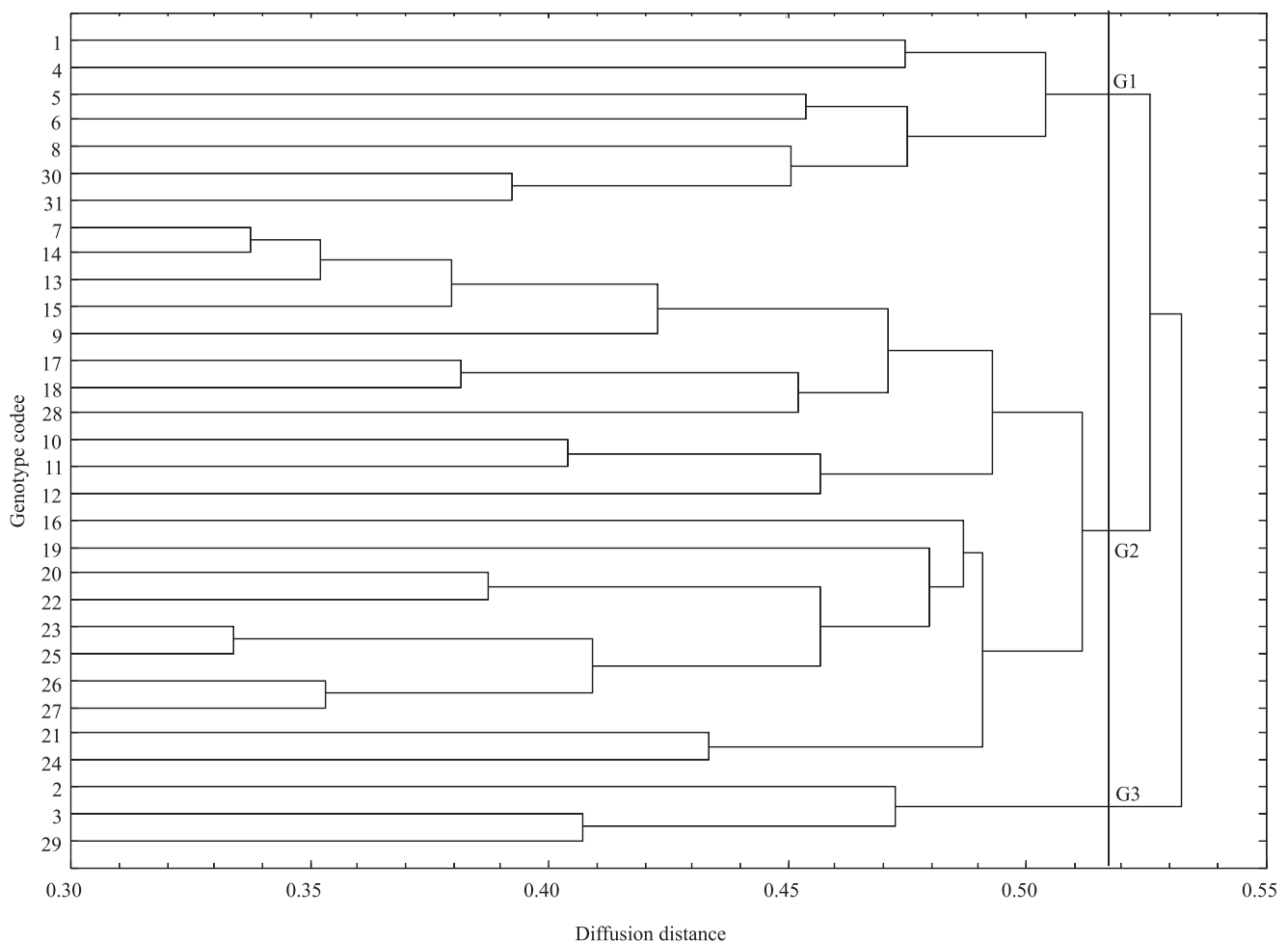
Therefore, the group formed may be associated with the small number of genitors involved in obtaining these hybrids. Similar behavior was observed in group 3, since the 028003-01 diploid is the male parent of the 086079-10 and 042079-06 genotypes.

Amorim et al. (2008) used SSR markers to quantify the genetic variability between cultivated, improved, and wild banana diploids. Cluster analysis based on the SSR polymorphism was not able to completely separate the improved, cultivated, and wild hybrids. Some diploids were grouped based on their geographic origin, such as *Musa ornata* Roxb. and IAC-1, and 'Tjau Lagada' and 'Lidi', whereas for others, no relationship was established. There was a tendency for clustering the improved diploids based on their genealogy.

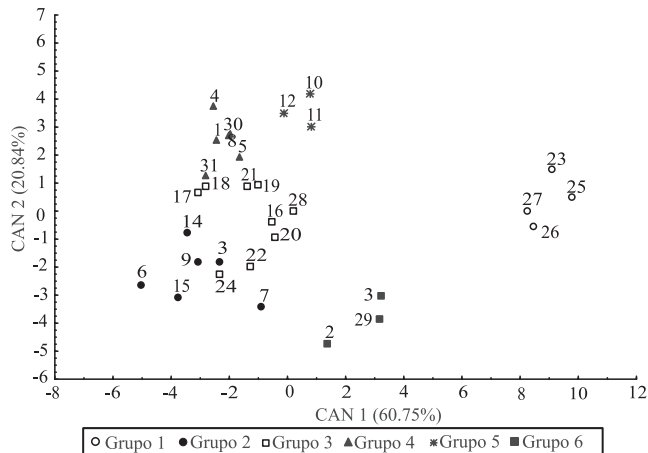
The first two canonical variables explained 81.59% of the variability between the six groups formed

(Figure 3). This value indicates that the graphic representation of the first two canonical variables was appropriate for visualizing the genetic relationship between the groups and between the accessions within the same group.

The length of stalk had the greatest correlation with the first canonical variable, followed by the number of tillers during flowering and the number of fruits in the second hand, with values of 0.70, 0.51, and -0.27, respectively. However, for the second canonical variable, the estimates of the highest correlations were found for the characteristics number of hands per bunch ( $r=0.56$ ) and number of fruits per bunch ( $r=0.46$ ). When comparing the groups formed by the Ward-MLM method (Figure 2) and the ones obtained by the canonical variables (Figure 3), the latter had a higher discriminatory power, since six groups were



**Figure 2.** Dendrogram constructed by the Ward MLM method using the genetic distances from 18 morphoagronomic characters from 31 improved banana diploids in the second production cycle. The genotype code is as shown in Table 1.



**Figure 3.** Dispersion of the first two canonical variables (CAN) with the formation of six groups (G1-G6) by the Ward-MLM strategy, considering 31 improved diploids in the first production cycle. Codes of the genotypes are as shown in Table 1.

formed in comparison to only three by the Ward-MLM method. The criteria used for separating the groups, considering the canonical variables, were associated to the genealogy of the diploids.

Currently, the banana genetic breeding program at Embrapa Mandioca e Fruticultura has 43 improved diploids with genetic resistance to most pests and diseases, including yellow and black Sigatoka and Fusarium wilt, which also present desirable agronomic characteristics, such as short stature, high yield, and drought tolerance.

## Conclusions

1. There is genetic variability between the 31 banana diploids developed by Embrapa Mandioca e Fruticultura, enabling crosses for the development of new cultivars.
2. Microsatellite markers are efficient in quantifying genetic variability.
3. Canonical variable analysis provides better clustering of the banana genotypes than the Ward-MLM method.

## Acknowledgements

To Conselho Nacional de Desenvolvimento Científico e Tecnológico, for financial support; and to the Fundação de Amparo à Pesquisa do Estado da Bahia, for scholarship granted.

## References

- AGRITEMPO. **AgriTempo**: sistema de monitoramento agrometeorológico. Disponível em: <<http://www.agritempo.gov.br/agroclima/sumario>>. Acesso em: 03 out. 2012.
- AMORIM, E.P.; AMORIM, V.B.O.; SILVA, S.O.; PILLAY, M. Quality improvement of cultivated *Musa*. In: PILLAY, M.; TENKOUANO, A. (Org.). **Banana breeding**: progress and challenges. New York: CRC Press, 2011. p.252-280.
- AMORIM, E.P.; LESSA, L.S.; LEDO, C.A. da S.; AMORIM, V.B. de O.; REIS, R.V. dos; SANTOS-seREJO, J.A. dos; SILVA, S. de O. e. Caracterização agrônômica e molecular de genótipos diploides melhorados de bananeira. **Revista Brasileira de Fruticultura**, v.31, p.154-161, 2009.
- AMORIM, E.P.; REIS, R.V. dos; SANTOS-seREJO, J.A. dos; AMORIM, V.B. de O.; SILVA, S. de O. e. Variabilidade genética estimada entre diploides de banana por meio de marcadores microsatélites. **Pesquisa Agropecuária Brasileira**, v.43, p.1045-1052, 2008.
- CABRAL, P.D.S.; SOARES, T.C.B.; GONÇALVES, L.S.A.; AMARAL JÚNIOR, A.T. do.; LIMA, A.B.P.; RODRIGUES, R.; MATTA, F. de P. Quantification of the diversity among common bean accessions using Ward-MLM strategy. **Pesquisa Agropecuária Brasileira**, v.45, p.1124-1132, 2010.
- CRESTE, S.; BENATTI, T.R.; ORSI, M.R.; RIESTERUCCI, A.M.; FIGUEIRA, A. Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. **Molecular Ecology Notes**, v.6, p.303-306, 2006.
- CRESTE, S.; TULMANN NETO, A.; SILVA, S. de O.; FIGUEIRA, A. Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. **Euphytica**, v.132, p.259-268, 2003.
- CRESTE, S.; TULMANN NETO, A.; VENCOSKY, R.; SILVA, S. de O.; FIGUEIRA, A. Genetic diversity of *Musa* diploid and triploid accessions from the Brazilian banana breeding program estimated by microsatellite markers. **Genetic Resources and Crop Evolution**, v.51, p.723-733, 2004.
- CROUCH, H.K.; CROUCH, J.H.; JARRET, R.L.; CREGAN, P.B.; ORTIZ, R. Segregation at microsatellite loci in haploid and diploid gametes of *Musa*. **Crop Science**, v.38, p.211-217, 1998.
- DOYLE, J.J.; DOYLE, J.L. Isolation of plant DNA from fresh tissue. **Focus**, v.12, p.13-15, 1990.
- DUARTE, J.B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293p. Tese (Doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.
- FEDERER, W.T. Augmented (or hoonuiaku) designs. **Hawaiian Planter's Records**, v.55, p.191-208, 1956.
- FRANCO, J.; CROSSA, J.; VILLASEÑOR, J.; TABA, S.; EBERHART, S.A. Classifying genetic resources by categorical and continuous variables. **Crop Science**, v.38, p.1688-1696, 1998.
- FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS. **FAOSTAT**. Available at: <<http://faostat.fao.org/>>. Accessed on: 15 Jan. 2012.



- GONÇALVES, L.S.A.; RODRIGUES, R.; AMARAL JÚNIOR, A.T. do; KARASAWA, M.; SUDRÉ, C.P. Heirloom tomato gene bank: assessing genetic divergence based on morphological, agronomic and molecular data using Ward-modified location model. **Genetics and Molecular Research**, v.8, p.364-374, 2009.
- GRAPIN, A.; NOYER, J.L.; CARREEL, F.; DAMBIER, D.; BAURENS, F.C.; LANAUD, C.; LAGODA, P.J.L. Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged microsatellite sites. **Electrophoresis**, v.19, p.1374-1380, 1998.
- KAEMMER, D.; FISCHER, D.; JARRET, R.L.; BAURENS, F.C.; GRAPIN, A.; DAMBIER, D.; NOYER, J.L.; LANAUD, C.; KAHL, G.; LAGODA, P.J.L. Molecular breeding in genus *Musa*: a strong case for STMS marker technology. **Euphytica**, v.96, p.49-63, 1997.
- KOSMAN, E.; LEONARD, K.J. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. **Molecular Ecology**, v.14, p.415-424, 2005.
- KRUSKAL, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. **Psychometrika**, v.29, p.1-27, 1964.
- LAGODA, P.J.L.; NOYER, J.L.; DAMBIER, D.; BAURENS, F.C.; GRAPIN, A.; LANAUD, C. Sequence tagged microsatellite site (STMS) markers in the *Musaceae*. **Molecular Ecology**, v.7, p.657-666, 1998.
- LESSA, L.S.; LEDO, C.A. da S.; SILVA, S. de. O. e; AMORIM, E.P.; OLIVEIRA, T.K. de. Características agrônômicas de híbridos diploides de bananeira em três ciclos de produção em Cruz das Almas, Bahia. **Revista Brasileira de Fruticultura**, v.32, p.213-221, 2010.
- MATTOS, L.A.; AMORIM, E.P.; AMORIM, V.B. de O.; COHEN, K. de O.; LEDO, C.A. da S.; SILVA, S. de O. e. Agronomical and molecular characterization of banana germplasm. **Pesquisa Agropecuária Brasileira**, v.45, p.146-154, 2010.
- NING, S.-P.; XU, L.-B.; LU, Y.; HUANG, B.Z.; GE, X.J. Genome composition and genetic diversity of *Musa* germplasm from China revealed by PCR-RFLP and SSR markers. **Scientia Horticulturae**, v.114, p.281-288, 2007.
- PESTANANA, R.K.N.; AMORIM, E.P.; FERREIRA, C.F.; AMORIM, V.B. de O.; OLIVEIRA, L.S.; LEDO, C.A. da S.; SILVA, S. de O. e. Agronomic and molecular characterization of gamma ray induced banana (*Musa* sp.) mutants using a multivariate statistical algorithm. **Euphytica**, v.178, p.151-158, 2011.
- SOUZA, E.; SORRELLS, M.E. Relationships among 70 North American oat germplasms. II. Cluster analysis using qualitative characters. **Crop Science**, v.31, p.605-612, 1991.
- STOVER, R.H. **Banana, plantain, and abaca diseases**. Kew: Commonwealth Mycological Institute, 1972. 316p.
- VAZ PATTO, M.C.; SATOVIC, Z.; PÊGO, S.; FEVEREIRO, P. Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. **Euphytica**, v.137, p.63-72, 2004.

---

Received on February 13, 2012 and accepted on September 3, 2012