

Molecular and morphological characterization of *Musa* spp. (Zingiberales: musaceae) cultivars*

María Angélica Buitrago-Bitar¹ Ayda Lilia Enriquez-Valencia², Jorge Mario Londoño-Cañedo³, Jaime Eduardo Muñoz-Flórez⁴, Bernardo Villegas-Estrada⁵, Gloria Esperanza Santana-Fonseca⁶

Abstract

Objectives: The overall goal was to analyze genetic diversity in cultivars of *Musa acuminata* (Colla) and *M. balbisiana* (Colla), commonly grown in farms from Caldas department. **Scope:** Characterization of the genetic variability, at the molecular and morphological level of cultivars of *M. acuminata* and *M. balbisiana*, found in farms from Caldas farmers using morphological descriptors and fluorescent microsatellites. **Methodology:** Phenotyping evaluations comprised 57 morphological characters following the descriptors proposed by IPGRI for the *Musa* genus, and for genotyping evaluations, nine fluorescent microsatellites (Simple Sequence Repeats-SSR) were used to allow the precise identification of alleles. Additionally, cluster analyses were carried out independently for both morphological and genotypic characterizations under Principal Component Analysis (PCA) and Bootstrapping methods respectively. **Main results:** Positive and negative highly significant correlations were found for the morphological descriptors, where traits such as presence/absence of male bud was the rule, as well as the diameter and perimeter of this trait, plus the diameter and perimeter of the peduncle, number of fruits, pseudostem height and fruit length contributed considerably to the variability among the cultivars allowing the discrimination of three main groups in the cluster analyzes. From the molecular perspective a total of 72 polymorphic alleles were obtained, with an average genetic diversity of 0,79, polymorphic information content (PIC) of 0,77 and heterozygosity of 0,48, showed a moderate degree of genetic differentiation ($F_{ST} = 0,061$) among *Musa* cultivars, generating three main sub-clusters based on their genetic dissimilarity. **Conclusions:** The identification of certain morphological traits showed to be suitable for the discrimination of *Musa* cultivars evaluated here. On the other hand, molecular characterization allowed to establish the genetic relationships among groups, also fluorescent SSR were highly informative and accurate, in such a way that can be considered suitable for characterizations in *Musa* varieties.

Key words: Genetic Diversity, Genotyping, Fluorescent Microsatellites, Phenotyping

* FR: 12-IX-2019. FA: 21-XI-2019.

¹ Magister. Investigadora grupo Diversidad Biológica. Universidad Nacional de Colombia. Palmira Colombia. E-mail: mabuitragob@unal.edu.co  0000-0003-3080-0669.

² Ingeniera Agrónoma. Investigadora grupo Diversidad Biológica. Universidad Nacional de Colombia. Palmira Colombia. E-mail: alenriquezv@unal.edu.co  0000-0003-1802-9613.

³ Magister. Investigador grupo Diversidad Biológica. Universidad Nacional de Colombia. Palmira Colombia. E-mail: jomlondonoca@unal.edu.co  0000-0001-9349-9725.

⁴ Doctor. Docente investigador grupo Diversidad Biológica. Universidad Nacional de Colombia. Palmira Colombia. E-mail: jemunozf@unal.edu.co  0000-0002-8237-0499.

⁵ Magister. Docente Investigador grupo InGene. Universidad de Caldas. Manizales, Caldas Colombia. E-mail: bernardo.villegas@ucaldas.edu.co  0000-0003-4441-5278.

⁶ Doctora. Docente investigador grupo Biodiversidad y Recursos fitogenéticos. Universidad de Caldas, Manizales, Caldas Colombia. E-mail: gloria.santana@ucaldas.edu.co  0000-0001-6467-3261.



CÓMO CITAR:

BUITRAGO-BITAR, M.A., ENRÍQUEZ-VALENCIA, A.L., et al., 2020.- Molecular and morphological characterization of *Musa* spp. (Zingiberales: musaceae) cultivars. *Bol. Cient. MusHist. Nat. U. de Caldas*, 24 (1): 33-47. DOI: 10.17151/bccm.2020.24.1.2



Caracterización molecular y morfológica de cultivares de *Musa* spp. (Zingiberales: musaceae)

Resumen

Objetivos: Analizar la diversidad genética de cultivares de *Musa acuminata* Colla y *Musa balbisiana* Colla (Musaceae), cultivados en el departamento de Caldas **Alcance:** Caracterización de la variabilidad genética, a nivel molecular y morfológico de cultivares de *M. acuminata* y *M. balbisiana*, encontrados en fincas de agricultores de Caldas, utilizando descriptores morfológicos y microsatélites fluorescentes. **Metodología:** Las evaluaciones fenotípicas comprendieron la evaluación de 57 caracteres morfológicos según los descriptores propuestos por el IPGRI para el género *Musa*, mientras la genotipificación comprendió el uso de nueve marcadores microsatélites fluorescentes (SSR) para la identificación correcta de los alelos amplificados. El análisis de clúster fue llevado a cabo independientemente para ambas caracterizaciones, morfológica y molecular, bajo Análisis de Componentes Principales (PCA) y métodos de probabilidad posterior (Bootstrapping) respectivamente. **Resultados principales:** Con los descriptores morfológicos se identificaron correlaciones directas e inversamente significativas, y se destacaron descriptores como tamaño de bellota masculina, diámetro del pedúnculo, número de frutos, altura de pseudotallo y longitud del fruto, para lo cual contribuyeron considerablemente a la variabilidad entre cultivares de plátano y banano permitiendo la discriminación de tres grupos principales en el análisis de clúster; desde la perspectiva molecular un total de 72 alelos polimórficos fueron obtenidos con una diversidad genética promedio de 0.79, Contenido de Información Polimórfica (PIC) de 0.77 y heterocigocidad de 0.48, mostrando un grado moderado de diferenciación genética ($F_{ST} = 0.061$) entre especies generando tres sub-grupos principales basados en su disimilaridad genética. **Conclusiones:** La identificación de ciertos rasgos morfológicos demostraron ser adecuados para la discriminación de los cultivares del género *Musa* evaluados en el presente estudio. Por otra parte, la caracterización molecular permitió establecer las relaciones genéticas dentro de los grupos, adicionalmente los SSRs evaluados presentaron alta precisión e información, siendo apropiados para su uso en caracterizaciones de variedades y cultivares del género *Musa*.

Palabras clave: Diversidad Genética, Fenotipificación, Genotipificación, Microsatélites Fluorescentes.

INTRODUCTION

Bananas (*Musa acuminata* Colla) and Plantains (*M. balbisiana* Colla) are the fourth most important food crop in developing countries and a basic food source for 400 million people worldwide. Currently, world production grew at a compound annual rate of 3,2 percent is about 114 million tons in 2017 (Agronet, 2017). Where the main producers are India, China, the Philippines, Ecuador, and Indonesia. In economic terms, this crop represent incomes of around USD\$8 billion annually. The main

exporters in Latin America and the Caribbean are Ecuador, Costa Rica and Colombia, accounting for more than 40% of production in the region from 2010 to 2016 share between 2010 and 2016, In Colombia the production is approximately 2,0 million tons annually with a yield of 24,88 tons per hectare, being the departments of Antioquia, Magdalena, and La Guajira are the main producers (Agronet, 2017).

Through genetic studies, some attempts have been made to clarify the controversy about the origin of domesticated cultivars of the *Musa* genus, for which four wild species [*M. acuminata* (AA), *M. balbisiana* (BB), *M. textilis* Neé and *M. schizocarpa* N.W. Simmonds] have been proposed as the source of genetic background, being *M. acuminata* and *M. balbisiana* the major contributors of the genomic introgressions (D'hont *et al.*, 2012). The origin of many cultivars (hybrids) resides mainly on crosses of genomes with different ploidy constitutions such as diploids (AA, AB, BB), triploids (AAA, AAB, ABB), and tetraploids (AAAA, AAAB, AABB, ABBB). This characteristic relies on the hypothesis that the hybrid cultivars have arisen through backcrossing between interspecific hybrids and parental species. This would have led to the generation of a complex spectrum of genotypes and cultivars (De Langhe *et al.*, 2010). In our view, we consider it necessary to know these variations through both morphological and genetic characterizations. These characterizations can be performed using traditional techniques based on morphological traits (phenotype), and through laboratory techniques with the application of molecular markers (genotype), being both processes valid and complementary to each other. Considering that the morphological characterization is the first step to study the genetic variability that provides key features like colors, shapes, smells, and textures, it is possible to infer about the potential uses of a particular species (Doyle, 1991). On the other hand, molecular methods provide specific genetic information based on the development and application of technologies used to identify genetic variability present in the *Musa* germplasm, offering information about of the polymorphism found at the DNA level allowing the genetic differentiation among individuals (Christelová *et al.*, 2017). Microsatellites are codominant molecular genetic markers based on different repetitive sequences present in the genomes. These genetic markers are universally scattered, possess a high level of polymorphism, and due to their small size are of easy detection (Vieira *et al.*, 2016).

Although several methods have been used for allele detection, the technology of fluorescently labeled microsatellites or Simple Sequence Repeats-SSR is an excellent alternative due to its detection accuracy (identifying polymorphisms so small like a base pair), as well as its high reproducibility with codominant characteristics allowing a big advantage with respect to other type of markers such as RAPDs (Random Amplified Polymorphic DNA) (BLAIR *et al.*, 2009; Blacket *et al.*, 2012). The genetic identification and the morphological characterizations are important steps to identify cultivars of interest with superior attributes that allow carry out next studies of

improvement and plant breeding. The goal of the present study was to characterize molecularly and morphologically *Musa* cultivars commonly grown in farms from Caldas department with the aim to identify their classification to evaluate a small panel of fluorescently labeled SSRs applicable for *Musa* cultivars identification and germplasm evaluation respect to their capability of discrimination and polymorphism detection.

MATERIALS AND METHODS

Plant material

Plant material of 12 *Musa* spp. cultivars were analyzed *in situ* in seven farms from San Jose municipality (Caldas, Colombia; Table 1). Morphological descriptions were registered from single individuals for every cultivar, from which fresh leaf tissue was harvested for downstream molecular analyses.

Table 1. Geographical Coordinates of the sampling area for the 12 *Musa* cultivars evaluated.

Coordinates	Village	Farm	Cultivar
5,081 N 75,786 W	La Paz	Sta. Teresita	Africa
5,081 N 75,786 W	La Paz	Sta. Teresita	Dominico
5,081 N 75,786 W	La Paz	Sta. Teresita	Dominico Hartón
5,089 N 75,773 W	La Paz	La Guardia	Guineo
5,081 N 75,786 W	La Paz	Sta. Teresita	Hartón
5,089 N 75,773 W	La Paz	La Guardia	Plátano Mocho
5,088 N 75,779 W	La Paz	Manga Larga	Plátano Trescientos
5,088 N 75,779 W	La Paz	Manga Larga	Urabeño
5,069 N 75,787 W	Arrayanes	El Rubí	Banano de Andes
5,087 N 75,799 W	El Contento	El Solvento	Banano Enano
5,086 N 75,798 W	El Contento	El Crédito	Banano NN
5,092 N 75,794 W	Pueblo Rico	La Huerta	Guayabo

Morphological characterization

The morphological characterization was carried out following the descriptors proposed by IPGRI (IPGRI, 1996). The trait registration was performed on plants with bunch completely developed under physiological maturity, which comprises fruits with little prominent ridges and a light-yellow tonality on the fingers of the first hand of the bunch. A total of 57 morphological variables were registered, being 38 qualitative and 19 quantitative, ordered in seven different clusters according to the organ of the plant evaluated: plant general appearance, pseudostem and sucker, leaf, male bud, bracts, male flowers and fruit.

Data analysis for morphological variables were performed on the statistical package SAS version 9,3 (2016). For quantitative variables, analysis of variance, correlation, principal components and clustering were carried out, whereas for the qualitative descriptors a principal component analysis (PCA) was employed to select the most informative variables, also correlation and clustering analyses were built based on these traits. These analyses allow estimating the parameters of each variable that contributes to the clustering definition and consequently the identification of the cultivars with similar characteristics (Doyle, 1991).

Molecular characterization

DNA extraction

Genomic DNA was extracted from young leaf tissue using the protocol for DNA extraction proposed by Doyle (1991). To verify the quality and concentration of the DNA, measurements on a spectrophotometer (Colibri Titertek Berthold) and visualization on agarose gel 1% (stained with GelRed® Biotium) were performed. Finally, DNA dilutions were performed to normalized to 20ng/μl for PCR amplification reactions.

Fluorescent microsatellites amplification

PCR amplifications were executed using 10 specific microsatellites (SSR) markers developed for the *Musa* genus from A genome (Creste *et al.*, 2006) and B genome (Buhariwalla *et al.*, 2005). All SSRs were fluorescently labeled at the 5' end of each forward primer (the reverse primer remains unmodified) and organized in 3 panels based on the fluorescence and allele sizes reported for each SSR to avoid overlapping effects, moreover by using differential fluorescence in every panel it allows a more precise allele calling for each sample (Table 2). The PCR reactions were performed in 96-well plates and amplified in a PTC-100 thermal cycler. The samples were adjusted to a final volume of 20μl, containing 20ng of genomic DNA, 5,4μL of ultrapure water, 0,8μL of each primer (10μM), 0,16μL of Taq DNA polymerase (0,05u/μL; Fermentas), 2,16μL of PCR Buffer (NH₄)₂SO₄ 10X, 1,6μL of MgCl₂ (25mM), 3,47μL dNTP's (5mM), 2μg/mL of BSA and 1,6μL of Trehalosa (10%). The amplification parameters were as follows: initial denaturation step at 94°C x 5 min, 32 cycles of 94°C x 1 min, 62°C for 1 min and 72°C for 1min, followed by a final extension cycle of 72°C x 10 min. The amplicons were visualized under UV light in agarose gels 1,5% stained with GelRed® (Biotium).

Panels organization, multiplexing and alleles detection

All primers were organized according to the fluorescent dye [6-FAM (Blue), NED(Yellow), (PET) Red and (VIC) Green] and allele sizes previously reported.

Three panels were organized with three pairs of primers each (forward and reverse) (Table 2). The multiplexing strategy was as follows: 1) once all SSRs from every panel were amplified and verified, the PCR products were mixed in a single 96-well plate (10 µl from every SSR) considering the position of every sample into the plate, in that sense three different dyes are combined in every well for every sample; 2) from the multiplexing plate an aliquot of 5 µL was transferred into a new plate; and 3) the fluorescence detection was carried out in Cornell University (USA) using the ABI 3730xl DNA analyzer Applied Biosystems following the manufacturer's instructions. The detection system was based on the capillarity technology where fragment sizes are separated according to the differential migration on each capillary, at the same time a molecular-weight size marker, with known fragment sizes, is added and analyzed to determine the specific size of alleles obtained for every single sample. This detection is performed by a laser that detects the signal emitted by the fluorophore contained in the amplicons as well as in the molecular-weight size marker.

Table 2. List of microsatellites markers used with their respective sequence, fluorescent dye, and panel organization.

SSR	Motif	Primer Sequence	Product size	Dye	Panel	Author
Mb1 – 63	(GA) ₂₀	F TCTCAGGAAGGGCAACAATC R GAATCAAAGAGGCGAGAAGACGA	191	FAM	1	Buhariwalla <i>et al.</i> , 2005
Mb1 – 69	(TC) ₁₅	F CTGCCTCTCCTTCTCCTTGAA R TCGGTGATGGCTCTGACTCA	386	PET	1	Buhariwalla <i>et al.</i> , 2005
Mb1 – 113	(GA) ₁₈	F AGGTGCCACACAGTTCAGACA R CAACCCAAACCTGTCGACCAA	399	VIC	3	Buhariwalla <i>et al.</i> , 2005
Mb1- 134	(GA) ₂₁	F ATGCCCAAGAAGGGAAGGGAA R TAATGCCGAGGATCAGTGTGA	398	PET	2	Buhariwalla <i>et al.</i> , 2005
MaOCEN01	(CT) ₁₇	F TCTCAGGAAGGGCAACAATC R GGACCAAAGGGAAGAAACC	210-250	NED	3	Creste <i>et al.</i> , 2006
MaOCEN04	(AG) ₂₀	F GCCGAGAGAGAAATGGAA R CTGCGAGGGTAGAGTGGATG	130-180	VIC	2	Creste <i>et al.</i> , 2006
MaOCEN08	(GA) ₁₆	F CTGCGAGGGTAGAGTGGATG R AATCTGGACGAGGGAGCA	240-270	FAM	2	Creste <i>et al.</i> , 2006
MaOCEN13	(TC) ₁₆	F GCTGCTATTTTGCTTGGTG R CTTGATGCTGGATCTGG	141-200	PET	3	Creste <i>et al.</i> , 2006
MaOCEN17	(TC) ₁₅	F CGCTCTGCCTTGTTCTGT R TGGGTTGACTCCTTTGTGG	150-190	VIC	1	Creste <i>et al.</i> , 2006

Data analysis

For the alleles calling, the software GeneMarker® (Soft Genetics State College, United States of America) was used specifying the panel, expected allelic sizes (in base pairs) and the dye to the respective fluorophore. Once the alleles were detected, a data matrix

was obtained and organized in Microsoft® Excel. For dissimilarity analysis, the software DARWIN v. 6.0.014 (Perrier *et al.*, 2003) was used, following the genetic distance methods and the Neighbor-Joining Tree algorithm based on a Bootstrapping of 3000 replicates, generating a Maximum Parsimony unrooted tree a priori which represents the best topology of the genetic diversity. For the identification of Polymorphic Information Content (PIC), the number of alleles, heterozygosity and genetic diversity, the software PowerMarker v. 3.25 (López & Hidalgo, 1994) was used. Additionally, the estimation of genetic distances between individuals and the Principal Coordinates Analysis (PCoA) were carried out in GenAlEx v. 6.0 (a complement program for Microsoft® Excel). To build a three-dimensional PCoA graphic, the software XLSTAT (Addinsoft, 1993) was employed, representing the distribution and relative spatial location of the cultivars according to their genetic distances. Finally, a hierarchical cluster analysis (dendrogram) was built with the statistical package R (2008) under Ward genetic dissimilarity algorithm.

RESULTS AND DISCUSSION

Morphological characterization

Correlation analysis

From a total of 57 morphological descriptors obtained for the 12 *Musa* cultivars, 19 quantitative variables were considered [five with banana type genome (A) and seven with plantain type genome (B)], for the simple correlation analysis which showed five highly significant ($P \leq 0,001$) and positively correlated variables such as: fruit length with number of empty nodules ($r^2 = 0,84$), diameter with perimeter of the peduncle ($r^2 = 0,99$), male bud perimeter with diameter ($r^2 = 1,0$), male bud length with perimeter ($r^2 = 0,92$), and male bud length with diameter ($r^2 = 0,92$). This indicates that such descriptors are very correlated and to consider them as stable descriptors, for future characterizations of related cultivars and to minimize the number of descriptors for this characters (Franco & Hidalgo, 2003; Batte *et al.*, 2018).

Among the significantly ($P \leq 0,05$) correlated variables, 16 of them were considered being 11 positively correlated and 5 negatively correlated. The variables for positive correlations ranged between $r^2 = 0,76$ (for male bud size with diameter) and $r^2 = 0,81$ (for blade length with pseudostem height), whereas the lowest correlations were between the variables: pseudostem height with aspect ($r^2 = 0,59$), blade length with peduncle length ($r^2 = 0,60$), and blade width with peduncle perimeter ($r^2 = 0,62$).

The negatively correlated variables comprised: number of fruits with leaf ratio ($r^2 = -0,61$), pseudostem height ($r^2 = -0,67$), petiole length ($r^2 = -0,61$), and fruit length ($r^2 = -0,71$), and finally fruit length with male bud size ($r^2 = -0,65$). These negative correlations highlight the influence between the number of fruits with fruit length,

considering that as the higher the number of fruits the lower the length of them, this a very important characteristic to take into account in plant breeding programs. Previous morphological studies have demonstrated the significant discrimination among *Musa* cultivars related to male bud, flower, and fruit (Daniells *et al.*, 2001) which agrees with the results obtained here.

Principal component analysis

For the principal component analysis, 74,26% of the variance was explained in the first three components, where the first contributed with 34,44% of the variance, being positive and negative correlations observed. For example, variables such as diameter (0,3237), perimeter (0,3236) and length (0,2933) of the male bud contributed positively to the component, whereas fruit length (-0,3057), number of nodules on the peduncle (-0,2411) and petiole length (-0,2389) contributed negatively. These variables represent a considerable level of differentiation among the *Musa* cultivars evaluated here, also the variables related to male bud are key factors to be considered in further analyses. For instance, Geronimo *et al.* (2013) who analyzed morphological characterizations on eight *Musa* varieties, found the highest discrimination influence related to male bud descriptors, also Brandão *et al.* (2013) identified the importance of this character for diversity quantification and variability applied in germplasm collections. These studies support the significance of male bud descriptors which, as reported here, provided the best discriminatory positive values for the cultivars analyzed.

The second component contributed with 26,07% of the variance where characteristics such as leaf length (0,4112), pseudostem height (0,4078) and aspect (0,3437) positively correlated with this component, and the negative correlations were for fruit number (-0,2347), male bud size (-0,1465) and number of nodules on the peduncle (-0,0206), being the positively correlated traits also important to be considered for cultivars discrimination. The third component, which contributed with 13,75%, is composed basically by petiole width (0,3586), perimeter (0,3301) and diameter of the peduncle (0,3280). The negative correlations comprised leaf ratio (-0,3174), male bud length (-0,2880) and petiole length (-0,2593). In general, the traits related to the bunch (i.e. male bud, peduncle, and fruit) demonstrated their influence in the three components, showing their importance for morphological characterizations.

Cluster analysis

Considering the Euclidian distance obtained for the cluster analysis for quantitative variables, a cutting line distance of seven was determined to generate four main clusters. These clusters showed a very low Euclidean distance within the groups (≤ 6). For sub-cluster 1 the plantain cultivars Africa, Plátano Trescientos, Dominico Hartón and

Hartón (known as “Horn Plantain” AAB genome) conform this group, where the most distinctive trait shared is the absence of male bud. The sub-cluster 2 is composed of plantain cultivars Guayabo, Plátano Mocho and Dominico, these cultivars are characterized by normal male bud and compact bunch appearance. For the sub-cluster 3 the cultivars Banano Urabeño, Banano NN, and Banano Andes conform the banana cluster, being B. Andes the most distant cultivar on which traits such as higher number of hands on the bunch (≥ 10), lower height, wax absence on pseudostem, long petioles and compact bunches support its differentiation, being these traits similar to those reported for the varieties known as “FHIA” (developed by FHIA organization) (Piña *et al.*, 2006). Similarly, B. NN is related to Gros Michel banana type respect to the number of hands and bunch shape (<http://www.promusa.org/Gros+Michel>, visited on July 20 – 2018). The sub-cluster 4 contained Guineo, Dominico Hartón and B. Enano, however, it is noticeable that Dominico Hartón clustered with the other cultivars classified as bananas, nevertheless traits related with male bud could be influencing this particular clustering (data not shown).

The qualitative analysis, in general, showed that intermediate leaf habit was the common condition for all cultivars surveyed (91,7%), and B. Enano was the only cultivar with erect leaf habit. Respect to pseudostem height 83,3% of the plants were above 3m (measured from the base of the pseudostem to the emerging point of the bunch), also 83,4% of them possessed pseudostem girth between 45 and 64 cm (measured at 1m from the base of pseudostem), however 8,3% of plants registered weak pseudostem (25 – 44cm) as well as 8,3% vigorous pseudostem (≥ 85 cm). Male bud was present in 53,3% of the cultivars and absent in 33,3%, although for one sample (B. Andes) there was no registration for this trait. Cluster analysis carried out for these variables showed three main sub-clusters (Euclidean cut off line distance of 10), being plantains grouped in sub-clusters 1 and 3 (highlighted in green) and bananas in sub-cluster 2 (highlighted in orange, Fig. 1). Characteristics such as presence or absence of male bud was the main trait that differentiated among the plantain cultivars into the sub-clusters 1 and 3 respectively. Moreover, additional traits related to pseudostem height and color, petiole canal, wax presence on leaf sheaths, blade symmetry, absence of pubescence in the peduncle, rachis position, transversal section of fruit and fruit apex, and length and width of the pedicel differentiated the sub-cluster 1; whereas for the sub-cluster 3, traits that corresponded to leaf habit and ratio, petiole length, peduncle diameter, male bud type, fruit position and peel thickness were differential for this sub-cluster. Finally, the differential descriptors for banana cluster included pseudostem aspect, color and pubescence of the peduncle, rachis type and position, male bud type, number of fruits in the bunch, fruit position and apex, presence of floral relicts, and fruit peel thickness. The descriptors included in this study were previously analyzed by Simmonds & Shepherd (1955) demonstrating its contribution to both banana and plantain cultivars discrimination for taxonomy purposes considering 15 morphological characters. However, recent studies (Batte *et al.*, 2018) have proposed

and demonstrated that stable characters should be considered *a priori* for any cultivar classification, which accounts: Pseudostem color, petiole sheath color, and rachis position, being these traits useful to distinguish the *M. acuminata* populations. The previously cited qualitative traits correlate with the findings of our study and all of which seem to be widely used at different levels. In spite of many morphological characters that differentiate *M. acuminata* and *M. balbisiana*, just a few characters distinguish them, however, imperative for *Musa* spp. subspecies discrimination (Rouard *et al.*, 2018). This bias on phenotyping or morphological characters are influenced by environmental factors as well as the complementary and polygenic nature of underlying genetic factors (Brown *et al.*, 2017).

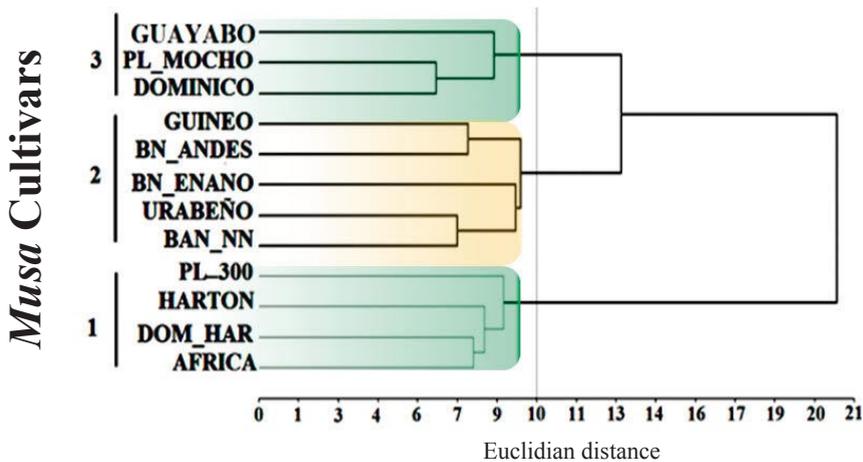


Figure 1. Cluster analysis for qualitative traits under Ward genetic dissimilarity analysis. In green, plantain sub-clusters and in yellow banana sub-cluster. The cutting line corresponds to the Euclidean distance. The numbers on the left corresponds to the sub-cluster assignment.

Molecular characterization

Genotyping analysis was carried out in 12 *Musa* cultivars using 10 fluorescently labeled Microsatellites (SSR) (See materials and methods for details). However, the SSR Mb1-139 could not be amplified even though when several thermal conditions and PCR additives were tested. In previous studies this marker has been reported as a difficult locus to be amplified (Ashikin *et al.*, 2012), probably issues related to the priming site such as mutations in the annealing sequence (Colson & Goldstein, 1999) null allele (Chapuis & Estoup, 2007), or competitive amplification (Hippolyte *et al.*, 2012) can be underlying these difficulties. For nine SSRs, a total 216 allelic data matrix was obtained (based on the triploid nature of the hybrids), where 72 alleles showed polymorphism.

The SSR MaOCEN08 provided the highest number of polymorphic alleles (12), whereas the SSR Mb1-63 contributed with the lowest number (4). In general, the average of polymorphic alleles per locus was 8. The polymorphic information content (PIC) detected by all SSRs showed an average of 0,77, where the marker Mb1-63 was the less polymorphic (PIC = 0,56) and the markers MaOCEN08 and Mb1-113 the most polymorphic ones (PIC = 0,87), being in general highly informative loci (Table 3). Additionally, these values correlate with the considerable heterozygosity ($\mu = 0,48$) and genetic diversity ($\mu = 0,79$) found in the population studied, enabling not only the discrimination of the cultivars surveyed here but also their applicability for germplasm characterizations. The remarkable polymorphism identified by the SSR employed in the present study could be related to, as proposed Rouard *et al.* (2018), paleopolyploid condition of banana, which experienced three independent whole-genome duplications generating some genome fractionations, therefore increasing genome variability among cultivars. These chromosome rearrangements and diploidization resulting in the functional divergence of subgenomes in polyploidy bananas, additionally genomic homologous exchanges may contribute to the diversity leading to a rapid genetic modification which in turn correlates to agronomic trait (Wang *et al.*, 2019).

Table 3. Microsatellites marker's parameters identified in 12 *Musa* cultivars.

SSR	Major Allele Frequency	Allele No.	Gene Diversity	Heterozygosity	PIC
Mb1-63	0.50	4	0.63	0.00	0.56
Mb1-69	0.32	7	0.80	0.27	0.78
Mb1 – 113	0.21	9	0.86	0.58	0.84
Mb1- 134	0.42	4	0.68	0.00	0.62
MaOCEN01	0.42	8	0.77	0.33	0.74
MaOCEN04	0.17	10	0.88	1.00	0.86
MaOCEN08	0.21	12	0.88	0.58	0.87
MaOCEN13	0.27	9	0.83	0.91	0.81
MaOCEN17	0.21	9	0.86	0.67	0.85
Mean	0.30	8	0.80	0.48	0.77

Genetic differentiation between plantain and banana [hereafter population one (P1) and population two (P2) respectively] based on allelic frequencies was $F_{ST} = 0,061$, and according to Wright, (1978) this value indicates a moderate genetic differentiation, where the most frequent alleles for P1 corresponded to SSRs Mb1-134 and Mb1-63, and for P2 was MaOCEN13.

Dissimilarity analysis obtained for molecular data revealed three main sub-clusters categorized by the genetic distance among cultivars (Fig. 2a). From bottom-up the first

cluster is composed basically by plantains under tree topology ((Dominico Hartón, Hartón) (B. Andes, Dominico)), the second sub-cluster showed a tree topology ((B. Urabeño) (B. NN, B. Enano)), and the third sub-cluster is composed by (Guayabo (Guineo, Africa) (Mocho, Trescientos)). Despite that Guayabo and Guineo are classified as bananas these cultivars are cooking bananas, characteristic shared with plantains and in which case, probably, the genetic background could be playing an important role in their relatedness. However, cluster comparisons demonstrated some plantain and banana cultivars well-differentiated, and this differentiation resides on the private alleles found for P1 at two loci MaOCEN04 (146, 160, 162bp) and MaOCEN08 (250, 258, 259bp), whereas for P2 the private alleles were found at three loci Mb1-34 (386), Mb1-113 (361) and MaOCEN01 (244). Additionally, the agreement between the sub-clusters for some banana cultivars (B. Enano, B. Urabeño and B. NN) obtained from quantitative and qualitative traits and molecular data represents the well-differentiated banana cultivars for both morphological and molecular analysis. However, some disagreements in tree topology between all datasets were found for the sample B. Andes (AAAA). This cultivar has not been characterized before being difficult to trace its origin. The molecular profile showed a considerable differentiation of B. Andes respect to banana cluster, instead, it is grouped with plantain cultivars. The uniqueness of this sample relies on the absence of two alleles (≥ 300 bp) at the locus Mb1-69 plus the private alleles for MaOCEN04 (144bp), MaOCEN08 (247 and 250 bp), MaOCEN13(151bp) and MaOCEN01 (238bp). However, the closeness of B. Andes with plantain cultivars reside basically on 5 private alleles shared between them, SSR Mb1-63 (384bp), MaOCEN17 (166bp), Mb1-134 (384bp) MaOCEN13 (156bp) and Mb1-113 (353bp), probably causing this association. Also due to the genomic origin of these SSRs (Buhariwalla *et al.*, 2005; Creste *et al.*, 2006) indicates that they are not related to phenotypic traits, consequently being under higher mutational events (XU *et al.* 2013; Vieira *et al.*, 2016). Moreover, studies carried by Buhariwalla *et al.* (2005) demonstrated the detection of homologous loci using SSRs designed in B genomes that later could be amplified in A genomes cultivars. Also, the hybrid origin of plantain cultivars [*M. acuminata* (A) x *M. balbisiana*(B)] could be related with the genomic background that exist among them, and finally a plausible explanation indicates that all alleles with the same length at the same locus could arise from different alleles with convergent evolution (i.e. homoplasy), condition that is widely known in microsatellites evolutionary nature (Estoup *et al.*, 2002; Hippolyte *et al.*, 2012), nevertheless the origin of this cultivar remains uncertain and further studies are required.

Genetic dissimilarity analysis

Under Neighbor-Joining (NJ) algorithm the genetic dissimilarity analysis showed three main sub-clusters that agree with the previously described cluster analysis, and again B. Andes is close to plantains demonstrating the influence of shared alleles among these cultivars. However, the only well supported sub-cluster under bootstrapping of posterior probability (3000 replicates) was the banana sub-cluster ($\geq 70\%$)(B.

Urabeño (B. Enano, B. NN)) indicating the high differentiation of these cultivars and the complex genetic background among plantains ($\leq 5\%$ of genetic dissimilarity) and the subsequent low bootstrapping support ($\leq 31\%$) for the basal nodes (Fig. 2b). On the other hand, Principal Coordinate Analysis (PCoA) explained 35,85% of the genetic variation among *Musa* cultivars through three main axes. The axis one explained 18,53% of the genetic variation, the axis two 17,4% and the axis three 13,54%. Although P1 and P2 populations are highly differentiated under this analysis, some cultivars can be identified as intermediates (Guineo and Guayabo) indicating the considerable hybridization present on them (Fig. 2c). In previous studies, Guineo and Guayabo cultivars were classified into the banana group based on the results obtained for morphological, physical, and chemical analysis (Corpoica C.I Palmira-Colombia, personal communication), nevertheless under molecular data these cultivars clustered within the plantain group, this could be a result of the genomic nature of the SSRs employed for this analysis and the hybrid nature of these cultivars (with a considerable introgression of A and B genomes). Respect to B. Andes it is grouped apart based on the differential traits that it possesses, showing disagreements among morphological and genetic analyses.

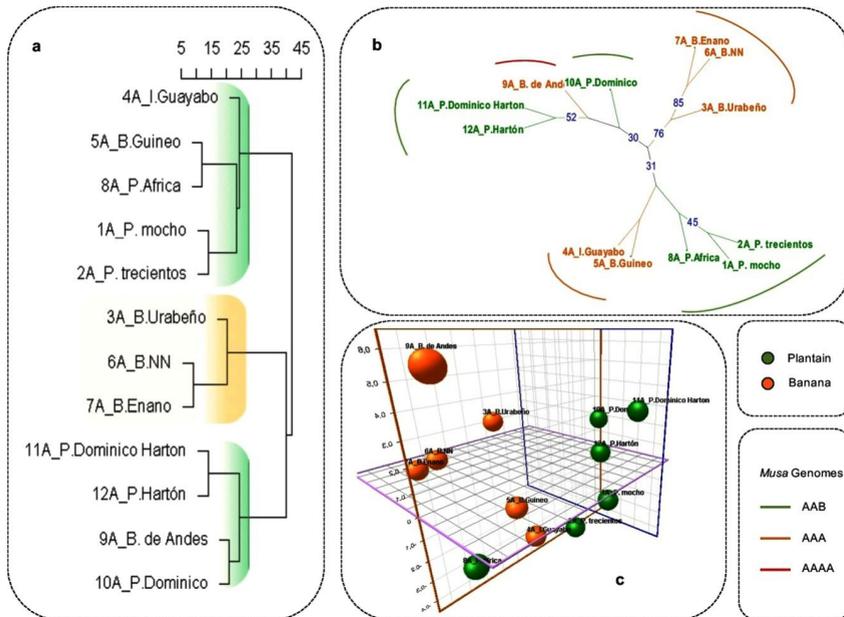


Figure 2. **a.** Ward dissimilarity analysis for 12 *Musa* cultivars using 9 SSRs, where the scale values are considered in percentage (highlighted in green are plantain clusters and in yellow banana cluster). **b.** Unrooted neighbor-joining tree where the bootstrap values are shown in blue. **c.** Three-dimensional graphic for Principal Coordinate Analysis (PCoA) showing the samples evaluated (plantains green and banana yellow) with their respective identifications.

CONCLUSIONS

The identification of a few differential morphological traits such as pubescence, apex fruit shape and waxy condition in pseudostem and leaves, are a very important tools for cultivars classification on the field, also highly applicable for *Musa* collections and germplasm banks to ease the codification of information inherent to specific cultivars. The set of SSR markers employed in the present study demonstrated the discrimination power, not only between *M. acuminata* and *M. balbisiana*, but also within species level and closely related cultivars. Together with automated detection systems, these fluorescently labeled SSRs can be useful for genetic diversity studies in *Musa* species and highly applicable for germplasm evaluations and genotyping. Nevertheless, more genetic markers need to be added to clarify the genetic origin of Guayabo, Guineo and B. Andes cultivars and to identify the genetic background underlying these intermediate hybrids.

ACKNOWLEDGEMENTS

Special thanks to Universidad de Caldas for the financial support under the Project “Aprovechamiento agroindustrial de los excedentes de producción de plátano y banano en San José-Caldas” and the national fund, Fondo de Ciencia, Tecnología e Innovación del Sistema General de Regalías under the project “Implementación de estrategias para el fortalecimiento tecnológico de la actividad agrícola y de transformación de plátano en Caldas”.

REFERENCES

- Addinsoft. (1993). Xlstat. Statistical Analysis in Excell. Addinsoft free trial version. All rights reserved. *Berkley California*.
- Agronet. (2017). Evaluaciones Agropecuarias Municipales. *Gobierno Nacional De Colombia*, 4.
- Ashikin, N., Abdullah, P., Saleh, G. Bin, Tarwaca, E., Putra, S., & Wahab, Z. Bin. (2012). Genetic relationship among *Musa* genotypes revealed by microsatellite markers. In *African Journal of Biotechnology* (Vol. 11). <https://doi.org/10.5897/AJB10.1319>
- Batte, M., Mukiibi, A., Swennen, R., Uwimana, B., Pocasangre, L., Hovmalm, H. P., ... & Ortiz, R. (2018). Suitability of existing *Musa* morphological descriptors to characterize East African highland ‘matooke’ bananas. *Genetic Resources and Crop Evolution*, 65(2), 645–657. <https://doi.org/10.1007/s10722-017-0562-9>
- Blacket, M., Robin, C., Good, R., Lee, S., & Miller, A. (2012). Universal primers for fluorescent labelling of PCR fragments—an efficient and cost-effective approach to genotyping by fluorescence. *Molecular Ecology Resources*, 12(3), 456–463; <https://doi.org/10.1111/j.1755-0998.2011.03104.x>
- Blair, M., Díaz, L., Buendía, H., & Duque, M. (2009). Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics*, 119(6): 955–972. <https://doi.org/10.1007/s00122-009-1064-8>
- Brandão, L., Souza, C., Pereira, V., Silva, S., Santos-Serejo, J., Ledo, C., & Amorim, E. (2013). Descriptor selection for banana accessions based on univariate and multivariate analysis. *Genetics and Molecular Research*, 12(2): 1603–1620. <https://doi.org/10.4238/2013.May.14.1>
- Brown, A., Tumuhimbise, R., Amah, D., Uwimana, B., Nyine, M., Mduma, H., ... Swennen, R. (2017). Bananas and Plantains (*Musa* spp.) (in) H. Campos & P. D. S. Caligari (Eds.), *Genetic Improvement of Tropical Crops* (1st ed., pp. 219–240). <https://doi.org/10.1007/978-3-319-59819-2>
- Buhariwalla, H., Jarret, R., Jayashree, B., Crouch, J., & Ortiz, R. (2005). Isolation and characterization of microsatellite markers from *Musa balbisiana*. *Molecular Ecology Notes*, 5(2): 327–330. <https://doi.org/10.1111/j.1471-8286.2005.00916.x>
- Chapuis, M.-P & Estoup, A. (2007). Microsatellite Null Alleles and Estimation of Population Differentiation. *Mol. Biol. Evol.*, 24(3), 621–631; <https://doi.org/10.1093/molbev/msl191>

- Christelová, P., De Langhe, E., Hřibová, E., Čížková, J., Sardos, J., Hušáková, M., ... Doležel, J. (2017). Molecular and cytological characterization of the global *Musa* germplasm collection provides insights into the treasure of banana diversity. *Biodiversity and Conservation*, 26(4): 801–824. <https://doi.org/10.1007/s10531-016-1273-9>
- Colson, I., & Goldstein, D. B. (1999). Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics*, 152(2): 617–627.
- Creste, S., Benatti, T., Orsi, M., Risterucci, A., & Figueira, A. (2006). Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. *Molecular Ecology Notes*, 6(2), 303–30. <https://doi.org/10.1111/j.1471-8286.2005.01209.x>
- D'hont, A., Denoeud, F., Aury, J. M., Baurens, F. C., Carreel, F., Garsmeur, O., ... & Wincker, P. (2012). The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature*, 488(7410): 213–217. <https://doi.org/10.1038/nature11241>
- Daniells, J., Jenny, C., Karamura, D., & Tomekpe, K. (2001). Musalogue: A Catalogue of *Musa* Germplasm, Diversity in the Genus *Musa*. INIBAP, Montpellier, France.
- De Langhe, E., Hřibová, E., Carpentier, S., Doležel, J., & Swennen, R. (2010). Did backcrossing contribute to the origin of hybrid edible bananas? *Annals of Botany*, 106(6): 849–857. <https://doi.org/10.1093/aob/mcq187>
- Doyle, J. (1991). DNA Protocols for Plants. In: Hewitt G.M., Johnston A.W.B., Young J.P.W. (eds) *Molecular Techniques in Taxonomy*. NATO ASI Series (Series H: Cell Biology), 57. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-83962-7_18
- Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular Ecology*, 11(9): 1591–1604. <https://doi.org/10.1046/j.1365-294X.2002.01576.x>
- Franco, T. L., & Hidalgo, R. (2003). Análisis estadístico de datos de caracterización morfológica de recursos fitogenéticos (in) *Boletín Técnico. no. 8, Instituto Internacional de Recursos Fitogenéticos (IPGRI)*.
- Geronimo, F., Ibarra, Q., Navia, M., & Aguirre, G. (2013). Caracterización morfológica de platano (*Musa paradisiaca* L.) en la provincia nor Yungas de la Paz y Provincia Chapare de Cochabamba Bolivia. *Agrociencias Amazonia*, 1(2), 35–41.
- Hippolyte, I., Jenny, C., Gardes, L., Bakry, F., Rivallan, R., Pomies, V., ... & Perrier, X. (2012). Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. *Annals of Botany*, 109(5): 937–951. <https://doi.org/10.1093/aob/mcs010>
- IPGRI. (1996). *Descriptors for banana (Musa spp.)*. International Network for the Improvement of Banana and Plantain (INIBAP)
- López, J., & Hidalgo, M. D. (1994). Análisis de componentes principales y análisis factorial. In *Fundamentos de estadística con Systat* (21): 457–503).
- Perrier, X., Flori, A., & Bonnot, F. (2003). Data analysis methods in *Genetic diversity of cultivated tropical plants*. 43-76. *Enfield, Science Publishers, Plymouth, UK*.
- Piña, G., Laborem, G., Sarga, J., Marín, C., Rangel, L., Espinoza, M., & Delgado, A. (2006). Atributos de calidad en frutos de híbridos FHIA (*Musa*) para tres ciclos de cosecha. *Rev. Fac. Agron. (LUZ)*, 23(1): 425–442.
- R core team (2008). R: A Language and Environment for Statistical. *R: A Language and Environment for Statistical Computing*, Vienna, Austria. Retrieved from <https://www.r-project.org/>
- Rouard, M., Droc, G., Martin, G., Sardos, J., Hueber, Y., Guignon, V., ... & Roux, N. (2018). Three new genome assemblies support a rapid radiation in *Musa acuminata* (Wild Banana). *Genome Biology and Evolution*, 10(12): 3129–3140. <https://doi.org/10.1093/gb/evy227>
- Simmonds, N., & Shepherd, B. (1955). The taxonomy and origins of the cultivated bananas. *Botanical Journal of the Linnean Society*, 55(359): 302–312. <https://doi.org/10.1111/j.1095-8339.1955.tb00015.x>
- Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. de F. (2016). Microsatellite markers: What they mean and why they are so useful. *Genetics and Molecular Biology*, 39(3): 312–328. <https://doi.org/10.1590/1678-4685-GMB-2016-0027>
- Wang, Z., Miao, H., Liu, J., Xu, B., Yao, X., Xu, C., ... & Jin, Z. (2019). *Musa balbisiana* genome reveals subgenome evolution and functional divergence. *Nat. Plants*, 5(8): 810–821. <https://doi.org/10.1038/s41477-019-0452-6>
- Wright, S. (1978). Variability Within and Among Natural Populations: 157 (in) *Evolution and the genetics of populations* (4). University of Chicago Press, Chicago.
- Xu, J., Liu, L., Xu, Y., Chen, C., Rong, T., Ali, F., ... & Lu, Y. (2013). Development and characterization of simple sequence repeat markers providing genome-wide coverage and high resolution in maize. *DNA Research*, 20(5): 497–509. <https://doi.org/10.1093/dnares/dst026>