RESEARCH

Musa spp. Germplasm Management: Microsatellite Fingerprinting of USDA–ARS National Plant Germplasm System Collection

Brian M. Irish,* Hugo E. Cuevas, Sheron A. Simpson, Brian E. Scheffler, Julie Sardos, Randy Ploetz, and Ricardo Goenaga

ABSTRACT

The USDA–ARS Tropical Agriculture Research Station (TARS) is responsible for conserving germplasm of a number of important agricultural crop species. Its banana (*Musa* spp.) collection is comprised of diploid, triploid, and tetraploid accessions of cultivated, ornamental, wild, and synthetic hybrid accessions. To estimate genetic diversity, identify gaps, determine integrity, and generate clonal reference multilocus DNA profiles for a total of 175 accessions in the collection, a set of 22 microsatellite markers developed in the framework of the Generation Challenge Program (www.generationcp.org/, accessed 30 June 2014) by Centre de Coopération Internationale en Recherche Agronomique pour le Développement and recommended by the Global Musa Genomics Consortium were screened with an additional 15 reference DNA samples. Twenty-one of the 22 microsatellite markers amplified well and generated a total of 302 alleles with an average number of 14.4 alleles per locus. In general, profiles were reproducible and consistent for the 21 loci when clonal field and tissue culture plants were compared with reference samples. The average number of alleles and gene diversity estimates demonstrated substantial genetic diversity in the collection. Principal coordinate and cluster analyses grouped accessions in the collection according to their ploidy level and genomic compositions. Markers that were used in the study distinguished accessions to the subgroup level and identified mislabeled accessions, notably in the tissue culture collection where phenotypic differences are difficult to observe. The accessions and fingerprint profiles for the TARS collection are available through the USDA National Plant Germplasm System, Germplasm Resource Information Network (GRIN-Global) database www.ars-grin.gov/ (accessed 30 June 2014).

B.M. Irish, H.E. Cuevas, and R. Goenaga, USDA–ARS Tropical Agriculture Research Station, 2200 Pedro Abizu Campos, Ave., Suite 201, Mayaguez, PR, 00680; B.E. Scheffler and S.A. Simpson, USDA–ARS Genomics and Bioinformatics Research Unit, 141 Experiment Station Rd., Stoneville, MS 38776; J. Sardos, Bioversity International, Parc Scientifique Agropolis II 1990 Boulevard de la Lironde, 34397 Montpellier Cedex 5, France; Randy Ploetz, Univ. of Florida, Tropical Research and Education Center, 18905 SW 280th St., Homestead, FL 33031. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. Received 5 Feb. 2014. *Corresponding author (Brian.Irish@ars.usda.gov).

Abbreviations: AFLP, amplified fragment length polymorphism; AMOVA, analysis of molecular variance; DArT, diversity arrays technology; GMGC, Global Musa Genomics Consortium; PCoA, principal coordinate analysis; PCR, polymerase chain reaction; PIC, polymorphic information content; RAPD, random amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; SSR, simple sequence repeat; TARS, Tropical Agriculture Research Station.

Bananas (*Musa* spp.), including plantains, are large monocotyledons of significant economic agricultural importance in the world. On the basis of genetic, linguistic, and archeological research, Perrier et al. (2011) suggest that bananas were domesticated at least 4000 yr ago in southern Asia and Melanesia. Since cultivation began, bananas have been used as important sources of food, in beverages, in medicine, in animal silage, for providing shelter and fiber, as ornamentals, and even in religious ceremonies (Nelson et al., 2006). The harvested fruit provides vital nutrition, especially in Africa and Asia, and is also associated with a large

Published in Crop Sci. 54:2140–2151 (2014).
doi: 10.2135/cropsci2014.02.0101

export industry that generates important revenue and jobs in tropical America. Modern-day cultivars, derived from those initial domestications, were grown on over 10.6 million hectares, with an estimated 144 million metric tons of fruit produced in 2011 (FAOSTAT, 2013).

Taxonomy of the edible bananas is complex since hybridization and polyploidization events have occurred naturally among *Musa* species and subspecies. Moreover, the recognition of specific banana cultivars is often confused by synonyms that exist for a given cultivar in different regions as well as by the range observed in many environmentally influenced morphological traits (Rossel, 1998; Ploetz et al., 2007; Musa Germplasm Information System (MGIS) database www.crop-diversity.org/banana/, accessed 16 June 2014).

Musa was divided into sections Eumusa and Rhodochlamys (22 chromosomes), Australimusa (20 chromosomes), Callimusa (18 or 20 chromosomes), and Ingentimusa (14 chromosomes). Recently, Häkkinen (2013) reduced section Rhodochlamys to synonymy with section Musa (Eumusa), and sections Australimusa and Ingentimusa to synonymy with section Callimusa. With the exception of the minor Fe'i bananas (Callimusa), edible bananas are hybrids between and among diploid *Musa acuminata* (A genome) and *M. balbisiana* (B) (Musa), and uncommon hybrids between these species and *M. schizocarpa* (S) (Musa) and *M. textilis* (T) (Callimusa) (Perrier et al., 2011). Natural, mainly diploid and triploid hybrids with the following genomic compositions are most important: AA, AB, AAA, AAB, and ABB. Synthetic, largely tetraploid hybrids (AAAA, AAAB, and AABB) that have been developed in plant breeding programs are less commonly cultivated (Escalant et al., 2002).

Conventional breeding in *Musa* is complicated by infertility, ploidy, parthenocarpy, and other genetic constraints (Dodds and Simmonds, 1948; Okoro et al., 2011). International breeding programs focus on the development of hybrids that could replace current industry standards (Rosales et al., 1999; Silva et al., 2001). Desirable traits include resistance to diseases and insect pests, agronomic performances, and postharvest qualities that are often first identified in parents in ex situ collections (Rowe and Rosales, 1993; Heslop-Harrison and Shwarzacher, 2007).

A number of important ex situ *Musa* germplasm collections, including the USDA–ARS National Plant Germplasm System collection at the Tropical Agriculture Research Station (TARS), exist worldwide (MGIS database www.crop-diversity.org/banana/). These collections are repositories of wild and cultivated germplasm, which provide propagative material for breeding and production. To be most valuable, ex situ collections should contain genetically diverse, accurately labeled, and true-to-type accessions for which important traits are known (FAO, 2013). Because these collections are maintained as live plants in the field, greenhouses, and laboratories (i.e., tissue culture), detailed records (e.g., labels, field maps, and databases) are essential. Nonetheless, errors can occur, which compromise the usefulness and integrity of collections. To help minimize these problems, molecular markers have been used to complement standard morphological characters that are used to manage *Musa* germplasm (Creste et al., 2003, 2004). Molecular markers can assist taxonomic queries, help identify germplasm errors, estimate genetic diversity and help identify gaps in germplasm collections (Kaemmer et al., 1997; Creste et al., 2003, 2004; Christelova et al., 2011b; Lorenzen et al., 2011; Perrier et al., 2009, 2011; Hippolyte et al., 2012; de Jesus et al., 2013).

Numerous molecular markers and techniques have been used to characterize *Musa* spp., including flow cytometry (Dolezel et al., 1997), restriction fragment length polymorphism (RFLP) (Bhat et al., 1994), random amplified polymorphic DNA (RAPD) (Pillay et al., 2001), amplified fragment length polymorphism (AFLPs) (Ude et al., 2002; Wongniam et al., 2010), polymerase chain reaction (PCR) RFLPs (Nwakanma et al., 2003; Irish et al., 2009), and microsatellites (Crouch et al., 1998; Lagoda et al., 1998; Hippolyte et al., 2010, 2012). Microsatellite markers (also known as simple sequence repeats or SSRs) are highly polymorphic, multiallelic, codominant, repeat DNA sequences of two to six base pairs that occur in eukaryote organisms; in practice, they are reproducible and easily scored. Recently, Christelova et al. (2011b), Hippolyte et al. (2012), and de Jesus et al. (2013) used microsatellites to estimate genetic diversity, identify gaps in collections, determine parentage, and identify propagation mistakes in cultivated and wild banana germplasm.

Although in some cases costly, and requiring specialized equipment and training, newer marker systems (technologies and techniques) with enhanced capacities are becoming available. For example, microsatellite markers developed in *Musa* gene regions can be employed in diversity assessment and marker assisted selection (Amorim et al., 2012). Diversity arrays technology (DArT) have been used to assess genetic diversity and in pedigree analysis in *Musa* spp. (Risterucci et al., 2009). Furthermore, with the advent of next-generation sequencing platforms and the release of the draft genome sequence for *Musa* (D'Hont et al., 2012), marker techniques with superior resolution (e.g., genotyping-by-sequencing [Chan et al., 2014]) are being expanded and could replace microsatellite markers in banana germplasm characterization.

Many of the *Musa* spp. accessions maintained by TARS have been previously characterized for morphological and agronomic field traits (www.ars-grin.gov/, accessed 16 June 2014). In addition, the ploidy and genomic composition for some of these materials were determined previously with flow cytometry and PCR RFLPs (Irish et al., 2009). Recently, a significant number of accessions were

incorporated into the collection. To assess genetic diversity in the enlarged TARS collection, all accessions were evaluated with microsatellite markers. The objectives of the study were to: (i) generate standard reference fingerprint profiles for accessions; (ii) estimate genetic relationships and diversity among accessions; (iii) identify gaps in coverage and redundancies; and (iv) verify genetic integrity of replicate plants in both field and in vitro collections.

MATERIALS AND METHODS

Plant and DNA Samples

The TARS *Musa* spp. collection consists of diploid (32), triploid (122), and tetraploid (21) accessions of cultivated, ornamental, wild, and hybrid accessions (Table S1). Plant material for DNA extractions was collected for a total of 175 accessions from three field plantings on the TARS research farm in Isabela, PR, as well as from the tissue culture collection maintained in the laboratories at the main TARS site in Mayaguez, PR. Cigar leaf samples were collected from all four replicates in the established field germplasm collection. Since plants in the in vitro tissue culture were micropropagated from a single adventitious meristem, DNA was extracted from one representative plantlet for each accession. DNA was extracted using an MP Biomedicals FastDNA Spin Kit (MP Biomedicals) following manufacturer instructions with the addition of 2% polyvynilpirolidine (Sigma-Aldrich) to the extraction buffer as the only modification. An additional 15 DNA samples were graciously provided by Dr. Jaroslav Dolezel and the Musa Genome Resource Centre (hosted by the Centre of Structural and Functional Genomics, Institute of Experimental Botany, Olomouc, Czech Republic) (Table S1). These DNA samples, a subset of a larger "mini-core" collection developed as part of the Generation Challenge Program (GCP), were included as references for subgroups of banana that were not represented in the TARS collection, to confirm matches with synonymously named accessions, and to confirm the reproducibility of the technique and the utilized microsatellites. All DNA samples were then shipped to the USDA–ARS Genomics and Bioinformatics Research Unit in Stoneville, MS, for quantification, normalization, PCR amplification, and subsequent electrophoresis.

Microsatellites

A set of 22 microsatellite primer sequences that were developed previously (Crouch et al., 1998; Lagoda et al., 1998; Hippolyte et al., 2010) was selected for allelic diversity and usefulness in distinguishing diverse *Musa* spp. Primers were fluorescently labeled, and PCR reactions were performed following the Christelova et al. (2011b) protocol, with minor volume modifications. The reaction was performed using 10-ng DNA template and Titanium *Taq* DNA Polymerase (Clontech) in a 5μL reaction. Each reaction contained 10 ng of DNA template, 1.0 pMol each of forward and reverse primer (fluorescently labeled), 1X Titanium *Taq* PCR Buffer, 5UTitanium *Taq* Polymerase, 1.0 μMol dNTP mix, in a 5μL volume. Polymerase chain reaction amplification was performed in three stages, with an initial denaturation step at 95°C for 1 min; followed by 30 cycles of 95°C for 30 s, 55°C for 30 s, and 68°C for 30 s; and a final extension step at 68°C for 3 min. Annealing temperature was the average annealing temperature reported for all primer combinations by Christelova et al. (2011b), and cycling conditions followed Clontech recommendations for use of Titanium *Taq* DNA Polymerase. Following amplification, PCR reactions were multiplexed according to expected fragment size and dye color and run on a 3730XL DNA Analyzer with an internal dye standard (GeneScan 500 ROX)(Applied Biosystems). Allele calls were performed using GeneMapper 4.0 software (Applied Biosystems) to determine alleles and fragment sizes.

Data Analysis

Pairwise matching, performed in Microsoft Excel with macros, for the multilocus microsatellite profiles was performed for the three field and single tissue culture plant DNA samples. Profiles matching allele calls at all loci within accessions were combined to produce a unique fingerprint profile. When a multilocus microsatellite profile within the five samples for a given accession did not match the others (i.e., allele calls at several loci were different), the genotype was not condensed and was treated as a unique sample in the downstream analyses. Distinctly named accessions sharing an exact multilocus profile were recorded, labeled as synonymous accessions, and condensed into a single genotype for the downstream analysis.

Summary descriptive statistics for total number of alleles and polymorphic information content (PIC) for each locus were determined with PowerMarker v3.25 software (Li and Muse 2005). Expected and observed heterozygosity (H_E and H_O, respectively) values were also generated in PowerMarker v3.25 and calculated only for 29 unique diploid genotypes. As the exact number of copies of individual alleles in polyploid species cannot be determined easily, genotypic data for all accessions (i.e., diploids, triploids, and tetraploids) was converted into binary code (expressed as 1, presence; 0, absence) and analyzed as a dominant marker's record (Weising et al., 2005; Hippolyte et al., 2012). First, a pairwise Euclidean genetic distance matrix was generated between all possible pairs of accessions with the binary data using the shared allele coefficient in PowerMarker v3.25 software (Li and Muse, 2005). The resulting distance matrix was subject to principal coordinate analysis (PCoA) as computed by GenAlEx 6.5 (Peakall and Smouse, 2006). The genetic structure of the collection was further investigated by an analysis of molecular variance (AMOVA) using GenAlEx 6.5 (Peakall and Smouse, 2006). Groups were defined on the basis of PCoA analysis, and significance of the partitioning of genetic variance among groups tested employing 1000 permutations. The same distance matrix was also utilized in phenetic clustering using the neighbor-joining algorithm with 1000 bootstrap iterations in PowerMarker v3.25. The Interactive Tree of Life software (Letunic and Bork, 2011) was used to display the dendrogram from the cluster analysis output.

RESULTS

Microsatellite Diversity Analysis

Of the 22 microsatellite markers tested, 21 amplified and generated reproducible results for further analysis. Marker mMaCIR195 was omitted from the analysis because of its poor amplification and difficulty in scoring alleles.

Table 1. Summary statistics and allele distribution across groups (defined by principal coordinate analysis [PIC]) for 22 microsatellite markers used for fingerprinting the USDA–ARS Tropical Agriculture Research Station *Musa* spp. collection.

						Allele									
						G-1§ *n* = 39		G-2 *n* = 13		G-3 *n* = 15		G-4 *n* = 36		G-5 *n* = 4	
	Allele														
Primer†	Range	No.	H_E‡	H_O‡	PIC	No.	PIC	No.	PIC	No.	PIC	No.	PIC	No.	PIC
mMaCIR01	220–294	24	0.91	0.62	0.97	19	0.94	7	0.60	7	0.70	16	0.94	1	0.00
mMaCIR03	90–113	10	0.81	0.62	0.94	9	0.89	5	0.37	6	0.77	5	0.80	2	0.30
mMaCIR07	121–169	21	0.87	0.52	0.96	20	0.95	7	0.37	8	0.69	13	0.86	2	0.30
mMaCIR08	231–251	11	0.79	0.62	0.81	11	0.88	2	0.13	5	0.55	5	0.59	2	0.38
mMaCIR13	246–286	21	0.89	0.52	0.96	19	0.95	7	0.69	8	0.58	11	0.89	1	0.00
mMaCIR24	216–263	13	0.73	0.31	0.90	12	0.88	4	0.44	5	0.50	7	0.75	2	0.38
mMaCIR27	207–232	15	0.80	0.60	0.90	14	0.91	6	0.68	4	0.41	5	0.70	0	0.00
mMaCIR39	309–352	22	0.86	0.72	0.90	18	0.92	6	0.54	7	0.58	11	0.67	2	0.38
mMaCIR40	150–189	16	0.88	0.41	0.95	14	0.89	6	0.45	6	0.78	9	0.86	3	0.55
mMaCIR45	254–275	9	0.80	0.55	0.91	7	0.91	3	0.26	5	0.62	8	0.75	1	0.00
mMaCIR150	237–249	6	0.64	0.31	0.80	6	0.76	3	0.26	3	0.20	4	0.76	2	0.38
mMaCIR152	137–176	13	0.75	0.28	0.93	10	0.82	3	0.54	5	0.50	10	0.88	2	0.30
mMaCIR164	278–389	14	0.89	0.75	0.95	10	0.90	4	0.34	6	0.64	9	0.89	3	0.55
mMaCIR195¶	–		–	–	–	–	–	–	–	–	–	–	–	–	–
mMaCIR196	143–186	18	0.62	0.34	0.96	18	0.95	5	0.44	8	0.69	10	0.84	4	0.55
mMaCIR214	100–108	6	0.90	0.72	0.86	3	0.66	4	0.13	4	0.12	6	0.76	2	0.38
mMaCIR231	215–260	16	0.75	0.45	0.95	16	0.95	7	0.68	6	0.51	8	0.83	2	0.30
mMaCIR260	185–234	11	0.90	0.45	0.86	7	0.82	3	0.13	4	0.39	7	0.87	3	0.38
mMaCIR264	214–331	17	0.61	0.28	0.95	17	0.94	8	0.69	5	0.58	12	0.84	1	0.00
mMaCIR307	141–153	5	0.91	0.72	0.83	5	0.75	3	0.13	4	0.39	3	0.50	1	0.00
Ma-1-32	196–243	19	0.88	0.52	0.95	16	0.94	6	0.37	8	0.70	12	0.79	2	0.30
Ma-3-90	127–160	15	0.91	0.38	0.97	14	0.95	6	0.68	8	0.74	8	0.90	2	0.55
Total		302				265		105		122		179		40	
Mean		14.4	0.81	0.51	0.91	12.6	0.88	5.0	0.4	5.8	0.55	8.5	0.79	1.9	0.28

† m = microsatellite; Ma = *Musa acuminata*; CIR = Coopération Internationale en Recherche Agronomique pour le Développement.

‡ Expected heterozygosity (H_E) and observed heterozygosity (H_O) are reported for the 29 unique diploid genotypes analyzed.

§ G = Group (e.g., G-1 = Group 1) defined in principal coordinate analysis (Fig. 1).

¶ Primer excluded from analysis owing to poor amplification and scoring results.

Amplification products were readily visualized following electrophoresis for all accessions tested except those that originated from *M. coccinea*, for which alleles at many loci failed to amplify. A total of 302 unique alleles with an average of 14.4, and range of 5 (mMaCIR307) to 24 alleles (mMaCIR01) were generated per locus (Table 1). Average PIC for the markers was high (0.91) and ranged from 0.80 (mMaCIR150) to 0.97 (mMaCIR01 and Ma-3-90) (Table 1). For diploid accessions, H_O varied considerably within the panel of markers, with an overall mean of 0.51 and range of 0.28 (mMaCIR152 and mMaCIR264) to 0.75 (mMaCIR164). The H_O values were on average lower (0.51) than the H_E values (0.81). Generally, the number of alleles were not evenly distributed across markers, ploidy, and genomic composition and showed a tendency to decrease as ploidy increased (Table S2).

Multilocus microsatellite profiles for DNA samples from clonal replicate plants from the field and from tissue culture, as well as from reference samples, matched well across all loci. When an exact match for all alleles was found for DNA samples from replicate field and tissue culture plants of a given accession, profiles were condensed to generate one unique consensus profile. Fingerprint profiles that matched across loci for individual accessions (i.e., accessions with different names) were also condensed, with the corresponding unique profile and noted. In all but four cases, replicate DNA samples from plants matched genotypes within accession. Differently named accessions that shared identical genotypes were considered to be synonymous; 105 accessions fell into 18 groups that contained 2 to 25 synonymous accessions (Table 2). Two of the synonymous groups with the largest number of accessions were the Cavendish and the Plantain subgroups, with all 25 and 21 accessions within each subgroup sharing identical profiles, respectively.

Principal Coordinate Analysis and Cluster Analysis

A total of 107 unique genotypes (i.e., genetic profiles) were used to generate a Euclidean pairwise genetic distance matrix. The genetic distance between accessions ranged from 0.0032 ('Monthan', 'Señorita', and 'Bluggoe')

Table 2. Eighteen synonymous groups (including 105 accessions) with their corresponding subgroup name within the USDA–ARS Mayaguez *Musa* spp. collection identified by microsatellite DNA analysis.

1-Red†	5-Cavendish	6-Gros Michel	10-Pisan Awak	15-Plantain	16-Maoli
Cuban Red	1-A	10-A	Bom	African Rhino	Hua Moa
Morado	2-A	2-R-2, 500	Fougamou 1	Chinga	Manini
Morado Enano	3-A	3-R-2, 500	Namwa Khom	Colombian dwarf	Mai Maoli Eka
Verdin	4-A	4-R-2, 500		Common dwarf	
	5-A	Gigante Blanco	**11-Mysore**	Common Harton	**17-Ney Manan**
2-Ibota	6-A	Gros Michel	Mysore	Congo –300	Blue Java
Kahin T. Ruang	8-A	Guaran Enano	Pisang Ceylan	Corozal sel. 25	B.T.S. Island
Yangambi Km5	D. Cavendish			D. superplantain	P. Abu Perak
	D. Valery	**7-Pome (short)**	**12-Ney Poovan**	Dominican dwarf	
3-Sucrier	Enano Gigante	Ant. Finger Rose	Ney Poovan	Dominico Harton	**18-Bluggoe**
Datil La Lima	G. Cavendish	Rajapuri	Safet Velchy	False-horn D. R.	Cacambou
Hapai	G. Governor			French dwarf	Dole
Maia Hapai	Grand Nain	**8-Pome (tall)**	**13-Silk**	French T. D. R.	Dw. Chamaluco
Mosslin	Guineo Doble	Foconah	Golden Pillow	Harton sel. Cha.	Dwarf Orinoco
Niño Comun	Guineo Enano	Hy Brazilian	Manzano	Ihitism	Gipumgusi
Niño Enano	Johnson	Pirineo		Maiden Plantain	Mafafo Adjuntas
	Mahoe	Poni	**14-Laknau**	Maricongo	Mafafo de Puerco
4-Mutika	Monte Cristo	Prata Aña	Dare	Obino l'Ewai	Mafafo Domin.
Igcpoca	M.C. Enano	Señortia	Laknau P.I. 23472	Orishele	Paka
Ignamico	Sc2T		Laknau P.I. 23479	Plantain w/o bud	
Nchumbahoka	T. Somaclone	**9-Pelipita**		Tall Superp.	
	Valery	Pelipita-C.R.			
	Volunteer *Musa*	Pelipita-Col.			
	Williams				
	Ziv				

† Accessions in the same synonymous group shared identical multilocus microsatellite profiles.

to 0.3019 ('Honduras' and 'FHIA-17') with an average distance of 0.18. The distance matrix was then used to generate a sample ordination using PCoA (Fig. 1). The first two principal coordinates explained over 56% of the variation, with approximately 75% of the variation explained in the first three coordinates (data not shown). Five major groups (related in ploidy level and genome composition) could be differentiated: (1) 39 accessions, both diploids and triploids, with a *M. acuminata* genetic background only; (2) 13 'Cavendish' and 'Gros Michel' triploids and derived tetraploid hybrids with a *M. acuminata* genetic background; (3) 15 Pome and Pome-derived tetraploid hybrids; (4) 36 natural diploid and triploid hybrids and bred tetraploids, and (5) three diploid *M. balbisiana* accessions (Fig. 1).

The AMOVA showed a significant variation (F_{ST} = 0.192; $p < 0.001$) among the five groups identified in the PCoA. Differences between groups accounted for 19% of the variation, while within-group diversity accounted for the other 81%. The two groups with the largest number of accessions (1 and 4) accounted for 32 and 28% of the variation, respectively. The other 21% of genetic variation was divided among Groups 2 (8%), 3 (12%), and 5 (1%). Pairwise F_{ST} was not significant between Groups 1 and 5 (F_{ST} = 0.05; p = 0.11), while the largest difference was observed between Groups 3 and 5 (F_{ST} = 0.31; p = 0.01) (data not shown). The number of alleles in each group appeared to be proportional to the number of accessions. Groups 1 and 4 included 265 and 179 alleles, while Groups 2, 3, and 5 contained 105, 122, and 40 alleles, respectively (Table 1).

Neighbor-joining clustering analysis agreed with the PCoA, in that accessions in the five major groups had similar ploidies and genomic compositions (Fig. 2). Cultivars with a pure *M. acuminata* genome formed a large group corresponding to the upper half on the dendrogram (Fig. 2; Panel A), whereas a second group in the lower part of the dendrogram corresponded to *M. balbisiana* and hybrid accessions (Fig. 2; Panel B). The resolution in cluster analysis allowed for further discrimination of eight subclusters (sub-clades) corresponding generally to cultivated subgroups of *Musa* spp., as well as to outlier ornamental species in section Musa. The eight subgroups included: (1) the 'Calcutta 4' AA accession and derived hybrids; (2) cultivated diploids and triploids with a *M. acuminata* and *M. acuminata* subsp. zebrina origin; (3) East African Highland Bananas (EAHB) AAA (Mutika/Lujugira subgroup), as well as several diploids derived from *M. acuminata* subsp. *Banskii*; (4) Cavendish AAA and Gros Michel AAA clones and derived tetraploid hybrids; (5) Pome AAB and Pome-derived tetraploid hybrids; (6) 'Maoli' AAB and Plantain AAB subgroups and derived tetraploid hybrids; (7) Bluggoe, 'Cardaba', 'Ney Mannan', and 'Saba' subgroups (all ABB); and (8) 'Mysore' AAB, 'Ney Poovan' AB, 'Pisang

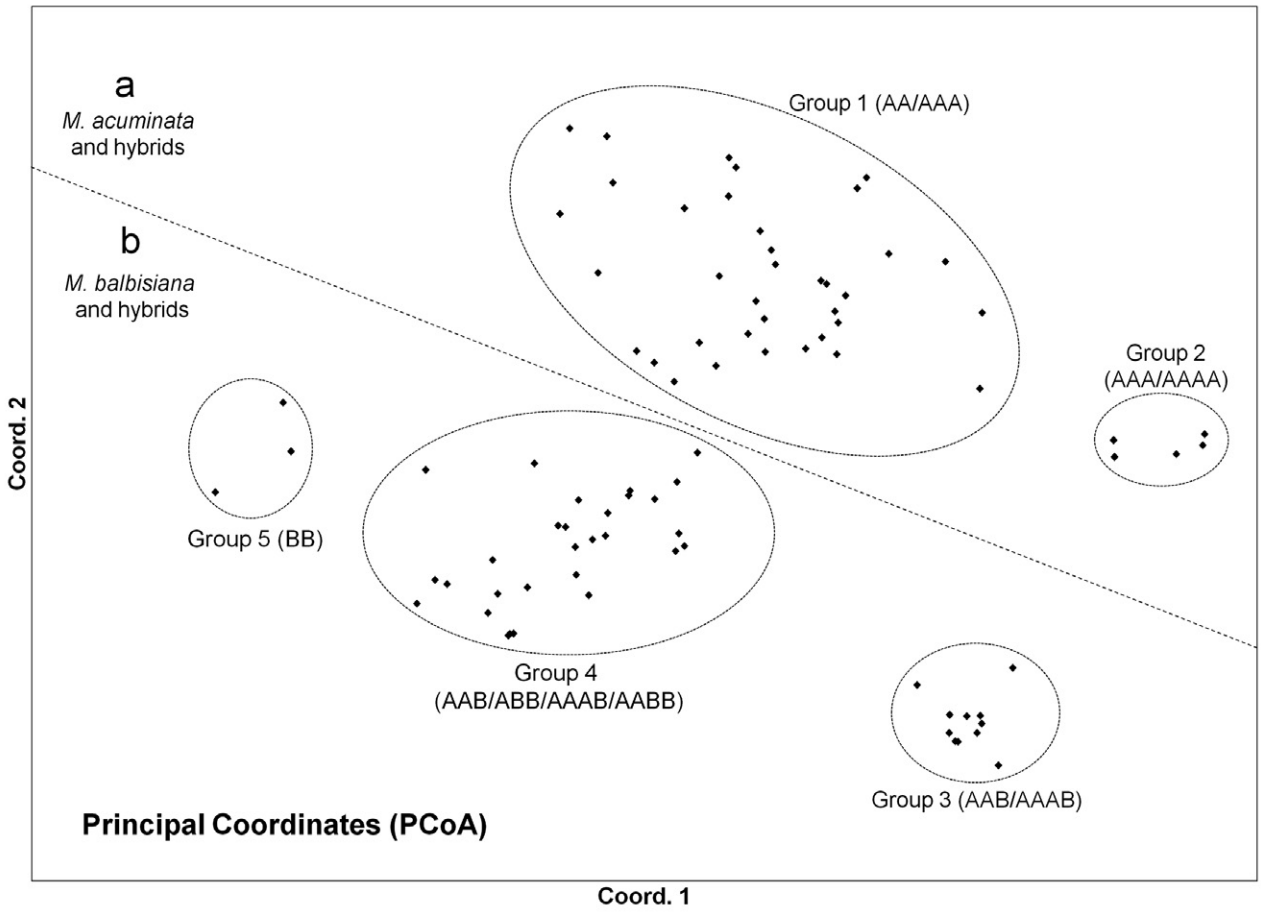

Figure 1. Principal coordinate analysis (PCoA), as computed by GenAlEx 6.5, of 107 unique genotypes belonging to 103 unique *Musa* spp. germplasm accessions maintained in the USDA–ARS Mayaguez collection. Five major groups are identified with the diagonal line separating the two major species contributing to edible banana cultivars: (a) *M. acuminata* and their hybrids, and (b) *M. balbisiana* and their hybrids.

Awak' AAB, 'Pelipita' ABB, and Silk AAB subgroups, as well as *M. balbisiana* accessions. Synthetic tetraploid hybrids grouped closely with their corresponding parental clones. All reference DNA samples matched corresponding accessions within the collection (Fig. 2). The dendrogram also includes the 18 synonymous groups (identified on the dendrogram by their corresponding subgroup name), where multiple accessions shared identical multilocus microsatellite profiles that correspond to the list in Table 2.

The markers were used to identify mislabeled accessions. Two types of "mistakes" were identified on the basis of genotype and clustering. One of the mistakes occurred when at least one of the five DNA samples genotyped from a given accession was different (i.e., an accession with multiple genotypes). A second was a genotype (accession or clone) in the collection that did not match its reported genomic composition and/or genetic group. Examples of both types of mistakes are identified in Fig. 1.

Mistakes in propagation included the tissue culture accession of 'Tigua' AAB, which was an exact match to one of the accessions in the EAHB subgroup, as well as the 'I-C-2' AAAA tissue culture genotype that did not match the field plants and had the same profile as the 'TMB 5295-1' AABB tetraploid accession. Two of the four field plants for the 'Señorita' AAB accession did not match and were identical to plants in the Maoli subgroup AAB. Last, the tissue culture accession of Señorita AAB was also incorrect on the basis of its genotype and shared an identical fingerprint profile to accessions in the Bluggoe subgroup ABB.

Several inaccuracies in matching an accession's genotype to reported passport information were identified. 'Pisang Klutuk Wulung' is reported to be a diploid *M. balbisiana* accession BB, but in our results it grouped with other diploid AA types. Passport information for '1-R-2, 500' indicates that this clone belongs to a series of local Puerto Rican selections in the Gros Michel subgroup, but groups close to other *M. acuminata* diploids in Group 1. The 'Thousand finger' accession, which is very close to the 1-R-2, 500, might also be incorrect, as it is reported to be a triploid AAB. The 'FHIA-25' accession appeared to be a mistake as well as the 'PITA 16' accession. Both of these accessions cluster closely together within the *M. acuminata* Group 1, but are supposed to be hybrids between the two major *Musa* species. On the basis of passport information FHIA-25 is an AAB triploid and PITA 16 an AAAB tetraploid hybrid. The 'Highgate' and '9-A' accessions are

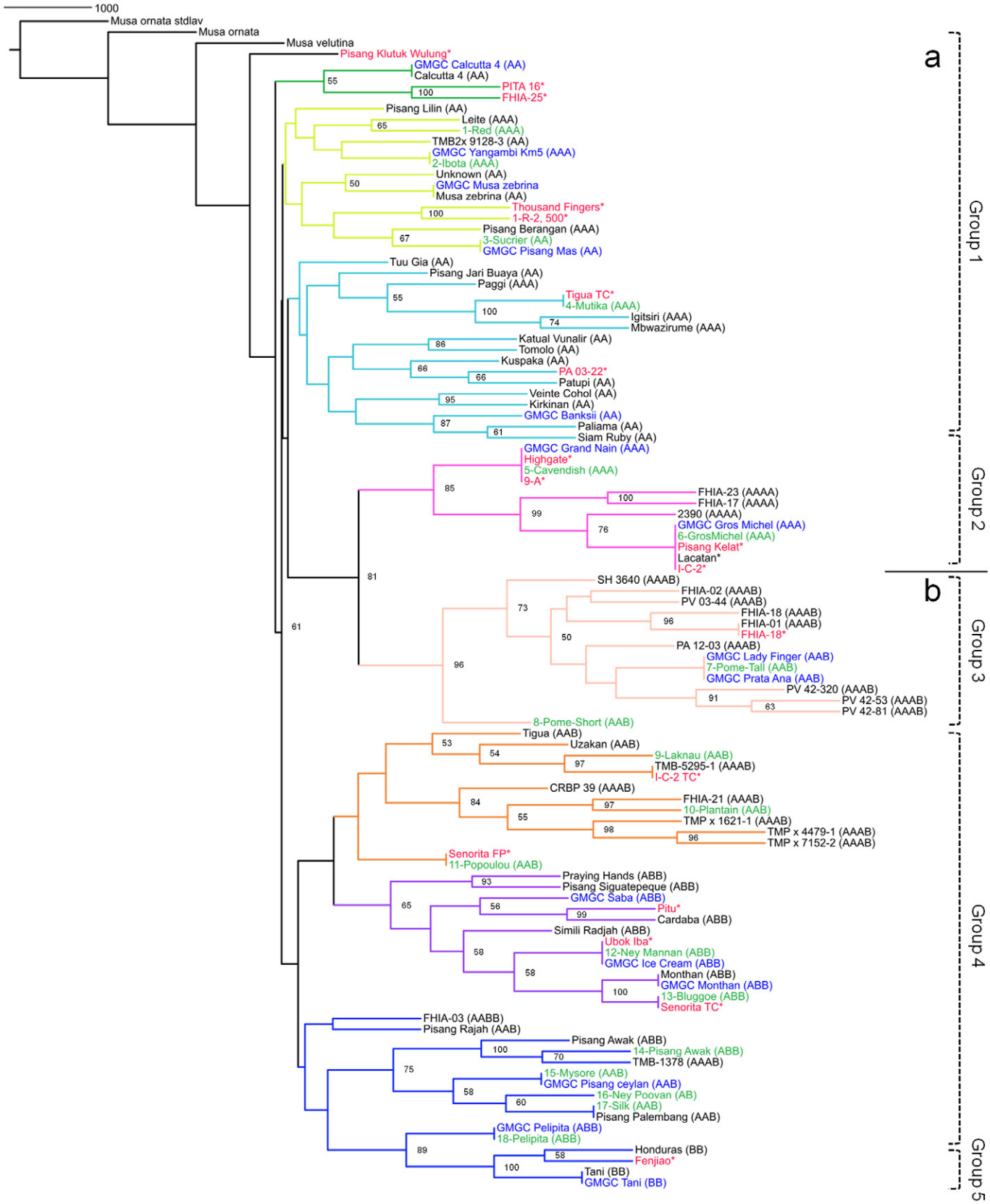

Figure 2. Dendrogram of 107 unique genotypes belonging to 103 unique *Musa* spp. germplasm accessions maintained in the USDA–ARS Mayaguez collection. Dendrogram includes 18 cases (numbered from top to bottom) in which multiple accessions shared identical multi-locus microsatellite profiles. Propagation and genotype mistakes are indicated (asterisk) as well as reference DNA samples (GMGC [Global Musa Genomics Consortium] included in label) from the GMGC. Identified in the figure are the five major groups, the eight subgroups, and the horizontal line separating "pure" *Musa acuminata* genome accessions from (a) *M. acuminata* and (b) *M. balbisiana* genome hybrids. Distance matrix and cluster analysis were performed in PowerMarker v3.25 and tree-visualized using the Interactive Tree of Life software.

supposed to be Gros Michel clones, but both group within the Cavendish group. Three accessions including, 'I-C-2', 'Lacatan', and 'Pisang Kelat' group within the Gros Michel subgroup, but on the basis of passport information should not; I-C-2 is reported to be an AAAA synthetic hybrid, Lacatan a triploid AAA Cavendish clone, and Pisang Kelat a triploid AAB hybrid. In the field collection, one of the two distinct plots for the 'FHIA-18' accession turned out to be 'FHIA-01'. 'Fenjiao', an AAB triploid on the basis of passport information, was also incorrect and grouped closely to other *M. balbisiana* accessions. 'Pitu' grouped closely to the 'Cardaba' accession in the Saba subgroup, but on the basis of passport information is supposed to be an AA diploid. Last, 'Ubok Iba' is thought to be an AAB plantain but grouped in our findings within the Ney Mannan subgroup ABB. In four cases, previously reported genomic compositions (on the basis of flow cytometry and PCR RFLPs [Irish et al., 2009]) were not in agreement with what was observed in the current study. 'Fenjiao' was reported as a diploid AA (*M. balbisiana* in current study). 'Ubok Iba' was a triploid AAB but appears to be a triploid ABB. 'FHIA 02' was a tetraploid AAAA, but a tetraploid AAAB in this study. 'PA 03' was reported as a tetraploid AAAB, but clusters here with AA diploids.

DISCUSSION

Molecular marker techniques have been used for many applications in plant sciences (Agarwal et al., 2008; Thomson et al., 2010). In plant breeding, molecular markers play important roles in marker-assisted selection (Gupta et al., 2010) but also in identification of propagation mistakes and rouges in breeding lines (Lucas et al., 2013). Lorenzen et al. (2011) and Perrier et al. (2009) reviewed the use of molecular markers in *Musa* research and described their various applications. To date, molecular markers have not been effectively employed for marker-assisted selection to improve *Musa*.

Microsatellites have been used to estimate genetic diversity, identify gaps in germplasm collections and resolve propagation mistakes in clonally propagated crops like apple (*Malus domestica* Borkh.) (Hokanson et al., 1998), orange [*Poncirus trifoliata* (L.) Raf.] (Fang et al., 1997), and cacao (*Theobroma cacao* L.) (Irish et al., 2010; Zhang et al., 2010). In *Musa*, microsatellites have been used to understand phylogenies and origins (Li et al., 2010; Perrier et al., 2011; Hippolyte et al., 2012) and to estimate genetic diversity and genetic integrity in germplasm (Creste et al., 2003, 2004; Christelova et al., 2011b; de Jesus et al., 2013). Unfortunately, the ability of microsatellite markers to estimate allelic relationships among genotypes is reduced in polyploid species, since these data must be converted into a binary format for genetic distance estimation, which does not account for allele dosage (Provan et al., 1996; Creste et al., 2003; Hippolyte et al., 2012).

The microsatellite markers used in the current study generated reproducible results across the diverse samples that were evaluated. Use of the Global Musa Genomics Consortium (GMGC)–suggested panel of microsatellite markers was beneficial, as they had been validated previously on a large genetically diverse population of *Musa* spp. as part of the Generation Challenge Program (GCP, www.generationcp.org/) (data can be publicly accessed at the GCP crop registry http://gcpcr.grinfo.net/, accessed 16 June 2014). Christelova et al. (2011b) used the same panel in their evaluations and reported difficulties in generating consistent and reproducible results with three of the markers (mMaCIR 27, mMaCIR195, and Ma-1-32), whereas no problems in amplification were reported for any of the markers in Hippolyte et al. (2012). Marker mMaCIR195 did not perform well in the current study and was excluded in the downstream analyses (Table 1). Unlike the findings of Creste et al. (2003), only the expected number of alleles was observed for each ploidy level in the evaluated germplasm. The multiple plant DNA replicate samples per accession as well as the reference samples made doubtful allele calls easier to eliminate or include, lessening the chances for false positives and/or false negatives. In addition, because much of the germplasm being evaluated had been previously characterized for ploidy and genomic composition using flow cytometry and PCR RFLPs (Irish et al., 2009), the alleles called during scoring were limited to the expected number. A few of the markers inconsistently amplified the outlier *Musa* species included in the study (e.g., *M. coccinea*, which was dropped from the analyses). Crouch et al. (1998) speculated that sequence differences at priming sites between different species might lead to poor amplification. Barbara et al. (2007) showed that transferability of microsatellites was inversely proportional to the distance between taxa for monocot species. The markers used in this study were developed in *M. acuminata* and *M. balbisiana* and consequently would have been expected to perform well within this and closely related species, but less effectively in more distantly related species (Christelova et al., 2011a).

The analyses conducted demonstrated that the National Plant Germplasm System *Musa* collection is genetically diverse and an important repository of plant genetic resources and alleles. The total number of unique alleles (302) and the average per locus (14.4) found in this study were higher than those reported by Creste et al. (2003, 2004). However, data in Creste et al. (2003, 2004) were generated with a smaller set of microsatellite markers, and the sample size was limited. In contrast, results were found to be similar to those reported by Christelova et al. (2011b), who screened a comparable sample population with the same marker panel. The average number of alleles was essentially the same (14.4 vs. 14.5) as reported by Hippolyte et al. (2012), despite their considerably larger sample

size (561 accessions). The total number of alleles (252 of a total of 302) for diploid accessions was high (Table S2), but the low H_O value (Table 1) suggested for this sample a departure from Hardy-Weinberg conditions. This pattern may result from the small sample size (n = 29) but also may be due to the "Wahlund effect," where samples originate from multiple, well-differentiated source populations. In the latter case, the high H_E would reflect the effective presence of numerous alleles that, fixed in their own population, would not reach the expected rate of heterozygotes. The high level of variation in H_O observed within the sample tends to support this hypothesis.

In addition, the large number of alleles in Groups 1 and 4 (265 and 179, respectively) as well as their high PIC value (0.88 and 0.79, respectively) (Table 1) indicate a high degree of genetic diversity in the collection. Generally, as ploidy increased, the number of unique alleles decreased. Creste et al. (2003) and de Jesus et al. (2013) suggested that this might be due to genome duplication events during meiosis in triploids and tetraploids (Table S2).

In both the PCoA (Fig. 1) and cluster analysis (Fig. 2), clear differences were resolved between the two species that are parents of most edible cultivated bananas, *M. acuminata* and *M. balbisiana*. Principal coordinate analysis (ordination methods) and cluster analysis (hierarchical algorithms) were two independent and harmonizing approaches used to show relationships within and between *Musa* spp. germplasm accessions. Messmer et al. (1992) explained that cluster analysis and PCoA complement each other, particularly when the first two coordinates explain over 25% of the original variation (56% in the current study). However, ordination has its limitations, especially when more than a few dimensions are needed to explain relationships among genotypes (Mohammadi and Prasanna, 2003). Results with the microsatellite markers distinguished most accessions to the subgroup level and agreed with previous morphological descriptions and classifications (Simmonds and Shepherd, 1955; IPGRI, 1996; Creste et al., 2003, 2004; Christelova et al., 2011b; Hippolyte et al., 2012; de Jesus et al., 2013). As expected, bred hybrids grouped within or close to their maternal parent in the cluster analysis (e.g., several FHIA hybrids grouped close to their maternal parent, Gros Michel). Likewise, Creste et al. (2003) found similar results when evaluating hybrids with parents in the Pome subgroup (e.g., PV and PA hybrids). The observed fixation index value (= 0.192) among the five groups identified in the PCoA analysis indicates a high degree of differentiation (Hart and Clark, 2007) and reflects the divergent backgrounds involved in modern day *Musa* spp. cultivars. It was also apparent that although diversity was high within the collection, as many subgroups were represented, gaps of other important cultivated (e.g., Fe'i bananas) and wild *Musa* germplasm still exist.

In spite of the high genetic diversity among subgroups, phenotypic differences that exist among accessions within subgroups could not be resolved with the microsatellite markers that were employed. No polymorphisms were observed within the Plantain subgroup, despite notable differences in plant height, pseudostem color, and flower type (Table 2). Our results agree with those of Noyer et al. (2005), who also did not observe differences within their Plantain samples. Although considerable morphological diversity has been described in this subgroup (Ortiz et al., 1998; De Lange et al., 2005), so far molecular data have failed to reflect this variation. Similarly, 25 different cultivars in the Cavendish subgroup shared identical multilocus fingerprint profiles (Table 2; Fig. 2) but varied in their phenotypes. Hippolyte et al. (2012) suggest that the lack of molecular diversity within subgroups could be the result of somatic variation in clonal accessions or those that originate from similar parental crosses. With a different panel of microsatellites, Creste et al. (2003) were unable to distinguish among accessions within the Cavendish and Pome subgroups, whereas de Jesus et al. (2013) found differences in several subgroups, including the Plantains. Christelova et al. (2011b) found polymorphisms among some of the accessions within particular subgroups (e.g., Cavendish) using the same panel of microsatellites used in the present study. Inconsistencies in reported differences (or lack thereof) among accessions within subgroups might be due to the use of diverse sets of microsatellite markers, as priming sites, and the number of possible loci amplified might vary. In addition, the use of different platforms (e.g., gel based vs. capillary sequencers) for amplification product visualization and scoring, missing data and calibration of size standards all could potentially explain some of these discrepancies. However, since the variation in phenotypes within subgroups is thought to be due to point mutations or epigenetic changes, microsatellite markers would not be expected to resolve these differences (Hippolyte et al., 2012). Other molecular markers known for their resolution but lack of reproducibility, including RAPDs (Agoreyo et al., 2008) and AFLPs (Ude et al., 2002), have identified genetic diversity in particular banana subgroups. Thus, although the microsatellites that were used in the present study were useful for identifying mislabels in the TARS collection, they were apparently unable to make genetic distinctions that were possible with the above markers. The recent draft genome of *Musa* (D'Hont et al., 2012) and new resequencing platforms as well as the genotyping-by-sequencing reduced representation genome sequencing approach (Chan et al., 2014) should help identify genetic variation within *Musa* subgroups.

Germplasm Management

Several types of mistakes in the TARS collection were identified, including propagation mistakes in both the field and in vitro (Table S1). All putative mistakes identified via

genetic analysis were verified by examining the source plants in the field and by growing plants from tissue culture in the field. In a few instances, morphological traits in the field allowed errors to be easily confirmed. For example, two erroneous plants of 'Hua Moa' (Maoli subgroup, AAB) were recognized in the Senorita accession (a tall Pome subgroup, AAB) field plot after they produced a bunch and fruit. However, morphology of different Cavendish cultivars could not be used to identify mislabels in this subgroup because of their very similar appearance. The markers also identified propagation mistakes in the in vitro tissue culture collection where no visible differences were apparent among accessions. All "mistakes" were verified in the field plantings, and all tissue culture accessions with questionable genetic integrity were reintroduced to the in vitro collection from true-to-type mother plants.

The microsatellite markers were useful in clarifying the ploidy and genomic composition of accessions in the collection. Irish et al. (2009) screened 135 accessions in the TARS collection for ploidy (flow cytometry) and genomic composition (PCR RFLPs). The present microsatellite evaluation was conducted because of an increase in accessions at TARS and the limited information that was provided by the previous work. Differences in ploidy and genomic composition were noted when data generated by flow cytometry and PCR RFLPs in Irish et al. (2009) were compared with the data generated in the current analysis (i.e., groupings of the PCoA and cluster analysis) (Table S1**)**. The discrepancy in ploidy and genomic composition between evaluations might have been due to the low resolution of the PCR RFLP technique and/or an erroneous original report. de Jesus et al. (2013) also had difficulty determining genomic compositions in triploids and tetraploid hybrids when using the PCR RFLP technique. In five cases in the present study ploidy and genomic composition did not agree with previous results (Irish et al., 2009), and in four of these cases a mistake in genetic integrity was also identified (Table S1). Considering these findings, the genetic integrity of *Musa* germplasm collections should be verified every five to seven years.

The markers used in the present study were practical and could be implemented in other significant ex situ genebanks for regular monitoring of genetic integrity. Genebanks that could benefit include the largest ex situ germplasm collection maintained by Bioversity International at the International Transit Center in Leuven, Belgium, as well as one of the major ex situ *Musa* spp. field collections conserved at the Centre Africain de Recherche sur Bananiers et Plantains in Njombe, Cameroon.

The GMGC-recommended panel of microsatellite markers worked well in the present study. When comparing results reported by Creste et al. (2003, 2004) and de Jesus et al. (2013), both groups obtained similar results in ordination. However, direct comparison of clonal profiles was not possible, as different sets of microsatellites were employed. Allelic profiles generated in the present study matched well and were directly comparable to those employed by Christelova et al. (2011b) and Hippolyte et al. (2012). A standardized set of microsatellite markers (e.g., the GCP developed and GMGC-recommended markers) would be useful when characterizing *Musa* germplasm, as they would enable clone comparisons, help examine synonymy, and allow comparisons to be made among laboratories. Although allele sizes at microsatellite loci may vary slightly when using different platforms (e.g., two to three base pair shifts), heterozygosity at given loci should not change. In addition, internal reference standards from specific subgroups (e.g., Cavendish and plantains) should be included for standardization purposes (Hippolyte et al., 2012).

The existence of numerous synonyms for a given cultivar and the problems that they can cause when one attempts to compare results in different banana-growing regions are well documented (Ploetz et al., 2007). As Rossel (1998, p. 2) indicated, "Musa cultivars are usually known under a multitude of vernacular names, reflecting the linguistic diversity rather than the Musa diversity in a given area. This makes the collection and identification of plantain cultivars very difficult and not only presents an obstacle in the communication and exchange of material between researchers, but also makes the maintenance of collections very costly in terms of space, time and money." The GCP-developed and GMGC-recommended panel of microsatellite markers could be used to address some of the above problems.

When compared with similar work (Creste et al., 2003, 2004; Christelova et al., 2011b; de Jesus et al., 2013; Hippolyte et al., 2012), genetic diversity within the current germplasm collection at the TARS seems to be high and encompasses much of the diversity found in cultivated *Musa*. Few "wild" genotypes or accessions are included in the TARS collection, as the collection's primary purpose is to introduce and evaluate cultivars for important agronomic traits. The present results provided an estimate of genetic diversity in the collection, helped identify gaps, and detected mislabeled accessions in both field and in vitro collections; they clearly complement continued evaluations of this germplasm in the field. The accessions and their microsatellite fingerprint profiles are available through the Germplasm Resource Information Network (GRIN) database www.ars-grin.gov/.

Supplementary Information

Two tables are submitted as supplementary information. Supplementary Table S1 contains information regarding list of the *Musa* spp. accessions evaluated and their genomic compositions (as determined in the current study) and Supplementary Table S2 provides additional information on allele distributions across ploidy levels.

Acknowledgments

The authors would like to thank Mr. Carlos Rios, Ms. Yaleidis Mendez, and Ms. Stephanie Cosme (information technologist and research technicians) for their help in generating data. In addition, the authors would like to acknowledge Courtney Irish as well as Drs. Dapeng Zhang and Osman Gutierrez for their critical internal review of the manuscript.

References

Agarwal, M., N. Shrivastava, and H. Padh. 2008. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep. 27:617–631. doi:10.1007/s00299-008-0507-z

Agoreyo, B.O., K.D. Golden, and S.E. Brown. 2008. Analysis of genetic variability among plantain cultivars (*Musa paradisiaca* L.) using arbitrarily primed PCR technique. Afr. J. Biotechnol. 7:1041–1045.

Amorim, E.P., P.H. Silva, C.F. Ferreira, V.B.O. Amorim, V.J. Santos, and A.D. Vilarinhos. 2012. New microsatellite markers for banana (*Musa* spp.). Genet. Mol. Res. 11:1093–1098. doi:10.4238/2012.April.27.8

Barbara, T., C. Plama-Silva, G.M. Paggi, F. Bered, M.F. Fay, and C. Lexer. 2007. Cross-species transfer of nuclear microsatellite markers: Potential and limitations. Mol. Ecol. 16:3759–3767. doi:10.1111/j.1365-294X.2007.03439.x

Bhat, K.V., R.L. Jarret, and Z.-W. Liu. 1994. RFLP characterization of Indian *Musa* germplasm for clonal identification and classification. Euphytica 80:95–103. doi:10.1007/BF00039303

Chan, A.P., X. Perrier, J. Sardos, J. Christophe, J.-P. Jacquemoud-Collet, N. Roux, et al. 2014. Using genotyping-by-sequencing to understand *Musa* diversity. In: Abstracts, Plant and Animal Genome, San Diego, CA. 10–15 Jan. 2014—P449.

Christelova, P., M. Valarik, E. Hribova, E. De Lange, and J. Dolezel. 2011a. A multi gene sequence-based phylogeny of the Musaceae (banana) family. BMC Evol. Biol. 11:103. doi:10.1186/1471-2148-11-103

Christelova, P., M. Valarik, E. Hribova, I. Van den Houwe, S. Channeliere, N. Roux, et al. 2011b. A platform for efficient genotyping in *Musa* using microsatellite markers. AoB PLANTS. doi:10.1093/aobpla/plr024

Creste, S., A. Tulmann, S. de Oliveira, and A. Figueira. 2003. Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. Euphytica 132:259–268. doi:10.1023/A:1025047421843

Creste, S., A. Tulmann, R. Vencovsky, S. de Olivera, and A. Figueira. 2004. Genetic diversity of *Musa* diploid and triploid accessions from the Brazilian banana breeding program estimated by microsatellite markers. Genet. Resour. Crop Evol. 51:723–733. doi:10.1023/B:GRES.0000034578.37951.c4

Crouch, J.H., H.K. Crouch, H. Constandt, A. Van Gysel, P. Breyne, M. Van Montagu, et al. 1998. Segregation of microsatellite loci in haploid and diploid gametes of *Musa*. Crop Sci. 38:211–217. doi:10.2135/cropsci1998.0011183X003800010035x

de Jesus, O.N., S. de Oliveira, E.P. Amorim, C.F. Ferreira, J.M. Salabert de Campos, G. de Gaspari-Silva, et al. 2013. Genetic diversity and population structure of *Musa* accessions in ex situ conservation. BMC Plant Biol. 13:41. doi:10.1186/1471-2229-13-41

De Lange, E., M. Pillay, A. Tenkounano, and R. Swennen. 2005. Integrating morphological and molecular taxonomy in *Musa*: The African plantains (*Musa* spp. AAB group). Plant Syst. Evol. 255:225–236. doi:10.1007/s00606-005-0346-0

D'Hont, A., F. Denoeud, J.-M. Aury, F.-C. Baurens, F. Carreel, O. Garsmeur, et al. 2012. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488:213–217. doi:10.1038/nature11241

Dodds, K.S., and N.W. Simmonds. 1948. Sterility and parthenocarpy in hybrids of *Musa*. Heredity 2:101–117. doi:10.1038/hdy.1948.6

Dolezel, J., M.A. Lysak, I. Van den Houwe, M. Dolezelova, and N. Roux. 1997. Use of flow cytometry for rapid ploidy determination in *Musa* species. Informusa. 6:6–9.

Escalant, J.V., S. Sharrock, and E. Frison. 2002. The genetic improvement of *Musa* using conventional breeding, and modern tools of molecular and cellular biology. IPGRI, Rome, Italy. p. 17.

Fang, D.Q., M.L. Roose, R.R. Krueger, and C.T. Federici. 1997. Fingerprinting trifoliate orange germplasm accessions with isozymes, RFLPs, and inter-simple sequence repeat markers. Theor. Appl. Genet. 95:211–219. doi:10.1007/s001220050550

FAO. 2013. Standards for plant genetic resources for food and agriculture. FAO, Rome, Italy.

FAOSTAT. 2013. http://faostat.fao.org/site/291/default.aspx (accessed 1 Dec. 2012).

Gupta, P.K., J. Kumar, R.R. Mir, and A. Kumar. 2010. Marker-assisted selection as a component of conventional plant breeding. In: J. Janick, editor, Plant breeding reviews. Vol. 33. John Wiley & Sons, Hoboken, NJ.

Häkkinen, M. 2013. Reappraisal of sectional taxonomy in *Musa* (*Musaceae*). Taxon 62:809–813. doi:10.12705/624.3

Hart, D.L., and A.G. Clark. 2007. Principles of population genetics. Sinauer Associates, Inc. Publishers, Sunderland, MA.

Heslop-Harrison, J.S., and T. Shwarzacher. 2007. Domestication, genomics and the future for banana. Ann. Bot. (Lond.) 100:1073–1084. doi:10.1093/aob/mcm191

Hippolyte, I., F. Bakry, M. Seguin, L. Gardes, R. Rivallan, A.-M. Risterucci, et al. 2010. A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. BMC Plant Biol. 10:65. doi:10.1186/1471-2229-10-65

Hippolyte, I., C. Jenny, L. Gardes, F. Bakry, R. Rivallan, V. Pomies, et al. 2012. Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. AoB PLANTS. doi:10.1093/aob/mcs010

Hokanson, S.C., A.K. Szewc-McFadden, W.F. Lamboy, and J.R. McFerson. 1998. Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus* × *domestica* Borkh. Core subset collection. Theor. Appl. Genet. 97:671–683. doi:10.1007/s001220050943

IPGRI. 1996. Descriptors for banana (*Musa* spp.). IPGRI, Rome, Italy. p. 55.

Irish, B.M., A. Crespo, R. Goenaga, and R. Neidz. 2009. Ploidy level and genomic composition of the USDA–ARS Tropical Agriculture Research Station *Musa* sp. germplasm collection. J. Agric. Univ. PR 93:1–22.

Irish, B.M., R. Goenaga, D. Zhang, R. Schnell, J.C. Motamayor, and S. Brown. 2010. Microsatellite fingerprinting of the USDA–ARS Tropical Agriculture Research Station cacao (*Theobroma cacao*) germplasm collection. Crop Sci. 50:656–667. doi:10.2135/cropsci2009.06.0299

Kaemmer, D., D. Fischer, R.L. Jarret, F.-C. Baurens, A. Grapin, D. Dambier, et al. 1997. Molecular breeding in the genus *Musa*: A strong case for STMS marker technology. Euphytica 96:49–63. doi:10.1023/A:1002922016294

Lagoda, P.J., J.L. Noyer, D. Dambier, F.-C. Baurens, A. Grapin, and C. Lanaud. 1998. Sequence tagged microsatellite site (STMS) markers in the Musaceae. Mol. Ecol. 7:657–666. doi:10.1046/j.1365-294X.1998.00340.x

Letunic, I., and P. Bork. 2011. Interactive Tree of Life v2: Online annotation and display of phylogenetic trees made easy. Nucleic Acids Res. 39:W475–W478 10.1093/nar/gkr201. doi:10.1093/nar/gkr201

Liu, K., and S.V. Muse. 2005. PowerMarker: An integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

Li, L.F., M. Häkkinen, Y.M. Yuan, G. Hao, and X.J. Ge. 2010. Molecular phylogeny and systematics of the banana family (Musaceae) inferred from multiple nuclear and chloroplast DNA fragments, with a special reference to the genus *Musa*. Mol. Phylogenet. Evol. 57:1–10. doi:10.1016/j.ympev.2010.06.021

Lorenzen, J., S. Hearne, G. Mbanjo, M. Nyine, and T. Close. 2011. Use of molecular markers in banana and plantain improvement. In: I. Van den Bergh et al., editors, Proc. Int'l ISHS-ProMusa Symp. on Global Perspectives on Asian Challenges, Acta Hort. 897, ISHS 2011.

Lucas, M.R., B. Huynh, J.D. Ehlers, P.A. Roberts, and T.J. Close. 2013. High-resolution single nucleotide polymorphism genotyping reveals a significant problem among breeder resources. Plant Gen. 6:1–5. doi:10.3835/plantgenome2012.08.0020

Messmer, M.M., A.E. Melchinger, J. Boppenmaier, R.G. Herrmann, and E. Brunklaus-Jung. 1992. RFLP analyses of early-maturing European maize germplasm: Genetic diversity among flint and dent inbreds. Crop Sci. 32:1301–1309. doi:10.2135/cropsci1992.0011183X003200060001x

Mohammadi, S.A., and B.M. Prasanna. 2003. Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop Sci. 43:1235–1248. doi:10.2135/cropsci2003.1235

Nelson, S.C., R.C. Ploetz, and A.K. Kepler. 2006. *Musa* species (bananas and plantains), ver. 2.2. In: C.R. Elevitch, editor, Species profiles for Pacific island agroforestry. Permanent Agriculture Resources (PAR), Hōlualoa, Hawai'i. www.traditionaltree.org, accessed 16 June 2014.

Noyer, J.L., S. Causse, K. Tomekpe, A. Bouet, and F.C. Baureans. 2005. A new image of plantain diversity assessed by SSR, AFLP and MSAP markers. Genetica (The Hague) 124:61–69.

Nwakanma, D.C., M. Pillay, B.E. Okoli, and A. Tenkouano. 2003. PCR-RFLP of the ribosomal DNA internal transcribed spacers (ITS) provides markers for the A and B genomes in *Musa* L. Theor. Appl. Genet. 108:154–159. doi:10.1007/s00122-003-1402-1

Ortiz, R., S. Madsen, and D. Vuylstke. 1998. Classification of African plantain landraces and banana cultivars using phenotypic distance index of quantitative descriptors. Theor. Appl. Genet. 96:904–911. doi:10.1007/s001220050818

Okoro, P., A.A. Shaibu, G. Ude, B.A. Olukolu, I. Ingelbrecht, A. Tenkouano, et al. 2011. Genetic evidence of developmental components of parthenocarpy in apomictic *Musa* species. J. Plant Breed. Crop Sci. 3:138–145.

Peakall, R., and P.E. Smouse. 2006. GenAlEx 6: Genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6:288–295. doi:10.1111/j.1471-8286.2005.01155.x

Perrier, X., F. Bakry, F. Carreel, C. Jenny, J-P. Horry, V. Lebot et al. 2009. Combining biological approaches to shed light on evolution of edible bananas. Ethno. Res. App. 7:199–216.

Perrier, X., E. De Lange, M. Donohue, C. Lentfer, L. Vrydaghs, F. Bakry, et al. 2011. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. Proc. Natl. Acad. Sci. USA 108:11311–11318. doi:10.1073/pnas.1102001108

Pillay, M., E. Ogundiwin, D.C. Nwakanma, G. Ude, and A. Tenkouano. 2001. Analysis of genetic diversity and relationships in East African banana germplasm. Theor. Appl. Genet. 102:965–970. doi:10.1007/s001220000500

Ploetz, R.C., A.K. Kepler, J. Daniells, and S.C. Nelson. 2007. Banana and plantain—an overview with emphasis on Pacific island cultivars, ver. 1. In: C.R. Elevitch, editor, Species profiles for Pacific island agroforestry. Permanent Agriculture Resources (PAR), Hōlualoa, Hawai'i. www.traditionaltree.org, accessed 16 June 2014.

Provan, J., W. Powell, and R. Waugh. 1996. Microsatellite analysis of relationships within cultivated potato (*Solanum tuberosum*). Theor. Appl. Genet. 92:1078–1084.

Risterucci, A.M., I. Hippolyte, X. Perrier, L. Xia, V. Caig, M. Evers, et al. 2009. Development and assessment of diversity arrays technology for high-throughput DNA analyses in *Musa*. Theor. Appl. Genet. 119:1093-1103.doi:10.1007/BF00224052

Rosales, F., E. Aranud, and J. Coto. 1999. A catalogue of wild and cultivated bananas—a tribute to the work of Paul Allen. International Network for the Improvement of Bananas and Plantain, Montpellier, France.

Rossel, G. 1998. Taxonomic-linguistic study of the plantain in Africa. CNWS. Leiden University, Leiden.

Rowe, P., and F. Rosales. 1993. Diploid breeding at FHIA and the development of Goldfinger (FHIA-01). Infomusa. 2:9–11.

Silva, S.O., M.T. Souza-Junior, E.J. Alves, J.R.S. Silveira, and M.B. Lima. 2001. Banana breeding program at Embrapa. Crop Breed. Appl. Biotechnol. 1:399–436. doi:10.13082/1984-7033.v01n04a09

Simmonds, N.W., and K. Shepherd. 1955. Taxonomy and origins of cultivated bananas. J. Linn. Soc. Bot. 55:302–312. doi:10.1111/j.1095-8339.1955.tb00015.x

Thomson, R., I.J. Wang, and J.R. Johnson. 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. Mol. Ecol. 19:2184–2195. doi:10.1111/j.1365-294X.2010.04650.x

Ude, G., M. Pillay, D. Nwakanma, and A. Tenkouano. 2002. Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. Theor. Appl. Genet. 104:1239–1245. doi:10.1007/s00122-001-0802-3

Weising, K., H. Nybom, K. Wolf, and G. Kahl. 2005. DNA fingerprinting in plants. In: Principles, methods and applications. 2nd ed. CRC Press, Taylor and Francis Group, Boca Raton, FL. p. 444.

Wongniam, S., J. Somana, S. Swangpol, T. Seelanan, P. Chareonsap, S. Chadchawan, et al. 2010. Genetic diversity and species-specific PCR-markers from AFLP analyses and Thai bananas. Biochem. Syst. Ecol. 38:416–427. doi:10.1016/j.bse.2010.03.015

Zhang, D., E. Arevalo-Gardini, L.A. Motilal, V. Baligar, B. Bailey, L. Zuñiga-Cenades, et al. 2010. Dissecting genetic structure in farmer selections of *Theobroma Cacao* in the Peruvian Amazon: Implications for on farm conservation and rehabilitation. Trop. Plant Biol. 4:106–116. doi:10.1007/s12042-010-9064-z