ISSN 1479-2621
Plant Genetic Resources: Characterization and Utilization (2013) **11**(1); 50–61
doi:10.1017/S1479262112000354

Genetic diversity among INERA-Mulungu (DR Congo) *Musa* spp. germplasm and their relatedness to those in Tanzania using numerical taxonomy

D. B. Nzawele*, C. L. Rweyemamu and A. P. Maerere

Department of Crop Science and Production, Faculty of Agriculture, Sokoine University of Agriculture, PO Box 3005, Chuo Kikuu, Morogoro, Tanzania

Received 17 July 2012; Accepted 29 September 2012 – First published online 29 November 2012

Abstract

Bananas and plantains (*Musa* spp.) constitute staple food for over 20 million people in Democratic Republic of Congo (DRC). Since 1960, DRC is considered as a secondary centre of plantain diversification with few unknown accessions kept in the INERA-Mulungu genebank. Through similarity coefficients, cluster (unweighted pair group method with arithmetic mean, single, complete, sequential, agglomerative, hierarchical and nested design/clustering procedure) and/or multivariate analyses, numerical morpho-taxonomy has established that this diversity is composed of 37 different accessions. Each accession expressed 98 characters among the 401 possible character states, thus providing 39,298 feature patterns (data points). The 98 characters included 32 vegetative and 66 male and female inflorescences. The accessions were clustered into three genomic groups (AAA, AAB and AABB). Subjective classification ascertained nine subgroups: AAB-Silk, AAB-Pome, AAB-Plantain, AABB-Pisang Awak, AAA-Cavendish, AAA-Ibota, AAA-Gros Michel, AAA-Green-Red and AAA-Lujugira-Mutika. Three subgroups were further divided into nine clone sets which consisted of: Dwarf and Giant Cavendish, French and Horn Plantains, and Musakala, Nfuuka, Nakitembe, Nakabululu and Beer/Mbidde within Lujugira-Mutika. Numerical morpho-taxonomy effectively indicated a relationship between the DRC and Tanzania's *Musa* diversity. For example, the accessions 'Kamaramasengi' and 'Isangi' were found to be similar to 'Kisukari' (AAB-Silk) and 'Ngego I' (AAB-French) common in the Tanzanian Southern Highland. Likewise, the accessions Kimalindi-fupi, Kimalindi-ndefu and Jamaica of Tanzania were duplicates of Bakurura (Kigurube), Cavendish of Butuza and Gros Michel in DRC, respectively. Moreover, numerical morpho-taxonomy confirmed the pedigree of AAB-Prata (Cibwalo) in FHIA 17 and FHIA 23 and the closeness of the ancestors of Yangambi Km5 and Gros Michel. Furthermore, numerical morpho-taxonomy established AA-Mshale malembo as one of the AAA-Lujugira-Mutika parents. Molecular investigations are finally required to confirm the genomes.

Keywords: banana; genebanks (DRC; Tanzania); genomic group; numerical morpho-taxonomy; plantain

Introduction

Bananas and plantains (*Musa* spp.) are produced in more than 120 tropical and subtropical countries, mainly by small-scale farmers (Pestana *et al.*, 2011). *Musa* is the fourth most important food commodity in the world

* Corresponding author. E-mail: dowiya@yahoo.com

(after rice, wheat and maize), providing a vital source of carbohydrates, fibre, minerals and vitamins (Miller *et al*., 2010; Christelová *et al*., 2011). *Musa* has a primary centre of domestication in South-East Asia, and has travelled to Africa through human migrations (De Langhe *et al*., 2005, 2010; Blench, 2009).

In the Democratic Republic of Congo (DRC), *Musa* constitutes the third staple food after *Manihot esculenta* Crantz and *Zea mays* L., but second after cassava in eastern DRC (Mobambo and Naku, 1993; Nzawele *et al*., 2009). The population of DRC is about 60 millions, of which 30% grow and integrate bananas in various farming and livestock production systems (Nzawele *et al*., 2010). In contrast, Tanzania has an annual production between 750,000 and 820,000 tonnes, ranking second in East Africa after Uganda. *Musa* constitutes the major staple food for 15–20% of the Tanzanian population (Msogoya *et al*., 2006; Maerere *et al*., 2010a).

For plantains, the secondary centre of diversity is in America and Africa, particularly in the areas of Yangambi in the rainforest basin of DRC (Mobambo *et al*., 2010); where De Langhe (1961) initially described 56 cultivars which did not exist in West Africa and elsewhere. De Langhe (1961) studied DRC's *Musa* using classical taxonomy (inflorescence and plant size) which showed no correlation with *Musa* diversity in the world (Stover and Simmonds, 1991). The reason being that two characters do not generate genetic variability, unlike numerical taxonomy using life-cycle characters as reliable phenetic markers. Following De Langhe (1961), information on the Research Institute (INERA-Mulungu) *Musa* genebank was provided by Sebasigari (1987) who partially classified the diversity into beer, boiling, flour and sweet bananas. To date, no research on DRC's *Musa* has been reported using numerical taxonomy. Conversely, several studies have been conducted on Tanzania's *Musa* (Stover and Simmonds, 1991; Evers, 1992; De Langhe *et al*., 2001).

Musa belong to the family *Musaceae* of the order *Zingiberales*. Most of the cultivated bananas are diploid (AA and AB) or triploid (AAA, AAB and ABB) originated from intra- and interspecific hybridizations between seed-bearing *Musa acuminata* (A genome) and *Musa balbisiana* (B genome) (Simmonds and Shepherd, 1955). The AAB group is subdivided into subgroups (Plantains, Mysore, Silk or Pome) through subjective classification using easily observable characters (De Langhe and Valmayor, 1980; Swennen and Vuylsteke, 1987; Swennen, 1990). Plantains are likewise subdivided into three clone sets (French, False Horn and Horn; Tezenas Du Montcel *et al*., 1983). AAA-Lujugira-Mutika is coined as a subgroup diversified through somaclonal mutation and similarly subdivided into five clone sets (Beer/Mbidde, Nakabululu, Musakala, Nakitembe and Nfuuka) (Karamura, 1999). This study aimed at determining the genomic groups, subgroups and clone sets of banana accessions kept in the INERA-Mulungu genebank and their correlation with known banana accessions in Tanzania. This correlation will allow the integration of DRC materials into the world's *Musa* diversity through reliable numerical taxonomic methods.

Materials and methods

The eastern DRC *Musa* spp. genebank was established in 1961 at the National Institute for Agronomic Study in the Belgian Congo (Institut National pour l'Etude Agronomique au Congo Belge) and maintained up to date at the National Agriculture Research Institute, Centre of Mulungu (INERA-M). The *Musa* spp. genebank at the INERA-M is located at 02°20′167″S and 028°47′541″E at 1686 m above sea level (masl) on a fertile soil described as volcanic soil by Pecrot (1958). The mean annual rainfall is around 1600 mm with a year-long growing season. There is currently no meteorological facility at the INERA-M apart from a rain gauge. The 1961 report shows that the mean temperature around the collection was 19°C, whereas the report of World (2012) shows, currently, a slight increase in the mean temperature of 19.7°C.

The Tanzania *Musa* spp. genebank was established in 2002 at the Horticulture Unit of Sokoine University of Agriculture (SUA) located in the plateau zone of Morogoro Urban District. The site is located between 06°50′ and 06°45′S, 37°35′ and 37°40′E and at 500 masl and experiences annual temperatures of 16–34°C (Maerere *et al*., 2010b). Rainfall distribution is bimodal with short rains between October and December followed by a dry spell starting from January to February. A long rainy season is between March and May while the dry season is from June to late October. According to Mtengeti (2008), the soil is kaolinitic well-drained clay. Due to climate change, the annual rainfall has decreased from an average of 2000 to 800 mm (Msogoya *et al*., 2010).

Forty-one accessions in the INERA-M genebank subdivided into beer (08), boiling (12), flour (08) and sweet (13) bananas were used in this study. The 13 sweet banana accessions were 'Cibwalo/Prata', 'Kamaramasengi/Kamelamasengi', 'Dwarf Cavendish', 'Bakurura/Kigurube', 'Lakatan(i)', 'Americani', 'Chindege/Naine de chine', 'Poyo', 'Cavendish B' (of Butuza), 'Chisukari', 'Chisukari red', 'Nyakitembe' and 'Gros Michel'. The eight flour banana accessions were 'Isangi', 'Chibulanana I' (one hand), 'Chibulanana II' (two hands), 'Chibulanana III' (three hands), 'Chibulanana VI'

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

(six hands), 'Walungu 16', 'FHIA 03' and 'Nyaluvu'. The 12 boiling banana accessions were 'Kagera masisi', 'Mbwazirume I', 'Ingagara', 'Inyamico', 'Incakara/Barhabesha', 'Igisahira', 'Muhuna binyoko', 'Inzirabushera', 'Inconnu INERA' (with yellow male bud), 'Muhorogo', 'Nyakitengwa' and 'Inyoya'. The eight beer banana accessions were 'Yangambi Km5', 'Nsinabuhaka', 'Ndundu', 'Munyamimba', 'Nshikazi', 'Intokatoke', 'Impysi' and 'Intembe'.

In the Tanzania genebank (SUA), there were 18 accessions characteristically divided into dessert (eight), roasting/cooking (seven), beer (one) and East African Highland cooking (two) bananas, which were used in this study. The seven accessions used as dessert included 'Yangambi Km5' (AAA-Ibota), 'Kisukari' (AAB-Silk), 'Kimalindi-fupi' (AAA-Dwarf Cavendish), 'Mtwike' (AAA-Cavendish), 'Kimalindi-ndefu' (AAA-Giant Cavendish), 'Jamaica' (AAA-Gros Michel), 'FHIA 17' and 'FHIA 23' (AAAA) (Stover and Simmonds, 1991; Maerere *et al.*, 2010b). The seven accessions used as roasting/cooking were 'Ngego I', 'Ngego Halisi', 'Mzuzu' (AAB-French Plantain), 'Unyoya' (ABB-Pisang Awak), 'Bokoboko' (ABB-Bluggoe), 'Mwajunjila' (with yellow male bud) and 'Mshale malembo' (AA-Pisang lilin) (Evers, 1992). The two AAA-EAHB cooking accessions were 'Embwailuma' and 'Bukoba'. The beer accession was 'Muhowe' belonging to AAA-EAHB (Stover and Simmonds, 1991; Evers, 1992; Maerere *et al.*, 2010b).

These *Musa* accessions were assigned to genomic groups, subgroups and clone sets from a scoring method and subjective classification based on quickly observable characters determined by Simmonds and Shepherd (1955), De Langhe and Valmayor (1980), Swennen and Vuylsteke (1987), Tezenas Du Montcel *et al.* (1983) and Karamura (1999) which are accepted worldwide. The numerical morpho-taxonomic characterization used 99 characters (Table S1, available online only at http://journals.cambridge.org) described by IPGRI (1996). The 99 characters encompassed the 15 for genomic group determination devised by Simmonds and Shepherd (1955). *Musa* characterization was done in the INERA-M genebank (Eastern DRC) from 16 January 2008 to 20 January 2009. Similarly, the Tanzania's *Musa* characterization for the purpose of their relatedness to those of eastern DRC was done from June to July 2008 and from July to November 2010. Two qualitative character states, green cigar or red purple cigar, were coded as binary, i.e. by scoring 1 for green cigar and 2 for red purple cigar. Ordered multistate characters were coded as a series of discrete attributes. Low intensities of a character were given lower scores so that the scores would increase with increasing intensity of the character. Non-ordered characters standing on their own, i.e. without an intensity level, were coded as a series of discrete states according to IPGRI (1996). Missing data were coded 999 as recommended by Rohlf (1993, cited by Karamura (1999)).

The product-moment correlation and distance coefficients were used to assess similarity among the banana accessions. The genetic similarity matrices were then used to construct the dendrogram (phenogram) with the single-linkage, complete-linkage and unweighted pair group method with arithmetic mean (UPGMA) algorithms, employing the sequential, agglomerative, hierarchical and nested design clustering procedure (Sneath and Sokal, 1973). Cophenetic correlation (coefficients) and principal component analysis (PCA) including the percentage trace and discriminant analyses were conducted using the NT-SYS package (version 2.1) according to Opara *et al.* (2010).

Results

The diversity of *Musa* kept in the INERA-M genebank was composed of 41 accessions (*prior to* numerical taxonomic analysis) which expressed, in total, 401 possible phenotypic expressions, hereafter referred to as 'character cases/states'. The appearance of the leaf upper surface was constant across the diversity, and hence not used in the analysis. So, each accession expressed 98 characters among the 401 character states, thus providing a total of 39,298 feature patterns (data points). The 98 characters (Table 1) were structurally composed of 32 vegetative and 66 male and female inflorescences (Table 1).

The contributions of qualitative and quantitative characters (Table 1) were of 78 and 22%, respectively, in clustering the diversity. Using the 98 characters, the numerical taxonomy (multivariate methods) effectively clustered the 41 accessions with regard to their genomic groups, subgroups and clone sets supporting subjective classification (Tables 2 and 3). The numerical taxonomy method used in this study related well the 41 accessions from DRC to the 18 known banana accessions from Tanzania maintained in the SUA genebank as revealed by the clustering methods and PCAs.

Diversity from the INERA-Mulungu genebank

The 41 accessions were grouped into three main clusters (A, B and C) as shown in the phenograms (Figs 1, S1 and S2, available online only at http://journals.cambridge.org). The first cluster was composed of two subclusters that were exclusively of heterogenomic groups from natural and artificial interspecific crosses between *M. acuminata* (A genome) and *M. balbisiana* (B genome). The results showed that the first subcluster (A 1) from 'Chibulanana I' to 'Chibulanana II' was made of five accessions belonging to the AAB-Horn Plantain

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

Table 1. The structure of the 98 characters expressed per banana cultivar

	Vegetative			Male inflorescence and fruit				
Character	Pseudostem	Sucker	Leaves	Male bud	Bract	Male flower	Fruit	Grand total
Type								
Quantitative	2	2	5	5	2	1	5	22
Qualitative	5	0	18	9	11	21	12	76
Subtotal	7	2	23	14	13	22	17	98
States								
Binary	1	0	5	3	0	1	2	12
Multistate ordered	5	2	17	5	5	12	8	54
Multistate unordered	1	0	1	6	8	9	7	32
Subtotal	7	2	23	14	13	22	17	98

clone set on the basis of subjective classification (Table 2). The five accessions were tied with strong similarity coefficients (R_s) of 0.83 or dissimilarity coefficients (R_d) of 0.63. The insignificant difference observed between the accessions was composed of qualitative characters and hence they were considered as different accessions. The second subcluster (A 2) was composed of AABB-Nyaluvu linked to AABB-FHIA 03 with a high R_d of 0.86. Both accessions were found to be linked to AAB-French 'Isangi' with a weak correlation coefficient (R_s) of 0.37. Based on the difference in character types, the banana accessions were considered to be different.

The second cluster (B) was exclusively found to be of triploid homogenomic composition from *M. acuminata* (A genome) and was subdivided into two subclusters (i.e. B 1 and B 2). The first subcluster (B 1) was composed of twelve accessions, in which the accession 'Inyoya' was found to be a duplicate of 'Igisahira' with a strong R_s of 0.88, with dissimilarity only among the quantitative characters (petiole, leaf and fruit). The 12 accessions were made of eight cooking and four beer types. The subjective classification (Table 3) showed that the first two cooking types were of the Musakala clone set (i.e. Muhorogo and Incakara), followed by three types of the Nakitembe clone set (i.e. Muhuna binyoko, Nyakitengwa and Mbwazirume). The last three accessions 'InconnuINERA (UnknwonINERA), 'Inzirabushera' and 'Igisahira' belonged to the Nfuuka clone set. The beer banana types were composed of two accessions (i.e. Nshikazi and Ndundu) belonging to the Nfuuka clone set and the other two belonging to the Nakitembe clone set which were 'Nsinabuhaka' and 'Munyamimba'. Beer bananas of the Nakitembe clone set did pair well ($R_s = 0.53$; Figs 1, S1 and S2, available online only at http://journals.cambridge.org) with their homologues from the cooking types. Six phenetic classifications were in line with subjective classifications of the accessions into the existing genomic groups, subgroups and clone sets (Tables 2 and 3).

The last subcluster (B 2) was composed of 14 accessions in which 'D-Cavendish', 'Americani' and 'Ingagara' were found to be duplicate of 'Bakurura/Kigurube', 'Lakatan' and 'Inyamico', respectively. The subjective classification showed that five accessions (i.e. Chindege, Bakurura/Kigurube, Lakatan, Poyo and Cavendish of Butuza; Table 2) were of the AAA-Cavendish subgroup linked to 'Chisukari' and 'Chisukari red' that belonged to the AAA-Green-Red subgroup. The Cavendish subgroup was found to be tied to Gros Michel (AAA), five AAA-EAHB and one outline of AAA-Green-Red (Nyakitembe) accessions with a slightly strong correlation coefficient ($R_s = 0.60$) of similarity. The five EAHB accessions were found to be composed of 'Kagera masisi' (i.e. Nakabululu clone set) and 'Inyamico' (i.e. Musakala clone set) that were linked to three beer types [that were Intokatoke, Impysi and Intembe (i.e. Nfuuka clone set); Table 3]. The second subcluster (B 2; Fig. 1) was weakly ($R_s = 0.42$) linked to the first subcluster (B 1). The third cluster (C) was composed of four accessions tied with moderate correlation coefficients ($R_s = 0.51$). The AAB-Pome accession 'Cibwalo' correlated moderately with the similarity (R_s) value of 0.59 with the AAB-Silk accession 'Kamaramasengi' (also named 'Namasugampene') and both tied with 'Yangambi Km5' (AAA-Ibota).

Relatedness between Musa *spp. from the INERA-M-DRC and SUA-Tanzania genebanks*

As shown in the phenograms (Figs 2 and S3, available online only at http://journals.cambridge.org) from the similarity and dissimilarity matrices, the 57 accessions were grouped into three main clusters marked as A, B and C, which also supported the subjective classification (Tables 2 and 3). The three main clusters were identified

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

Table 2. Tie from numerical taxonomy between genomic and subjective classification

Genomic taxa	Subjective classification[a]	SC[b]	DC[c]	Name of accessions	Conclusion
AAB	Pome (Prata)	–	–	Cibwalo	1 different clone
AAB	Silk/Kamara	0.93	0.32	Kamaramasengi and Kisukari-URT	2 are duplicates/ramets
AAB	French Plantain	0.93	0.37	Isangi and Ngego I-URT	2 are duplicates
AAA	Dwarf Cavendish	0.93	0.34	D-Cavendish[d], Kimalindi-fupi-URT and Bakurura	3 are duplicates
AAA	Giant Cavendish	0.96	0.29	Lakatan(i) and Americani	2 are duplicates
AAA	Ibota	0.75	0.70	Yangambi Km5 of DRC and Yangambi Km5-URT	2 are duplicates
AAB	Horn Plantain	0.82	0.57	Chibulanana I, II, III, VI and Walungu 16	5 different clones
AAB	French Plantain	0.70	0.77	Ngego Halisi and Mzuzu UR[e]	2 different clones
AABB, ABB	Pisang awak	0.69	0.73	Nyaluvu and Unyoya-URT	2 different clones
ABB	Bluggoe			Bokoboko-URT	1 different clone
AABB	FHIA			FHIA 03	1 different clone
AAAA	FHIA	0.72	0.75	FHIA 17 and FHIA 23 (URT)	2 different clones
AAA	Dwarf Cavendish			Chindege	1 different clone
AAA	Giant Cavendish			Poyo or Mtwike-URT, Cavendish of Butuza or Kimalindi-ndefu-URT	2 different clones
AAA	Green-Red			Chisukari[f], Chisukari red and Nyakitembe	3 different clones
AAA	Gros Michel			Gros Michel or Jamaica-URT	1 different clone
AA	Pisang lilin			Mshale malembo-URT	1 different clone
6 groups	13 subgroups				27 clones

[a] At the subgroup or clone set level. [b] Similarity coefficient (Fig. S3). [c] Dissimilarity coefficient (Fig. 2). [d] Dwarf Cavendish. [e] Both of United Republic of Tanzania (URT). [f] Also written as 'Kisukari' in DRC which is called 'Mzungu mweupe' or 'Buki' in URT.

as the 'A' cluster from both matrices starting with the accession 'Chibulanana I' to 'FHIA 03' having three subclusters that were exclusively of heterogenomes from natural and artificial interspecific crosses between *M. acuminata* (A genome) and *M. balbisiana* (B genome).

The relatedness was observed only within the second and third subclusters. The second subcluster A 2 (Fig. 2) (i.e. from 'Unyoya' to 'Kisukari') was made of five accessions. There were two accessions, 'Unyoya' and 'Bokoboko', from Tanzania related to 'Nyaluvu' from DRC that were of the ABB and AABB genomic groups, respectively (Table 2). The last two accessions were of AAB-Silk, which also showed that the DRC's accession 'Kamaramasengi' was highly similar to the Tanzania's 'Kisukari' with a R_s of 0.93 (Fig. S3, available online only at http://journals.cambridge.org). While a lesser R_d (Fig. 2) of 0.32, between these accession pairs (Kamaramasengi–Kisukari) was observed (Table 2). The similarity between 'Kamaramasengi' and 'Kisukari' was verified from six phenetic classifications (Figs 2, S3, S4 and S5, available online only at http://journals.cambridge.org), suggesting them as a ramet. On the other hand, the accession 'Cibwalo' (AAB) did not cluster with the varieties mentioned above.

The third subcluster A 3 (Figs 2 and S3, available online only at http://journals.cambridge.org) was made of four French Plantain accessions (i.e. Isangi, Ngego I, Ngego Halisi and Mzuzu) and AABB-FHIA 03 from both DRC and Tanzania genebanks (Table 2). The accession 'Ngego I', commonly grown in the Southern Highlands of Tanzania, strongly correlated with 'Isangi' from DRC with a R_s of 0.93 but had a weak R_d of 0.37, showing these as a ramet which was verified using six different clustering methods (Figs 2, S3, S4 and S5, available online only at http://journals.cambridge.org).

The second cluster (B) was exclusively of EAHB (i.e. beer and cooking types) and was composed of 16 accessions paired according to the overall relatedness of phenotypic expressions. The important information was that the accession 'Mbwazirume I' from DRC paired with 'Embwailuma' from Tanzania with a strong R_s of 0.87 (Fig. S3, available online only at http://journals.cambridge.org) but with a weak dissimilarity (R_d) of 0.47 (Fig. 2), suggesting these as a ramet (Table 3) as also shown in other phenetic classifications (Figs S3–S5, available online only at http://journals.cambridge.org). The differences were in quantitative characters of pseudostem and leaf. The accession 'Bukoba' from Tanzania was found to be within the same cluster with the DRC's 'Igisahira'.

The third cluster (C) was found to be made of dessert and EAHB that was subdivided into three subclusters: C 1, C 2 and C 3. The first subcluster (C 1) (Fig. 2) was composed of 11 accessions. The subjective classification showed that nine accessions named 'Chindege',

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

Table 3. Tie from numerical taxonomy between AAA-EAHB and subjective classification

Use of accessions	Clone set	SC[a]	DC[b]	Name of accessions	Conclusion
Cooking	Nakitembe	0.86	0.47	Mbwazirume I and Embwailuma-URT	2 are duplicates
Beer making	Beer/Mbidde			Nsinabuhaka, Ndundu, Munyamimba, Nshikazi, Intokatoke, Impysi, Intembe and Muhowe-URT	8 different clones
Cooking	Musakala	0.85	0.49	Ingagara and Inyamico	2 are duplicates
Cooking	Musakala	–	–	Incakara/Barhabesha, Bukoba and Muhorogo	3 different clones
Cooking	Nakitembe	–	–	Mwajunjila-URT, Muhuna binyoko and Nyakitengwa	3 different clones
Cooking	Nfuuka	0.86	0.47	Inyoya and Igisahira	2 are duplicates
Cooking	Nfuuka	–	–	Inzirabushera and In[c].INERA	2 different clones
Cooking	Nakabululu	–	–	Kagera masisi	1 different clone
	5 clone sets				20 clones

[a] Similarity coefficient (Fig. S3). [b] Dissimilarity coefficient (Fig. 2). [c] Inconnu means unknown.

'D-Cavendish', 'Kimalindi-fupi', 'Bakurura', 'Lakatan', 'Americani', 'Poyo', 'Mtwike' and 'Cavendish B' belonged to AAA-Cavendish (Table 2). The remaining two accessions (of Green-Red) were mentioned earlier. The accession named 'Kimalindi-fupi' (Dwarf Cavendish) from Tanzania was found to be a ramet of 'Bakurura/Kigurube' from DRC. The AAA-Cavendish accessions named 'Mtwike', 'Kimalindi-ndefu' and AAA 'Jamaica' from Tanzania were found to be 100% similar to 'Poyo', 'Cavendish B' and 'Gros Michel' from DRC in 2008 and were omitted in some analyses such as ordinations because Eigen vector analysis is based on the variation of character states between the accessions (Sneath and Sokal, 1973). The last subcluster (C 3) was a mixture of genomic groups in which the two accessions named AAA-EAHB 'Muhowe' and 'AA-Mshale malembo' from Tanzania and 'Nyakitembe' from DRC (INERA-M) clustered with both 'Yangambi Km5' (AAA-Ibota, from DRC and Tanzania). The AAB-Pome/Prata accession 'Cibwalo' clumped with FHIA 17 and FHIA 23 (both were of the AAAA genome). The relatedness between the 59 accessions based on the 98 characters using multivariate analysis had ascertained 47 different accessions, and established the remaining as their duplicates (Tables 2 and 3). The leaf upper surface was invariable and deleted by the NT-SYS package (during the standardization process). The high distance coefficient between the ramets was explained by the 22% of the quantitative characters used that were not considered as the element of genetic variability.

The results of the cophenetic (Ultrametric D) matrix presented a cophenetic correlation of 83% (i.e. $R_u = 0.83$, $N = 41$, $P = 0.008$) for the 41 accessions characterized from the DRC genebank (Table S2, available online only at http://journals.cambridge.org). The current results showed a good fit ($R_u > 0.8$) of the measure of the cluster analysis. A further addition of four accessions from the Tanzania genebank that have been characterized since 2008 had established a cophenetic correlation of 82% (i.e. $R_u = 0.82$, $N = 45$, $P = 0.008$) for the accessions characterized from both genebanks (INERA-M and SUA). These results also showed a good fit ($R_u > 0.8$) of the measure of the cluster analysis. Furthermore, the addition from the characterization of the 11 banana accessions from Tanzania (SUA Horticulture Unit) in 2010 (from July to November) had lowered the goodness of fit to 77% ($R_u = 0.77$, $N = 56$, $P = 0.008$). These were due to the fact that the genebank faced a drought spell. Therefore, the results of clustering analysis have ascertained and pointed out the deviation of character expression due to drought. This has once again established the effectiveness of numerical taxonomy to detect a change in gene expressions in relation to ecological variation. However, the phenetic classifications produced using the single-linkage and complete-linkage methods on the product-moment correlation and distance coefficients established a very poor fit ($R_u < 0.8$) of cophenetic values (Table S2, available online only at http://journals.cambridge.org).

The grouping of the accessions in PCA, shown in Fig. 3, was in line with that in clustering analyses in that the accessions also formed three clusters related to the genomic groups. High loaded character values (>0.5) with a percentage trace of 31.0, 4.75 and 0.13% for principal components (PCs) I, II and III, respectively, were observed (Table S3, available online only at http://journals.cambridge.org). However, the total percentage trace of PCs I, II and III was 43.29, 9.34 and 8.92%, respectively, giving 62%. These results confirm that every succeeding component contained less and less of the total variability and hence ascertain yet again the goodness of fit. As shown in PC axis I (31%) (Table S3, available online only at http://journals.cambridge.org), apart from the peduncle hairiness, bunch and rachis positions that had positive

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

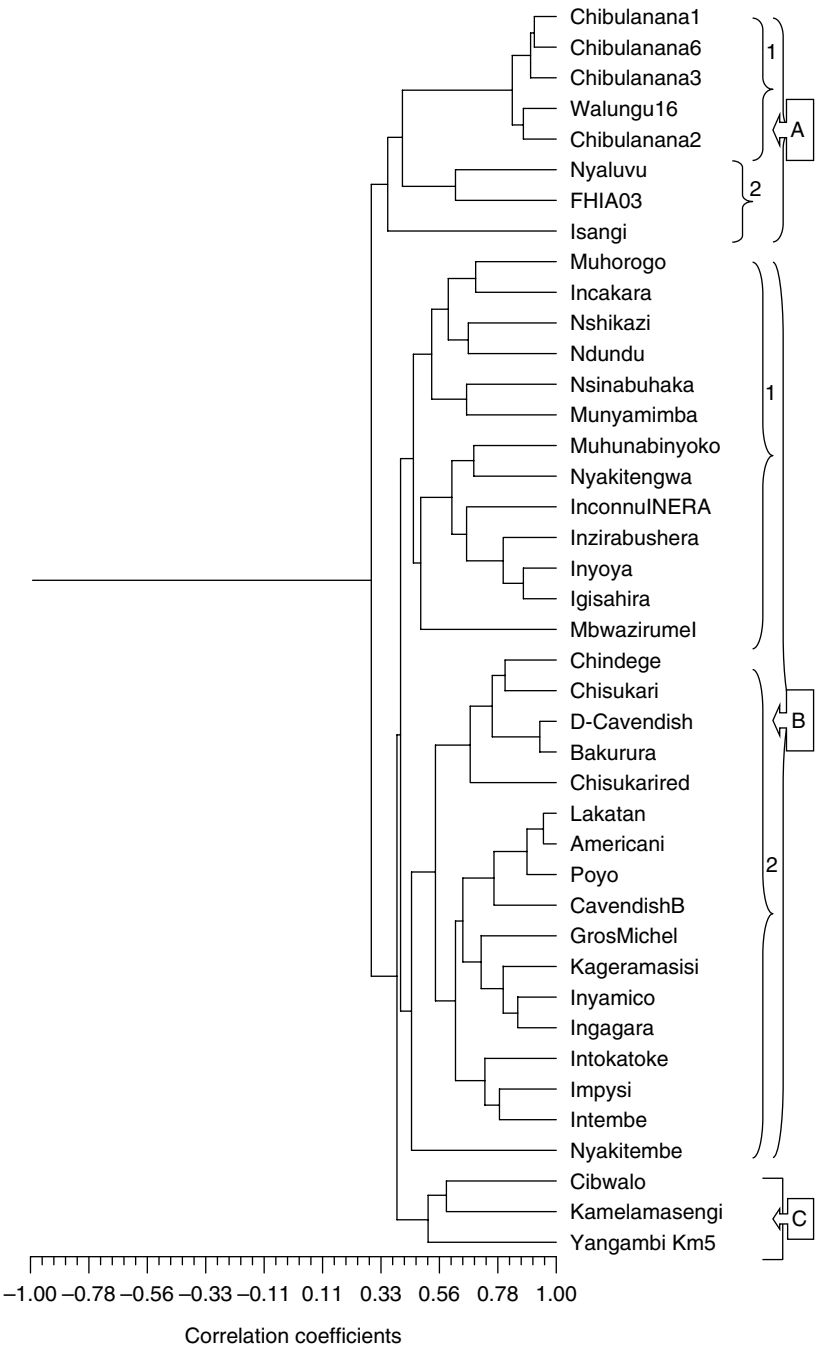

Fig. 1. Phenogram from UPGMA clustering of the correlation coefficient between the 41 banana accessions (Y) (INERA-M) based on the 98 characters (X).

loaded values, the entire remaining characters had negative, high loaded values. All high loaded characters (48) in this component started from leaves and petiole continued to male flower, bract, male bud and fruit that ended with a predominant taste. PC I showed that the heterogenomic bananas (AAB and AABB) clustered on the left side while AAA-EAHB were in the middle and the dessert bananas (AAA) on the right side (Figs 3, S6 and S7, available online only at http://journals.cambridge.org). Hence, these 48 high loaded characters constitute the best in genomic groups' separation.

Eleven high loaded (>0.5) characters in PC axis II (Table S3, available online only at http://journals.cambridge.org) were from pseudostem and suckers. These had low loaded values in PC I. A positive, high loaded value was observed on pseudostem height

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

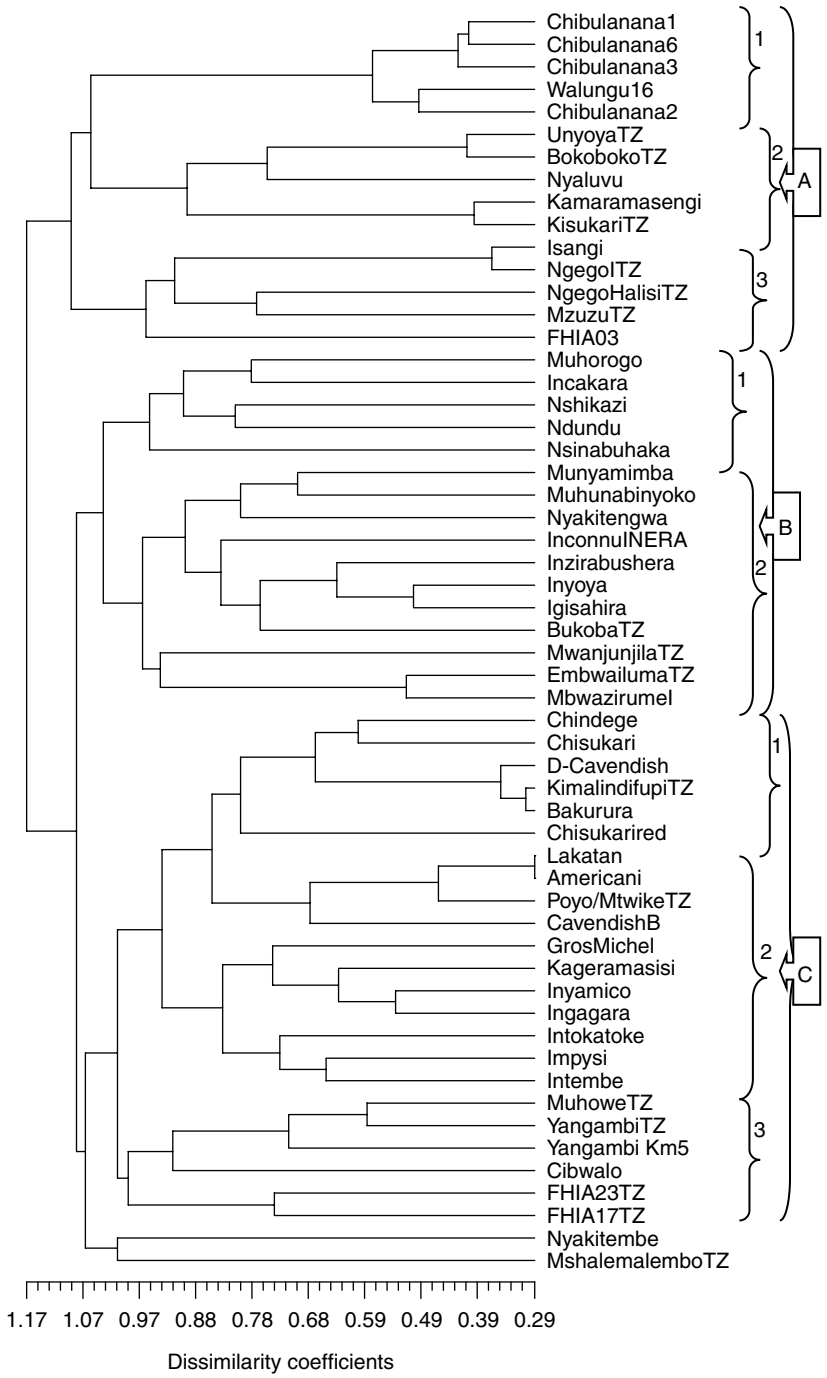

Fig. 2. Phenogram from UPGMA clustering of the dissimilarity coefficient between the 57 *Musa* accessions (41 from INERA-M and 16 from SUA).

(0.54), blotches at the petiole (0.56) and ovary basic colour (0.50). The negative characters were from other pseudostem and suckers' characters, leaf blade width (−0.54), fruit's pedicel surface (−0.61) and fusion of pedicels (−0.72). These 11 characters (PC II) showed the separation between AAB-Plantains and AABB, whereas the EAHB sharing many of these characters with dessert bananas were mixed (Figs 3 and S6, available online only at http://journals.cambridge.org). Five characters in PC axis III (Table S3, available online only at http://journals.cambridge.org) with a high loaded value (>0.5) in which a positive, high loaded value was observed for leaf blade length (0.59) but a negative loaded value observed for dwarfism character (−0.53).

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

These characters showed low loaded values in PCs I and II. There was a contrast between the petiole length with a negative, high loaded value in PC I but a positive, high loaded value in PC III, and peduncle hairiness with a positive high loaded value in PC I but a negative high loaded value in PC III. The PC III showed the separation between the EAHB and dessert bananas (Fig. S7, available online only at http://journals.cambridge.org).

Therefore, these results show that the PC were independent from each other in the case that all characters with a very low loaded value in PC I were high in the subsequent PC and no character had the same loaded value in the two PCs. These confirmed the validity of clustering analysis done on the banana accessions from DRC (INERA-M) and successfully related these accessions to their homologues from Tanzania (SUA genebank). Furthermore, high loaded character values with exponent 'a' (Table S3, available online only at http://journals.cambridge.org) were obtained after the addition of the 11 accessions (under drought in 2010) at SUA in Tanzania. In fact, there was a cut-off below 0.5 of the 11 high loaded characters, and a little addition of four other high loaded (>0.5) characters such as leaf habit, peduncle width (cm), bunch shape and fruit apex. Therefore, 15 characters were severely affected by the effects of drought (climate change). The total percentage trace of PCs I, II and III was 27.46, 8.87 and 7.21%, respectively, that gave a sum of 43.54% (~44%).

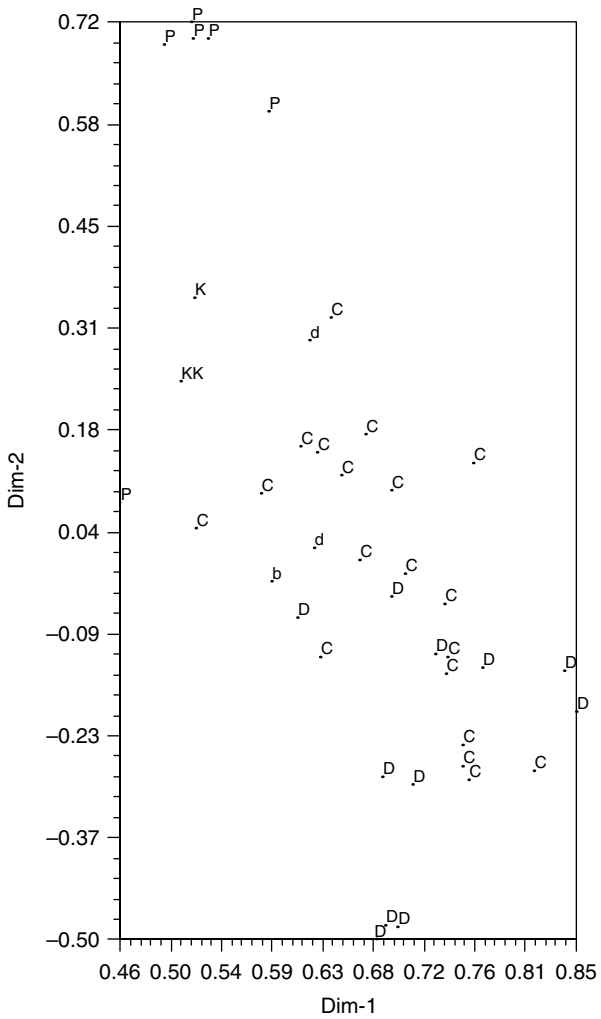

Fig. 3. PCA showing the relative positions on the first (Dim-1) and second (Dim-2) PCs of the 41 banana accessions of the INERA-M genebank (DR Congo) [6 AAB-Plantains (P), 2 AAB-Kamaramasengi (d), 1 AAA-Ibota (b), 1 AAA-Gros Michel (d), 3 AAA-Green/Red (D), 1 AABB-Pisang Awak (K), 1 AABB-FHIA 03 (KK), 19 AAA-EAHB (C) and 7 AAA-Cavendish (D)].

Discussion

The current banana genebank kept in INERA-Mulungu contain only 37 accessions in which six are plantain accessions remaining from the 56 cultivars collected by De Langhe (1961). This shows the existence of genetic erosion. The observed 78% minimum value of difference between the accessions based on qualitative characters is supported by Sneath and Sokal (1973) who found that the quantitative characters in bananas are influenced by the ecological variation. Even using only six quantitative characters (lengths, width and weight of fruit), Nsabimana (2006) did not succeed in separating the Nfuuka clone set from the Nakitembe clone set in similar ecology (genebank), which shows its weakness in classification. Karamura (1999) has shown that both have similar fruit length and the difference is between persistence flower and bracts on the rachis. Therefore, the use of both quantitative and qualitative characters is a very important method that allows having a good classification. The 39,298 feature patterns (data points) considered as phenetic markers are also supported by the reports of De Langhe *et al.* (2005) and Amorim *et al.* (2009), and imply that there are as many phenotypic expressions as there are molecular markers, and hence nullify the concept of the lack of data in phenotypic/phenetic study. The results of the three main clusters A, B and C in phenograms (Figs 1 and 2), which showed different genomic taxa following their genetic resemblance, are in line with the group, subgroup and clone set so far created by De Langhe and Valmayor (1980), Tezenas Du Montcel *et al.* (1983), Swennen and Vuylsteke (1987), Swennen (1990) and Karamura (1999). The six accessions of plantains, one belonging to the French clone set and five were of the Horn clone set, show that all the accessions belonging to the False Horn clone set have disappeared.

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

This clarifies the genetic loss in plantain diversity. The accession 'Isangi' is found to be a synonym of 'Isansi' described by Swennen (1990), while the current 'Walungu 16' belonging to the Horn clone set is in contrast with that described by the author (as belonging to French). The separation of AAB-Horn Plantain (in cluster A) from AAB-Silk and Pome (in cluster C; Figs 1 and 2) by numerical taxonomy is in line with the difference in mitochondrial and chloroplast gene pools reported by Boonruangrod (2008). The author has found two *M. acuminata*-type cytoplasms (III and VIII) in AAB (plantain and Silk) and mixed cytotypes (XII, XIII and XIV) in the ABB genomic groups that constitute the source of genetic variability. The accessions AAB-Pome/Prata 'Cibwalo' and AAA-Dwarf Cavendish 'Bakurura' were named 'Prata aña' and 'Kigurube' by Sebasigari (1987). The DRC's banana accessions clustered well with their homologues from Tanzania. These results ascertained the effectiveness of numerical taxonomy through the mathematical process (i.e. algorithm) to cluster individual banana accessions in the existing *Musa* genomic groups, subgroups and clone sets regardless of their different agro-ecological conditions. This is supported by Karamura (1999) from two genebanks in Uganda. The Horn accessions 'Chibulanana one hand (I)', 'Chibulanana two hands (II)' and 'Chibulanana three hands (III)' are found to be a synonym of 'Mkono wa tembo Chana moja', 'Mkono wa tembo Chana mbili' and 'Mkono wa tembo Chana tatu' out of the comparison with the characterization by Swennen (1990) and Evers (1992) in Nigeria and Tanzania, respectively. The accessions 'Mtwike' and 'Kimalindi-ndefu' were highly similar (100%) to 'Poyo' and 'Cavendish B' and hence omitted in the subsequent analysis. The reason being that it is based on the variation of characters among the accessions (Sneath and Sokal, 1973). Both 'Mtwike' and 'Poyo' are found to be a synonym of 'Grand(e) Nain(e)' described by Stover and Simmonds (1991). The accession 'Gros Michel' from DRC was found as a duplicate of 'Jamaica' in Tanzania. The clustering of the accession 'Cibwalo' (i.e. AAB-Prata by Sebasigari, 1987) to the AAAA-FHIA genomic group confirmed the presence of a similarity between their A genome from a breeding programme in Honduras where Prata aña (AAB) was a female plant (Robinson, 1996). The observed cluster of 'Gros Michel' with 'Yangambi Km5', with the highest distance coefficient (R_d) of 1.00 (Fig. 2), is in line with the upshot of Ude *et al.* (2002) and Onyango *et al.* (2010) who had reported through molecular characterization (using amplified fragment length polymorphism and simple sequence repeat methods, respectively) that both accessions clustered because of their similar source of some A genome. The current clustering of the AA-Mshale genomic group with the AAA genomic group including the AAA-EAHB subgroups (Figs 2 and S3, available online only at http://journals.cambridge.org) indicates that AA-Mshale malembo may be one of the ancestors of Lujugira-Mutika, which is supported by De Langhe *et al.* (2001). The accession 'Mshale' is similar to *Musa acuminata* spp. *malecensis*, which was reported by Ude *et al.* (2002) to be one of the ancestors of 'Yangambi Km5' and 'Gros Michel'. However, the exact contribution by AA-'Mshale malembo' as one of the ancestors of the Lujugira-Mutika subgroup requires a molecular investigation to confirm the current phylogeny. The clumping of AAA dessert bananas (Cavendish and Gros Michel) with AAA-EAHB is in agreement with Onyango *et al.* (2010) who found similar results from microsatellite analysis but was in contrast with Karamura's (1999) phenetic finding from Ugandan banana collections where these dessert accessions clustered with AAB-Plantain. This contrast may be due to the disparity between the number of characters (73) used by Karamura (1999) and those in this study (98 characters). The percentage trace (62%) being superior to 50% reconfirmed the validity of the cluster analysis between 50 and 95% as described by Wiley (1981). The observation of the distance between Gros Michel and FHIA 03 in the phenograms (Figs 1 and 2) is explained by somaclonal mutation that changed 'High gate' features (to dwarf type) reported by Robinson (1996). The total percentage trace (44%) after adding the 11 Tanzania accessions (in 2010) being inferior to 50% established the slight fit of the cluster analysis as described by Wiley (1981). The weak fit explains that the accessions with known character states, once under drought spell, express different character states. PC I (Fig. 3) represented nearly the plant's life cycle and explained up to 43.29% of the observed variation; and did not reach 50% because of the presence of both positive and negative values. This result is supported by the Palmer (2012) report on the validity of PC I. Moreover, the results (Table S3, available online only at http://journals.cambridge.org) show that there were 12 characters not used by Karamura (1999) that had a high loaded value (>0.5) influencing the clustering of AAA-Cavendish to AAA-EAHB. Among these characters were even those which are thought to have high discriminant values described by IPGRI (1996). These were qualitative flowers and fruit pedicel characters that are highly heritable and hence very important in banana breeding. They included the characters reported by De Langhe (1961), Tezenas Du Montcel *et al.* (1983), Vuylsteke *et al.* (1991) and De Langhe *et al.* (2005) that were used in the subjective classification of the AAB-Plantain subgroup. Numerical taxonomy therefore constitutes the most reliable way for assessing the genetic relationship among diversity using characters covering the crop's life cycle.

https://doi.org/10.1017/S1479262112000354 Published online by Cambridge University Press

Acknowledgements

The authors would like to thank the Canadian International Development Agency (CIDA) (through Bioscience in East and Central Africa Network/NEPAD) for funding this research.

References

Amorim EP, Vilarinhos AD, Cohen KO, Amorim VBO, Santos-Serejo JA, Silva SO, Pestana KN, Santos VJ, Paes NS, Monte DC and Reis RV (2009) Genetic diversity of carotenoid-rich bananas evaluated by diversity arrays technology (DArT). *Genetics and Molecular Biology* 32: 96–103.

Blench R (2009) Bananas and plantains in Africa: re-interpreting the linguistic evidence. *Ethnobotany Research & Applications* 7: 363–380.

Boonruangrod R (2008) Application of cytoplasmic and nuclear genome based marker systems for studying the diversity of present day edible bananas. PhD Thesis, Universität für Bodenkultur.

Christelová P, Valárik M, Hribová E, De Langhe E and Doležel J (2011) A multi gene sequence-based phylogeny of the *Musaceae* (banana) family. *BMC Evolutionary Biology* 11: 103–115.

De Langhe E (1961) La taxonomie du bananier plantain en Afrique équatoriale. *Journal d'Agriculture Tropicale et de Botanique Appliquée* 8: 417–449.

De Langhe E and Valmayor RV (1980) French plantains in South East Asia. *IBPGR/SEAN* 4: 3–4.

De Langhe E, Karamura D and Mbwana A (2001) *Tanzania Musa Expedition 2001*. Rome: INIBAP/IPGRI, Future harvest, 107 pp.

De Langhe E, Pillay M, Tenkouano A and Swennen R (2005) Integrating morphological and molecular taxonomy in *Musa*: the African plantains (*Musa* spp. AAB group). *Plant Systematics and Evolution* 255: 225–236.

De Langhe E, Hribová E, Carpentier S, Doležel J and Swennen R (2010) Did backcrossing contribute to the origin of hybrid edible bananas? *Annals of Botany* 106: 849–857.

Evers G (1992) Banana cultivar diversity in the area of Morogoro, Tanzania. *Fruits* 47: 377–391.

IPGRI (1996) *Descriptors for Banana* (Musa *spp.*). Rome: IPGRI.

Karamura D (1999) Numerical taxonomic studies of the East African highland bananas (*Musa* AAA-East African) in Uganda. PhD Thesis, University of Reading.

Maerere AP, Rweyemamu CL, Sibuga KP, Mgembe ER, Rwambali E, Nchimbi-Msolla S (2010a) Analysis of the agriculture science, technology and innovation systems: a case study of banana in Tanzania. *Acta Horticulturae* 879: 851–858.

Maerere AP, Msogoya TJ, Mgembe ER, Mwaitulo S, Mtui HD, Mbilinyi L (2010b) Comparison of yield performance of improved and local banana (*Musa*) cultivars in Eastern zone of Tanzania. In: Batamuzi EK, Gimbi DM, Max RA, Nombo CI, Msoffe GE (eds) *Proceedings of the Fourth Annual PANTIL Scientific Conference*, 19–21, October 2009, Morogoro, Tanzania, pp. 172–177.

Miller RNG, Passos MAN, Menezes NNP, Souza MT Jr., Costa MMC, Azevedo VCR, Amorim EP, Pappas GJ Jr. and Ciampi AY (2010) Characterization of novel microsatellite markers in *Musa acuminata* subsp. *burmannicoides*, var. Calcutta 4. *BMC Research Notes* 3: 148–153.

Mobambo KN and Naku M (1993) Situation de la cercosporiose noire des bananiers et plantains *Musa* spp. Sous différents systèmes de culture à Yangambi, Haut-Zaïre. *Tropicultura* 11: 7–10.

Mobambo P, Staver C, Hauser S, Dheda B and Vangu G (2010) An innovation capacity analysis to identify strategies for improving plantain and banana (*Musa* spp.) productivity and value addition in the Democratic Republic of Congo. *Acta Horticulturae* 879: 821–828.

Msogoya TJ, Maerere AP, Kusolwa PM and Nsemwa LT (2006) Field performance of improved banana cv FHIA 17 and FHIA 23 in the Eastern zone of Tanzania. *Journal of Agronomy* 5: 533–535.

Msogoya TJ, Grout B and Maerere AP (2010) Altered yield and fruit quality of micropropagation-induced off-type of East African Highland banana (*Musa* AAA East Africa). In: Batamuzi EK, Gimbi DM, Max RA, Nombo CI and Msoffe GE (eds) *Proceedings of the Fourth Annual PANTIL Scientific Conference,* 19–21, October 2009. Morogoro, Tanzania, pp. 231–235.

Mtengeti EE (2008) Residual effects of minjingu phosphate rock and triple superphosphate in some highly weathered and volcanic soils of Tanzania. PhD Thesis, Sokoine University of Agriculture.

Nsabimana A (2006) Establishing genetic diversity of Rwanda highland banana using random amplified polymorphism DNA markers. PhD Thesis, University of Kwazulu-Natal.

Nzawele DB, Rweyemamu CL and Maerere AP (2009) Banana (*Musa* spp. Colla) cropping systems, production constraints and cultivar preferences in Eastern Democratic Republic of Congo. *Journal of Animal & Plant Sciences* 4: 341–356.

Nzawele DB, Blomme G, Dheda DB, Rweyemam CL, Velly D, Vigheri N, Milango A, Eden-Green S, Karamura E and Maerere AP (2010) The alarming spread of banana *Xanthomonas* wilt in eastern Democratic Republic of Congo and it impact on food security and income. *Tree and Forestery Science and Biotechnology* 4: 56–59.

Onyango M, Haymer D, Keeley S and Manshardt R (2010) Analysis of genetic diversity and relationships in East African 'Apple Banana' (AAB genome) and 'Muraru' (AA genome) dessert bananas using microsatellite markers. *Acta Horticulturae* 879: 623–636.

Opara UL, Jacobson D and Al-saady NA (2010) Analysis of genetic diversity in banana cultivars (*Musa* cvs.) from the South of Oman using AFLP markers and classification by phylogenetic, hierarchical clustering and principal component analyses. *Journal of Zhejiang Uni 332 v-Sci B (Biomed & Biotechnol)* 11: 332–341.

Palmer M (2012) Principal component analysis. Available at www.ordination.okstate.edu/PCA.htm

Pecrot A (1958) *Carte des sols et de la végétation du Congo Belge et du Rwanda Urundi. 16 Dorsale du Kivu. A. sols, carte de reconnaissance*. Bruxelles: INEAC.

Pestana RKN, Amorim EP, Ferreira CF, Amorim VBO, Oliveira LS, Ledo CAS and Silva SO (2011) Genetic dissimilarity of putative gamma-ray-induced 'Preciosa – AAAB-Pome type' banana (*Musa* sp) mutants based on multivariate statistical analysis. *Genetic and Molecular Research* 10: 3976–3986.

Robinson JC (1996) *Bananas and Plantains*. Wallingford: CAB International.

Sebasigari K (1987) Morphological taxonomy of *Musa* in Eastern Africa. In: Persley GJ and De Langhe EA (eds) *Proceedings of an International Workshop on Banana and Plantains*

Breeding Strategies, 13–17 October 1986, Cairns, Australia, pp. 172–176.

Simmonds NW and Shepherd K (1955) The taxonomy and origins of the cultivated bananas. *Journal of Linnean Society of London Botany* 55: 302–312.

Sneath PHA and Sokal RR (1973) *Numerical Taxonomy: the Principles and Practice of Numerical Classification*. San Francisco, CA: Freeman WH and Company.

Stover RH and Simmonds NW (1991) *Bananas. Tropical Agriculture Series*, 3rd edn. London: Longman Scientific & Technical.

Swennen R (1990) Limits of morphotaxonomy: names and synonyms of plantains in Africa and elsewhere. In: Jarret RL (ed.) *Proceeding of Identification of Genetic Diversity in the Genus Musa,* 5–10 September 1988, Los Banos, Philippines, pp. 172–176.

Swennen R and Vuylsteke D (1987) Morphological taxonomy of plantains (*Musa* cultivar AAB) in West Africa. In: Persley G and De Langhe E (eds) *Proceeding of Banana and Plantains Breeding,* 13–17 October 1986, Cairns, Australia, pp. 165–171.

Tezenas Du Montcel H, De Langhe E and Swennen R (1983) Essai de Classification des bananiers plantains (AAB). *Fruits* 38: 461–474.

Ude G, Pillay M, Nwakanma D and Tenkouano A (2002) Genetic Diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP markers. *Theoretical and Applied Genetics* 104: 1246–1252.

Vuylsteke D, Swennen R and De Langhe E (1991) Somaclonal variation in plantains (*Musa* spp. AAB group) derived from shoot-tip culture. *Fruits* 46: 429–439.

Wiley EO (1981) *Phylogenetics: the Theory and Practice of Phylogenetic Systematics*. New York/Chichester/Brisbane/Toronto: John Wiley & Sons, Inc, pp. 348–365.

World, CG (2012) Available at http://www.worldclimateguide.co.uk/climateguides/democraticrepublicofthe congo/bukavu.php