Annals of Botany **123**: 19–36, 2019
doi: 10.1093/aob/mcy156, available online at www.academic.oup.com/aob

ANNALS OF BOTANY
Founded 1887

RESEARCH IN CONTEXT

East African diploid and triploid bananas: a genetic complex transported from South-East Asia

Xavier Perrier[1,2,*], Christophe Jenny[1,2], Frédéric Bakry[1,2], Deborah Karamura[3], Mercy Kitavi[4], Cécile Dubois[1,2], Catherine Hervouet[1,2], Gérard Philippson[5,6] and Edmond De Langhe[7]

[1]CIRAD, UMR AGAP, F-34398 Montpellier, France, [2]AGAP, Université de Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France, [3]Bioversity International Uganda Office, PO Box 24384, Kampala, Uganda, [4]International Potato Center, PO Box 25171, Nairobi, Kenya, [5]Institut National des Langues et Civilisations Orientales, Paris, France and [6]Laboratoire Dynamique du Langage CNRS, Université Lyon 2, France and [7]Katholieke Universiteit Leuven (KUL), Belgium
**For correspondence. E-mail xavier.perrier@cirad.fr*

Received: 20 April 2018 Returned for revision: 18 June 2018 Editorial decision: 20 July 2018 Accepted: 27 July 2018
Published electronically 20 September 2018

• **Background and Aims** Besides bananas belonging to the AAA triploid Mutika subgroup, which predominates in the Great Lakes countries, other AAA triploids as well as edible AA diploids, locally of considerable cultural weight, are cultivated in East Africa and in the nearby Indian Ocean islands as far as Madagascar. All these varieties call for the genetic identification and characterization of their interrelations on account of their regional socio-economic significance and their potential for banana breeding strategies.
• **Methods** An extensive sampling of all traditional bananas in East Africa and near Indian Ocean islands was genotyped with simple sequence repeat (SSR) markers, with particular emphasis on the diploid forms and on the bananas of the Indian Ocean islands, which remain poorly characterized.
• **Key Results** All the edible AA varieties studied here are genetically homogeneous, constituting a unique subgroup, here called 'Mchare', despite high phenotypic variation and adaptions to highly diverse ecological zones. At triploid level, and besides the well-known AAA Mutika subgroup, at least two other genetically related AAA subgroups specific to this region are identified. Neither of these East African AAA genotypes can be derived directly from the local AA Mchare diploids. However, it is demonstrated that the East African diploids and triploids together belong to the same genetic complex. The geographical distribution of their wild *acuminata* relatives allowed identification of the original area of this complex in a restricted part of island South-East Asia. The inferred origin leads to consideration of the history of banana introduction in Africa. Linked to biological features, documentation on the embedding of bananas in founding legends and myths and convincing linguistic elements were informative regarding the period and the peoples who introduced these Asian plants into Africa. The results point to the role of Austronesian-speaking peoples who colonized the Indian Ocean islands, particularly Madagascar, and reached the East African coasts.
• **Conclusions** Understanding of the relations between the components of this complex and identifying their Asian wild relatives and related cultivars will be a valuable asset in breeding programmes and will boost the genetic improvement of East African bananas, but also of other globally important subgroups, in particular the AAA Cavendish.

Key words: *Musa*, banana, Mutika, Mchare, East Africa, Indian Ocean, nuclear SSR, linguistic, human migrations, genetic diversity, plant diffusion, genetic complex.

INTRODUCTION

Bananas contribute greatly to the diet in large regions of Africa. Yet the banana does not originate from Africa. No species of *Musa* is native to the continent and only the neighbouring genus *Ensete* is naturally present in Africa and Madagascar. The success of bananas in Africa arises from highly suited ecologies in the rainforest zones of Cameroon, Central African Republic, Congo, etc., with extensions into western Africa. In the more marginal climatic zones of East Africa, such suitable conditions are found at high altitude, with three major zones along the East African Rift: the Ethiopian Highlands, the vast zone of the Great Lakes and, to the south, the upland areas of Malawi. While the cultivation of *Ensete* dominates in Ethiopia, banana cultivation is core in Rwanda, Burundi, north-western Tanzania and Uganda, and a little more limited in Malawi. This crop even extends to the Indian Ocean coastline, in Kenya or Tanzania, benefiting from the mountain chains around the great volcanoes such as Mount Kenya and Kilimanjaro. This crop also exists, although it is not a staple, along the coast, on the offshore islands, such as Zanzibar or Pemba, and further out to sea in the Comoro archipelago and Madagascar.

In addition to its climatic adaptability, the major asset of banana is its high yield capacity per unit area with little seasonality; this asset becomes decisive in zones with a very high population density. Such is the case in all the upland zones mentioned, with, for example, densities exceeding 500, sometimes

1000, inhabitants per km^2 (FAO, 1986) on the southern and south-eastern slopes of Mount Kilimanjaro, at an altitude of between 1000 and 1800 m. Although the climatic conditions there are not optimal, the population densities found on the islands (Pemba, Zanzibar, Comoros) no doubt explain the frequent existence of banana crops there too.

Compared with the very great diversity of wild and cultivated bananas found in South-East Asia, or even in India, which is known to be a secondary region of diversification, the genetic diversity in Africa is very limited (excluding here the numerous exchanges of the modern era). It mainly involves sterile parthenocarpic triploid forms, which share, along with all the cultivated bananas, pulp development without seeds, a particularity that is the major domestication syndrome. As edible derivatives of the wild species *Musa acuminata* (genome A), the AAA triploids of the Mutika subgroup, initially called Mutika-Lujugira by Shepherd (1957) and currently often called EAHB (East African Highland bananas), are very widely grown in East Africa, especially in the Great Lakes region, where they are the main crop in many areas. The highly starchy fruits are mostly consumed cooked but also after fermentation in the form of banana beer. Consumption is estimated at almost 250 kg per inhabitant per year in Uganda, Burundi and Rwanda. In Central and West Africa, the traditionally cultivated bananas belong to the plantain subgroup, triploid AAB interspecific hybrids with the species *Musa balbisiana* (genome B). The AAB Plantain fruits are mostly roasted but also cooked, with consumption levels that can reach 100 kg per inhabitant per year. In both cases, Mutika and Plantain, almost all production is devoted to both on-farm consumption and local markets.

In addition to these traditional crops, various AAA, AAB or ABB genotypes have been introduced under Arabian and colonial influence. A most notable production is that of bananas of the AAA Cavendish dessert type, which makes up most of the international market and which can be found in estate plantations in Central and West Africa.

Despite their low genetic diversity, the traditional African production of AAA Mutika as well as AAB Plantain subgroups relies on a large number of varieties that differ in their colour, shape, taste, etc., sometimes with extremely marked differences. This phenotypic differentiation has resulted from a long selection process involving somatic variations in the absence of any sexuality for these sterile bananas, the variations being maintained by vegetative propagation.

However, the assertion of limited African genetic diversity needs to be put in perspective. The inventory and survey works in East Africa (De Langhe *et al.*, 2001; Byabachwezi *et al.*, 2005; Karamura *et al.*, 2006) show that the AAA Mutika of the Great Lakes region are rare in the highlands of the Kenya–Tanzania–Malawi zone, where other AAA triploid types are cultivated that could not be classified as Mutika. Moreover, in the same Kenya–Tanzania–Malawi zone, several phenotypically diversified AA diploid varieties (Fig. 1) are grown and are of considerable cultural importance for some ethnic groups, such as the Chaggas of Kilimanjaro (Montlahuc and Philippson, 2003). Cultivation of these AAA or AA varieties extends even into the lowland zone of these countries, and down to the coast, as well as on the islands of the Indian Ocean, as far as Madagascar.

In East Africa, and apart from some pioneering work (Baker and Simmonds, 1951; 1952; Shepherd, 1957), studies have often been focused on the AAA Mutika varieties: their morphological diversity, the diversity of their uses, or the cropping systems (Karamura, 1999; Tugume *et al.*, 2002; Nsabimana *et al.*, 2010) and genetic diversity (Pillay *et al.*, 2001; Onguso *et al.*, 2004; Onyango *et al.*, 2010; Karamura *et al.*, 2016; Kitavi *et al.*, 2016).

The other AA and traditional AAA varieties specific to East Africa appear incidentally in several of the previously mentioned papers, but they remain largely less well known despite their great potential for providing information useful for understanding the genetic structure and evolutionary history of the East African banana varieties. In addition to central questions on the domestication, diversification and diffusion of bananas and their drivers, this understanding has concrete applications, in particular for varietal improvement within Mutika as well as Cavendish subgroups, where several serious diseases, causing drastic decline in yields, are a threat to farmers and consumers, as will be explained later in this article.

This work therefore deals with all the traditionally cultivated bananas in East Africa but devotes special attention to the diploid forms with consequent emphasis on the Indian Ocean islands and Madagascar, where they are abundant. The purpose is to characterize, beyond the observed phenotypic diversity, the underlying genetic diversity and its organization, in particular the kinship relationships between AA and AAA genomes. The traditional banana varieties in East Africa will be shown to constitute a particular genetic complex with its origin in a well-defined part of South-East Asia. This finding will shed a new light not only on questions about the routes of introduction of the banana in Africa, but also on the people who drove this process and on the time at which it occurred. Indeed, a broad review of archaeological, cultural and linguistic evidence, combined with our genetic results, will show how the history of banana introduction is entangled with the history of the migrant peoples who colonized the islands of the Indian Ocean, particularly Madagascar, and reached the East African coasts.

MATERIALS AND METHODS

Plant materials

A sample of 89 accessions was genotyped with 23 simple sequence repeat (SSR) markers. Sampling was targeted on the diploid and triploid bananas traditionally grown in East Africa and on the Indian Ocean islands as far as Madagascar (Table 1 and Fig. 2).

The diversity of the *acuminata* diploid and triploid bananas of Tanzania was sampled during a collecting mission carried out in the Kilimanjaro, Usambara and Pare regions of Tanzania (De Langhe *et al.*, 2001), later completed by a collection in the Morogoro region (Byabachwezi *et al.*, 2005). The accessions were introduced into the Bioversity International *Musa* Germplasm Transit Centre (ITC) in Belgium. For this study, eight of these accessions were supplied by ITC in the form of proliferating shoot clusters, allowing ploidy determination after culture, and four as freeze-dried leaves. Two older introductions in the CIRAD collection in Guadeloupe were also added to the sample: the ‘Mjenga’ and ‘Paka’ accessions identified by Baker and Simmonds (1952) in Zanzibar. A small population of seedy bananas from the island of Pemba in Tanzania, around 50 km off the African coast, is mentioned by Simmonds and Shepherd (1952) at two sites in northern Pemba, the Ngezi Forest and Wete, but also on Funzi Island, which is very close to the continent. The seedy fruits of these wild bananas, named

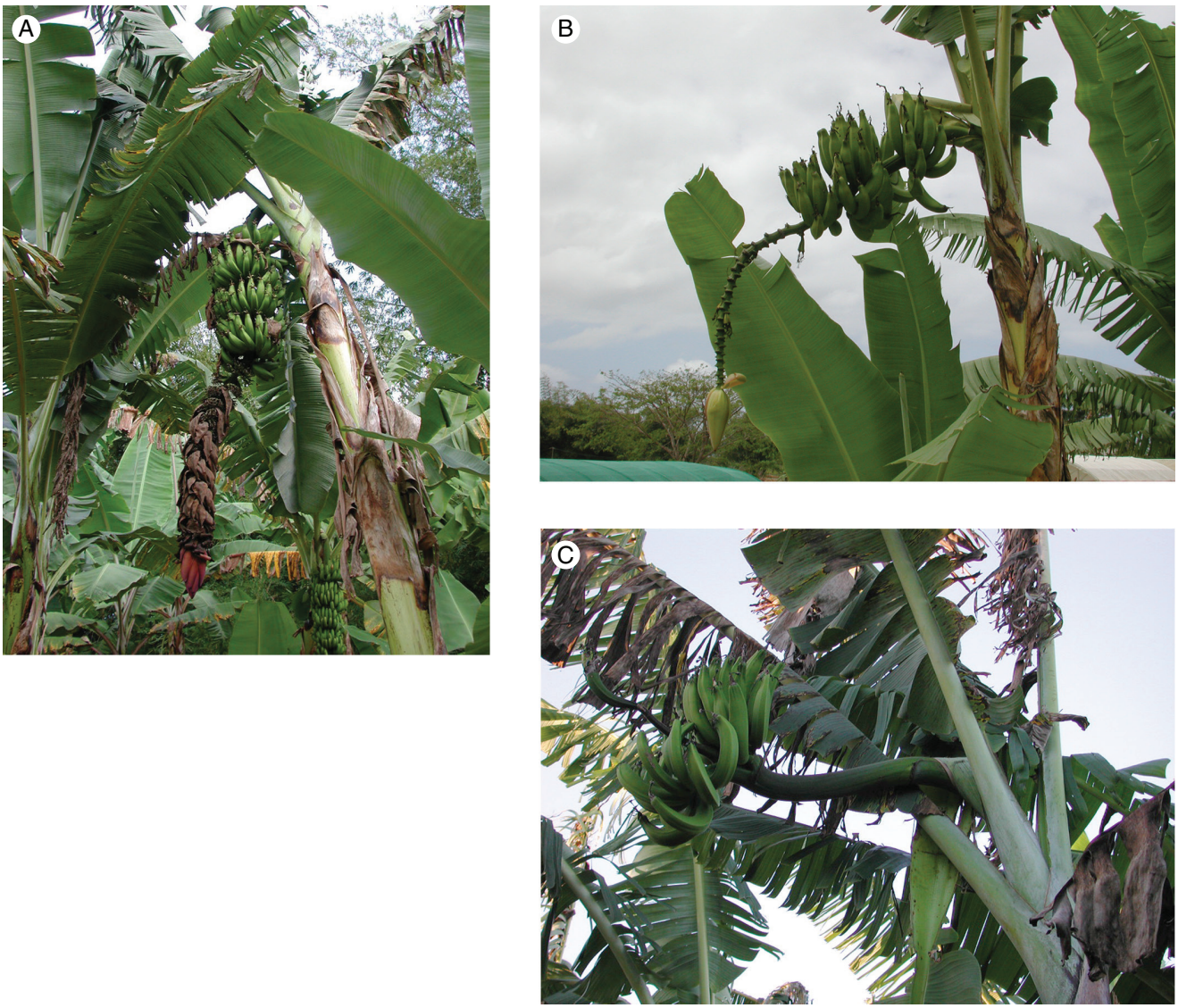

FIG. 1. Three contrasting morphotypes of the Mchare diploid AA subgroup from the Comoros. (A) 'Mlali Angaia'; (B) 'Mlali Mshia Wa Komba'; (C) 'Pima Moja' (photographs: Thierry Lescot).

'Mgomba tumbili' or 'Mgomba kofi' in Swahili, were used as famine food in the past and are still eaten by children (Walsh, 2009). For this study, three plants around 3 km from each other were sampled by Alison Crowther in the Ngezi Forest as part of the Sealinks project funded by the European Research Council (ERC). Two plants provided analysable DNA.

Several representatives of the diploid cultivars, known in Kenya under the generic term 'Muraru', were selected by the late Dr Margaret Onyango [Kenya Agricultural Research Institute (KARI), now known as Kenya Agricultural Research and Livestock Organization (KARLO)] from the collection in Kisii. Ten accessions were sent as green leaves, thus allowing ploidy measurement, but only five provided DNA of sufficient quality for genotyping.

The AAA Mutika subgroup was represented by three accessions: two from Burundi and one from the Democratic Republic of Congo. A fourth, 'Biu-Se-Ed', morphologically described as a typical Mutika, was collected in Bali. It was added here to check its genetic attribution. The material came from the CIRAD collection in Guadeloupe.

As part of the Sealinks project mentioned above, several banana plants were sampled by Michele Wollstonecroft in the Gamo Gofa region of Ethiopia and five of these were included in this analysis as potential controls of the northern extension of East African bananas.

The bananas from the Comoro Islands came from a mission carried out on Mwali (Moheli) in the Comoros archipelago (Mzemouigni, 2007). The accessions were brought back in fresh leaf form to measure ploidy by flow cytometry prior to freezing. Fourteen of these frozen accessions were genotyped. In addition, the 'Mnalouki' accession, formerly introduced in the CIRAD collection in Guadeloupe from the Comoro Islands, was also genotyped.

The accessions from Mayotte were sampled in the collection at the Dembeni agricultural research station and in various banana plantations in the centre and south of the island (survey

TABLE 1. *Analysed accessions: name, classification, source and geographical origin*

nid	Accession	Received as			Confirmed as			Plant material source		Origin of accession when known		
		Ploidy	Group	Subgroup	Ploidy*	Group	Subgroup	Collecting mission	Gene bank** (code)	Village	Region	Country
1	Pima Moja 6 mains	2	AA	Mlali	2	AA	Mchare		CAPAM (SD1)		Mayotte	Mayotte
2	Pima Moja vrai	2	AA	Mlali	2	AA	Mchare		CAPAM (SD2)		Mayotte	Mayotte
3	Mlali Pima 2 mains	2	AA	Mlali	2	AA	Mchare		CAPAM (SD3)		Mayotte	Mayotte
13	Mlali Popo	3	AA	Mlali	2	AA	Mchare		CAPAM (SD14)		Mayotte	Mayotte
34	Mlali Makoutri	2	AA	Mlali	2	AA	Mchare		CAPAM (C33L2)		Mayotte	Mayotte
36	Mlali Mainty Régime noir	2	AA	Mlali	2	AA	Mchare		CAPAM (C31L3)		Mayotte	Mayotte
38	Mlali Angaia	2	AA	Mlali	2	AA	Mchare	Mayotte		Coconi	Mayotte	Mayotte
39	Mlali Mshia Wa Komba	2	AA	Mlali	2	AA	Mchare	Mayotte		Coconi	Mayotte	Mayotte
51	Chikame	2	AA	Mlali	2[c]	AA	Mchare	Comoros		Hoani	Mwali	Comoros
54	Chimoili Kana Nkoboï	2	AA	Mlali	2[c]	AA	Mchare	Comoros		Hayiraha	Mwali	Comoros
58	Mdjenga Maoua	2	AA	Mlali	2[c]	AA	Mchare	Comoros		Hoani	Mwali	Comoros
92	Pimodja	3			2	AA	Mchare	Comoros		Hoani	Mwali	Comoros
112	Muraru	2	AA	Mlali	2[c]	AA	Mchare		KARI (Kari 2)			Kenya
114	Muraru Mshare	2	AA	Mlali	2[c]	AA	Mchare		KARI (Kari 4)			Kenya
118	Kamunyilya	2	AA	Mlali	2[c]	AA	Mchare		KARI (Kari 8)			Kenya
119	Maji Maji	2	AA	Mlali	2[c]	AA	Mchare		KARI (Kari 9)			Kenya
120	Makyughu	2	AA	Mlali	2[c]	AA	Mchare		KARI (Kari 10)			Kenya
121	Mshale	2	AA	Mlali	2	AA	Mchare		ITC (ITC1223)		Morogoro	Tanzania
124	Huti (Shumba Nyelu)	2	AA	Mlali	2[c]	AA	Mchare		ITC (ITC1452)	Mbokoi	Usambara	Tanzania
126	Mshale Mlelembo	2	AA	Mlali	2	AA	Mchare		ITC (ITC1455)	Uduru	Kilimanjaro	Tanzania
128	Ntindi II	2	AA	Mlali	2[c]	AA	Figue		ITC (ITC1463)	Bethaniya	East Usambara	Tanzania
130	Kisanga Machi	2	AA	Mlali	2[c]	AA	Figue		ITC (ITC1467)	Bethaniya	East Usambara	Tanzania
131	Kahuti	2	AA	Mlali	2[c]	AA	Mchare		ITC (ITC1468)	Maguzoni	West Usambara	Tanzania
133	Ndyali	2/4	AA	Mlali	2[c]	AA	Mchare		ITC (ITC1552)	Malangali	Mbeya	Tanzania
134	Huti green bell	2	AA	Mlali	2[c]	AA	Mchare		ITC (ITC1559)			Tanzania
136	Mshare Mrerembo	2	AA	Mlali	2	AA	Mchare		ITC (ITC1579)			Tanzania
138	Mshare	2	AA	Mlali	2	AA	Mchare		ITC (ITC1594)			Tanzania
167	Akondro Mainty	2	AA	Mlali	2	AA	Mchare		CIRAD (PT-BA-00010)			Madagascar
168	Mjenga	2	AA	Mlali	2	AA	Mchare		CIRAD (PT-BA-00205)		Pemba	Tanzania
169	Paka	2	AA	Mlali	2	AA	EAB AAcv ind.		CIRAD (PT-BA-00270)		Zanzibar	Tanzania
62	Mumbwa Dzu	2			2[c]	AB ?	AB ind.?	Comoros		Hayiraha	Mwali	Comoros
142	AC12-002	2	AAw		2	AAw	EAB AAw	Pemba, TZA		Ngezi Forest	Pemba	Tanzania
143	AC12-003	2	AAw		2	AAw	EAB AAw	Pemba, TZA		Ngezi Forest	Pemba	Tanzania
152	Ambihy-P1	2	AAw		2	AAw	EAB AAw	Madagascar			Ambihy	Madagascar
155	Ambihy-P4	2	AAw		2	AAw	EAB AAw	Madagascar			Ambihy	Madagascar
156	Ambihy-P5	2	AAw		2	AAw	EAB AAw	Madagascar			Ambihy	Madagascar
157	Ambihy-P6	2	AAw		2	AAw	EAB AAw	Madagascar			Ambihy	Madagascar
178	Mbwazirume	3	AAA	Mutika	3	AAA	Mutika		CIRAD (PT-BA-00201)			Burundi
187	Bolo Bigouyo	3	AAA	Mutika	3	AAA	Mutika		CIRAD (PT-BA-00038)			Congo DR
188	Biu-Se-Ed	3	AAA	Mutika	3	AAA	Mutika		CIRAD (PT-BA-00045)		Bali	Indonesia
190	Nshika	3	AAA	Mutika	3	AAA	Mutika		CIRAD (PT-BA-00246)			Burundi
28	Shiwendre	3			3	AAA	Ilalyi		CAPAM (SD42)		Mayotte	Mayotte
29	Koja	3			3	AAA	Ilalyi		CAPAM (SD43)		Mayotte	Mayotte
56	Dimba	2			3[c]	AAA	Ilalyi	Comoros		Hoani	Mwali	Comoros
132	Mlelembo	3	AAA	AAA ind.	3[c]	AAA	Ilalyi		ITC (ITC1544)		Kilimanjaro	Tanzania
140	Mlali/Ilalyi	3	AAA	AAA ind.	3	AAA	Ilalyi		ITC (ITC1564)		Kilimanjaro	Tanzania
75	Irumbe/Dabe	3	AAA	AAA ind.	3[c]	AAA	AAA ind.	Comoros		Wala II	Mwali	Comoros
105	Padji	3	AAA	AAA ind.	3[c]	AAA	AAA ind.	Comoros		Hoani	Mwali	Comoros
27	Mdzo Wa Djeni	3			3	AAA	Cavendish		CAPAM (SD40)		Mayotte	Mayotte
103	Kontrike	3	AAA	Cavendish	3[c]	AAA	Cavendish	Comoros		Hoani	Mwali	Comoros
20	Menaluki Dzilu	2			3	AAB	EAB AAB ind.		CAPAM (SD23)		Mayotte	Mayotte
21	Dzu Mwessi	3			3	AAB	Plantain		CAPAM (SD29)		Mayotte	Mayotte

TABLE 1. *Continued*

nid	Accession	Received as			Confirmed as			Plant material source		Origin of accession when known		
		Ploidy	Group	Subgroup	Ploidy*	Group	Subgroup	Collecting mission	Gene bank** (code)	Village	Region	Country
22	Dzu Dzilu	3			3	AAB	Plantain		CAPAM (SD30)		Mayotte	Mayotte
23	Dzu Djeu	3			3	AAB	Plantain		CAPAM (SD31)		Mayotte	Mayotte
94	Swankoubou	3			3[c]	AAB	Plantain	Comoros		Hayiraha	Mwali	Comoros
98	Betaloundu	3	AAB	Plantain	3[c]	AAB	Plantain	Comoros		Hoani	Mwali	Comoros
100	Dzu Moegne	3	AAB	Plantain	3[c]	AAB	Plantain	Comoros		Hoani	Mwali	Comoros
101	Dzu Mossi	3	AAB	Plantain	3[c]	AAB	Plantain	Comoros		Hoani	Mwali	Comoros
104	Minaluki	3	AAB	AAB ind.	3[c]	AAB	EAB AAB ind.	Comoros		Hoani	Mwali	Comoros
196	Mnalouki Gua	3	AAB	AAB ind.	3	AAB	EAB AAB ind.		CIRAD (PT-BA-00207)			Comoros
24	Shari'a	3	AAB	Pome	3	AAB	Pome		CAPAM (SD32)		Mayotte	Mayotte
25	Zabi	3			3	AAB	Mysore?		CAPAM (SD33)		Mayotte	Mayotte
26	Kissoukari (Dembeni)	3	AAB	AAB ind.	3	AAB	AAB ind.		CAPAM (SD34)		Mayotte	Mayotte
144	MW12-7				3	AAB	AAB ind.	Ethiopia		Gamo Gofa		Ethiopia
145	MW12-30				3	AAA	Mutika	Ethiopia		Gamo Gofa		Ethiopia
146	MW12-22				3	AAA	Mutika	Ethiopia		Gamo Gofa		Ethiopia
147	MW12-127				3	AAA	Cavendish	Ethiopia		Gamo Gofa		Ethiopia
172	Pahang	2	AAw	ssp. *malaccensis*.	2	AAw	ssp. *malaccensis*		ITC (ITC0609)			Malaysia
173	Hawain 2	2	AAw	ssp. *banksii*	2	AAw	ssp. *banksii*		CIRAD (PT-BA-00113)			Papua New Guinea
170	Zebrina	2	AAw	ssp. *zebrina*	2	AAw	ssp. *zebrina*		CIRAD (PT-BA-00433)			
191	EN13	2	AAw	ssp. *microcarpa*	2	AAw	ssp. *microcarpa*		CIRAD (PT-BA-00078)			Indonesia
192	PNG151	2	AAw	ssp. *banksii*	2	AAw	ssp. *banksii*		ITC (ITC1157)			
174	Pisang Klutuk Wulung	2	BB	BB	2	BB	BB		CIRAD (PT-BA-00302)			
175	Lal Velchi	2	BB	BB	2	BB	BB		CIRAD (PT-BA-00172)			
166	Balbisiana	2	BB	BB	2	BB	BB		CARBAP			
193	Pisang Mas	2	AA		2	AA	AAcv		CIRAD (PT-BA-00305)			
194	IDN110	2	AA		2	AA	AAcv		CIRAD (PT-BA-00131)			
176	Grande Naine	3	AAA	Cavendish	3	AAA	Cavendish		CIRAD (PT-BA-00104)			
181	Gros Michel	3	AAA	Gros-Michel	3	AAA	Gros-Michel		CIRAD (PT-BA-00106)			
182	Green Red	3	AAA	Red	3	AAA	Red		CARBAP			
183	Pisang Umbuk	3	AAA	Ambon	3	AAA	Ambon		CIRAD (PT-BA-00326)			
184	Merik	3	AAA	AAA PNG	3	AAA	AAA PNG		CARBAP		East Sepik	Papua N.G.
185	Mise Ehina	3	AAA	AAA PNG	3	AAA	AAA PNG		CARBAP			Papua N.G.
186	Pagatau	3	AAA	AAA PNG	3	AAA	AAA PNG		CARBAP			Papua N.G.
179	Yankambi km5	3	AAA	Ibota	3	AAA	Ibota		CIRAD (PT-BA-00426)			Congo D.R.
171	Pisang Kapas	3	AAB	Laknao	3	AAB	Laknao		CIRAD (PT-BA-00295)			Indonesia
162	Maritu	3	AAB	Iholena	3	AAB	Iholena		ITC (ITC0639)			Colombia
177	French clair	3	AAB	Plantain	3	AAB	Plantain		CIRAD (PT-BA-00094)			Cameroun
195	Ambo I	3	AAB	Mysore	3	AAB	Mysore		CARBAP			Congo D.R.

* 'c' indicates ploidy values confirmed by cytometry during this study.

** CAPAM, Chambre d'Agriculture de la Pêche et de l'Aquaculture de Mayotte; KARI, Kenya Agricultural Research Institute; CARBAP, Centre Africain de Recherches sur Bananiers et Plantains; ITC, Bioversity International Transit Centre; CIRAD, Centre de coopération internationale en recherche agronomique pour le développement.

AAw, wild AA; AAcv, cultivated AA variety; ind., indeterminate; nid, internal identification number.

mission by F. Bakry in 2010). DNA of 19 of these accessions was extracted for analysis from frozen leaf samples.

The 'Ambihy' seedy bananas, grow over 40 km along the Lokoho valley in north-eastern Madagascar, from Maroambihy up to Andapa, and are becoming a plague for the local populations (J.M. Leong, pers. comm., FOFIFA, Madagascar). This banana is represented in the herbarium of the Museum d'Histoire Naturelle in Paris (MNHN), from the survey by Humbert in 1948 in the same Lokoho valley. He declared it already abundant and invading the fresh lands along rivers and lands abandoned after cropping. It was not possible to extract DNA of sufficient quality from the leaf sample that MNHN agreed to take for analysis. However, leaf samples were taken for this study from six different plants in the Maroambihy region (J.M. Leong, 'Forests and Biodiversity' platform in partnership in Madagascar), of which four provided some analysable DNA. In addition, the AA 'Akondro Mainty' cultivar was supplied for analysis by the CIRAD collection in Guadeloupe, where it had been introduced from Madagascar.

A further 22 accessions of varied genotypes and origins were included in the analysis as references for the overall diversity of cultivated bananas. Thus, the various subspecies of *acuminata*, the species *balbisiana*, some cultivated AA of the *malaccensis* type and the main AAA and AAB subgroups were represented. The material was supplied by the CARBAP collection in Cameroon and the CIRAD collection in Guadeloupe, with some additions from ITC.

According to national policies, the dispatch of leaf tissue samples was covered by Material Transfer Agreements.

Ploidy analysis

Total DNA content was estimated by flow cytometry only on fresh leaf samples. Leaf samples were chopped together with leaf tissue of *Oryza sativa japonica* 'Nipponbare' as reference (haploid genome size 0.91 pg per nucleus) or AAA Cavendish banana (triploid genome size 1.72 pg per nucleus) as internal reference. Nucleus extraction was performed in 1 mL of a modified LB01 buffer (Dolezel *et al.*, 1989) in which mercaptoethanol was replaced by 40 mM sodium sulphite (Na_2SO_3). After filtration with a 30 µm nylon mesh to eliminate cell debris, 50 ppm of propidium iodide (PI) + 50 ppm of RNase was added to stain the nuclei. After 1 h in the dark, analyses were performed with a PAS II flow cytometer (Partec, Münster, Germany) using an argon ion laser (488 nm) for PI excitation (Dolezel and Bartos, 2005).

DNA content was measured on the accessions of Comoros and Mayotte for which the ploidy was not known but only inferred from morphological characters. In addition, accessions from Kenya and Tanzania were also measured to confirm previous estimates.

DNA extraction and SSR genotyping

Leaf samples were received at the CIRAD genotyping and robotics platform (Montpellier, France) for DNA extraction and genotyping. Leaf fragments from fresh as well as dried samples were ground in liquid nitrogen and genomic DNA was extracted following the MATAB protocol described by Risterucci *et al.* (2000).

A set of 23 SSR markers (primers are listed in Supplementary Data Table S1) was genotyped in the 89 accessions. This set of markers was chosen to optimize the coverage among the 11 chromosomes, using information about SSR marker positions on the DH-Pahang reference genome (D'Hont *et al.*, 2012; Martin *et al.*, 2016). From a previous set of 22 markers used to classify *Musa* germplasm (Grapin *et al.*, 1998; Lagoda *et al.*, 1998; Hippolyte *et al.*, 2010; Christelova *et al.*, 2011), 16 markers were kept for their ability to discriminate the main *Musa* subgroups. Nine other markers were added (Hippolyte *et al.*, 2010; D'Hont *et al.*, 2012), aiming to cover each chromosome with at least two markers, possibly one per arm, but two of them were discarded due to difficult allele scoring.

The PCR reaction mix (10 µL total volume) contained: 1.0 µL of PCR buffer (10 mM Tris-HCl, 50 mM KCl, 2 mM $MgCl_2$, 0.001% glycerol), 0.8 µL of dNTP (Jena Bioscience, Jena, Germany), 0.1 µL of $MgCl_2$, 0.08 µL of 10 µM forward primer with an M13 tail at the 5′ end (5′-CACGACGTTGTAAAACGAC-3′), 0.1 µL of 10 µM reverse primer, 0.1 µL of M13 tail fluorescently labelled with four dyes [6-FAM for blue, NED for yellow, VIC for green and PET for red (Applied Biosystems, Foster City, CA, USA)], 1 unit of Taq DNA polymerase (Sigma–Aldrich, St Louis, MO, USA), 5 µL of template DNA (5 ng μL^{-1}) and 2.32 µL of H_2O.

An initial denaturation at 94 °C for 5 min was followed by a touchdown PCR consisting of 35 cycles with denaturation at 95 °C for 45 s, annealing for 1 min with temperature decreasing by 0.5 °C every cycle from 55 to 50 °C (ten cycles), then 25 cycles at 50 °C, and a final extension at 72 °C for 10 min. PCR products were pooled in 10 µL of H_2O as follows: 2 µL of 6-FAM, 2 µL of VIC, 2.5 µL of NED and 3.5 µL of PET. Then, 2 µL of this solution was diluted 6-fold in Hi-Di formamide containing 0.12 µL of GeneScan 600 LIZ Size Standard and analysed on an ABI 3500xL Genetic Analyzer (Life Technologies, Carlsbad, CA, USA).

Alleles were scored using GeneMapper v4.1 software (www.appliedbiosystems.com), with careful manual corrections of automatic scoring. Synthetic parameters for the 23 SSRs, when observed on the whole set of 89 accessions, are listed in Supplementary Data Table S2.

Data analysis

Data analyses were performed using DARwin 5.6 software (Perrier and Jacquemoud, 2006).

The mixture of different ploidies and the reading ambiguity for triploid loci with two scored alleles (xy can be read as xxy as well as xyy) ruled out the use of a classic genetic distance for codominant markers. Considering these constraints, each marker was coded on as many columns as the alleles observed in the population for that marker, with a score of 1 or 0 if the allele was present or not (Supplementary Data File S1). From this table, a Dice index was used to calculate the dissimilarities for all possible pairs of accessions two-by-two on the set of markers without missing data for the two accessions.

Diversity trees were constructed by using the weighted neighbour-joining (NJ) algorithm (Saitou and Nei, 1987). The *balbisiana* cluster was considered as an outgroup for tree rooting.

Due to the disjunctive coding of the SSR data, the columns corresponding to the alleles of the same marker were

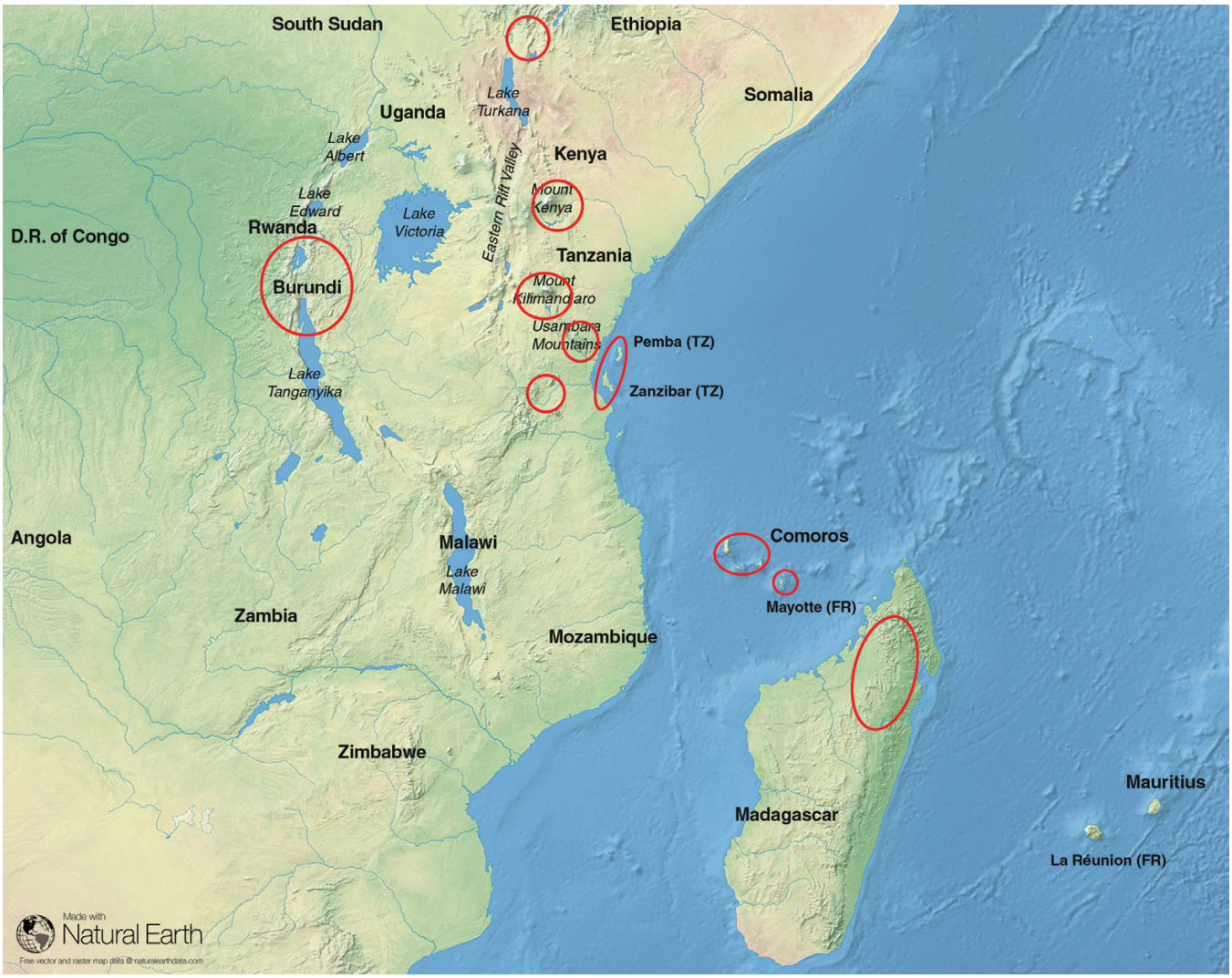

FIG. 2. The main sampling areas (base map from Natural Earth).

clearly not independent. Hence, random resampling on the columns, as done for bootstrap analysis, had no real meaning. However, the robustness of the inferred tree was checked using the 'influential unit detection' function proposed in the DARwin software. This is a jackknife-like procedure that iteratively removes one of the accessions, the resulting partial tree being compared with the whole tree in order to infer the contribution of this accession to the structure of the tree. A particularly high contribution indicates a questionable accession since a very different structure is produced when discarding it.

RESULTS

Ploidy of accessions

The accessions known as 'Mlali' at Comoros and Mayotte were diploid with DNA content ranging from 1.12 to 1.21 pg per nucleus. These values are common for edible diploids in banana. The observed tiny variations likely reflected experimental variation rather than actual inter-accession differences. The other analysed accessions were found to be triploid with a DNA content varying from 1.82 to 1.87 pg per nucleus, which are common values for triploid bananas (Table 1).

Distribution of dissimilarities

The dissimilarities calculated on all pairs of accessions two-by-two displayed a multimodal distribution (Fig. 3). A majority of dissimilarities was distributed according to a bell curve with a mode around 0.6. This distribution characterized the distances between individuals of a population with sexual reproduction. The secondary peak of this distribution around 0.8 corresponded to pairs opposing the two species *acuminata* and *balbisiana*.

Then, a group of values close to zero was identified where pairs were found between accessions of the same diploid AA or triploid AAA (Cavendish, for example) or AAB (Plantain, for example) subgroup. This group could be interpreted as a population with vegetative propagation. It was possible to fix the threshold between vegetative propagation and sexual reproduction at 0.1.

The secondary peak between 0.1 and 0.2 corresponded to some dissimilarities between highly related subgroups. For example, it included the distances between AAA Cavendish and AAA Gros-Michel accessions, included in this study as references, which are the main varieties feeding the word trade in sweet bananas. They are known to share the same AA 2n gamete donor parent, with the third genome A coming from two different but genetically related diploid AAs.

Diversity tree

A diversity tree (Fig. 4) was built from the dissimilarity matrix. It was rooted on the *M. balbisiana* cluster whose divergence from *M. acuminata* is ancient and necessarily prior to the creation of the cultivated diploids or triploids.

At the top of the tree appeared a large cluster grouping most of the cultivated AA accessions collected in the study zone (Mlali, Muraru, Mchare, etc.). The dissimilarities were extremely weak, evidencing a set in strict clonal propagation. No sub-structuring was detectable and the geographical origins (Kenya, Tanzania, Comoros, Madagascar) were completely intermingled.

The triploid AAA Gros-Michel and Cavendish were close to the cluster of diploids. Such proximity was expected since they share the same AA genome of Mlali/Mchare origin (Raboin *et al.*, 2005). The closely related unknown accessions 'Mdzo Wa Djeni' from Mayotte and MW12-127 from Ethiopia were therefore identified as AAA Cavendish, equivalent to 'Kontrike' from the Comoro Islands.

Also included in this set was the 'Sharia' accession from Mayotte, which is known as an AAB of the Pome subgroup of Indian origin. The position of this subgroup here was not surprising since the *acuminata* AA component has been shown to be of Mlali/Mchare origin (Perrier *et al.*, 2009). To this Pome accession was linked the 'Mumbwa Dzu' accession from the Comoro Islands. However, this accession was measured as diploid, which was confirmed by the absence of any SSR marker with more than two alleles, while all the triploids in our sample had several markers with three alleles. Moreover, the accession had several alleles which, in our sample, were specific to *balbisiana*: allele 166 of mMaCIR0307 and allele 175 of mMaCIR0196, for example. This accession was therefore identified as an AB, with an A genome close to Mchare/Mlali. The AB group is little represented in this geographical zone and only 'Kisubi', introduced to the ITC collection from Burundi, can be mentioned. The most conclusive references should be sought in India, where several AB varieties are cultivated, such as 'Safet Velchi' and 'Kunnan'.

A second large group, formed of two clusters, was linked to this first large set. The first was that of the wild AA forms of Madagascar and Pemba. The populations of the two islands appeared to belong to the same genetic population because the distances between accessions, while quite low, are high enough to exclude a population in vegetative propagation. It was interesting to find the 'Paka' accession included in this wild AA population. Paka, found in Zanzibar, was therefore the only form of cultivated AA to be outside the preceding subgroup of the AA Mchare.

The second cluster contained only AAA accessions with a uniform first sub-cluster of the AAA Mutika from the Great Lakes region and a second distinct sub-cluster, which was very uniform too, of the AAAs from Tanzania, Mayotte and the Comoro Islands. The 'Merik' accession, collected in Papua New Guinea (PNG), was grafted onto this cluster, along with the 'Padji' accession from the Comoro Islands, which appeared to be genetically different from the other Comoro AAAs. Two Ethiopian accessions were completely inserted in the Mutika subgroup and could thus be attached to that subgroup. Incidentally, accession MW12-30 was also said to be 'from Uganda' by the farmer.

The large set of bananas we have just examined then joined a cluster of AAB triploids. A first uniform sub-cluster of several accessions from the Comoro Islands and Mayotte was identifiable by the 'French Clair' reference accession belonging to the

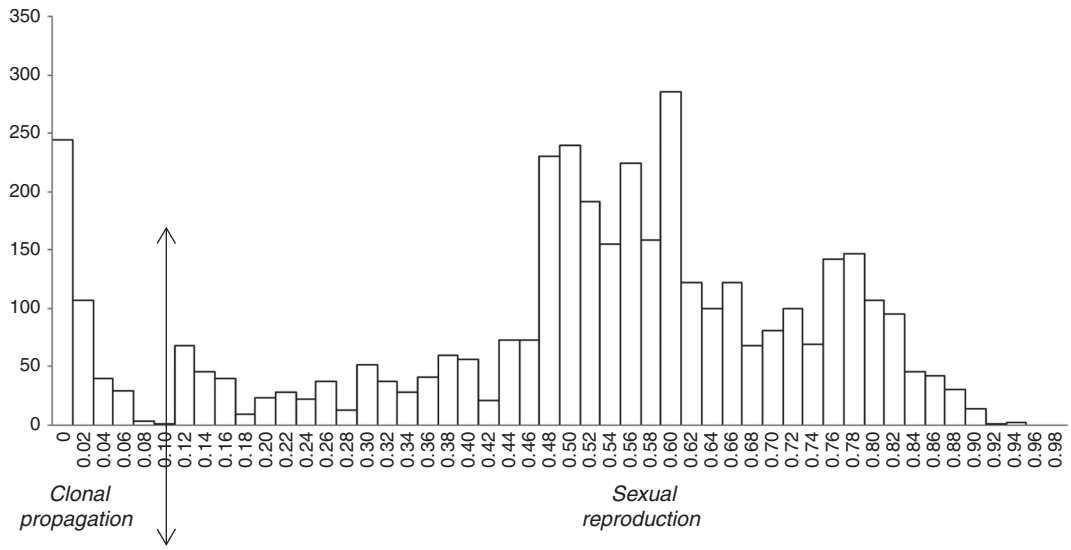

FIG. 3. Distribution in number of pairs of the dissimilarities (Dice index on disjunctive coding) calculated on the 89 accessions.

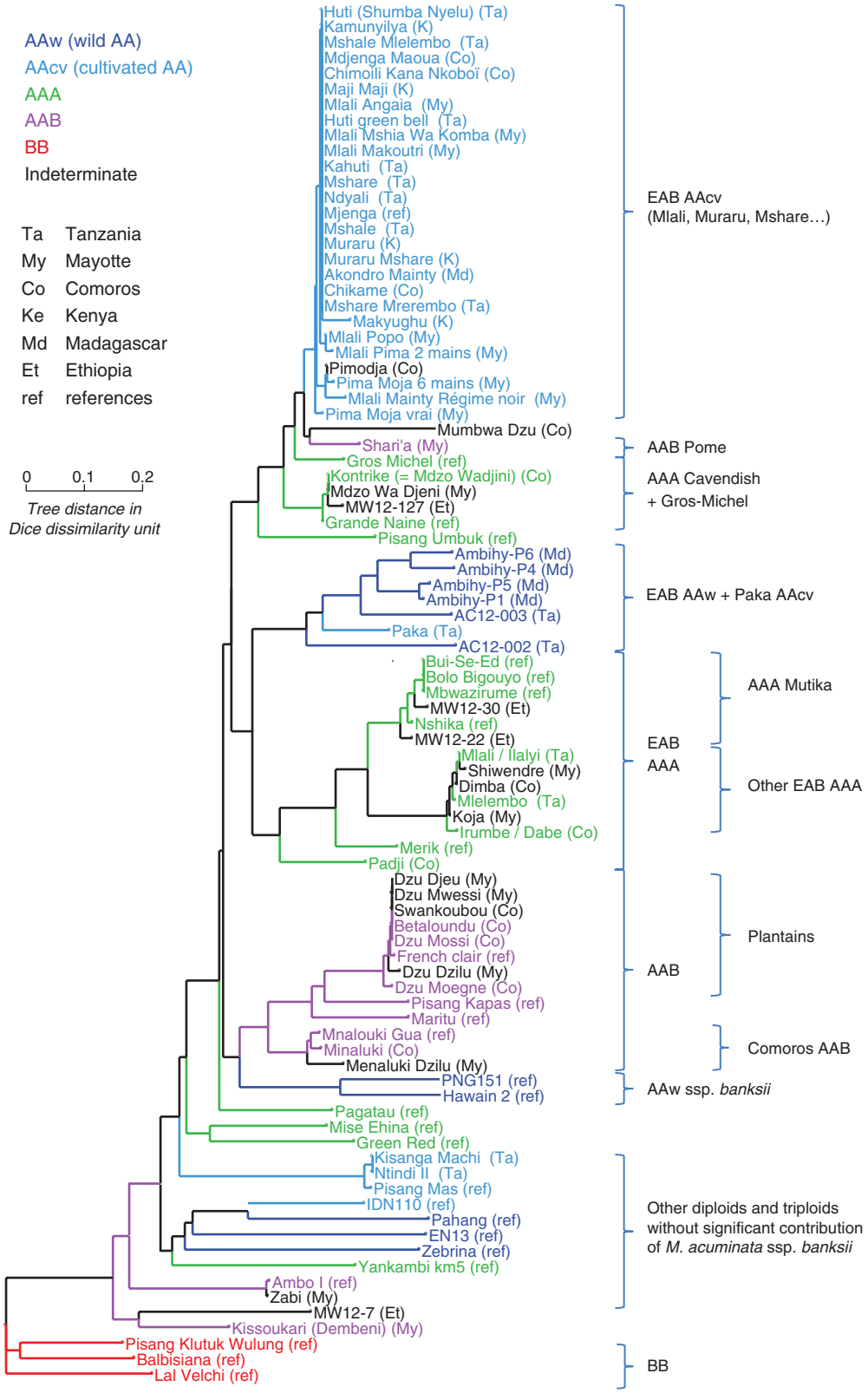

FIG. 4. Diversity tree (NJtree, DARwin software) from the dissimilarity matrix (Dice index on disjunctive coding) on 89 accessions and 23 SSR markers (rooted on the BB cluster).

Plantain subgroup. This subgroup was joined by the Laknao type accessions ('Pisang Kapas') then the AAB Iholena type ('Maritu'), which are known to be genetically close to the Plantains. More surprising was the proximity to these Plantains of the three representatives of the 'Mnaluki' accession (with some orthographical variations) collected in Mayotte and the Comoro Islands, along with the individual from the collection in Guadeloupe that had been formerly introduced from the Comoro Islands too. It was therefore shown that these AABs were close to the AAB Plantain, but that they were clearly detached from them and were even more distant than the AAB Laknao or Iholena. These AAB Plantain varieties have the subspecies *banksii* as their *acuminata* parent and the two *banksii* references were therefore logically found in proximity.

The rest of the tree no longer contained any specifically East African forms but made it possible to position the diversity of these East African bananas in relation to the representatives of the overall diversity. The only exceptions were forms imported, no doubt recently, from other regions. For instance, 'Kisanga Machi' and 'Ntindi' from Tanzania are typical representatives of the 'Sucrier' bananas, an AA subgroup that is widespread in all geographical zones. 'Zabi', collected in Mayotte, is an AAB Mysore like the 'Ambo I' reference. 'Kissoukari' from Mayotte is probably an AAB but not related to any of the main AAB subgroups. Today, the classification of these bananas remains vague, no doubt because the term 'kissoukari' in the sense of 'sweet' is used for some varieties that are clearly genetically different.

The search for some influential units of the tree detected two accessions playing a more marked role in the structure of the tree (data not shown). The first was the AAA 'Green Red', whose position in relation to the AAA from PNG was unstable, but this reference accession is of no direct interest here. The second was the AAA 'Padji' accession from the Comoro Islands. Its grafting *a posteriori* onto the tree constructed without the 'Padji' accession did not change its relations with the other East African AAAs but reversed the relation between the three large clusters, with the wild AA cluster now attached to the AA Mchare cluster before becoming attached to the AAA cluster. This instability was an invitation to view the three clusters rather on the same dissimilarity level.

DISCUSSION

A set of genetically homogeneous AA diploids

Whether they are called Mlali in the Comoro Islands, Mchare in Tanzania or Muraru in Kenya, all these edible AA diploids are genetically similar. The genetic dissimilarities between accessions are always very weak, and are interpreted as somatic variations within a vegetatively propagating population. The only exception was the 'Paka' cultivar, which is discussed later in this article.

Some earlier partial results (Onyango *et al.*, 2010; Hippolyte *et al.*, 2012) are thus confirmed on a more genetically wide and geographically extensive sample. The very high genetic homogeneity points to all these particular AA cultivars as belonging to the same subgroup, and it is proposed to classify it as the Mchare subgroup. The term 'Mchare' is widely used in Tanzania and often adopted in the literature to qualify these diploids (De Langhe *et al.*, 2001). The definition criteria are consistent with those adopted to identify subgroups among the triploid cultivars, the varieties of a subgroup arising from somatic variations starting from an initial successful genotype, either because it was much more appreciated than the others for a use of interest (e.g. the AAAs), or because of a lack of local competition over long periods, usually in zones very remote from the endemic zones (e.g. the AAB Plantains in Africa). Such subgroups are much rarer among the diploids (Daniells *et al.*, 2001). The Jari Buaya subgroup of bananas is consensual, but these accessions are very particular AA varieties which probably incorporate genome elements of another species, *Musa schizocarpa*. The subgroup AA Sucrier is also often recognized; its highly appreciated flavour qualities have prompted a very wide geographical dissemination, albeit recently and with limited somaclonal diversification – like the Cavendish subgroup, for example, in the AAA group. This Mchare subgroup is morphologically much more diversified and corresponds more to isolation far from the zone of origin, and in this sense is comparable to the Plantain subgroup in the triploids.

It has been shown that this Mchare subgroup shared some alleles with both the subspecies *zebrina* and *banksii* of *M. acuminata*, with a likely origin towards Borneo, Java or Sumatra (Perrier *et al.*, 2009; Hippolyte *et al.*, 2012), but that today no equivalent diploid forms are known in that region. It is therefore possible to put forward the hypothesis of a single-cultivar introduction from South-East Asia followed by wide dissemination on the islands of the Indian Ocean and in East Africa, resulting in many somatic mutants.

A Mchare subgroup with high phenotypic variability

The Mchare subgroup is characterized by many morphological traits: a slender pseudostem, often pale in colour but with dark blotches in some cultivars, like the 'Ijugu inkundu' (De Langhe *et al.*, 2001), an erect foliage habit, a subhorizontal to pendulous bunch bearing fruits with marked ridges and a pronounced bottlenecked apex, much recurved on a tightly packed bunch, slightly resembling the fruits of the AAB Plantain subgroup, though a little smaller.

But the subgroup displays extraordinarily wide diversity at inflorescence and fruit level, with a consequent distinct vernacular name for each variety that often refers to a morphological particularity (Fig. 1) (C. Jenny, unpubl. res.). For instance, incomplete bunches are found as in the AAB plantains of the Horn type, with one or two hands at the most, and an abruptly truncated rachis below the last hand ('Mlali Pima Moja'). The 'Samba' type presents a bunch structure resembling that of the False-Horn AAB plantains, with a rachis degenerating after the last hand and a virtually non-existent male bud. When the bunch is complete, the colouring of the bud bracts can vary from the classic violet to a discoloured pink ('Mlali Angaia'). Some yellow bract forms have also been described ('Shumba nyeelu', 'Ijugu inkundu'; De Langhe *et al.*, 2001). The rachis below the bunch may keep all its old bracts ('Mlali Mshia Wa Komba', 'Mlali Angaia') or be completely bare ('Mlali Makoutri', 'Mlali Commun'). The fruits are often pale, but 'Mlali Mulu' (also called 'Mlali Mainty') bears darker fruits. The size of the fruits is fairly variable and it can be very short in 'Mlali Makoutri' and 'Mlali Popo'.

It is interesting to note the morphological convergence in bunch morphology in different genetic backgrounds; the variability observed in the Mchare subgroup is also found within the AAB Plantain subgroup, and partially in other AAB and ABB subgroups.

The use of Mchare fruits can vary depending on the region; for instance, in the Comoro Islands and in Mayotte, Mchare fruits are harvested before full maturity and preferentially eaten cooked (fried or braised), whereas they are regarded as sweet bananas, like the AAB Silk, in East Africa (Karamura *et al.*, 2012). In the Kilimanjaro region they are roasted like the plantains in East Africa (De Langhe *et al.*, 2001).

A Mchare subgroup adapted to highly diverse ecological zones

The cultivation zone of these diploids covers Madagascar, the islands of the Comoro archipelago, the coastal zone of East Africa, including offshore islands such as Pemba and Zanzibar, and the hinterland of East Africa, especially in sufficiently humid altitudinal zones – the Pare Mountains, Kilimanjaro and Mount Kenya. But they appear only in few spots near the border between Uganda and Kenya and in the central region of Uganda, where they are commonly referred to as 'Bogoya atayengera', meaning 'Gros-Michel which does not ripen' (see below for the relation with AAA Cavendish and Gros-Michel) (D. Karamura, pers. comm.). Mchare accessions, despite having identical genotypes, can thus grow under very different ecological conditions, ranging from sea level in Zanzibar or Mayotte up to elevations of >1500 m (De Langhe *et al.*, 2001). This adaptive capacity is often highlighted for the triploid subgroups, attributing it to the presence of three versions of each gene, but here we have an example of how diploids can perform. It should be noted that no link could be seen between morphological variability and geographical or ecological distribution.

A Mchare subgroup with considerable cultural weight

The place of this diploid subgroup in cropping systems is highly variable. It is marginal in Madagascar, where 'Akondro Mainty' is the best known. In the Comoro Islands, the crop is ancient, with a large number of cultivars distinguished by farmers and identified with local names, but it remains marginal compared with the triploids. On the other hand, in the altitudinal zones of East Africa (Usambara Hills, Pare Mountains, Kilimanjaro) these bananas play a considerable cultural role. The Chaggas of the southern slopes of Kilimanjaro called themselves the 'Banana People', living in the 'Land of Bananas'; they grow various bananas in irrigated and very carefully tended home gardens, intercropped with tubers, coffee, etc. (Montlahuc and Philippson, 2003). Stabled cattle produce manure that enriches the already rich soil of the home gardens. This extremely elaborate cropping system would seem to have been established in the 12th century (Hemp and Hemp, 2008). The AA diploids are the preferred food of the Chaggas, the other bananas being much less appreciated; they also have many other uses. The banana lies at the heart of traditions and rituals, as well as myths and legends, indicating a fairly ancient adoption of this culture.

A companion set of seedy AA bananas

The wild seedy AA bananas from Pemba and Madagascar were found to be genetically close to each other, and thus probably form the same genetic population. They are bound to be plants introduced by human intervention, because the geographical areas involved are very distant from the endemic zones of wild *Musa* in Asia and near Oceania. Neither can it be a matter of feral forms escaped from a crop, as no reversion from the parthenocarpic form to the seedy form is known for banana. These plants may have been introduced accidentally in seed form. However, finding them on both islands would indicate that they were intentionally transported from South-East Asia, maybe for uses other than fruit, such as for fibre, in view of the importance of textile activities in the past in Madagascar.

This set was genetically linked to the AA Mchare subgroup but was distant enough from it to rule out a relation of direct kinship. In contrast, the 'Paka' cultivar, a parthenocarpic diploid grown in Zanzibar, was found to be genetically very close to this AA wild population. This cultivar cannot be attached to the Mchare subgroup and, as knowledge stands, it remains an isolated form. Its genetic proximity with the wild AA, as well as its geographical proximity (spatially restricted to the offshore islands of Tanzania), leads one to wonder about the possible local generation of this AA cultivar from the wild forms. However, not all 'Paka' alleles are found in the wild population and it can be shown that, for almost all the SSR markers, 'Paka' could be reconstructed from an AA of this wild population and an AA Mchare. These results are merely indicative since the wild population is only known from six representatives that cannot claim to represent the total allelic diversity of the population. The parthenocarpy of 'Paka' might thus be explained by hybridization between a fertile plant of this AA wild population and a genetically more distant AA Mchare form displaying residual male fertility. If this were to prove correct, it would be a unique case, to our knowledge, of the creation of a diploid AA cultivar outside the South-East Asian endemic zone for banana.

It should be remembered that this 'Paka' cultivar displays some traits of resistance to Sigatoka and black leaf streak diseases and it was used very early in sweet AAA banana breeding programmes (Champion, 1967). Understanding how it was formed could lead to other combinations of parents bearing these traits of interest.

A diversified set of East Africa-specific AAA varieties

While several triploid varieties (AAA Cavendish, AAB from India, some ABB, etc.) were introduced in more modern times, the traditionally grown AAA bananas are specific to East Africa and must have a long history there.

The AAA Mutika subgroup is found throughout the region of the Great Lakes (Uganda, Burundi, Rwanda). It is genetically very homogeneous (Karamura *et al.*, 2016; Kitavi *et al.*, 2016) but displays a multitude of morphologically distinguishable, recognized and named varieties (Karamura *et al.*, 2012), derived by somatic variations from perhaps even a single initial genotype, as already observed by Champion as early as 1970. This subgroup has been without doubt recorded in Africa only. The 'Biu-Se-Ed' variety, surveyed in 1961 on the island of Bali

in Indonesia (Rosales *et al.*, 1999), which is known as a Mutika for both its phenotype and genotype, would be a notable exception. However, the history of the specimens presently available in collections is not clear and this accession must be viewed with great caution.

The phenotypic diversity of these AAA Mutika has been extensively explored (Karamura, 1999) and five subsets, or 'clone sets', have thus been defined on morphological grounds. However, the wide observed phenotypic diversity is not supported by molecular markers (Karamura, 1999; Pickersgill and Karamura, 1999; Karamura and Pickersgill, 2000; Kitavi *et al.*, 2016).

This Mutika subgroup is not strictly limited to the Great Lakes region. Several varieties are grown around the northern side of Lake Malawi and a few are present in the Kilimanjaro or Pare Mountains region, particularly used for beer. However, in the latter regions these varieties are accompanied by a diversity of AAA bananas that do not belong to the Mutika subgroup. A survey team in Tanzania found in 2001 that 'sufficient indications pointed to the existence of cultivars grown in the Tanzanian highlands other than those of the Great Lakes region', and a specific subgroup was proposed with the name 'Ilalyi' (De Langhe *et al.*, 2001). Here these AAA cultivars formed a genetically homogeneous set, linked to the Mutika subgroup, but distant enough to rule out clonal relations. The validity of the proposed Ilalyi subgroup is thereby confirmed and mostly represented by the variety called 'Ilalyi' but also 'Mlali' in the Kilimanjaro region (Mochi and Rombo dialects of Chagga; hence the possible ambiguities since this term has been devolved to the AA Mchare in the Comoro Islands and Mayotte), or 'Mnyerere' in the Pare Mountains of Tanzania, where they are common. This subgroup includes also some accessions from Mayotte and from the Comoro Islands, in completely different ecologies, showing the high plasticity of this subgroup. It is shown here that the Mutika and Ilalyi subgroups have two of their three alleles in common for all the genotyped SSR markers. This suggests that both subgroups were formed from the same diploid non-reduced gamete donor, the other gamete coming from two different diploids. This would thus express the same type of relation as that shown, for example, between the AAAs Cavendish and Gros-Michel.

The 'Merik' accession, which will be mentioned later, was also attached to these two subgroups, along with the 'Padji' accession collected in the Comoro Islands, whose triploidy was confirmed and which has been morphologically identified as AAA; it may represent another subsampled AAA subgroup.

This diversity of the AAA genotypes was confirmed by the results of a recently published analysis (Karamura *et al.*, 2016) with 64 SSR markers in a set of 53 accessions: seven diploid accessions originally from Tanzania, 24 triploids from the East African highlands, 13 other triploids from Tanzania and nine exotic diploids or triploids as references. The diversity tree built from these data (Fig. 5) clearly isolated the Mutika subgroup; the Ilalyi subgroup was close to it but clearly distinct. A cluster of four other triploid varieties that may also be interpreted as a particular subgroup, here called Kitombo, was also attached to this set. Lastly, the 'Luholele' accession appeared to be a little different and was grafted onto the base of the Kitombo cluster. These Ilalyi and Kitombo subgroups, which differed from each other and differed from the AAA Mutika, also appeared in the diversity tree constructed on 171 triploids of the ITC collection from 19 SSR markers (Christelová *et al.*, 2011).

It therefore clearly appears that, beyond the Mutika of the Great Lakes countries, there exists to the east of the eastern Rift Valley and as far as the Comoro archipelago a diversity of other triploid AAAs that are genetically related to but different from the Mutika.

Some earlier results (Hippolyte *et al.*, 2012) showed contributions mainly of the *M. acuminata banksii* and *zebrina* subspecies for the genomes of these East African AAA cultivars. Another confirmation is the type Vβ of their cytoplasmic genomes (Carreel *et al.*, 2002), the chloroplastic V type being typical of the subspecies *banksii* and the mitochondrial β typical of the subspecies *zebrina*. Muiruri *et al.* (2017) also confirmed the contribution of subspecies *zebrina* to 'Mbwazirume', a popular Mutika accession, and to Ngombe, an AAA accession from Kenya. Therefore, the region of the basic origin of all these AAAs is likely the contact zone of the *banksii* and *zebrina* subspecies in present southern Indonesia.

The AA Mchare are not the direct parents of the East African AAAs

Since all the Mutika bananas have the same genotype, the subgroup is extremely vulnerable to pests and diseases as well as climate change. Perhaps alternative forms can be found among other East African AAA genotypes that would be better adapted to particular constraints. Varietal improvement of Mutika is another solution and important breeding programmes have been initiated.

Given their common geographical distribution, it has been proposed that the AA Mchare should be considered as the diploid ancestors of the East African AAA triploids, and it has even been suggested that the switch to the triploid state took place in East Africa, making this region a secondary centre of diversification. Such is not the case. The genetic distance between these two sets was too great and the number of alleles present in the triploids, but absent from the diploid forms, was too large.

In fact, we know of other AAA subgroups that actually derive from the AA Mchare subgroup and they do not look like Mutika at all. Indeed these AA Mchare have been proved to be genetically close to the AAAs Cavendish and Gros-Michel (Onyango *et al.*, 2010). Using restriction fragment length polymorphism (RFLP) markers, it had previously been shown (Raboin *et al.*, 2005) that these triploids incorporate the whole genome of 'Akondro Mainty', an edible AA accession from Madagascar, subsequently classified as a member of the Mchare subgroup. It was then shown that the Mlali accessions from Mayotte, also from the Mchare subgroup, were also good candidates as potential parents of these AAAs and that, in addition, they were also parents of the interspecific AAB triploids of the Pome subgroup from India (Hippolyte *et al.*, 2012). The other diploid parent of the Cavendish needs to be sought among the AAs derived from *M. acuminata* ssp. *malaccensis* from continental South-East Asia (Thailand, Cambodia, Vietnam). Likewise, the complementary diploid for the formation of the AAB Pome is of the 'Lal Velchi' type, a typical Indian *M. balbisiana* (Perrier *et al.*, 2009). By inferring an origin of the Mchare AAs in southern Indonesia, the hybridization with *malaccensis*-based AAs in mainland South-East Asia and with BBs in India implies the migration of these AA Mchare over a large distance. It has been shown that such movements of plant material can be related to human population movements, as indicated by the linguistic

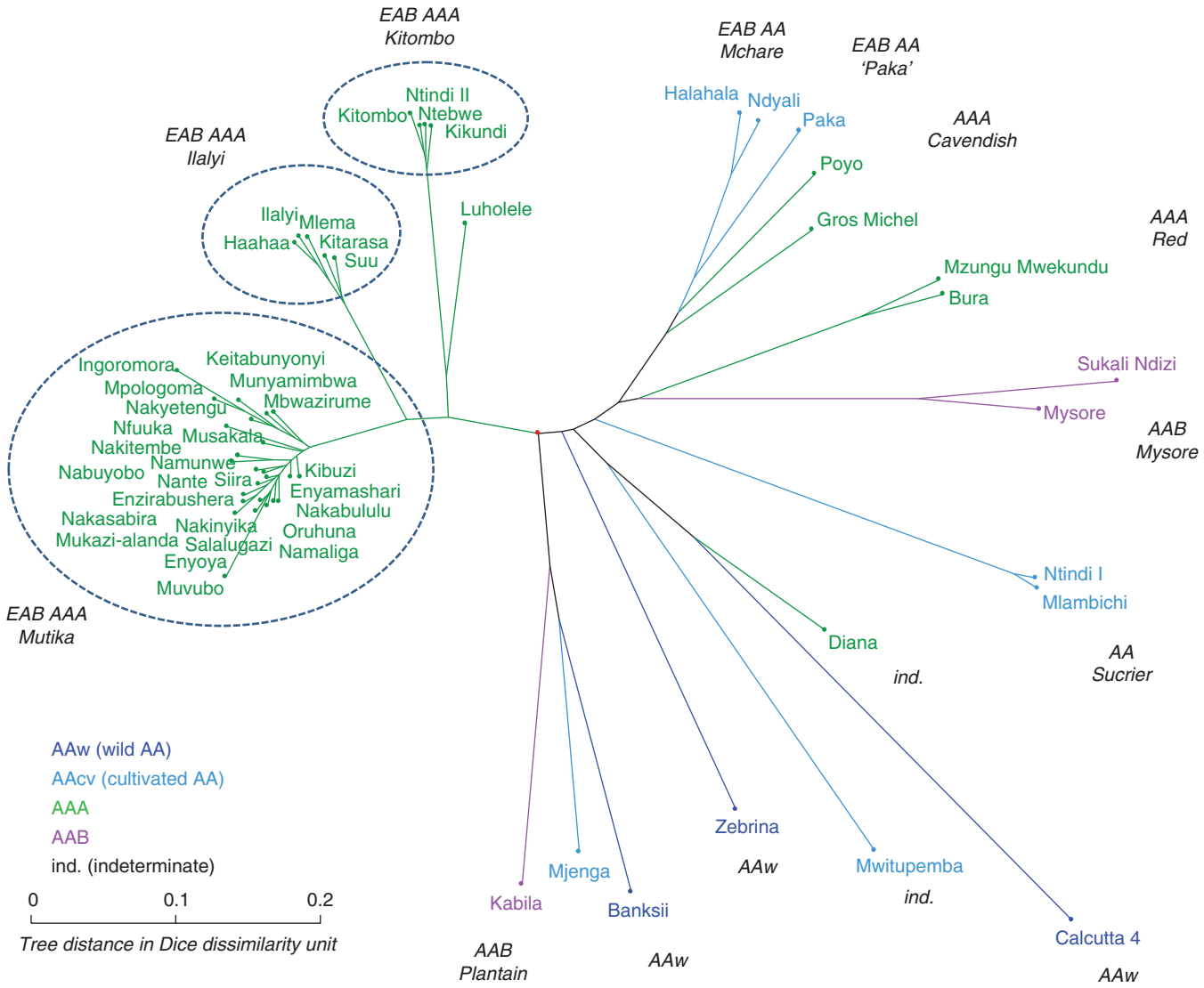

FIG. 5. Diversity tree (NJTree, DARwin software) from a Dice dissimilarity matrix on 53 accessions and 64 SSR markers. Data from Karamura *et al.* (2016).

analysis of the terminology specific to bananas in the various languages of the countries involved (Perrier *et al.*, 2011).

These diploids therefore have a remarkable history, since, in addition to their migrations through the South-East Asian mainland up to India, one must add the move across the Indian Ocean to East Africa. This gives rise to questions regarding the causes of such success, which, elsewhere, is only known for triploid subgroups.

A banana complex originating from the South-East Asian islands

Whether they are called Mchare by the Chaggas, Mnyali in Gweno, Ijighu in Asu, Huti by the Shambaas, Muraru in the Kikuyu language, Mhalihali in Zanzibar or Mlali on Pemba and in the Comoro Islands of Akondro in the centre of Madagascar, all these AA diploids are derived through somatic variations from a unique genotype that originated in the zone of contact between the subspecies *banksii* and *zebrina* in the southern Indonesia region. However, no AA Mchare are known today in that zone of origin, or in mainland South-East Asia and India, where they would have participated in the formation of some triploids. The East African Mchare are therefore, at the present time, the only representatives of this so successful subgroup in the past, which was at the origin of some of the most widely grown triploid forms today, the Cavendish and the Gros-Michel.

The genetic contributions of *banksii* and *zebrina* to the specific AAA East African subgroups were mentioned above and a zone of origin similar to that of the AA Mchare can be deduced. The only known reference in Asia for these triploids is the 'Biu-Se-Ed' accession from Bali, but its credibility is questioned above. Also known is the AAA 'Merik' accession, molecular markers of which have shown that it is genetically close to these AAAs (Fig. 4). This cultivar was surveyed in northern PNG (Sharrock *et al.*, 1989), but it is not a pure *banksii*-based AAA like the other AAAs collected in PNG. It probably incorporates another genome, possibly *zebrina*, in line with its localization at the eastern tip of the *banksii* and *zebrina* contact zone. Morphologically, it shares several traits with the East African AAAs, particularly the discolouration of the male bud and the stem. In any event, this accession attests to the

formation of this type of triploid in South-East Asia. Similarly, the Tala banana, in south-east Borneo, which has been recently described morphologically (Sunaryo *et al.*, 2017), could be a representative of these local AAAs because they were confirmed as being from traditional culture and not a recent introduction (Sunaryo *et al.*, 2017, pers. comm., Universitas Mulawarman, East Kalimantan, Indonesia). Indeed, from the very detailed description and photographs published, the resemblance to East African AAA Mutika is striking (C. Jenny and E. De Langhe, pers. comm.) but calls for genetic validation. The seedy bananas of Madagascar and Pemba were found to be genetically attached to these triploids (Fig. 4) and it could also be concluded that they have a similar geographical origin. Consequently, the notion of a particular banana complex emerges, which we propose to call 'EAB', for 'Eastern African Bananas'. This complex, which links seedy with parthenocarpic diploid as well as triploid forms, has a current spatial distribution strictly restricted to the East Africa region, and a common geographical origin in the southern Indonesian zone within the same *banksii* and *zebrina* inter-subspecific pool. One can thus assume that the transfer of this EAB complex results from one or several human migratory waves from south-eastern Indonesia to Africa across the Indian Ocean.

Uncertain archaeological dating

The question then arises of the dating of these movements of populations and of the cultivated plants they carried with them. Physical traces of banana cultivation in a remote past are extremely difficult to find because the edible banana does not produce durable seed, pollen or wood. Only phytoliths and starch grains can provide direct proof of an ancient banana as a crop. Thus, the archaeological digs undertaken at Munsa in Uganda, around 100 km west of Kampala and Lake Victoria, would have provided both *Musa* and *Ensete* phytoliths from swamp sediment cores, in levels dated to the fourth millennium BCE (Lejju *et al.*, 2006). This very remote dating, which was to have considerable implications for the history of plant domestication in Africa, has largely been challenged, both regarding the identification of the phytoliths and the dating proposed (Neumann and Hildebrand, 2009). Moreover, the authors expressed themselves with great caution.

The only other similar discovery comes from Nkang in Cameroon and would seem to indicate the presence of bananas, probably an AAB Plantain given the location, ~2500 BP in Central and West Africa (Mbida *et al.*, 2000). The exact dating is hampered by a large plateau in the calibration curve and is in conflict with other archaeological data indicating, at that time, a strongly seasonal climate very unsuitable for banana; dating around 2000 BP could therefore be more realistic (Neumann and Hildebrand, 2009).

The very great morphological diversity among >100 known AAB Plantain varieties argues in favour of a long period of accumulation of somatic variations, as estimated by a mutational model by De Langhe *et al.* (1994). However, apart from western Uganda and a few notable exceptions elsewhere (particularly the Nyakyusas, a Bantu cultivator group, in South Tanzania, north of Lake Malawi; Maruo, 2007), there is little Plantain cultivation in East Africa. Conversely, EAB are rare or absent in the Congo Basin and West Africa. It would be risky to extend this archaeological dating point specific to the Plantain subgroup to the EAB. Questions therefore need to be specifically asked about the age of the introduction of this EAB complex in agricultural landscapes in East Africa, known to emerge at the end of the first millennium BCE and involving domesticated plants in Africa, such as sorghum, finger millet and yam.

As with the AAB Plantain, the large somatic diversification, at least for AA Mchare and AAA Mutika, suggests a long period of selection by farmers. However the difficulties in knowing the actual mutation rate, the intensity of selection pressure or the rate of spatial dissemination prevent any reliable estimation of these durations.

Bananas in founding legends and myths

The reference to bananas in founding legends and myths could also be indicative of the age of introduction. For the Chaggas of Kilimanjaro, banana is completely embedded in traditions and is seen to be as ancient as the people themselves. Some legends recount that on their arrival the country was occupied by men of small stature, the Vakoningos, who had mastered iron making and banana cultivation (Montlahuc and Philippson, 2003). If this legend is to be believed, it would indicate an introduction of banana prior to the arrival of these Bantu populations in the first millennium.

As regards AAA Mutika, they are absent from the legends of western Uganda, Burundi and Rwanda, which were largely spread as support for the legendary Bacwesi empire (Chretien, 1985). This is not surprising because banana cultivation, although extremely relevant in these regions, only took on its importance in the colonial periods (Cochet, 2001). On the other hand, in Buganda or Busoga, in the northern zone of Lake Victoria, the legend of the founding hero Kintu explicitly mentions banana among the benefits he brought, along with iron, cattle, yam, etc.; it is specified that Nambi, the wife of Kintu, brought a banana plant from an area near Mount Elgon in the north-east of Uganda and that all the current banana plants are derived from that plant (McMaster, 1963). This legend, even if it accurately describes what is observed, i.e. a very strong bottleneck with the introduction of bananas restricted to one or a few basic Mutika plants followed by intense diversification and diffusion via vegetative propagation, does not tell us anything about the dating of these events.

Some convincing linguistic elements

Ultimately, linguistics remains at present the most informative approach. It seems acknowledged that the first Bantu populations who arrived from the West/Central African forest zone at the turn of our era did not know banana, as none of the banana-related terms has a root in the ancestral Bantu languages. For example, in Chagga languages none of the words for banana belong to the Bantu family and have thus all been borrowed, no doubt since early times because they have mostly followed the evolutionary rules of Bantu words, whereas terms relative to sweet potato or tobacco, introduced more recently, did not (Philippson, 2007). The oldest names relative to banana seem to derive from those already used for *Ensete*. *Ensete*, which is still widely cultivated in

Ethiopia, once formed a veritable '*Ensete* belt' in East Africa (De Langhe *et al.*, 1994), of which relicts can still be seen in altitudinal places down to the Usambara Hills, and was used as a food in times of scarcity. The plant maintains a profound cultural significance all over East Africa. For instance, among the Chaggas of Kilimanjaro, it is an attribute of power strictly reserved for the garden of the chief (Montlahuc and Philippson, 2003). It has been suggested that this ancient care of *Ensete* in Africa contributed to the rapid and widespread adoption of the bananas arriving from Asia (De Langhe *et al.*, 1994).

The comparative analysis of the vocabulary in the different Great Lakes languages led Schoenbrun (1993) to suggest that the initial stages of AAA cooking/beer banana growing took place on the East African coast during the first half of the first millennium. The Bantu-speaking peoples who settled in the region at the turn of our era then incorporated banana into their cropping system. Banana cultivation then crossed the obstacle of the Eastern Rift and reached Lake Victoria before the end of the first millennium. It then spread westwards and southwards in the following centuries, in the Great Lakes zone, as attested for example by the creation of entirely new terms for banana by the Rutaran-speaking peoples, in northern Tanzania, between Rwanda and Lake Victoria. The major development of banana cultivation, hence its cultural weight, is then reflected in the enrichment of the dedicated vocabulary. These terms are specific to the different languages and therefore were innovated after the divergence of those languages, so they can be dated for the Great Lakes region to the period circa AD 1200–1500. However, the greatest diversification of the vocabulary occurs post-15th century in Buganda, even more recently, as we have seen, in Rwanda and Burundi, and is no doubt still going on. Today, this scenario is often accepted with slightly earlier dates (Stephens, 2015). However, one cannot rule out that bananas arrived prior to the settlement of the Bantu populations on the coast of East Africa and were early adopted by local populations, such as the Vakinongos of Chagga legends.

Dissemination pathways of the EAB complex in Africa

While genetic diversity is broad east of the East Rift valley, it is drastically reduced to the AAA Mutika alone in the Great Lakes zone. This filter does not appear to come from particular ecological constraints and the adaptive capacity of both the AAs and the AAAs has already been highlighted. It seems rather to be down to the history of the concerned societies, with partial isolation between the coastal zone and the hinterland (Chami, 1994).

The pathways of movement westwards are still disputed. The legend of Kintu would seem to indicate an arrival in Buganda, with later dissemination to the edges of the Congo, via the north of Lake Victoria, from Mount Elgon. A pathway via the South is also suggested based on linguistic arguments (Karamura, 1999). 'Tooke', the general word for Mutika bananas in Uganda, is found in a broad corridor through western Tanzania and northern Malawi and reaches the Tanzanian coast via the Ruvuma valley, where 'Tooke' covers more than just the AAA Mutika subgroup.

The place of introduction into continental Africa might be indicated by the term 'Huti', which designates 'bananas cooked with their skins' and is used for the AA Mchare subgroup by the Shambaa- and Bondei-speaking Bantu, respectively, on the Usambara Hills and on the lowland area to the coast (De Langhe *et al.*, 2001). This term is a reflex of the Malayo-Polynesian *punti for 'banana', widespread in central/eastern Indonesia (Donohue and Denham, 2009). Similarly, 'hontsy' and 'fontsy', other reflexes from *punti, are used for 'banana' in many parts of Madagascar (Beaujard, 2017).

The role of the Indian Ocean islands

The offshore islands such as Zanzibar, Pemba and Mafia were certainly favoured entryways but they do not inform about the introduction into Africa. More questions arise as to the role of the islands of the Comoro archipelago and, further away, Madagascar as stepping stones for the introduction of plants from South-East Asia: banana, taro (*Colocasia esculenta*) or water yam (*Dioscorea alata*) according to Murdock's (1959) hypothesis of the 'Tropical Food Kit'.

Even though these islands were probably visited episodically >4000 years ago (Gommery *et al.*, 2011; Dewar *et al.*, 2013), the beginnings of significant and permanent occupation for agriculture is set rather at the end of the first millennium. Indeed, populations from South-East Asia speaking Austronesian languages related to the Barito languages of south-eastern Borneo, at the origin of the current Malagasy (Dahl, 1951, cited for example in Beaujard, 2011), arrived in Madagascar and the Comoros, between the eighth and ninth centuries. Moreover, Malagasy includes many Malayan terms relative, in particular, to navigation (Adelaar, 2006), along with terms specific to the Celebic languages of Sulawesi (Blench, 2015). On this point, the Asian origins of the Malagasy terms 'ontsy' and 'fontsy' should be remembered here. The Malagasy terms 'ambihy' can be added, referring to the seedy bananas that would derive from the South-East Asian term 'vihy', 'bigi' meaning 'seed' (Beaujard, 2011). Contacts with the African continent subsequently developed, leading to the incorporation of numerous Bantu terms in Malagasy, including the Malagasy term 'akondro' used for several varieties, derived from the Bantu root 'kondo'. This multiple origin is confirmed by the genomic analysis of the current human populations showing admixture between African and South-East Asian components (Kusuma *et al.*, 2015; Brucato *et al.*, 2016). The South-East Asian component corresponds more precisely to the Banjar people of south-east Borneo, who result from admixture between a group from south-east Borneo, Southern Sulawesi or Maluku, and Malay populations. The Malay kingdoms, which mastered sea navigation very early, had effectively established trading posts on Java and Sumatra as early as the second half of the first millennium.

This plurality of human population origins is also found in the Comoro Islands (Msaidie *et al.*, 2011), but the current language there is a Bantu one related to the Swahili dialects of the African coast (Grollemund *et al.*, 2015), likely as result of intensive contacts.

These South-East Asian components cut rather precisely across the assumed zone of EAB origin, along with the period of their arrival, which tallies with the dates provided for the arrival of EAB on the African continent, making the Austronesian

populations who crossed the Indian Ocean highly credible candidates for the introduction of EAB in East Africa.

In fact, this Austronesian colonization could have reached Madagascar and the Comoros before Africa. Crowther *et al.* (2016), analysing crop seeds found at archaeological sites from the 8th to the 12th century CE, showed that African crops (mainly pearl millet, sorghum and finger millet) were largely in the majority at the African sites and on the offshore islands and that Asian crops (mainly rice, cotton and mung bean) did not appear until the 11th century. Conversely, Asian crops dominated throughout the Comoros and the Malagasy sites as early as the eighth century, indicating an earlier arrival. Of course, these results cannot take into account the non-seedy Asian crops like banana, water yam and taro, which may have arrived much earlier than seedy crops. However, the presence only in Indian Ocean islands of seedy diploids is an argument for this proposition. Similarly, even if the importance of banana, when compared with other crops, is presently limited in Malagasy myths and rituals, traces of an ancient influence are still clearly legible, indicating banana was more important at an earlier time (Beaujard, 2017). For example, the traditions of the Betsileo (in southern central Madagascar) very explicitly refer to banana, particularly in funeral rites. There is also mention of slaves assigned to guarding banana and taro crops (Rangan *et al.*, 2015).

Several introduction events

The different proto-subgroups of the EAB complex, i.e. their basic genotypes, were not necessarily introduced at the same time, and may have been carried during the different waves of migration across the Indian Ocean. These various waves did not necessarily pass through Madagascar or the Comoros and direct arrivals can also be assumed in the ports of Zanzibar or Pemba, opposite Shamba'a-Bondei country, maybe even before the settlement of the Austronesians in Madagascar (Blench, 2010).

We propose here an initial arrival of the AA proto-Mchare, followed at later stages by the different AAA subgroup proto-cultivars. This would explain the deep cultural involvement of these Mchare for several populations in Tanzania or Kenya. The triploid forms are not much appreciated in these countries and are cultivated only because of their greater robustness, as for example the Mchare plots that are frequently surrounded by triploids for wind protection. Further west, the barrier of the Rift Valley halted the spread of these diploids into the Great Lakes region and it is probably slightly by chance that, among the various AAAs, a 'proto-Mutika' reached the vacant niche of the Great Lakes region as an improved substitute for *Ensete*, with the success we now know. Subsequent introductions of attractive varieties would have taken place from the African coast to the Comoros and Madagascar, as illustrated by the numerous banana-related terms of Bantu origin.

CONCLUSIONS

This contribution reveals the existence of an East Africa-typical (EAB) complex including at least three distinct triploid subgroups but also a diploid subgroup, the AA Mchare. Members of this subgroup are already of particular interest for breeding programmes focused on the sweet AAA bananas of the Cavendish subgroup. The insertion of Mchare into a larger EAB complex including the Mutika AAA subgroup opens new perspectives on how the Mutika subgroup, so important in East Africa, can also be improved. The phylogenetic relations existing between the different EAB components need to be more precisely established and their relations to the ancestral wild *acuminata* forms from South-East Asia remain to be specified. Technologies available today that provide a large number of single-nucleotide polymorphisms should enable these necessary advances.

In addition, this diploid Mchare subgroup could be used as a model for several popular subgroups that display a similar wide phenotype diversity due to centuries-long vegetative propagation of an initial single genotype, such as AAB Plantain, AAB Maoli-Popoulu and ABB Bluggoe. These subgroups are usually triploid, which greatly complicates the study of the genetic or epigenetic determinism of such phenotypes.

Finally, our tentative reconstruction of the history of the EAB complex should be useful in the effort to elucidate the history of the other popular starchy bananas in Africa, the AAB Plantain subgroup. It has become clear that, although the two entities, EAB and Plantain, share a similar initial movement from Southeast Asia to Africa, they underwent their evolution in different places and probably at different times as well. The basic Plantain genotype was generated in the Philippines or nearby (De Langhe *et al.*, 2015), and the subgroup got its extraordinary diversity in the rainforest zone of Africa (De Langhe, 2007). In contrast, the EAB originated in southeastern Indonesia and diversified in the more altitudinal East Africa. Consequently, one can now approach the problem more precisely with the question of whether the plantains were introduced before, during or after the EAB event, the period of the latter being estimated at early first millennium CE, and what kind of Austronesian-speaking ancestors were involved.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following. File S1: SSR data for the 89 genotyped accessions (data_23SSR_89accessions.txt) Table S1: SSR markers, chromosomal position on the *Musa* genome reference, motif, forward and reverse primers, and origin. Table S2: synthetic parameters for the 23 SSR markers on the 89 *Musa* accessions.

ACKNOWLEDGEMENTS

We would particularly like to thank the late Dr Margaret Onyango [Kenya Agricultural Research Institute (KARI), now Kenya Agricultural Research and Livestock Organization (KARLO)], who selected from the collection in Kisii the accessions to be analysed and ensured the preparation of the material. She assured us from the very beginning of her interest in this study and her intention to participate. The 'collectivité territoriale de Mayotte' must also be thanked for providing access to the material from the island. We thank Jean-Michel Leong Pock Tsy (FOFIFA, Centre National de la Recherche Appliquée au Développement Rural, Madagascar) who, especially for this study, went to the Lokoho Valley to collect

leaf material from wild banana plants. We also thank Alison Crowther (School of Social Science at The University of Queensland, Australia) and Michele Wollstonecroft (Institute of Archaeology at University College London). They provided us with material collected during the Sealinks project (funded by the European Research Council), which focused on the earliest maritime connections that linked up societies around the Indian Ocean; Alison collected accessions from the Pemba Island, and Michele from Ethiopia. We also thank the managers of the three major international collections that kindly provided us with plant material:

The Bioversity International Musa Germplasm Transit Centre (ITC), which ensures the conservation of more than 1500 banana accessions kept *in vitro*, with a frozen duplicate for security. The collection is hosted at the Katholieke Universiteit Leuven (KU Leuven). The conserved germplasm is placed in the Multilateral System of Access and Benefit Sharing of the International Treaty on Plant Genetic Resources for Food and Agriculture.

The 'Tropical plants' Biological Resource Centre in Guadeloupe, which maintains *in vivo* cultivars and crop wild relatives for several tropical plants. For banana, around 450 accessions representative of worldwide diversity, are maintained in the field.

The CARBAP (African Centre for Research on Bananas and Plantains) banana germplasm collection at Njombe (Cameroon). It is one of the most important *in vivo* collections, with more than 700 accessions that have been maintained and observed in the field for many years.

We also want to thank Steeve Joseph for the many hours spent on DNA extraction, and to make a special mention of Christian Leclerc (UMR AGAP, CIRAD), who made a significant contribution to improving the writing of the document.

LITERATURE CITED

Adelaar A. 2006. The Indonesian migrations to Madagascar: making sense of the multidisciplinary evidence. In: Simanjuntak T, Pojoh IHE, Hisyam M. eds. *Austronesian Diaspora and the Ethnogeneses of People in Indonesian Archipelago. Proceedings of the International Symposium*. Jakarta: LIPI Press, 205–231.

Baker RED, **Simmonds NW. 1951**. Bananas in East Africa. Pt. I. The botanical and agricultural status of the crop. *Empire Journal of Experimental Agriculture* **19**: 283–290.

Baker RED, **Simmonds NW. 1952**. Bananas in East Africa. Pt. II. Annotated list of varieties. *Empire Journal of Experimental Agriculture* **20**: 66–76.

Beaujard P. 2011. The first migrants to Madagascar and their introduction of plants: linguistic and ethnological evidence. *Azania* **46**: 169–189.

Beaujard P. 2017. *Histoire et voyages des plantes cultivées à Madagascar avant le XVIe siècle*. Paris: Karthala.

Blench R. 2010. Evidence for the Austronesian voyages in the Indian Ocean. In: Anderson A, Barrett JH, Boyle KV, eds. *The global origins and development of seafaring*. Cambridge, UK: McDonald Institute for Archaeological Research, 239–248.

Blench R. 2015. Interdisciplinary approaches to stratifying the peopling of Madagascar. In: Akshay Sarathi ed. *Proceedings of the Indian Ocean Conference, Madison, Wisconsin*. Oxford: Archaeopress.

Brucato N, **Kusuma P**, **Cox MP**, ***et al.* 2016**. Malagasy genetic ancestry comes from an historical Malay trading post in southeast Borneo. *Molecular Biology and Evolution* **33**: 2396–2400.

Byabachwezi M, **Shaban M**, **Karamura D. 2005**. *Tanzania Musa expedition 2005 - Morogoro-Mbeya region*. Uganda: INIBAP, Kampala.

Carreel F, **Gonzalez de Leon D**, **Lagoda PJL**, ***et al.* 2002**. Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. *Genome* **45**: 679–692.

Chami F. 1994. *The Tanzanian coast in the first millennium AD. An archaeology of the iron-working, farming communities*. Uppsala: Societas Archaeological Upsaliensis.

Champion J. 1967. *Les bananiers et leur culture. Vol. 1. Botanique et génétique des bananiers*. Paris: I.F.A.C, Editions SETCO.

Champion J. 1970. Culture du bananier au Rwanda. *Fruits* **25**: 161–168.

Chretien JP. 1985. L'empire des Bacwezi. La construction d'un imaginaire géopolitique. *Annales Économies, Sociétés, Civilisations* **40**: 1335–1377.

Christelová P, **Valarik M**, **Hribova E**, ***et al.* 2011**. A platform for efficient genotyping in *Musa* using microsatellite markers. *AoB Plants*: plr024.

Cochet H. 2001. *Crises et révolutions agricoles au Burundi*. Paris: Karthala.

Crowther A, **Lucas L**, **Helm R**, ***et al.* 2016**. Ancient crops provide first archaeological signature of the westward Austronesian expansion. *Proceedings of the National Academy of Sciences of the USA* **113**: 6635–6640.

Daniells JW, **Jenny C**, **Karamura D**, **Tomekpé K. 2001**. *Musalogue: a catalogue of Musa germplasm. Diversity in the genus Musa*. Montpellier: INIBAP.

Dewar RE, **Radimilahy C**, **Wright HT**, **Jacobs Z**, **Kelly GO**, **Berna F. 2013**. Stone tools and foraging in northern Madagascar challenge Holocene extinction models. *Proceedings of the National Academy of Sciences of the USA* **110**: 12583–12588.

D'Hont A, **Denoeud F**, **Aury JM *et al*. 2012**. The banana *Musa acuminata* genome and the evolution of monocotyledonous plants. *Nature* **488**: 213–219.

Dolezel J, **Bartos J. 2005**. Plant DNA flow cytometry and estimation of nuclear genome size. *Annals of Botany* **95**: 99–110.

Dolezel J, **Binarova P**, **Lucretti S. 1989**. Analysis of nuclear DNA content in plant cells by flow cytometry. *Biologia Plantarum* **31**: 113–120.

Donohue M, **Denham T. 2009**. Banana (*Musa* spp.) domestication in the Asia-Pacific region: linguistic and archaeobotanical perspectives. *Ethnobotany Research & Applications* **7**: 293–332.

FAO. 1986. *Rehabilitation of traditional schemes (Phase I)*. Tanzania AG: TCP/URT/4522, Mission Report, Rome.

Gommery D, **Ramanivosoa B**, **Faure M**, ***et al.* 2011**. Les plus anciennes traces d'activités anthropiques de Madagascar sur des ossements d'hippopotames subfossiles d'Anjohibe (Province de Mahajanga). *Comptes Rendus Palevol* **10**: 271–278.

Grapin A, **Noyer JL**, **Carreel F**, ***et al.* 1998**. Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged microsatellite sites. *Electrophoresis* **19**: 1374–1380.

Grollemund R, **Branford S**, **Bostoen K**, **Meade A**, **Venditti C**, **Pagel M. 2015**. Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences of the USA* **112**: 13296–13301.

Hemp C, **Hemp A. 2008**. The Chagga homegardens on Kilimanjaro. *IHDP Update* **2**: 12–17.

Hippolyte I, **Bakry F**, **Seguin M**, ***et al.* 2010**. A saturated SSR/DArT linkage map of *Musa acuminata* addressing genome rearrangements among bananas. *BMC Plant Biology* **10**: 65.

Hippolyte I, **Jenny C**, **Gardes L**, ***et al.* 2012**. Foundation characteristics of edible *Musa* triploids revealed from allelic distribution of SSR markers. *Annals of Botany* **109**: 937–951.

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

Karamura DA, **Pickersgill B**, **Vuylsteke D**, **Gold CS**, **Karamura EB**, **Kiggundu A. 2000**. Multivariate analyses of supposedly duplicate accessions of East African Highland bananas in germplasm collections in Uganda. In: Craenen K, Ortiz R, Karamura EB, Vuylsteke D, eds. *Proceedings of the International Conference on Banana and Plantain for Africa.* Leuven: International Society for Horticultural Science. *Acta Horticulturae* **540**: 89–90.

Karamura D, **Njuguna J**, **Nyamongo D. 2006**. *Kenyan Musa expedition 2006*. Montpellier: Bioversity.

Karamura D, **Karamura E**, **Tinzaara W. 2012**. *Banana cultivar names, synonyms and their usage in East Africa*. Kampala: Bioversity International.

Karamura D, **Kitavi M**, **Nyine M**, ***et al.* 2016**. Genotyping the local banana landrace groups of East Africa. *Acta Horticulturae* **1114**: 67–74.

Kitavi M, **Downing T**, **Lorenzen J**, ***et al.* 2016**. The triploid East African Highland Banana (EAHB) genepool is genetically uniform arising from a single ancestral clone that underwent population expansion by vegetative propagation. *Theoretical and Applied Genetics* **129**: 547–561.

Kusuma P, **Cox MP**, **Pierron D**, ***et al.* 2015**. Mitochondrial DNA and the Y chromosome suggest the settlement of Madagascar by Indonesian sea nomad populations. *BMC Genomics* **16**: 191.

Lagoda PJL, **Noyer JL**, **Dambier D**, **Baurens FC**, **Grapin A**, **Lanaud C. 1998**. Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Molecular Ecology* **7**: 659–663.

De Langhe E. 2007. The establishment of traditional plantain cultivation in the African rain forest: a working hypothesis. In: Denham T, Irarte J, Vrydaghs L. eds. *Rethinking agriculture*. Walnut Creek, CA: Left Coast Press, 361–370.

De Langhe E, **Swennen R**, **Vuylsteke D. 1994**. Plantain in the early Bantu world. *Azania* **29–30**: 147–160.

De Langhe E, **Karamura D**, **Mbwana A. 2001**. *Tanzania Musa expedition 2001*. Kampala: INIBAP.

De Langhe E, **Perrier X**, **Donohue M**, **Denham T. 2015**. The original banana split: multi-disciplinary implications of the generation of African and Pacific plantains in island Southeast Asia. *Ethnobotany Research & Applications* **14**: 299–312.

Lejju BJ, **Robertshaw P**, **Taylor D. 2006**. Africa's earliest bananas? *Journal of Archaeological Science* **33**: 102–113.

Martin G, **Baurens FC**, **Droc G**, ***et al.* 2016**. Improvement of the banana "*Musa acuminata*" reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics* **17**: 243.

Maruo S. 2007. Development of the plantain-based culture of the Nyakyusa of southern Tanzania. *African Study Monographs* **34** (Supplement): 21–38.

Mbida CM, **Van Neer W**, **Doutrelepont H**, **Vrydaghs L. 2000**. Evidence for banana cultivation and animal husbandry during the first millennium BC in the forest of southern Cameroon. *Journal of Archaeological Science* **27**: 151–162.

McMaster DN. 1963. Speculations on the coming of the banana to Uganda. *Uganda Journal* **27**: 163–175.

Montlahuc ML, **Philippson G. 2003**. Avant le café : le lexique de l'agriculture pré-coloniale. In: Bart F, Mbonile J, Devenne F. eds. *Kilimandjaro. Montagne, mémoire, modernité*. Espaces Tropicaux No. 17. Bordeaux: Presse Universitaire de Bordeaux, 81–88.

Msaidie S, **Ducourneau A**, **Boetsch G**, ***et al.* 2011**. Genetic diversity on the Comoros Islands shows early seafaring as major determinant of human biocultural evolution in the Western Indian Ocean. *European Journal of Human Genetics* **19**: 89–94.

Muiruri KS, **Britt A**, **Amugune NO**, **Nguu E**, **Chan S**, **Tripathi L. 2017**. Dominant allele phylogeny and constitutive subgenome haplotype inference in bananas using mitochondrial and nuclear markers. *Genome Biology and Evolution* **9**: 2510–2521.

Murdock GP. 1959. *Africa: its peoples and their culture history*. New York: McGraw Hill.

Mzemouigni SM. 2007. *Caractérisation morpho-taxonomique et cytogénétique des bananiers des Comores*. MSc Thesis, Université Pierre et Marie Curie, France.

Neumann K, **Hildebrand E. 2009**. Early bananas in Africa: the state of the art. *Ethnobotany Research & Applications* **7**: 353–362.

Nsabimana A, **Gaidashova SV**, **Nantale G**, **Karamura D**, **Van Staden J. 2010**. Banana cultivar distribution in Rwanda. *African Crop Science Journal* **16**: 1–8.

Onguso JM, **Kahangi EM**, **Ndiritu DW**, **Mizutani F. 2004**. Genetic characterization of cultivated bananas and plantains in Kenya by RAPD markers. *Scientia Horticulturae* **99**: 9–20.

Onyango M, **Haymer D**, **Keeley S**, **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.

Perrier X, **Jacquemoud-Collet JP. 2006**. *DARwin software*. http://darwin.cirad.fr/

Perrier X, **Bakry F**, **Carreel F**, ***et al.* 2009**. Combining biological approaches to shed light on the evolution of edible bananas. *Ethnobotany Research & Applications* **7**: 199–216.

Perrier X, **De Langhe E**, **Donohue M**, ***et al.* 2011**. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. *Proceedings of the National Academy of Sciences of the USA* **108**: 11311–11318.

Philippson G. 2007. Langue et histoire au Kilimanjaro et alentour: diffusion démique et diffusion culturelle. In: Deslaurier C, Juhé-Beaulaton D. eds. *Afrique, terre d'histoire. Au coeur de la recherche avec Jean-Pierre Chrétien*. Paris: Karthala, 50–71.

Pickersgill B, **Karamura D. 1999**. Issues and options in the classification of cultivated bananas with particular reference to the East African Highland bananas. In: Andrews S, Leslie AC, Alexander C. eds. *Taxonomy of cultivated plants. Third International Symposium*. Kew: Royal Botanic Gardens, 159–167.

Pillay M, **Ogundiwin E**, **Nwakanma DC**, **Ude G**, **Tenkouano A. 2001**. Analysis of genetic diversity and relationships in East African banana germplasm. *Theoretical and Applied Genetics* **102**: 965–970.

Raboin LM, **Carreel F**, **Noyer JL** ***et al*. 2005**. Diploid ancestors of triploid export banana cultivars: molecular identification of 2n restitution gamete donors and n gamete donors. *Molecular Breeding* **16**: 333–341.

Rangan H, **Alpers EA**, **Denham T**, **Kull CA**, **Carney J. 2015**. Food traditions and landscape histories of the Indian Ocean world: theoretical and methodological reflections. *Environment and History* **21**: 135–157.

Risterucci AM, **Grivet L**, **N'Goran JA**, **Pieretti I**, **Flament MH**, **Lanaud C. 2000**. A high-density linkage map of *Theobroma cacao* L. *Theoretical and Applied Genetics* **101**: 948–955.

Rosales F, **Arnaud E**, **Coto J. 1999**. *A tribute to the work of Paul Allen. A catalogue of wild and cultivated bananas.* Montpellier: INIBAP.

Saitou N, **Nei M. 1987**. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.

Schoenbrun DL. 1993. Cattle herds and banana gardens: the historical geography of the western Great Lakes region, ca AD 800–1500. *African Archaeological Review* **11**: 39–72.

Sharrock S. 1989. *Report on the fourth IBPGR/QDPI banana germplasm collecting mission to Papua New Guinea, 04/6/1989 to 9/7/1989*. Nambour, Australia: IBPGR.

Shepherd K. 1957. Banana cultivars in East Africa. *Tropical Agriculture* **34**: 277–286.

Simmonds NW, **Shepherd K. 1952**. An Asian banana (*Musa acuminata*) in Pemba, Zanzibar Protectorate. *Nature* **169**: 507–508.

Stephens R. 2015. *A history of African Motherhood: the case of Uganda, 700–1900*. New York: Cambridge University Press.

Sunaryo W, **Nurhasana, Rahman**, **Sugiarto A. 2017**. Identification and characterization of Talas banana, a superior local cultivar from East Kalimantan (Indonesia), based on morphological and agronomical characters. *Biodiversitas* **18**: 1414–1423.

Tugume AK, **Lubega GW**, **Rubaihayo PR. 2002**. Genetic diversity of East African Highland bananas using AFLP. *InfoMusa* **11**: 28–32.

Walsh M. 2009. The use of wild and cultivated plants as famine foods on Pemba Island, Zanzibar. *Études Océan Indien* **42–43**: 217–241.