



Combining Biological Approaches to Shed Light on the Evolution of Edible Bananas

Xavier Perrier, Frédéric Bakry, Françoise Carreel, Christophe Jenny, Jean-Pierre Horry, Vincent Lebot and Isabelle Hippolyte

Research

Abstract

Deciphering the diversity of the banana complex needs a joint characterization and analysis of the original wild species and their relatives, primitive diploid forms and triploid derived varieties. Sexuality, the primary source of diversity, is strongly disrupted in the cultivated varieties (sterility, parthenocarpy and vegetative propagation) by human selection of vegetatively maintained punctuated mutations. Many biological tools are available for characterizing this diversity, each one illustrating some peculiar facets, and we show that their joint analysis enables an evolutionary reading of this diversity. We propose various scenarios regarding the structure of wild species, on the domestication of the edible diploids from hybrids between wild forms, on the direct ancestry of triploids from cultivated diploids, and on the ancient migrations dispersing cultivated forms around the world. The comparison with data from archaeology, linguistics and human genetics will enable the validation, refinement and dating of the proposed domestication process.

Introduction

Banana is a major plant for humanity. Dessert or cooking triploid banana is an international traded commodity, and also a staple food for millions of people in Oceania, Asia, India, Africa and Latin America. Moreover, multi-use diploids remain popular in numerous Asian traditional societies. This importance stimulated considerable research during the last century. Simmonds (1962) presented an impressive synthesis of the salient results concerning botanical classification, the relative contributions of species *Musa acuminata* Colla and *Musa balbisiana* Colla, plant ecology, polyploidy, parthenocarpy and so on. More recent works have clarified and extended several aspects of Simmonds' work, but without negating his general findings (Jenny *et al.* 2003). However, results obtained over the last ten years lead us to fundamentally reconsider some of the previous conclusions regarding the historical process

of evolution that led to the present diversity, including the role of people in the various processes of domestication. These major revisions have several causes. First, germplasm resources have been greatly extended, particularly with the intensive collecting mission in Papua New Guinea (PNG) initiated by the INIBAP network (Sharrock 1990). It thus appeared that the long isolation of human societies in New Guinea (NG) and surrounding islands has maintained some important traces of banana evolution (Tezenas du Montcel 1990), which are now available for analysis. Second, new biological tools have been developed, which give an insight into the genome itself and on relationships between genomes. Finally, findings in archaeology, linguistics, ethnobotany and other social sciences have added to our understanding (Denham *et al.* 2003, Donohue & Denham 2009, Fuller & Madella 2009, Mbida *et al.* 2001, Neumann & Hildebrandt 2009).

We focus here on the contribution of recent biological methods to the understanding of the process of banana domestication. In the first part, we present a condensed history of the research based on biological methods and

Correspondence

Xavier Perrier, Frédéric Bakry, Jean-Pierre Horry, Isabelle Hippolyte, CIRAD (Centre de coopération internationale en recherche agronomique pour le développement), Avenue Agropolis - TAA-75 / 02 - 34398 Montpellier Cedex 5, FRANCE.

Xavier Perrier, xavier.perrier@cirad.fr

Françoise Carreel, CIRAD. Campus international de Baillarguet - TAA-75 / 02 - 34398 Montpellier Cedex 5, FRANCE.

Christophe Jenny, CIRAD. Station de Neufchâteau, Sainte-Marie, 97130 Capesterre-Belle-Eau, GUADELOUPE.

Vincent Lebot, CIRAD. CARFV, Department of Agriculture, PMB 946 Port Vila, VANUATU.

Ethnobotany Research & Applications 7:199-216 (2009)

Published: July 30, 2009

www.ethnobotanyjournal.org/vol7/i1547-3465-07-199.pdf

their respective contributions. In the second part, we examine how new biological techniques and their results can shed light on evolutionary and domesticatory processes. We conclude with a call for closer collaboration between the biological and human sciences to elucidate outstanding questions regarding the history of banana domestication.

Biological Tools to Study Evolution

Biological tools developed to explore the diversity of a species and to infer its evolutionary history exploit mutational events that have affected the genome. If not lethal or deleterious, these modifications are fixed and sexually transmissible to the descendants. Phylogenetic analyses are fundamentally based on this property of inheritance and assume that two individuals sharing a mutation inherit this from a common ancestor. Combining several mutations enables the building up of resolved phylogenies. Nevertheless, different mutations may produce the same apparent effect. This situation, called homoplasy, can lead to the mistaken association of individuals derived from different ancestors, thereby producing inaccurate phylogenetic constructions. Another difficulty of phylogenetic analysis is sexuality, which disperses genomic modifications to different lineages.

Studies of the genome give different types of information depending on the level considered. They range from the whole genome to a single allele, passing through chromosome, gene and gene expression.

Structural events at the genome level

At the level of the whole genome, cytological studies published as early as 1910 by Tischler and in 1928 by White suggested that the basic haploid chromosome number in bananas was 4. Subsequently, Cheesman (1932) proposed that the basic haploid number was equal to 11 and that the most common condition among edible bananas was triploidy and not octoploidy as previously stated. These results were later confirmed by Cheesman and Larter (1935) who provided the true chromosome number of most wild species and cultivated varieties.

Major modifications of the whole genome are deletions or insertions of large genomic sequences that modify the size of the genome. Thus it was shown that the DNA quantity per nucleus in *M. balbisiana* is lower than in *M. acuminata* (Dolezel *et al.* 1994). Significant variation in DNA content have also been found among the different subspecies of *M. acuminata* (Jenny & Carreel n.d.).

More rare events concerning the whole genome modify the chromosome structure, such as duplication of a short segment of the same chromosome or transfer of a segment from one chromosome to another. These chromosomal rearrangements are revealed by observations of

chromosome pairing during meiosis. They are quite frequent in *M. acuminata*. Shepherd (1999) reported that wild accessions within *M. acuminata* subspecies do not show structural heterozygosity, meaning that pairs of homologous chromosomes have the same sequence alignment. However, hybrids between two different structural groups display chromosomal heterozygosity with rearrangements between pairs of homologous chromosomes.

Mutations in DNA sequences and their markers

Besides these structural modifications, mutations in DNA sequence may arise due to errors in DNA duplication leading to the substitution of one base by another, or to the insertion (or the deletion) of a few consecutive nucleotides. These mutations are spread throughout the genome and are assumed to occur by chance, producing an approximately linear evolutionary process through time. These mutations are mainly neutral, but can also modify the morphology or the metabolism when gene-coding regions are involved. To detect these DNA mutations, or nucleotide/genetic polymorphisms, molecular markers are now widely used in diversity and phylogenetic analyses.

The earliest markers used to detect polymorphism between individuals were morphological traits. The in-depth analysis of these morphological markers by several generations of scientists led to clear characterizations of the wild species and the cultivated forms of bananas, and to the establishment of the basic botanical classification, which is still in large part valid and has been only marginally corrected since. In 1865, Kurz proposed for the first time the hypothesis of a bispecific origin of cultivars: *M. acuminata*, which donated the so-called A genome; and, *M. balbisiana*, which donated the so-called B genome. Simmonds and Shepherd (1955), using a taxonomic scoring method, specified the relative contribution of the two species in the constitution of current varieties. The broad diversity of the species *M. acuminata* was classified into inter-fertile subspecies on the basis of morphological characteristics and geographic distributions. The classification was later refined by Shepherd (1999) who divided this species into seven groups based upon chromosomal structure.

Variations in morphological traits points to gene mutations, unfortunately of unknown nature and location in the genome. Other major limitations are that morphology is susceptible to environmental conditions and that multi-genic regulations can frequently induce homoplasy. Moreover, the traits may be strongly biased by selection, which filters a large part of mutations.

Novel methods arose as biochemistry developed. They enabled studies at the metabolic level, analyzing different kinds of metabolites, their concentration and the different forms of enzymes involved in their metabolism (Markert & Moller 1959). These methods still revealed the expression of the genome rather than the genome itself, but they indicated more directly genome mutations and the risk of

homoplasmy was limited. They remain under selective influence. In *Musa*, the analysis of polyphenols suggested for the first time several centers of domestication for *M. acuminata* varieties (Horry & Jay 1988). In addition, this study revealed several steps of evolution for the anthocyanin patterns among *M. acuminata* subspecies. Isozymes also contributed to the understanding of the *Musa* complex (Horry 1989, Jarret & Litz 1986, Lebot *et al.* 1993), confirming the morpho-differentiation of the complex into major genome groups. Subspecies-specific band patterns were identified in *M. acuminata* which readily distinguished itself from *M. balbisiana*. These works evidenced for the first time the contribution of *M. acuminata* ssp. *banksii* (F. Muell.) Simmonds derived forms from New Guinea to the plantains of Africa and the Pacific.

More recently, several molecular markers have been used to detect polymorphisms directly at the DNA level.

The types of polymorphism and the techniques of detection are briefly described on Table 1. Unaffected by the environment, these markers have proven to be effective tools for distinguishing between genotypes, even closely related ones. These mutations can affect any part of the genome, but they are in general considered to be neutral in relation to natural or human selection.

Musa has been successfully studied with different molecular markers such as RFLP (Carreel *et al.* 1994, Gawel *et al.* 1992, Jarret *et al.* 1994), AFLP (Noyer *et al.* 2005, Ude *et al.* 2002a, b, Wong *et al.* 2001, 2002); SSR (Creste *et al.* 2004, 2006, Grapin *et al.* 1998, Lagoda *et al.* 1998) and DArT (Risterucci, in press). In association with ploidy level, all these molecular methods led to clearer representations of the *Musa* complex, specifying relations between its components. Besides the nuclear genome, chloroplasts and mitochondria genomes are also subject to mutational

Table 1. Molecular marker techniques available for diversity analysis.

RFLP (Restriction Length Fragment Polymorphisms) The RFLP technique (Botstein *et al.* 1980) was largely used during the last decade. The whole DNA sequence is cut into numerous fragments by an enzyme, which recognizes specific short nucleotide sequences (3-6 base pairs). The generated fragments are separated by differential migration according to their length. If a fragment is labeled by a marked probe, it becomes readable as a band on the migration profile. In case of mutation – an insertion/deletion or a substitution altering the cutting sequence and thus merging two consecutive fragments – the length of the fragment is modified, giving a band at a different position on the migration image and so a different allele. Several combinations of restriction enzymes and probes are used, detecting different mutation loci. RFLP markers are codominant and give access to all the alleles of a locus. They have demonstrated their efficiency for diversity analysis; their main limitation is to be time consuming and labor intensive.

PCR (Polymerase Chain Reaction) PCR is not directly a molecular marker technique. It is a method of DNA amplification based on the ability of a specific enzyme to generate copies of a targeted sequence of DNA. This sequence duplicated a million times is thus easily discriminated from the rest of the genome. The PCR technique was a decisive advance for molecular marker analysis, reducing drastically time and expense for analysis.

AFLP (Amplified Fragment Length Polymorphisms) As for RFLP, DNA is cut into fragments by restriction enzymes, and a large subset of fragments is then amplified by PCR (Vos *et al.* 1995). After migration and coloration, these fragments are directly readable. AFLPs detect the same polymorphism as RFLPs but give access to a greater number of mutation sites. However, the alleles of the same locus cannot be associated, limiting the genetic information.

DArT (Diversity Arrays Technology) DArT technology is a recent application of hybridization methods (Jaccoud *et al.* 2001) that is also based on DNA restriction by specific enzymes. Alleles identified in a set of accessions representative of the diversity are initially spotted on a small array. This array is used to detect the presence of these alleles in any new accession. The result is an image which can be automatically analyzed, making DArT a high throughput technique enabling the rapid detection of thousands of alleles on thousands of accessions, but as with AFLP the alleles of the same locus are not identified.

SSR (Simple Sequence Repeats) Instead of detecting variations in length of DNA fragments, SSR markers pinpoint polymorphisms in numbers of repeats for stretches of consecutively repeated small units (1 to 6 nucleotides) (Condit & Hubbell 1991). Variations in the number of repeats induce length variations which are detected after PCR amplification. The mutation rate is higher than for point mutations (approximately 10^{-3} against 10^{-9}). Consequently, SSR is able to generate many more polymorphic markers compared to other methods and able to detect recent polymorphism between closely related accessions. These molecular markers are widely used in genetic studies because they are spread across the whole genome, quite easy to generate and codominant. They can be automated and so have great potential in large-scale studies.

SNP (Single Nucleotide Polymorphism) SNP detects point substitutions to a single nucleotide and these are identified by genome sequencing (Brookes 1999). Automation by high density array and great availability in all genomes (more than 1 per 1000 base pairs) makes the technique very attractive for diversity analyses. Until now these markers have not been used for banana diversity. The scheduled project of the sequencing of the whole *Musa* genome will enable in the close future considerable development of this technique.

events. These cytoplasmic genomes are haploid and are inherited from only one of the two parents. In *Musa*, the chloroplast genome is transmitted by the mother and the mitochondrial genome by the father (Fauré *et al.* 1994). This single-parent inheritance is particularly effective to trace lineages. Cytoplasmic genome analysis played a decisive role in elucidating the relationship between diploids and triploids (Carreel *et al.* 2002). Besides punctual gene mutations, modifications in mechanisms regulating the gene transcription can induce phenotypic polymorphism in groups that are solely vegetatively propagated (plantains for example), which can be tracked using specific molecular markers (Baurens *et al.* 2003).

Methodology

In the following section, we will try to synthesize previous analytical results in a comprehensive and plausible history of domestication. We will successively consider the wild species, the relations between cultivated diploids and wild relatives, and finally the evolution from diploids to triploids. The phylogenetic relations concern primarily the *M. acuminata* components and only marginally the *M. balbisiana* components. This is due to the limited number of available accessions of the latter in germplasm collections. Although *M. balbisiana* is said to be less variable and consequently to play a lesser part in the diversity of cultivars, recent results suggest a more important diversity which should be investigated (Ge *et al.* 2005).

We will strengthen this synthesis with the unpublished results of a study with 22 SSR markers on a wide sample including nearly 500 accessions comprising wild and cultivated forms, as well as diploids and triploids from collections in Guadeloupe, Cameroon and Nigeria (Generation Challenge Programme Project comprising CIRAD, International Institute of Tropical Agriculture, Bioversity International). Several statistical methods were used to analyze these data: factorial analysis that depicts in low dimensional space the dissimilarities between accessions, and phylogenetic analysis that infers evolutionary trees best fitted to these dissimilarities. A difficulty arose from the mixture of several levels of ploidy. Therefore, we defined a general measure of dissimilarity as a probability of parentage between two accessions, regardless of the ploidy level. Population genetic methods were used on the basis of allelic frequencies observed in the defined groups. In addition, a specific method was developed to detect direct filiations between accessions. For a target diploid compared to a pair of potential diploid parents, a marker was regarded as positive if one of its alleles was found in one of the putative parents and the other one in the other putative parent. Each diploid was successively taken as a target and a kinship score was calculated for each pair of diploids as the proportion of positive markers. High score values indicated that the two diploids were potential parents or, more exactly, were closely related to these parents. For a triploid, a marker was positive when two alleles

were found in the first parent, regarded as the 2N gamete donor, and the third one was in the second parent. When no complete filiation was detected, the constraint was relaxed to identify at least the 2N gamete donor (for example for an AAB, the two A genomes coming from the same parent) then, if necessary, to identify two distinct diploids transmitting two alleles of the triploid (for an AAB, the two A genomes coming from two different parents).

Evolution and Domestication

Recently, a relevant terminology was proposed for the main steps of banana domestication (De Langhe *et al.* 2009). We will follow this terminology to show how biological results can support the hypothesized transformations between these variously termed bananas.

Wild AA diploids

Biologists cannot address directly the distinction between wild and cultivated in the sense of De Langhe's terminology. They can only approach it by observing levels in seed fertility. As a consequence, we will consider seeded forms among *M. acuminata* diploids first.

Nuclear molecular markers (RFLP and SSR) differentiate *M. acuminata* seeded diploids into four distinct clusters (Figure 1), which fit with four of the subspecies defined using morphological characters:

- (1) ssp. *banksii* cluster from New Guinea.
- (2) ssp. *malaccensis* (Ridl.) Simmonds cluster from Malayan Peninsula.
- (3) ssp. *burmanica* Simmonds, *burmanicoides* De Langhe and *siamea* Simmonds. They cannot be discriminated by their nuclear genome and they form a complex that covers the northern part of the *acuminata* area centered on northeast India, Burma, southern China and Thailand (Figure 2). This complex and the *malaccensis* cluster are genetically closer than they are to other subspecies. They are geographical neighbors and several accessions from Thailand appear to be hybrids between them.
- (4) ssp. *zebrina* (Van Houtte ex J. É. Planchon cluster from Java. Its genome size is significantly, c. 10%, higher than the three other groups (Jenny & Carreel n.d.). Regarding polyphenol components (Horry & Jay 1988) *zebrina* appears as the most primitive form, followed by *banksii*, the other subspecies being at a more advanced stage of evolution.

These four groups are clearly discriminated by nuclear markers; they correspond with specific chloroplast and mitochondrial types (Figure 1) as well as with strong chromosome rearrangements (Shepherd 1999) and specific

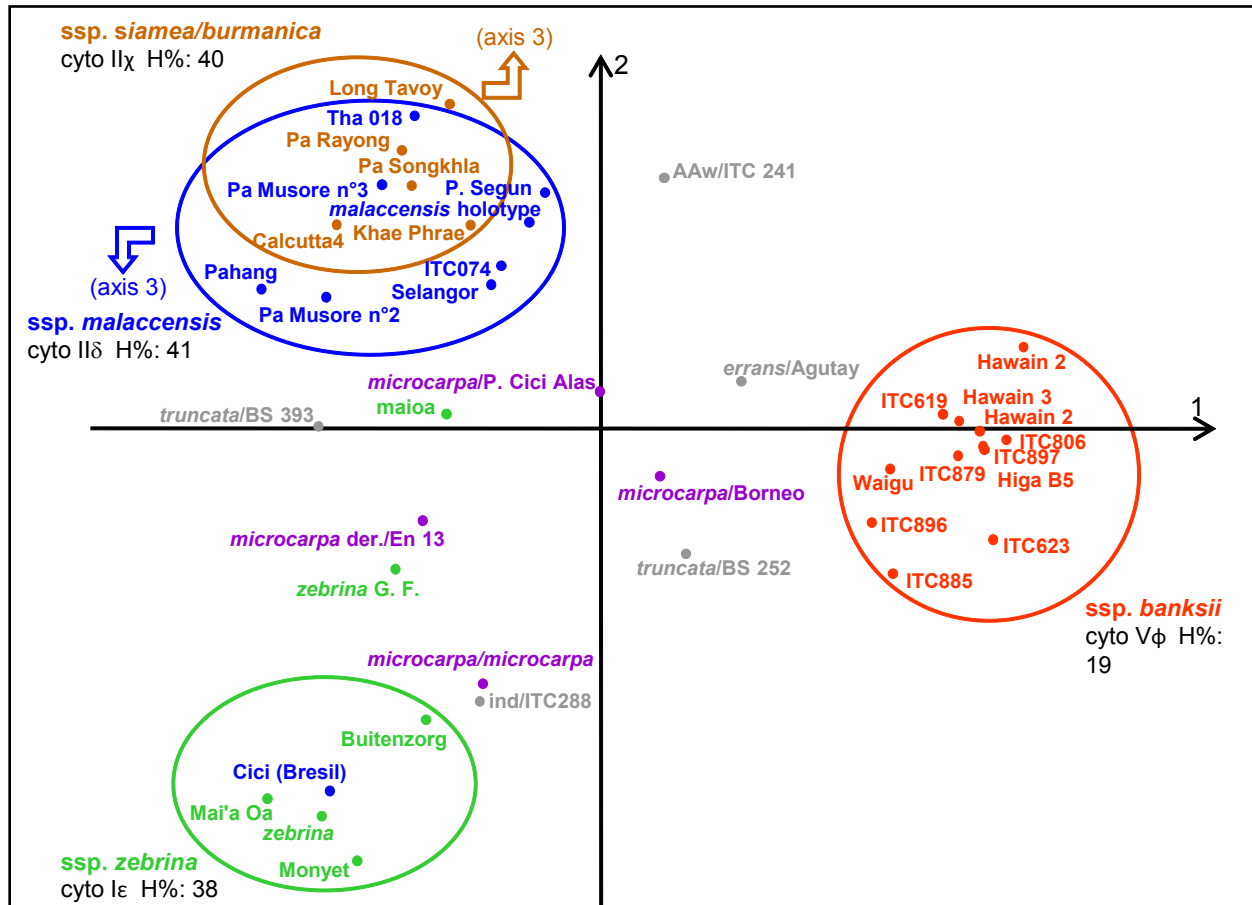


Figure 1. Factorial analysis from 22 SSR markers on a set of 41 semiferous diploid acuminata (plan 1-2, axis 1: 20.6%, axis 2: 11.2%). The cytoplasmic type (cyto) and the mean heterozygosity (H%) are given for the 4 main groups. *Ssp. siamea/burmanica* and *ssp. malaccensis* are separated on the third axis.

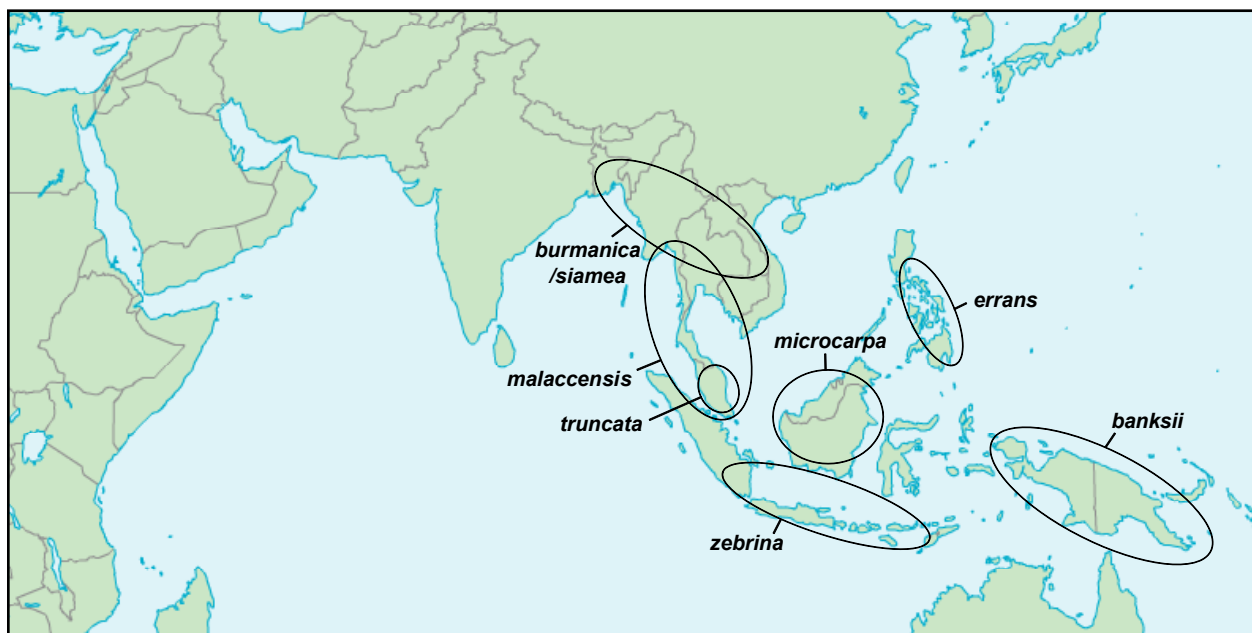


Figure 2. Geographical distribution of the subspecies of semiferous diploid *Musa acuminata*.

geographical distributions (Figure 2). These lines of evidence justify the distinction of subspecies within *M. acuminata*.

Some accessions cannot be associated with any of the aforementioned basic groups. This is the case for accessions classified as '*microcarpa*' which are more heterozygous than the representatives of the other groups. In addition they show a high molecular heterogeneity; the 'Borneo' accession shares several alleles with *banksii* while the *microcarpa* accession is closer to *zebrina*. In addition, several so-called *microcarpa* accessions share with *zebrina* the higher genome size. Thus, these *microcarpa* accessions appear to be derived from crosses between the other subspecies in the central Indonesian area around Borneo where these subspecies could be in contact. However, isozyme data (Horry 1989) and PCR-RFLP markers of chloroplast genome (Carreel unpublished results) reveal a specific pattern for *microcarpa*. Häkkinen (De Langhe pers. comm.) identified in the north of Borneo several wild populations with distinct morphological characters. So one can hypothesize that *microcarpa* (Becc.) Simmonds is a true subspecies, probably related to *zebrina*, and can adopt the concept of a *zebrina/microcarpa* complex among which only hybrid forms have been sampled.

The representativeness of the two *truncata* [*M. acuminata* ssp. *truncata* (Ridl.) Kiew.] accessions in the SSR analysis is dubious. This form, endemic to the highlands of Peninsular Malaysia, was classified as *malaccensis* or as *microcarpa* on morphological characters. However, studies based on RFLP markers concluded that *truncata* is a real subspecies (Carreel *et al.* 1994).

We have also to mention the *M. acuminata* ssp. *errans* (Blanco) R.V. Valmayor because of its particular mitochondrial type α , although only one accession has been studied. It is the only wild *M. acuminata* with such type, otherwise very common in diploid or triploid cultivars. Morphologically, it was first classified as *banksii* but subsequently was given the status of a subspecies of *M. acuminata* (Valmayor 2001). Using SSR markers, the association with *banksii* is clear even if several alleles of the *burmanica/burmanicoides/siamea* complex can be found.

Species of the section Eumusa, other than *M. acuminata* or *M. balbisiana*, and even species of other sections, share several nuclear alleles with *M. acuminata*. *M. schizocarpa* Simmonds is close to ssp. *banksii* and covers the same geographical area (hybrids between ssp. *banksii* and *M. schizocarpa* are known). *Musa laterita* Cheesman, from the Rhodoclamys section, shares alleles with the *burmanica/burmanicoides/siamea* complex, and they are geographically proximal. These interspecific crosses are also attested by direct observation of the chromosomes (D'Hont *et al.* 2000). The contribution of DNA from other species to the *M. acuminata* genome is certainly un-

derestimated and requires more attention, particularly in the area of Assam, Burma and Thailand, which appears to be the native area of *M. acuminata* (Simmonds 1962) and is likely to have been a hot spot of evolution with frequent interspecific genetic exchanges.

In summary, at least four main subspecies of *M. acuminata* are clearly identifiable, and possibly five with *microcarpa*. The native areas of each subspecies are potential centers for domestication of edible diploids (Figure 2).

Towards edible AA cultivars

A general scenario is commonly proposed concerning the evolutionary process leading from *M. acuminata* seeded types to cultivated edibles. The emergence of proto-agriculture initiated a long-term selection process with progressive evolution from wilds to cultiwilds to basic cultivars. This process would have occurred in parallel and independently in the various areas of *M. acuminata* subspecies (Figure 2). Later on, cultivated forms were spread by humans as they explored surrounding territories. Hybridizations between clones derived from different subspecies generated new cultivars feeding the selection process and leading to the diversity of edible cultivars displayed today.

The principal outstanding question in this scenario is when and where did the edible character arise in domesticated AAs, namely, an edible fruit being acceptably non-seeded and with sufficient pulp. As a general rule, the development of fruit tissues is initiated by the development of the seeds, with the exception of the rare parthenocarpic plants which produce fruits even when viable seeds are absent. The absence of seeds can be the consequence of inefficient fertilization, early abortion of the seed after fertilization, or sterile gametes. As wild bananas follow the general rule, the key point of banana domestication for fruit edibility is the establishment of spontaneous parthenocarpic associating abundant pulp development and partial or total sterility. Simmonds (1962), who discussed in depth these two notions, emphasized that pulp development and sterility are two independent characters: sterility does not induce pulp development and similarly the converse is true, even though pulp development per se could cause some sterility through hormonal interactions. These characters are linked by a parallel selection but their evolutionary process is not necessarily the same. Having stated these preliminaries, we will propose some key ideas to elucidate the complex evolution of diploid cultivars.

A first and important result is that many, if not all, currently known AA cultivars are the result of hybridizations between AA subspecies. The SSR analysis divides the set of 131 AA cultivars into several homogeneous subgroups that can be positioned according to the range exhibited by wild accessions (Figure 3). These AA cultivar subgroups never overlap with seeded subgroups and are always very

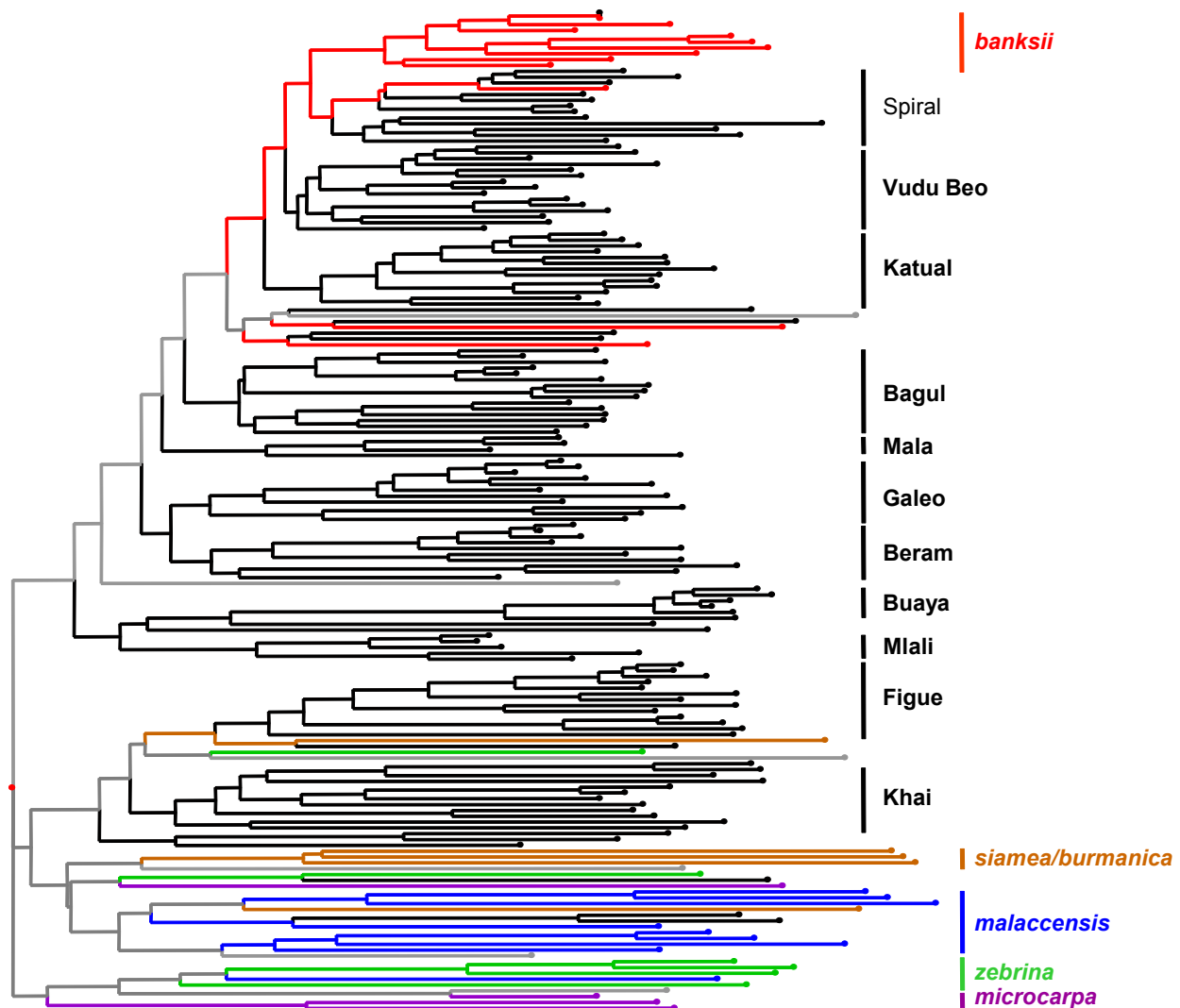


Figure 3. Neighbor-Joining tree built from 22 SSR markers on a set of 131 AA cv (in black) and 41 AA w (in color, according to figure 1), rooted on *Musa balbisiana* (not shown). The inferred subgroups of AA cv are identified by the name of one of their accessions.

heterozygous, suggesting hybrids between genetically distant forms. The allelic contributions of the main subspecies to these subgroups (Table 2) often show the contribution of at least two subspecies. The cytoplasmic types confirm the frequent crossings between subspecies. The set of AA cultivars collected in PNG is particularly illustrative. It shows a series of subgroups with an increasing heterozygosity that coincides with a decreasing contribution of *banksii*, balanced by an increasing contribution of *zebrina/microcarpa*. In parallel, the frequency of *banksii* cytoplasmic type $V\phi$ decreases to the benefit of hybrid forms $V\alpha$ ($=V\phi \times II\alpha$), or more rarely $II\phi$ ($=II\alpha \times V\phi$). So, even in New Guinea, the native area of the subspecies *banksii*, most of the collected parthenocarpic AA cultivars are hybrid forms integrating part of a *zebrina/microcarpa* exogenous genome.

AA cultivars are structurally heterozygous hybrids of *M. acuminata* subspecies with different genome structures. Cytogenetic studies have shown that structural heterozygosity is an important factor of male and female sterility, because it disrupts the formation of the gametes (Dessau 1988). Consequently, sterility, mainly male sterility, of the AA cultivars is certainly in large part a consequence of this heterozygosity. However, it does not guarantee complete sterility and most AA cultivars bear some seeds when pollinated. Conversely, fertility is not synonymous with homozygosity. For example, the wild accession *microcarpa* 'Borneo', although fertile, was found structurally heterozygous by genetic mapping (Hippolyte pers. comm).

If inter-subspecificity and structural heterozygosity appear as common features to all cultivars, edibility is probably also under genetic control. Genetic factors are able to dis-

Table 2. Allelic contributions of the 4 wild groups (*banksii*, *zebrina*, *malaccensis*, *burmanica+siamea*) to edible diploid or triploid subgroups or accessions. For each subgroup, the heterozygosity (H%) (for diploids), the cytoplasmic types (cyto) and the number of accessions into the subgroup (nb) are given. (Appearing on two pages, to be continued on page 207)

	Group	H%	cyto	nb	Scores			
					<i>banksii</i>	<i>zebrina</i>	<i>malaccensis</i>	<i>burmanica & siamea</i>
AAw	<i>banksii</i>	19	Vφ	9	0.75	0.16	0.22	0.13
	<i>zebrina</i>	38	Iε	3	0.17	0.68	0.14	0.10
	<i>malaccensis</i>	41	IIδ	5	0.23	0.16	0.50	0.14
	<i>burmanica/ siamea</i>	40	IIχ	3	0.14	0.11	0.13	0.45
AAcv	Spiral	30	Vα, Vφ	13	0.58	0.19	0.19	0.12
	Vudi beo	43	Vα, Vφ	15	0.58	0.20	0.18	0.11
	Katual	38	Vα	13	0.50	0.21	0.16	0.12
	Bagul	54	Vα, Vφ, IIα	15	0.48	0.24	0.19	0.12
	Mala	54	IIφ	4	0.44	0.27	0.17	0.09
	Galeo	70	Vα	11	0.35	0.23	0.23	0.10
	Beram	55	Vα	10	0.40	0.23	0.18	0.12
	Guyod (11-33)	62	IIα	1	0.38	0.33	0.22	0.10
	Agutay	20	IIα	1	0.37	0.15	0.16	0.14
	MIali	55	IIα	5	0.35	0.20	0.18	0.14
	Buaya	81	IIβ	6	0.20	0.17	0.16	0.12
	Figue	59	IIα	15	0.16	0.25	0.20	0.13
	Khai	65	IIα, IIδ	12	0.17	0.17	0.24	0.14
AAA	Cavendish		IIα	26	0.37	0.16	0.21	0.15
	Gros Michel		IIα	5	0.32	0.15	0.20	0.09
	Mutika/Lujugira		Vε	25	0.30	0.30	0.14	0.08
	Ambon		IIα	1	0.31	0.22	0.23	0.15
	Rio		IIα	1	0.29	0.20	0.22	0.10
	Orotava		IIα	5	0.26	0.22	0.19	0.10
	Red		IIα	7	0.23	0.17	0.17	0.12
	Ibota		IIα	8	0.17	0.19	0.25	0.13
AAB	Iholena		Vφ	5	0.42	0.16	0.14	0.08
	Popoulou/ Mai'a Maoli		Vφ	7	0.39	0.15	0.12	0.07
	Plantain		Vφ	79	0.31	0.16	0.12	0.07
	Laknao		Vφ	5	0.29	0.22	0.14	0.09
	Nendra Padaththi		IIα	2	0.29	0.21	0.16	0.12
	Nadan		IIα	2	0.28	0.21	0.15	0.11
	Pome/Prata		IIα	6	0.29	0.17	0.16	0.10
	Pisang Rajah		VIIIα	1	0.23	0.11	0.15	0.11
	Mysore		IIα	1	0.14	0.21	0.19	0.12
	Silk		IIδ	5	0.11	0.14	0.18	0.10

	Group	H%	cyto	nb	Scores			
					<i>banksii</i>	<i>zebrina</i>	<i>malaccensis</i>	<i>burmanica & siamea</i>
ABB	Ney Mannan		II τ	4	0.41	0.16	0.17	0.10
	Saba		II τ	1	0.40	0.19	0.14	0.10
	Bluggoe		II τ	6	0.38	0.16	0.14	0.11
	Monthan		-	2	0.36	0.15	0.14	0.10
	Peyan		VIII τ	2	0.22	0.12	0.12	0.08
	Pisang Awak		VIII τ	5	0.16	0.08	0.16	0.08
	Pelipita		V τ	1	0.14	0.10	0.06	0.06

rupt the mechanisms of fertilization and to decrease fertility; it is then possible that human selection was practised among cultiwilds for this character. Pulp development is certainly under genetic control and may have also undergone a positive selection during the cultiwild phase. A proof of this selection might be a wild Samoan *banksii* that shows a partial pulp development although it is perfectly fertile (Simmonds 1962). The same case has been observed for the wild *malaccensis* 'Pisang Serun' (Rosales *et al.* 1999). The AA diploids that people exchanged and that were implied in these inter-subspecific crosses were certainly not wild forms; they were already improved forms with some pulp development and reduced fertility. They were certainly not sterile as they hybridized later with other subspecies. Given that these diploids are no longer cultiwilds and are not yet basic cultivars, we have introduced the term protocultivars for these intermediate forms.

Examples of such protocultivars are rare, probably because there was no reason for maintaining them. No parthenocarpic AA close to *zebrina* or *siamealburmanica* or *microcarpa* was found in the analyzed sets. Accessions from PNG belonging to the 'Spiral' or 'Vudu Beo' subgroups could be potential candidates (Figure 3). Although parthenocarpic, they are genetically close to wild fertile *banksii* for their nuclear genome, their cytoplasmic type V ϕ and their low heterozygosity. We can regard them as an advanced step to edibility deriving directly from wild *banksii*. However it cannot be excluded that they incorporate a low level of *zebrina* component and some of their accessions have a cytoplasm V α . So these accessions could result from several back-crosses with the *banksii* parent, restoring the *banksii* type and possibly the cytoplasm V ϕ (V ϕ =V α x V ϕ).

Although belonging to the Australimusa section, the Fe'i group, a group with characteristic erect bunches and orange pulp that is widely cultivated in the Pacific Islands, could be an example of this endogenous process towards parthenocarpy. These edible parthenocarpic forms derive directly from seeded forms related to *Musa maclayi* F. Muell. ex Mikl.-Maclay. However, little is known about their genome and it has been sometimes suggested that cross-

es with other wild species could play a part in their evolution (Constantine 1999).

Geographically, hybridizations between protocultivars derived from different subspecies do not imply large-scale migrations, but only interactions due to geographical proximity. For hybrids between *banksii* and the *zebrina/microcarpa* complex, it is supposed that Papuans visited the islands west of NG, including Molucca Islands, Sulawesi and, to the south, the Lesser Sunda Islands such as Timor. They could have carried with them *banksii* forms and brought back *zebrina* or *microcarpa* forms. It is not possible to date these exchanges but these movements between islands could be as ancient as what specialists call 'the Trans New Guinea migration', which dates to c. 6000 years ago and dispersal occurred from the eastern highlands of NG to the west coast and onward to neighboring islands (Denham & Donohue 2009). This date is consistent with possible diploid banana cultivation at Kuk Swamp in the highlands of Papua New Guinea from c. 7000 years ago (Denham *et al.* 2003). In the same way, we can assume contacts between NG and the Philippines that are represented by intersubspecific hybridizations between *banksii* and northern forms (*errans?*). Contacts between *zebrina/microcarpa* and *malaccensis* generated the subgroups 'Khai' or 'Figue', or maybe with *siamealburmanica*, the subgroup 'Buaya'. The case of subgroup 'Galeo' from NG is special since it simultaneously includes genome contributions from *banksii*, *zebrina* and *malaccensis*. Certain indications point to 'Galeo' being a recent creation following the introduction of *malaccensis* to NG in historical times.

At this point, it is interesting to focus on the subgroup 'Mlali' (Jenny pers. comm.). It includes a set of accessions (Akondro Mainty, Paka, Nzumoheli) that are only found in East Africa and neighboring islands: Madagascar, Comoros, Zanzibar and Pemba. It is clearly an ancient group; its genome is based on *banksii* and *zebrina/microcarpa* and originates from somewhere between Java, Borneo and NG where it is no longer recorded. However, it must have been popular in the past since it is the unique case of ancient diploid migration far from the native area



Figure 4. Proposed geographical distribution of the main AACv subgroups (plus the particular accession Guyod) as identified on Figure 3. The arrows indicate long term migration of the Mlali subgroup.

and, as we will see later, it contributed to some major AAA and AAB triploids.

The major contribution of the *banksii* subspecies to AA cultivars, as well as to AAA or AAB triploids, is often raised. From this observation, it has been proposed that the development of parthenocarp is restricted to this subspecies, making New Guinea the cradle of banana domestication. Indeed, the only conceivable example of parthenocarp in cultiwilds comes from *banksii*. Moreover, it was shown from RFLP markers that almost all the AA cultivars have at least some alleles specific to *banksii* (Carreel *et al.* 1994). Several *banksii* alleles are also found in important triploid subgroups like AAA Cavendish or AAB Plantains. This single emergence would imply the diffusion of *banksii* accessions from NG to the whole of Southeast Asia for hybridizations with local seeded forms giving birth to the present diversity of AA cultivars. For SSR markers (Table 2), the *banksii* contribution is verified for AA cultivars from PNG for the 'Mlali' subgroup and for the particular accession 'Guyod' from Philippines. However this *banksii* contribution is not found in the northern 'Khái' and 'Figue' subgroups. Moreover, as shown later, some triploid subgroups like AAA *Ibota* or AAB Silk do not exhibit a *banksii* contribution.

So it is clear that *banksii* contributed to most domesticated bananas and the data indicate the *banksii* area (New Guinea) as the earliest area of domestication and certainly, the most active one. However this contribution is probably not exclusive and other selection processes, in other areas and for other subspecies, are necessary to explain the present diversity of AA cultivars and triploids.

In summary, even if a progressive selection for edibility during the cultiwild phase is likely, most of the current edible forms do not result directly from this initial and local selection. Rather, they derive from hybridizations between AA within different subspecies in the areas of contact between subspecies, hybridizations enabled by human movements and exchanges (Figure 4).

From Diploids to Triploids

Triploid bananas are known to result from the fertilization between a non reduced 2N gamete and an N gamete. The 2N gametes result from accidents during meiosis which arise when the two parent genomes are too different, as is the case for interspecific *M. acuminata-balbisiana* AB or even for intersubspecific *M. acuminata* AA. So intersubspecificity, which has been demonstrated at the origin of AA cultivars, also occurs at the origin of triploids through the 2N gametes. Thereby, triploids follow diploids in a common and coherent history.

Triploids are today cultivated throughout the world. The success of triploids, such as the AAA Cavendish, is due to their high vigor and fruit growth rate, as well as gametic sterility. For AAB or ABB, the *balbisiana* genome adds characters like tolerances to abiotic stresses. These favorable characteristics led to an extension far beyond the primary center of diversity, as convincingly demonstrated by AAB plantains. However triploids did not supersede the traditional diploids within the primary center. This suggests another characteristic of triploids that is rarely emphasized; they show an impressive plasticity and can grow in a large range of environmental conditions. Thus,

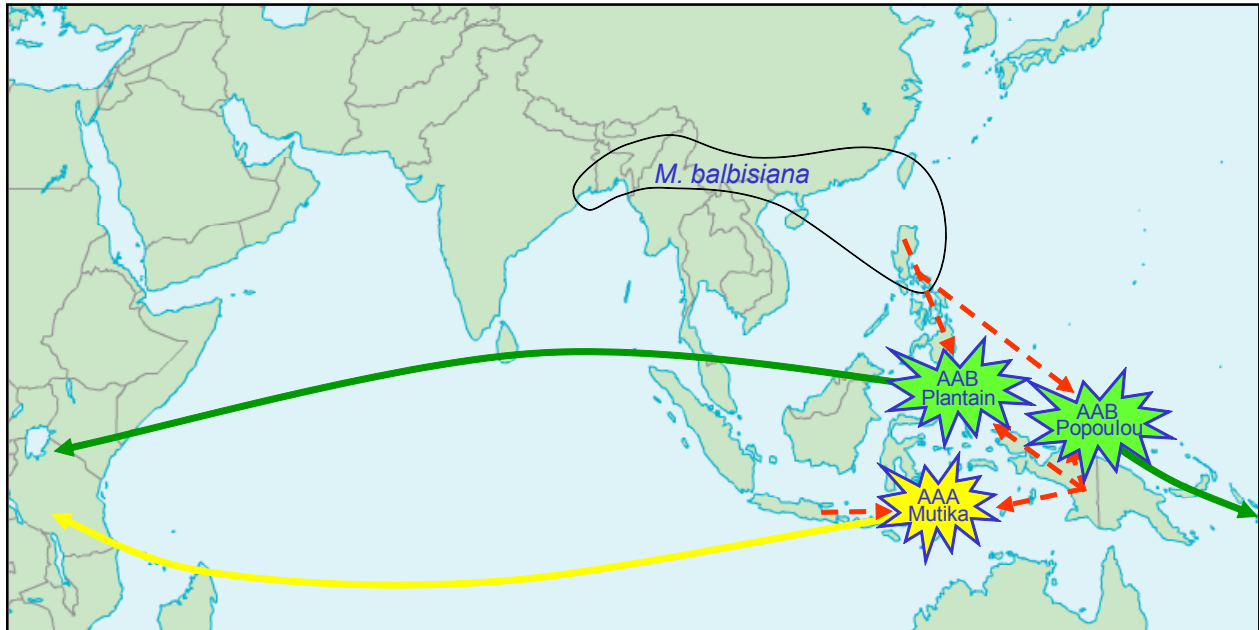


Figure 5. First wave of triploids, creation and geographical diffusion. Plain arrows indicate long term migrations of triploids subgroups: Plantain and **Mutika** triploids to Africa or **Popoulou** to Oceania. Dotted arrows illustrate the contacts between diploids at the origin of these triploids.

triploids are usually better adapted than diploids to conditions deviating from the native ones, which is probably due to the potential of three different alleles for the same gene.

Diversity analysis of the triploid set compared to diploids shows a structure in groups that roughly agrees with groups defined from morphological characteristics (Jenny *et al.* 2003). The first axis of a factorial analysis on the 22 SSR markers reveals the respective numbers of *acuminata* and *balbisiana* genome from AAA to ABB. The second axis, independent of the first one, accounts for *acuminata* subspecies contributions, from *banksii* to *burmanica/burmanicoides/siamea*, through *zebrina* and *malaccensis*. These contributions are confirmed by allelic frequencies and cytoplasmic types (Table 2). These results will not be detailed here and we will focus mainly on mechanisms of triploidization from diploids.

For the main triploid groups, we propose two main waves of triploidization that we call “ancient” and “recent”, without being in a position to give precise chronological ranges for these terms. Two arguments contribute to distinguish these waves. The first one is biological. When looking for direct parents of a triploid with the kinship score method, identifying two putative diploid parents with a high score is interpreted as a sign of recent formation. The parents, or close relatives, are still recognizable by their alleles and have probably not undergone a long period of evolution. Conversely, absence of a clear relationship with present diploid germplasm means that the parents have evolved to an unrecognizable degree and that the formation is probably ancient. Of course, actual diploid parents may be absent due to incomplete sampling but it is very im-

probable that all closely related forms are simultaneously absent, especially for studies based on large samples. The second argument pertains to human prehistory. The ancient wave corresponds to triploids that have an area of diffusion distant from the primary diversity center and that were diffused by human migrations in remote times. The second wave corresponds to triploids that remained in the center, or nearby, and thus probably were generated more recently, even if some of them have been widely dispersed in modern times.

Before expanding on these two waves, we should mention that triploidization is probably a very common event. The collecting missions in PNG provided a lot of AAA triploids. The SSR pattern of some of them can almost perfectly be built from PNG AA genomes. The 2N gamete comes generally from a hybrid AA with a $V\alpha$ cytoplasmic type and the N gamete is a $V\phi$ *banksii* (see for example AAA **Mise’Ehina** from AA **Pitu** and AA **Yendisi**, Table 3). According to our previous argument, these AAA are probably recent and we can assume that triploidization is a continuous process in PNG, whereby triploids have not been heavily selected by people and have not replaced diploids. The main conclusion is that triploidization is not a unique or improbable event and probably occurred in several places and genetic backgrounds.

The ancient wave (Figure 5)

Mutika/Lujugira from East-Africa are AAA including *banksii* and *zebrina* genome, although close diploid parents cannot be identified. The cytoplasm type is V like *banksii* and ϵ like *zebrina*. An origin somewhere in Lesser Sunda Islands, between NG and Java is likely. No accession

close to Mutika has been collected in this area, but some AAA of PNG, like **Merik**, are linked to this **Mutika** subgroup. The ancestral forms of this subgroup came with people across the Indian Ocean to the African coast.

Plantains of Africa and the **Iholena** and **Popoulou/Maia Maoli** complex (Lebot *et al.* 1994), also called Pacific plantains, are AABs where the *M. acuminata* component has a predominant *banksii* origin. *Musa balbisiana* has a northern distribution; its particularly wide distribution implies human dispersal of *M. balbisiana* cultiwilds, which is so far poorly known. We suggest that contacts with *banksii* imply a migration southward from southern China through the Philippines. For African plantains, direct diploid parents cannot be identified but an approximate reconstruction is obtained with an A genome of **Meleng**, a *banksii* AA cultivar of PNG, and **Guyod**, a diploid AA accession from the Philippines. No close AAB was found in NG but plantains have similarities with AAB **Laknao** from the Philippines. We propose an origin for the plantains in an intermediary area between northern NG and the Philippines, maybe on Halmahera Island, followed by a long term migration to Africa and to India.

Popoulou/Maia Maoli and **Iholena** subgroups are close to plantains but they do not share the resemblances with **Laknao** or **Guyod**. Accessions close to **Iholena** were found in NG. We propose a center of origin, different to that of plantains, located in the eastern part of NG or the Bismarck Archipelago, before dissemination to Pacific Islands. The arrival of these AAB on the American continent remains an open question.

Thus, at least two human dispersals are responsible for banana cultivation in Africa. The first one concerns the East African coast, up to the Great Lakes, and neighboring islands. This dispersal accounts for populations of AAA **Mutika**, some AA cultivars - particularly the '**Mlali**' subgroup - and potentially the enigmatic wild population of Pemba Island (De Langhe 2009), which shares the translocation of *zebrina*, and the small isolate population identified on Madagascar (T. Lescot pers. com. 2007). All these forms are based on *banksii* and *zebrinal/microcarpa* genomes and originate from the contact area between NG and Java; they probably arrived in Africa as a result of human movements around the Indian Ocean (Donohue & Denham 2009, Fuller & Madella 2009). The second migration concerns plantains from the Philippines/northern NG to Congo Basin (Blench 2009). Recent phytolith evidence (Mbida *et al.* 2001) confirms the hypothesis of De Langhe *et al.* (1996) for an introduction of plantains to Africa at least 2500 years ago. It is generally considered that plantains were the first bananas introduced to Africa, earlier than **Mutika** (De Langhe *et al.* 1996). However we believe that the cross between *banksii* and *zebrina* at the origin of the **Mutika** group is probably the first event of hybridization and predates the introduction of *M. balbisiana*, which occurs at the origin of the plantains and may be associ-

ated with the migrations of Austronesian language speakers. This chronology is suggestive of an earlier introduction of **Mutika**. The contentious archaeological results at Munsu, north of Lake Victoria and dated to 5000 years B.P. (Lejju *et al.* 2006), could also invert the previously accepted chronology (see Neumann & Hildebrand 2009). The current dating of the introduction of bananas to Africa, namely before 2500 years ago, is consistent with that of other species. Exchanges around the Indian Ocean are frequent at this time: rice, taro and sugarcane dispersed from Asia to Africa; and, millet and sorghum dispersed from Africa to Asia (Blench 1996, Fuller 2003).

In summary, triploids of this first wave resulted from ancient crosses that occurred in the primary diversity centers and hence very far from their present area of distribution. The high diversity observed in these subgroups is a consequence of a necessarily long phenotypic differentiation by vegetative propagation from a small number of types initially introduced. These assumptions are supported by linguistic and archaeological data (De Langhe & de Maret 1999).

The recent wave (Figure 6).

The important AAA Cavendish and Gros-Michel subgroups appeared in this second wave. The SSR pattern of Cavendish is almost perfectly restored (Table 3) taking as 2N gamete the accession **Akondro Mainty**, or more generally an accession of the subgroup '**Mlali**', and as N gamete the **Pisang Pipit**, an accession of the '**Khai**' subgroup, or **Pisang Madu**. This result is statistically significant and cannot be obtained by chance, nor can it be obtained with any other combination of diploids. Moreover it confirms a previous result based on RFLP markers (Raboin *et al.* 2005). Therefore, the contribution of the '**Mlali**' subgroup as 2N gamete donor seems well established, although the 2N parent is not one of the current '**Mlali**' accessions but one of their close ancestors. We previously saw that this subgroup is found today only on the East African coast but originates probably from an area between NG, Borneo and Java. The N gamete was delivered by an AA diploid with a *malaccensis* affiliation. **Pisang Madu** and **Pisang Pipit** are two good candidates, in fact they share half of their alleles inherited from a common parent, which is probably the other direct ancestor of Cavendish. Since these accessions originated on the Malay Peninsula or from southern Thailand, Cavendish probably generated around that region as well. As a consequence, we can assume that the **Mlali** ancestor moved from a southern area - New Guinea, Borneo, Java - to hybridize with a local AA cultivar in the northern region (Figure 4). Until recently, Cavendish forms stayed in their native area but underwent moderate expansion, especially towards South China where they are abundant.

The contributions of '**Mlali**', of *zebrinal/microcarpa* and *banksii* ascendance, and '**Khai**', of *malaccensis* ascen-



Figure 6. Second wave of triploids. ‘AAA Cav.’ includes Cavendish and Gros-Michel cultivars. Plain arrows indicate long term migration of AA **Mlali**. Dotted arrows illustrate the contacts between diploids at the origin of the triploids.

dance, make the Cavendish a three way hybrid, which might explain in part its favorable characteristics. The primitive state of the polyphenol metabolites found in the Cavendish is not a sign of ancient origin of these triploids; it is rather inherited from *zebrina* through ‘**Mlali**’.

Gros-Michel AAA are close to Cavendish and share the same **Mlali** ancestor. The N gamete however is **Sa** or **Khai-Nai-On**, two very similar accessions of subgroup ‘**Khai**’ (Table 3). Gros-Michel AAA certainly originate from the same area as Cavendish. Besides these subgroups, other crossings with a ‘**Mlali**’ 2N gamete donor and a *malaccensis* N gamete donor took place in the same area, the accession **Hom Thong Moko** being an example from Thailand. The AAA **Orotava** subgroup may have a similar origin with another 2N gamete donor, but still with *banksii* and *zebrina/microcarpa* contributions. Finally, Red and **Ibota** AAA subgroups are rather of restricted *malaccensis* ascendance.

The expansion of ‘**Mlali**’ must have been even more important since it reached India. Indeed, it is surprising to find ‘**Mlali**’ probably also at the origin of the sweet acid AAB family of India. The accessions belonging to the **Pome/Prata** subgroup, and its relatives **Nendra Padathi** and **Nadan**, result with a very high probability from a **Mlali** 2N gamete and an N gamete coming from an Indian *M. balbisiana* like **Lal Velchi** (Table 3). *M. balbisiana* is native in the north of India and that area appears as a second center of contact between *M. acuminata* and *M. balbisiana*, quite distinct from the eastern NG/Philippines region. Other Indian AAB genotypes, especially those of the Silk subgroup, are based on the *malaccensis* type without obvious contribution of *banksii* or *zebrina*.

The case of ABB triploids is less clear. We place them in the second wave mainly because of their limited geographical distribution (before modern expansion). However, ABB related to *banksii* (**Bluggoe**, **Ney Mannan**, **Monthan**) probably have a common eastern origin with plantain and **Popoulou**. Indian ABB triploids, ‘**Awak**’ and ‘**Peyan**’ subgroups, display a *M. acuminata* component of *malaccensis* type, and they probably arose with AAB **Pome** or AB types during the secondary phase of diversification in India.

In summary, triploidization is a common event which occurs in different areas. The oldest generation wave occurred in the south or east of the *acuminata* region and includes AAA **Mutika**, the African plantains and the Pacific plantains. They integrate the genetic contribution of the primitive diploids *zebrina/microcarpa* and *banksii*. They were largely spread by important human migrations and are popular today in regions quite remote from their native location. The second wave originated in more northern areas. Contacts between ancestors of ‘**Mlali**’ coming from the south and local *malaccensis* produced the Cavendish family. Farther to the west in India, these same ‘**Mlali**’ diploids produced the sweet Indian triploids through combination with *M. balbisiana*. The triploids arising during this second wave remained close to their native area compared to the wide dispersals of the first wave, before some of them were recently spread throughout the world.

Towards a Multidisciplinary Approach

The history of *Musa* domestication, as proposed here, still suffers from long shadows and large uncertainties. Progress in biological tools will greatly refine our knowledge of

Table 3. Reconstruction of triploid SSR allelic pattern from two diploids: one as 2N gamete and the other as N gamete donors. Keys: blue: matching alleles; '-': missing data; and, '?': partially heterozygous triploids, the third allele may be any of the two read alleles.

	C03	C07	C08	C13	C39	C40	C45	C152	C195	C214	1_32	3_90	M01	M24	M27	M150	M164	M196	M231	M260	M264	M307
3X Mise'ehina AAA	124	170	273	286	335	180	289	165	298	123	251	166	270	281	241	257	403	180	250	210	266	164
	122	158	261	286	329	178	284	158	258	123	247	162	266	247	235	257	403	168	242	210	260	162
	120	?	?	286	?	?	?	?	?	123	235	?	?	?	?	257	403	?	?	210	251	?
	122	158	273	-	335	180	289	165	298	123	247	166	270	281	235	257	-	180	250	210	-	-
2X Pitu AA	120	158	261	-	329	178	284	165	258	123	235	162	266	247	235	257	-	168	242	210	-	-
X Yendisi AA	124	170	261	286	335	180	284	158	258	123	251	162	266	247	245	257	-	180	242	210	264	164
	122	158	261	286	331	178	284	158	296	123	235	168	250	253	245	257	-	180	252	210	264	164
3X Cavendish AAA	C03	C07	C08	C13	C39	C40	C45	C152	C195	C214	1_32	3_90	M01	M24	M27	M150	M164	M196	M231	M260	M264	M307
	124	170	267	286	335	180	289	164	298	123	245	168	264	253	245	261	407	180	276	212	258	164
	122	158	261	286	331	178	284	158	298	119	235	162	258	247	243	257	401	168	250	212	250	162
	?	?	?	286	?	176	?	?	?	298	?	?	150	254	237	?	?	?	242	212	?	?
2X Akondro Mainty AA	124	170	267	286	335	180	289	164	298	123	245	168	258	253	241	261	401	180	278	210	258	164
	122	158	261	286	335	178	284	164	298	119	245	150	254	247	241	257	401	180	242	206	258	164
X P. Pipit AA	124	170	261	286	349	176	289	164	258	119	235	-	264	-	243	257	407	168	242	212	250	162
	120	172	276	286	341	178	284	167	258	128	254	-	250	-	235	257	407	168	270	208	264	160
3X Gros Michel AAA	C03	C07	C08	C13	C39	C40	C45	C152	C195	C214	1_32	3_90	M01	M24	M27	M150	M164	M196	M231	M260	M264	M307
	124	170	267	286	335	180	289	164	298	123	245	168	270	253	243	261	401	180	276	212	266	164
	122	158	261	286	329	178	284	164	258	119	235	150	258	247	243	257	401	176	252	212	258	162
	?	?	?	286	?	176	?	?	?	?	?	?	254	?	243	?	?	?	242	212	?	?
2X Akondro Mainty AA	124	170	267	286	335	180	289	164	298	123	245	168	258	253	241	261	401	180	278	210	258	164
	122	158	261	286	335	178	284	164	298	119	245	150	254	247	241	257	401	180	242	206	258	164
X Khai Nai On AA	124	170	261	286	329	176	289	165	258	119	235	168	270	-	241	257	403	176	252	210	266	162
	120	170	273	286	331	178	284	161	258	115	254	162	250	-	235	257	403	168	250	210	252	162
3X Pome AAB	C03	C07	C08	C13	C39	C40	C45	C152	C195	C214	1_32	3_90	M01	M24	M27	M150	M164	M196	M231	M260	M264	M307
	124	170	267	290	357	180	289	182	298	125	245	168	290	253	245	261	401	180	276	225	258	168
	122	166	261	286	335	178	284	164	298	123	239	152	258	247	243	257	298	177	252	210	242	164
	121	158	?	?	?	?	274	?	298	119	?	?	254	241	?	?	?	?	242	?	?	?
2X Akondro Mainty AA	124	170	267	286	335	180	289	164	298	123	245	168	258	253	245	261	401	180	276	210	258	164
	122	158	261	286	335	178	284	164	298	119	245	150	252	247	243	257	401	180	242	210	258	164
X Lai Veichi BB	121	166	261	290	357	180	274	182	296	125	239	152	290	241	235	261	298	177	249	212	242	168
	121	166	255	282	357	180	274	173	296	123	239	152	310	243	235	257	313	182	249	212	242	168

the banana genomes and of their expression. But whatever the power of these tools, a major limitation is the germplasm available for analysis and its representativeness of the diversity. Although our SSR sample includes nearly 500 accessions, it is obvious that some important parts of the diversity are not covered. Particularly, the very important phase between cultiwild and basic cultivars requires an extension of the available germplasm by new collecting missions to the probable contact areas: Lesser Sunda, Sulawesi, Moluccas and southern Philippines. Another challenge that calls for new collecting missions is the mysterious mitochondrial type α in numerous edible diploids and triploids. Its distribution seems to indicate an origin in the central Indonesian area, maybe somewhere around the Celebes Sea. The unsolved question of the *microcarpa* subspecies case, which is perhaps related to the α mitotype question, has also to be clarified. For wild forms, the contribution of other species, maybe even of other sections, to the *acuminata* genome should be more accurately investigated.

However, many outstanding questions concerning *Musa* domestication will not be answered by biological research alone. Domestication is essentially a human process and biology can only address its ultimate results, the current plants. We miss the extinct ancestral forms, even if cloning was able to maintain some ancient genotypes. The obscure parts of the history of banana domestication can, however, be enlightened through approaches that address the role of human history.

Archaeology, linguistics and ethnobotany can provide arguments for the dating and localization of the main supposed stages of the domestication process, the role of human migrations and dispersals, and so on. Interactions between the human sciences and biology are clearly the way to progress in the understanding of banana domestication. It is also evident that this investigation cannot be strictly limited to *Musa*. *Ensete* could be a valuable model for understanding the initial stages of banana domestication.

In a larger context, bananas are associated with the earliest agricultural systems in Melanesia and South Asia, along with several fruit trees, sugarcane and tubers like yam or taro. All are vegetatively propagated and were cultivated, selected, spread through the region and/or beyond by the same human populations. Consequently, a better understanding of the domestication of these plants could derive from research at the level of the agricultural system as a whole rather than from the analysis of each crop separately.

Acknowledgment

We would like to thank Edmond De Langhe for the fruitful discussions which accompanied this work and Anne

Vezina and Tim Denham for their support in the edition of this paper.

Literature Cited

Baurens, F.-C., F. Bonnot, D. Bienvenu, S. Causse & T. Legavre. 2003. Using SD-AFLP and MSAP to assess CCGG methylation in the banana genome. *Plant Molecular Biology Reporter* 21:339-348

Blench, R. 1996. The ethnographic evidence for long-distance contacts between Oceania and East Africa. Pp. 417-438 in *The Indian Ocean in Antiquity*. Edited by J. Reade. Kegan Paul International/ British Museum Press, London.

Blench, R. 2009. Bananas and plantains in Africa: Re-interpreting the linguistic evidence. *Ethnobotany Research and Applications* 7:363-380.

Brookes, A.J. 1999. The essence of SNPs. *Gene* 234:177-186.

Botstein, D., R.L. White, M. Skolnick & R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* 32:314-331.

Carreel, F., S. Fauré, D. Gonzalez De Leon, P. Lagoda, X. Perrier, F. Bakry, H. Tézenas du Montcel, C. Lanaud & J.-P. Horry. 1994. Evaluation de la diversité génétique chez les bananiers diploïdes (*Musa* sp.). *Genetics Selection Evolution* 26:125-136.

Carreel, F., D. Gonzalez De Leon, P. Lagoda, C. Lanaud, C. Jenny, J.-P. Horry & H. Tézenas du Montcel. 2002. Ascertaining maternal and paternal lineage within *Musa* by chloroplast and mitochondrial DNA RFLP analyses. *Genome* 45:679-692.

Cheesman, E.E. 1932. Genetical and cytological studies of *Musa*. I. Certain hybrids of the Gros Michel banana. *Journal of Genetics* 26:291-312.

Cheesman E.E. & N.H. Larter 1935. Genetical and cytological studies of *Musa*. III Chromosome numbers in the Musaceae. *Journal of Genetics* 30:31-52.

Condit, R. & S.P. Hubbell. 1991. Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. *Genome* 34:66-71.

Constantine, D. 1999. *Musa*. An annotated list of the species of *Musa*. www.users.globalnet.co.uk/~drc/musa_home.htm#The_genus

Creste, S., A. Tulmann Neto, R. Vencovsky, S.O. Silva & A. Figueira. 2004. Genetic diversity of *Musa* diploid and

- triploid accessions from the Brazilian banana breeding program estimated by microsatellite markers. *Genetic Resources and Crop Evolution* 51:723-733.
- Creste, S., T.R. Benatti, M.R. Orsi, A.M. Risterucci & A. Figueira. 2006. Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. *Molecular Ecology Notes* 6:303-306.
- De Langhe, E. 2009. Relevance of banana seeds in archaeology. *Ethnobotany Research and Applications* 7:271-281.
- De Langhe, E. & P. de Maret. 1999. Tracking the banana: its significance in Early Agriculture. Pp. 377-396 in *The Prehistory of Food*. Edited by C. Godsen, & J.G. Hather. Routledge, London.
- De Langhe, E., R. Swennen & D. Vuylsteke. 1996. Plantain in early Bantu world. Azania. *The Journal of the British Institute in Eastern Africa* 29-30:147-160.
- De Langhe, E., L. Vrydaghs, P. de Maret, X. Perrier & T. Denham. 2009. Why bananas matter: An introduction to the history of banana domestication. *Ethnobotany Research and Applications* 7:165-177.
- Denham, T. & M. Donohue. 2009. Pre-Austronesian dispersal of banana cultivars west from New Guinea: Linguistic relics from Eastern Indonesia. *Archaeology in Oceania* 44:18-28.
- Denham, T., S.G. Haberle, C. Lentfer, R. Fullagar, J. Field, M. Therin, N. Porch & B. Winstborough. 2003. Origins of agriculture at Kuk Swamp in the Highlands of New Guinea. *Science* 301:189-193.
- Dessauw, D. 1988. Etude des facteurs de la stérilité du bananier (*Musa* spp.) et des relations cytotaxinomiques entre *M. acuminata* Colla et *M. balbisiana* Colla. *Fruits* 27:539-558.
- D'Hont, A., A. Paget-Goy, J. Escoute & F. Carreel. 2000. The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genomic DNA *in situ* hybridization. *Theoretical and Applied Genetics* 100:177-183.
- Dolezel, J., M. Dolezelova & F.J. Novak. 1994. Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biologia Plantarum* 36:351-357.
- Donohue, M. & T.P. Denham. 2009. Banana (*Musa* spp.) domestication in the Asia-Pacific region: Linguistic and archaeobotanical perspectives. *Ethnobotany Research and Applications* 7:293-332.
- Fauré, S., J.L. Noyer, F. Carreel, J.-P. Horry, F. Bakry & C. Lanaud. 1994. Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). *Current Genetics* 25:265-269.
- Fuller, D.Q. 2003. African crops in prehistoric South Asia: a critical review. Pp. 239-272 in *Food, Fuel and Fields: Progress in African archaeobotany*. Edited by K. Neumann, A. Butler & S. Kahlheber. Heinrich-Barth Institut, Köln.
- Fuller, D. & M. Madella. 2009. Banana cultivation in South Asia and East Asia: A review of the evidence from archaeology and linguistics. *Ethnobotany Research and Applications* 7:333-351.
- Gawel, N.J., R. Jarret & A.P. Wittemore. 1992. Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. *Theoretical and Applied Genetics* 84:286-290.
- Ge, X.J., M.H. Liu, W.K. Wang, B.A. Schaal & T.Y. Chiang. 2005. Population structure of wild bananas, *Musa balbisiana*, in China determined by SSR fingerprinting and cp-DNA PCR-RFLP. *Molecular Ecology* 14:933-944.
- Grabin, A., J.-L. Noyer, F. Carreel, D. Dambier, F.-C. Baurens, C. Lanaud & P. J. L. Lagoda. 1998. Diploid *Musa acuminata* genetic diversity assayed with sequence-tagged microsatellite sites. *Electrophoresis* 19:1374-1380.
- Horry, J.-P. 1989. Chimiotaxonomie et organisation génétique dans le genre *Musa* (II). *Fruits* 44:509-520.
- Horry, J.-P. & M. Jay. 1988. Distribution of anthocyanins in wild and cultivated banana varieties. *Phytochemistry* 27:2667-2672.
- Jaccoud, D., K. Peng, D. Feinstein & A. Kilian. 2001. Diversity arrays: A solid state technology for sequence information independent genotyping. *Nucleic Acids Research* 29:e25.
- Jarret, R., R.E. Litz. 1986. Isozymes as genetic markers in bananas and plantains. *Euphytica* 35:539-549.
- Jarret, R., N. Gawel, A.P. Wittemore & S. Sharrock. 1994. RFLP-based phylogeny of *Musa* species in Papua New Guinea. *Theoretical and Applied Genetics* 84:579-584.
- Jenny, C., F. Carreel, K. Tomekpé, X. Perrier, C. Dubois, J.-P. Horry, H. Tézenas du Montcel. 2003. Banana. Pp. 99-124 in *Genetic Diversity of Cultivated Tropical Plants*. Edited by P. Hamon, M. Seguin, X. Perrier, J.C. Glaszmann. Enfield, Science Publishers, Montpellier.

- Jenny, C. F. Carreel. n.d. *Final Completion Report. BIP 94.41 Project: Field and laboratory evaluation of diploid bananas for their use in breeding schemes.*
- Kurz, S. 1865. Note on the plantains of the Indian Archipelago. *Journal of Agricultural and Horticultural Society of India*. 14:295-301.
- Lagoda, P.J.L., J. L. Noyer, D. Dambier, F.-C. Baurens, A. Grapin & C. Lanaud. 1998. Primer note - Sequence tagged microsatellite site (STMS) markers in the Musaceae. *Molecular Ecology* 7:659-663.
- Lebot, V., K. M. Aradhya, R. Manshardt & B. Meilleur. 1993. Genetic relationships among cultivated bananas and plantains from Asia and the Pacific. *Euphytica* 67:163-175.
- Lebot, V., B.A. Meilleur & R.M. Manshardt. 1994. Genetic diversity in Eastern Polynesian Eumusa bananas. *Pacific Science* 48:16-31.
- Lejju, B.J., P. Robertshaw & D. Taylor. 2006. Africa's earliest bananas? *Journal of Archaeological Science* 33:102-113.
- Markert, C. & F. Moller. 1959. Multiple Forms of Enzymes. *Biochemistry and Cell Biology* 45:753-762.
- Mbida, Ch., H. Doutrelepont, L. Vrydaghs, Ro. Swennen, Ru. Swennen, H. Beeckman, E. De Langhe & P. de Maret. 2001. First archaeological evidence of banana cultivation in central Africa during the third millennium before present. *Vegetation History and Archaeobotany* 10:1-6.
- Neumann, K. & E. Hildebrand 2009. Early bananas in Africa: The state of the art. *Ethnobotany Research and Applications* 7:353-362.
- Noyer, J.L., S. Causse, K. Tomekpe, A. Bouet & F.C. Baurens. 2005. A new image of plantain diversity assessed by SSR, AFLP and MSAP markers. *Genetica* 124:1-61.
- Raboin, L.M., F. Carreel, J.L. Noyer, F.C. Baurens, J.P. Horry, F. Bakry, H. Tezenas Du Montcel, J. Ganry, C. Lanaud & P.J.L. Lagoda. 2005. Diploid ancestors of triploid export banana cultivars: molecular identification of 2n restitution gamete donors and n gamete donors. *Molecular Breeding* 16:333-341.
- Risterucci, A.-M., I. Hippolyte, X. Perrier, L. Xia, V. Caig, M. Evers, E. Huttner, A. Kilian & J.-Ch. Glaszmann. in press. Development and assessment of diversity arrays technology (DArT) for high-throughput DNA analyses in *Musa*. Accepted for *Theoretical and Applied Genetics*.
- Rosales, F., E. Arnaud & J. Coto. 1999. Editors of *A Tribute to the Work of Paul Allen: A catalogue of wild and cultivated bananas*. International Network for the Improvement of Banana and Plantain Annual Report, Montpellier.
- Sharrock, S. 1990. Collecting *Musa* in Papua New Guinea. Pp. 140–157 in *Identification of Genetic Diversity in the Genus Musa*. International Network for the Improvement of Banana and Plantain, Montpellier.
- Shepherd, K. 1999. *Cytogenetics of the Genus Musa*. International Network for the Improvement of Banana and Plantain, Montpellier.
- Simmonds, N.W. 1962. *The Evolution of the Bananas*. Longmans, London.
- Simmonds, N.W. & K. Shepherd. 1955. The taxonomy and origins of the cultivated bananas. *Journal of the Linnean Society (Botany)*. 55:302-312.
- Tézenas du Montcel, H. 1990. *M. acuminata* subspecies *banksii*: Status and diversity. Pp. 211–218 in *Identification of Genetic Diversity in the Genus Musa*. Edited by R.L. Jarret. International workshop proceeding, Los Baños, Philippines, 5–10 September 1988. International Network for the Improvement of Banana and Plantain, Montpellier.
- Tischler, G. 1910. Untersuchungen über die Entwicklung des Bananen-Pollens. *Archiv für experimentelle Zellforschung* 5:622-670.
- Ude, G., M. Pillay, D. Nwakanma & A. Tenkouano. 2002a. Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. *Theoretical and Applied Genetics* 104:1239-1245.
- Ude, G., M. Pillay, D. Nwakanma & A. Tenkouano. 2002b. Genetic diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP markers. *Theoretical and Applied Genetics* 104:1246-1252.
- Valmayor, R.V. 2001. Classification and characterization of *Musa exotica*, *M. alinsanaya* and *M. acuminata* ssp. *errans*. *InfoMusa* 10:35-39.
- Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Friters, J. Pot, J. Paleman, M. Kuiper & M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Research* 23:4407-4414.
- White, P.R. 1928. Studies on the banana. An investigation of the floral morphology and cytology of certain types of the genus *Musa* L. *Zeitschrift für Zellforschung und Mikroskopische Anatomie* 7:673-733.
- Wong, C., R. Kiew, J.P. Loh, L.H. Gan, O. Set, S.K. Lee, S. Lum & Y.Y. Gan. 2001. Genetic diversity of the wild ba-

nana *Musa acuminata* Colla in Malaysia as evidenced by AFLP. *Annals of Botany* 88:1017-1025.

Wong, C., R. Kiew, G. Argent, O. Set, S.K. Lee & Y.Y. Gan. 2002. Assessment of the validity of the sections in *Musa* (Musaceae) using AFLP. *Annals of Botany* 90:231-238.