Screening for species-specific DNA families in *Musa acuminata*

Franc-Christophe Baurens[a]
Jean-Louis Noyer[b]
Claire Lanaud[b]
Pierre-Jean Laurent Lagoda[b*]

[a] Vitropic SA, ZAE les avants, 34270 Saint-Mathieu-de-Treviers, France

[b] Cirad-amis, Agetrop Laboratory, Avenue d'Agropolis, BP 5035, 34032 Montpellier cedex 01, France (lagoda@cirad.fr)

* Correspondence and reprints

Received 19 January 1999
Accepted 16 September 1999

Fruits, 2000, vol. 55, p. 3–15
© 2000 Éditions scientifiques et médicales Elsevier SAS
All rights reserved

RESUMEN ESPAÑOL, p. 15

Screening for species-specific DNA families in *Musa acuminata*.

Abstract — Introduction. Repetitive sequences are of great interest to developers of molecular markers because, in general, they cover large portions of the genome of higher plants. Plant genomes carry repetitive elements of various copy numbers and different levels of similarity between repeat families and between plant species. We propose a method of targeting species-specific repetitive elements for cloning and subsequent development of molecular marker systems based on the Polymerase Chain Reaction. **Materials and methods**. Genomic libraries were constructed and screened by Southern hybridization. Species-specificity was estimated using a specificity index (*Si*) based on differences in intensities of hybridization signals. **Results and discussion**. Four different *M. acuminata* subspecies, three different *M. balbisiana* type and one *M. schizocarpa* genomic libraries were constructed, characterized and screened for species specific probes. The *Si* proved paramount to discriminate species specific from non-specific repetitive DNA fragments. The total repetitive DNA content of the *Musa* genome could be assessed. Several A genome repetitive elements could be identified and are described. Within the *Eumusa* section, *Musa acuminata* species-specific DNA elements have been identified as either short interspersed elements (SINEs) or copia-like interspersed sequences. The banana genome is composed of 77 % repetitive elements and 23 % single copy sequences. **Conclusion and prospects**. The strategy presented allows for idenfication of repetitive elements with copy numbers above 1 000. These may be used to study the genomic composition of complex banana polyploid cultivars by in situ hybridization. © Éditions scientifiques et médicales Elsevier SAS

***Musa* / molecular biology / genetic markers / DNA / nucleotide sequence / identification / genomes / France**

Recherche de familles d'ADN spécifiques à l'espèce *Musa acuminata*.

Résumé — Introduction. Les séquences répétées sont d'un grand intérêt pour les utilisateurs de marqueurs moléculaires parce qu'en général, elles couvrent de grandes parties du génome des plantes supérieures. Le génome des plantes porte des éléments répétitifs en nombre variable et présente différents niveaux de similitude selon les familles de répétition et l'espèce végétale considérée. Nous proposons une méthode pour rechercher des éléments répétitifs spécifiques à une espèce pour le clonage suivi du développement de systèmes de marqueurs moléculaires, basée sur la réaction en chaîne de la polymérase. **Matériel et méthodes**. Des banques génomiques ont été construites et criblées par hybridations. La spécificité de l'espèce a été estimée en utilisant un index de spécificité (*Si*) basé sur la différence d'intensité des signaux d'hybridation. **Résultats et discussion**. Des banques génomiques de quatre sous-espèces différentes de *Musa acuminata*, de trois types différents de *M. balbisiana* et d'un *M. schizocarpa* ont été construites puis caractérisées, et des sondes spécifiques de ces espèces ont été recherchées. L'index de spécificité s'est avéré primordial pour distinguer les fragments d'ADN répétitifs des espèces, en fonction de leur spécificité. L'ensemble de l'ADN répétitif du génome de *Musa* a pu être évalué. Plusieurs des éléments répétitifs du génome A ont pu être identifiés et décrits. Dans la section d'*Eumusa*, des éléments d'ADN spécifiques à l'espèce *M. acuminata* ont été identifiés comme de courts éléments dispersés (SINEs) ou comme des séquences dispersées de type Copia. Le génome du bananier se compose de 77 % d'éléments répétés et de 23 % de séquences correspondant à de simples copies. **Conclusion et perspectives**. La stratégie présentée permet d'identifier les éléments répétés à plus de 1 000 copies. Ils peuvent être utilisés pour étudier, par hybridation in situ, la composition génomique de cultivars de bananes polyploïdes complexes. © Éditions scientifiques et médicales Elsevier SAS

***Musa* / biologie moléculaire / marqueur génétique / ADN / séquence nucléotidique / identification / génome / France**

1. introduction

Modern crop improvement is based on molecular marker assisted selection and introgression of genes of agronomic traits of interest, such as resistance to diseases and pests or quality enhancing genes. Marker-assisted breeding needs 'gene donors' chosen from core collections. The identification of potential progenitors based on agronomic traits, considering the relative difficulty of crossing banana due to the sterility and parthenocarpy of cultivars, requires a better knowledge of the genetic diversity and phylogenetic relationships in the *Musaceae*.

Repeated sequences, present in numbers varying between 100 and several million copies per genome, represent a large proportion of the genome and could be an efficient source for Polymerase Chain Reaction (PCR) based molecular markers. PCR amplification of these repetitive elements needs very small amounts of DNA compared to Restriction Fragment Length Polymorphism (RFLP) analyses, and only 'low quality' DNA (i.e., partially sheared), due to the relatively low size of the target amplicon (from 10 bp to few kb) and the efficiency of PCR. Thus, the search for repetitive sequences in the banana genome is of great interest for molecular breeding based strategies. PCR-based analysis of repetitive sequences should provide an efficient means of distinguishing the different individuals involved in complex cultivars or hybrids [1]. The presence of Variable Number of Dinucleotide Repeats (VNDR or microsatellites) with AT, CT, CA and GC motifs has been reported and complex core sequences have also been detected [2]. These microsatellites are randomly distributed in the banana genome, as has been shown with mapping experiments [3]. Variable Number of Tandem Repeat (VNTR) loci have also been reported and they are suitable markers for cultivar fingerprinting [4]. Independent species-specific repetitive DNA families have also been studied [1, 5]. It was demonstrated that decamer-primers defined in a species-specific repetitive element were able to detect polymorphism in banana cultivars [5]. An unrelated species-specific DNA family has also been assessed [6]. The purpose of the present paper is to establish and evaluate a protocol for the cloning of species-specific repetitive elements in banana. Finally, we will present an overview of repetitive sequences identified to date in the banana genome.

Table I.
Wild type diploid banana accessions used to sreen for species specific repeated DNA families in *Musa*, section *Eumusa*.

	Species	Subspecies	Name
1	*schizocarpa*	*schizocarpa*	Schizocarpa
2	*acuminata*	*banksii*	Madang
3	*acuminata*	*zebrina*	Monyet
4	*acuminata*	*malaccensis*	Malaccensis
5	*acuminata*	*burmannicoides*	Calcutta 4
6	*balbisiana*	*type IV*	Pisang Batu
7	*balbisiana*	*type I*	Honduras
8	*balbisiana*	–	Butuhan

2. materials and methods

2.1. plant material and DNA extraction

Total DNA was extracted by the cetyl-trimethyl ammonium bromide (CTAB) method [7] modified by Fauré et al. [8]. Five accessions representing the genetic diversity of *Musa acuminata*, three *Musa balbisiana* and one *Musa schizocarpa* were used in this study (*table I*). All accessions were obtained from the French West Indies' (Guadeloupe) collection.

2.2. restriction digest of genomic DNA

Four bp cutters (*Mbo* I, *Rsa* I, *Sau* 3A1) and 6 bp cutters (*Bam* HI, *Dra* I, *Eco* RI, *Eco* RV, *Pst* I) have been used to cut genomic DNA according to the instructions of the furnisher (BRL) for complete restriction.

2.3. RFLP analysis

RFLP analyses were performed according to Fauré et al. [8] modified as follows: 1.5 µg of total restricted DNA per lane were

loaded onto 1 % agarose gels and migrated at 120 V for 4 h in TBE 1X buffer (89 mM tris HCl pH = 8, 2 mM EDTA). Pre-hybridization was performed overnight, followed by a 4 h hybridization period. Final stringent wash was performed at 68 °C in 0.1 SSPE, 0.1 SDS. X-ray films were exposed for 16 h at –80 °C.

2.4. library construction and characterization

Eco RI and *Sau* 3A1 genomic DNA libraries were constructed on the five *Musa acuminata* subspecies. One µg of total genomic banana DNA was ligated onto 200 ng of appropriate (*Eco* RI or *Bam* HI) dephosphorylated pUC19 plasmid. Ligation was performed in 50 µl using 2 units of T4 DNA ligase, according to the specifications of the supplier (BRL). Ten aliquots of *E.coli* DH5α competent cells [9] were each transformed with 1/10th of the ligation product. Transformed bacteria were pooled before phenotype expression, and the libraries were amplified for three doubling cycles after antibiotic selection. Bacteria were harvested by centrifugation and re-suspended in 4 ml of LB medium. Two separate aliquots of 500 µl of bacterial solution corresponding to one library equivalent were both plated onto X-Gal, ampicillin, 2YT medium for recombinant plasmid selection [10]. Plates were incubated overnight at 37 °C and lifted or used for random choice of white colonies.

2.5. library screening

2.5.1. colony lift

Colonies were transferred onto Hybond® N+ nylon membranes according to the recommendations of the supplier (Amersham). Membranes were 'baked' in a microwave oven for 5 min at 1 000 W between two moistened Whatman 3M paper sheets. Membranes were pre-washed in 7 % SDS with vigorous shaking to remove contaminating bacterial compounds. Hybridizations were performed using the previously described protocol for RFLP analysis.

To probe the membrane, 200 ng of total genomic DNA from the homologous banana accession was used. Strong hybridization signals obtained after an overnight exposure were identified and relevant colonies were collected and grown overnight, then plasmid DNA was recovered using DNA mini-preparation [10].

2.5.2. random choice

Randomly-chosen white colony inserts were amplified by PCR using standard – 21M13 and reverse primers, and amplification products were blotted onto a Hybond® N+ membrane for further hybridization screening steps.

2.5.3. hybridization with chloroplast DNA

To measure chloroplast DNA proportion in total genomic DNA libraries, 100 ng of *Lotus* sp. total chloroplast DNA was hybridized following RFLP hybridization procedures described by Fauré et al. [8].

2.5.4. hybridization with total genomic DNA

Inserts of probes collected from colony lift screenings, as well as those randomly chosen clones from the libraries, were amplified by PCR using the standard M13 and Reverse universal primers. Amplification products were blotted onto nylon membranes as previously described. All inserts were hybridized using total genomic DNA of the eight banana accessions studied (*table I*). Hybridization signals obtained after 16 h exposure were graded using a scale from 0 (no signal) to 3 (strong signal) for each genomic DNA probe.

For all inserts, we determined a specificity index (*SI*) as the ratio between the number of specific signals [difference between A (from the *acuminata* genome) + S (from the *schizocarpa* genome) and B (from the *balbisiana* genome) signals] and the sum of the signals. Due to the difference in number of representatives of the A, B and S genomes, numbers of specific signals were normalized :

$$[[(A+S)/5]-(B/3)]/[[(A+S)/5]\pm(B/3)]$$

giving the final expression :

$$[3(A+S)-5B]/[3(A+S)+5B],$$

where: *'A'* is the sum of the signal intensities obtained for the four *Musa acuminata* genomic DNA probes, *'B'* is the sum of the signal intensities obtained for the three *Musa balbisiana* signal intensities, and *'S'* corresponds to the *Musa schizocarpa* signal intensity.

Characteristics of this *Si* indicator will be extensively discussed. Empirically, a cut-off was determined and inserts with $Si > 0.4$ were collected and hybridized as RFLP probes to verify specificity. Probes with a clear species-specific signal were sent to a subcontractor (genome express SA) for sequencing.

3. results and discussion

In banana, complete restriction of total DNA using 4 bp cutters (*Mbo* I, *Rsa* I, *Sau* 3A1) and 6 bp cutters (*Bam* HI, *Dra* I, *Eco* RI, *Eco* RV, *Pst* I) produces smears with weak underlying banding patterns in ethidium bromide (EtBr) stained gels (data not shown). These might represent equally important multiple repetitive sequence families or organelle DNA fragments. As no strong bands are apparent that allow discrimination of the different species or subspecies, it is very difficult to clone this type of sequence conventionally directly from the gel, as has proven possible in other species (*figure 1*). Thus, we chose to construct genomic libraries for differential screening. Genome size of the banana is lower than 1 Gbp [11]. Based on a 1 kb sequence with a copy number of up to 100 000, a library covering 0.1 % of the genome or 1 000 kbp should be sufficient to detect representatives of such sequences.

3.1. choice of banana species for library construction and screening

The genome of most cultivated bananas is derived from two major species identified as *Musa acuminata* (A genome) and *M. balbisiana* (B genome). It is well known that there exists considerable intraspecific genetic variation within the *M. acuminata* species complex. Recently, genetic diversity studies identified four major wild type groups of *M. acuminata* subspecies that are most important for cultivar breeding: ssp. *banksii*, ssp. *malaccensis*, ssp. *zebrina* and ssp. *burmannicoides* [12]. One representative of each of these groups was used in this study. *M. balbisiana* is the second most important species involved in the genomic composition of cultivars [12]; representatives of this species were therefore used for all screening steps. As *M. schizocarpa* was found to be closely related to *M. acuminata* [13] and because it is involved in the complex genomic formula of some cultivars [12], a *M. schizocarpa* (S genome) representative was also used for library construction and for the subsequent screening steps.

3.2. library construction and characterization

Eco RI and *Sau* 3A1 libraries were constructed using each of the four defined *M. acuminata* and *M. schizocarpa* accessions. Most of the inserts covered more than the limit size previously discussed for detecting repetitive sequences and represent approximately 1 % of the *M. acuminata* genome (*table II*). Due to the ligation conditions, we expected the occurrence of chimeric inserts. *Eco* RI libraries were tested for the presence of such inserts. The analysis of more than 100 clones identified only one putative chimera, thus a very low rate of chimerization is apparent. For the *Sau* 3A1 libraries, the proportion of chimeric inserts was greater, but they could be eliminated following the recommendations of Sambrook et al. [10] by adjusting the ratio between inserts and plasmids in the ligation mix.

Due to the great number of clones, exhaustive analysis was not possible and the different screening steps had to be evaluated to filter out species-specific sequences.

3.3. screening libraries for species specific probes

In a preliminary approach, screening steps were based on the method developed by Cai and Bullen [14] on *Phleum* sp. Unfor-

tunately, differential colony lifts failed to detect strong species-specific signals (data not shown), probably due to a combination of several factors, including copy number of putative species-specific sequences, which seems to be quite low in banana (less than 10 000 copies per genome). Efficacy, measured as the yield of candidate clones identified by differential colony lift, depends critically on copy number. Thus, colonies differing in stage (number of bacteria per colony and / or mean number of plasmids per cell) will have a decisive influence on differential signal intensities from plated bacteria, which are never perfectly homogeneous. This putative heterogeneity of the plated colonies might be sufficient to hide differential signals from medium repetitive elements in the background hybridization. For these reasons, each library was only screened for repetitive sequences with total homologous DNA.

In a second approach, 96 clones were randomly-chosen from each of the *Eco* RI libraries listed in *table II* (except for library No. 4 where tested clones issued from the *Sau* 3A1 library) and blotted for additional screening steps by hybridization.

An important screening step, when starting from total genomic DNA, is to assess the quantity of chloroplast genomic DNA, since it represents a sizeable proportion of the crude extracted DNA. This could lead to high intensity hybridization signals due to the high intracellular copy number of this genome. As the chloroplast genome seems to be well conserved among plants [15, 16], *Lotus* sp. total chloroplast DNA was used as a probe for identifying inserts of chloroplast origin. Our data suggest that, in banana extracts, the chloroplast genome represents up to 4.5 % of total DNA.

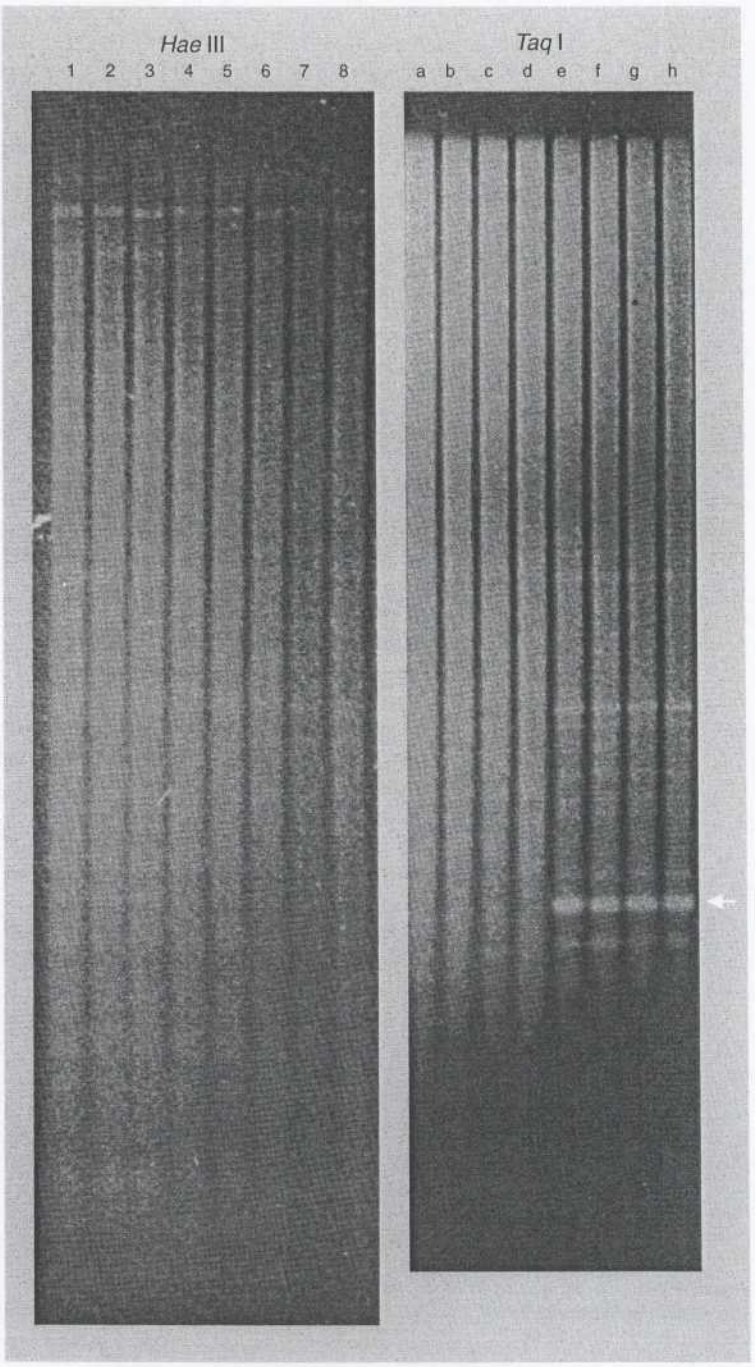

Figure 1.
Complete DNA restriction assay of species-specific repetitive sequences content of *Musa* and *Saccharum /Erianthus* complex species.
Lanes 1 to 8 corespond to banana accessions listed in *table I*. The restriction enzyme used is *Hae* III.
Lanes a to h correspond to *Saccharum barberi* "Paunra", *Saccharum sinense* "Oshima", *Saccharum edule* "NG 28201", *Miscanthus sinensis*, *Erianthus arundinaceus* "IK 7624", "IK 7648", "IS 76176", "NG 28.7", respectively. The restriction enzyme used is *Taq* a1
The arrowhead identifies a species-specific highly repetitive element.

Table II.
Library average insert size and genome coverage for the *Musa* accessions studied.

Banana accession[1]	Clone number *Eco* RI	Clone number *Sau* 3A1	Average insert size (in bp)	Coverage of the genome (in percent)
1	2 000	6 000	350	0.3
2	2 000	6 000	500	0.4
3	6 000	nd	250	0.2
4	150	3 000	400	0.1
5	6 000	70	500	0.3

[1] Banana accessions are those listed in *table I*.
The two restriction enzymes used for library construction (see materials and methods) are indicated.
Average insert size: ten randomly chosen clones were chosen from each library, inserts were amplified by PCR and migrated in agarose gel. After EtBr staining, insert size was determined in reference to a molecular weight standard and corrected for the size of the multiple cloning site.
Coverage of the genome was calculated assuming no duplicates, random distribution of the probes, and a genome size value of approximately 1 Gbp [19].
nd: not done.

3.4. use of the *Si* specificity index to assess species specific pattern of repetitive sequences

The *Si* indicator was developed to select inserts with species-specific signals. Due to the difference in number between the A, S and B representatives, total intensities were normalized in the final calculation of *Si*. Normalization was chosen to give a zero index for the inserts having proportionally the same number of A and B signals. If no signal appears on one or other species, thus detecting a species-specific probe, the absolute value of the *Si* (| *Si* |) is 1.0. Intermediate values of *Si* indicate partial specificity of the probe. *Si* variation is steeper for small than for high values of the graded signals (*figure 2*). Thus, probe discrimination, as compared to the previously described

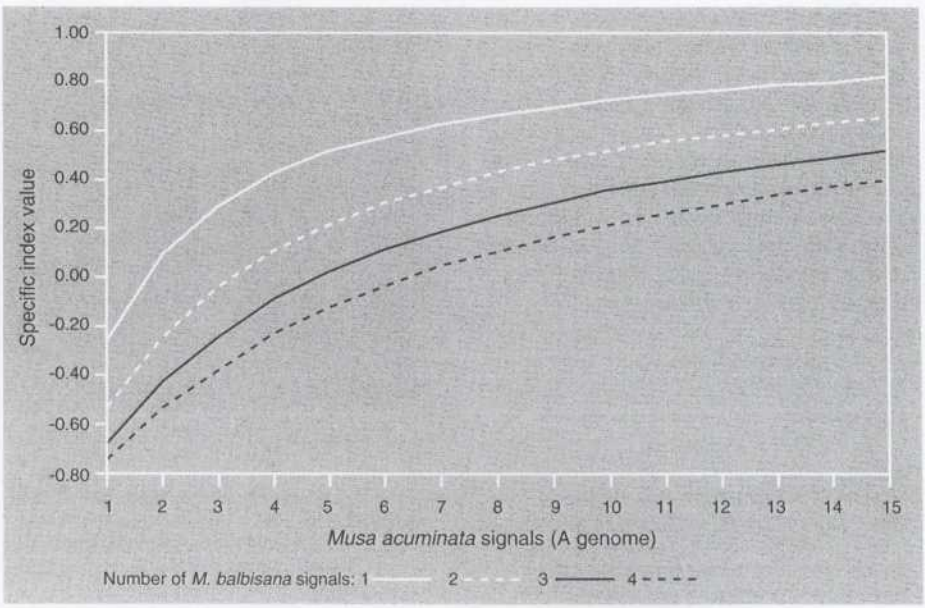

Figure 2.
The specificity index (*Si*) value variation: simulation of the *Si* value variation as a function of the number of *Musa acuminata* and *M. balbisiana* signals. Note that the *Si* variation is better for the small values of A signals.

method, is enhanced. A cut-off was empirically determined to select subsets of the tested probes. The choice of the cut-off followed two major rules: (i) the number of selected inserts should decrease significantly without excluding partially-specific probes, (ii) account should be made of the putative false positives (or negatives) due to the different hybridization steps and conditions (e.g., membrane ageing). Cut-off was made at an *Si* value of 0.4, corresponding to a selection of inserts exhibiting four times more signals (both intensity and number) in one species than in the other. A last characteristic of the *Si* value should be considered: 1 (or –1) corresponds to a species-specific probe regardless of the intensity of the signal. A probe exhibiting a weak signal with only one tested DNA will reach the same value as a probe with high intensity signals on all the representatives of one species.

Use of this cut-off value reduced the number of candidate probes to 226 representing about 40 % of the candidates issued from the previous step. The pre-selected probes, hybridized to restricted DNA blots from the test panel of banana species, exhibited a multiband pattern with varying copy numbers among the banana accessions (*figures 3, 4*). In the following paragraphs, representatives of the two classes of detected repetitive elements (i.e., non-specific and species-specific) are discussed.

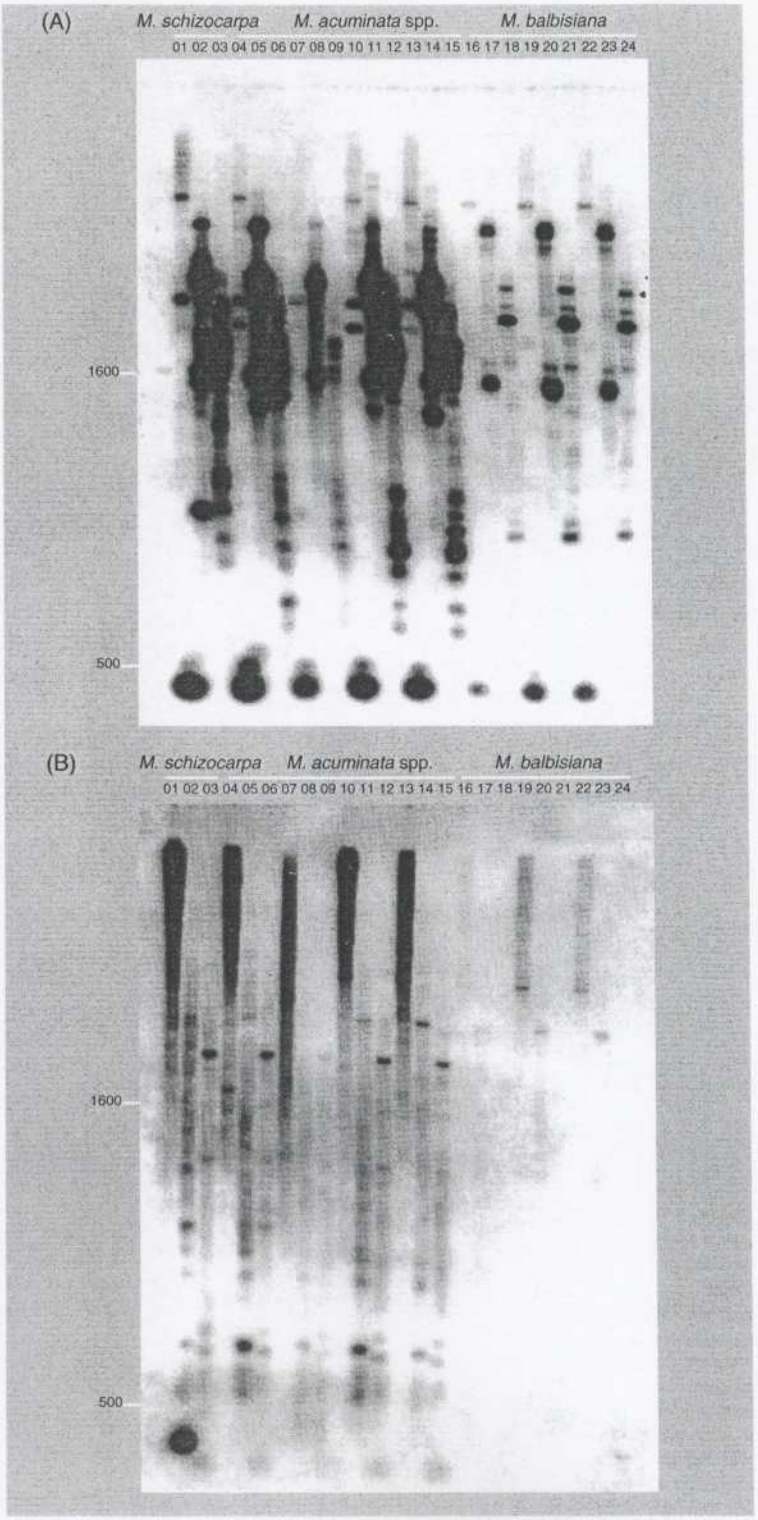

Figure 3.
RFLP hybridization of non-specific probes: probes pMaCIR 21 577 (A) and pMaCIR 40 564 (B) were used as RFLP probes on total genomic banana DNAs. The *Musa* species DNA are restricted with, respectively, the *Eco* RI, *Hae* III and *Sau* 3A restriction enzymes from left to right. Sizes in kb are indicated on the left.
Lanes 1–3: *M. schizocarpa*;
lanes 4–6: *M. acuminata* ssp. *banksii*;
lanes 7–9: *M. acuminata* ssp. *zebrina*;
lanes 10–12: *M. acuminata* ssp. *malaccensis*;
lanes 13–15: *M. acuminata* ssp. *burmannicoides*;
lanes 16–18: *M. balbisiana* "Pisang Batu";
lanes 19–2: *M. balbisiana* "Honduras";
lanes 22–24: *M. balbisiana* "Butuhan".

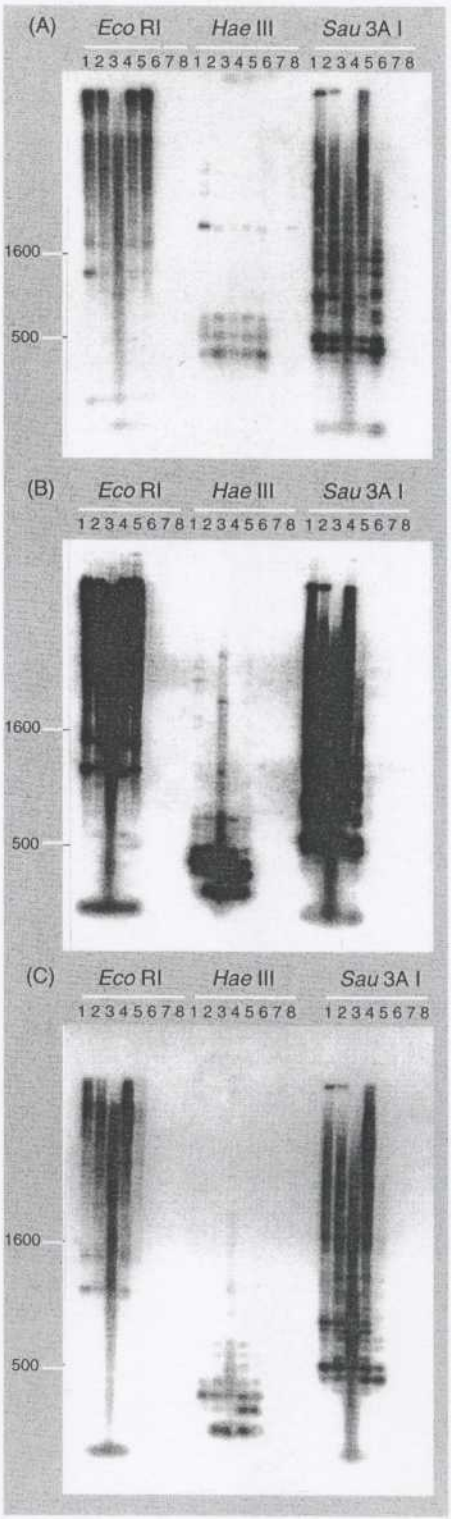

Figure 4.
RFLP hybridization of species-specific probes: probes pMaCIR 51 522 (A) and pMaCIR 51 517 (B) pMaCIR 21 589 (C) were used as RFLP probes on total genomic banana DNAs. Lanes 1 to 8 correspond to accessions listed in *table I*. Restriction enzymes are indicated above. Size in bp are indicated on the left.

3.5. non-specific probes

Most of these non-specific probes produce hybridization patterns with species-specific copy numbers, based on signal intensities. Moreover, the pattern is composed of a very low number of bands common to all the tested clones (*table III*). Thus, their use for genetic profiling is suggested. A direct consequence of this observation is that a higher cut-off value would have led to a loss of potentially useful probes. On the other hand, an additional screening step was necessary to increase selectivity.

3.6. species-specific repetitive sequences of the A genome

The efficacy of our screening method could be measured (*table IV*). Species-specific probes represented 15 % of our candidate clones, considerably more than for randomly-selected probes from a previously described [8, 12] size-selected *Pst* I library (< 0.1 %, i.e., 0 out of 724 probes) and an *Eco* RI library (1 %, i.e., 1 out of 98 probes). Strong differences are observed comparing efficiencies of the same screening steps between randomly-chosen and pre-selected probes. First, the repetitive sequence proportion is lower for the subset pre-selected by colony-lift. For practical purposes, areas with multiple colonies, sometimes inseparable, were chosen in the majority of cases. Thus, rapid analysis of the autoradiographs did not systematically enable unambiguous identification. This procedure leads to a decrease of the repetitive sequence proportion and could be eliminated using the standard method of replating the colonies from the interesting areas. This would, however, result in an increase of hybridization steps and time required.

The chloroplast genome copy number in crude DNA extracts is high, thus hybridization procedures produce strong signals which could be mistaken for signals arising from nuclear repetitive elements. Chloroplast probes are preferentially selected by the colony-lift screening procedure. This particular step could enable estimation of the colony-lift efficiency: the chloroplast proportion is doubled by this step (*table IV*).

Table III.
RFLP pattern analysis of three selected probes with no species-specific signal in *Musa*.

Probe	Number of bands[1]	Banding level[2]			*Si*[3]
		Common bands	Only A	Only B	
pMaCIR 31 501	19	1	0	0	1
pMaCIR 10 009	13	2	2	4	0.59
pMaCIR 20 009	20	1	0	2	1

[1] The number of bands were analysed on the *Eco* RI restriction.
[2] Banding level is 'common' if all tested individuals are positives; 'only A', if it appears only on all *M. acuminata* genotypes; 'only B', if it appears only on all *M. balbisiana* genotypes.
[3] *Si*: specific index. It is the ratio between the number of specific signals [difference between A (from the *acuminata* genome) + S (from the *schizocarpa* genome) and B (from the *balbisiana* genome) signals] and the sum of the signals.

Table IV.
Comparison of the efficiency at each step of the screening process with and without previous 'colony lift' screen from probe clone libraries in *Musa* (libraries are those previously described).

Steps of the screening procedure	Preselected probes		Randomly-chosen probes	
	Number of selected clones	%[1]	Number of selected clones	%[1]
Library	40 000	–	20 000	–
Random choice	–	–	414	2.1
Colony lift screening	325	0.81	–	–
Hybridization with total DNA	208	64.00	331	79.9
Hybridization with cpDNA	181	87.00	312	94.0
Si cut-off	30	16.60	196	62.8
RFLP analysis	0	–	29	14.8[2]

[1] Percentage of selected clones after the pre-screening step, in reference to the number of clones selected at the previous step.
[2] Percentage value calculated from 54 tested clones.

The selection potential of the *Si* indicator varies between the two approaches. It was found to be more effective for the pre-selected probes. In fact, the colony-lift screening step increases the number of all repetitive elements regardless of their specificity. This was evident from the higher cumulated average value of hybridization intensities per probe (see materials and methods). It reached a mean value of 5.4 units per pre-selected probe (colony lift), whereas this value was only 3.5 units per randomly-chosen probe. The pre-selected subset is composed in a great majority of non-specific repetitive elements with high copy numbers. As the *Si* indicator was built to discriminate repetitive elements based on their specificity, screening the pre-enriched subset of the library leads to a drastic decrease in the number of potentially positive probes. As a comparison, the randomly-chosen insert subset is more representative for probes with intermediate and low copy numbers, and was found to be more appropriate for our purposes, since most of the species-specific

Table V.
Evaluation of the proportion of species-specific repetitive sequence proportion in the banana genome.

Repetitive element	Number of non redundant clones	Percent of the genome
Brep 1	3	0.80
Copia-like	3	0.80
Others	1	0.25

Brep 1: Banana repetitive element 1 family; copia-like: sequences homologous to copia-like elements described by Turcich et al. [17]; others: not yet characterized repetitive sequences (i.e., pMaCIR 60 207).

sequences fell in this range of copy number. A differential screening step with a good predictive index allows the number of candidate RFLP probes to be greatly reduced.

Based upon the criteria discussed above, six candidate inserts were sequenced. Similarity searches for these species-specific sequences were performed on nucleic acid databases (EMBL). These sequences could thus be classified into different groups (*table V*). The most important species-specific sequence family was the Brep 1 family which have been extensively described elsewhere [6]. It corresponds to a Short Interspersed Element (SINE) family present in the A genome with various copy numbers [1]. A second sequence family composed of three representatives could be identified as a transposon-like element. Two of the selected probes (pMaCIR 21 589 and pMaCIR 51 522) were highly homologous to the copia-type retroelement PREM-2 sequence of *Zea mays* (maize) previously identified by Turcich et al. [17]. The pMaCIR 51 517 probe is less homologous to the maize copia-like element, but is highly similar to the other banana selected probes. Hybridization of the three probes showed that clearly they have a much greater abundance in the A and S genomes than in the B genome. In fact, only a very faint band appears (*figure 4a*, lane 7, and *figure 4c*, lanes 7 and 8). This transposon-like related repeated sequence family could be considered as species specific. Comparison between the three homologous representatives of this family revealed no clear differences in the species-specificity ot their RFLP patterns, but nucleotide variations affecting the sequences were identified. Moreover, distribution of representatives of this family in the A genome seems to be random, as indicated by multiple polymorphic bands discriminating subspecies (*figure 4*). The pMaCIR 60 207 exhibits no significant similarity to other characterized sequences of DNA in the databases searched. Further, this probe seems to be clustered, as three restriction enzymes tested exhibit one major hybridization signal (data not shown). Considering only one restriction enzyme, increases in nucleotide substitutions might cause such patterns as well, and explanation of these data would not require a change in location. But the data on three different restriction enzymes suggest that species-specific sequence families could either be dispersed along chromosomes as well as clustered in a few loci in the banana genome. However, further investigations are required for a better understanding of

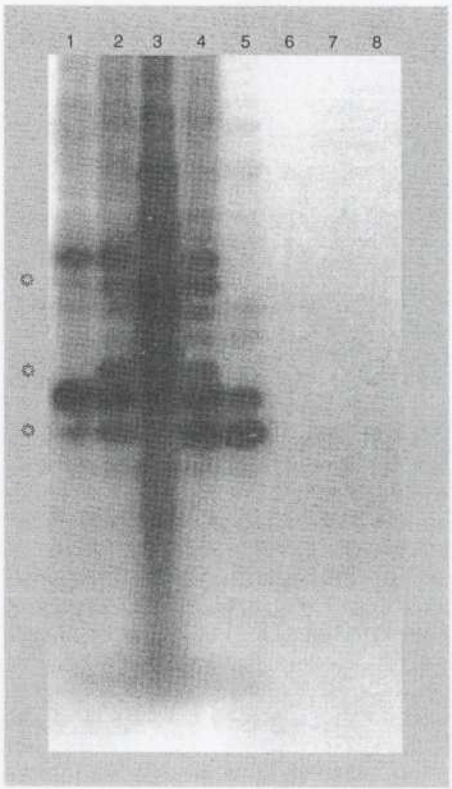

Figure 5.
Detail of the hybridization of pMaCIR 21 589 probe on *Sau* 3A1 total banana DNAs. Lanes 1 to 8 correspond to DNAs listed in *table I*. Stars indicate polymorphic bands discriminating *Musa acuminata* spp. and *M. schizocarpa*.

Table VI.
Evaluation of the repetitive sequence proportion in the banana genome. Repetitive nature of probes was determined by hybridization with total genomic DNA and calculated excluding the chloroplastic probes from the total. Chloroplast origin of probes was determined using chloroplast DNA from *Lotus* species.

Cloning enzyme	Number of tested clones	Repetitive DNA (%)	Unique sequences (%)	Chloroplastic DNA (%)
Eco RI	320	75.0	20.0	5
Eco RI*	257	77.0	23.0	nd
Sau 3A	94	77.0	19.0	4
Pst I*	257	76.5	23.5	nd
Mean	–	76 ± 1	21 ± 2	4.5

nd: not determined.
* Indicates compiled data from previously described partial libraries, Fauré et al. [8].

the genomic location of the different families of species-specific sequences or discussions of genomic rearrangements or transposon activity.

3.7. repetitive DNA proportion in the banana genome

We estimated roughly to what extent the libraries are representative. Two parameters are worthy of further discussion: (i) non randomly distributed restriction sites could have an influence (the use of two different restriction endonucleases for library construction should mitigate the consequences of this potential bias), (ii) methylation sensitivity of the restriction enzyme could also have an influence. Comparing two enzymes sensitive to different methylation positions of the same recognition site, *Sau* 3A1 and *Mbo* I, restrictions of total banana DNA produced no apparent differences (data not shown). If data from another methylation sensitive enzyme, *Pst* I, and from *Eco* RI are compared, the relative proportions of repetitive and single copy sequences are not affected (*table VI*). Thus, methylation is not of major concern here. Similar results were previously published by Baker et al. [18] on cotton.

Our data were compared to those obtained on a previously described, nuclei-enriched, partial banana library [8]. The single copy sequence percentage is slightly increased for the DNA extracted from the nuclei. This could be partially due to undetected chloroplast contaminants remaining in the single copy subset of the nuclei-enriched libraries. In fact, one chloroplastic probe has subsequently been detected in the single copy subset of the *Eco* RI library [12] and four in the single copy subset of the *Pst* I library (data not shown). Chloroplast material was found to represent up to 4.5 % of the libraries from total banana DNA.

In summary, roughly 80 % of the banana genome appears to be repetitive DNA. As a comparison, the cotton genome is composed of 70 % repetitive sequences [18].

Based on the measured 2C value [11], this finding is in accordance with cell DNA content of banana and cotton, assuming a strong correlation between genome size and the proportion of repetitive sequences.

4. conclusion and prospects

Our strategy to target copy numbers above 1 000 produced a sizeable amount of species-specific repetitive elements (*tables IV, V*). As evidenced by Jarret et al. [5], different species-specific elements exist in the banana genome. Some of these elements belong to the DNA fraction with less than

1 000 copies [1]. To estimate sequence family copy numbers, the number of the homologous clones per family was counted and compared to the number of tested clones. Data suggest that roughly 7 % of the A genome is composed of species-specific sequences (*table IV*) whereas 2 % of the A genome is represented by two major families (*table V*). Species-specific repetitive sequence families may explain part of the variation in the genome size (40 %) observed between the A and B genomes [19] and between A subspecies (10 %). The method proposed here is an efficient way of obtaining species-specific sequences of the banana genome. It could be used as an alternative to previously described methods [14], in particular when the analysed genome is poor in these elements. In the banana genome, the presence of multiple non-related species-specific repetitive elements could thus be demonstrated. The interest of these sequences for phylogeny analyses or fingerprinting are discussed elsewhere [14, 20, 21]. Sequence data from these elements may foster PCR based molecular marker development using primers anchored in the core sequences, according to previously described approaches for typing, fingerprinting or subspecies identification [1, 4, 5, 22]. This is an important prerequisite to PCR tests discriminating between the different components of the complex genome structure of current cultivars (species and subspecies hybrids). These DNA elements are sure to become a powerful tool for in situ hybridization. Used in this manner, repetitive DNA elements will facilitate the study of interspecific crosses for a better understanding of the genomic composition of complex polyploid banana cultivars.

acknowledgements

We would like to thank P. Gauthier for providing purified *Lotus* sp. chloroplast DNA, K. Alix for data concerning *Erianthus* sp. highly repetitive element. We are specially indebted to D. Jarrell for his helpful suggestions and text improvements.

This work was supported by a Cirad grant.

references

[1] Baurens F.-C., Noyer J.-L., Lanaud C., Lagoda P.J.L., Competitive PCR assays of repetitive DNA sequences in banana, Mol. Gen. Genet. 253 (1996) 57–64.

[2] Lagoda P.J.L., Noyer J.-L., Dambier D., Baurens F.-C., Lanaud C., Abundance and distribution of simple sequence repeats in the Musaceae family: microsatellite markers to map the banana genome, In: International symposium on the use of induced mutations and molecular techniques for crop improvement, International Atomic Energy Agency and FAO, Vienna, Austria, 1995, pp. 287–295.

[3] Lagoda P.J.L., Noyer J.-L., Update on *Musa* genome mapping at Cirad-Agetrop, Infomusa 2 (1994) 4.

[4] Kaemmer D., Afza R., Weising K., Kahl G., Novak F.J., Oligonucleotide and amplification fingerprinting of wild species and cultivars of banana (*Musa* spp.), Bio/technology 10 (1992) 1030–1035.

[5] Jarret J.L., Vuylsteke D.R., Gawel N.J., Pimentel R.B., Dunbar L.J., Detecting genetic diversity in diploid bananas using PCR and primers from a highly repetitive DNA sequence, Euphytica 68 (1993) 69–76.

[6] Baurens F.-C., Noyer J.-L., Lanaud C., Lagoda P.J.L., Assessment of a repetitive DNA family (Brep1) in *Musa acuminata*, Theor. Appl. Genet. 95 (1997) 922–931.

[7] Gawel N.J., Jarret R.L., A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea*, Plant Mol. Biol. Rep. 9 (1991) 262–266.

[8] Fauré S., Noyer J.-L., Horry J.-P., Bakry F., Lanaud C., González de León D., A molecular marker-based linkage map of diploid bananas (*Musa acuminata*), Theor. Appl. Genet. 87 (1993) 517–526.

[9] Hanahan D., Studies on transformation of *E. coli* with plasmids, J. Mol. Biol. 66 (1983) 557–580.

[10] Sambrook J., Fritsch E.F., Maniatis T., Molecular cloning: a laboratory manual, Cold Spring Harbor Laboratory Press, New York, USA, 1989.

[11] Bennett M.D., Leitch I.J., Nuclear DNA amount in angiosperms, Ann. Bot. 76 (1995) 113–176.

[12] Carreel F., Étude de la diversité génétique des bananiers (genre *Musa*) à l'aide de marqueurs RFLP, Thesis, Ina Paris-Grignon, Paris, France, 1994.

[13] Simmonds N.W., The evolution of the bananas, Longmans, Tropical Sciences Series, London, United Kingdom, 1962.

[14] Cai Q., Bullen M.R., Analysis of genome-specific sequences in *Phleum* species: identification and use for study of genomic relationships, Theor. Appl. Genet. 88 (1994) 831–837.

[15] Clegg M.T., Gaut B.S., Learn G.H. Jr., Morton B.R., Rates and patterns of chloroplast DNA evolution, Proc. Nat. Acad. Sci. USA 91 (1994) 6795–6801.

[16] Demesure B., Sodzi N., Petit J., A set of universal primers for amplification of polymorphic non-coding regions of mitochondrial and chloroplast DNA in plants, Mol. Ecol. 4 (1995) 129–131.

[17] Turcich M.P., He C., Hamilton D.A., Mascarenhas J.P., PREM-2, a copia-type retroelement in maize is expressed preferentially in early microspores, Sex. Plant Reprod. 9 (1996) 65–74.

[18] Baker R.J., Longmire J.L., Van Der Bussche R.A., Organization of repetitive elements in the Upland cotton genome (*Gossypium hirsutum*), J. Hered. 86 (1995) 178–185.

[19] Dolezel J., Dolezelova M., Novak F. J., Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*), Biol. Plant. 36 (1994) 351–357.

[20] Crowhurst R.N., Gardner R.C., A genome specific repeat sequence from kiwifruit (*Actinidia deliciosa* var. *deliciosa*), Theor. Appl. Genet. 81 (1991) 71–78.

[21] Baurens F.-C., Noyer J.-L., Lanaud C., Lagoda P.J.L, Inter-*Alu* like profiling in banana, Euphytica 99 (1998) 137–142.

[22] Jeffreys A.J., MacLeod A., Tamaki K., Neil D.L., Monckton D.G., Minisatellite repeat coding as a digital approach to DNA typing, Nature 354 (1991) 204–209.

Búsqueda de familias de ADN específicas de la especie *Musa acuminata*.

Resumen — Introducción. Las secuencias repetidas presentan un gran interés para los utilizadores de marcadores moleculares porque, en general, cubren amplias partes del genoma de las plantas superiores. El genoma de las plantas contiene un número variable de elementos repetitivos y presenta diferentes niveles de similitud según las familias de repetición y la especie vegetal considerada. Aquí se propone un método, basado en la reacción en cadena de la polimerasa, a fin de buscar los elementos repetitivos específicos de una especie para el clonado seguido del desarrollo de sistemas marcadores moleculares. **Material y métodos**. Se construyeron bibliotecas genómicas que se exploraron mediante hibridación. La especificidad de la especie se estimó utilizando un índice de especificidad (*SI*) basado en la diferencia de intensidad de las señales de hibridación. **Resultados y discusión**. Se construyeron y caracterizaron bibliotecas genómicas de cuatro subespecies diferentes de *Musa acuminata*, de tres tipos diferentes de *M. balbisiana* y de un *M. schizocarpa*; se buscaron sondas específicas de dichas especies. El índice de especificidad se reveló primordial para distinguir los fragmentos de ADN repetitivos de las especies en función de su especificidad. Se pudo evaluar el conjunto del ADN repetitivo del genoma de *Musa*. Se identificaron y describieron varios elementos repetitivos del genoma A. En la sección de *Eumusa*, se identificaron elementos de ADN específicos de la especie *M. acuminata* como elementos cortos entremezclados (SINEs) o como secuencias entremezcladas asimiladas a copias. El genoma del banano se compone de un 77 % de elementos repetidos y un 23 % de secuencias que corresponden a simples copias. **Conclusión y perspectivas**. El método presentado permite la identificación de los elementos repetidos hasta más de 1 000 copias. Dichos elementos pueden utilizarse para estudiar, mediante hibridación in situ, la composición genómica de cultivares de bananos poliploides complejos. © Éditions scientifiques et médicales Elsevier SAS

***Musa* / biología molecular / marcadores genéticos / ADN / sequencia nucleotídica / identificación / genomas / Francia**