Original Article

Cytogenet Genome Res 119:268–274 (2007)
DOI: 10.1159/000112073

Cytogenetic and Genome Research

Isolation and characterization of the highly repeated fraction of the banana genome

E. Hřibová[a] M. Doleželová[a] C.D. Town[b] J. Macas[c] J. Doležel[a,d]

[a]Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Olomouc (Czech Republic)
[b]The Institute for Genomic Research, A Division of the J. Craig Venter Institute, Rockville, MD (USA)
[c]Institute of Plant Molecular Biology, Biological Centre ASCR, České Budějovice, and
[d]Department of Cell Biology and Genetics, Palacký University, Olomouc (Czech Republic)

Accepted in revised form for publication by B. Friebe, 14 August 2007.

Abstract. Although the nuclear genome of banana (*Musa* spp.) is relatively small (1C ~ 610 Mbp for *M. acuminata*), the results obtained from other sequenced genomes suggest that more than half of the banana genome may be composed of repetitive and non-coding DNA sequences. Knowledge of repetitive DNA can facilitate mapping of important traits, phylogenetic studies, BAC-based physical mapping, and genome sequencing/annotation. However, only a few repetitive DNA sequences have been characterized in banana. In this work, we used DNA reassociation kinetics to isolate the highly repeated fraction of the banana genome (*M. acuminata* 'Calcutta 4'). Two libraries, one prepared from Cot ≤0.05 DNA (2,688 clones) and one from Cot ≤0.1 sequences (4,608 clones), were constructed, and 614 DNA clones were chosen randomly for sequencing and further characterization. Dot-plot analysis revealed that 14% of the sequenced clones contained various semi-tandem and palindromic repeated sequences. 'BLAST' homology searches showed that, in addition to tandem repeats, the Cot libraries were composed mainly of different types of retrotransposons, the most frequent being the *Ty3/gypsy* type *monkey* retrotransposon. Selected sequences displaying tandem organization properties were mapped by PRimed IN Situ DNA labeling (PRINS) to the secondary constriction on metaphase chromosomes of *M. acuminata* 'Calcutta 4'. Southern hybridization with selected BAC clones carrying 45S rDNA confirmed the presence of the tandem repeats in the 45S rDNA unit. This work significantly expands the knowledge of the repetitive fraction of the *Musa* genome and organization of its chromosomes.

Bananas (*Musa* spp.) are giant perennial herbs growing in humid tropics and subtropics. Their fruits are a major commodity in international trade with a net value second only to citrus (INFO COMM, 2005 http://r0.unctad.org/infocomm/anglais/banana/market.htm). Out of the 106 million tons produced annually, only about 13% of bananas are exported (FAO, 2005 http://faostat.fao.org/); the majority of bananas are consumed by locals who rely upon this plant as a staple in their diets. This critical role has been threatened recently by viral and fungal diseases, nematodes, and pests (Robinson, 1996). Their spread has been accelerated by the asexual nature of banana cultivars, all of which are seed-sterile diploid, triploid and tetraploid clones originating from hybridization between diploid progenitors *Musa acuminata* Colla and *Musa balbisiana* Colla (Simmonds and Shepherd, 1955). While traditional breeding has been hampered by seed sterility and limited knowledge of genetic diversity, lack of information about the nature of important traits has hindered the application of molecular tools.

This work was supported by Academy of Sciences of the Czech Republic (grant no. A600380703), Czech Science Foundation (204/04/1207), International Atomic Energy Agency (Research Contract no. 12230) and Generation Challenge Programme (LOA INIB 2006/15).

Request reprints from Jaroslav Doležel
Institute of Experimental Botany
Laboratory of Molecular Cytogenetics and Cytometry
Sokolovská 6, CZ–77200 Olomouc (Czech Republic)
telephone: +420 585 205 852; fax: +420 585 205 853
e-mail: dolezel@ueb.cas.cz

1424–8581/07/1194–0268$23.50/0

Accessible online at:
www.karger.com/cgr

Despite its enormous socioeconomic importance, banana remains an under-researched crop that, like citrus, is in considerable danger due to its dependence upon asexual reproduction. Clearly, greater knowledge of the banana genome is needed to facilitate breeding improved cultivars. Moreover, the evolution of bananas and their biology provide unique opportunities to study traits not available in other model monocot species, such as different forms of seed sterility combined with parthenocarpy as well as fruit physiology (Simmonds, 1953), and to unravel the role of interspecific hybridization and polyploidization in plant evolution. The importance of analyzing the banana genome is further illuminated by the fact that it has been considered an attractive candidate for angiosperm comparative genomics. While being a monocot species, banana is only distantly related to rice, sorghum, maize, and wheat, and represents an important node in the study of monocot evolution as an outgroup to the grasses (Paterson et al., 2004).

Although the nuclear genomes of *M. acuminata* and *M. balbisiana* are small (1C ~ 610 and 550 Mbp, respectively) (Doležel et al., 1994; Lysák et al., 1999; Bartoš et al., 2005), sequencing research on banana has been limited (Aert et al., 2004; Cheung and Town, 2007). Moreover, banana is not expected to be the subject of whole genome sequencing in the near future. Of particular note, there is little information on the repetitive sequences that comprise a significant fraction of the banana genome. Up to now, only a few repetitive DNA sequences of banana have been studied. Valárik et al. (2002) identified and characterized a set of 12 clones containing repetitive DNA sequences. In addition to these '*Radka*' repeat sequences, only a species-specific element (Brep-1; Baurens et al., 1997), the *Ty3/gypsy*-type retrotransposon *monkey* (Balint-Kurti et al., 2000), and a *Copia*-like element (Baurens et al., 1996) have been described for *Musa*.

Recently, there has been renewed interest in the use of DNA reassociation kinetics techniques (Britten and Kohne, 1968) to fractionate genomes into repetitive and low-copy components. These methods are based on the observation that specific sequences reassociate at different rates in proportion to their copy number (Britten and Kohne, 1968; Britten and Davidson, 1976). DNA reassociation of a eukaryotic genome can be described by a Cot curve that consists of one or more fast-reassociating repetitive DNA components and a slow-reassociating single/low-copy component, where Co is the starting concentration of nucleotides and t is the reassociation time. Low Cot values correspond to highly repetitive DNA while high Cot values are associated with single and low-copy DNA sequences.

To date, Cot techniques have been used to isolate and sequence repetitive DNA in plants to a limited extent (Peterson et al., 2002; Ho and Leung, 2002). Here we report on its application to investigate highly-repetitive DNA sequences from banana with the aim of characterizing the repetitive landscape of the genome. Moreover, we describe repetitive DNA sequences suitable for use as chromosome-specific landmarks.

Materials and methods

Plant material

In vitro rooted plants of *Musa acuminata* 'Calcutta 4' (2n = 2x = 22, AA) (ITC 0249) and *M. balbisiana* 'Tani' (2n = 2x = 22, BB) (ITC 1120) were obtained from the INIBAP Musa Transit Centre (ITC, Katholieke Universitiet, Leuven, Belgium). Plants were transferred to soil and maintained in a greenhouse.

Shearing and ressociation of DNA

Total cellular genomic DNA of *M. acuminata* 'Calcutta 4' was isolated from fresh cigar leaves using a modified CTAB method (Rogers and Bendich, 1985) and sheared by sonication to approximately 250–400 bp fragments. Two 50 μg aliquots of sheared DNA in 0.12 M sodium phosphate buffer (SPB; pH 6.8) were denatured by heating in boiling water and allowed to reassociate at 60°C to Cot values of 0.05 and 0.1 M.s, respectively (Britten and Kohne, 1968). For each sample, reannealed double-stranded DNA was separated from single-stranded DNA by hydroxyapatite chromatography (Peterson et al., 2002). Briefly, the hydroxyapatite column was washed three times with pre-warmed (60°C) 0.17 M SPB to remove single-stranded DNA. Double-stranded DNA was then eluted by washing the column with 0.3 M SPB (60°C).

Construction of Cot libraries

The double stranded DNA eluates were desalted and the DNA concentrated using a Microcon YM-30 spin column (Millipore, USA). The two double stranded DNA samples were treated with mung bean nuclease (Amersham, UK) for 1 h at 30°C to generate blunt-ended molecules, purified by phenol-chlorophorm extraction, and concentrated using Microcon YM-30 spin columns. For each sample, a 40 ng aliquot was used in a ligation reaction with *Sma*I-digested, dephosphorylated pBluescript SK^+ plasmid vector (Stratagene, USA), to produce recombinant molecules. The recombinant molecules were used to transform *E. coli* (XL1 blue MRF), and recombinant clones were picked into 384-well plates containing 2YT medium with 15% glycerol and stored at –80°C until use.

Screening of DNA libraries

Cot libraries were spotted onto Hybond-N+ filters (Amersham) and screened with probes for *Radka*1 and *Radka*2 sequences, which carry parts of 45S and 5S rDNA, respectively, and *Radka*3, *Radka*4, *Radka*5, *Radka*6, *Radka*8, *Radka*10, *Radka*12 and *Radka*14 sequences (Valárik et al., 2002). A genomic BAC library of *M. acuminata* 'Calcutta 4' (MA4, Vilarinhos et al., 2003) was spotted onto Hybond-N+ filters and screened with a probe for *Radka*1, to identify clones carrying units of 45S rDNA. In both cases, the probes were labeled with alkaline phosphatase and hybridization was performed using the AlkPhos Direct Kit (Amersham) according to manufacturer's instructions. Positive clones were detected using chemiluminescent substrate (CDP Star, Amersham).

Sequence analysis

DNA clones selected from Cot libraries were sequenced at the University of Arizona (Tucson) and at the Institute for Genomic Research (Rockville, MD, USA). Sequence data were compared using the Dotter software (Sonnhammer and Durbin, 1995) and searched for homology to sequences in the GenBank database using 'BLAST 2.0.2' (Altschul et al., 1997).

Genomic organization of tandem repeats

Aliquots of genomic DNA of *M. acuminata* 'Calcutta 4' were digested using *Rsa*I, *Hae*III, *Mse*I, *Alu*I, *Eco*RI, *Sma*I, *Sac*I and *Dra*I. BAC clones carrying 45S rDNA that were selected from the MA4 BAC library were digested using *Dra*I, *Sac*I and *Taq*I. Digested genomic DNA and BAC DNA were size-fractionated by 1.5% agarose gel electrophoresis and transferred to Hybond N+ nylon membranes (Amersham). Clones from the Cot ≤0.05 library containing tandemly organized sequence units (C427, 2F10, 4E2, C444 and 7D20) as revealed by Dotter software were labeled using alkaline phosphatase (AlkPhos Direct Kit, Amersham) and used as probes for Southern blots of genomic DNA. The probes for C427, 2F10, 4E2, 18S rDNA and 26S rDNA respectively, were used for hybridization with Southern blots of digested DNA of BAC clones.

Table 1. Oligonucleotide primers used for PRINS on mitotic chromosomes

Tandem repeat	Primer sequence
MaTR1	5′-GTA ACC CGT CTG CGT CGT-3′ 5′-TTA CCG CGT AGG TGT GGT CT-3′
MaSTR1	5′-CTACTCGCCGCTGTCACAAT-3′ 5′-AGAAGACCAAGGCTGCTGAG-3′
MaSTR2	5′-GCACTGCCAGAGCTCAAGTT-3′ 5′-GGCAGAGCTCGAAGACTGAG-3′

Copy number of tandem repeats

The copy number of newly identified repetitive DNA sequences was estimated for the genomes of *M. acuminata* 'Calcutta 4' (AA genome) and *M. balbisiana* 'Tani' (BB genome). Serial dilutions of genomic DNA and PCR products of isolated repeats used as standards were dot-blotted onto Hybond-N+ membranes (Amersham). The PCR products were labeled using alkaline phosphatase and used as hybridization probes. Dots of genomic DNA and standards that gave the same intensity of hybridization signals were identified after visual inspection. Copy numbers of individual probes were estimated assuming that 1 pg of genomic DNA equals 0.978×10^9 bp (Doležel et al., 2003).

Cytogenetic mapping

Metaphase spreads were prepared from root tips of *M. acuminata* 'Calcutta 4' according to Doleželová et al. (1998). Localization of DNA sequences was carried out using direct C-PRINS (Cycling PRimed IN Situ DNA labeling) as described by Kubaláková et al. (1997). Briefly, the reaction mixture consisted of 0.1 mM dATP, dCTP, dGTP and 0.01 mM Alexa Fluor 488-5-dUTP, 0.017 mM dTTP, 2.5 mM $MgCl_2$ and 3 U/40 μl of Taq polymerase (Finnzymes, Finland) in 1× PCR buffer. Synthetic oligonucleotides specific for tandem repeats C427, 4E8 and 2F10, respectively, were designed using the Primer3 software (Rozen and Skaletsky, 2000) and were used as primers (Table 1) at 1 μM concentration. The temperature profile of the reaction consisted of denaturation at 92°C for 1 min, primer annealing at 58°C for 50 s, and extension at 72°C for 1 min. After 20 reaction cycles, the chromosome preparations were counterstained with DAPI (0.2 μg/ml) and mounted in Vectashield antifade solution (Vector Laboratories, UK). The slides were examined with an Olympus AX70 fluorescence microscope, and the images of DAPI and Alexa Fluor 488-5-dUTP fluorescence were acquired separately with a cooled high-resolution black and white CCD camera that was interfaced to a PC running the MicroImage software (Olympus, Japan). The images were superimposed after contrast and background optimization.

Results and discussion

Here we report on the first systematic attempt to characterize the repeated portion of the banana genome. Although it is not known how much of the banana genome is repetitive, its size may be estimated by considering a species with a similar genome size. Rice *(Oryza sativa)*, until now the only monocot species with a genome sequence near completion, possesses a small genome (1C = 490 Mbp; Bennett and Smith, 1976) whose size is not very different from that of banana. Sequence analysis revealed that about 35% of the rice genome accounted for non-coding sequences (Vij et al., 2006). Provided that the number and size of banana genes do not differ significantly from rice, the banana genome should contain about 55% non-coding DNA accounting for about 300 Mbp. However, this may be still an underestimate as a significant fraction of rice heterochromatin has not yet been sequenced.

Although the principle of Cot analysis has been known for a long time (Britten and Kohne, 1968; Britten and Davidson, 1976), its recent revival was stimulated by a need to fractionate nuclear genomes into various classes of DNA differing in their redundancy for further characterization and sequencing (Peterson et al., 2002; Yuan et al., 2003). With the exception of Ho and Leung (2002), the application of Cot fractionation in plant genomics has focused largely on study of low- and single-copy DNA ('genic' sequences) (Yuan et al., 2003; Lamoureux et al., 2005). While others have used Cot techniques to isolate repetitive DNA (Hanson et al., 1995; Zhang et al., 2004), sequencing of isolated repeats has been uncommon. Our results suggest that Cot analysis is useful for fractionation of banana genomic DNA and for isolation of repetitive DNA for sequencing.

In this work we created two low-Cot libraries from *M. acuminata* 'Calcutta 4'. The Cot ≤0.05 and Cot ≤0.1 libraries consisted of 2,688 and 4,608 clones, respectively. In order to avoid analysis of already known sequences, we screened both libraries for the presence of clones carrying repetitive DNA sequences isolated and characterized in our previous study (Valárik et al., 2002). Colony filters were screened with probes for *Radka*1 and *Radka*2 sequences, carrying parts of banana 45S and 5S rDNA units, respectively. The second screening step was done with a pooled probe prepared from other *Radka* repetitive sequences (see Materials and methods). Positive clones identified with the *Radka* probes accounted for approximately 2.5% of clones in both Cot-libraries.

The Cot ≤0.05 library, which was expected to contain more repetitive DNA clones, was then characterized in more detail. One thousand clones that did not hybridize with the *Radka* repetitive sequences were selected randomly and amplified using PCR. Out of them 614 clones gave PCR products with sizes ranging from 300 to 900 bp and were selected for sequencing. All sequences were deposited in GenBank (accession numbers ED827164 to ED827777).

Out of the 614 sequenced clones, 48% represented novel and previously undescribed sequences (Fig. 1). The remaining clones carried sequences homologous to DNA sequences already present in GenBank. Some of the clones were suspected to be chimeric as they contained two parts, each homologous to a different DNA sequence. Use of standard blunt-end cloning is likely the reason for the chimeric molecules, and this approach should not be used in future work. Despite negative selection of all clones hybridizing with the *Radka* probes prior to sequencing, 9.8% of sequenced clones were homologous to rDNA sequences and 3.7% of sequenced clones were homologous to other *Radka* repetitive sequences. The presence of *Radka* sequences in the newly sequenced clones could be due to the occurrence of other parts of rDNA sequences than those present in the *Radka* clones or due to the occurrence of chimeric clones.

Table 2. Classification of DNA clones sequenced from the Cot ≤0.05 library based on homology to DNA sequences deposited in GenBank

DNA sequence	Homologous clones, %
Retrotransposons	24.1
rDNA	9.8
Radka repetitive DNA sequences[a]	3.7
Other *Musa* repetitive elements[b]	7.7
Telomere	0.9
Chloroplast DNA	4.2
tRNA	0.3
mRNA	1.3
Total	52

[a] *Radka*3, *Radka*4, *Radka*5, *Radka*6, *Radka*8, *Radka*10, *Radka*12 and *Radka*14 (Valárik et al., 2002).
[b] MUSA6 clone (Ndowora et al., 1999), MUSA8 clone (Ndowora et al., 1999), *Musa acuminata* repetitive element (Y10143, Baurens, unpublished) and Brep1 (Baurens et al., 1997).

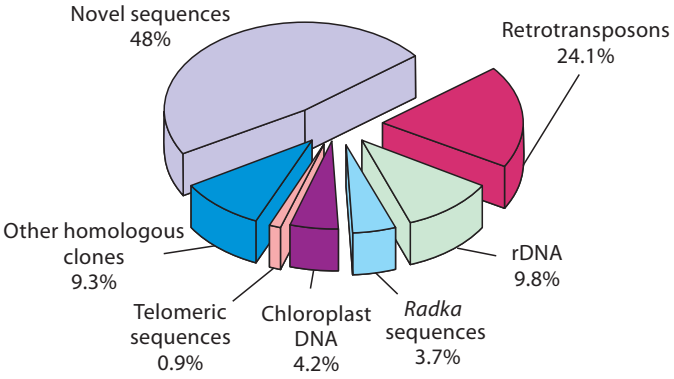

Fig. 1. Homology of DNA clones isolated from the Cot ≤0.05 library to sequences deposited in GenBank. Out of the 614 sequenced clones, 48% clones represented novel sequences. The majority of clones homologous to known sequences, i.e. 24.1% of all clones, showed homology to different types of retrotransposons. 9.8% of all sequenced clones were homologous to rDNA sequences and 3.7% of all sequenced clones showed homology to other *Radka* sequences that were characterized previously (Valárik et al., 2002). 4.2% of all sequenced clones were homologous to chloroplast DNA and 0.9% of sequenced clones carried telomeric sequences. The remaining clones (9.3% of all sequenced clones) were homologous to various sequences without a common character.

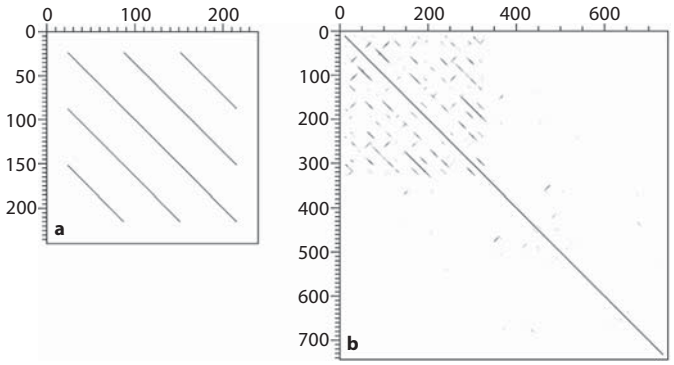

Fig. 2. Dot-blot analysis was used to compare the areas of internal sequence similarity of clone C427 carrying MaTR1 DNA sequence (**A**) and clone 4E2 that contains MaSTR2 DNA sequence (**B**). Similarities are represented by dots or lines.

Out of the 52% clones homologous to known sequences, a majority (24.1%) showed homology to various types of retrotransposons, the *Ty3/gypsy* type *monkey* retrotransposon (Balint-Kurti et al., 2000) being the most frequent among them (16%). 4.2% of sequenced clones were homologous to chloroplast DNA (Table 2). The genomic DNA used in this study was isolated from fresh cigar leaves which could explain the contamination by chloroplast DNA.

A significant portion of clones sequenced from the Cot ≤0.05 library (ca 13%) showed homology to '*Musa acuminata* clone MuG9, genomic sequence' (Aert et al., 2004), a bacterial artificial chromosome (BAC) clone that is rich in repetitive DNA and in part shows homology to the retrotransposon *monkey*. Out of the 13% repetitive sequences homologous to MuG9, 39% clones were found to be homologous to putative LTR sequences and 47.5% of the clones were homologous to the 3′ non-coding part of the *monkey* retrotransposon (Balint-Kurti et al., 2000) carrying the banana repetitive sequences *Radka*8 and *Radka*9 (Valárik et al., 2002). The remaining 13.5% of clones showed homology to polyprotein-like sequence *G9pp4*. 4.2% of sequenced clones showed homology to the repetitive part of '*M. acuminata* class III acidic chitinase gene, complete cds' (AY525367) and 19.5 % of sequenced clones showed homology to different types of the sequences deposited in GenBank (Table 2).

The fact that most of the sequenced clones were homologous to various retrotransposons and *Radka* sequences confirmed the highly repetitive nature of the genome fraction that we obtained after the Cot fractionations. Our data also confirm the presence of different types of LTR sequences in DNA clone MuG9 and provide further support for the observation that the retrotransposon *monkey* is the most abundant mobile element in the banana genome.

Dot-blot analysis revealed that 87 (14%) of the sequenced clones contained various semi-tandem and palindromic repeated sequences (Fig. 2). Of these, 30 clones contained semi-tandem organized repeats, 28 clones carried microsatellite sequences, 6 clones carried telomeric sequences, and 22 clones contained palindromic sequences. Isolation of the highly repeated portion of the genome using Cot fractionation therefore seems feasible to target these classes of repeats. Nevertheless, only a few tandem organized repeats were isolated, indicating a low abundance of this class of repeats in the relatively small banana genome. This observation contrasts with the studies on large genomes, where clusters of tandem repeats often form large interstitial heterochromatin blocks (Macas et al., 2000).

As clusters of tandem repeats may be useful as cytogenetic markers (Pedersen et al., 1996; Tsujimoto et al., 1997;

Table 3. Copy number and genomic distribution of selected tandem repeats

DNA clone	DNA sequence	Insert length (bp)	Copy number per genome (1C)		GenBank accession number	Genomic distribution as determined by PRINS
			M. acuminata 'Calcutta 4'	*M. acuminata* 'Tani'		
C427	MaTR1	60	$1.0 \times 10^2 - 2.5 \times 10^2$	$0.9 \times 10^1 - 4.0 \times 10^1$	ED827195	Strong signal in secondary constriction and weak signals in centromeric regions
2F10	MaSTR1	185	$4.0 \times 10^2 - 6.0 \times 10^2$	$0.6 \times 10^2 - 1.0 \times 10^2$	ED827350	Strong signal in secondary constriction and other signals clustered in pericentromeric and telomeric regions of most of the chromosomes
4E2	MaSTR2	158	$5.5 \times 10^3 - 7.0 \times 10^3$	$4.0 \times 10^2 - 6.0 \times 10^2$	ED827432	

Navrátilová et al., 2003) five clones with tandemly organized repetitive units longer than 50 bp (C427, 2F10, 4E2, C444 and 7D20) that showed no homology to sequences deposited in GenBank were selected for further characterization. With the aim of verifying their genomic organization, the clones were hybridized as probes to genomic DNA digested with eight restriction endonucleases (*Rsa*I, *Hae*III, *Mse*I, *Alu*I, *Eco*RI, *Sma*I, *Sac*I and *Dra*I). Intensely hybridizing bands that showed ladder-like patterns were observed for clone 4E2 only; clones 2F10 and C427 gave weak ladder-like hybridization signals (data not shown) indicating lower abundance in the genome of *M. acuminata* 'Calcutta 4'. Clone C427 contains a DNA sequence that we named MaTR1 (GB accession number: ED827195) and the two other clones carry sequences that we named MaSTR1 (clone 2F10, GB accession number: ED827350) and MaSTR2 (clone 4E2, GB accession number: ED827432).

Southern analysis supported the observations of dot-blot analysis, which was used to determine the copy number of these three clones. For the remaining two clones (C444, 7D20) the hybridization signals were below the detection limit, indicating a low number of copies in the genome, an observation that was confirmed by dot-blot analysis. Copy numbers of tandem organized sequences were estimated in two accessions representing the A and B genomes of *Musa*: *M. acuminata* 'Calcutta 4' and *M. balbisiana* 'Tani', respectively. The number of copies of the C427 clone (MaTR1 DNA sequence) ranged from 100 to 250 per unreplicated haploid A genome (1C) and 9 to 40 in the B genome. Even higher differences in copy number between the A and B genomes were observed for clones 2F10 and 4E8 (MaSTR1 and MaSTR2 DNA sequences, respectively, Table 3). However, at the absence of sequence data we cannot exclude that the differences in hybridization signal intensity were due to sequence divergence between the A and B genomes.

Our results indicate that the clones C444 and 7D20 contain sequences that are low-copy in the banana genome. As the Cot filtration is an enrichment technique, one may expect contamination of repeat components by low-copy sequences. An alternative explanation could be that these clones were artifacts that originated during the reassociation step of the Cot fractionation. However, even if we did not estimate a total number of artifacts in the Cot ≤ 0.05 library, we expect their proportion to be low due to the short reassociation time. Other tandem repeats (MaTR1, MaSTR1 and MaSTR2) were found to be moderately abundant. All three of these sequences appear to be less frequent in the B genome of *M. balbisiana* as compared to the A genome of *M. acuminata*. Similar difference in the frequencies of repetitive DNA sequences between the two genomes were observed previously (Valárik et al., 2002) and are consistent with the smaller genome size of *M. balbisiana* (Doležel et al., 1994).

In order to characterize chromosome distribution of tandem repeats MaTR1, MaSTR1 and MaSTR2, we used PRINS rather than FISH, as the former is suitable for rapid localization of tandem organized repeats on plant chromosomes (Kubaláková et al., 1997; Menke et al., 1998). The genomic distribution of MaTR1, MaSTR1 and MaSTR2 was analyzed on mitotic metaphase chromosomes of *M. acuminata* 'Calcutta 4'. C-PRINS with primers specific for MaTR1 resulted in strong fluorescence signals in the secondary constriction. In addition, weak signals were observed in centromeric regions of all chromosomes (Fig. 3A–D). However, we cannot exclude a possibility that the weak signals were due to non-specific labeling after 20 cycles of the PRINS reaction (Kubaláková et al., 1997). Interestingly, MaSTR1 and MaSTR2 were also localized in the secondary constrictions. However, unlike the MaTR1 sequence, relatively strong signals were observed in pericentromeric regions and other distal parts of chromosomes (Fig. 3E, F). The signals of MaSTR2 were stronger than signals from MaSTR1, confirming the copy number estimates based on Southern blotting (Table 3).

Despite the results of the 'BLAST' search for tandem repeats MaTR1, MaSTR1 and MaSTR2, which showed no homology to sequences deposited in GenBank, the results obtained by PRINS on mitotic chromosomes indicated co-localization with 45S rDNA and suggested that these sequences could be a part of the 45S rDNA unit. To confirm this observation, three BAC clones carrying 45S rDNA units were digested using *Dra*I, *Sac*I and *Taq*I and hybridized separately with probes for MaTR1, MaSTR1 and MaSTR2, as well as a combined probe for 18S rDNA and 26S rDNA. Southern hybridization confirmed that the clones MaTR1 (Fig. 4), MaSTR1 and MaSTR2 (data not shown) hybridized with the bands of the restricted BAC clones and indicated that the clones were parts of the 45S rDNA unit localized in the secondary constriction.

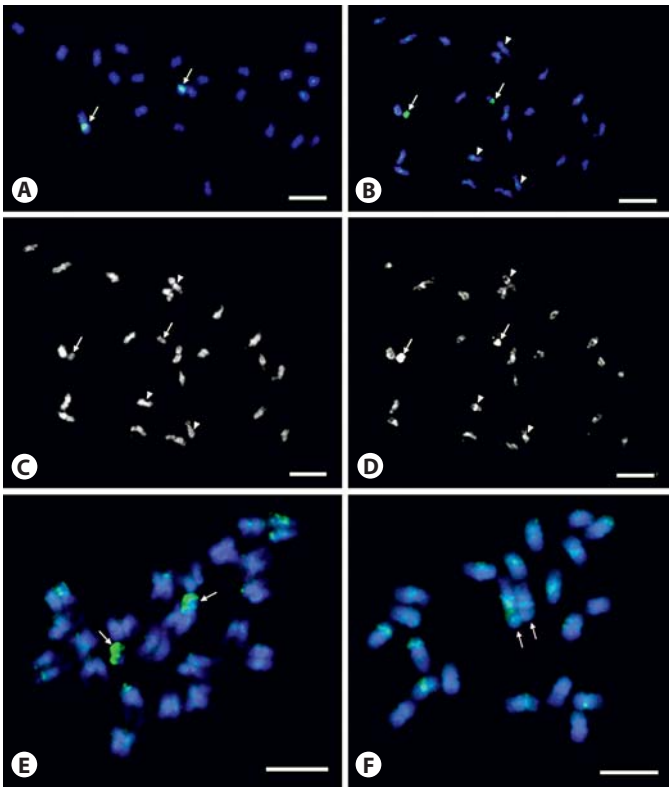

Fig. 3. Cytogenetic mapping of tandem organized repeats using C-PRINS. The repeat MaTR1 localized preferentially in the secondary constriction (**A**, arrows). In addition, weak signals were observed in centromeric regions (arrowheads) of all chromosomes (**B**, **C**, **D**). Also other two repeats (MaSTR1 and MaSTR2) were localized in the secondary constriction (arrows) and relatively strong signals were observed in pericentromeric regions and other distal parts of chromosomes (**E**, **F**). The signals of clone 4E2 containing MaSTR2 DNA sequence were stronger and better visible (**F**) than signals of clone 2F10 containing MaSTR1 DNA sequence (**E**), which was in agreement with their copy number. Tandem repeats were visualized using Alexa Fluor 488 (yellow-green color) and the chromosomes were counterstained by DAPI (blue color). Bar = 5μm.

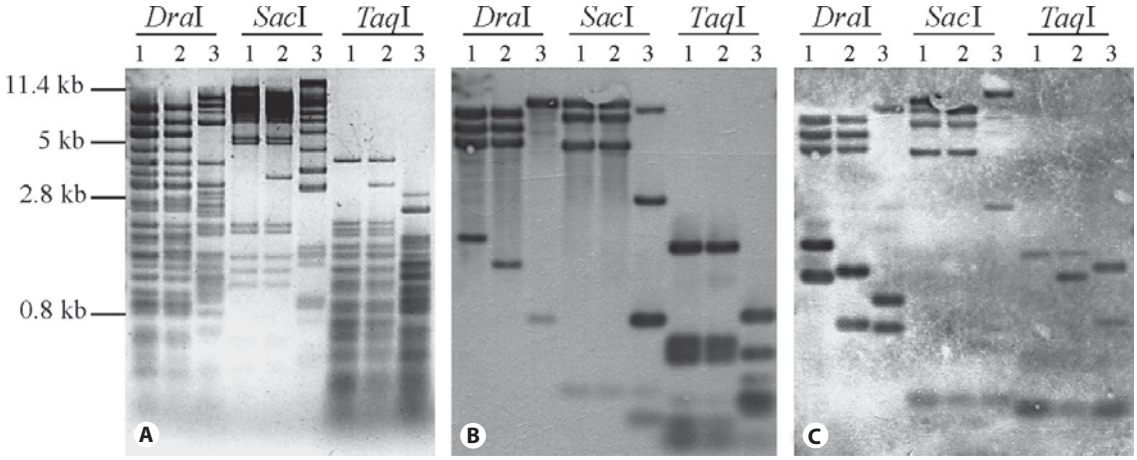

Fig. 4. Arrangement of the new tandem repeats with respect to 45S rDNA. BAC library of *M. acuminata* 'Calcutta 4' was screened with a probe for 26S rDNA to select BAC clones carrying 45S rDNA units. Selected BAC clones (1: MA4_1P13; 2: MA4_1J14, 3: MA4_1B22) were digested using *Dra*I, *Sac*I and *Taq*I and size fractionated on agarose gel (**A**). Southern hybridization with probes for clone C427 (**B**) and with probes for 18S rDNA and 26S rDNA (**C**) confirmed the presence of the new tandem organized repeats in the 45S rDNA unit.

Together with the previous studies of the locus carrying 18S-5.8S-26S rDNA in different plant species (McMullen et al., 1986; Rogers and Bendich, 1987; Ueki et al., 1992; Suzuki et al., 1996), our results obtained after Southern hybridization with the rDNA bearing BAC clone indicated that the newly identified tandem repeats were repetitive parts of non-transcribed intergenic spacers (IGS) of 45S rDNA. Our results suggest that the MaSTR1 and MaSTR2 tandem repeats are parts of IGS and are localized in distal parts of all chromosomes of *M. acuminata*. This raises a question about their evolution and movement in the genome. One possibility is that they originated from IGS regions that were amplified as extrachromosomal elements, and were reintegrated into the genome (Lohe and Roberts, 1990). Alternatively, they could be derived from a retrotransposon, e.g. *monkey*, which is also localized in the secondary constriction.

In conclusion, this study expands significantly our knowledge of the low-Cot (repetitive) fraction of the banana genome. We isolated and sequenced 614 repetitive DNA clones and found 370 sequences with no known homology to sequences in the GenBank. In addition to providing a broad picture of the repetitive component of the *Musa* genome, the results will support new sequence annotation and the analysis of genetic diversity with the aim of improving the current classification and taxonomy of *Musa*.

Acknowledgements

We thank Ir. I. van den Houwe, Musa curator of the INIBAP Transit Centre at the Katholieke Universiteit, Leuven (Belgium) for the supply of plant material. We acknowledge the receipt of a copy of genomic BAC library of *M. acuminata* 'Calcutta4' (MA4) from the Musa Genome Resource Centre http://www.musagenomics.org/. We are grateful to Dr. Marie Kubaláková for preparing the color plate, and our colleagues R. Tušková and M. Sekerová, Bc. for excellent technical assistance.

References

Aert R, Sagi L, Volckaert G: Gene content and density in banana *(Musa acuminata)* as revealed by genomic sequencing of BAC clones. Theor Appl Genet 109:129–139 (2004).

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402 (1997).

Balint-Kurti PJ, Clendennen SK, Doleželová M, Valárik M, Doležel J, et al: Identification and chromosomal localization of the monkey retrotransposon in *Musa* sp. Mol Gen Genet 263: 908–915 (2000).

Bartoš J, Alkhimova O, Doleželová M, De Langhe E, Doležel J: Nuclear genome size and genomic distribution of ribosomal DNA in *Musa* and *Ensete* (Musaceae): taxonomic implications. Cytogenet Genome Res 109:50–57 (2005).

Baurens FC, Noyer JL, Lanaud C, Lagoda PJL: Use of competitive PCR to essay copy number of repetitive elements in banana. Mol Gen Genet 253:57–64 (1996).

Baurens FC, Noyer JL, Lanaud C, Lagoda PJL: Assessment of a species-specific element (Brep 1) in banana. Theor Appl Genet 95:922–931 (1997).

Bennett MD, Smith JB: Nuclear DNA amounts in angiosperms. Phil Trans R Soc Lond B Biol Sci 274:227–274 (1976).

Britten RJ, Davidson EH: Studies on nucleic acid reassociation kinetics: empirical equations describing DNA reassociation. Proc Natl Acad Sci USA 73:415–419 (1976).

Britten RJ, Kohne DE: Repeated sequences in DNA. Science 161:529–540 (1968).

Cheung F, Town CD: A BAC end view of *Musa acuminata* genome. BMC Plant Biol 7:29 (2007).

Doležel J, Doleželová M, Novák FJ: Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). Biol Plant 36:351–357 (1994).

Doležel J, Bartoš J, Voglmayr H, Greilhuber J: Nuclear DNA content and genome size of trout and human. Cytometry 51:127–128 (2003).

Doleželová M, Valárik M, Swennen R, Horry JP, Doležel J: Physical mapping of the 18S-25S and 5S ribosomal RNA genes in diploid bananas. Biol Plant 41:497–505 (1998).

Hanson RE, Zwick MS, Choi SD, Islam-Faridi MN, McKnight TD, et al: Fluorescent in situ hybridization of a bacterial artificial chromosome. Genome 38:646–651 (1995).

Ho IS, Leung FC: Isolation and characterization of repetitive DNA sequences from *Panax ginseng*. Mol Genet Genomics 266:951–961 (2002).

Kubaláková M, Macas J, Doležel J: Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. Theor Appl Genet 94: 758–763 (1997).

Lamoureux D, Petersen DG, Li WL, Fellers JP, Gill BS: The efficacy of Cot-based gene enrichment in wheat (*Triticum aestivum* L.). Genome 48: 1120–1126 (2005).

Lohe AR, Roberts PA: An unusual Y chromosome of *Drosophila simulans* carrying amplified rDNA spacer without RNA genes. Genetics 125: 399–406 (1990).

Lysák MA, Doleželová M, Horry JP, Swennen R, Doležel J: Flow cytometric analysis of nuclear DNA content in *Musa*. Theor Appl Genet 98: 1344–1350 (1999).

Macas J, Požárková D, Navrátilová A, Nouzová M, Neumann P: Two new families of tandem repeats isolated from genus *Vicia* using genomic self-priming PCR. Mol Gen Genet 263:741–751 (2000).

McMullen MD, Hunter B, Phillips RL, Rubenstein I: The structure of the maize ribosomal DNA spacer region. Nucleic Acids Res 14:4953–68 (1986).

Menke M, Fuchs J, Schubert I: A comparison of sequence resolution on plant chromosomes: PRINS versus FISH. Theor Appl Genet 97: 1314–1320 (1998).

Navrátilová A, Neumann P, Macas J: Karyotype analysis of four *Vicia* species using in situ hybridization with repetitive sequences. Ann Bot (Lond) 91:921–926 (2003).

Ndowora T, Dahal G, LaFleur D, Harper G, Hull R, Olszewski NE, Lockhart B: Evidence that badnavirus infection in *Musa* can originate from integrated pararetroviral sequences. Virology 255:214–220 (1999).

Paterson AH, Bowers JE, Chapman BA: Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics. Proc Natl Acad Sci USA 101:9903–9908 (2004).

Pedersen C, Rasmussen SK, Linde-Laursen I: Genome and chromosome identification in cultivated barley and related species of the Triticeae (Poaceae) by in situ hybridization with the GAA-satellite sequence. Genome 39:93–104 (1996).

Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, et al: Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res 12:795–807 (2002).

Robinson JC: Bananas and Plantains (CAB International, Oxon 1996).

Rogers SO, Bendich AJ: Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. Plant Mol Biol 5:69–76 (1985).

Rogers SO, Bendich AJ: Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Mol Biol 9:509–520 (1987).

Rozen S, Skaletsky HJ: Primer3 on the WWW for general users and for biologist programmers, in Krawetz S, Misener S (eds): Bioinformatics: Methods and Protocols: Methods in Molecular Biology, pp 365–386 (Humana Press, Totowa 2000).

Simmonds NW: The development of the banana fruit. J Exp Bot 4:87–105 (1953).

Simmonds NW, K Shepherd: The taxonomy and origins of the cultivated bananas. Bot J Linn Soc 55:302–312 (1955).

Sonnhammer ELL, Durbin R: A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167:GC1–GC10 (1995).

Sun M, Chen H, Leung FC: Low-Cot DNA sequences for fingerprinting analysis of germplasm diversity and relationships in *Amaranthus*. Theor Appl Genet 99:464–472 (1999).

Suzuki A, Tanifuji S, Komeda Y, Kato A: Structural and functional characterization of the intergenic spacer region of the rDNA in *Daucus carota*. Plant Cell Physiol 37:233–238 (1996).

Tsujimoto H, Mukai Y, Akagawa K, Nagaki K, Fujigaki J, et al: Identification of individual barley chromosomes based on repetitive sequences: Conservative distribution of Afa-family repetitive sequences on the chromosomes of barley and wheat. Gene Genet Syst 72:3003–3009 (1997).

Ueki M, Uchizawa E, Yakura K: The nucleotide sequence of the rDNA intergenic spacer region in a wild species of the genus *Vicia, V. angustifolia*. Plant Mol Biol 18:175–178 (1992).

Valárik M, Šimková H, Hřibová E, Šafář J, Doleželová M, Doležel J: Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (*Musa* spp.). Chromosome Res 10:89–100 (2002).

Vij S, Gupta V, Kumar D, Vydianathan R, Raghuvanshi S, et al: Decoding the rice genome. Bioassays 28:421–432 (2006).

Vilarinhos AD, Piffanelli P, Lagoda P, Thibivilliers S, Sabau X, et al: Construction and characterization of a bacterial artificial chromosome library of banana (*Musa acuminata* Colla). Theor Appl Genet 106:1102–1106 (2003).

Yuan Y, SanMiguel PJ, Bennetzen JL: High-Cot sequence analysis of the maize genome. Plant J 34: 249–255 (2003).

Zhang P, Li WL, Fellers J, Friebe B, Gill BS: BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements. Chromosoma 112:288–299 (2004).