Mol Gen Genet (2000) 263: 908–915 © Springer-Verlag 2000

ORIGINAL PAPER

P. J. Balint-Kurti · S. K. Clendennen · M. Doleželová
M. Valárik · J. Doležel · P. R. Beetham · G. D. May

Identification and chromosomal localization of the *monkey* retrotransposon in *Musa* sp.

Received: 30 August 1999 / Accepted: 7 April 2000

Abstract Retroelements are ubiquitous features of eukaryotic genomes, often accounting for a substantial fraction of their total DNA content. One major group of retroelements, which includes the *gypsy* and *copia*-like elements, is distinguished by the presence of long terminal repeats (LTRs). We have identified and partially characterized a sequence from banana (*Musa acuminata* cv. Grand Nain) which shows significant homology to *gypsy*-like LTR retroelements from other species. The element, named *monkey*, shows a high degree of homology to the reverse transcriptase, RNase H and integrase genes of retroelements from plants, fungi and yeast. However, several stop codons are present in the major ORF of this element, suggesting that this copy of *monkey*, if functional, is non-autonomous. Southern analysis indicated that *monkey* is present in both the A and B genomes of *Musa*, and that it is found in 200–500 copies per haploid genome in cv. Grand Nain. Chromosomal localization by fluorescent in-situ hybridization indicates that copies of *monkey* are concentrated in the nucleolar organizer regions and colocalize with rRNA genes. Other copies of *monkey* appear to be dispersed throughout the genome.

Key words *Skippy* · Nucleolar organizer region (NOR) · Somaclonal variation · Repetitive sequence · In situ hybridization

Communicated by H. Saedler

P. J. Balint-Kurti (✉)
DNA Plant Technologies, 6701 San Pablo Ave.,
Oakland CA 94608, USA
e-mail: balint-kurti@DNAP.com
Fax: +1-510-5472817

P. J. Balint-Kurti · S. K. Clendennen[1]
P. R. Beetham[2]
Boyce Thompson Institute, Tower Rd.,
Ithaca, NY 14853, USA

G. D. May
Noble Foundation Inc., P.O. Box 2180,
2510 Sam Noble Parkway, Ardmore, OK 73402, USA

M. Doleželová · M. Valárik · J. Doležel
Institute of Experimental Botany,
Laboratory of Molecular Cytogenetics and Cytometry,
Sokolovská 6, CZ-77200 Olomouc, Czech Republic

Present addresses:
[1] Agritope Inc., 16160 SW Upper Boones Ferry Rd.,
Portland, Oregon 97224-7744, USA
[2] Kimeragen, 300 Pheasant Run, Newtown, PA 18940, USA

The first two authors contributed equally to this work

Introduction

It has become clear in the last few years that retroelements are major constituents of most plant genomes. An early report identified a *gypsy*-like element in lily (Smyth et al. 1989), and LTR (long terminal repeat) retroelements have since been characterized from a number of plant species, including maize, tomato (Su and Brown 1997) and pine (Kossack and Kinlaw 1999). In maize, over half the ~2500-Mb genome is accounted for by characterized LTR retrotransposon families (SanMiguel et al. 1996). Plants with smaller genome sizes (e.g. *Arabidopsis* – approximately 150 Mb) tend to accumulate fewer copies of this type of element. However, *Arabidopsis* does contain members of the two major classes of LTR retroelements commonly found in plants – *copia*- and *gypsy*-like retroelements (Tutois et al 1999).

LTR retrotransposons are thought to be closely related to retroviruses. Retroviruses possess an RNA genome which encodes the genes responsible for completion of its replication cycle. The *gag* domain encodes proteins that form the nucleocapsid core. The protease (*pr*) domain encodes a protein responsible for the maturation of the different proteins from an initial polyprotein. The product of the reverse transcriptase (RT)/RNase H domain transcribes the RNA genome into DNA prior to insertion in the host genome, and the *endo* domain encodes the integrase protein which has endonuclease activity and is required for the integration of

the reverse-transcribed element into the host genome (Kunze et al. 1997; Grandbastien 1998). The main difference between retrotransposons and retroviruses seems to be that retrotransposons do not possess an *env* domain, which is responsible for the production of envelope proteins needed for the formation of extracellular infectious virions. Thus, LTR retrotransposons can perform all the functions of classical retroviruses, save for the fact that they cannot move between cells. Classical retroviruses have not yet been identified in plants. Plant pararetroviruses differ from retroviruses in that they do not encode an integrase function (for an exception, see Richert-Pöggler and Shepherd 1997) nor do they produce long terminal repeats necessary for integration in the host genome. Recently it was discovered that the genomes of many *Musa* cultivars contain integrated copies of the banana streak badnavirus (a pararetrovirus) genome. There is compelling evidence to suggest that, during tissue culture, these integrated copies can give rise to full-blown banana streak disease (Harper et al. 1999; Ndowora et al. 1999).

We report here on an LTR retrotransposon, *monkey*, that constitutes up to ~0.5% of the *Musa* genome. Banana (*Musa* sp.) has a relatively small genome at about 500–600 Mb/haploid genome (Lysak et al. 1999). Baurens et al. (1997) previously reported on a *copia* homolog which was repetitive and highly polymorphic between different banana accessions. However, this is the first report of an LTR retrotransposon of the *gypsy* class in *Musa* spp.

Bananas are a clonally propagated crop and somaclonal variants are often encountered during propagation. It has been suggested that the activation of retrotransposons during tissue culture may account for somaclonal variation in plants, and furthermore, that active retroelements may serve as a tool for the functional analysis of genes by insertional mutagenesis (Hirochika et al. 1996). The identification and characterization of banana retrotransposons is a significant step towards investigating these hypotheses in *Musa* spp.

Materials and methods

Plant materials

All *Musa* species and clones used in the characterization of the *monkey* retrotransposon were obtained from the INIBAP Transit Centre (Katholieke Universiteit, Leuven, Belgium) as in vitro rooted plantlets. After transfer to soil, plants were maintained in a greenhouse. For the in situ hybridizations the diploid *M. acuminata* cv. Niyarma Yik (ITC 0269) and *M. balbisiana* cv. Singapuri (ITC 0248) species were used.

Genomic library screening

A *M. acuminata* cv. Grand Nain (AAA) genomic library was constructed in the λFIX cloning vector (Stratagene, La Jolla, Calif.) according to the manufacturer's instructions. Using heterologous probes, a number of genomic isolates were characterized, including clone T46, which was determined to contain sequences homologous to previously characterized retroelements from phytopathogenic fungi and plants using the BLASTX search algorithm to search GenBank (Gish and States 1993). Similar results were obtained using the BLASTN search algorithm (Altschul et al. 1990).

Mapping, subcloning, sequencing

T46 contained an insert of approximately 25 kb. The region around the putative retroelement was mapped with restriction endonucleases. Three subclones were generated in pBluescript (Stratagene): containing a 1.5-kb and a 1.8-kb *Sal*I fragment, and a 4.0-kb *Eco*RI fragment, respectively (see Fig. 1). Together these subclones span the region containing the retroelement homology.

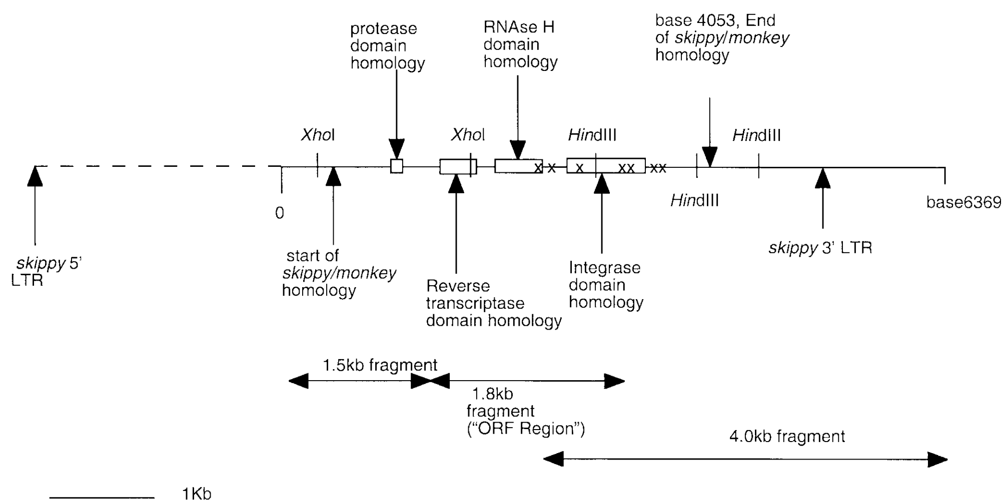


Fig. 1 A schematic representation of the characterized fragment containing the retroelement *monkey*. The gene structure of the banana element is compared to the *skippy* retrotransposon from *F. oxysporum* f sp. *lycopersicii*. The 6.369 kb of *monkey* sequence is represented as a *solid line*, while regions homologous to functional domains are represented as *open boxes*. The *dotted line* indicates the extent of *skippy* sequence for which no corresponding *monkey* sequence has been determined. The limits of *skippy*/*monkey* homology are indicated. The relative positions of the subcloned 1.5- and 1.8-kb *Sal*I and the 4.0-kb *Eco*RI fragments from the original T46 lambda clone used as probes for the genomic Southern analysis and FISH are indicated *below* the line as *double-ended arrows*. The recognition sequences for the restriction enzymes *Hin*dIII and *Xho*I are also indicated. The approximate positions of stop codons within the major ORF are indicated (x)

Plasmid DNA for sequencing was prepared using a midi-prep kit (Qiagen, Valencia, Calif.). Sequencing was performed following the manufacturer's recommendations using the BigDye terminator cycle sequencing kit and an ABI 373A automated sequencer (PE Applied Biosystems, Foster City, Calif.).

Isolation of genomic DNA

Genomic DNA was isolated from banana leaves, either from greenhouse-grown plants or from green shoots propagated in vitro, using a cetyl trimethylammonium bromide (CTAB) isolation protocol. Leaves or shoots were ground to a fine powder in liquid nitrogen using a mortar and pestle. The frozen powder was added to 2× CTAB buffer [100 mM TRIS-HCl pH 8.0, 1.4 M NaCl, 20 mM EDTA, 2% (w/v) CTAB and 1 µl/ml *β*-mercaptoethanol] at a ratio of approximately 5 ml buffer per gram (fresh weight) of tissue. The slurry was incubated at 65 °C for 1 h, then extracted with one volume of chloroform. After the phases were separated by centrifugation (3000 ×*g* for 10 min at room temperature), DNA was precipitated from the aqueous phase by the addition of one volume of isopropanol. The nucleic acids were either spooled out using a glass rod or pelleted by centrifugation (3000 ×*g*, 10 min at 4 °C). The DNA was air dried and resuspended in 2 ml of TE; RNase A was added to a final concentration of 10 mg/ml and the mixture was incubated at 37 °C for at least 30 min. The DNA was extracted gently with one volume of phenol:chloroform (1:1), then with one volume of chloroform, and precipitated by the addition of 0.1 volume of 3 M sodium acetate (pH 4.5) and two volumes of ethanol, again spooling the DNA out if possible or pelleting by centrifugation. Finally the DNA was resuspended in TE and stored at 4 °C.

Southern hybridization

For genomic Southern analysis, 20 µg aliquots of DNA were digested with restriction endonucleases, fractionated on a 1% agarose gel, and transferred to Nytran Plus membrane (Schleicher and Schuell, Keene, N.H.) as described in Sambrook et al. (1989). Probes (~50 ng) were labeled with [^{32}P]dCTP (NEN, Boston, Mass.) using the Rediprime II kit (Amersham, Piscataway, N.J.). Hybridization conditions were as described by Church and Gilbert (1984). Slot blots were prepared using an Evenflow capillary slot blot apparatus (Laboratory Product Supplies, Rochester, N.Y.) and Nytran Plus membrane (Schleicher and Schuell, Keene, N.H.), following the manufacturer's instructions. Hybridization conditions were as described by Church and Gilbert (1984). High-stringency washes were carried out at 65 °C in 1% SDS, 40 mM sodium phosphate buffer pH 7, 1 mM EDTA.

Chromosome preparations

Metaphase spreads were prepared as described by Doležel et al. (1998). Root tips were collected in 50 mM phosphate buffer (pH 7.0) containing 0.2% *β*-mercaptoethanol. The roots were fixed in 3:1 ethanol:acetic acid after a pre-treatment in 0.05% 8-hydroxyquinoline for 3 h. Fixed roots were washed in 75 mM KCl-7.5 mM EDTA (pH 4). Meristem tips were digested in an enzyme mixture (2% pectinase and 2% cellulase) for 90 min at 30 °C. The protoplast suspension was then filtered through a 150 µm nylon mesh and pelleted. The pellet was resuspended in 75 mM KCl-7.5 mM EDTA (pH 4) and incubated for 5 min at room temperature. After pelleting, the protoplasts were washed three times with 70% ethanol, and 5 µl of the suspension was applied to a slide. Shortly before the drop dried, 5 µl of 3:1 fixative was added to induce protoplast bursting. Finally, the slide was rinsed in 100% ethanol and air-dried.

In situ hybridization

Localization of DNA probes was performed according to Doleželová et al. (1998). Plasmids containing 1.5-kb, 1.8-kb and 4.8-kb fragments of *monkey* and the plasmid VER17 containing the 18S-5.8S-25S rDNA of *Vicia faba* (Yakura and Tanifuji 1983) were labeled either with biotin or digoxigenin, using a Random Primed DNA Labeling Kit (Boehringer Mannheim, Mannheim, Germany). The probe VER17 has been found to be specific for secondary constrictions (NORs) of several species, including *Musa* (Doleželová et al. 1998). Chromosome preparations were denatured in a solution of 70% formamide in 2× SSC for 2 min at 72 °C, and 20 µl of hybridization mixture was applied to each slide. The hybridization mixture consisted of 50% formamide, 10% dextran sulfate in 2× SSC, 250 µg/ml sheared calf thymus DNA, and 2 µg/ml labeled probe. The mixture was denatured at 76 °C for 15 min shortly before use. Hybridization was carried out at 37 °C overnight. After removing the cover slips, slides were subjected to high-stringency washes. High-stringency washes were performed at 42 °C in 0.1× SSC and 35% formamide, allowing only target sequences that are more than 95% homologous to remain hybridized. The sites of digoxigenin-labeled probe hybridization were detected using anti-digoxigenin-FITC, and the signal was amplified using anti-sheep-FITC. Sites of biotin-labeled probe hybridization were detected using avidin-Cy3 and the signal was amplified using biotinylated anti-avidin and avidin-Cy3. Chromosome preparations were counterstained with 0.2 µg/ml DAPI (4′,6-diamidino-2-phenylindole) and mounted in Vectashield antifade solution (Vector Laboratories, Burlingame, Calif.).

Fluorescence microscopy

The slides were examined with an Olympus BX 60 epifluorescence microscope and the images of DAPI, FITC and Cy3 fluorescence were acquired separately with a black and white CCD camera. The camera was interfaced to a PC running the ISIS software (Metasystems, Belmont, Calif.). Image processing consisted exclusively of signal intensity, contrast, and background adjustments that affected the whole image.

Results

Sequence of *monkey*

A ~25 kb lambda clone insert from a *M. acuminata* genomic library was sub-cloned and a contiguous region of 6369 bp was entirely sequenced (GenBank Accession No. AF143332). BLAST analysis (Altschul et al. 1990; Gish and States 1993) and the deduced gene structure within the retroelement revealed homology to the *gypsy* class of LTR retrotransposons. We named this banana retrotransposon homolog *monkey*. The highest homologies were found to sequences from various unrelated plant and fungal species (see Table 1), including the plants pineapple, tomato, *Arabidopsis* and lily, and the plant pathogenic fungi *Fusarium oxysporum*, *Cladosporium fulvum* and *Magnaportha grisea*.

The most extensive homology identified was to the LTR retrotransposon *skippy* from *F. oxysporum*. (Anaya and Roncero 1995). Homology ranging between 25 and 48% identity at the amino acid level was identified between *monkey* and *skippy* in the reverse transcriptase, RNase H and integrase domains. Furthermore, *monkey* and *skippy* were essentially colinear along their ~3.5 kb length of homology (see Fig. 1). This colinearity, and comparison with the *skippy* sequence, also suggested that the original *λ* clone did not contain the entire *monkey* sequence but was missing ~2 kb of 5′ sequence,

Table 1 Sequences in Genbank that display the highest homology to the entire 6.369-kb fragment of the banana retroelement *monkey*

Description[a]	Species	Smallest sum probability	Genbank Accession No.
1. Polyprotein	*Lycopersicon esculentum*	7.0 e^{-181}	AF119040
2. Polyprotein	*Ananas comosus*	4.5 e^{-162}	Y12432
3. Putative reverse transcriptase	*Arabidopsis thaliana*	9.3 e^{-162}	AC005398
4. Endonuclease	*Magnaporthe grisea*	9.8 e^{-155}	L35053
5. Putative reverse transcriptase	*A. thaliana*	2.3 e^{-147}	AF077407
6. *skippy* retrotransposon	*Fusarium oxysporum* f. sp. *lycopersicae*	2.1 e^{-146}	L34658
7. Pol polyprotien	*Fugu rubripes*	6.0 e^{-139}	AF030881
8. Retrotransposon *del1-46*	*Lilium henryi*	8.8 e^{-139}	prf 1510387A
9. *Cft-1* retrotransposon	*Cladosporium fulvum*	3.0 e^{-126}	Z11866
10. Putative POL3 protein	*A. thaliana*	9.0 e^{-124}	AC005561

[a] The search was performed using the BLASTX search algorithm (Gish and States 1993). The putative gene (product) identity, species of origin, and the GenBank Accession No. are listed, together with the probability of the match occurring by chance (Altschul et al. 1990; Gish and States 1993). In all cases the homologies were extensive – greater than 30% amino acid sequence identity over at least 2 kb of nucleotide sequence

including the 5′ LTR. No significant homologies were found in the database to bases 1–670 or 4053–6369 of the *monkey* sequence.

Analysis of the putative ORFs in *monkey* showed that several stop codons were present in what should be the major ORF of this element (Fig. 1). This particular copy of *monkey* is therefore a non-functional element, though theoretically it might be activated in *trans* via the expression of other functional copies of *monkey* present in the genome.

Copy number of *monkey* in the *M. acuminata* genome

A common feature of retrotransposons is their extremely high copy number in host genomes. Southern hybridization analysis of *M. acuminata* DNA using a probe from the major ORF of the *monkey* sequence revealed that *monkey* is indeed present in multiple copies in the *M. acuminata* genome (Fig. 2). Regardless of the restriction enzyme used, there were always several clearly distinguishable bands against a background of relatively much weaker hybridization running the entire length of the lane. These results suggest that there are groups of highly conserved, highly repeated *monkey*-like sequences in the *Musa* genome, but that there is also considerable sequence variation amongst *monkey* elements. Rapid sequence divergence is another common feature of plant retrotransposons (Kunze et al. 1997).

Slot-blot analysis under high stringency was carried out to estimate the copy number of *monkey* in the *M. acuminata* genome. The strength of the hybridization signal of a *monkey* probe (encompassing most of the major ORF; see Fig. 1) obtained with a known number of *M. acuminata* genomes was compared the intensity of hybridization of the same probe to a known number of plasmids containing the probe sequence. It was determined that *monkey* was present in approximately 200–500 copies per haploid genome (data not shown), corresponding to between 600 and 1500 copies per triploid nucleus. Assuming that most of the hybridizing copies of *monkey* were full length (i.e. about 6 kb),

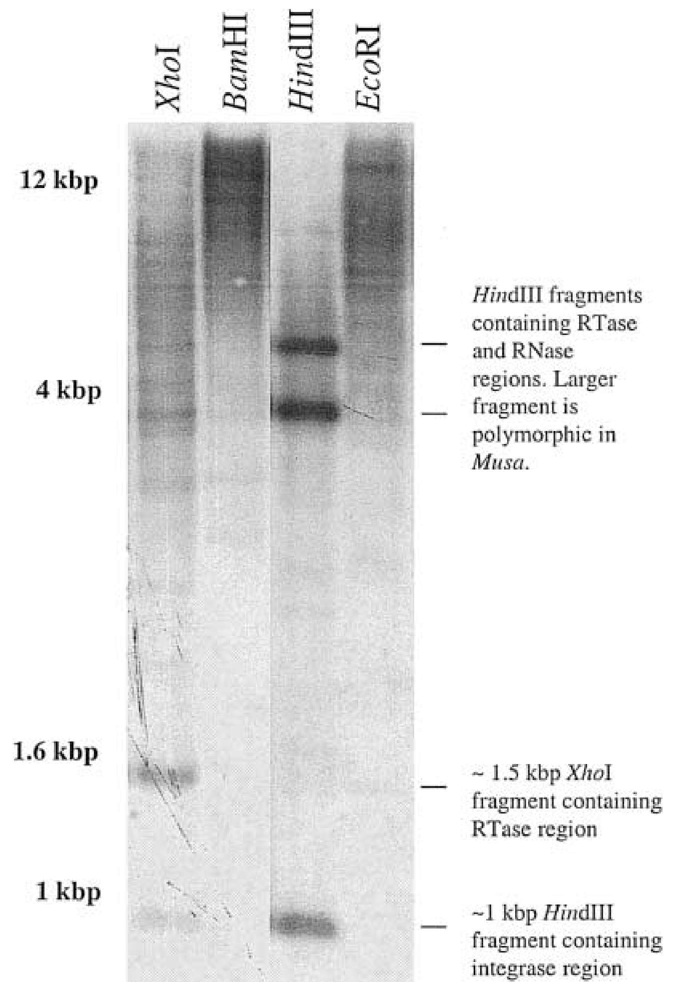


Fig. 2 Genomic Southern analysis. Aliquots (20 µg) of *M. acuminata* genomic DNA were digested with the restriction enzymes indicated and probed with a labeled 1.8-kb *Sal*I subcloned fragment encompassing the majority of the putative major ORF (see Fig. 1), which includes the RTase, RNase H and integrase domains. The fragment used as a probe contains internal restriction sites for *Xho*I, and *Hin*dIII as indicated in the *inset*. The positions of molecular weight markers are indicated on the *left*

copies of *monkey* constitute about 1.2–3 Mb per haploid genome or about 0.2–0.5% of the *M. acuminata* genome.

monkey is ubiquitous and polymorphic in *Musa* genomes

The genomes of cultivated bananas and plantains are generally triploid. While banana genomes tend to consist of three A genomes from *M. acuminata* (AAA genomes), plantain genomes (AAB or ABB) often contain one or more B genomes from the closely related *M. balbisiana* (Osuji et al. 1997). DNA was isolated from various banana and plantain cultivars representing both A and B genomes, as well as from an ornamental banana species (*M. vellutina*; non-A, non-B genome), and subjected to Southern hybridization analysis using the major ORF region of *monkey* as a probe. The *monkey* probe hybridized to genomic DNA from all the cultivars and species used, indicating that *monkey* was introduced into *Musa* prior to the divergence of *M. acuminata, M. balbisiana* and *M. vellutina* (Fig. 3).

There appears to be a close association between a *Hin*dIII fragment that hybridizes to the *monkey* major ORF probe and the B genome, while another *monkey* fragment is associated with the A and B genomes but not with the *M. vellutina* genome (Fig. 3). The hybridization pattern for one cultivar, cv. Chakrakeli (AAB), appears to be inconsistent with this hypothesis, as it lacks the putative B-specific fragment. It is possible that cv. Chakrakeli is actually of AAA constitution. This possibility is made more likely by the observation that hybridization patterns with an unrelated genomic probe reveal a similar set of A and B genome-associated restriction fragments, with cv. Chakrakeli again being the only inconsistency (P. Balint-Kurti, unpublished results).

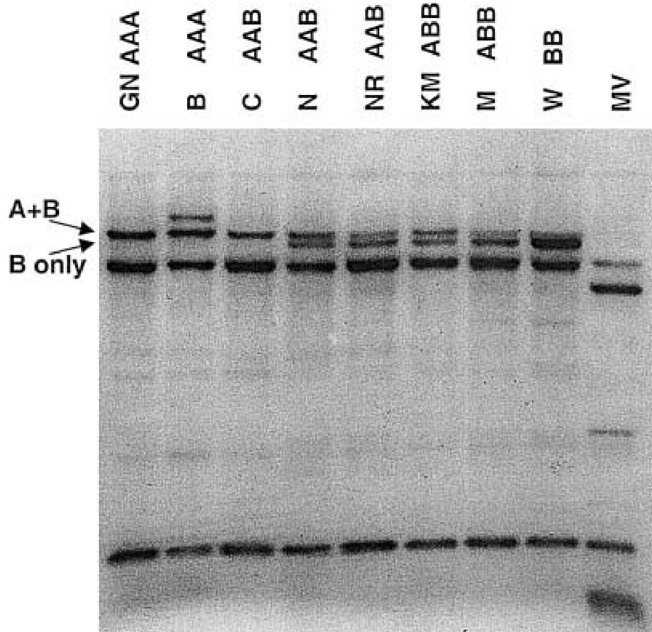


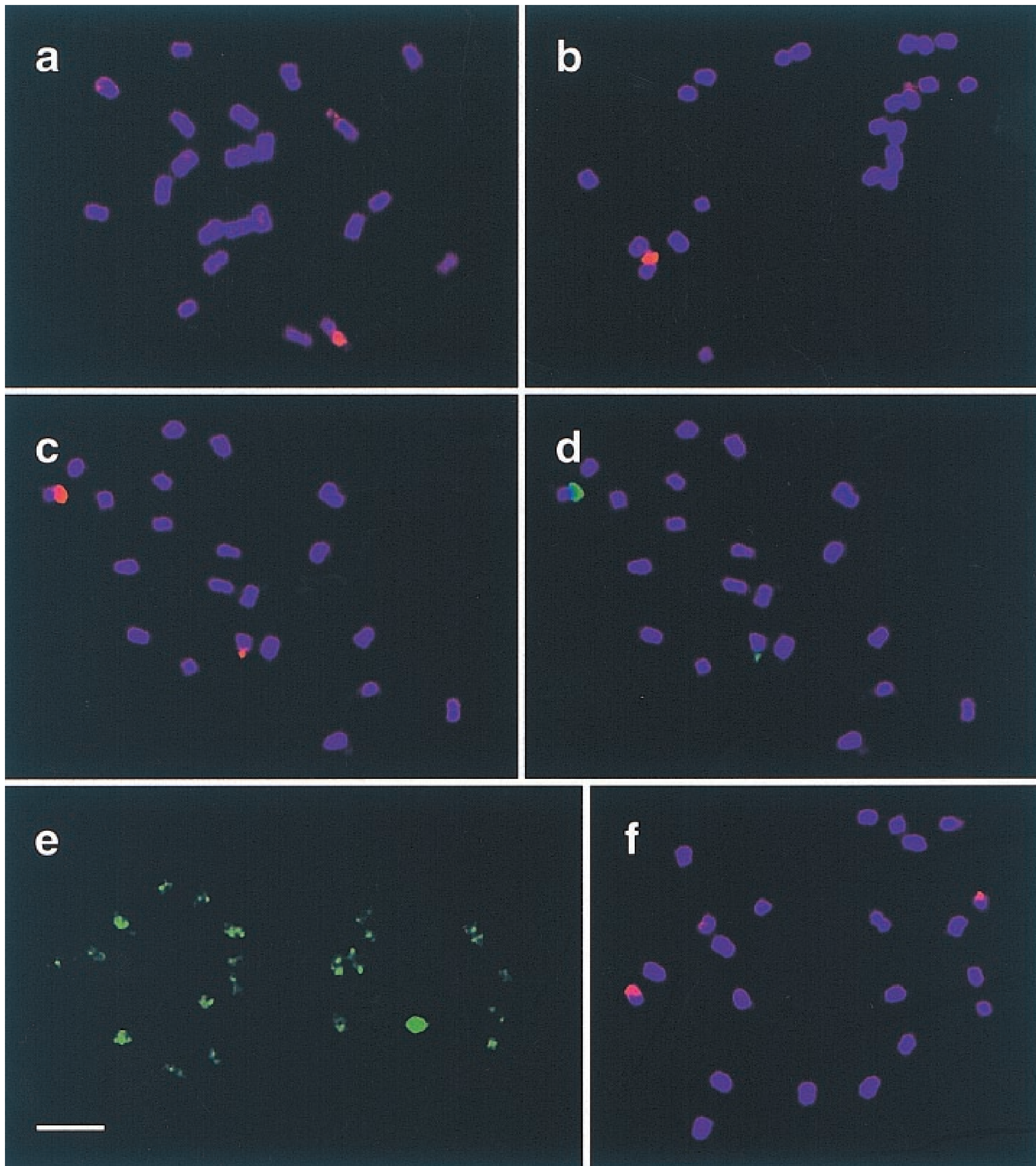
Fig. 3 Southern analysis of genomic DNAs from nine different banana cultivars, digested with *Hin*dIII and probed with the 1.8-kb fragment of the major ORF (see Fig. 1). GN, cv. Grand Nain (genomic composition AAA); B, cv. Basrai (AAA); C, cv. Chakrakeli (AAB); N, cv. Nendran (AAB); NR, cv. Rasthali (AAB); KM, cv. Karibali Monthan (also known as Bluggoe, plantain, ABB); M, cv. Monthan (plantain, ABB); W, *M. balbisiana* (BB); MV, *M. vellutina* . The polymorphic A and B genome-associated fragments are indicated by the *arrows*

Fig. 4a–f Localization of *monkey* retrotransposon components on chromosomes of diploid ($2n = 2x = 22$) *Musa* species using fluorescence in situ hybridization: **a**, **c–f** *M. acuminata* cv. Niyarma Yik (ITC 0269). **b** *M. balbisiana* cv. Singapuri (ITC 0248). Chromosomes were counterstained with DAPI (*blue fluorescence*). In **a** and **b** the 4.0-kb subclone of *monkey* was labeled with biotin and detected using Cy3 (*red signals*). In both *Musa* species, the probe hybridized preferentially to nucleolus organizer regions (NORs) on one pair of chromosomes. Note that the homologs differ in signal strength, indicating differences in the number of *monkey* copies. **c, d** Double in situ hybridization using a biotin-labeled probe for 18S-5.8S-25S rDNA genes and a digoxigenin-labeled 4.0-kb subclone of *monkey*. In **c** the probe for rDNA genes hybridized strongly to NOR regions on one pair of chromosomes (*red signals* obtained after detection with Cy3). **d** The probe for *monkey* hybridized to the same sites (*green signals* obtained after detection with FITC), confirming the preferential localization of *monkey* to NORs. Note that the signal strength correlates with the amount of hybridized rDNA probe. **e** Distribution of sequences homologous to the 4.0-kb subclone of *monkey* observed after signal amplification (the probe was labeled with digoxigenin and detected using FITC). Note that the signal appears to be homogeneously distributed, apart from preferential binding to NOR. **f** The 1.8-kb subclone of *monkey* was labeled with biotin and detected using Cy3 (*red signals*). The probe hybridized preferentially to NORs on one pair of chromosomes. *Bar* 5 μm

Genomic DNAs from several other plant and fungal species, including tomato, strawberry, pea, *Mycosphaerella fijiensis* (the causative agent of black Sigatoka disease of banana) and *F. oxysporum* f. sp. *cubense* (the causative agent of Panama disease of banana), were subjected to Southern analysis using the *monkey* major ORF region as a probe under a range of different stringencies. No significant hybridization was detected (data not shown).

FISH analysis of *monkey* localization

The three subclones of the *monkey* element (1.5 kb, 1.8 kb and 4.0 kb; see Fig. 1) were used as probes for FISH on *M.* chromosomes. The largest clone (4.0 kb) hybridized strongly to the secondary constriction of a pair of NOR-bearing chromosomes in diploid *M. acuminata* and *M. balbisiana* (Fig. 4a, b). Localization to the NOR region was confirmed by a dual-color FISH analysis in which the probes for *monkey* and for 18S-5.8S-25S rRNA genes were localized simultaneously (Fig. 4c, d). At high signal amplification, minor sites of *monkey* were detected on all chromosomes (Fig. 4e). These sites were dispersed through the whole genome and no clear pattern of distribution could be established. The medium-sized clone (1.8 kb) showed virtually the same pattern of distribution in both *Musa* species as the 4.0 kb clone



(Fig. 4f). Clear differences in signal intensity between homologs were observed after FISH with the probe for rRNA genes. The same differences were observed with the probes for *monkey* (clones 1.8 kb and 4.0 kb). Although these differences were estimated only semi-quantitatively, the signal intensities of the two were clearly correlated.

In contrast to the 1.8 kb and 4.0 kb clones, only weak signals were observed with the smallest (1.5 kb) clone. These signals were observed only at high signal amplification and no clear pattern of chromosome distribution could be established. Furthermore, no labeling was found at NOR loci. The sequence of the 1.5 kb clone shows none or very low homology with known retrotransposon sequences over most of its length. This clone may carry a large amount of sequence not shared with most of the other copies of *monkey* in the genome. This would explain the weak hybridization observed.

Discussion

We report the identification and initial characterization of a *gypsy*-class LTR retrotransposon element from banana. The sequence of *monkey* exhibits strong homology to the reverse transcriptase, RNase H and integrase genes of *gypsy*-like retroelements identified in fungi and plants. *gypsy* retrotransposons were recently shown to exist in a wide variety of different plant species (Suoniemi et al. 1998); however, the elements *monkey*

most closely resembles are from plant pathogenic fungi and unrelated plant species (Table 1).

The sequence with the most extensive homology to *monkey* is the *skippy* retrotransposon from *F. oxysporum* f. sp. *lycopersicii*. *F. oxysporum* f. sp. *cubense*, which is associated with Panama disease in banana, is an extremely important banana pathogen with complex evolutionary origins (O'Donnell et al. 1998). The observation that *monkey* shows highest sequence homology to retroelements from plants and fungi unrelated to banana suggests that *monkey* was introduced into *Musa* spp. by horizontal transfer of DNA from another species, possibly from a *F. oxysporum* pathogen during the intimate pathogenic interaction. The fact that retroelements share so many similarities with infectious RNA viruses makes this possibility all the more compelling. In this case we have no direct evidence for horizontal transfer of DNA between plant and fungal or viral pathogens besides the highly suggestive sequence similarities; however, movement of retrotransposons between species has been documented (Jordan et al. 1999).

Southern hybridization analysis indicated that *monkey* was present in multiple copies in both the A and B genomes of *Musa*, and also in a more distantly related species, *M. vellutina* (Figs. 2 and 3). Furthermore, a fragment corresponding to the major ORF of *monkey* is polymorphic in *Musa* species, notably between cultivars that carry or lack B genomes (Fig. 3). Such analysis may be useful in determining the genetic composition of poorly characterized cultivars. While it is clear that the *monkey* sequence was first introduced into the *Musa* genome prior to the divergence of *M. acuminata* and *M. balbisiana*, there remain highly repeated sequence variants of *monkey* that are specific to each of the species. Estimates from slot-blot analysis suggested that 0.2–0.5% of the *M. acuminata* genome is made up of *monkey*-like sequences, suggesting that *monkey* may have had a substantial impact on the evolution of the cultivated banana.

FISH with the probes for *monkey* confirmed that *monkey* is present in multiple copies and showed that these copies are preferentially localized within the NOR regions of *Musa* chromosomes. This is in contrast to many observations reported to date; plant retroelements that have been physically mapped exhibit relatively uniform distributions over all chromosomes, with the exception of centromeres, telomeres or the NORs. This is the case for *Ty1/copia*-like elements in *Avena* (Katsiotis et al. 1996), *Beta vulgaris* (Schmidt et al. 1995), *Pinus elliottii* (Kamm et al. 1996) and *Allium cepa* (Pearce et al. 1996). A similar distribution was found also for the *copia*-like retroelement *BARE-1* in *Hordeum vulgare* (Suoniemi et al. 1996) and the *Ty3/gypsy*-like element in *Lilium henryi* (Sentry and Smyth 1985). However, the concept of dispersed localization of retroelements within plant genomes may be misleading, as only a few elements and few species have been studied so far. For instance, Brandes et al. (1997) found clustering of *Ty1/copia*-like elements in AT-rich regions at the centromeres of *Cicer arietinum* and *Arabidopsis thaliana*.

In a recent report, Presting et al. (1998) described preferential localization of a *Ty3/gypsy* retrotransposon-like sequence named *cereba* to centromeric regions of cereal chromosomes. Our results add a second chromosome domain to the list of sites of preferential retroelement accumulation in plants. Brandes et al. (1997) speculate that in the case of species with small genomes that contain only a small proportion of repetitive DNA sequences (e.g. *A. thaliana*), retrotransposons have a limited number of sites at which they may insert without disrupting genes that confer a selective advantage. However, while centromeric repeats represent genetically silent heterochromatin, the NOR contains many copies of rRNA genes. It is thus not clear how the *monkey* sequences are organized within the rDNA repeats. It should be noted that, although our data suggest preferential localization of *monkey* to NOR regions of *Musa*, dispersed minor sites observed at high signal amplification may represent single elements inserted along the chromosome arms. However, we cannot exclude the possibility that these sites represent background due to non-specific binding of probes and not additional copies of the retroelement.

The differences in the size of the secondary constriction and in the intensity of signals obtained after FISH with the probe for 18S-5.8S-25S rRNA genes have been described previously (Doleželová et al. 1998). These differences were observed also in this study. It is interesting to note that the same differences were observed with the *monkey* element. The localization of many copies of *monkey* in NORs, and the correlation between the number of copies of rRNA genes and the number of *monkey* retroelements, suggests the simultaneous evolution and amplification (e.g. by unequal recombination) of *monkey* and rRNA genes. It is noteworthy that similar localization patterns are observed in the A and B genomes. This suggests that either this localization was established prior to the divergence of the genomes or that *monkey* is preferentially inserted in the NOR.

Several reports indicate that expression of plant retrotransposons is induced during tissue culture (for review see Kunze et al. 1997). It has been proposed that this activation and the subsequent insertion of new retrotransposon copies in the genome can account for the phenomenon of somaclonal variation seen in plants propagated via tissue culture (Pouteau et al. 1991). Somaclonal variation is a particularly important concern in commercial banana propagation; however, we were not able to detect, by northern analysis, expression of *monkey* in tissue culture cells or in mature plants. It is possible that further work, using RNA isolated from plants that have been subjected to various kinds of stress (e.g. UV light, pathogenesis) and using more sensitive assays, will help to characterize *monkey* expression.

Acknowledgments We thank I. Van den houwe, curator of the INIBAP Transit Centre at the Katholieke Universiteit, Leuven (Belgium) for the supply of plant material. We also thank Professor S. Tanifuji and Dr. I. Schubert for providing the VER17 rDNA probe. We are grateful to Ms. R. Tuzková and Ms. J. Weiserová

for excellent technical assistance, and to J.E. Keen for valuable creative input. Part of this study was undertaken as a part of the Global Programme for *Musa* Improvement (PRO*MUSA*) and was supported by a Research Contract No. 8145/RB from the International Atomic Energy Agency, Vienna. This research was also supported in part by the Boyce Thompson Institute and in part by the Banana Improvement Program, sponsored by the Common Fund for Commodities, the FOA Inter-Governmental Group on Bananas (FOA/IGB) and the World Bank. All experiments reported here comply with current New York state and US Federal law.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410

Anaya N, Roncero MIG (1995) *skippy*, a retrotransposon from the fungal plant pathogen *Fusarium oxysporum*. Mol Gen Genet 249: 637–647

Baurens FC, Noyer JL, Lanaud C, Lagoda PJL (1997) A repetitive sequence family of banana (*Musa spp*) shows homology to *Copia*-like elements. J Genet Breed 51: 135–142

Brandes A, Heslop-Harrison JS, Kamm A, Kubis S, Doudrick RL, Schmidt T (1997) Comparative analysis of the chromosomal and genomic organization of *Ty*1-*copia*-like retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant Mol Biol 33: 11–21

Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci USA 81: 1991–1995

Doležel J, Doleželová M, Roux N, Van den houwe I (1998) A novel method to prepare slides for high resolution chromosome studies in *Musa spp*. Infomusa 7: 3–4

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

Gish W, States DJ (1993) Identification of protein coding regions by database similarity search. Nat Genet 3: 266–272

Grandbastien M-A (1998) Activation of plant retrotransposons under stress conditions. Trends Plant Sci 3: 181–187

Harper G, Osuji JO, Heslop-Harrison JS, Hull R (1999) Integration of Banana streak badnavirus into the *Musa* genome: molecular and cytogenetic evidence. Virology 255: 207–213

Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93: 7783–7788

Jordan IK, Matyunina LV, McDonald JF (1999) Evidence for the recent horizontal transfer of long terminal repeat transposon. Proc Natl Acad Sci USA 96: 12621–12625

Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of *Ty*1-*copia*-like sequences as a component of large genomes in *Pinus elliottii* var *elliottii* and other gymnosperms. Proc Natl Acad Sci USA 93: 2708–2713

Katsiotis A, Schmidt T, Heslop-Harrison JS (1996) Chromosomal and genomic organization of *Ty*1-*copia*-like retrotransposon sequences in the genus *Avena*. Genome 39: 410–417

Kossack DS, Kinlaw CS (1999) IFG, a gypsy-like retrotransposon in *Pinus* (Pinaceae), has an extensive history in pines. Plant Mol Biol 39: 417–426

Kunze R, Saedler H, Lönnig W-E (1997) Plant transposable elements. Adv Bot Res 27: 332–469

Lysak MA, Doleželova M, Horry JP, Swennen R, Doležel J (2000) Flow cytometric analysis of nuclear DNA content in *Musa*. Theor Appl Genet, in press

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

O'Donnell K, Kistler HC, Cigelnik E, Ploetz RC (1998) Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial genomes. Proc Natl Acad Sci USA 95: 2044–2049

Osuji JO, Harrison G, Crouch J, Heslop-Harrison JS (1997) Identification of the genomic constitution on *Musa* L lines (bananas, plantains and hybrids) using molecular cytogenetics. Ann Bot 80: 787–793

Pearce SR, Pich U, Harrison G, Flavell AJ, Heslop-Harrison JS, Schubert I, Kumar A (1996) The Ty1-*copia* group retrotransposons of *Allium cepa* are distributed throughout the chromosomes but are enriched in the terminal heterochromatin. Chromosome Res 5: 357–364

Pouteau S, Huttner E, Grandbastien MA, Caboche M (1991) Specific expression of the tobacco *Tnt*1 retrotransposon in protoplasts. EMBO J 7: 1911–1918

Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A *TY3/gypsy* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J 16: 721–728

Rrichert-Pöggler KR, Shepherd RJ (1997) Petunia vein-clearing virus A plant pararetrovirus with core sequences for an integrase function. Virology 236: 137–146

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual (2nd edn). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

SanMiguel P, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765–768

Schmidt T, Kubis S, Heslop-Harrison JS (1995) Analysis and chromosomal localization of retrotransposons in sugar beet (*Beta vulgaris* L.): LINEs and *Ty*1-*copia*-like elements as major components of the genome. Chromosome Res 3: 335–345

Sentry JW, Smyth DR (1985) A family of repeated sequences dispersed through the genome of *Lilium henryi*. Chromosoma 92: 149–155

Smyth DR, Kalitsis P, Joseph JL, Sentry JW (1989) Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the gypsy group of *Drosophila*. Proc Natl Acad Sci USA 86: 5015–5019

Su PY, Brown TA (1997) Ty3/gypsy-loke retrotransposon sequences in tomato. Plasmid 38: 148–157

Suoniemi A, Anamthawat-Jónsson K, Arna T, Schulman AH (1996) Retrotransposon *BARE*-1 is a major, dispersed component of the barley (*Hordeum vulgare* L) genome. Plant Mol Biol 30: 1321–1329

Suoniemi A, Tanskanen J, Schulman AH (1998) *Gypsy*-like retrotransposons are widespread in the plant kingdom. Plant J 13: 699–705

Tutois S, Cloix C, Cuvillier C, Espagnol MC, Lafleuriel J, Picard G, Tourmente S (1999) Structural analysis and physical mapping of a pericentromeric region of chromosome 5 of *Arabidopsis thaliana*. Chromosome Res 7: 143–156

Yakura K, Tanifuji S (1983) Molecular cloning and restriction analysis of *Eco*RI fragments of *Vicia faba* ribosomal DNA. Plant Cell Physiol 24: 1327–1330