Theor Appl Genet (1997) 95:922–931

F.-C. Baurens · J.-L. Noyer
C. Lanaud · P. J. L. Lagoda

Assessment of a species-specific element (Brep 1) in banana

Received: 17 January 1997 / Accepted: 7 March 1997

Abstract The nuclear genome of wild-type banana accessions was investigated for repetitive elements. We report here the occurrence, in the banana genome, of a sequence family of species-specific repetitive elements: Brep 1. This sequence family is distributed throughout the *Musaceae* with various copy numbers. The two species *Musa acuminata* and *M. schizocarpa* carry the highest copy numbers in contrast to *M. balbisiana* and tested representatives of different other sections. PCR primers were defined in the core consensus sequence for specific amplifications, which allow representatives of this sequence family to be easily detected in wild and cultivated banana clones. Sequence data were analysed and hypotheses on the evolution of banana cultivars from the wild-type banana clones are discussed.

Key words Repetitive DNA · Banana · *Musa* · Genetic relationships · Species-specific sequences

Introduction

Banana is a monocotyledonous plant of the Musaceae family (order Zingiberale). The early taxo-morphologic classification of banana (Cheesman 1947; Simmonds and Shepherd 1955; Simmonds 1962) is now adopted worldwide with only minor corrections and additions. The wild diploid bananas contain about 30 *Musa* species grouped into four sections: *Australimusa*, *Callimusa*, *Eumusa* and *Rhodochlamys*. The *Eumusa* section contains mainly the *Musa acuminata* species, whose great phenotypic diversity allows the subdivision into distinct subspecies, the *Musa schizocarpa* species (related to the *M. acuminata* species complex) and the *Musa balbisiana* species. Next to nothing is known about evolution of the *Musa* because of a lack of fossils. Cultivated clones are diploids or polyploids, and it has been established that parthenocarpic clones of agronomic interest originated from the Mendelian hybridization of two major species, *M. acuminata* (A genome) and *M. balbisiana* (B genome), belonging to the *Eumusa* section (Simmonds and Shepherd 1955). However, recent advances in diversity studies using molecular markers demonstrate that *M. schizocarpa* (S genome) and representatives from the *Australimusa* section (T genome) are involved in the genomic composition of both the diploid and the polyploid cultivated clones (Carreel 1994). *M. acuminata* is known to be extremely diversified, and several subspecies can be distinguished both morphologically and by molecular markers (Lebot et al. 1993; Jarret et al. 1993; Carreel 1994; Bhat and Jarret 1995). It has been proposed that specific subspecies contributed to different cultivated banana types (Lebot et al. 1993; Carreel 1994). Differences observed between the different subspecies do not only affect the genic sequences. Large structural rearrangements of the genome have also been reported, and specific translocations have been detected by studying crosses between wild *M. acuminata* ssp. by metaphase spreads (Dessauw 1987). Furthermore, the size of the nuclear genome varies (Dolezel et al. 1994) between the different subspecies. This could be partly due to variations in the copy number of nuclear repetitive sequences (Baurens et al. 1996).

Repetitive elements have been used in different plant species as molecular markers for fingerprinting and as a basis for molecular phylogeny (Cai and Bullen 1994; Li et al. 1996). Due to the sequence copy number varying between 100 and several millions, repetitive

Communicated by R. Hagemann

F.-C. Baurens · J.-L. Noyer · C. Lanaud · P. J. L. Lagoda (✉)
CIRAD (Centre de coopération Internationale en Recherche Agronomique pour le Développement)-GERDAT (Département de Gestion, Recherche, Documentation et Appui Technique), AGETROP (Analyse du Génome des Espèces Tropicales) Laboratory, avenue du val de Montferrand, BP 5035, 34032 Montpellier Cedex 01, France

sequences often form the greatest proportion of plant genomes, the exact extent depending on the animal or plant species considered. The use of site-specific primers anchored in nuclear repetitive DNA allows rapid and efficient access to the nuclear genome. Considering the exponential nature of polymerase chain reaction (PCR) amplification, only very small amounts of DNA template are required compared to restriction fragment length polymorphisms (RFLPs) or other hybridization-based methods. Moreover, the relatively low size of the amplicon, generally comprised between 100 bp and a few kilobases, enables the use of poor quality or even sheared or partially degraded DNA.

Repeated sequences have been successfully used for typing banana clones with hybridization-based methods (Kaemmer et al. 1992) and using decameric PCR primers located in these elements (Jarret et al. 1993). The purpose of the study described here was to identify suitable repetitive elements and to assess the potentials of site-specific PCR-based markers with respect to molecular breeding strategies in banana improvement programmes, especially for distinguishing the individual genotypes involved in the cultivars' genome complex using PCR tests. *Musa acuminata* and *M. schizocapa* genomic libraries were screened for species-specific repetitive elements. These elements were used to determine site-specific PCR primers for detecting these sequences in different representatives of the Musaceae family. Sequence data from specific isoforms were subjected to phylogenetic analysis and complex polyploid cultivar genome analysis.

Materials and methods

Plant material and DNA extraction

Total DNA was extracted by the modified CTAB method (Gawel and Jarret 1991; Fauré et al. 1993a). The *Musa acuminata* ssp. *M. balbisiana* and *M. schizocarpa* accessions used in this study (Table 1) were obtained from the French West Indies' (Guadeloupe) collection. DNAs from *Ensete giletii*, *Musa jackeyi*, *M. textilis*, *M. becarrii*, *M. angustigemma* and *M. laterita* were kindly provided by F. Carreel.

Table 1 Wild-type banana clones used in this study

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

Library construction and screening

M. schizocarpa and *M. acuminata Eco*RI genomic DNA libraries were constructed as previously described (Baurens et al. 1996). Inserts were amplified by PCR using the M13 universal primer pair, and amplification products were blotted onto a Hybond® N+ membrane and hybridized with total genomic DNA from 8 wild banana clones (Table 1). Probes with differential hybridization signals, obtained after overnight exposure, were collected and used as RFLP probes on total restricted genomic DNA to confirm specificity. Only probes with a clear species-specific signal were sequenced.

RFLP analysis

RFLP analyses were performed according to Fauré et al. (1993a) with minor modifications (Baurens et al. 1996).

Primer design and PCR amplifications

Primer design and calculation of annealing temperature (Ta) were performed using the OLIGO™ program. PCR was performed on a 50-µl aliquot containing 0.2 µ*M* of each primer and 200 m*M* of dNTP (Pharmacia) in a buffer of 67 m*M* TRIS-HCl pH 8.0, 16.6 m*M* $(NH_4)_2SO_4$, 1.5 m*M* $MgCl_2$ and 1 U of BIOTAQ® (Eurobio). Aliquots of 50 ng of total DNA were used in each reaction as template. PCR was computed on a PTC 100 thermocycler

Table 2 Name and sequence of primers used in this study

Name[a]	Sequence 5′ → 3′	Direction[b]	Ta[c](°C)
AGRP07	ACC CTA TCG TCT GTT GAA TCC	F	54
AGRP08	CCA GTG GCA CAT CAA ACA AA	R	53
AGRP09	TTC AAG ATT TCT GAC CTT TCG	F	54
AGRP54	CTT TTG ACA CCG ATT TTG TAG A	F	53
AGRP55	GCT CCA ATA CCC ATA AGA AGG	R	60
AGRP60	ACT CCC TGT TGA AAA TTC TGA	R	52
AGRP61	AAG TAT TCG GTG TCC AAA ATC	F	52

[a] Expected product sizes on consensus sequence using primer pairs AGRP07-AGRP08, AGRP08-AGRP09 and AGRP54-AGRP55 are, respectively, 200 nt, 365 nt and 451 nt. PCR amplifications were performed using the highest annealing temperature from the primer pair
[b] F, forward primer; R, reverse primer
[c] Ta, annealing temprature in °C

(MJ Research) using the following amplification programme: 94°C for 4 min, 35 × (94°C for 30 s; T for 30 s; 72°C for 45 s), 72°C for 4 min (For Ta values, see Table 2). For each PCR amplification, two repetitions were processed using Ta values for high-stringency specific amplification and an annealing temperature 5°C below the optimum Ta of low-stringency PCR. Half the reaction products were loaded onto a 2% agarose gel, analysed by electrophoresis and visualized by ethidium-bromide staining.

Phylogenetic tree building

Sequence data from the repeated sequence family representatives were computed with PAUP software (Swofford 1993) using an exhaustive search option for minimal tree length and DNAPARS software (PHYLIP software package, Felsenstein 1993). 1000 bootstraps were processed on the consensus tree obtained using default parameters. Genetic distances were obtained by the DNADIST software using Kimura 2 distance parameters.

Results

Identification of species-specific repetitive elements was achieved by differentially screening genomic libraries. Three clones were selected for strong species-specific hybridization signals of similar intensities and sequenced: pMaCIR 11114 (Y08644), pMaCIR 11145 (Y08645), pMaCIR 11161 (Y08646). All of these sequences exhibited a strong hybridization signal with *M. acuminata* and *M. schizocarpa* accessions but not with *M. balbisiana*. We intend to name these sequences Brep 1 (for *B*anana *rep*etitive DNA family no. *1*).

Sequence comparison

Sequence data from these species-specific RFLP probes were compared with the EMBL database content using both BLAST (Altschul et al. 1990) and FASTA (Pearson and Lipman 1988) algorithms. Interestingly, they were found to be related to a previously described *M. acuminata* repetitive sequence (EMBL X91626, Baurens et al. 1996) but do not display substantial homology with any other plant or animal repetitive DNA sequences.

RFLP pattern analysis

A member of this sequence family, pMaCIR 1115 probe, was hybridized on restricted DNA from 8 individuals of wild-type representatives of the *Eumusa* section (Fig. 1). A strong signal was observed for the *M. acuminata* and *M. schizocarpa* genotypes but not for *M. balbisiana*. Analysis of the hybridization pattern, after 2 h of exposure, suggested that the core sequence might be a 500-bp *Eco*RI fragment. In fact, other fragments did appear in the genome with variable intensities after a longer exposure (e.g. 250-bp and 3-kbp

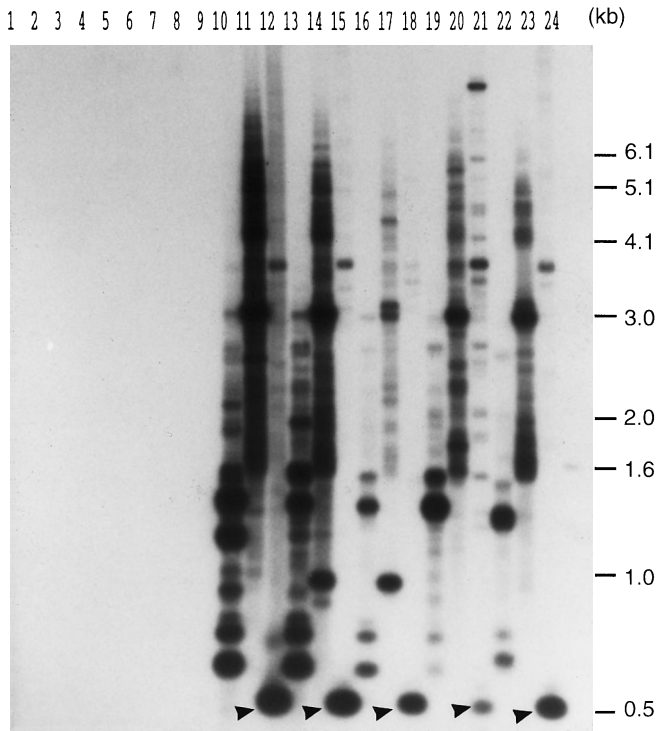

Fig. 1 Southern Blot hybridization of the pMaCIR1115 representative of the Brep 1 family. Overnight exposure of the pMaCIR 1115 probe on total endonuclease-digested DNA of several representatives of the *Eumusa* section. *Lanes 1–3 Musa balbisiana* Butuhan DNA restricted with *Eco*RI, *Hae*III and *Sau*3AI respectively, *lanes 4–6 M. balbisiana* Honduras DNA restricted with *Eco*RI, *Hae*III and *Sau*3AI respectively, *lanes 7–9 M. balbisiana* Pisang Batu DNA restricted with *Eco*RI, *Hae*III and *Sau*3AI respectively, *lanes 10–12 M. acuminata* ssp. *burmannicoides* restricted with *Eco*RI, *Hae*III and *Sau*3AI respectively, *lanes 13–15 M. acuminata* ssp. *malaccensis* restricted with *Eco*RI, *Hae*III and *Sau*3AI respectively, *lanes 16–18 M. acuminata* ssp. *zebrina* restricted with *Eco*RI, *Hae*III and *Sau*3AI respectively, *lanes 19–21 M. acuminata* ssp. *banksii* restricted with *Eco*RI, *Hae*III and *Sau*3AI respectively, *lanes 22–24 M. schizocarpa* restricted with *Eco*RI, *Hae*III and *Sau*3AI respectively. Sizes in kilobases are indicated on the *right*. *Arrowheads* indicate the conserved core *Eco*RI "cassette" of the Brep 1 family revealed after a short-time exposure

*Eco*RI fragments). Analysis of other restrictions with *Hae*III and *Sau*3AI did not exhibit single high-intensity fragments but only fragments with variable intensities, as described for *Eco*RI.

Sequence structure of the Brep 1 family

Sequence alignments of the representatives of this sequence family are shown in Fig. 2. The homologous region of the pMaCIR 11161 clone sequence, representing the first 500 nucleotides, is shown. The nature of the remainder of this sequence is discussed below. All of the representatives can be seen to exhibit homology on the 500-bp *Eco*RI "cassette" except clone pMaCIR 11145 which is shorter due to the presence of an *Eco*RI

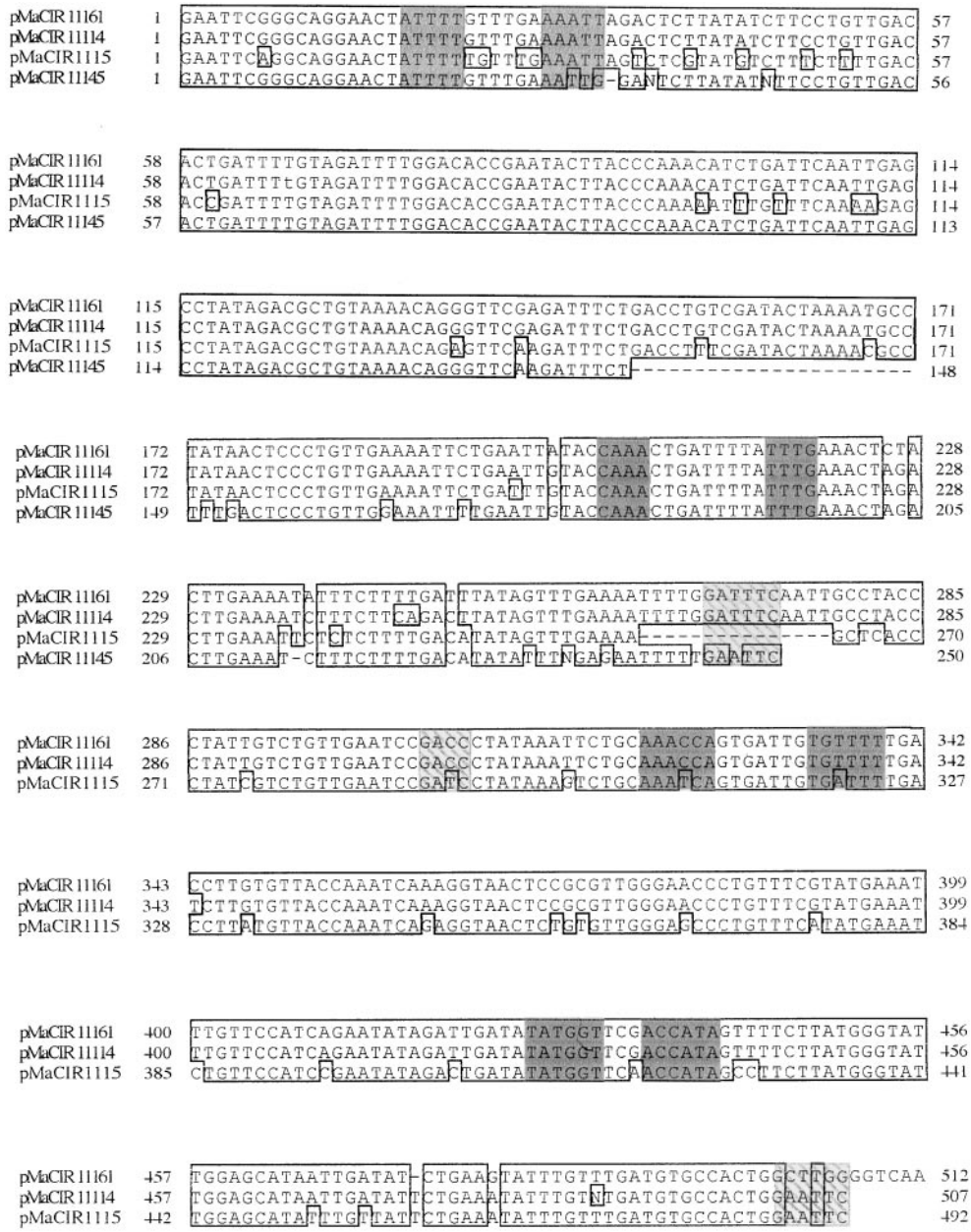

Fig. 2 Sequence alignment of Brep 1 family members. Sequence data are available on EMBL databank under Y08644, Y08645, Y08646 and X91626 accession numbers. *Dashed squares* potential polymorphic enzyme sites, *shaded regions* autocomplementary regions of potential loops, – gap

site that is cryptic or deleted in other representatives (Fig. 2). Another insertion-deletion polymorphism occurs at positions 150–172 on the pMaCIR 11145 probe. Point-mutations are randomly distributed throughout the sequence. Most of the variable regions are between positions 20 and 60, 98 and 111, 235 and 250. Sequence data analysis allowed the detection of mutations at positions responsible for the creation or destruction of restriction endonuclease restriction sites for *Eco*RI and *Sau*3AI (Fig. 2). There was no strong relationship between hot-spot mutation points or gap positions and potential loops in the DNA structure (Fig. 2), suggesting that insertion-excision processes may not be involved in generating polymorphism. A consensus sequence of the *Eco*RI element was built, and PCR primers were defined in the most-conserved regions (Fig. 3). Tentative polymorphic restriction enzyme sites are indicated (squares).

Site-specific PCR amplification

Different primer pairs (Table 2) were used to evaluate the occurrence of the core conserved sequence in different banana genotypes. Results are listed in Table 3. When we used the AGRP 08-AGRP 09 and AGRP 08-AGRP 07 primer pairs, amplification products were detectable in all of the positive genotypes tested by

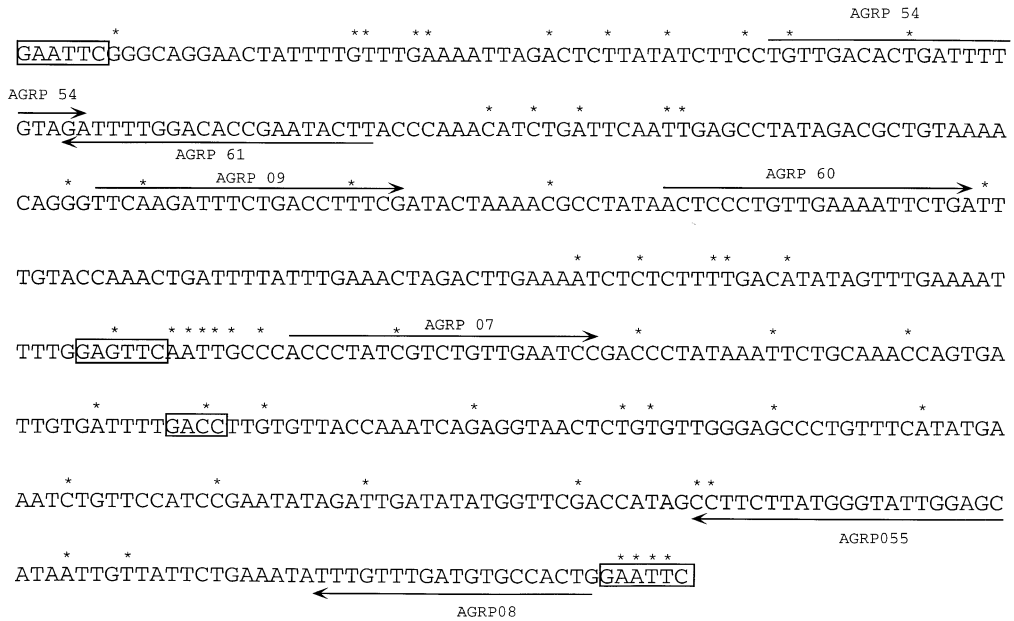

Fig. 3 Consensus sequence of the *Eco*RI core conserved sequence and primer localization. The consensus sequence was obtained by alignment of all the available sequence data on the *Eco*RI core sequence, including PCR products from *M. acuminata* ssp. and *M. Schizocarpa*. *Arrows* indicate primer locations, *squares* indicate polymorphic restriction enzyme recognition sites, * indicates the presence of a nucleotide variation (gap or substitutions) in at least one of the family members

Table 3 PCR detection of the Brep 1 family in the *Musaceae* in comparison with RFLP

Species	Section[a]	RFLP[b]	PCR primer pairs[b]			
			07-08	08-09	54-55	60-61
Ensete giletii	En	−	−	−	nd	nd
Musa jackeyi	A	−	(+)	−	nd	nd
M. textilis	A	−	−	−	nd	nd
M. beccarii	C	−	+	+	nd	nd
M. angustigemma	P	−	−	−	nd	nd
M. laterita	R	+	+	+	nd	nd
M. acuminata spp. *banksii*	Eu	+	+	+	+	−
M. a. spp. *burmannicoides*	Eu	+	+	+	+	+
M. a. spp. *malaccensis*	Eu	+	+	+	+	+
M. a. spp. *zebrina*	Eu	+	+	+	+	−
M. balbisiana	Eu	−	−	−	−	−
M. b. type I	Eu	−	−	−	−	−
M. b. type IV	Eu	−	(+)	−	(+)	−
M. schizocarpa	Eu	+	+	+	(+)	+

[a] Taxonomic sections, En, *Ensete*; *A*, *Australimusa*; C, *Callimusa*; P, *Peekeli*; R, *Rhodochlamys*; Eu, *Eumusa*

[b] +, Presence of an amplification product or hybridization signal; −, absence of an amplification product or hybridization signal; (+), presence of PCR amplification product only under low stringency PCR (5°C below Ta value); nd, not done

RFLP probe hybridization. *M. beccarii* belonging to the *Callimusa* section also exhibited an amplification product with the expected size, reflecting the presence of representatives of the Brep 1 sequence family in this section (Fig. 4). The AGRP 54-AGRP 55 primer pair was designed to specifically amplify representatives of *M. acuminata*. The GG dinucleotide at the 3′ end of the AGRP 55 primer was considered as potentially species-specific following sequence data analysis. In fact, the GG dinucleotide appeared on all of the fragments cloned from *M. acuminata* but not in those from *M. schizocarpa*. This particular position was selected for designing primers because it was the only occurrence of a specific dinucleotide site that clearly distinguished the two species, and it allowed us to design a species-specific primer in an approach similar to "digital DNA typing" within the human genome described by Jeffreys et al. (1991). This primer was tested on all genotypes in association with a primer designed in a highly conserved region (i.e. AGRP 54, see Fig. 3). PCR conducted at the defined annealing temperature enabled the detection of products in the expected

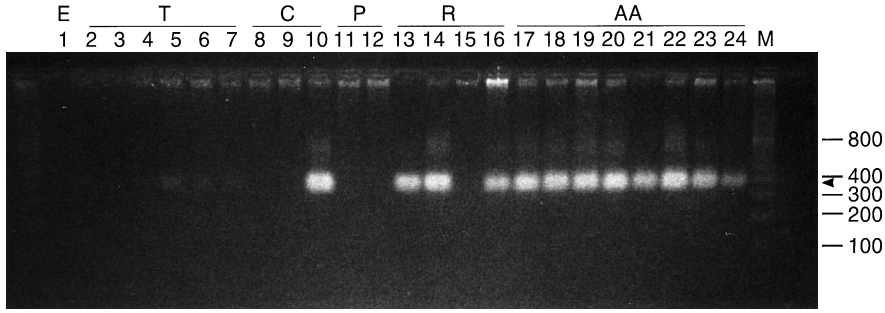

Fig. 4 PCR amplification using the AGRP08/AGRP09 primer pair. Agarose gel electrophoresis of AGRP08/AGRP09 PCR amplification using DNA from a panel of different accessions from the *Musaceae* family. *Lane 1 Ensete* sp. (*E*), *lanes 2–7 Australimusa* section (*T*) *M. boman*, *M. jackeyi*, *M. textilis*, *M. maclayi* 'Huang Si', *M. maclayi* 'Skai', *M. lolodensis* respectively, *lanes 8–10 Callimusa* section (*C*) *M. coccinea*, *M. gracilis*, *M. beccarii*, respectively, *lanes 11–12* Section *Peekeli* (*P*) *M. angustigemma*, *lanes 13–16 Rhodochlamys* section (*R*) *M. ornata*, *M. laterita*, *M. sanguinea* 'Sanguinea', *M. sanguinea* 'ITC0543', respectively, *lanes 17–24 Musa acuminata* spp. *banksii* (*AA*), *lane M* molecular marker (sizes are indicated in bp on the *right*). *Arrowhead* indicates the expected PCR product

size range for the *M. acuminata* representatives but not for *M. schizocarpa* and *M. balbisiana* (Table 3). The AGRP 60-AGRP 61 primer pair was designed to amplify the amplicon in order to test the hypothesis that the Brep 1 family was tandemly repeated. When either primer was used alone, no amplification product could be detected (data not shown), suggesting that this sequence is not invertedly repeated throughout the banana genome. With the AGRP 60 - AGRP 61 primer pair, only a few genotypes exhibited a faint amplification product (Table 3). The sizes of the PCR products were variable: 400 bp in Schizocarpa, 1 kbp in Calcutta 4 and Malaccensis.

Occurrence of Brep 1 in the Musaceae family

The use of previously described PCR primer pairs, AGRP 08-AGRP 09 and AGRP 08-AGRP 07, with lower annealing temperatures (see Materials and methods) enabled detection of a product in other genotypes which exhibited no hybridization signal with the pMaCIR 1115 probe even after a long exposure (3 days). Interestingly, PCR amplification products were detectable in *M. balbisiana* Pisang Batu with very low intensities but not in *M. balbisiana* Butuhan nor Honduras. In the *Australimusa* section, similar weak products were also detected. This reflects the presence of representatives of the Brep 1 family in these clones with a very low-copy number. The primer pair AGRP 54-AGRP 55 was also used under low-stringency conditions. Amplification products appeared in *M. balbisiana* Pisang Batu but not in the other Balbisiana clones tested. A decrease in PCR stringency enabled the detection of slightly degenerated Brep 1 representatives in other species, demonstrating the molecular diversity of this sequence throughout Musaceae.

PCR amplification of subspecies-specific isoforms

The most specific internal AGRP 08-AGRP 09 primer pair was used to amplify the Brep 1 region in different banana clones. A single amplification product appeared at the expected size only in *M. acuminata* and *M. schizocarpa* and no amplification products could be detected in *M. balbisiana*. This leads to easy access of sequence data in different banana genotypes. PCR amplification products of the five positively tested accessions were purified and sequenced. Data from *M. schizocarpa* (Y08647), *M. acuminata* ssp. *banksii* (Y08648) and *M. acuminata* ssp. *zebrina* (Y08649), *M. acuminata* ssp. *malaccensis* (Y08650) and *M. acuminata* ssp. *burmanicoides* (Y08651) are available. Sequence data showed that all these products are uniformly of the same size, 365 bp, and that the expected sequence was amplified. PCR amplification products are not considered to be excellent templates for sequencing due to the risk of nucleotide mis-incorporation by the thermostable polymerase. In order to check sequencing information, PCR products were completely sequenced in both strands. The relatively small size of the products should mitigate the consequences of artefacts due to polymerase errors. Thus, nucleotide positions where point mutation occurred between the different subspecies were identified, suggesting the presence of different isoforms. In addition, triploid banana cultivars were subjected to the same procedure and, interestingly, in contrast to the diploids, no unambiguous sequence data could be obtained due to locally overlapping nucleotide signals.

Evolution of the Brep 1 family

The five PCR isoforms of the *M. acuminata* subspecies and *M. schizocarpa* were used to build a phylogenetic tree (Fig. 5a). *M. acuminata* isoforms were found to be not strongly related to the *M. schizocarpa* isoforms.

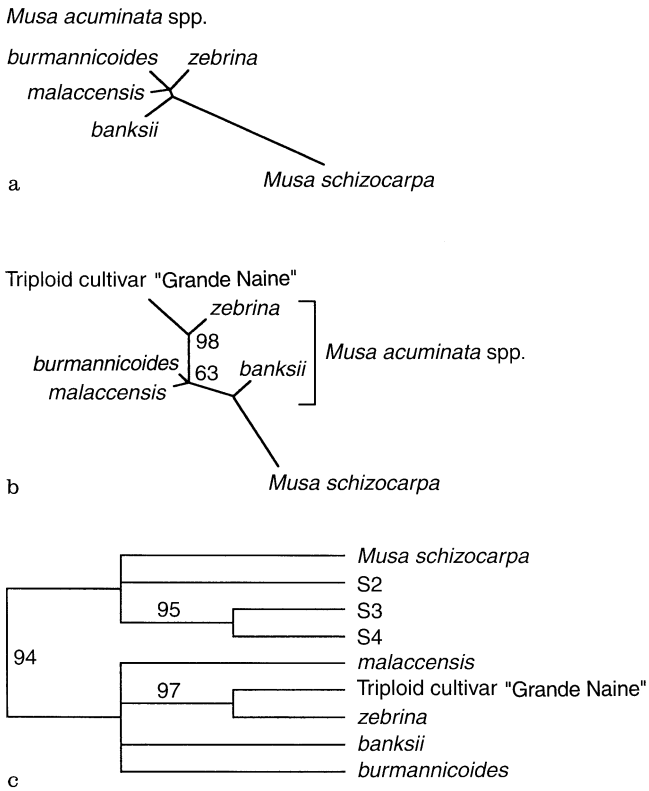

Fig. 5a–c Unrooted phylogenetic tree constructed with nucleotide sequence of Brep 1. **a** Data from the nucleotide sequence of the PCR 365-bp fragment of *M. acuminata* subspecies and *M. schizocarpa* were used to build a phylogenetic tree using the parsimony method (PAUP software). The strictly consensual tree obtained was assayed by bootstrapping (1000), and values are indicated on the appropriate node. **b** Addition of the sequence data from the triploid cultivar 'Grande Naine' (pMaCIR 1115), **c** Dendrogram obtained by adding the homologous fragment of *M. schizocarpa* cloned probes to the tree of **b**. Representation was assayed by bootstraping (1000), and values are indicated on the appropriate node

Genetic distances (Kimura 1980) given in Table 4 show that the Schizocarpa isoform was well-separated from the others, suggesting a strong divergence between the two species. Inside the *M. acuminata* group, Malaccensis and Calcutta 4 were closely related whereas Banksii and Monyet were on separate branches. Four *M. acuminata* isoforms were grouped with a bootstrap value of 63. The use of other phylogenetic tree building methods with different weight in tranversion/transition events gave similar results (data not shown).

In a second approach, an isoform cloned from the genomic DNA of the triploid AAA clone cv 'Grande Naine' was added to the tree (Fig. 5b). Interestingly, it located in the *M. acuminata* part of the tree, strongly related (bootstrap value 97) to the *zebrina* ssp.

Other representatives of the Brep 1 sequence family from *M. schizocarpa* added to the phenogram all fell into the Schizocarpa group (Fig. 5c). This strongly suggests that all Schizocarpa isoforms are related and could originate from a specific form of the Brep 1 family.

Discussion

The occurrence of four highly homologous sequences found in the medium-repeated DNA fraction of the banana genome strongly suggests the existence of a repeated DNA sequence family named Brep 1.

Genomic organization of the Brep 1 sequence family

On the basis of the hybridization pattern, the Brep 1 family seems to be interspersed: an overall conserved core sequence, identified as a 500-bp *Eco*RI fragment containing highly conserved and more variable regions, probably flanked by variable DNA segments. Furthermore, in the conserved core sequence, polymorphic restriction enzyme sites are present. This, added to the interspersed organization of the family, contributed to the production of a complex hybridization pattern. This showed various strong signals of the *Eco*RI fragments produced by cuts of the core repeat at variable positions, thus producing a variety of inserts recoverable by cloning. Short inserts due to the creation of an *Eco*RI site at position 272 were identified (pMaCIR 11145). In the same way, the pMaCIR 11161 clone, lacking the final *Eco*RI site, may identify a specific flanking locus, the 3′ part of this sequence not being related to the Brep 1 family core sequence. Furthermore, hybridization patterns using this probe revealed the strongest signal in *M. schizocarpa* compared to other *M. acuminata* clones (data not shown). This could reflect the presence of species-specific flanking regions in *M. schizocarpa*.

PCR amplification data did not reveal any occurrence of inverted repeats of this core sequence. Intensities of the PCR products obtained using the AGRP 60–AGRP 61 primer pair did not vary between positive clones, whereas differences in copy number, measured using the AGRP 08-AGRP 09 primer pair, ranged over an order of magnitude (Baurens et al. 1996). Moreover, the amplification of single-size spacers in each of these clones reflected the occurrence of reiterations of the conserved core Brep 1 sequence with conservation of the spacer size. The fact that few of the genotypes exhibited this kind of genomic organization and that the low intensity of PCR products might reflect a single locus amplificate is consistent with an interspersed nature of the Brep 1 sequence family. Repetitive sequences are generally classified into two groups: satellite sequences (e.g. Reddy et al. 1993; Schmidt and Heslop-Harrison 1993; King et al. 1995; Kolshinsky and Gresshof 1995) and retroelements (Voytas et al. 1992; White et al. 1994; Aledo et al. 1995). Our data are in accordance with an interspersed nature of the Brep 1 element, but no significant homologies to other known retroposon-like DNA families could be detected.

Specific isoforms of the Brep 1 sequence family

Site-specific PCR amplification of part of this *Eco*RI core sequence enabled rapid and simple access to the isoforms present in a banana population. The sequences of the PCR products must however, be considered with care. They represent the major representatives of the sequence family present in the tested genotypes. If alterations in the sequence appear in a subset of this family, only isoforms with the highest copy number will be amplified whereas minor ones will be obscured. This assumption could be tested analysing different representatives of *M. schizocarpa*. Cloned representatives of the Brep 1 sequence family from *M. schizocarpa* accessions showed the greatest homology (Fig. 5c) with the PCR product sequence of the same species compared with other members of the *Eumusa* section. Nevertheless, no completely identical clones were found. The fact that no clear data sequence could be obtained from triploid cultivars could be interpreted in the same way, suggesting that no real major isoform appears in these due to the combination of different genomes with various copy numbers of the Brep 1 sequence family, characterized by specific nulceotidic alterations. Furthermore, analysis of the sequence data strongly suggests the appearance of genome-specific alterations in the 'Schizocarpa' species. This is supported by the genetic distances separating representatives of *M. schizocarpa* from those of *M. acuminata* (Table 4). Since considerable intra-species variation in this sequence occurs, the sequence divergence observed between *M. schizocarpa* and *M. acuminata* representatives is significant and indicative of the molecular evolution of the Brep 1 sequence family.

Evolution of the Brep 1 sequence family

The Brep 1 sequence family was present in all the representatives of *M. acuminata*, *M. schizocarpa* and in the *Rhodochlamys* section with variable copy numbers. Considering the absence of hybridization signals with the RFLP probe on the *Australimusa* and *Callimusa* sections and in *M. balbisiana* and the production of amplification products using low-stringency PCR amplifications, we suggest that the Brep 1 family was present at extremely low copy numbers in these clones. Brep 1 was not detected in *Ensete*, which is considered to be the species most closely related to the putative common ancestor for the Musaceae family (Cheesman 1947). It is clear that the appearance of the Brep 1 sequence family preceded speciation of the *Callimusa*, *Australimusa* and the representatives of the *Eumusa* sections. In *M. acuminata* and *M. schizocarpa* and in the *Rhodochlomys* section, copy numbers of the Brep 1 sequence family were considerably increased. Our data on the Brep 1 sequence family support the hypothesis of Simmonds (1962) concerning the evolution of the different sections of the Musaceae.

In view of the absence of homology of the Brep 1 family to other known plant or animal interspersed elements, the most probable hypothesis concerning the molecular evolution of its sequence is that the single-ancestor sequence duplicated during evolution. As copy numbers of the Brep 1 family in *M. balbisiana* or in the *Australimusa* and *Callimusa* sections were quite low, representatives from these accessions could be more related to the hypothetical ancestor because of the absence of multiplication of the sequence. Analysis of the PCR amplification products suggests that the occurrence of tandem repeats of this family is not the general case but occurred at a very low rate in the genome. This suggests a complex mechanism of duplication which could lead to changes in the sequence structure except for the conserved core sequence found in all the Musaceae. This was confirmed by the diversity of sequences found in *M. schizocarpa* compared to the major isoforms obtained by PCR amplification. Further investigation is, however, needed to confirm the founder-isoform hypothesis. However, a possible horizontal transfer from one species to another cannot be completely excluded since horizontal transmission of Gypsy transposons among different fly species has been proposed (Mizrokhi and Mazo 1991).

M. schizocarpa is known to have diverged from the *Eumusa* section earlier than *M. acuminata* (Simmonds 1962). The evidence of Schizocarpa species-specific point mutations could be interpreted as the differential evolution of the ancestral sequence within the two species and could have contributed to their speciation (Radman and Wagner 1993). This is to be compared to the difference in copy number as evidenced by the analysis of RFLP hybridization patterns, the relative intensities of PCR products and the results of low-stringency PCR. The *Australimusa* and *Callimusa* sections separated before the appearance of the *Eumusa* section. Copy numbers of the Brep 1 sequence family in

Table 4 Pairwise Kimura genetic distance between wild and cultivated bananas

	1	2	3	4	5	AAA
1	–	0.0673	0.0900	0.0673	0.0672	0.0891
2		–	0.0285	0.0056	0.0055	0.0292
3			–	0.0197	0.0226	0.0116
4				–	0	0.0174
5					–	0.0231
AAA						–

Sequence data of the 365-bp fragment of PCR-amplified alleles of clones 1–5 and the homologous regions of the pMaCIR 1115 cloned probe from the triploid cultivar 'Grande Naine' (AAA) were used to calculate genetic distances using the DNADIST software (PHYLIP package)

[a] Numbers of the wild banana clones are those listed in Table 1

the Musaceae family seem to have increased during evolution. Thus, the presence in *M. balbisiana* of a Brep 1 representative, with very low copy numbers, is in accordance with the hypothesis of the divergence of *M. balbisiana* prior to other representatives of the *Eumusa* and *Rhodochlamys* sections.

In the *M. acuminata* species complex, both copy number and nucleotidic sequences divergence could be considered for phylogeny. Given that the *M. schizocarpa* species separated from the others earlier (Simmonds 1962), a hypothetical root can be placed on the tree between these two species. The tree constructed with nucleotidic data shows that the *zebrina* subspecies seems to have diverged more recently, whereas the subspecies *banksii* differentiated early. The *malaccensis* and *burmannicoides* subspecies cannot be separated using these data. Interestingly, copy-number data show that the *malaccensis* subspecies carries six times more copies of the sequence than *banksii* (Baurens et al. 1996). This suggests that the sequence duplication process remains complex with the selection both high- and low-copy numbers.

Another interesting characteristic of this sequence family is the genetic relationship between its representatives among the *M. acuminata* species complex. Sequence data could be used for analysis of the complex polyploid genomes of some banana cultivars. Carreel (1994) demonstrated, using RFLP analysis of cytosolic probes, that the triploid (AAA) Cavendish group was related to *M. acuminata* ssp. *errans* and *M. acuminata* ssp. *malaccensis* whereas RFLP data from the nuclear genome did not allow for any strong association with either subspecies. In spite of the high number of nuclear probes used in this study, it was not possible to regroup this triploid clone to any diploid wild-type subspecies. Repeated sequences, on the other hand, form a large portion of the genome. In this particular case, it was possible to link the triploid cultivar to the *zebrina* ssp. This leads to the hypothesis that three *M. acuminata* subspecies are included in the cultivar 'Grande Naine': *errans* ssp., *malaccensis* ssp. and *zebrina* ssp. The use of cytoplasmic probes in banana is very interesting due to the fact that the two cytoplasmic genomes are differentially inherited in wild diploid *Eumusa* (Fauré et al. 1993b). If, as a working hypothesis, the strictly paternal inheritance of the mitochondrial genome and maternal transmission of the chloroplast genome could be extrapolated to all the cultivated clones the history of the cross that originally produced the AAA cultivar 'Grande Naine' might be explained.

As *zebrina* and *malaccensis* are sympatric, the first original cross could have been female *malaccensis* × male *zebrina*. This hybrid should have produced a diploid female gamete which subsequently crossed with an *errans* male donor would have produced a triploid AAA individual with an *errans* ssp. mitotype, a *malaccensis* ssp. chlorotype and a *zebrina* ssp. nuclear Brep 1 sequence family type. The reality of this cross history is surely not so simple. Hybrids most probably replaced the putatively "pure" subspecies parents, giving broader genetic bases to the cultivar. Moreover, it has been suggested that the effect of repeated selection/mutation on the putative original cultivar was conducive to strong differentiation from its ancestral progenitors (Bhat and Jarret 1995). This could also have enhanced the actual genetic diversity of 'Grande Naine', Considering the genome representativity of a repeated sequence and the fact that Brep 1 could be useful for phylogeny, this sequence should be a good marker and useful for the genetic improvement of banana with molecular breeding strategies.

Conclusion and prospects

Variable copy numbers of the Brep 1 interspersed repetitive element occur in the Musaceae family. Starting from only a few repetitions in *Musa beccarii* in the *Autralimusa* section and *Musa balbisiana* in the *Eumusa* section, copy numbers increase in the *Rhodochlamys* section, to reach a maximum in the *Musa acuminata* and *M. schizocarpa* species from the *Eumusa* section. Significant variations in copy number allow the different subspecies of *M. acuminata* to be distinguished (Baurens et al. 1996).

It has been shown that repetitive elements are a basis for efficient molecular markers in banana (Jarret et al. 1993; Kaemmer et al. 1992). Sequence and molecular organization data on the Brep 1 sequence family members open up new opportunities for the development of site-specific PCR. A PCR-typing approach, analogous to those developed by Jeffreys et al. (1991), could be applied using carefully chosen primers that carry a species-specific point mutation at their 3′ end. The isoforms of *M. schizocarpa* and those of *M. acuminata* are unambiguously grouped in two separate parts of the phylogenetic tree synoptically representing our data. Moreover, the addition to the tree of data corresponding to the cloned probes of *M. schizocarpa* confirms through the identification of a strong Schizocarpa isoform group (Fig. 3) that some nucleotide positions may be characteristic of different species of the Musaceae family. Based on this, and bearing in mind all of the difficulties due to ambiguous nucleotide positions from PCR product sequence data acquisition, primer positions have to be carefully defined for the detection of species-specific point mutations in the core of the Brep 1 family. Accumulation of data corresponding to the *M. acuminata* subspecies-specific isoforms could be very informative and might open the way to distinguishing the *M. acuminata* subspecies involved in cultivars using PCR tests.

Copy numbers of the Brep 1 sequence family previously assayed in *M. acuminata* ssp. showed a large variation, depending on the *M. acuminata* subspecies, and throughout the whole Musaceae family itself.

Additional investigations are, however, needed to clarify the nature of the Brep 1 repetitive element and any potential correlation between copy number of the sequence and the expression of agronomical traits. Overall, the relationship between repeated sequences, speciation and genomic structure alterations has been discussed elsewhere (Radman and Wagner 1993), and a link to parthenocarpy or sterility of banana is not to be excluded. This can be further explored by in situ hybridization of this sequence family for precise physical mapping and the study of interspecific crosses.

Acknowledgements F. Carreel is gratefully acknowledged for providing DNA samples and for the pMaCIR 1115 hybridization data. E. Jenczewski and J. Ronfort are also thanked for helpful discussions and N. Wynn for English corrections. This work was supported by a grant of CIRAD.

References

Aledo R, Raz R, Montfor A, Vicient CM, Puigdomenech P, Martinez-Izquierdo JA (1995) Chromosome localization and characterization of a family of long interspersed repetitive DNA elements from the genus *Zea*. Theor Appl Genet 90:1094–1100

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic Local Alignment Search Tool. J Mol Biol 215:403–410

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

Bhat KV, Jarret RL (1995) Random amplified polymorphic DNA and genetic diversity in Indian Musa germplasm. Genet Resource Crop Evol 42:107–118

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

Carreel F (1994) Etude de la diversite génétique (genre *Musa*) à l'aide des RFLP. PhD thesis, Institut National Agronomique Paris-Grignon, France

Cheesman EE (1947) Classification of the bananas. Kew Bull 2: 97–117

Dessauw D (1987) Etude de la stérilité du bananier (*Musa* ssp.) et des relations cytotaxonomiques entre *M. acuminata* et *M. balbisiana* Colla. PhD thesis, Paris XI University, France

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

Fauré S, Noyer JL, Horry JP, Bakry F, Lanaud C, González de León D (1993a) A molecular marker-based linkage map of diploid bananas (*Musa acuminata*). Theor Appl Genet 87: 517–526

Fauré S, Noyer JL, Carreel F, Horry JP, Bakry F, Lanaud C (1993b) Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas (*Musa acuminata*). Curr Genet 25:265–269

Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, Wash.

Gawel NJ, Jarret RL (1991) A modified CTAB DNA extraction procedure for *Musa* and *Ipomoea* Plant Mol Biol Rep 9:262–266

Jarret RL, Vuylsteke DR, Gawel NJ, Pimentel RB, Dunbar LJ (1993) Detecting genetic diversity in diploid bananas using PCR and primers from a highly repetitive DNA sequence. Euphytica 68:69–76

Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. Nature 354:204–209

Kaemmer D, Afza R, Weising K, Kahl G, Novak FJ (1992) Oligonucleotide and amplification fingerprinting of wild species and cultivars of banana (*Musa* spp). Bio/Technology 10: 1030–1035

Kimura M (1980) A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

King K, Jobst J, Hemleben V (1995) Differential homogenization and amplification of two satellite DNAs in the genus *Cucurbita* (Cucurbitaceae). J Mol Evol 41:996–1005

Kolchinsky A, Gresshoff PM (1995) A major satellite DNA of soybean is a 92-base pairs tandem repeat. Theor Appl Genet 90:621–626

Lebot V, Aradhya KM, Manshardt R, Meilleur B (1993) Genetic relationships among cultivated bananas and plantains from Asia and the Pacific. Euphytica 67:163–175

Li X-B, Liang C-Z, Wu H-G, Zhai W-X, Huang N, Zhu L-H (1996) Isolation and identification of a non-specific tandemly repeated DNA sequence in *Oryza* species. Theor Appl Genet 92:702–708

Mizrokhi LJ, Mazo AM (1991) Cloning and analysis of the mobile element gypsy from *D. viridis*. Nucleic Acids Res 19:913–916

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. Proc Natl Acad Sci USA 85:2444–2448

Radman M, Wagner R (1993) Mismatch recognition in chromosomal interactions and speciation. Chromosoma 102:369–373

Reddy AS, Kiefer-Meyer MC, and Delseny M (1993) Characterization of new variants of a satellite DNA from *Oryza officinalis*, specific for the CC genome of wild rice. Genome 36:750–761

Schmidt T, Helsop-Harrison JS (1993) Variability and evolution of highly repeated DNA sequences in the genus *Beta*. Genome 36:1074–1079

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

Simmonds NW (1962) The evolution of bananas. Longman Group UK Limited, London, pp 268

Swofford DL (1993) PAUP 3.1 (Phylogenetic Analysis Using Parsimony), Laboratory of molecular systematics, Smithsonian Institution, Washington, D.C.

Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) Copia-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89:7124–7128

White SE, Habera LF, Wessler SR (1994) Retrotransposons in the flanking regions of normal plants genes: a role of copia-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91:11792–11796