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) · Somatic variation · Repetitive sequence · In situ hybridization

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. Niymara 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 *FIX* 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 *Sall* fragment, and a 4.0-kb *EcoRI* 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* fsp. *lycopersici*. 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/m*onkey homology are indicated. The relative positions of the subcloned 1.5- and 1.8-kb *Sall* and the 4.0-kb *EcoRI* 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 *HindIII* and *Xhol* are also indicated. The approximate positions of stop codons within the major ORF are indicated (x).