Thesis submitted for the degree of

Doctor of Philosophy

at the University of Leicester

by

Chee How Teo

Department of Biology

University of Leicester

PhD 2007

UMI Number: U231436

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI U231436
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

I hereby declare that no part of this thesis has been previously submitted to this or any other University as part of the requirements for a higher degree. The content of this thesis is the result of my own work unless otherwise acknowledged in the text or by reference.

The work was conducted in the Department of Biology, University of Leicester during the period January 2004 to January 2007.

Signed.....

Chee How Teo, October 2007

LTR retrotransposons and tandem repeats in *Musa* genomes and their contribution to *Musa* diversity and genome evolution

By

Chee How Teo

Abstract

Musa is a large genus that includes many important staple crops particular in Asia and Africa. *Pseudoviridae*, LARD-like and three different groups of *Metaviridae* elements were identified based on their RT sequence heterogeneity. Additionally, full length elements were identified in published BAC sequences. Most of the LTR-retrotransposons were degenerate and disrupted at different positions, by other types of repetitive DNAs (retrotransposons and tandem repeats). *Metaviridae* elements are generally present in higher copy numbers than *Pseudoviridae* elements, and the A genome contains more retrotransposons than the B genome, possibly explaining the larger genome size of the A genome. However, genome specific LTR-retrotransposons were not detected. Retrotransposons were clustered in the centromeric regions, *Pseudoviridae* elements showing a more dispersed pattern while LARD-like and some of the *Metaviridae* elements were also found at the Nucleolar Organizing Region (NOR). The application of LTR-retrotransposons as molecular markers based on insertion polymorphism allowed identification of different *Musa* taxa. Two genome-specific tandem repeat DNA (CoTR and MuTR) families were identified. The MuTR repeats, organised in long arrays are located at the NOR loci of seven *Musa* taxa and additionally at some to all centromeres of selected taxa. Different degrees of methylation of LTR retrotransposons and tandem repeat DNA, the latter generally being under-methylated were observed. The transcribed copies of the 45SrDNA sequences at the NOR are interspersed by MuTR repeats and weakly methylated, while the distal part of the NOR region and the chromosomal satellite beyond contains MuTR, *Metavirdae* and LARD-like sequences that are heavily methylated. The close chromosomal proximity and insertion within each other postulates a possible link in the evolution of LTR-retrotransposons, tandem repeats and 5S rRNA genes, but also a role of retrotransposon sequences in gene regulation. RNA transcripts of both *Metaviridae* and MuTR repeats indicates their activity, but also that they might be involved in siRNA silencing mechanisms.

Dedication

I would like to dedicate this thesis to my lovely wife Foong Chuen Yi for her support, encouragement and motivation that made this possible. It is also to my parents Teo Man Sing and Kee Wuang Hiah for their education, my in laws and my brothers and sisters for their support.

Acknowledgements

First of all, I would like to acknowledge my supervisor Dr. Trude Schwarzacher for her assistance, advice, guidance, help and support during this project.

I would also like to acknowledge numerous people for their theoretical and technical assistance of this work. It begins with the help and encouragement from Dr. John Bailey followed by Dr. Richard Gornall, Prof. John Seymour (Pat) Heslop-Harrison, Dr. Celia Hansen, Dr. Alessandra Contento, Prof. Ashalatha Nair, Dr. Bohumil Mandák, Dr. Gustavo Kuhn, Dr. Kateřina Bímová, Dr. Pratap Handique, Prof. Takato Koba, all the members of the Fifth EU Framework Programme, PARADIGM and Department of Biology, University of Leicester.

Particular thanks to Jean Liggins for her help and creating a harmony and efficient working environment in the lab, followed by Judy Horner, Sue Odgen, Penny Butler, Judy Jackson, Steve Ison, Penny Chamberlain, Elaine Rawding, Lesley Barnett for their assistance, hospitality and friendship.

I would like to express my sincere gratitude to all my friends all around the world that I met in the Laboratory Molecular Cytogenetics for all the cheerful company, laugh, motivation and help. This includes Néstor Santiago Gonzalvo, Assunta Biscotti, Johanna Wikgren, Mariangela Vilckas, Emmanuel Plas Otwe, Saadiah Jamli, Ulrike Anhalt, Alexandros Bousios, Xu Xiang Ming, Makoto Fujiwara, Maria Maron, Filomena Filadega, Hojatolah Saiedi, Said Hafidh, Jennifer Haught, Ramesh Patel, Maurice Andrews, Husni Ibrahim, Haniza Moh'd Zain, Azhar Mohamad and many more.

I would like to convey my gratitude to Prof. Rofina Yasmin Othman, Dr. Tan Siang Hee, Mr. Markku Häkkinen and International Transit Centre for providing the banana materials and helping me with their expertise in bananas.

Least but not less, a since appreciation to the Fifth EU Framework Programme, PARADIGM and Generation Challenge Program of CGIAR for the consumable support and the Global *Musa* Genomics Consortium for the genomic resources.

Abbreviations

AFLP	amplified fragment length polymorphism
AP	aspartic protease
BAC	Bacterial artificial chromosome
BCIP	5-bromo-4-chloro-2-indolyphosphate
bp	base pair
BSA	bovine serum albumin
BSV	*banana streak virus*
CD	chromodomain
cDNA	complementary DNA
CSSV	*cacao swollen shoot virus*
CTAB	cetyltrimethyl-ammoniumbromide
°C	degree Celsius
DAPI	4',6-diamidino-2-phenylindole
DTT	1,4-dithio-DL-threitol
dH_2O	distilled water
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotidetriphosphates
DR	direct repeat
EDTA	ethylenediamine tetraacetic aicd
env	envelope
EST	expressed sequence tag
FAO	Food and Agricultute Organization of the United Nations
FISH	fluorescent *in situ* hybridization
FITC	fluorescein isothiocyanate
FLE	full length element
g, µg, mg	gram, microgram, milligram
H3	histone 3
hrs	hours
GAG	group-specific antigen
ICTV	International Committee on the Taxonomy of Viruses
INT	integrase
IPTG	isopropyl-β-thiogalactosid
IR	inverted repeat
IRAP	inter-retrotransposon amplified polymorphism
ISTR	inverse sequence-tagged repeats
ITC	International Trade Centre
ITS	internal transcribed spacer
µl, ml	microliter, milliliter
LARD	large retrotransposon derivatives
LB	Luria-Bertani
LTR	long terminal repeat
M, mM	molar, millimolar
MARDI	Malaysian Agricultural Research and Development Institute
MAS	marker assisted selection
min	minutes
MRIS	*Musa* retrotransposon insertion study
mRNA	messager RNA

NaCl	sodium chloride
NBT	4-nitroblue tetrazolium chloride
NIB	nuclei isolation buffer
NOR	nucleolar organizing region
nt	nucleotide
NTS	non-transcribed spacer
ORF	open reading frame
PBS	primer binding site, phosphate-buffered saline
PCR	polymerase chain reaction
%	percent
POL	polyprotein
PPT	polypurine tract
PVP-40	polyvinylpyrrolidone, molecular weight 40,000
R	repeated RNA
RAPD	random amplified polymorphic DNA
RBIP	retrotransposon-based insertion polymorphisms
rDNA	ribosomal DNA
REMAP	retrotransposon-microsatellites amplified polymorphism
RFLP	restriction fragment length polymorphism
RH	ribonuclease N/RNase H
RNA	ribonucleic acid
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	reverse transcriptase
RT-PCR	reverse transcriptase PCR
SDS	sodium dodecyl sulphate
SINE	short interspersed nuclear element
sp, ssp	species, subspecies
S-SAP	sequence-specific amplified polymorphism
SSC	sodium chloride sodium citrate/saline sodium citrate
SSR	simple sequence repeat
STS	sequence-tagged site
TAE	Tris-acetate-EDTA
TAS	telomeric associated sequence
TDR	terminal direct repeat
TE	Tris-EDTA buffer
T_m, T_a	melting temperature, annealing temperature
TRIM	terminal-repeat retrotransposon in miniature
tRNA	transfer RNA
TSD	target site duplication
U	Unit
UTS	un-translated sequence
U3	3' untranslated region
U5	5' untranslated region
V	volt
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosid

TABLE OF CONTENT

1.0 CHAPTER I: Introduction

1.1 *Musa* species

1.1.1 Origin and diversity of *Musa* species, varieties and cultivars

The genus *Musa* contains about 50 diploid species and can be divided into four sections: *Eumusa* (*2n=2x*=22) including *Musa acuminata* and *M. balbisiana*, edible bananas and plantains, *Rhodochlamys* (*2n=2x*=22), *Callimusa* (*2n=2x*=20, except for *Musa beccarii* with *2n=2x*=18) and *Australimusa* (*2n=2x*=20) (Figure 1.1). *Musa* species are thought to originate in Southeast Asia, where they are native, and distributed from India to Polynesia (Simmonds, 1962). The discovery of new wild *Musa* species mainly in the *Callimusa* section has increased the number of wild *Musa* species in the International Transit Centre (ITC) germplasm collection and this will help to increase the genetic diversity of *Musa* species (Valmayor, 2001; Häkkinen and De Langhe, 2001; Häkkinen, 2001; 2003; Häkkinen and Sharrock, 2002; Valmayor *et al.*, 2004; Häkkinen, 2004a, b; Häkkinen, 2005; Häkkinen and Meekiong, 2004; 2005; Häkkinen *et al.*, 2005; Häkkinen, 2006a, b).

Bananas and plantains are the fourth most important crop in developing countries, where they are an important staple food in the humid tropics and also an important source of income in rural areas (FAO, 2007; Heslop-Harrison and Schwarzacher, 2007). The modern edible bananas and plantains are derived from the inter- or intra-species hybrids and polyploids of two wild, seeded species, *Musa acuminata* and *M. balbisiana*, which are native to Southeast Asia and India, respectively. *M. acuminata* has the genome designation AA and *M. balbisiana* with the genome designation BB are both diploids with *2n =2x*=22 (Figure 1.1). The word 'plantain' is very ambiguous as the term implies an cooking banana, but in Spanish the word can also be used to mean dessert forms (Gowen, 1995). There is thus

considerable overlap between bananas and plantains dependent on the way they are consumed.

The centre of diversity of edible bananas and plantains is concentrated in Malaysia and Indonesia (Daniellis *et al.*, 2001), where they were first recorded several thousand years ago and have consequently spread in the hands of travellers, eastwards to the remoter Pacific Islands and westwards to Africa, probably via Madagascar (Simmonds, 1962; 1976; Gowen, 1995). European visitors first found them in West Africa and it was from there that several clones were taken to the New World. The history of the crop in Africa still presents many uncertainties and it is not clear why the varieties grown in eastern and western Africa can be so different if they had a common ancestor (Simmonds, 1962; 1976). Generally, the first step of evolution of edible bananas and plantains involved the development of parthenocarpy and seed-sterility in *M. acuminata* under human selection, which subsequently gave rise to different types of edible diploid *M. acuminata* cultivars in Southeast Asia. From the AA cultivars, by chromosome restitution at meiosis or unreduced gametes, AAA (*acuminata*) triploids resulted that are the basis of all sweet dessert bananas exported to the west. Another step in the evolution of edible bananas and plantains in Asia, was the crossing of AA cultivars with *M. balbisiana* BB wild species from India, which generated the inter-specific hybrids (AB, ABB, AAB, ABBB, AABB and AAAB). *M. balbisiana* is a hardier and more drought-tolerant plant compared to *M. acuminata*, and hybrids have been selected to extend the geographical range of the bananas out of the wetter tropics into the seasonally drier zone (Gowen, 1995).

1.1.2 Classification of banana cultivars

1.1.2.1 Taxonomic classification by morphological traits

The generic name *Musa* was derived from the Arabic word *mouz*. Bananas were known to the early Arabs and appear in the Koran as the 'tree of paradise'. The earliest 'scientific' classification of bananas was made by Linnaeus in 1783 who gave the name *Musa sapientium* to all dessert bananas which are sweet when ripe and eaten raw. The name *M. paradisiaca* was given to the plantain group that are cooked and consumed while still starchy (Figure 1.1). However, it is now known that they are not species but both refer to closely related interspecific triploid hybrids of the AAB group (Simmonds and Shepherd, 1955).

The modern classification of edible bananas (see Purseglove, 1972; Stover and Simmonds, 1987; Valmayor *et al.*, 1990 and Valmayor *et al.*, 1991) was devised by Simmonds and Shepherd (1955) and is based on the relative contribution of the two wild species *M. acuminata* (AA) and *M. balbisiana* (BB) to the genomic constitution of the cultivars (Figure 1.1). Simmonds and Shepherd (1955) showed that the contributions of the two species could be clearly distinguished by using 15 diagnostic characters. For each character in which cultivars agreed completely with wild *acuminata*, a score of 1 was given, and for each character in which the cultivars agreed with *balbisiana*, a score of 5 was given. Intermediate expressions of the character were assigned a score of 2, 3 or 4, according to its intensity. The scoring technique allows for a range of scores from 15 (pure *M. acuminata*) to 75 (pure *M. balbisiana*). Simmonds and Shepherd (1955) and Stover and Simmonds (1987) used these scoring techniques to classify a range of edible bananas. Silayoi and Chomchalow (1987) classified 137 accessions in the Thai banana genebank on this basis with some modifications.

Espino and Pimental (1990) demonstrated the ability of isozyme technology to distinguish clones of pure *acuminata*, pure *balbisiana*, and their hybrids from one another. They found broad bands of malate dehydrogenase activity which were unique to pure *balbisiana*, and other bands which were indicative of the presence of an *acuminata* genome. They concluded that BB and BBB cultivars were unique and distinct from hybrid ABB clones. Valmayor *et al.* (1991) concluded the classification of edible bananas by endorsing the continued adoption of Simmonds and Shepherd's classification scheme, but also recommended that it be expanded and fine-tuned to accommodate the great diversity of banana cultivars in the centre of origin, namely Southeast Asia.

1.1.2.2 Polyploidy in bananas and their molecular cytogenetic aspect

As mentioned above, most cultivated bananas belonging to the section *Eumusa* are polyploids with 33 or 44 chromosomes with haploid number of 11, and most of these cultivars are vegetatively propagated triploids, with the remainder being mostly diploids while tetraploid clones are very rare (Gowen, 1995). The problems during meiosis due to the uneven numbers of chromosomes gave rise to the sterile triploid cultivars. Ortiz (1997) reported that the unilateral sexual polyploidization ($2nxn$) might be involved in the origin of triploid *Musa* species and at least one dominant gene controls the production of $2n$ pollen in *Musa* species based on the segregation data. In addition, aneuploidy, hypotriploidy (2n=28) and polyploidy were detected at the cell suspension level of triploid *Musa* hybrids (Sandoval *et al.*, 1996; Roux *et al.*, 2004). This leads to the production of tetraploid *Musa* hybrids, which exhibit some degree of female fertility and has been considered to be an important breakthrough in banana breeding programme. As ploidy is concerned, it is necessary to know the

ploidy of a clone before it can be correctly classified. This can only be done cytologically by making a chromosome count (Cheesman, 1932; Simmonds, 1954) or by flow cytometry (Doležel *et al.*, 1994; Lysák *et al.*, 1999; Roux *et al.*, 2004).

Musa species has a small genome size of 500 – 600 Mbp per haploid genome with different chromosome numbers observed in species of different sections. The small genome size (Dolezel *et al.*, 1994) and the small chromosomes (Isobe and Hashimoto, 1994; Osuji *et al.*, 1996a) together with the difficulty of obtaining dividing cells (Osuji *et al.*, 1996b) have limited the application of traditional cytogenetic techniques. However, the development of Fluorescent *In Situ* Hybridization using repetitive DNA probes has enabled characterization of *Musa* genomes (Osuji *et al.*, 1997). Genomic *in situ* hybridization techniques have been applied to *Musa* taxa in order to identify the genome constitution of *Musa* lines (bananas, plantains and artificial hybrids)(Osuji *et al.*, 1997). Double target *in situ* hybridization using labelled total genomic DNA from two *Musa* species, *M. acuminata* (A) and *M. balbisiana* (B), enabled the identification of *Musa* genome constitution with A genome probe hybridized to a broad centromeric region of *M. acuminata* 'Calcutta 4' and B genome probe hybridized strongly to the centromeric regions of *M. balbisiana* 'Butohan 2'. Labelled A and B genome DNAs were able to distinguish the A and B genome chromosomes in interspecies hybrids. This will help the identification of chromosome origin and also help to characterize the cultivars and hybrids produced in different *Musa* breeding programs.

Flow cytometry has been applied to *Musa* species to estimate the genome size of different *Musa* species (Asif *et al.*, 2001; Bartoš *et al.*, 2005 and Pillay *et al.*, 2006). In addition, this technique has been used to detect ploidy level instability in *Musa* embryogenic cell suspension cultures (Roux *et* al, 2004). Bartoš *et al.* (2005)

showed that nuclear genome size estimation of *Musa* species together with the chromosomal localization of 45S and 5S rDNA loci allow the clear distinction of different *Musa* species. This suggested that a combined approach using flow cytometry and FISH with repetitive DNA probes can help to characterize the genome constitution of *Musa* species more precisely. Flow cytometry showed that *M. balbisiana* had the lowest 2C value (1.13 – 1.133 pg) of *Musa* taxa studied whereas *M. beccarii* ($2n=2x=20$) had the highest 2C value of 1.561. Together with the results from the 45S and 5S rDNA loci, where *Musa balbisiana* contained only one pair of NORs per chromosome set and *Musa beccarii* contained three pairs, suggested that repetitive DNA, to some extent, plays an important role in the genome size expansion of *Musa* species.

1.1.2.3 Application of the molecular markers to *Musa* classification

Molecular markers can be used for 'genetic mapping' purposes in which the segregation of agronomic traits, such as resistance to Black Sigatoka disease, can be identified for breeding programmes. Molecular markers can also be used in plant breeding in several ways: (1) *Marker-assisted selection* (MAS), where the information of the molecular markers at the DNA-level, such as whether they are closely linked to, or even located within, one or more quantitative trait loci (QTL), can be used directly to increase the response to selection of commercially important bananas. (2) *Marker-assisted introgression*, where molecular markers are used to increase the efficiency of introgression when the gene of interest is introduced from wild relatives into modern plant varieties. (3) *Studies of genetic diversity*, where molecular markers can be used to elucidate the taxonomic/phylogenetic relationships between *Musa* species or varieties or between populations within species. (4) *Studies of biological processes*, where molecular markers can be used to study the genetic

mechanisms that are involved in the physiological trait changes such as mating systems, pollen movement or seed dispersal (Gowen, 1995).

One of the first molecular marker techniques is 'restriction fragment length polymorphism' (RFLPs). RFLP of diverse germplasm has been used to study the taxonomy and phylogeny of *Musa* species (Jarret *et al.*, 1992; Gawel *et al.*, 1992; Lanaud *et al.*, 1992; Carreel *et al.*, 1994; Jenny *et al.*, 1997) and variation in the chloroplast genome within the *Musa* genus (Gawel and Jarret, 1991; Baurens *et al.*, 1997). RFLPs are sensitive and can be used to determine close relationships in accessions of similar genome and ploidy level. However, there is only one report of their use to distinguish more closely related material in *Musa* (Bhat *et al.*, 1994) and they are laborious and expensive to develop and use as they require large amount of DNA. Hence, for high throughput and routine breeding applications, researchers have concentrated on the use of the polymerase chain reaction (PCR).

The Random Amplified Polymorphic DNA (RAPD) technique has been successfully used to distinguish *Musa* genome groups from diverse *Musa* germplasms (Howell *et al.*, 1994; Bhat and Jarret, 1995; Crouch *et al.*, 1998). In addition, a molecular linkage map has also been developed using a variety of marker systems including RAPDs (Faure *et al.*, 1993). RAPD assays are particularly useful, as they require no prior knowledge of the organism. However, RAPD analysis has several disadvantages including the dominant nature of the marker system and reproducibility problems, which may limit their application in MAS.

Simple sequence repeats (SSRs) or microsatellites are regions of short tandemly repeated DNA motifs (generally less than or equal to 4 bp) with an overall length in the order of tens of base pairs. SSRs have been reported to be highly abundant and randomly dispersed throughout the genomes of many plant species.

Characteristic variation in the number of times the motif is repeated is found in population and species, and is thought to arise through slippage errors during DNA replication, thus, creating length polymorphisms. Microsatellite length polymorphisms may occur even between closely related individuals. Microsatellite markers are generated by highly specific PCR amplification and therefore should not suffer from the reproducibility problems experienced with RAPD analysis (Crouch *et al.*, 1998). Microsatellite analysis has been shown to detect a high level of polymorphism between individuals of *Musa* breeding populations (Jarret *et al.*, 1994; Crouch *et al.*, 1998). However, the isolation of microsatellites is time consuming and expensive. Nonetheless, the isolation of SSRs is becoming increasingly routine with the availability of automated DNA sequencing facilities; improved techniques for the construction of genomic libraries enriched for SSRs and improved techniques for the screening of appropriate clones (Jarret *et al.*, 1994; Kaemmer *et al.*, 1997; Crouch *et al.*, 1998; Creste *et al.*, 2001; 2003a, b; 2006; Ge *et al.*, 2005; Oriero *et al.*, 2006). *Musa* microsatellites have been considered as optimum markers compared to other systems due to their abundance, polymorphism and reliability (Jarret *et al.*, 1994; Kaemmer *et al.*, 1997).

The technique of amplified fragment length polymorphisms (AFLP) which is very robust and able to produce large numbers of polymorphic bands without any prior knowledge of the organism, has been applied to *Musa* (Loh *et al.*, 2000; Wong *et al.*, 2002; Ude *et al.*, 2002; Bhat *et al.*, 2004). Software has been developed to distinguish homozygotes and heterozygotes on the basis of band intensity, although such an approach may be frequently confounded by the presence of bands of intermediate intensity. However, AFLP assays are also technically demanding and expensive in that they require a number of DNA manipulations and a complex

visualization procedure. In addition, they require relatively large amounts of reasonably high quality DNA. The use of poor quality DNA may lead to incomplete digestion, which can result in artefactual polymorphisms (Bhat *et al.*, 1995).

Despite the fact that the different types of molecular markers mentioned above are suitable for *Musa* species, it is important to choose the most suitable molecular marker system to help the breeding programmes. Further molecular marker development based on the repetitive DNA such as retrotransposons is needed to provide another perspective for *Musa* classification and breeding studies.

1.2 Retrotransposons

Studies of different types of retroelements (retroviruses, pararetroviruses, retroposons and retrotransposons) in different organisms has shown that they are the most common group of eukaryotic transposable elements (Michel and Lang, 1985; Boeke and Corces, 1989; Doolittle *et al.*, 1989; Meunier *et al.*, 1990; Xiong and Eickbush, 1990; McClure, 1991; Lambowitz and Belfort, 1993; Schmidt and Heslop-Harrison, 1998; Kumar and Bennetzen, 1999; Hohn *et al.*, 2007).

Genome sequencing projects of plant species have further enhanced the understanding of retroelement diversity and evolutionary trends. Among all the retroelements studied, retrotransposons have been shown to occupy the largest proportion of plant genomes together with DNA transposons and have become an important component of the genome (SanMiguel and Bennetzen, 1998; Yu *et al.*, 2002; Ma *et al.*, 2004; Ma and Bennetzen, 2004; Nagaki *et al.*, 2004; Wong and Choo, 2004; Lin *et al.*, 2005; Yang *et al.*, 2005; Vitte and Bennetzen, 2006). Retroelements are distinguished from DNA transposons by this ability to transpose via an RNA intermediate, which they convert to DNA by reverse transcription prior

to insertion into another locus in the genome, so the original copy is retained at the same genomic location. This mechanism of replication is similar to retroviruses with the difference that retrotransposons do not form the infectious particle that allows retroviruses to leave the cell and infect other cells (Boeke & Corces, 1989; Xiong and Eickbush, 1990). In most plant species, *Pseudoviridae* (Ty1-*copia* group) elements occupy a considerable proportion of the plant genomes (Moore *et al.*, 1991; Manninen and Schulman, 1993; *et al.*, Pearce 1996a, 1996b; Teo *et al.*, 2002; Bousios *et al.*, 2007). This is in marked contrast to insect and fungal species, where their populations tend to be of much lower copy number and more homogeneous. Consequently amplification and dispersion of retrotransposons play an important role in plant-genome evolution (Smyth, 1993; Flavell *et al.*, 1994; Kumar, 1996; Heslop-Harrison *et al.*, 1997).

In contrast to yeast and *Drosophila* retrotransposons, most of the plant retrotransposons examined were inactive under normal growth conditions but can be activated under various abiotic and biotic stresses (Grandbastien *et al.*, 1989; Hirochika, 1993). Hirochika (1993) showed that tobacco retrotransposon *Tto*1 can be activated by tissue culture and further enhanced by protoplast formation. The rice Ty1-*copia* group LTR retrotransposons (*Tos10*, *Tos17* and *Tos 19*) are activated under tissue culture conditions but not enhanced by protoplast formation (Hirochika *et al.*, 1996). The tobacco Ty1-*copia* group LTR retrotransposon, *Tnt1*, has been shown to be actively transcribed in protoplasts (Pouteau *et al.*, 1991). Kimura *et al.* (2001) showed that OARE-1, a Ty1-*copia* group retrotransposon is activated by abiotic and biotic stresses. On the other hand, Turcich and Mascarenhas (1994) reported that PREM-1 retrotransposons are preferentially transcribed during the pollen development of maize. Turcich *et al.* (1996) also showed that PREM-2 which

is a type of Ty1-*copia* group LTR retrotransposon is expressed in a tissue-specific manner only in the early microspores but is not detected in more mature pollen or any of the vegetative tissues.

1.2.1 Taxonomical classification, type and structure of LTR retrotransposon

In previous classifications, retrotransposons are divided into two groups: LTR (long terminal repeat) and non-LTR retrotransposons (Boeke & Corces, 1989; Xiong and Eickbush, 1990; Grandbastien, 1992). LTR retrotransposons are flanked by LTRs that provide *cis*-regulatory sequences such as promoters which are required for transcription of an RNA intermediate (Boeke & Corces, 1989). The internal sequences of these elements encode polyproteins in one or two ORFs (Gag, GAG; aspartic protease, AP; integrase, INT; reverse transcriptase, RT; and RNase H, RH; Figure 1.2) necessary for reverse transcription and integration of these elements into other locations in the genome.

Boeke & Corces, (1989) first divided LTR retrotransposons into two groups based on the arrangement of their protein-coding domains, named after the *Drosophila* retrotransposons *copia* and *gypsy*. The integrase domain is positioned at the 5' of reverse transcriptase in *copia*-like retrotransposons (Figure 1.2A), and the 3' of the RNaseH domain in *gypsy*-like retrotransposons (Figure 1.2B). The most characterized group of LTR retrotransposons among these two groups is the Ty1-*copia* group, named after the best studied elements in *Saccharomyces cerevisiae* and *Drosophila melanogaster* (Pearce *et al.*, 1997). LTR retrotransposons have been shown to be ubiquitous in plants (Voytas *et al.*, 1992) and some elements have been characterized at the sequence level (Grandbastien, 1992). The summary of full-length elements of LTR retrotransposons in different plant species is given in Table 1.1.

Phylogenetic analysis based on the similarity of their reverse transcriptase (RT) sequences has shown that LTR retrotransposons evolved from non-LTR retrotransposons by acquiring the LTR sequence during evolution and retroviruses evolved from LTR retrotransposons by acquiring envelope (*env*) gene (Xiong and Eickbush, 1990, Hansen and Heslop-Harrison, 2004). The close relationship between plant endogenous pararetroviruses (EPRV) and *Metaviridae* elements is evidenced by the same order of *pol* gene domains (Hansen and Heslop-Harrison, 2004), although the integrase domain is missing in plant EPRVs. The RT domain of plant EPRVs and *Metaviridae* elements are extremely conserved at the amino acid level while the RNAseH domains contain several significant motifs (Hohn *et al.*, 2007). Malik, Henikoff and Eickbush (2000) showed that infectious retroviruses evolved independently several times through the acquisition of the *env* gene by the retrotransposon. Although retrotransposons in plants in general do not encode any *env* gene, there are few reports of retrotransposons which contained *env* gene in their ORF (Laten *et al.*, 1998; Wright and Voytas, 1998).

The classification of retroelements has been revised by Hull (1999; 2001), where all these elements are classified into one large class, the *Retroelementopsida* (Figure 1.3) and further divided into two orders, the *Retrovirales* and the *Retrales*. All the non-LTR retrotransposons (suborder *Retroposineae*) and Group II mitochondrial introns (suborder *Retronineae*) are classified under *Retrales* (Figure 1.3). The *Retrovirales* consists of three distinct suborders: the *Orthoretrovirineae*, the *Pararetrovirineae* and the *Retrotransposineae*. The *Orthoretrovirineae* consists of all the vertebrate retroviruses (Figure 1.3). The *Hepadnaviridae* and *Caulimoviridae* are classified under *Pararetrovirineae* suborder whereas the LTR retrotransposons, both *Pseudoviridae* and *Metaviridae*, are classified under

Retrotransposineae suborder. Both *Pseudoviridae* and *Metaviridae* are further divided into three genera each (Figure 1.3).

1.2.2 *Pseudoviridae* (Ty1-*copia* group)

Pseudoviridae (Ty1-*copia* group) retrotransposons were originally divided into two genera, *Hemivirus* and *Pseudovirus* (Figure 1.2A; 1.3), which are distinguished by the primer used to initiate reverse transcription. Pseudoviruses use the terminal 3' residue of initiator tRNA to prime DNA synthesis whereas hemiviruses prime the DNA synthesis by utilizing a half-tRNA generated by cleavage within the anticodon stem-loop (Peterson-Burch and Voytas, 2002). Peterson-Burch and Voytas (2002) showed that no hemiviruses were found in plants after they characterized the coding regions of *Pseudoviridae* elements retrieved from GenBank and the complete genome sequences of five eukaryotes representing plants, humans, animals and fungi whereas another distinct lineage of plant elements containing *env*-like ORFs were found in their study and have been designated as *Agrovirus* (now under *Sirevirus*; Figure 1.2A). There are five members of *Agrovirus*: Endovir1-1 of *Arabidopsis thaliana*, Opie-2 and PREM-2 of *Zea mays*, *SIRE-1* of *Glycine max* and ToRTL1 of *Lycopersicon esculentum*. The length of the *env*-like ORF in these elements varies from 476 to 668 amino acids and, they are separated from the polyprotein ORF by distances from 27 to over 1,000 nucleotides. No conserved splice acceptor sites have been detected in the *env* gene of *Pseudoviridae* elements indicating that alternative splicing does not play a role in the expression of the *Pseudoviridae* element's *env* gene. The *env* gene of *Pseudoviridae* elements has predicted transmembrane domains (N-terminal and C-terminal TMD) but the organization of these domains varied among three members of *Agrovirus* (Peterson-Burch and Voytas, 2002). The N-

terminal TMD showed higher sequence homology compared to the C-terminal which is enriched in secondary structural features.

An obvious observation of agroviruses (now under *Sirevirus*) characteristic is that the *gag* domain has undergone considerable expansion. The *gag* domain of other *Pseudoviridae* elements has only one RNA-binding motif, RB ($Cx_2Cx_4Hx_4C$) whereas agroviruses (now under *Sirevirus*) have a second RB with variable sequence length adjacent to the RB. Multiple RBs were also observed in retroviruses, where tandem pairs help packaging of the viral genome within the virion (Peterson-Burch and Voytas, 2002). Elements of two members of the *Pseudoviridae* contain a tandem pair of RBs: Endovir1-1 (*Agrovirus*, now under *Sirevirus*) and Tpv2-6 (*Pseudovirus*). Tpv2-6 is closely related to agroviruses (now under *Sirevirus*) based on the RT phylogenetic analysis and suggested that duplication of RB may have occurred before the acquisition of the *env*-like gene in *Pseudoviridae* elements (Peterson-Burch and Voytas, 2002). The International Committee on Taxonomy of Viruses has included the agroviruses in the *Sirevirus* genus (Figure 1.3; Boeke *et al.*, 2004a; Havecker *et al.*, 2004; Syomin and Ilyin, 2005). The *Sirevirus* genus is derived from plant hosts and makes up a distinct lineage according to its RT amino-acid sequences. There are 400 *Pseudoviridae* sequences present in the Genbank database, of which *Sirevirus* make up half.

1.2.3 *Metaviridae* (Ty3-*gypsy* group)

Metaviridae (Ty3-*gypsy* group) retrotransposons encode the same type of gene domains as *Pseudoviridae* (Ty1-*copia* group) retrotransposon. The significant difference between these two classes of LTR retrotransposons is the RT and RH gene domains, those of *Metaviridae* (Ty3-*gypsy* group) preceding the INT gene domain

(Figure 1.2B). Analysis of the RT sequences of these two classes of LTR retrotransposon showed they form a different lineages and animal retroviruses have probably evolved from *Metaviridae* elements by the acquisition of an *env* gene (Xiong and Eickbush, 1990). *Env* is usually the last protein coding sequence in the POL domain next to the 3' LTR, and allows them to be packaged into viral particles permitting cell-to-cell infectivity (Temin, 1980; Xiong and Eickbush, 1990; Kumar and Bennetzen, 1999). There is some argument over whether *Metaviridae* elements actually are a type of retrotransposon or retrovirus (Bennetzen, 1996; Kumar, 1998).

The *Metaviridae* elements comprises of three genera: the *Metavirus*, the *Errantivirus* and the *Semotivirus,* also called BEL retrotransposon (Figure 1.3; Malik *et al.*, 2000; Havecker *et al.*, 2004; Syomin and Ilyin, 2004). The errantiviruses possess the *env*-like gene in their sequence whereas the metaviruses and semotiviruses do not have the *env*-like gene in their sequence (Figure 1.2B). Some of the metaviruses possess a CHROMO domain at the end of their integrase domain (chromointegrase) have been designated as the *Chromovirus* (Malik and Eickbush, 1999; Marin and Llorens, 2000; Gorinšek *et al.*. 2004). There is another distinct lineage of *Metaviridae* elements which has yet to be added to the taxonomic framework; the DIRS group (Figure 1.3; named after the founding member from *Dictyostelium discoideum*). DIRS elements have some unusual features such as lacking a protease and having a tyrosine recombinase instead of an integrase (Goodwin and Poulter, 2001; 2004). Malik *et al.* (2000) proposed that the non-viral ancestor of *Errantivirus* (*Drosophila* specific *gypsy*-like virus) acquired the *env* gene from another family of double-strand DNA insect viruses, the *Baculoviridae* due to the shared sequence features of the *env* gene from these two viruses. Moreover, baculoviruses were found to harbour LTR retrotransposons in their sequence which

provided some indication of acquisition of the *env* gene by the LTR retrotransposon (Hansen and Heslop-Harrison, 2004). The presence of the *env* gene is not limited to *Errantivirus*; genomic studies have revealed that *env*-like ORFs are also found in *Metavirus* and *Semotivirus* (Figure 1.2B; Malik *et al.*, 2000; Eickbush and Malik, 2002, ICTVdB, 2006). The best characterized LTR retrotransposons containing *env*-like genes in their sequence are the *Drosophila* errantiviruses, including *gypsy* and *ZAM* (Leblanc *et al.*, 2000; Pelisson *et al.*, 2002). The life cycle of these two elements has been characterized in detail and *gypsy* has been shown to be infectious by Kim *et al.* (1994) and Song *et al.* (1994). Recently, a *Metavirus* isolated from a human blood fluke has been shown to have an *env*-like ORF (Copeland *et al*, 2003). The *Athila* element from *Arabidopsis thaliana* which is a type of *Metavirus* also possesses an *env*-like gene (Wright and Voytas, 2002). Vicient *et al.*, 2001 reported that the *env*-like transcript of *Bagy*-2 elements from barley is spliced, similarly to what has been observed in the *env* transcript of retroviruses. Semotiviruses with *env*-like ORFs have also been observed and described in nematode, pufferfish and *Drosophila* genomes (Bowen and McDonald, 1999; Frame *et al.*, 2001). *Metaviridae* elements with other novel ORFs of unknown function upstream of its *gag* gene have been identified in rice; the *RIRE2* element of rice belongs to the *Metavirus* genus (Kumekawa *et al.*, 1999). Others with ORFs that are antisense to the genomic RNA transcript, including *RIRE2* and *Grande1* have been detected in rice and maize genomes, respectively (Martinez-Izquierdo *et al.*, 1997; Ohtsubo *et al.*, 1999). The function of these antisense ORFs is still unknown. The *Bs1* elements of maize have transduced a cellular gene sequence which is a part of a gene encoding for an ATPase (Bureau *et al.*, 1994; Jin and Bennetzen, 1994).

Genomic studies have shown that plant genomes contained medium to high copy numbers of *Metaviridae* elements (SanMiguel and Bennetzen, 1998; Balint-Kurti, 2000; Yang *et al.*, 2005; Peigu *et al.*, 2006) with different preferential chromosomal localization. The centromeric retrotransposon (CR) of *Metaviridae* has been identified in a range of grass species and preferentially transposes into the centromere in the grass chromosome (e.g. Jiang *et al.*, 2003). The CR retrotransposons have been observed in both monocot and dicot species and represent a distinct clade in the *Metaviridae* (Gorinšek *et al.*, 2004). Balint-Kurti *et al.* (2000) showed that a *Metaviridae* element is localized at the Nucleolar Organized Region (NOR) of *M. acuminata*.

1.2.4 New type of LTR retrotransposon without ORFs

Two new types of LTR retrotransposon (Terminal-repeat Retrotransposons In Miniature, TRIM and LArge Retrotransposon Derivatives, LARD; Figure 1.2C) that share some of the features of the classical LTR retrotransposon have been detected in certain plant species (Witte *et al.*, 2001; Jiang *et al.*, 2002a; b; Kalendar *et al.*, 2004; Sabot *et al.*, 2005; Antonius-Klemola *et al.*, 2006). These LTR retrotransposons contained both 5' and 3' LTR sequences with different lengths and also target site duplication (TSD) located immediately upstream of the 5' LTR and downstream of the 3' LTR (Figure 1.2C). Like the *Metaviridae*, *Pseudoviridae* and retroviruses, they also have primer binding sites (PBS), located immediately downstream of the 5' LTR and complementary to the methionine tRNA, and polypurine tracts (PPT) which located immediately upstream of the 3' LTR. The only things that distinguish them from classical LTR retrotransposon and retroviruses are that they do not encode any mobility-related ORFs in their internal domain (Figure 1.2C).

1.2.4.1 Terminal-repeat Retrotransposon In Miniature, TRIM

TRIMs (Figure 1.2C) were first detected during the sequence analysis of a genomic clone containing the *Solanum tuberosum* urease gene (Witte *et al.*, 2001). They are also found in the *Arabidopsis* genome sequence, where they were referred to as *Katydid* (*Arabidopsis* Genome Initiative, 2000).

Several TRIM elements from monocot and dicot plant species have been identified with iterative database analyses using TRIM sequences identified in the *Arabidopsis* Genome Initiative annotation effort (Witte *et al.*, 2001). TRIM elements contains a PBS and PPT and they are short in size, most about 350 bp, and contain a pair of TDR of 100-250 bp (on average < 140 bp). They do not encode any of the ORFs of the classical LTR retrotransposon. The intact elements are flanked by 5-bp TSD which is generated after the insertion into a new location in the genome. TRIM elements are present in multiple copies and also dispersed within the host chromosomes (Witte *et al.*, 2001; Antonius-Klemola *et al.*, 2006). They are ubiquitous in the plant kingdom and were found in both monocotyledonous and dicotyledonous species. The TDR sequences of TRIM elements from different plant species share some sequence homology. TRIM elements are able to transpose despite the fact that they do not have any mobility-related ORFs in their internal domain which make them the smallest active LTR retrotransposons known to date (Witte *et al.*, 2001; Antonius-Klemola *et al.*, 2006). A TRIM element (*Katydid*-A1) inserted into the mitochondrial genome of *Arabidopsis* was identified and it is truncated and degenerate (Witte *et al.*, 2001). Different types of classical LTR and non-LTR retrotransposon have been shown to insert within the internal domain of TRIMs in potato and maize, a complete *Pseudoviridae* (~ 9 kb) was inserted within the 3' TDR of a TRIM. Expressed Sequence Tag (EST) analysis of *Solanaceae* and *Fabaceae*

revealed that TRIM elements are inserted into transcribed coding regions of genes (Witte *et al.*, 2001).

Recently, usually long TRIM elements have been detected in wheat genomes and have been designated as *Veju*_L (long form) and *Veju*_S (short form). They are formed by the joining of unidentified segments which have been labelled as *unknown DNA*, flanked by the borders of the classical *Veju* TRIM element (Sabot *et al.*, 2005).

1.2.4.2 LArge Retrotransposon Derivatives, LARD

Kalendar *et al.* (2004) reported a group of retrotransposons in 13 species and four genera of the grass tribe *Triticeae* including barley, with long LTR (~ 4.4 kb in length) which were formerly designated as *Sukkula* elements. These elements have been newly designated as LArge Retrotransposon Derivatives (LARD). LARD elements contain a central domain of ~ 3.5 kb including reverse transcriptase priming sites. They are well conserved in sequence but do not contain ORFs encoding classical retrotransposon proteins but they have specific well-conserved RNA secondary structures.

LARD elements have a pair of long LTRs ranging from 3130 to 5605 bp which contain a highly conserved segment extending from LTR 5' terminus for ~ 400bp into the internal domain of the element. Alignment of the succeeding segment of the sequence showed a highly conserved stretch extending inward for ~ 1880 bp from the 3' end of the LTR, which together with the highly conserved 5' terminus sequence indicated that 50% of the LTR is conserved (Kalendar *et al.*, 2004). LARD elements contained an imperfect 6-bp terminal inverted repeat and their internal domain is highly conserved but non-coding. Similar to other LTR retrotransposons, LARD elements also contain highly conserved PBS and PPT, which suggested that

LARD may have the ability to be reverse-transcribed. DNA sequence analysis showed that LARD elements in barley and related species belong to the *Metaviridae* (Kalendar *et al.*, 2004). Homology search on four EST databases using complete LARD identified ESTs matching both LTR and internal domain of these elements, consistent with LARD elements having been integrationally active since genetic separation of the parents in the mapping crosses.

LARD elements are polymorphic in their insertion sites and serve well in recombinational map construction (Manninen *et al.*, 2000; Boyko *et al.*, 2002). They also behave like other types of LTR retrotransposon in barley (Shirasu *et al.*, 2000) and other plant species (SanMiguel *et al.*, 1996; Wicker *et al.*, 2001) which form a nest of inserted elements. The LTR probe of LARD elements uniformly labelled the chromosome arms except at the telomere, NOR and centromere (Kalendar *et al.*, 2004). Another LARD element has been detected in the rice genome and has been designated as *Dasheng* (Jiang *et al.*, 2002a; b).

1.2.5 Hotspot or insertion site preferences of LTR retrotransposon

Many studies have shown that LTR retrotransposons exhibit insertion site preference, in that they target specific parts of the genome of different eukaryotes for their integration (Labrador and Corces, 1997; Craig, 1997). This is generally known as 'Hotspots' or 'Nests' of LTR retrotransposons. Different types of LTR retrotransposon do exhibit different insertion site preferences. In some cases, some of the LTR retrotransposons are preferentially inserted into actively transcribed chromosomal regions, particularly into or near the promoter region of the genes (Hirochika *et al.*, 1996b; Garber *et al.*, 1999; Kumar and Bennetzen, 1999). On the other hand, other LTR retrotransposons target the heterochromatin, which contains

many different type of repetitive DNA. In yeast, the Ty1, Ty2 and Ty3 elements preferentially integrate into upstream regions of genes transcribed by RNA polymerase III, including tRNA, 5S rDNA and U6 genes, whereas Ty5 elements insert primarily into silenced regions of the yeast genome, including silenced mating-type cassettes and telomeric regions (Kirchner *et al.*, 1995; Zou and Voytas, 1997; Gai and Voytas, 1998).

Genome studies of eukaryotes have revealed the association of retrotransposons with other type of repetitive DNA such as tandem repeat or satellite and microsatellite DNA (Hisatomi *et al.*, 1997; Makunin *et al.*, 1999; Ramsay *et al.*, 1999; Kapitonov and Jurka, 2003; Burke *et al.*, 2003; Kojima and Fujiwara, 2004). In plants, many *Pseudoviridae* elements are concentrated in euchromatin (Garber *et al.*, 1999; Hirochika *et al.*, 1996a; b) whereas in some of the cereals *Metaviridae* elements accumulate within heterochromatin, such as near the centromere or the knob DNA in maize. Some LTR retrotransposons share a similar insertion site preference such as *Grande*, *Zeon-1* and RE-15 elements which have been shown to insert at the same site in the 180 bp repeat sequence of the knob DNA (Ananiev *et al.*, 1998a). In some cases, LTR retrotransposons and other classes of retrotransposon such as LINE elements have been shown to preferentially jump into themselves (Higashiyama *et al.*, 1995; SanMiguel *et al.*, 1998).

It is still not clear whether this biased accumulation behaviour is due to preferential insertion, a less active process of elimination from these regions, or to less fitness selection acting against elements in these regions. Labrador and Corces (1997) and Craig (1997) suggested that the selection of the insertion sites may be made by the integration complex which consists of retrotransposon-encoded, reverse transcribed cDNA copy and various host-encoded proteins.

1.2.6 Retrotransposons, genome expansion and rearrangement

LTR retrotransposons can play an important role in genome rearrangement in the host whether they are active or inactive. For instance, the *Bs1* element in the maize genome has acquired part of another gene and is then amplified through its transposition ability (Johns *et al.*, 1985; Bureau *et al.*, 1994; Jin and Bennetzen, 1994). Drouin and Dover (1987) and Loguercio and Wilkins (1998) showed the creation of intronless pseudogenes in the host genome by the rare action of retrotransposon reverse transcription and integration functions on normal cellular mRNAs. After the amplification and dispersion of these genes to new locations in the genome, they can then serve as template for evolution of new genes (Kumar and Bennetzen, 1999).

Retrotransposons can also serve as sites for unequal or ectopic recombination events due to the dispersed and high-copy-number nature of the retrotransposon. For example, the high copy number of solo LTRs present in the plant genomes provides evidence that unequal recombination between the LTR retrotransposons has occurred to create the solo LTRs reported in different plant species (Shepherd *et al.*, 1984; SanMiguel *et al.*, 1996; Chen *et al.*, 1998; Vicient *et al.*, 1999, Vitte and Panaud, 2003). Different types of genome rearrangement such as ectopic recombination and reciprocal translocation can occur in the eukaryotic genome depending on the orientation of the retrotransposon insertion and also its location within the genome (Williamson, 1983; Lim and Simmons, 1994; Kumar and Bennetzen, 1999; Vitte and Panaud, 2005). Kumar (1996) suggested that retrotransposons can play an important role in genome size determination. The small genome size of *Arabidopsis thaliana* might be due to the lack of retrotransposon accumulation (Wright *et al.*, 1996) whereas plants with large genomes such as broad bean and maize, might be the

product of retrotransposon colonization and amplification (Pearce *et al.*, 1996; SanMiguel *et al.*, 1996, SanMiguel and Bennetzen, 1998). Varies studies have shown that increase in plant genome size was due to the accumulation of retrotransposons (Hulbert *et al.*, 1990; Moore *et al.*, 1995; Avramova *et al.*, 1996; Bennetzen, 1998; SanMiguel *et al.*, 1998; SanMiguel and Bennetzen, 1998; Kalendar *et al.*, 2000; Hill *et al.*, 2005; Piegu *et al.*, 2006).

Although unequal recombination between the LTRs of an LTR retrotransposon which generated the solo LTRs can result in retrotransposon sequence elimination and maybe genome size reduction (Chen *et al.*, 1998; Vicient *et al.*, 1999), the dispersed nature of LTR retrotransposon makes this unlikely to happen without causing deleterious effects on the survival of the host due to concurrent elimination of the genes that are interspersed with LTR retrotransposon (Bennetzen and Kellogg, 1997; Kumar *et al.*, 1997).

1.2.7 Application of LTR retrotransposons for molecular marker purposes

Since the first application of an LTR retrotransposon sequence as a molecular marker by Rodhe *et al.* (1995), different types of retrotransposon-based molecular marker have been developed in different plant species. Rodhe *et al.* (1995) described a novel strategy to detect polymorphism by inverse sequence-tagged repeats (ISTR) by using PCR primers which are complementary to the highly repetitive *pseudoviridae* (Tyl-*copia* group) *Eco*RI elements to amplify the spacer region between individual *Eco*RI repetitive units. *Eco*RI elements were detected in coconut by *Eco*RI restriction digestions of genomic DNA and sequence analysis showed a high homology to *Pseudoviridae*. Rodhe (1996) extended the application of ISTR markers into a generally applicable strategy for plant and animal genome analysis.

ISTR markers have been applied to a wide range of genomic DNA from plants and animals using identical primers or primer pairs derived from *Pseudoviridae.* Modification of the technique allowed the nonradioactive detection of DNA fragments in the sequencing gel without membrane transfer. The genome or species-specific bands could then be re-isolated from the gel, re-amplified, sequenced and converted to PCR-based sequence-tagged site (STS) analysis. The comparative study of the discrimination capacity of AFLP and ISTR markers by Demey *et al.* (2004) showed that ISTR's generated more polymorphic markers than AFLP's and had a better capacity to quantify the genetic diversity and also showed superior discriminatory capacity.

Waugh *et al.* (1997) described a modified version of the AFLP marker technique using both AFLP adaptor primers and retrotransposon primers, based on *BARE-1* LTR sequences, to amplify the flanking region of LTR retrotransposons. This technique has been designated as sequence-specific amplification polymorphisms (S-SAP). Genetic distribution of *BARE-1* elements in the barley genome has been revealed by using this technique. S-SAP amplifies the region between each insertion of a particular retrotransposon and a neighbouring frequently occurring restriction site (Bousios *et al.*, 2007). New banding patterns will be observed in the S-SAP profiles of different plants species if primers are designed based on the retrotransposon sequence of an active element. Bousios *et al.* (2007) have developed the S-SAP marker based on the *pseudoviridae* (Ty1-*copia* group) LTR retrotransposon which are highly polymorphic and present in high copy number that are conserved throughout the *Agavaceae.* Ellis *et al.* (1998) reported a modified version of the S-SAP marker technique by using primers designed from the polypurine tract (PPT) instead of LTR sequence of *PDR1* element, a *pseudoviridae*

(Ty1-*copia* group) LTR retrotransposon of pea. Most of the *PDR1* S-SAP products were shown to be derived from *PDR1* elements in the pea genome. The *PDR1* S-SAP markers have been shown to be more informative than AFLP and RFLP markers and are distributed throughout the genome. These characteristics make them suitable for integrating genetic maps derived from related crosses (Ellis *et al.*, 1998).

Flavell *et al.* (1998) developed a molecular marker technique that is based upon the polymorphic *PDR1* retrotransposon insertion site in *Pisum sativum*. This technique involved PCR amplification using primers derived from *PDR1* retrotransposon LTR sequence and its flanking regions and has been called retrotransposon-based insertion polymorphisms (RBIP). This technique requires sequence information of retrotransposon flanking region for primer design and then produces a co-dominant marker where different allelic states at one locus can be revealed by PCR amplification. This technique can be applied to any retrotransposon in any plant species if the retrotransposon flanking sequence information is known. Provan *et al.* (1999) have developed a PCR-based assay using a combination of two classes of repetitive DNA found in eukaryotic genomes, namely *pseudoviridae* (Ty1-*copia* group) LTR retrotransposon and simple sequence repeats (SSR) to amplify the flanking region between these two type of repetitive DNAs. This technique enables the mapping of seven markers to four different chromosome arms using a barley double haploid mapping population and is also able to generate both dominant and co-dominant genetic markers for mapping and diversity studies.

Kalendar *et al.* (1999) described two novel molecular markers using primers designed from LTR sequences and SSR repeat sequences. The first technique is Inter-Retrotransposon Amplified Polymorphism (IRAP) which uses outward primers designed from the LTR sequence of the *BARE-1* retrotransposon. This technique

allows PCR amplification with a single LTR primer dependent on the LTR retrotransposon insertion orientation. Most of the LTRs will insert in a head-to-head, head-to-tail and tail-to-tail orientation in the host genome and the use of single IRAP primer allows amplification of nested insertion events (≤ 2 kb). The LTR primers of the barley *BARE-1* retrotransposon have been applied to *Musa* species and showed high levels of polymorphism, which allowed the identification of a B genome specific band (Teo *et al.*, 2005). The second technique developed by Kalendar *et al.* (1999) is Retrotransposon-Microsatellites Amplified Polymorphism (REMAP) where the first primer was designed from the LTR sequence of *BARE-1* retrotransposon and the second primer was designed from the SSR unit such as (AG)n or (CT)n(AT)n. REMAP markers allow amplification of the region between LTR retrotransposons and SSRs and also provide information regarding the insertion preference of LTR retrotransposons. Markers generated by both techniques can be used for mapping and genetic diversity studies in plants.

1.3 Tandem repetitive DNA

The second class of repetitive DNA is the tandemly repeated or satellite DNA (Schmidt and Heslop-Harrison, 1998; Heslop-Harrison, 2000a; b; Heslop-Harrison *et al.*, 2003; Schwarzacher, 2003). Tandem repeat DNA is composed of identical or nearly identical repeat units of various lengths that are repeated many thousand of times. Tandem repeat DNAs with large repeat subunits, are often called satellite DNA, that is derived from the centrifugation of DNA in buoyant density gradients which results in a minor 'satellite' or shoulder in the DNA density histogram due to large number of DNA fragment with a different G+C content to the main DNA (Kit, 1961; Bond *et al.*, 1967; Corneo *et al.*, 1967; 1968; Arrighi *et al.*, 1970). The

terminology of satellite DNA has been broadened to include similar repetitive sequences that don't necessarily form 'satellite' bands. Satellite DNAs form large arrays and are abundant at the centromere, telomere, NORs and constitutive heterochromatin in eukaryotes. The repetitive nature of satellite DNAs in the centromere and constitutive heterochromatin has limited the whole genome sequencing effort to focusing only on euchromatin regions or regions with low copy number of satellite DNA.

Satellite DNAs usually have repeat units of 160-180 bp or 320-360 bp, which might be due to the natural fit of DNA sequences of these lengths coiled around the nucleosome core. They have been postulated to play a key role in stabilizing DNA packaging and higher order chromatin condensation (Heslop-Harrison, 2000a; Schwarzacher, 2003).

1.3.1 Centromeric tandem repeat DNA

In the majority of eukaryotes, centromeres are embedded in long tracks of highly repetive DNA with satellite DNA often the major DNA component (Csink and Henikoff, 1998). For example, the α-satellite, a 171-bp tandem repeat flanked by 0.5 Mbp of diverged satellite DNA from the same family, is located in the centromeres of all human chromosomes (Schueler, *et al.*, 2001; Nagaki *et al.*, 2003a, b) and can function as centromere in human artificial chromosomes (Harrington *et al.*, 1997; Ikeno *et al.*, 1998; Henning *et al.*, 1999).

In plants, the centromeres of *Arabidopsis thaliana* are among the most studied, and have been mapped genetically using tetrad-based genetic mapping technology (Copenhaver *et al.*, 1999). The highest copy number satellite DNA in *Arabidopsis thaliana* is the pAL1 repeat (Martinez-Zapater *et al.*, 1986; Maluszynska and Heslop-

Harrison, 1991; Round *et al.*, 1997; Jackson *et al.*, 1998; Kumekawa *et al.*, 2000; 2001), which is a 180-bp satellite DNA family that may be interrupted by LTR retrotransposons (Fransz *et al.*, 2000; Kumekawa *et al.*, 2000; 2001) and other repeats (Brandes *et al.*, 1997a). The chromosomal location of the pAL1 repeats coincides with the centromeric H3 histone (Talbert *et al.*, 2002). Two highly conserved satellite DNAs were reported in the centromeric region of grass chromosomes (Aragon-Alcaide *et al.*, 1996; Jiang *et al.*, 1996) which are derived from *Metaviridae* elements (Miller *et al.*, 1998; Presting *et al.*, 1998; Langdon *et al.*, 2000; Cheng and Murata, 2003). Satellite DNAs have been isolated from many different plant species including rice (Dong *et al.*, 1998; Nonomura and Kurata, 1999), barley (Presting *et al.*, 1998) and maize (Ananiev *et al.*, 1998a; b) and are often associated with centromeric regions (Alfenito and Birchler, 1993; Harrison and Heslop-Harrison, 1995; Kaszas and Birchler, 1996; 1998; Ananiev *et al.*, 1998a, b; Miller *et al.*, 1998; Nagaki *et al.*, 1998; 2003a; Francki *et al.*, 2001; Gindullis *et al.*, 2001; Hudakova *et al.*, 2001; Kishii *et al.*, 2001; Page *et al.*, 2001; Saunders and Houben, 2001).

In other cases, chromosomes that only contain small number of satellite DNA sequences have been reported (rice CentO; Cheng *et al.*, 2002). This might be due to the centromeric misdivisions that give rise to telocentric chromosomes with reduced copy number of CentO (Cheng *et al.*, 2002). The low copy number of CentO repeats on *Cen8* allows the contiguous tilling of BAC clones that span *Cen8* and also cytological mapping the CentO arrays to a region in the BAC contig (Nagaki *et al.*, 2004).

1.3.2 Telomeric tandem repeat DNA

Telomeres are DNA and protein structures that interact with each other to form a complex structure that protects the chromosomes of all living organisms. In general, telomeric DNA is composed of short repeats (5 to 26 bp) typically with G-clusters, and is arranged in large tandem arrays (Klobutcher *et al.*, 1981; Henderson and Blackburn, 1989; Wellinger *et al.*, 1993). They are recognized by telomerase, a cellular ribonucleoprotein reverse transcriptase that is responsible for elongation of one strand of the telomere (McEachern *et al.*, 2000). Telomeric tandem repeat DNA appears to consist of a *cis*-acting sequence which is necessary for telomere function. Telomeric tandem repeat DNAs are very conserved throughout all eukaryotes, but show some different monomeric motifs in animal and plant species.

Several groups of animal telomeres consist of telomeric tandem repeat DNA with conserved monomeric motif of TTAGGG (Meyne *et al.*, 1989). This indicates that they shared a common ancestor over 400 million years ago. It is still not clear how this repeat motif can be conserved over such a long time. The possible explanation suggested by McEachern *et al.* (2000) was the need for telomeric tandem repeat DNA to act as specific binding sites for a number of different proteins or to form G-quartet or other noncanonical base interactions. Furthermore, constraints on the functioning of telomerase may also play an important role in telomeric tandem repeat DNA sequence conservation (Gilley *et al.*, 1995; Gilley and Blackburn, 1996; Cohn *et al.*, 1998; d'Adda di Fagagna *et al.*, 1999).

In plants such as *Arabidopsis thaliana*, the telomeric tandem repeat DNA with the repeat motif of TTTAGGG was observed by Richard *et al.* (1992). Although it is generally thought that the *Arabidopsis thaliana*-type telomeric tandem repeat motif cap the end of all plant chromosomes (Cox *et al.*, 1993), studies on telomere

sequences of species from different families showed that telomeres may also be capped by other sequences including retroelements (Fay and Chase, 1996; Pich *et al.*, 1996; Pich and Schubert, 1998; Adams *et al.*, 2000; 2001; Sỳkorova *et al.*, 2003; 2006). In rice, the Telomeric Associated Sequence (TAS) without the typical telomeric tandem repeat motif were hybridized to the termini of one or several rice chromosome arms and showed a high level of polymorphism between two rice varieties (Ashikawa *et al.*, 1994). Vershinin *et al.* (1995) showed that two tandem repeat DNA families, pSc200 (550 bp) and pSc250 (379 bp) with repeat motifs of several hundred base pairs containing degenerate telomere repeat motifs were located in mainly single prominent bands very close to the end of ~18 chromosome arms. The TASs are often more structurally variable and polymorphic than other regions of plant genomes (Contento *et al.*, 2005). They are highly repeated and arranged in a location separated from telomeres by a spacer region or directly adjacent to them (Ganal *et al.*, 1991; Fajkus *et al.*, 1995; Vershinin *et al.*, 1995; Contento *et al.*, 2005)

Not all telomeric tandem repeat DNA are located close to the termini of the chromosomes. Richards *et al.* (1991) reported that the centromeric region of *Arabidopsis thaliana* chromosome 1 consists of sequences similar to telomeric tandem repeat motifs. Furthermore, telomere repeats at the centromeric regions of human chromosome 2 and also at intercalary positions of several other vertebrate chromosomes were observed (Allshire *et al.*, 1988; Meyne *et al.*, 1990). Indeed, some of the earliest characterized centromeric satellite DNA was found to be composed of the human telomeric tandem repeat motif, TTAGGG (Southern, 1970; Fry and Salser, 1977).

1.3.3 Ribosomal DNA repeats

Another class of tandem repeat DNA with high copy number in eukaryotic species is the ribosomal DNA: 45S and 5S rDNA. They are the most accessible genes for gene expression study at the molecular, structural and ultra-structural levels (Heslop-Harrison, 2000a). Both 45S and 5S rRNA genes encode the structural RNA component of the ribosome and they are highly repetitive forming large arrays with one or more loci per chromosome set (Heslop-Harrison, 2000b). The 45S rRNA genes are composed of repeat units in tandem arrays and consist of 18S, 5.8S and 26S rRNA genes with the transcribed and non-transcribed spacers (Beebee and Rowe, 2004). The 45rRNA genes are about 10 kbp in length and are repeated many thousands of times in plant genomes. The 45 rRNA gene loci are located at the Nucleolar Organizing Region (NOR) of the chromosomes, although some rDNA may not organize a nucleolus in some or all tissues. The 5S rRNA genes also form large arrays although generally smaller than the 45S rDNA array (Schwarzacher, 2003).

Plant ribosomal RNA genes are highly conserved in their coding sequence. However, DNA sequence differences found in the Internal Transcribed Spacers (ITS) and Non-Transcribed Spacers (NTS) have been used as molecular markers to study the diversity of different organisms and also for phylogeny reconstruction (Beebee and Rowe, 2004). Differences in chromosomal localization of rDNA arrays have been applied for understanding evolutionary trends in many different plant genomes including in *Triticeae* (Castillo and Heslop-Harrison, 1995; Flavell *et al.*, 1998; Taketa *et al.*, 1999) and *Musa* (Doleželová *et al.*, 1998; Osuji *et al.*, 1998; Doležel *et al.*, 2004). Physical mapping of the 45S and 5S rRNA gene loci have been studied in many different organisms. For instance, Raina *et al.* (2001) physically mapped the

45S and 5S rRNA gene loci on *Vicia* species and their allied taxa using 45S and 5S rDNA clones from wheat, pTa71 and pTa794, respectively, with fluorescent *in situ* hybridization. The variability in copy number, size and chromosomal location of two ribosomal DNA probes have distinguish the species within the *Vicia bithynica* and *Vicia hybrida*.

Eukaryote genomes contain active and inactive or silent rDNA loci. The transcription of active and silent rRNA genes is usually coexists in most genomes (López-León *et al.*, 1999). Mukai *et al.* (1991) showed that 'Chinese Spring' wheat has 10 rDNA loci but only those located on chromosome 6B and 1B are normally actively involved in nucleolus formation (Jordan *et al.*, 1982; Martini and Flavell, 1985). The grasshopper, *Stauroderus scalaris*, has unusually high amounts of inactive ribosomal DNA (López-León *et al.*, 1999).

1.3.4 *Musa* repetitive DNA

Only a few studies of repetitive sequences in *Musa* have been reported (Doleželová *et al.*, 1998; Osuji *et al.*, 1998; Balint-Kurti *et al.*, 2000; Teo *et al.*, 2002; Valárik *et al.*, 2002; Doležel *et al.*, 2004, Bartoš *et al.*, 2005). Applications of repetitive DNA as probes for molecular cytogenetic studies have enabled the understanding of the long range organization of chromosome structure in *Musa*.

Several repetitive DNAs have been isolated from *Musa acuminata* and *Musa balbisiana* (Valárik *et al.*, 2002) and have been used as probes for FISH. Four different types of genome distribution patterns were observed for twelve repetitive DNA clones (Valárik *et al.*, 2002). Radka2 sequences were localized on eight chromosomes of *M. acuminata* 'Calcutta 4' and six chromosomes of *M. balbisiana* 'Tani'. Radka1, 7 and 14 clones showed strong signal at the NOR region on one pair of chromosomes in both 'Calcutta 4' and 'Tani'. The third type of distribution was

observed for clones Radka3, 5, 6, 8, 9 and 12, which preferentially localized to the centromeric regions of all chromosomes and exhibited dispersed distribution throughout the genome with the exception of four pairs of chromosomes, where only one arm was labelled. Radka4 and Radka10 which showed preferential location on centromeric regions of all chromosomes with otherwise weak dispersed signal represented the fourth type (Valárik *et al.*, 2002).

Osuji *et al.* (1998) used 45S and 5S rDNA clones isolated from wheat, and telomeric repeat probes, to study the chromosomal location of these repetitive DNAs on *Musa* chromosomes. Two 45S rDNA loci were observed on the short arm of the nucleolar organizing chromosomes. FISH showed that chromosomal localization of 45S rDNA loci is useful for quick ploidy level determination even using the interphase nuclei from slowly growing tissue culture material. Different numbers of 5S rDNA loci were observed in *Musa*, with 8 loci detected on *Musa acuminata* 'Calcutta 4' and 6 loci detected on *Musa balbisiana* 'Butohan 2'. Triploid lines showed 6 – 9 loci of 5S rDNA with varying signal intensity, whereas for diploid hybrids, 5 – 9 were detected and the tetraploid line had 11 loci. Furthermore, the chromosomal locations of 5S and 45s rDNAs in different *Musa* taxa has been studied (Doleželová *et al.*, 1998; Doležel *et al.*, 2004, Bartoš *et al.*, 2005).

1.4 Aims

The overall aim of my work was to investigate the genome organization of *Musa* species using different molecular biological, cytogenetic and bioinformatic techniques. The small genome size of 500-600Mp, the high commercial value of some species and its important phylogenetic position as a genus within the *Zingiberales*, a sister group to the *Poales* (grasses) within the Commelinids, makes

Musa an important species group to be chosen as model plant for fundamental and comparative genomic studies. Furthermore, within the genus *Musa*, a great number of species and inter- and intra-species hybrids are present displaying wide morphological and genomic diversity that underpin the study of genome organization, evolution and function.

More specifically, the objectives of the current work are

1. The study of the long range organization, diversity and phylogenetic relationships of different types of repetitive DNA such as LTR retrotransposons (Chapter 3) and tandemly organized repeats (Chapter 4) using PCR amplification, cloning and sequencing of repetitive DNA sequences. Extensive bioinformatic analysis of the DNA and amino-acid sequences from my clones and those from BAC sequences in published databases will allow a detailed comprehensive survey of *Musa* repetitive sequences.
2. Understanding the co-evolution of repetitive DNA and *Musa* genomes such as genome expansion based on insertion site polymorphisms of repetitive DNA. (Chapter 5). Also of interest are retrotransposon copy numbers in *Musa* genomes and their application as molecular markers for distinguishing different *Musa* taxa.
3. The investigation of the chromosomal distribution and function of repetitive DNA sequences using fluorescent *in situ* hybridization (FISH) and immuno-cytochemistry with anti-5-methylcytosine antibodies (Chapter 6). These have been combined with DNA methylation sensitive PCR and isolation of RNA transcripts as well as fine scale analysis using extended DNA fibers
4. Finally, a unifying model of *Musa* genome organization, amplification and evolution will be put together (Chapter 7).

Table 1.1: *Pseudoviridae* and *Metaviridae* elements found in different plant species.

Species	*Pseudoviridae*	*Metaviridae*	References
Nicotiana sp	*Tnt1, Tto1-Tto3*	-	Grandbastien *et al.*, 1989; Hirochika, 1993
Arabidopsis sp	*Ta1-Ta3, AtC1-AtC18, AtRE1, Evelknievel, Meta-1*	*Athila, Athila1-2, Athila2, Athila3, Tat1, Tat4, Gimli, Gloin, Legolas, Tft1, Tma1-Tma4*	Voytas and Ausubel, 1988; Konieczny *et al.*, 1991; Pelissier, *et al.*, 1995; Henikoff and Comai, 1998; Wright and Voytas, 1998; Kapitonov and Jurka, 1999; Marín and Lloréns, 2000; Terol *et al.*, 2001
Zea sp	*PREM-2, Bs1, Stonor, Opie, Ji, Fourf, Victim*	*PREM-1, Zdel, Huck, Cinful, Tekay, Grande-1, Grande-zm, CentA, CRM, Reina, Zeon-1, magellan*	Shepherd *et al.*, 1984 ; Jin and Bennetzen, 1989 ; Varagona *et al.*, 1992 ; Purugganan and Wessler, 1994 ; Turcich and Mascarenhas, 1994 ; Jiang *et al.*, 1996 ; SanMiguel *et al.*, 1996 ; Turcich *et al.*, 1996 ; Vicient and Martínez-Izquierdo, *et al.*, 1997 ; Ananiev *et al.*, 1998b ; Miller *et al.*, 1998 ; SanMiguel and Bennetzen, 1998 ; Meyers *et al.*, 2001 ; Zhong *et al.*, 2002
Oryza sp	*Tos3, Tos5, Tos10, Tos17, Tos19, RIRE1, Osr1-Osr24*	*RIRE2, RIRE3, RIRE7, RIRE8/Osr33, Kangourou, Wallabi, Dingo, CRR, Osr28-Osr36, Osr38-Osr42*	Hirochika *et al.*, 1992 ; Hirochika *et al.*, 1996 ; Nakajima *et al.*, 1996 ; Noma, *et al.*, 1997 ; Ohtsubo *et al.*, 1999 ; Kumekawa *et al.*, 1999 ; 2001 ; McCarthy *et al*, 2002 ; Nagaki *et al.*, 2003a ; Piegu, *et al.*, 2006
Triticum sp	*Wis-2, Angela,*	*Daniela, Erika-1, Fatima, Latidu, Nusif, Wham*	Moore *et al.*, 1991; Wicker *et al.*, 2001
Solanum sp	*Tst-1*	*Sore1*	Camirand *et al.*, 1990; Tek *et al.*, 2005
Pisum sp	*PDR-1*	*Cyclops*	Lee *et al.*, 1990; Chavanne *et al.*, 1998
Hordeum sp	*BARE-1, BARE-2, Thv19*	*Cereba, BAGY-1, BAGY-2*	Manninen and Schulman, 1993; Panstruga *et al.*, 1998; Presting *et al.*, 1998; Gribbon *et al.*, 1999; Shirasu *et al.*, 2000; Tanskanen *et al.*, 2007
Lilium sp	-	*Del1-46*	Smyth *et al.*, 1989
Glycine sp	*SIRE*-1, *Tmgr*	*Diaspora*	Laten, *et al.*, 1998; Vicient *et al.*, 2001; Yano *et al.*, 2005
Pinus sp	-	*IFG*	Kossack and Kinlaw, 1999

Iris sp	-	*IRRE1*	Kentner *et al.*, 2003
Phaseolus sp	*Tpv2*	-	Garber *et al.*, 1999
Cajanus sp	*Panzee*	-	Lall *et al.*, 2002
Lycopersicon sp	*Lere1*, *ACO1RPT*, *Retrolyc1*	-	Blume *et al.*, 1997; Mao *et al.*, 2001; Costa *et al.*, 1999
Ipomoea sp	-	*RTip1*	Hisatomi *et al.*, 1997
Silene sp	-	*Retand*	Kejnovsky *et al.*, 2006
Cucumis sp	-	*CURE*	van Leeuwen *et al.*, 2003
Vitis sp	*Tvv1*, *Vine-1*	-	Pelsy and Merdinoglu, 2002; Verriès *et al.*, 2000
Avena sp	*OARE-1*	-	Kimura *et al.*, 2001

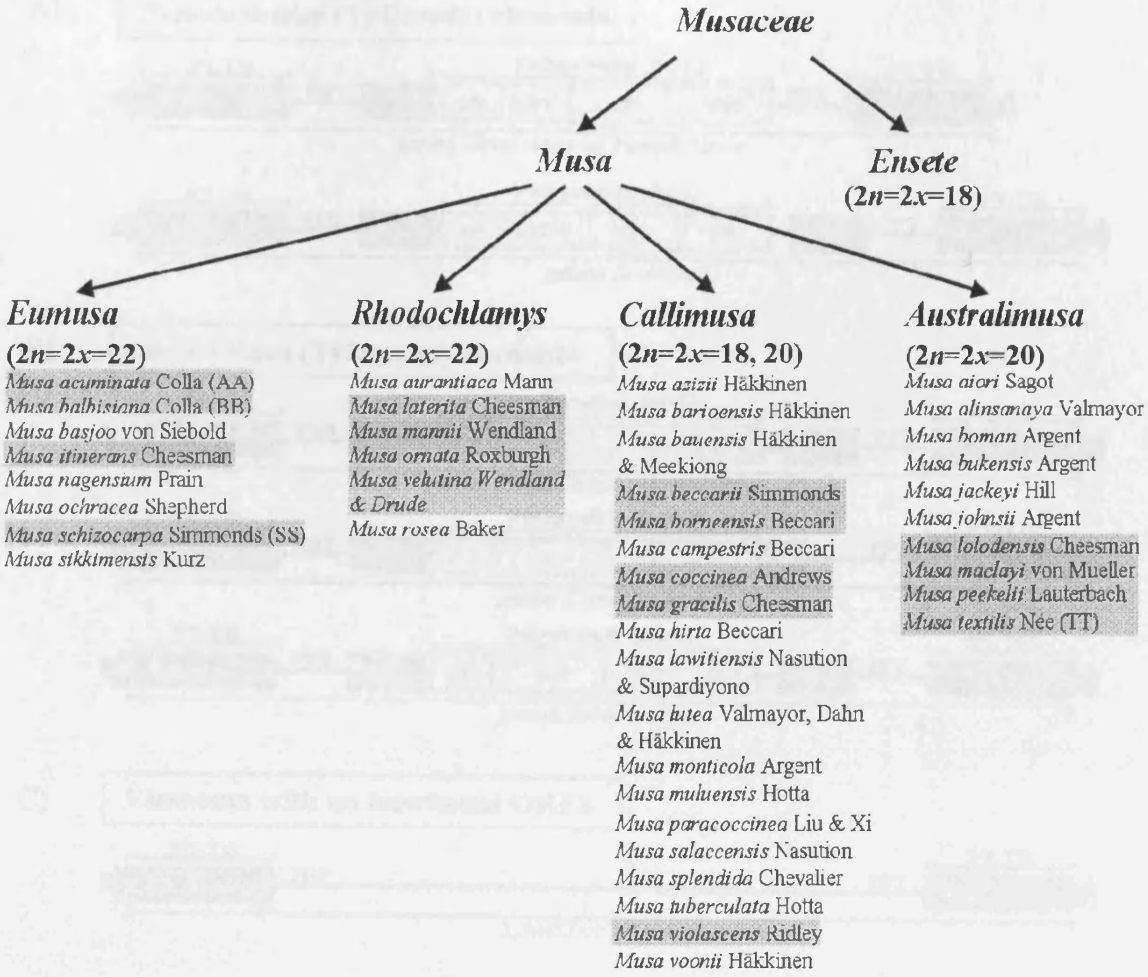


Figure 1.1: *Musa* Linnaeus species classification based on morphological traits (Simmonds and Shepherd, 1955; Stover and Simmonds, 1987; Silayoi and Chomchalow, 1987; Espino and Pimental, 1990; Valmayor *et al.*, 1991). The chromosome numbers of each section and the genome constitutions of four species (AA, BB, SS and TT) are shown. *Musa* species that are used in this study are highlighted. Fourteen new wild *Musa* species mainly in *Callimusa* section have been discovered and added into this classification (Valmayor, 2001; Häkkinen and De Langhe, 2001; Häkkinen, 2001; 2003; Häkkinen and Sharrock, 2002; Valmayor *et al.*, 2004; Häkkinen, 2004a, b; Häkkinen, 2005; Häkkinen and Meekiong, 2004; 2005; Häkkinen *et al.*, 2005; Häkkinen, 2006a, b).

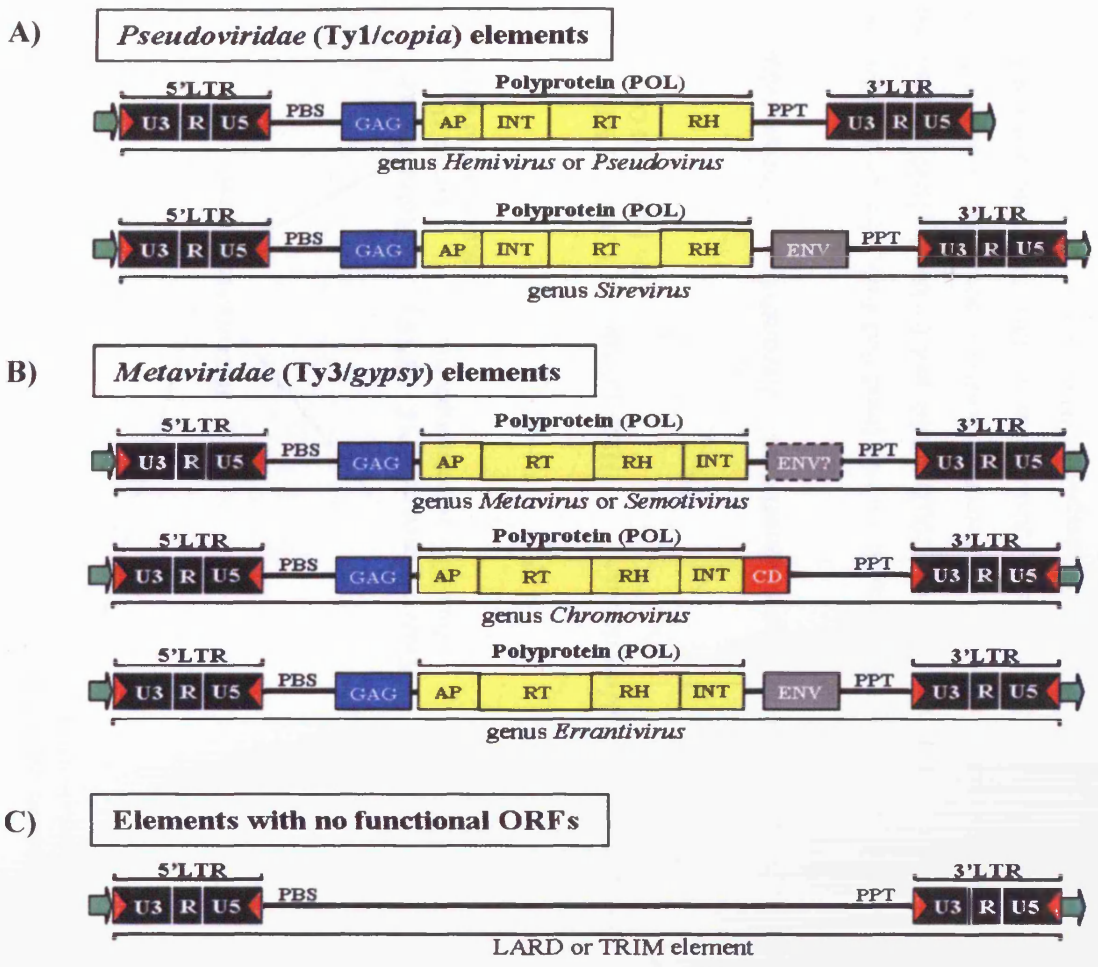


Figure 1.2: Gene domain organization and structure of LTR retrotransposons, **A** *Pseudoviridae*, **B** *Metaviridae* and **C** elements with no functional ORFs. LTR retrotransposons contain Long Terminal Repeats (LTR) at each end. The signals for initiation and termination of transcription are located at the U3 (3' untranslated region), R (repeated RNA), and U5 (5' untranslated region) regions within the LTRs. The GAG domain consists of a capsid-like protein and a nucleic acid binding moiety. The envelope (ENV) domain is found in *Pseudoviridae*, genus *Sirevirus* and in *Metaviridae*, genus *Errantivirus*. Some members of *Metavirus* and *Semotivirus* also contain ENV-like (ENV?) domain (see text). Genus *Chromovirus* contains CHROMO domain (CD) at the end of its INT. The other genes within the retrotransposons: INTegrase (INT), Aspartic Protease (AP), Reverse Transcriptase (RT), and RNAaseH (RH), form a large domain often as a single ORF, called POLyprotein (POL). Note that the gene order in the POL region is different for *Pseudoviridae* and *Metaviridae*. Other sequence features are the Primer Binding Site (PBS), the PolyPurine Tract (PPT), small Inverted Repeats (IR, red arrowheads) and flanking genomic DNA target Direct Repeat (DR, green arrows). These figures are modified from Kumar and Bennetzen, (1999) and are not drawn to scale but typical LTR retrotransposons range from a few kb up to 15 kb in size, with LTRs from a few hundreds bp to few kb.

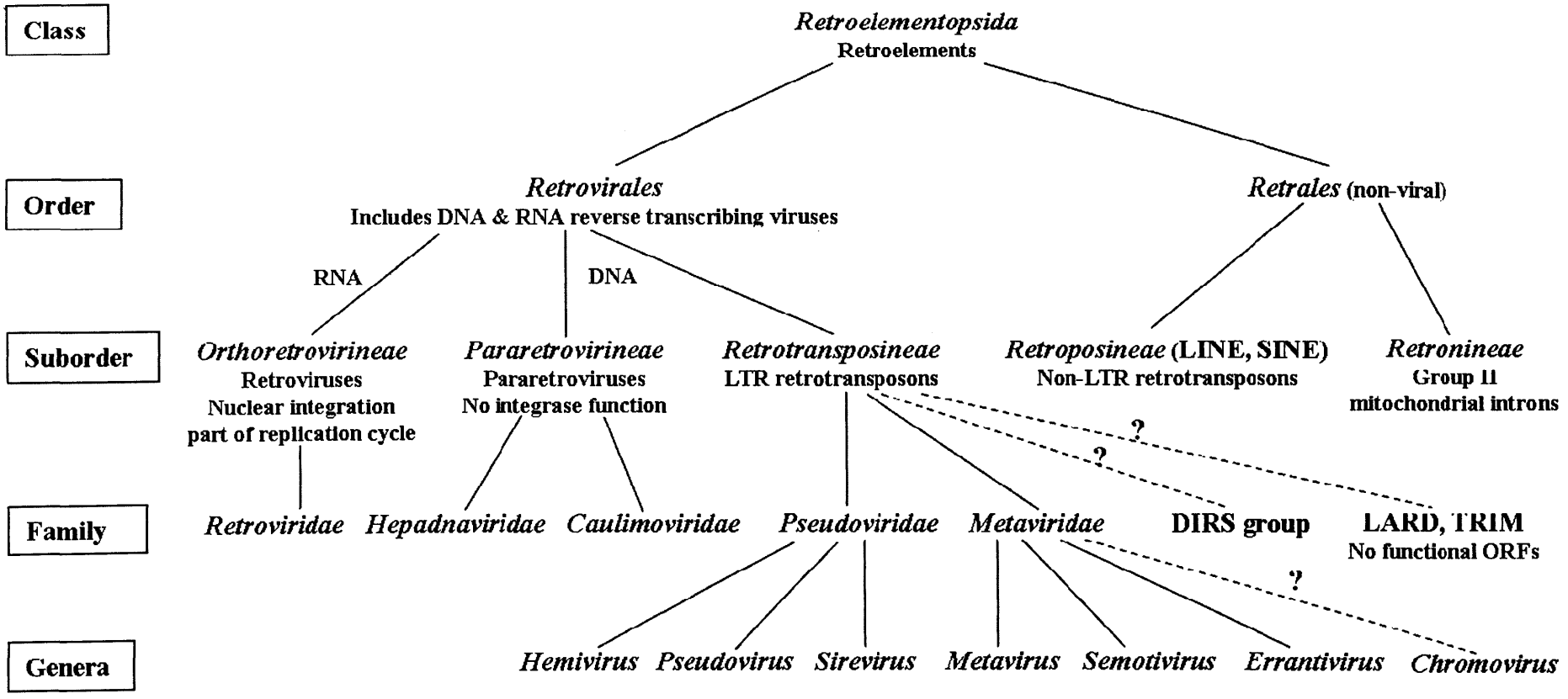


Figure 1.3: Relationship of retroelements (LTR and non-LTR containing elements), retroviruses and pararetroviruses (modified from Hansen and Heslop-Harrison, 2004). This classification is based on reports from Hull (1999; 2001), Peterson-Burch and Voytas (2002), Malik and Eickbush (1999), Marin and Llorens (2000), Havecker *et al.* (2004) and ICTVcB (2006). For gene domain and structure organization of *Retrotransposineae*, see Figure 1.3. The DIRS group, LARD and TRIM elements are yet to be added within the taxonomic framework. The relationship of *Chromovirus* to other three genera is not clear. The classification described above does not correspond directly with the phylogenetic relationship of the retrotransposons (Havecker *et al.*, 2004).

2.0 CHAPTER II: MATERIALS AND METHODS

2.1 Materials

2.1.1 Plant Materials

One hundred and nine *Musa* accessions were used in the present study. Their ploidy levels and genomic constitutions are given in (Table 2.1 and 2.2).

Leaf samples from these bananas were from the germplasm collection of Dr. Siti Hawa Jamaluddin of MARDI, Serdang, Malaysia; Professor Rofina Yasmin Othman of University Malaya, Petaling Jaya, Malaysia; Mr. Markku Häkkinen of University of Helsinki Botanical Garden, Helsinki, Finland; Professor Ashalatha Nair of Kerala Agricultural University, Kerala, India and International Transit Centre (ITC), Leuven, Belgium.

2.1.2 Solutions and media

1x TAE buffer
40 mM Tris-acetate
1 mM EDTA

6x loading dye
0.25% Bromophenol blue
0.25% Xylene cyanol FF
60% Glycerol

LB Broth Media
1% (w/v) Bacto-Tryptone
0.5% (w/v) Yeast Extract
86 mM NaCl
20 mM glucose

LB agar
1.5% (w/v) Bacto-agar
20 mM glucose

20x SSC (saline sodium citrate)
0.3 M NaCl
0.03 M sodium citrate

10x PBS (phosphate-buffered saline)
1.3 M NaCl
70 mM Na_2HPO_4
30 mM NaH_2PO_4

10x enzyme buffer
40 mM citric acid
60 mM tri-sodium citrate

10x TE buffer
100 mM Tris(tris-hydroxymethylamino-methane)-HCL, pH8.0
10 mM EDTA (ethylene-diamine-tetra-acetic acid), pH8.0

2.2 Methods

2.2.1 Extraction of total genomic DNA

Total genomic DNA was extracted from young leaves using a modified cetyltrimethylammonium bromide (CTAB) method (Ausubel *et al.*, 1995) with RNase A treatment as follows. 10 g of tissue was ground to powder in liquid nitrogen and transferred into 20 ml CTAB buffer, preheated at 65°C 2% (w/v) CTAB; 100 mM Tris, pH 8.0; 20 mM EDTA, pH 8.0; 1.4 M NaCl; 6% (w/v) PVP-40 and 5% (v/v) β-mercaptoethanol) and mixed gently. The mixture was incubated at 65°C for 2 hrs. Two thirds of the volume of wet chloroform (ratio of chloroform to isoamyl alcohol is 24:1, v/v) was added, gently mixed and centrifuged at 12,000 rpm, room temperature for 15 min to pellet down the cell debris. The clear top aqueous phase was transferred to a new tube. Two thirds of the volume of prechilled isopropanol was added to the supernatant, followed by incubation at room temperature overnight. Precipitated DNA was pelleted down by centrifugation at 12000 rpm, 4°C for 15 min.

After washing with wash buffer (76 % ethanol and 10 mM ammonium acetate), the pellet was dissolved in 1 - 2 ml of 1x TE buffer and RNase A (0.1 mg/ml) was added and incubated at 37°C for 30 min. The total genomic DNA samples were kept at –20°C until needed. Total genomic DNA was quantified by spectrophotometer (Helyos) using a 1:40 sample dilution, (total volume 200 μl). The ratio 260/280 was used to check the quantity, quality and purity of isolated total genomic DNA. Two microliters of total genomic DNA was loaded on a 0.7% (w/v) agarose (Molecular Biology Grade, Melford) gel with loading dye and run in 1x TAE buffer and detected by ethidium bromide staining for checking integrity and purity. HyperLadder™ I

(BOLINE) DNA ladder was used as marker. The agarose gel was observed using GENE FLASH (Syngene).

2.2.2 Polymerase Chain Reaction (PCR)

Different types of repetitive DNA (LTR retrotransposon and tandem organized repeats) were amplified from *Musa* total genomic DNA using specific and degenerate primers. Primers and annealing temperature are described in the result chapters. The PCRs were performed in a TGradient Thermocycler (Biometra) in 25 µl reaction typically containing 1x PCR buffer (10 mM Tris-HCl, pH 9.0, 50 mM KCl and 0.1% Triton® X-100 (Promega)), 1.5 to 2.0 mM $MgCl_2$, 200 µM of dNTPs (Bioline or YorkBio), 10 pmoles of each primer (forward and reverse), 1.5 U of *Taq* DNA polymerase in storage buffer A (50 mM Tris-HCl, pH 8.0, 100 mN NaCl, 0.1 mM EDTA, 1 mM DTT, 50% Glycerol and 1% Triton® X-100 (Promega)) and 50 – 100 ng of DNA template or water as control. The *Taq* DNA polymerase in storage Buffer A must be used with the reaction buffer provided containing Triton® X-100. Use of other reaction buffers without 0.1% Triton® X-100 will result in inactivation of the *Taq* DNA polymerase. The PCR cycle's parameters were (unless specifically mentioned in the result chapters): 95°C for 5 minutes, 30 cycles of 95°C for 30 seconds, T_a °C for 30 seconds and 72°C for 2 minutes; followed by 72°C for 10 minutes.

2.2.3 Gel electrophoresis and purification of PCR products

After the PCR reactions, 5 µl of 6x gel-loading dye were added directly into PCR products (25 µl) and mixed gently by using micropipetters. The mixed PCR products were separated by 1 - 2% (w/v) agarose gel electrophoresis, premixed with

0.5 μg/ml) ethidium bromide, in 1x TAE buffer at 5V/cm for about 30 minutes. The agarose gel was observed using GENE FLASH (Syngene). DNA ladder, Hyperladder I (Bioline) and Q-step 2 (YorkBio), were used to obtain the molecular weight and the concentration of PCR products. Selected PCR bands were excised from the agarose gel by using a sterile blade and the PCR products were purified from the agarose gel using QIAquick Gel Extraction Kit (Qiagen) according to manufacturer's instruction.

2.2.4 Cloning of PCR products

Purified PCR products were ligated to pGEM-Teasy commercially available vectors by using pGEM-T Easy Vector System I (Promega) according to manufacturer's instructions. The *Taq* DNA polymerase generates single base 3' adenine overhangs at the end of extended products which are suitable for use in TA cloning vectors. The pGEM-Teasy vector has a single base 3' thymidine overhang which can then be ligated to single base 3' adenine overhangs of PCR products generated by *Taq* DNA polymerase.

The ligation reactions were performed in 15 μl reaction mixtures containing: 1x Rapid Ligation Buffer (30 mM Tris-HCL, pH 7.8; 10 mM $MgCl_2$; 10 mM DTT; 1 mM ATP; 5% PEG from Promega); 50 ng of the pGEM-Teasy vector; 3 Weiss units of T4 DNA Ligase ; 250 – 500 ng of purified PCR product. The reaction mixture was incubated at room temperature for 1 hour and then at 4°C overnight to increase the copy number of recombinant clones.

E. coli strain XL1-Blue competent cells were transformed with the ligation mixtures using the protocol provided in pGEM-T Easy Vector System I (Promega) with minor modification. 7.5 μl of the overnight ligation reactions were added to 100 μl of the competent cells, prepared according to modified protocol of Nishimura *et*

al. (1990). The remaining ligation mixture was kept in -20°C. The competent cells were incubated on ice for 30 minutes to allow the slow closure of cell membrane poles and then incubated at 42°C for 90 seconds for heat shock. The tubes were then placed on ice for 5 minutes before 1 ml of LB medium were added to the tubes. The LB medium was incubated at 37°C for 90 minutes with continuous orbital shaking at 200 rpm to allow the growth of transformed *E. coli* competent cells. After the incubation period, the competent cells were pelleted down by centrifugation at 5,000 rpm for 5 minutes and 1 ml of supernatant was carefully discarded. The remaining LB medium and cells were resuspended slowly with a micropipettor and plated on LB agar plates containing: 100 μg/ml ampicillin, 40 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactosidase (X-gal) and 500 μM isopropyl-β-Δ-thiogalactopyranoside (IPTG). The plates were incubated at 37°C for 14 – 16 hours.

Recombinant colonies were screened from the LB agar plate using the blue-white screening method. The pGEM-Teasy vectors contained lacZ gene encoded for β-galactosidase which breaks down the chromogenic X-gal substrates and results in blue colonies on antibiotic LB agar plate. On the other hands, ligation of PCR products into this region of the pGEM-Teasy vector disrupts the open reading frame of the encoded lacZ gene which then produces white colonies on antibiotic LB agar plate. Only white colonies with recombinant plasmids were picked from antibiotic LB agar plates and inoculated in 10 ml LB medium with 40 μg/ml ampicillin and incubated overnight at 37°C in an orbital incubator shaking at 250 rpm.

M13-PCR amplifications were carried out to confirm the presence of recombinant plasmids in white colonies. PCR reactions were carried out in TGradient Thermocycler (Biometra) in 12.5 μl reaction mixtures containing 1x PCR buffer containing 10 mM Tris-HCl, pH 8.9, 50 mM KCl and 0.1% Triton® X-100

(YorkBio), 1.5 mM $MgCl_2$, 200 μM of dNTPs (Bioline), 5 pmole of each M13 primer (M13 Forward: 5'-GTAAAACGACGGCCAGT-3' and M13 Reverse: 5'-GGAAACAGCTATGACCATG-3'), 0.65 U of *Taq* DNA Polymerase in storage buffer containing 50 M Tris-HCl, pH 8.8, 50 mM NaCl, 0.1 mM EDTA, , 1 mM DTT, 1% Triton® X-100 and 50% Glycerol (YorkBio) and 0.5 μl of LB medium culture. The PCR cycles parameters contained 95°C for 3 minutes, 25 cycles of 95°C for 30 seconds, 50 °C for 30 seconds, 72°C for 2 minutes and then followed by 72°C for 5 minutes. The M13-PCR products were analyzed on 1% (w/v) agarose gel electrophoresis in 1x TAE buffer for 30 minutes at 8V/cm together with DNA ladder, Hyperladder I (Bioline).

After gel electrophoresis, the positive recombinant clones were isolated by using QIAprep® Spin Miniprep Kits according to manufacturer's instruction. The miniprep products (1 μl) were analyzed on 1% (w/v) agarose gel electrophoresis to checking the integrity and purity of isolated recombinant plasmids.

2.2.5 DNA sequencing analysis and phylogenetic analysis

2.2.5.1 DNA sequence homology search

Recombinant clones (200 ng) were sent to the DNA sequencing service provider at John Innes Center Genome Laboratory, Norwich and also Protein and Nucleic Acid Chemistry Laboratory (PNACL), University of Leicester using universal M13 forward and reverse primer pairs. Sequence chromatograms were retrieved from John Innes Centre Genome Laboratory and PNACL websites.

The retrieved DNA sequence chromatograms were opened using bioinformatics software, Chromas version 1.45 (Conor McCarthy, Griffith University, Australia). The DNA sequences were exported to FASTA format using sequence export tool

embedded in Chromas version 1.45. The pGEM-Teasy plasmid sequence flanking the inserted sequence were removed from original FASTA file and saved as new file under FASTA format. Homology analyses were carried out using BlastN and BlastX services provided by National Center for Biotechnology Information (NCBI) which is established in 1988 as a national resource for molecular biology information.

2.2.5.2 Phylogenetic analysis of DNA sequences

DNA sequence (in FASTA format) was exported into BioEdit Sequence Alignment Editor version 7.0.5.3 (Hall, 1999) and multiple sequence alignment was carried out using ClustalW Multiple Alignment software embedded in the BioEdit Sequence Alignment Editor version 7.0.5.3 (Hall, 1999). The multiple sequence alignment file was saved in NEXUS format for Maximum Likehood (ML) phylogenetic analysis using PAUP version 4.0 beta 10 win (Sinauer Associates, Inc. Publisher) and in FASTA format for other methods of phylogenetic analysis (Neighbor-Joining, NJ; Maximum Parsimony, MP; Minimum Evolution, ME and Unweighted Pair Group Method with Arithmatic Mean, UPGMA) implemented in MEGA version 3.1 (Kumar *et al.*, 2004). For phylogeny reconstruction strategy of repetitive DNA sequences, the method described by Harrison and Langdale (2006) was used.

2.2.6 Total RNA isolation of *Musa acuminata* ssp. *malaccensis*

The inflorescence and young leaves of *Musa acuminata* ssp. *malaccensis* were collected from Rimba Ilmu, University of Malaysia and immediately put on ice to avoid degradation of RNAs. The outer layers of the inflorescence were removed from the male bud. Buds and leaf samples were washed under running tap water to

carefully remove dirt and other organisms like insects before being cleaned with 100% ethanol by using clean paper towels. The samples were immediately frozen in liquid nitrogen and stored in -80°C in a clean zip bag.

Before RNA extraction, all apparatus was swabbed with RNaseZap (Ambion) or RNaseAWAY and rinsed with RNase-free water or DEPC-treated water. For 2 g of the samples, 10 ml of CTAB extraction buffer (2% (w/v) CTAB; 2% (w/v) PVP-40; 100 mM Tris-HCl, pH 8.0; 25 mM EDTA, pH 8.0; 2 M NaCl; 0.5 g/L (w/v) Spermidine; 0.1 g/L (w/v) Proteinase K) together with 0.2% (w/v) of β-mercaptoethanol in 50 ml polypropylene tube preheated at 65°C were used to isolate RNA. The pestle and mortar were pre-cooled in -80°C freezer. 2g samples were ground to a fine powder in liquid nitrogen and then transferred immediately to 50 ml polypropylene tubes containing 10 ml of CTAB extraction buffer preheated to 65°C. The mixtures were incubated at 65°C for 2 hours. After the incubation period, an equal amount of Chloroform-Isoamyl alcohol (24:1) was added and the tubes vortexed until the mixture became milky. The mixture was centrifuged at 10,000 rpm for 15 minutes at room temperature and then the aqueous phase was transferred to new polypropylene tubes. Equal amounts of 4 M Lithium Chloride were added and mixed well before incubation overnight or 2 – 4 hours at 4°C.

After the incubation period, the mixture was centrifuged at 10,000 rpm at 4°C for 30 minutes and the supernatant was discarded. The pellets were resuspended in 5 ml NTES buffer containing 1 M NaCl, 10 mM Tris-HCL (pH 8.0), 1 mM EDTA (pH 8.0) and 0.5% (w/v) SDS and an equal amount of Chloroform-Isoamyl alcohol (24:1) was added to the tubes and vortexed until milky. The mixture was centrifuged at 10,000 rpm for 15 minutes at room temperature and the aqueous phase was transferred to new polypropylene tubes. 0.1 volume of 3 M sodium acetate (pH 5.2)

and 2.5 volume of 100% ethanol (pre-cooled at -20°C) were added and mixed thoroughly before incubation overnight at -80°C. The tubes were centrifuged at 10,000 rpm at 4°C for 30 minutes and the supernatant discarded. At this stage, a transparent/semi-transparent pellet was observed at the bottom of the polypropylene tube. The pellet was washed with 5 – 10 ml of 70% (v/v) ice-cooled ethanol and then centrifuged at 10,000 rpm at 4°C for 5 minutes. The pellet was air-dried at room temperature and then dissolved in DEPC-treated water/ Nuclease-free water. Isolated total RNA was stored in -80°C until it was needed for subsequent procedures.

2.2.7 mRNA enrichment from *Musa* total RNA samples

Plant mRNAs contain a stretch of poly(A) residues located at their 3' ends. This stretch of poly(A) allows the separation of mRNA from other RNAs in the total RNA such as ribosomal and transfer RNA. MicroPoly(A)Purist™ mRNA purification kit from Ambion was used to enrich the mRNAs from total RNA of *Musa* inflorescence and leaf samples on a small scale according to manufacturer's instruction. The recovered poly(A) RNA was stored at -80°C until it was needed.

2.2.8 First-strand synthesis of cRNA and reverse-transcriptase-PCR (RT-PCR)

Purified mRNAs were used as starting material for the first-strand synthesis of cDNA and followed by reverse-transcriptase-PCR. The SuperScript™ III Reverse Transcriptase (RT) (Invitrogen) was used to synthesize the first-strand of cDNA from purified mRNA of *Musa* inflorescence according to manufacturer's instruction. For RT-PCR reaction, the PCR_x Enhancer System (Invitrogen) was used with gene-specific primers according to manufacturer's instruction. The RT-PCR products were run on 1% (w/v) agarose gel electrophoresis at 5 V/cm to separate the bands and

bands of the expected size were excised from the gel and purified (see section 2.2.3). The purified products were cloned into the commercial vector, pGEM-Teasy, using pGEM-T Easy Vector System I (Promega) (see section 2.2.4). The positive clones were sent for DNA sequencing (see section 2.2.5).

2.2.9 DNA labeling

DNA fragments can be labeled using radioactive and non-radioactive protocols. The demand for non-radioactive labeling protocols, due to the dangerous properties of radioactive isotopes, has allowed the development of indirect and direct fluorophore labeling (Schwarzacher and Heslop-Harrison, 2000). In indirect labeling, fluorophores are conjugated to avidin or streptavidin to detect biotin, or anti-digoxigenin to detect digoxigenin that was incorporated into the probe DNA linked to dUTP. No antibodies are needed in direct fluorophore labeling as the nucleotides have been conjugated directly with fluorophores.

2.2.9.1 M13-PCR labeling

Selected clones were labeled by PCR amplification using universal M13 primer pair as described in section 2.2.4, together with 10 μM biotin-16-dUTP (Roche) or digoxigenin-11-dUTP (Roche). 3 μl of M13-PCR labelled products were loaded on a 2% agarose gel to check the labeling efficiency and product concentration of labeled probes. Labeled probes were purified using ethanol precipitation method (see above section 2.2.9.1).

2.2.9.2 Random primer labeling

Prior to labeling selected clones were amplified by M13-PCR (see section 2.2.4). Random primer labeling of large LTR retrotransposon clones (more than 500

bp) was carried using BioPrime® DNA Labeling Kit (Invitrogen) for biotin incorporation and Random Primer DNA Labeling System (Invitrogen) for digoxigenin incorporation according to manufacturer's instructions.

200 ng of the purified M13-PCR products were mixed together with 20 μl 2.5x Random Primer Solution and heat-denatured in boiling water for 5 minutes and then cooled down on ice for 5 minutes. 5 μl of 10x dNTP mixture and 1 μl of 40U Klenow Fragment were added to the reaction tubes and incubated overnight at 37°C. For digoxigenin labeling, 200 ng of denatured M13-PCR products were added to the tubes together with 20 μl Random Primer Solution. 2 μl of dATP, dCTP and dGTP together with 1 μl dTTP and 1 μl digoxigenin-11-dUTP were used in this labeling reaction. The enzyme, Klenow Fragment (3U, 1 μl), was added to the reaction last. The reaction tubes were incubated at 25°C overnight. Labeled probes were purified using the ethanol precipitation method (see above section 2.2.9.1) and the labeling efficiency was checked by dot-blot detection method.

A small piece (3 cm x 3cm) of positive charged nylon membrane (Roche) was soaked in buffer 1 (100 mM Tris-HCl, pH7.5; 15 mM NaCl) for 5 min and semi-dried between two filter papers (Whatman). The labeled probes were spotted on the membrane and air-dried for 5 minutes and then soaked in buffer 1 for 1 minute. The membrane was incubated in Blocking Reagent (Roche), 0.5% (w/v) in buffer 1, at room temperature for 30 minutes with slow shaking and then incubated in antibody solution containing: anti-biotin-alkaline phosphatase (Roche) and anti-digoxigenin-alkaline phosphatase (Boehringer Mannheim) conjugated antibody, diluted to 0.75U/ml in buffer 1, at 37°C for 30 minutes with slow agitation before being washed in buffer 1. The membrane was equilibrated in buffer 3 (100 mM Tris-HCl, pH9.5; 100 mM NaCl; 50 mM $MgCl_2$) and then detected in detection solution containing:

NBT (4-nitroblue tetrazolium chloride) and BCIP (5-bromo-4-chloro-2-indolyphosphate) detection reagents (Life Technologies) diluted in Buffer 3 to final concentration of 75 mg/ml and 50 mg/ml, respectively, for 5 – 10 minutes in the dark at room temperature. Successfully labeled probe showed dark as brown dots on the membrane due to the colorimetric reaction developed by the detection solution.

2.2.10 Fluorescent *In Situ* Hybridization (FISH)

FISH protocols for metaphase chromosomes slide preparation and *in situ* hybridization followed Schwarzacher and Heslop-Harrison (2000) with some minor modification to the metaphase chromosome accumulation steps and the proteolytic enzyme solution.

2.2.10.1 Fixation of *Musa* root tips

Actively growing *Musa* root tips of medium size (about 2 mm in width) were collected from plants growing in University Botanical Garden and put into the metaphase arresting agent, 2 mM 8-hydroxyquinoline, for 2 hours at glass house temperature and then at 4°C for 2.5 hours. The root tips were partially blotted dry on filter paper before transferring into freshly prepared fixative consisting of 100% (v/v) ethanol and glacial acetic acid in 3:1 ratio and incubated at room temperature for 30 minutes before transfer to new fixative solution at 4°C for short term storage and -20°C for long term storage.

2.2.10.2 Metaphase chromosomes preparation

Fixed root tips were washed in 1x Enzyme Buffer (4 mM citric acid, 6 mM tri-sodium acetate, pH 4.8) twice for 10 minutes at room temperature and then digested

in 1.5 ml of proteolytic enzyme solution containing: 5% (v/v) P2 Pectinase (Sigma, 450 units/mL), 1.8% (w/v) Cellulase (Calbiochem, 4000 units/g) and 0.2% (w/v) Cellulase (Onozuka, 5000 units/g) and 60 μl of Viscozyme® L solution (Novozymes, ~1.2 g/ml) in 1x Enzyme buffer for 1 – 2 hours at 37°C depending on the thickness of the root tips. Viscozyme® L solution is a multi-enzyme complex containing a wide range of carbohydrases, including arabanase, cellulase, β-glucanase, hemicellulase and xylanase. The digested root tips were washed twice with 1x Enzyme buffer for 15 minutes each and stored overnight at 4°C to allow the slow digestion of the inner part of the root tips. For chromosome spread preparation, squash preparation method was used. The root tips were removed from the 1x Enzyme buffer and put in a small petri dish with 45% glacial acetic acid and incubated for 10 minutes before transferring onto a chromic acid cleaned microscope slides. One drop of 60% glacial acetic acid was dropped onto the root tip and the root cap and other differentiated tissues were removed using fine needles and fine forceps. Cell suspensions containing meristematic cells were made and a small amount of the cell suspension was transferred to a new slide. A glass coverslip was put onto the cell suspension and then squashed firmly. The squash procedure was repeated for the remaining cell suspension. The coverslips were removed after freezing with dry ice. The slides were air-dried at room temperature and then stored at -20°C with silica gel until needed.

2.2.10.3 *In situ* hybridization

The slides were re-fixed in freshly prepared fixative (100% (v/v) ethanol and glacial acetic acid in 3:1 ratio) for 10 minutes and then dehydrated in 100% ethanol twice for 10 minutes each. The air-dried slides were subjected to RNAse A treatment (100 μg/ml) in Tris-HCl buffer, pH 8.0, for 30 – 60 minutes at 37°C in a humid

chamber. The slides were then washed twice in 2x SSC for 5 minutes each at room temperature and equilibrated in 10 mM HCl for 5 minutes at room temperature before applying 10 μg/ml pepsin (3200-4500 units mg protein^{-1}) prepared in 10 mM HCl onto the chromosome spread area and then covered with a plastic coverslip. The chromosome spread was incubated at 37°C in humid chamber for 10 – 20 minutes and then washed twice with 2x SSC for 5 minutes. The slides were re-fixed with freshly prepared 4% (w/v) paraformaldehyde, pH 8.0 for 10 minutes at room temperature, washed twice with 2x SSC for 5 minutes each and then dehydrated in an ethanol series (70%, 90% and 100% ethanol) for 2 minutes each and air-dried.

For fluorescent *in situ* hybridization, the hybridization mixture (50 μl per slide) contained: 40 or 50 % (v/v) formamide, 1x or 2x SSC, 10% (v/v) dextran sulphate, 0.125% (w/v) SDS, 0.125 mM EDTA, 1 μg of denatured salmon sperm DNA as blocking DNA and 100-200 ng labeled probes. The hybridization mixture was denatured at 85°C for 10 minutes and then immediately cooled on ice for 10 minutes. The denatured hybridization mixture was added to the slides and then covered with a plastic coverslip before being denatured at 75°C for 6 – 8 minutes in a ThermoHybaid HyPro-20 with vibration system set to 3, and incubated overnight at 37°C.

After the overnight hybridization, coverslips were allowed to float off in 2x SSC, and the chromosome preparations were washed twice in the stringent wash solution consisting of 20% (v/v) formamide in 0.1x SSC at 42°C for 5 minutes with agitation and then followed by 0.1x SSC twice at 42°C for 5 minutes. Finally, the slides were washed in 2x SSC twice for 5 minutes and then cooled down to room temperature before incubation in detection buffer (4x SSC and 0.1% (v/v) Tween 20) for 5 minutes. Slides were incubated with 5% bovine serum albumin (BSA, prepared

with detection buffer) block solution for 30 minutes to block unspecific sites. The hybridization signal was detected by incubation with 2 μg/ml streptavidin conjugated to Alexa 594 (Molecular Probes) and 3 μg/ml FITC (Fluorescein isothiocyanate) conjugated to anti-digoxigenin antibody (Roche) in 5% BSA block solution for 1 hour at 37°C in humid chamber. The slides were washed twice in detection buffer at 42°C for 10 minutes each with agitation and then counterstained with 4 μg/ml of DAPI (4', 6-diamidino-2-phenylindole, Sigma) prepared in McIlvaine's buffer, pH7.0. Finally, the slides were mounted in antifade solution (AF1, Agar Scientific) to prevent the fast fading of the fluorescence signal when the slides were observed under fluorescence microscope. The slides were stored in 4°C overnight to allow the slow binding of the antifade solution to the fluorophores.

2.2.10.4 Photography and image processing of *in situ* hybridization slides

The *in situ* hybridization slides were analyzed with an Axioplan 2 epifluorescence microscope with single band pass filters (Zeiss) equipped with a cooled CCD camera (ProgRes™ C12, JENOPTIK Laser, Optik, Systeme GmbH). The *in situ* hybridization signals were analyzed using Filter Set 10 (excitation = BP450-490, bean splitter = FT510 and emission = BP515-565) for digoxigenin-labeled probe and Filter Set 15 (excitation = BP546/12, bean splitter = FT580 and emission = LP590) for biotin-labeled probe whereas the DAPI-stained chromosomes were analyzed with UV band pass filter (Filter Set 01; excitation = BP365/12, bean splitter = FT395 and emission = LP397). The images of each metaphase were captured using the 3 different filter sets on the cooled CCD camera and the location on the slides was recorded to allow the easy finding of the recorded metaphase. The 3 *in situ* hybridization images for each metaphase were overlayed using Adobe

Photoshop 6.0 imaging software using only those functions that are applied to all pixels in the image.

2.2.11 Nuclear DNA fiber isolation and extended DNA fiber preparation

2.2.11.1 Plant nuclei isolation

The nuclear DNA isolation protocol from Schwarzacher and Heslop-Harrison (2000) with some minor modifications was used to obtain the nuclei from *Musa balbisiana* 'Butohan 2'. The Nuclei Isolation Buffer (NIB) stock buffer containing: 10 mM Tris-HCl, ph 9.5; 10 mM EDTA; 100 mM KCl; 500 mM sucrose; 4 mM spermidine and 1 mM spermine was freshly prepared and filtered sterilzed through 0.2 μm sterile filter and stored in 4°C.

Young leaf material was collected from *Musa balbisiana* 'Butohan 2' growing in the glass house. 2 g of young leaves was ground into a fine powder in liquid nitrogen with pre-cooled pestle and mortar and transferred into 20 ml NIB stock buffer. β-mercaptoethanol (0.1%, w/v) was added to the tubes in the fumehood and mixed gently to break up the leaf powder clumps. The leaf tissue suspension was incubated on ice for 5 minutes and then filtered through a 64 μm filter into pre-cooled 50 ml polypropylene tube using a cut-end 1 ml pipette tip. The filtration step separates the leaf tissue from the nuclei in the suspension. 1 ml of pre-mixed NIB containing 0.5% (v/v) Triton X-100 (Sigma) was added to the filtrate and then mixed gently. Triton X-100 helps to remove the chloroplasts from the nuclei suspension. The mixture was centrifuged at 2000 x *g* at 4°C for 5 minutes to pellet down the nuclei in the filtrate and then the supernatant was discarded from the tube. 20 ml of NIB with 0.1% β-mercaptoethanol was added to resuspend the pellet and then filtered through a 30 μm filter followed by centrifugation at 2000 x *g* at 4°C for 5

minutes. The pellet was resuspended gently in 1 ml of NIB containing 0.5% (v/v) Triton- X-100 and then centrifuged at 2000 x *g* at 4°C for 5 minutes. The supernatant was discarded before the pellet was resuspended carefully with 200 μl of NIB containing 100% glycerol in 1:1 ratio without any β-mercaptoethanol and Triton X-100. The isolated nuclei (1 μl) were pipetted on a microscope slides and stained with DAPI to check the quality of the isolated nuclei under an Axioplan 2 epifluorescence microscope with UV band pass filter (Zeiss, Filter Set 01). The remaining nuclei were stored at -20°C until needed for extended DNA fiber preparation.

2.2.11.2 Preparation of extended DNA fiber from *Musa* nuclei

The nuclei suspension stock was carefully removed from -20°C freezer to avoid mixing of the nuclei, which are settled at the bottom of the polypropylene tube with storage buffer and then transferred on ice. The end of a yellow pipette tip was cut to avoid shearing the nuclei, which would result in fragmentation of the DNA fibers in subsequent steps. In a microcentrifuge tube, 10 μl nuclei suspension stock was added to 100 μl NIB stock buffer to dilute the glycerol content in the nuclei suspension stock and gently mixed before centrifuged at 3000 rpm for 5 minutes at 4°C. The supernatant was carefully removed from the tube with micropippette as the pellet was not attached firmly to the bottom of the microcentrifuge tube. The pellet was resuspended in 1x PBS buffer and 2 μl nuclei suspension was pipetted across one end of a cleaned poly-L-lysine side (Sigma) and partially air-dried until it became sticky. 8 μl freshly prepared STE lysis buffer (0.5% SDS, 5 mM EDTA, 100 mM Tris-HCl, pH 7.0) was pipetted on top of the sticky nuclei and incubated at room temperature for 5 – 10 minutes. The nuclei suspension was dragged down slowly to the other end of the slide with the edge of a cleaned glass coverslip without touching the surface of

the slide. The slide was air-dried for 10 – 15 minutes and then fixed with freshly prepared fixative (100% (v/v) ethanol and glacial acetic acid in 3:1 ratio) for 10 minutes. The slides were then baked at 60°C for 30 minutes before being used for *in situ* hybridization or stored at -20°C for up to several weeks.

2.2.11.3 Extended DNA fiber-FISH

For DNA fiber-FISH hybridization, the chromosome FISH protocol described in section 2.2.10.3 was followed in detail without any chromosome pre-treatment step. Formamide was not used for any steps of the post-hybridization washes. After the hybridization step, the slides were washed in 2x SSC for 5 minutes to remove the plastic coverslip and then further washed in 2x SSC at 42°C for 10 minutes followed by 5 minutes. The slides were equilibrated in detection buffer (4x SSC and 0.05% Tween 20) for 5 minutes and then 3% BSA blocking solution was added on top of the slide before being covered with a plastic coverslip. The slides were incubated at 37°C for 30 minutes and then washed in detection buffer for 2 minutes. The antibody solution prepared (see section 2.2.10.3) in 3% BSA blocking solution was applied to the slide and incubated at 37°C for 1 hour followed by two washes in detection buffer at 42°C of 10 minutes each. The subsequent steps of DAPI staining, slide mounting and observation followed the protocol described in section 2.2.10.3 and 2.2.10.4 with no modification.

2.2.12 Fluorescent *in situ* hybridization and immunostaining with anti-5-methylcytosine antibody

Metaphase chromosome preparations of root tips fixed in ethanol:glacial acetic acid (3:1) from section 2.2.10.1 and 2.2.10.2 were used. Only DNA probes labeled with biotin-16-dUTP (Roche) were used for *in situ* hybridization in this section. The

probe denaturation and hybridization steps of FISH described in section 2.2.13.3 were followed in detail before the immunostaining of the metaphase chromosomes with anti-5-methylcytosine antibody. After the post-hybridization washes, slides were equilibrated in 1x PBS buffer-Tween 20 0.3% (v/v) for 5 minutes at room temperature. 3% (w/v) BSA (Sigma) blocking solution prepared in 1x PBS buffer containing 0.3% (v/v) Tween 20 was added to the slide and incubated for 30 minutes at 37°C. 5 μg/ml per slide of the monoclonal anti-5-methylcytosine primary antibody (Mouse)(Research Products, Cat. No. NA81) prepared in 1x PBS buffer (Oncogene) was added to the slide and covered with a plastic coverslip followed by incubation at 37°C for 45 – 60 minutes in a humid chamber. After incubation, slides were washed with 1x PBS containing 0.3% (v/v) Tween 20 twice for 5 minutes at room temperature with agitation. Alexa 488-conjugated goat anti-mouse secondary antibody (Molecular Probes)(1:200 dilution in 1x PBS buffer containing 0.3% (v/v) Tween 20 together with 2 μg/ml Alexa 594-conjugated streptavidin (Molecular Probe) was added onto the slides and incubated at 37°C for 1 hour in a humid chamber. After incubation, the slides were washed in 4x SSC and 0.3% (v/v) Tween 20 twice at room temperature for 10 minutes each with constant agitation. The DAPI staining, slide mounting and slide observation steps were as described in section 2.2.10.3.

For immunostaining of anti-5-methylcytosine antibody on extended DNA fiber, the hybridization and immunostaining protocol described in section 2.2.15 was followed in detail with minor modification in post-hybridization steps where no formamide was used in any of the washing steps.

2.2.13 Microarray analysis of *Musa* LTR retrotransposons

2.2.13.1 Total Genomic DNA labeling for microarray hybridization

Total genomic DNA fragmentation prior to random primer labeling is the most crucial step in optimising of the total genomic DNA probe for hybridization to microarrays. The classical method suggests boiling total genomic DNA in water or weakly buffered solution for 2-5 min followed by sonication to allow fragmentation of the DNA to an average size of 500 – 1500bp. Another method which has been proposed by Amon and Ivanov (2003) employs an alkaline denaturation step before the random primer reaction.

In this study, the alkaline denaturation step was used to denature the total genomic DNA of three different *Musa* species before random primer labeling with random hexamers. All the buffers used for the alkaline denaturation step were prepared with autoclaved double deionised water to avoid contamination. Total genomic DNA (200 – 300 ng) was incubated with 200 mM NaOH in a total volume of 100 μl for 10 min in a 70°C water bath followed by precipitation with 40 μl 5M ammonium acetate (pH7.5) and 350 μl chilled absolute ethanol for 2h at -20°C or overnight at -80°C. The tubes were allowed to thaw at room temperature for 5 min to remove the ice formed on the tube wall before being centrifugated at 13,000 rpm for 30 minutes at 4°C followed by washing with ice-chilled 70% ethanol to remove the remaining ammonium acetate salt in the tubes. The pellets were resuspended in 50 μl of 10 mM Tris-HCl, pH 8.0. Fragmented genomic DNA (5 μl) was run on 1% agarose gel electrophoresis to check the DNA molecular weight range with DNA ladder, Hyperladder I (Bioline).

The direct labeling approach was used to label the denatured genomic DNA for *Musa* repetitive DNA arrays. Labeling nucleotide mixture was prepared by mixing

0.35 mM dUTP conjugated with Alexa fluorophore (Alexa 488 and 594) with 0.15 mM dTTP and adding 0.5 mM each of the remaining three nucleotides (dATP, dCTP and dGTP). The direct labeling of genomic DNA was carried out using the Random Primer DNA Labeling System kit (Invitrogen) according to manufacturer's instruction. Genomic DNA (200 ng) was denatured in boiling water for 5 min together with 15 μl Random Primer Buffer (containing random hexamers) and then cooled on ice for 5 min. 8 μl of pre-mixed labeling nucleotide mixture (above) was added together with 1 μl of 3U Klenow Fragment and nuclease-free water to a total volume of 50 μl, and incubated at 25°C overnight. The labeled genomic DNA was purified using ethanol precipitation. 5 μl of 3M sodium acetate, pH5.2 and 100 μl of ice-chilled absolute ethanol were added and mixed vigorously by vortexing for 30 seconds. The tubes were incubated at -20°C for 4 hours before being centrifuged at 13,000 rpm for 30 minutes at 4°C to precipitate the labeled genomic DNA. The pellets were resuspended in 50 μl Tris-HCl, pH8. The labeled genomic DNA was checked by spotting 0.5 μl labeled genomic DNA on poly-L-lysine coated slides and viewing under UV using GENE FLASH (Syngene). The remaining labeled probes were stored at -20°C for future use.

2.2.13.2 Amplification of the LTR retrotransposon clones and microarray slide preparation

DNA clones were amplified from the plasmid using universal M13 primer pair and 2 μl of the PCR products were then run on gel to verify these products. The remaining PCR products were precipitated using 0.1 volumes of 3M Potassium acetate, pH 4.8 and 2 volumes of ice-chilled 100% ethanol, centrifuged and the pellets were resuspended in 10 μl of sterile distilled water. The precipitated PCR

products were further diluted to 200 ng/μl with sterile distilled water based on the result obtained from spectrophotometer and 1 μg of these products were mixed with printing buffer (DMSO) to obtain the final concentration of 50% DMSO and then transferred to 96-well arraying plate (Genetix Ltd). Clones were printed on Nexterion™ Slide A (SCHOTT) coated with aminosilane substrate using the in house-built microarray printing robot modulated from the Stanford design (http://smd.stanford.edu). The printed microarray slides were then allowed to air-dry for 10 minutes and rehydrated the printed surface by holding the slides over a heated water bath (95°C) for 2-3 seconds. The slides were then snap-dried with DNA side up by placing on a heat block at 85°C for 10 seconds. Finally, the slides were exposed to UV light using UV crosslinker at a setting of 600 mJ or baked at 80°C for 2 hours to fix the DNA to the aminosilane-coated slides.

2.2.13.3 Microarray hybridization and data analysis

Microarray hybridization followed the principals of *in situ* hybridization and protocols were adapted from those in Section 2.2.10.3. The hybridization solution consisted of 40% formamide, 1x SSC, 0,125% SDS, 1x Denhart's solution, 1 μg of salmon sperm DNA as blocking DNA and 200 ng each of Alexa fluorophore-labeled genomic DNA of two *Musa* species, *M. acuminata* ssp. *malaccensis* (Alexa 594) and *M. balbisiana* 'Pisang Gala' (Alexa 488). The hybridization solution was denatured in boiling waterbath for 10 minutes and then immediately transferred on ice for 10 minutes. The edge of printed area of the microarray slides were marked with white color fluorescent maker pen before adding the denatured hybridization solution on to the printed area. The printed area of the slide was covered with plastic coverslip. The slides were then denatured at 72°C for 5 minutes and hybridized overnight at 37°C in

ThermoHybaid HyPro-20 (Hybaid) with vibration system set to 3. After the overnight hybridization, the slides were washed with washing solution (0.2x SSC and 0.05% SDS) for high stringency wash at room temperature until the plastic coverslip fell off from the slide and then further washed in the same solution for 2 minutes, followed by 0.2X SSC without any SDS for 2 minutes. The slides were dried by centrifugation at 1000 rpm for 1 minute and scanned using ScanArray® 4000XL scanner (Packard BioScience, USA) with appropriate filter sets. The intensities of the microarray spots were obtained using the Fixed Circle method and then normalized using Background Subtraction method described in QuantArray® Microarray Analysis Software version 3.0 (Packard BioScience, USA).

Table 2.1: *Musa* species used.

Species	Subspecies	Accession Name	ITC code	Genome constitution	Abbreviation used
Section *Rodochlamys*					
Musa ornata Roxburgh	*ornata*	Ornata	0370	-	MO, ORN
Musa velutina Wendland & Drude	*velutina*	Velutina	0638	-	VEL
Musa laterita Cheesman	*laterita*	Laterita	0627	-	LAT
Musa mannii Wendland	*sanguinea*	Sanguinea	0534	-	SAN
Musa mannii Wendland	*mannii*	Mannii	-	-	MAN
Section *Callimusa*					
Musa beccarii Simmonds	*beccarii*	Beccarii	1070	-	BEC
Musa beccarii Simmonds	*beccarii*	Hottana	-	-	-
Musa borneensis Beccari	*borneensis*	Borneensis	-	-	BOR
Musa coccinea Andrews	*coccinea*	Coccinea	0012	-	COC
Musa gracilis Cheesman	*gracillis*	Gracillis	-	-	GRA
Musa violascens Ridley	*violascens*	Violascens	-	-	VIO
Section *Australimusa*					
Musa peekelii Lauterbach	*peekelii*	Peekelii	0916	-	PEE
Musa maclayi von Mueller	*maclayi*	Maclayi	0933	-	MAC
Musa lolodensis Cheesman	*lolodensis*	Lolodensis	0956	-	LOL
Musa textilis Née	*textilis*	Textilis	0539	TT	TEX
Section *Eumusa*					
Musa schizocarpa Simmonds	*schizocarpa*	Schizocarpa	0599	SS	SCH
Musa itinerans Cheesman	*itinerans*	Itinerans	-	-	ITI
Musa acuminata Colla	*malaccensis*	Malaccensis	1060	AA	MAL
Musa acuminata Colla	*banksii*	Banksii 623	0623	AA	BAN
Musa acuminata Colla	*banksii*	Paliama, PNG067	0766	AA	-
Musa acuminata Colla	*microcarpa*	Borneo	0253	AA	-
Musa acuminata Colla	*burmannicoides*	Calcutta 4	0249	AA	-
Musa acuminata Colla	*errans*	Agutay	1028	AA	-
Musa acuminata Colla	*siamea*	Khae (Phrae), THA 015	0660	AA	-
Musa acuminata Colla	*burmannica*	Long Tavoy pied	0283	AA	-
Musa acuminata Colla	*zebrina*	Zebrina	1177	AA	-
Musa balbisiana Colla	-	Pisang Gala	-	BB	-
Musa balbisiana Colla	-	Balbisiana	-	BB	-
Musa balbisiana Colla	Type 4	Pisang Klutuk Wulung, IDN056	1063	BB	-
Musa balbisiana Colla	Type 4	Pisang Batu, IDN080	1156	BB	-
Musa balbisiana Colla	-	Tani	1120	BB	-
Musa balbisiana Colla	Type 1	Honduras	0247	BB	-
Musa balbisiana Colla	Type 3	Lal Velchi	-	BB	-
Musa balbisiana Colla	-	Butohan 2	0565	BB	-
Musa balbisiana Colla	-	Javan	-	BB	-

Table 2.2: *Musa* accessions (section *Eumusa*) used.

Group	Subgroup	Accession Name	ITC code	Genome constitution	Abbreviation used
AAcv	-	Kadali	-	AA	-
AAcv	-	Cherukadali	-	AA	-
AAcv	-	Sanna chenkadali	-	AA	-
AAcv	-	Matti	-	AA	-
AAcv	-	Pisang Bulin	-	AA	-
AAcv	-	Pisang Lemak Manis Terengganu	1183	AA	-
AAcv	-	Pisang Kra (7732)	1345	AA	-
AAcv	Cooking AA	Tomolo, PNG023	1187	AA	-
AAcv	Pisang Mas	Pisang Mas	0653	AA	-
AAcv	Pisang Mas	Pisang Mas Tissue Culture	-	AA	-
AAcv	Pisang Jari Buaya	Pisang Jari Buaya, BS312	0312	AA	-
AAcv	Pisang Lilin	Pisang Lilin	1400	AA	-
AB	-	Adukkan	-	AB	-
AB	-	Padala mukili	-	AB	-
AB	-	Njali povan	-	AB	-
AAA	-	Pisang Emdun	-	AAA	-
AAA	-	Pisang Montel	-	AAA	-
AAA	-	Pisang Berangan	1287	AAA	-
AAA	-	Pisang Berangan Tissue Culture	-	AAA	-
AAA	-	Mano rajitham	-	ABB*	-
AAA	Pisang Nangka	Pisang Nangka Pulau Tioman	0004	AAA	-
AAA	Red/Green Red	Green Red	0485	AAA	-
AAA	Red/Green Red	Red	-	AAA	-
AAA	Gros Michel	Gros Michel	1122	AAA	-
AAA	Cavendish	Dwarf Cavendish	0548	AAA	-
AAA	Cavendish	Giant Cavendish	0346	AAA	-
AAA	Cavendish	Grande Naine	0180	AAA	-
AAA	Cavendish	Robusta	0574	AAA	-
AAA	Cavendish	Mon Maris	0552	AAA	-
AAA	Cavendish	Petite Naine	0654	AAA	-
AAA	Cavendish	Poyo	-	AAA	-
AAA	Cavendish	Novaria_N1	-	AAA	-
AAA	Cavendish	Novaria_N2	-	AAA	-
AAA	Cavendish	Novaria_N3	-	AAA	-
AAA	Cavendish	Novaria_N4	-	AAA	-
AAA	Orotava	Pisang Kayu, IDN098	0420	AAA	-
AAA	Ambon	Pisang Bakar, IDN106	1064	AAA	-
AAA	Rio	Leite	0277	AAA	-
AAA	Lujugira/Mutika	Mbwazirume	0084	AAA	-
AAA	Lujugira/Mutika	Intokatoke	0082	AAA	-
AAA	Ibota	Yangambi KM5	1123	AAA	-
AAB	-	Pisang Rastali	-	AAB	-
AAB	-	Pisang Mutiara M1	-	AAB	-
AAB	-	Pisang Mutiara M2	-	AAB	-
AAB	-	Nendran	-	AAB	-
AAB	-	Kosta bontha	-	AAB	-
AAB	-	Motta povan	-	AAB	-
AAB	-	Karim kadali	-	AAB	-
AAB	-	Perum padalli	-	AAB	-

AAB	-	Palayam codan	-	AAB	-
AAB	-	Mysoore ettan	-	AAB	-
AAB	-	Krishna vazhai	-	AAB	-
AAB	-	Poovan	-	AAB	-
AAB	-	Chara padati	-	AAB	-
AAB	-	Kumbilla kannan	-	AAB	-
AAB	-	Poomkalli	-	AAB	-
AAB	-	Dooth Sagar	-	ABB*	-
AAB	-	China li	-	AAB	-
AAB	Pome/Prata	Prata Ana	0962	AAB	-
AAB	Pisang Rajah	Etta padati/Nendra padaththi	0243	AAB	-
ABB	-	Kosta bontha	-	ABB	-
ABB	-	Kanchi kela	-	ABB	-
ABB	-	Boothi bale	-	ABB	-
ABB	-	Velli padati	-	AAB*	-
ABB	-	Padati	-	AAB*	-
ABB	-	Karpooravalli	-	ABB	-
ABB	-	Monthan	0046	AAB*	-
ABB	Simili Radjah	Peyan	0123	ABB	-
ABBB	-	Pisang Nipah	-	ABBB	-
ABBB	-	Pisang Abu Siam Selangor (7765)	-	ABBB	-
ABBB	Pisang Awak	Pisang Awak	0213	ABBB	Aw
AABB	-	Pisang Abu Baru	-	AABB	-
AABB	-	Pisang Abu Baru Tissue Culture	-	AABB	-
AAAB	FHIA	Gold Finger Tissue Culture (FHIA-01)	0504	AAAB	-

* revised genome constitution by Nair and Teo *et al.*, 2005

3.0 Chapter III: LTR retrotransposon diversity in *Musa*

3.1 Summary

Sequence analysis based on the *Metaviridae* RT gene from different *Musa* taxa identified three different types. Group I elements showed high level of heterogeneity and three subgroups were identified by phylogenetic analysis. Some of the RT genes from Group I elements are disrupted by other types of repetitive DNA, including Radka3 repetitive sequence and unknown genomic sequence. Group II elements were identified from Bacterial Artificial Chromosome (BAC) clones using LTR_STRUC software and harboured a long stretch of 35 monomeric repeats of ~ 60 bp minisatellite in the region between the 5' LTR and POL gene domain. Group III elements showed high homology to genus *Chromovirus* of the *Metaviridae* and contained sequences homologous to the chromodomain (CHROMO) downstream of the integrase (INT) gene. Finally, the diversity of chromoviruses from different organisms and the *Musa Chromovirus* sequences are discussed.

3.2 Introduction

Retrotransposons are mobile elements that transpose through reverse transcription of an RNA intermediate that is converted into extrachromosomal DNA by the encoded reverse transcriptase (RT)/RNaseH (RH) enzymes prior to reinsertion into the genome (Bingham & Zachar, 1989; Boeke & Corces, 1989; Kumar & Bennetzen, 1999; Heslop-Harrison, 2000). This type of replication mode can rapidly increase the copy number of this type of element and hence greatly increase the genome size (Kumar, 1996; SanMiguel & Bennetzen, 1998). Retrotransposons can generate mutations by inserting within or near genes and this type of retrotransposon-

induced mutation is relatively stable because they transpose through replication and the sequences at the insertion sites are retained.

The LTR retrotransposons have, as the name implies, direct long terminal repeats that can range from a few 100 bp to over 5 kb in size. The LTRs do not encode any known proteins but they do contain the promoters and terminators associated with their own transcription (Xiong and Eickbush, 1990). There are several conserved features of LTR retrotransposons (Xiong and Eickbush, 1990), such as 5' and 3' LTR sequences flanking the ORF, 2 bp Inverted Repeats (IR) at the start and end of the LTR sequences, Primer Binding Site (PBS), PolyPurine Tract (PPT) and 4 or 5 bp Target Site Duplication (TSD) flanking the FLEs allow the identification of Full-Length Elements (FLEs) using bioinformatic tools (LTR_STRUC software; McCarthy and McDonald, 2003). LTR retrotransposons are further subdivided into *Pseudoviridae* (Ty1-*copia* elements) and *Metaviridae* (Ty3-*gypsy* elements) that differ from each other in both their degree of sequence similarity (Xiong and Eickbush, 1990; Hansen and Heslop-Harrison, 2004) and the order of their encoded genes. *Pseudoviridae* elements have been further subdivided into three genera: *Hemivirus*, *Pseudovirus* and *Sirevirus*. Similarly to *Pseudoviridae* elements, *Metaviridae* elements are subdivided into three genera: *Metavirus* (some chromoviruses are under this genus), *Errantivirus* and *Semotivirus*; a further recently identified, the DIRS group still need to be added within the taxonomic framework. *Pseudoviridae* are present throughout the plant kingdom, in species ranging from single-cell algae to bryophytes, gymnosperms and angiosperms (Voytas *et al.*, 1990; Eickbush, 1992). The *Pseudoviridae* are characterized by a distinctive *pol* domain order, wherein coding sequences for *int* precede those for RT/RH (Figure 1.1)(Peterson-Burch and Voytas, 2002). *Metaviridae* are also widely distributed in

the plant kingdom, including gymnosperms and angiosperms (Kossack & Kinlaw, 1999; Kumekawa *et al.*, 1999). In *Metaviridae*, the coding sequences of the RT/RH domains precede those that of the *int* domain (Figure 1.2).

In this chapter, different types of *Musa* LTR retrotransposons were isolated by PCR amplification, cloned and sequenced in order to study their diversity. FLEs of LTR retrotransposons were identified from the Bacterial Artificial Chromosome (BAC) clones of *Musa acuminata* 'Calcutta 4' (MA4_78I12, MA4_112I10, MA4_64C22, MA4_54B05, MA4_111B14, MuG9) and *M. balbisiana* 'Pisang Klutuk Wulung' (MBP_31O07) using bioinformatic software. Similarity and phylogenetic analyses were carried out to investigate their relationships.

3.3 Materials and Methods

3.3.1 PCR amplification and primer design

Total genomic DNA from of 18 *Musa* taxa (Table 2.1) was isolated from young leaves (see section 2.2.1) and diluted to 25 ng/μl for PCR amplification with degenerate and specific primers. Degenerate primer pairs (Table 3.1) to amplify the conserved regions of reverse transcriptase (RT) gene of *Metaviridae* elements were previously described by Kubis *et al.* (1998) and Friesen *et al.* (2001). The *Musa* specific primer pairs were designed from the conserved region of the RT gene of the full-length *Metaviridae* element, eCoTR-REs, identified in BAC clone, MA4_78I12 (see Results section). PCR amplification was carried out using PCR parameters and cycling conditions described in section 2.2.2. The primer sequences and melting temperatures (T_m) are provided in Table 3.1. For PCR product purification, cloning and sequencing, see sections 2.2.5 and 2.2.9.

3.3.2 Primer design for amplification of full-length *Metaviridae* element

For full-length *Metaviridae* element amplification, RT and LTR sequences were identified in the *M. acuminata Monkey* retrotransposon (Genbank accession number: AF 143332) using bioinformatic software (BioEdit) and database search (BlastN, Genbank). Two pairs of specific primers were then designed to amplify the region corresponding to RT and LTR sequence of elements. PCR amplification was carried using 2 mM $MgCl_2$, 400 mM dNTPs and 15 pmole of each primer (LTRF + RTR or RTF + LTRR)(Table 3.2) in 50 μl reaction mixtures. The timing of the PCR extension step was set to 4 minutes to allow amplification of longer PCR products. PCR parameters, cloning and sequencing of PCR products are described in sections 2.2.5 and 2.2.9.

3.3.3 Bioinformatic tools

LTR_STRUC software (McCarthy and McDonald, 2003) was used to identify the FLEs of LTR retrotransposons in the seven publicly available sequences of BAC clones. Dotplot analysis using JDOTTER (Sonnhammer and Durbin, 1995) was carried out to search for other sequence features such as minisatellite repeats and tandem repeats in the FLEs. DNA-to-protein homology search using the BlastX algorithm against the Genbank database was carried out to identify the gene domains of LTR retrotransposons. Multiple sequence alignments and phylogenetic analysis were carried out using BioEdit Sequence Alignment Editor version 7.0.5.3 (Hall, 1999) and MEGA version 3.1 (Kumar *et al.*, 2004), respectively.

3.4 Results

3.4.1 Identification of full-length LTR retrotransposons from published BAC DNA sequences

Several FLEs representing four different types of LTR retrotransposons: four *Metaviridae*, three *Pseudoviridae* , two *Morgane*-like and three LARD-like elements (Table 3.2), could be identified in the seven publicly available sequences of BAC clones using LTR_STRUC software (McCarthy and McDonald, 2003) based on the conserved features of LTR retrotransposons (see section 3.2).

The ORFs of the identified FLEs were further characterized by comparison with LTR retrotransposon sequences from other species using BlastX analysis (Table 3.2). The organization of all 12 FLEs is presented in Figure 3.1. LTR sequences with different lengths (Table 3.2) were observed in all FLEs and showed low DNA homology between LTR sequences from different FLE types. Their gene domains (GAG, AP, RT, RH and INT) are encoded in a single long ORF (Figure 3.1). In all 12 FLEs sequences (Figure 3.2; Table 3.2), various different minisatellite sequences with different monomer length and small number of repeats at different positions were detected in the LTR region and 3' Un-Translated Sequence (3'UTS) using JDOTTER (Sonnhammer and Durbin, 1995).

In one *Metaviridae* element, a long stretch of minisatellite repeated sequence consisting of 35 repeats of a ~ 60 bp monomer was identified between 3' LTR and the POL domain. This *Metaviridae* element has been designated as element Colocalized Tandem Repeat REtroelement sequence (eCoTR-REs)(Figure 3.1b and 3.2g). The MuMeta3 FLE was identified in BAC clone, MA4_78I12, due to the insertion of eCoTR-REs FLE and then reconstructed by deleting the eCoTR-REs FLE sequence and one TSD sequence. New types of LTR retrotransposon were

detected in the *Musa* BAC clones which structurally resemble *Morgane*, in wheat (Sabot *et al.*, 2006) and LARD, in barley (Kalendar *et al.*, 2004) like elements (Figure 3.1c). Detection of a long stretch of the ~ 60 bp minisatellite repeats in the *Morgane*-like element that showed DNA-level homology to eCoTR-REs element, together with their high LTR sequences homology indicate that this *Morgane*-like element might be derived from eCoTR-REs element and was therefore designated as *Meta*Morgane like element (Figure 3.1d and 3.2h). The other *Morgane*-like element detected in the BACs contained a POL-like domain with high homology to the POL domain of classical *Pseudoviridae* elements. This suggests that this *Morgane*-like element was derived from a *Pseudoviridae* element and is designated as *Pseudo*Morgane like element (Figure 3.1d and 3.2i).

Reverse-transcriptase (RT) and RNaseH (RH) gene sequence analysis of different *Metaviridae* elements identified in the BACs and selected genomic DNA (see sections 3.4.1 to 3.4.5), showed that the *Metaviridae* elements in *Musa* genomes fall into three different groups, named Group I, Group II and Group III (Figure 3.3). Group I elements represent the *Metaviridae* elements isolated from genomic DNA using the degenerate *Metaviridae* primers, GyRT1 and GyRT4, (see below section 3.4.2) together with three *Metaviridae* FLEs sequences, MuMeta1, MuMeta2 and MuMeta3. Group II elements are represented by the eCoTR-REs elements identified in the BAC clone. The representative of Group III elements is the *M. acuminata Monkey* element (Balint-Kurti *et al.*, 2000) which contains a Chromodomain (CHROMO) at the 3'end of the INT gene. Group III elements are members of the genus *Chromovirus* within *Metaviridae*. No chromodomain was identified in *Metaviridae* Group I and II elements and no envelope (*env*) ORF was detected in all three *Metaviridae* groups.

3.4.2 Sequence diversity of *Metaviridae* Group I elements

Degenerate *Metaviridae* element primers, GyRT1 and GyRT4, amplify the conserved domain of the RT gene of *Metaviridae* elements from different plant species (Kubis *et al.*, 1998; Friesen *et al.*, 2001) and normally generate a PCR product of 430 bp which corresponds to the conserved RT gene of *Metaviridae* elements. However, no PCR product with the expected size of 430 bp was observed in any of the 15 *Musa* samples tested. Instead, multiple PCR bands were observed with size ranging from ~ 200 bp to ~ 1 kbp, together with high molecular weight smearing. Different band patterns were observed between the *Musa* taxa (Figure 3.4a).

In order to validate the PCR products, bands with different sizes from one DNA sample, *M. balbisiana* 'Pisang Gala', were cloned and 12 positive clones were sent for sequencing. Homology search of the clone sequences using BlastN and BlastX analysis against the Genbank database showed homology to the RT gene of *Metaviridae* elements from other plant species. For sequences longer than 500 bp, the degenerate primers amplified the RT and RNaseH gene of *Metaviridae* elements whereas only the RT gene was amplified for sequences shorter than 500 bp. After this validation, more putative *Metaviridae* element clones from 7 different *Musa* taxa were isolated and 44 clones from different *Musa* taxa were sent for sequencing. In total, 55 *Metaviridae* element clones were identified and showed different levels of homology (ranging from 42 - 82%) to *Metaviridae* element RT genes using DNA-protein homology search (BlastX) against Genbank database. Searches against the local *Musa* DNA database showed higher homology of the cloned sequences to *Metaviridae* Group I elements (MuMeta1, MuMeta2 and MuMeta3) compared to *M. acuminata Monkey* retrotransposon or eCoTR-REs. Frame shifts and stop codons of

the RT ORF were observed in 34 clones out of the 55 *Metaviridae* Group I element clones. One *Metaviridae* element sequence (891GTY3-15) isolated from *M. balbisiana* 'Pisang Gala' showed 50% protein-protein homology to the RT gene sequence of the *Monkey* retrotransposon. This clone was subsequently used for a phylogenetic study of *Metaviridae* Group III elements (genus *Chromovirus*) in this chapter (see section 3.4.4).

All 54 *Metaviridae* element sequences identified were translated into amino acids and manually adjusted where necessary to take into account various frame shifts, and the stop codons were replaced with 'X'. Multiple amino acid sequence alignment was carried out using BioEdit software. Alignment results showed that most of the *Metaviridae* element sequences contained either deletions at the conserved DD motif (D = aspartic acid) at position 104 and 105 of the alignment, which is the active site of RT gene (Xiong and Eickbush, 1990; Hansen and Heslop-Harrison, 2004) or the sequence stopped before the conserved DD motif (Figure 3.4b and Appendix 1). Further sequence analysis showed that the RT domain of 15 clones out of 54 clones was truncated by the insertion of other types of repetitive DNA (Figure 3.4b and Table 3.3).

Phylogenetic analysis of the amino acid sequence of RT/RH region of the *Metaviridae* element clones and *M. acuminata Metaviridae* Group I FLEs was carried out together with a *gypsy* element from *Arabidopsis* (AAG51046) as an out-group (Figure 3.5). The Maximum Parsimony (MP) tree showed that they were divided into 3 subgroups. Subgroup A is the biggest consisting of 37 *Metaviridae* element clones together with *Metaviridae* element sequence, CSSV-Gypsy (AF106947), that flanks the *Cacao Swollen Shoot Virus* (*CSSV*) and MuMeta3 *Metaviridae* Group I FLE whereas MuMeta1 *Metaviridae* Group I FLE is grouped

together with *Metaviridae* element clones, 1176GTY3-10 (*M. balbisiana* 'Pisang Gala') and 808B5-2 (*M. paradisiaca* 'Pisang Abu Baru', tissue-cultured material), in Subgroup B. The members of Subgroup C consist of MuMeta2 *Metaviridae* Group I FLE and 17 *Metaviridae* element clones from total genomic DNA.

3.4.3 Sequence diversity of *Metaviridae* Group II elements

Primers (CoTR-RTF and CoTR-RTR) were designed from the conserved RT domain of the *Metaviridae* Group II element (eCoTR-Res)(Table 3.1). The POL gene domain (BlastX) between *Metaviridae* Group I and Group II elements showed low homology levels of 31%. This supports the notion that *Metaviridae* Group I and Group II elements belong to different groups of *Metaviridae* elements. The expected PCR product size is 330 bp based on the conserved RT gene (Table 3.1). This primer pair was used to amplify eCoTR-REs RT gene domain fragment from *M. acuminata* ssp. *malaccensis* and *M. balbisiana* 'Pisang Gala'. A sharp distinct PCR product band with expected size of 330 bp was observed in both species, purified from the gel before cloning and sequencing analysis.

DNA-protein homology search, BlastX, was carried out to verify the clones. All twelve sequences, 6 from each species, showed homology to *Metaviridae* RT genes. The percentage of amino acid homology ranged from 57% - 72% with clones from *malaccensis* showing higher homology to each other than clones from 'Pisang Gala'. No frame shifts and stop codons were observed in clones from *malaccensis* whereas both frame shifts (one clone) and stop codons (3 clones) were observed in clones from 'Pisang Gala'. Three clones had no frame shifts and stop codons.

Multiple DNA sequence alignments are shown in Figure 3.6. The phylogenetic tree constructed using the Maximum Parsimony (MP) method (Kumar *et al.*,

2004)(Figure 3.7) showed that *Metaviridae* Group II element clones from *M. balbisiana* 'Pisang Gala' formed two subgroups (II and III) and all the clones from *M. acuminata* ssp. *malaccensis* formed subgroup I together with eCoTR-REs sequence identified in the BAC clone.

3.4.4 Sequence analysis of *Metaviridae* Group III element LTR sequences

The 3' LTR sequence of the *M. acuminata Monkey* retrotransposon, AF143332 (genus *Chromovirus*, *Metaviridae*) was identified by comparison with LTR sequence features of *Chromovirus* elements from other species to determine the U3, R and U5 region of the LTR sequence (Figure 3.8)(Xiong and Eickbush, 1990). Further analysis of the U3 and U5 region allowed the detection of the 7 bp repeats and minisatellite repeats at its U3 and U5 regions, respectively. Three copies of the 7 bp repeats sequence were observed in the U3 region of the *Monkey* retrotransposon 3' LTR sequence and four copies of the tandem repeat were observed in the U5 region, together with the putative *cis*-regulatory elements (Figure 3.8).

Primers (LTRF and LTRR) specific to LTR sequences of *M. acuminata Monkey* retrotransposon were designed and used for amplification of LTR sequences from different bananas. The expected PCR product size for a single LTR sequence is 800 bp based on the length of LTR sequence in *Monkey* retrotransposon (Table 3.1). After the PCR amplification, two distinct bands with similar size (~ 770 and ~ 790 bp) were observed. PCR bands were cloned and 52 positive clones from three species, *Musa acuminata* ssp. *malaccensis* (A genome), *M. balbisiana* 'Pisang Klutuk Wulung' (B genome) and *M. ornata*, were sequenced. Sequencing showed that the size of LTR clones ranged from 680 to 794 bp, DNA-DNA homology search BlastN analysis against Genbank database showed that all 52 LTR clones are highly

homologous to the LTR sequence of the original *Monkey* retrotransposon with percentages of homology ranging from 67% to 94%. Two LTR clones, Mal-LTR1a and PKW-LTR1c, from *malaccensis* and 'Pisang Klutuk Wulung', showed the same length and 100% homology to each other. Two LTR clones each from 'Pisang Klutuk Wulung' (PKW-LTR1e and PKW-LTR1g) and *M. ornata* (Mo-LTR1f and Mo-LTR1h), were identical. Two LTR clones (PKW-LTR2e and PKW-LTR2f) from 'Pisang Klutuk Wulung' with identical sequence length showed 99.6% homology with a C↔T transition at three positions in the sequences (Figure 3.9a). Furthermore, two LTR sequences with 99.5% homology (two A↔G transitions and two single nucleotide deletions/insertions (indels)) were observed in 'Pisang Klutuk Wulung' (Figure 3.9b).

Multiple DNA sequence alignments in Appendix 2 showed that LTR sequences from three *Musa* species are highly variable in their U5 region with different copies of tandem repeats at their 3' end. Although the U3 region of the LTR sequences are more conserved than the U5 region, different copies of the 7 bp repeats were observed in all the LTR sequences (Appendix 2). The most conserved part of the LTR sequences is located at their R region. The phylogenetic tree based on Maximum Parsimony showed that most of the LTR sequences formed three groups with high bootstrap values and some of these sequences are phylogenetically unresolved due to low bootstrap supports (Figure 3.10). No species-specific LTR groups were present and sequences from all species were scattered throughout the tree but generally LTR sequences within one species are more closely related to each other than to sequences from the other species (Figure 3.10).

3.4.5 Isolation of full-length *Metaviridae* Group III element

In order to isolate full-length *Metaviridae* Group III elements, a strategy to use primers that bind to the RT/RH gene domain in both directions and the LTR region in both directions was designed (Figure 3.11a). Forward and reverse primers were selected to amplify an overlapping fragment of ~ 1.2 kbp in the RT/RH domain and to extend to either the 5' or 3' LTR regions. Sequence comparison of the overlapping domain should then allow identification of full-length *Metaviridae* Group III elements, if those clones came from the same full-length fragment. Primers with high melting temperature, T_m, (Table 3.1) were designed to have higher specificity and PCR efficiency to target only *Metaviridae* Group III elements in *M. balbisiana* 'Pisang Gala'.

After temperature gradient PCR amplification of *M. balbisiana* 'Pisang Gala', genomic DNA with long extension times (4 minutes), PCR bands with the expected size of ~ 4 kbp for the region between 5'LTR and RT (fragment A) and ~ 3.5 kbp for the region between RT and 3'LTR (fragment B) were observed on the agarose gel together with other bands of various size (Figure 3.11b). These two large bands were excised from the gel and cloned. PCR amplification using corresponding LTR (LTRF and LTRR, Table 3.1) and RT (RTF and RTR, Table 3.1) primer pairs was carried out to verify that the clones indeed contained LTRs and RT genes. PCR products with expected size for LTR (~ 790 bp) and RT (~ 1.2 kbp) confirmed this. Five clones of each A and B fragments were selected and sent for sequencing in both directions using universal M13 primers. Primer walking strategy was carried out to sequence the entire clones. Specific primers were designed based on the sequencing results and used as sequencing primers (Table 3.1) in the next round of sequencing.

Three rounds of sequencing with each round giving ~ 800 bp were needed to fully sequence each of the A and B fragments of the *Metaviridae* Group III element.

Sequences were assembled and showed that the DNA fragment A (Appendix 3a; Figure 3.12) contained 5' LTR sequences, *gag*, AP, RT and RH genes of *Metaviridae* Group III element whereas the DNA fragment B (Appendix 3b; Figure 3.12) consisted of RT, RH, INT and CHROMO genes together with the 3'LTR sequence of *Metaviridae* Group III elements. High homology (84% – 94%) was observed in all 10 clones to *M. acuminata Monkey* retrotransposon using DNA-DNA BlastN homology searches. However, all fragment A clones showed higher sequence homology (86% - 94%) than fragment B clones (84% - 90%). One particular sequence, MbTy31a5, showed highest DNA sequence homology (94%) to the complete POL sequence of the *Monkey* retrotransposon. High homology was observed especially in the ORF regions while LTR sequences with different length were observed and included small duplications and deletions at different positions. One sequence contained a deletion at the downstream sequence of *gag* and upstream sequence of AP gene domain, MbTY31a2, and another clone, MbTY31a1, showed deletion in the middle of the AP gene domain. Furthermore, deletions were observed in the RT sequence of MbTY32a2 and MbTY31a2 clones (Appendix 3a, b; Figure 3.12). Sequences were translated into amino acids, frame shift of the POL ORF was observed in 8 clones out of 10 clones whereas several stop codons at different positions were detected in nine clones. Only one clone, MbTY32a4, showed no frame shift and stop codon in the sequence.

3.4.6 Diversity of chromoviruses in different organisms

Four complete RT genes of chromoviruses were isolated from *M. paradisiaca* 'Pisang Awak' (ABBB) using RTF and RTR primers (Table 3.1). One hundred and two amino acid *Chromovirus* sequences from different organisms were retrieved from Genbank database and used for multiple alignments together with four *Musa Chromovirus* RT sequences (Appendix 4) and a Maximum Parsimony (MP) tree was constructed (Figure 3.13). The *Osvaldo* retrotransposon (genus *Metavirus*) from *Drosophila buzzatii* (AJ133521) was used as an out-group *Metaviridae* element for this study. Five groups of chromoviruses were identified in the phylogenetic tree based on classification designated by Gorinšek *et al.* (2004). All chromoviruses from three different *Musa* species: *M. acuminata* (AA), *M. balbisiana* 'Pisang Gala' and *M. paradisiaca* 'Pisang Awak' (ABBB) were group into the Galadriel subgroup (Figure 3.13).

3.4.7 Sequence analysis of *Pseudoviridae* elements

Three full-length *Pseudoviridae* elements had been detected in BAC clones (MA4_78I12 and MA4_112I10) using LTR_STRUC software (Figure 3.1a and Table 3.2). Multiple amino acid sequence alignment was carried out together with partial RT sequences of *Pseudoviridae* elements previously isolated (Teo *et al.*, 2002). Three variable regions were observed in the RT domain at amino acid position 21 to 28, 56 to 66 and 73 to 81 (Figure 3.14). RT sequences isolated from the same *Musa* taxa (for example, *M. acuminata* 'Pisang Lemak Manis Terengganu' and *M. paradisiaca* 'Pisang Awak Nipah') tended to share the same amino acid residues in these three variable regions. Phylogenetic analysis using the Maximum Parsimony (MP) method distinguished three subgroups of *Pseudoviridae* elements

(Figure 3.15). *Pseudoviridae* elements isolated from total genomic DNA of the same species tended to cluster together in one subgroup (Subgroup I and II), while the *Pseudoviridae* FLEs detected in the BAC clones were clustered together and formed Subgroup III (Figure 3.15).

3.5 Discussion

Metaviridae elements have been classified into three genera on the basis of the presence (*Errantivirus*) or absence (*Metavirus*, *Semotivirus*) of an envelope gene (*env*) and the presence of chromodomain (chromointegrase) (*Chromovirus*) (Malik & Eickbush, 1999; Marin & Llorens, 2000). Nine clades of *Metaviridae* elements have been identified on the basis of phylogenetic analysis of combined RT, ribonuclease H (RH) and integrase (INT) domains (Malik & Eickbush, 1999; Bae *et al.*, 2001). *Chromovirus* (Marin & Llorens, 2000) is the most widespread genus of *Metaviridae* with Eukaryota-wide distribution (Gorinšek *et al.*, 2004).

In my study of *Musa* LTR retrotransposons using molecular biology and database mining strategies, four different classes have been detected: two classical elements, *Metaviridae* and *Pseudoviridae*, and two new *Musa* types, LARD-like and *Morgane*-like elements (Figure 3.1 and 3.2). Furthermore, the *Metaviridae* elements in *Musa* can be divided into three different groups based on the amino acid sequence homology of the RT/RH genes: Group I, II and III (Figure 3.3). The absence of the *env* and CHROMO domains in both *Metaviridae* Group I and II elements suggests that these elements belong to the genus *Metavirus* (Figure 3.1b). Low sequence homology (DNA and amino acid) of the RT/RH genes between these two elements allowed them to be classified into two groups (Figure 3.3). As for the *Pseudoviridae*

elements, three different groups have been detected in different *Musa* taxa (Figure 3.15).

3.5.1 The organization and diversity of RT genes in the *Metaviridae* Group I elements of *Musa*

Group I elements were isolated using the published primers (Kubis *et al.*, 1998; Friesen *et al.*, 2001). This resulted in clones of diverse length and sequence (Figure 3.4), which can be subdivided into 3 subgroups by phylogenetic analysis (Figure 3.5). After close analysis of *Metaviridae* Group I element sequences to find the forward and reverse degenerate primer binding sites, all the sequences were actually amplified only with the forward primer, GyRT1, except for one clone, 308B3-2. Sequence analysis showed that forward primer, GyRT1, has homologies in its inverse orientation at several other sites within the retrotransposons giving rise to fragments of different length between ~ 200 bp to ~ 1kbp. For clone 308B3-2, both the forward and reverse primer binding sites were observed in the sequence. However, instead of the expected 430 bp size, the clone is 308 bp long, with a 69 amino acid residue deletion in the internal region, including the DD motif which is conserved active sites of RT gene (Xiong and Eickbush, 1990; Hansen and Heslop-Harrison, 2004). The forward degenerate primer, GyRT1, was designed from a more conserved *Metaviridae* RT gene motif (RMCVDYR) compared to the reverse primer, GyRT4 (YAKLSKC) (Friesen *et al.*. 2001; Kubis *et al.*, 1998) but both show high degeneracy, 40% and 52.6%, respectively, to cater for the high sequence variability found in these elements. The high degeneracy allows the forward primer to bind any DNA sequence with only 40% homology, although, it was designed to bind the most conserved RT gene motif. Hence, the forward primer sequence binds to other

positions in *Metaviridae* element's RT genes and also binds to DNA sequences found in other type of repetitive DNA, e.g. the Radka3 repetitive sequence (Valárik *et al.*, 2002), if it is inserted into the RT gene of *Metaviridae* elements (Figure 3.4b and Table 3.3). The use of higher $MgCl_2$ concentration (3.5 mM) in the reaction mixture together with low annealing temperature (T_a = 39°C) further lowered the stringency level of the reaction which resulted in amplification of multiple bands. Cloning and sequencing of these bands showed that they all contained *Metaviridae* element sequences, but with a high range of diversity (Figure 3.4b; Appendix 1). These longer clones allowed us to analyze not only the RT domain but also part of the RNaseH domains.

The loss of the conserved DD motif of the active RT site together with frame shifts and stop codons in *Metaviridae* Group I elements isolated using the degenerate primer pairs, suggest that these RT genes have lost their reverse transcription ability. Hence, they are most likely defective elements. Fourteen out of the 23 clones showed a large deletion at position 72, of the conserved motif, MPFGL and stopped at the position 140 of the amino acid alignment (Appendix 1) including the DD motif. This corresponds to a 69 amino acid deletion. Although 18 *Metaviridae* Group I element clones (about half of the isolated sequences) showed no deletion at the conserved DD motif, the presence of frame shifts and stop codons in different position of the RT gene, which will result in pre-mature termination of mRNA transcript suggest that these 18 *Metaviridae* Group I element clones have also lost their reverse transcription ability. The presence of other repetitive DNA sequence (Radka3; Valárik *et al.*, 2002) and unknown sequence, at the 3'end of fifteen *Metaviridae* Group I element's RT clones (Table 3.3), which disrupted the RT gene domain,

provided further evidence that these *Metaviridae* Group I element's RT clones are defective elements.

Many members of the retrotransposon family differ from each other due to the large internal insertion or deletions (Hu *et al.*, 1995; Marillonnet and Wessler, 1998). For example, the truncated insertion at the 5'end of LINE elements suggested that the internal variation was generated during the transposition process. The low replication fidelity of reverse transcriptase will also increase the rate of internal mutation (Kumar and Bennetzen, 1999). The absence of complete coding capacity of a retrotransposon does not necessarily mean that it will not transpose. The *Bs1* element from cereal that lacks the *pol* coding potential was first detected as a *de novo* insertion mutagen, indicating that it had utilized the POL function from some other retrotransposon (Johns *et al.*, 1985; Jin and Bennetzen, 1989). Tanskanen *et al.* (2007) reported the parasitism of a chimeric *BARE-2* element from barley, which can not synthesize its own GAG protein, but uses the GAG protein of the active *BARE-1* element. The *BARE-2* elements share high homology in the PBS (Primer Binding Site), PSI (Packaging Signal and DIS (DImerization Signal) with active *BARE-1* element. These motifs have been shown to allow parasitism among the *lentiviruses* by using the machinery of an active element (Tanskanen *et al.*, 2007).

3.5.2 The organization and diversity of RT genes in the *Metaviridae* Group II elements of *Musa*

Another scenario was observed in *Metaviridae* Group II element RT clones isolated using specific primers. Only four of the 12 clones contained frame shifts and stop codons in the RT sequences together with a single deletion detected in one clone, GALcoTR-RT3 (Figure 3.6), indicates that active Group II elements might be still

present in the *Musa* genome, particularly in *M. acuminata* (A genome) where higher homology levels (72%) were observed in the RT clones isolated. This level of homology allowed the differentiation of the A genome clones from the B genome clones (Figure 3.7). The very high homology (96%) between the 5'LTR and 3'LTR sequence of the full-length *Metaviridae* Group II element (eCoTR-REs, Figure 3.1b and Table 3.2) found in the BAC clone from *M. acuminata* 'Calcutta 4' provides further evidence that this type of element was still active in the recent past in the *M. acuminata* genome. The two LTR sequences of a retrotransposon are synthesized from a single template during their replication cycle. As a result, the 5'LTR is usually identical to the 3'LTR at the time of transposition. Therefore, the age of the insertion can be estimated from the sequence divergence between the two LTR sequences by calculating the mutation or substitution rate of two LTR sequences, assuming that the rates of substitution are consistent since the insertion date (Kumar and Bennetzen, 1999). Hence a retrotransposon with LTR sequences that share high homology suggests a very recent transposition.

3.5.3 The organization and diversity of *Metaviridae* Group III element LTR sequences

Fifty two LTR sequences from 3 representative *Musa* species (*M. acuminata*, *M. balbisiana* and *M. ornata*) *Metaviridae* Group III elements showed conserved length of 680 – 794 bp and variable DNA sequence homology between 67 – 94% (Appendix 4). One pair of identical LTR sequences was found in *M. balbisiana* 'Pisang Klutuk Wulung' and one pair in *M. ornata* (Figure 3.9). These identical LTRs (a sign of recent activity, see above), although we do not know whether they have come from the same *Metaviridae* element, indicate that the *Metaviridae* Group

III elements could still be active or were active very recently. Furthermore, identical LTR sequences from different species were identified one from *M. acuminata* ssp. *malaccensis* (Mal-LTR1a) and one from *M. balbisiana* 'Pisang Klutuk Wulung' (PKW-LTR1c). These could have arisen from transposition before the split of the two species or be a result of horizontal transfer as described by Kumar and Bennetzen (1999). Three groups of the LTR sequences of different lengths with high bootstrap supports (see Figure 3.10 & Appendix 4) was observed indicate high variability in the LTR region of *Metaviridae* Group III elements in *Musa*. Detection of various copy numbers of the 7 bp repeats in the U3 region, tandem repeats in U5 region and *cis*-regulatory elements of the LTR sequences of *Metaviridae* Group III element (Figure 3.8), together with the identification of identical LTR sequence of different lengths in three *Musa* species (Figure 3.9; Appendix 4), suggest that *Metaviridae* Group III elements in *Musa* genome might use the high sequence heterogeneity to allow them to evolve regulatory mechanisms for their survival in different *Musa* genomes.

The LTR region provides the promoter, terminator and the polyadenylation signal which is necessary for their replicational life cycle in the host genome. Studies focusing on the expression of the LTR sequence of retrotransposons showed that their length and *cis*-regulatory motif plays an important role in species-specific, tissue-specific and stress-specific expression of a retrotransposon (Matyunina *et al.*, 1996; Vicient *et al.*, 2005; Takeda *et al.*, 1999; Beguiristain *et al.*, 2001). The tobacco *Tnt1* elements showed sequence heterogeneity in the coding domains and U3 regulatory sequences of the LTR (Casacuberta *et al.*, 1997; Vernhettes *et al.*, 1998), which allow *Tnt1* elements to be further divided into three subfamilies: *Tnt1A*, *Tnt1B* and *Tnt1C*. The high sequence variability could allow retrotransposons to evolve

regulatory mechanisms in order to optimize their coexistence with their host genome (Vernhettes *et al.*, 1998). Furthermore, Beguiristain *et al.* (2001) reported that all three *Tnt1* subfamilies are expressed in *Nicotiana tabacum* and the differences in stress-associated pattern of expression is correlated with U3 region variability. Vicient *et al.* (2005) examined the variability and evolution of a *BARE*-1-like element in barley by focusing on their LTR sequences. In their study, three different groups of LTR sequences with different length were found, corresponding to each of the *Hordeum* genome types analyzed, which predate the divergence of these types. Tandem repeats near the 3' end of the LTR sequence which are located in the most variable region of the LTR sequence were observed in the *Hordeum* genome types examined (Vicient *et al.*, 2005).

3.5.4 *Metaviridae* Group III element (*Chromovirus*) diversity

The partial sequence of *Metaviridae* Group III element, *M. acuminata Monkey* retrotransposon, was first identified by Balint-Kurti *et al.* (2000) using a genomic library construction strategy. In my study, primer walking of two large fragments (A and B) that overlap at their RT/RH domains (Figure 3.11a) allowed the characterization of the putative full-length element in *M. balbisiana* genome. High homology between overlapping region (RT/RH domain) of ten clones, 5 clones from each fragment, together with the presence of the CHROMO sequence at the 3' end of the integrase domain (Appendix 3a, b; Figure 3.12) allowed them to be classified as members of the genus *Chromovirus* (Malik & Eickbush, 1999; Marin & Llorens, 2000), and also indicate that they belonged to the same *Metaviridae* Group III element family as the *Monkey* retrotransposon. However, no identical RT/RH domains were observed in these clones suggesting that they were amplified from 10

different *Metaviridae* Group III elements. Two fragment A clones contained deletions at their AP and GAG gene domains (Appendix 3a; Figure 3.12) and a large deletion in the RT domain was also observed in clone MbTy32A2 (Appendix 3b; Figure 3.12). Together with these deletions, the presence of frame shifts and stop codons in the ORF in 9 out of 10 clones, suggest that the isolated *Metaviridae* Group III element clones are most likely defective elements in *M. balbisiana* genome.

Phylogenetic analysis using amino acid sequence of the RT/RH gene domain of 102 *Chromovirus* sequences (Appendix 4) based on Maximum Parsimony (MP) algorithm (Figure 3.13) was used to reveal the true tree topology of chromoviruses and showed similar tree topology to that reported in Gorinšek *et al.* (2004). However, in our analysis the *Tekay* group is divided into two subgroups, *Tekay* I and II, separated by members of the Galadriel group indicating that high sequence diversity within the *Tekay* group is present (Figure 3.13). The *Tekay* I subgroup contains members of the previously described *Tekay* group identified by Marin & Llorens (2000) and Gorinšek *et al.* (2004) whereas the *Tekay* II subgroup consists of three members of the previously described *Tekay* group and three newly identified members (Medicago AC126790, Oryza AP002071c and Lotus chromovir9a AP004967) retrieved from BAC sequences of three different plant species using retrotransposon detection software, LTR_STRUC (McCarthy and McDonald, 2003).

3.5.5 LTR retrotransposon evolution in *Musa* genomes

Plant species often harbour different types of LTR retrotransposons in their genomes. For example, the wheat genome contains *Pseudoviridae* element, *Wis-2* element, (Moore *et al.*, 1991) and *Morgane*-like elements which were first detected in wheat by Sabot *et al.* (2006). The barley genome consists of high copy number of

Pseudoviridae element, *BARE-1* (Manninen and Schulman, 1993) and *BARE-2* element (Tanskanen *et al.*, 2007) and the LARD-like element, which was first discovered in barley by Kalendar *et al.* (2004). In tobacco, two highly active *Pseudoviridae* elements, the *Tnt1* (Grandbastien *et al.*, 1989) and *Tto1-Tto3* (Hirochika, 1993) elements, were detected. The *Tnt1* elements comprise of three members, Tnt1A, Tnt1B and Tnt1c, which show high variability in the U3 region of their LTR regulatory sequences (Vernhettes *et al.*, 1998). *Tnt1* elements have been shown to express in protoplast and roots as a heterogeneous population of RNA molecules that resemble retroviral quasispecies (Casacuberta *et al.*, 1995). Several different types of LTR retrotransposon were detected in the rice genome before and during the rice genome project: the *Tos1-Tos3* (Hirochika *et al.*, 1992), *RIRE1* (Noma *et al.*, 1997), *RIRE2* (Ohtsubo *et al.*, 1999), *Kangourou* and *Wallabi* elements (Piegu *et al.*, 2006). The same scenario was predicted for *Musa* where highly heterogeneous populations of LTR retrotransposons were detected using different LTR retrotransposon isolation and database mining strategies.

Metaviridae Group II elements (Figure 3.1b; 3.2g) and *Meta*Morgane-like elements (Figure 3.1d; 3.2h) might be the types of *Metaviridae* elements that carried minisatellite repeats in the region between 3' LTR and POL domain, although further database mining searches are needed to be carried out to confirm this. Minisatellite sequences and tandem repeats were also detected in the LTRs of other LTR retrotransposons, except the LARD-like elements, detected in BAC clones (Figure 3.2) and genomic DNA (Figure 3.8). This suggests that LTR retrotransposons might act as the minisatellites or tandem repeat carriers to help the dispersal of these two types of repetitive DNA to other locations in the *Musa* genome. In rice, the role of transposable elements (retrotransposons and DNA transposons) in the dispersal of

the minisatellite sequences has been studied by Inukai (2004). In his survey of minisatellites in 5.3 Mb of randomly selected rice DNA sequences from the public database, 82% of the minisatellite sequences were found in multiple copies and all were parts of transposable elements. Inukai (2004) also reported that at least 61 different transposable elements were acting as the carriers of minisatellites. The results obtained in this study together with the results reported by Inukai (2004) which represented two different plant genomes using different strategies pose an interesting question as to the nature of the dispersal of transposable elements and other type of repetitive DNAs and their relationship, whether the action of one type of repetitive DNA will help or prevent the survival of another type of repetitive DNA.

Table 3.1 PCR primers used for isolation of LTR retrotransposon and sequencing of *Metaviridae* FLEs from *Musa* species.

Primer Name	Sequence (5' to 3')	T_m (°c)	Binding site of the primers	PCR product size	Use or the primers
GyRT1	MRNATGTGYGTNGAYTAYMG	50.7	RT domain, *Metaviridae* Group I	430 bp	Retrotransposon isolation
GyRT4	RCAYTTNSWNARYTTNGCR	51.7	RT domain, *Metaviridae* Group I		Retrotransposon isolation
RTF	AGATGGGAGCCTCCGATTATGCGTC	74.1	RT domain, *Metaviridae* Group III	1000 bp	Retrotransposon isolation
RTR	CTTCCGACTTAATGCATCGGCCACG	75.8	RT domain, *Metaviridae* Group III		Retrotransposon isolation
LTRF	TGTCACGAACGGTCGTCG	67.5	LTR region, *Metaviridae* Group III	800 bp	Retrotransposon isolation
LTRR	TGTCACGGACTTAGCTGG	60.0	LTR region, *Metaviridae* Group I		Retrotransposon isolation
CoTR-RTF	AGGATGTGCGTTGACTACACTAGTC	64.3	RT domain, *Metaviridae* Group II	330 bp	Retrotransposon isolation
CoTR-RTR	GCTTTTCACGATCATGTCATCCACG	72.1	RT domain, *Metaviridae* Group II		Retrotransposon isolation
RTF01	TGATGGCTTCATTCGCATGGAC	71.2	RT domain, *Metaviridae* Group III	-	Sequencing
RTR01	RGAGTTCGCCATTCCGYAAC	64.8	RT domain, *Metaviridae* Group III	-	Sequencing
LTRF01	YRCTGAATTGCTTGTAAAGC	57.1	LTR region, *Metaviridae* Group III	-	Sequencing
LTRF02	CACTRAATGCGGGAGGAAGC	65.6	LTR region, *Metaviridae* Group III	-	Sequencing
RTF02	AGCAGTTATCCTGATGCAGC	61.9	RT domain, *Metaviridae* Group III	-	Sequencing
RTR02	AGAGTCTTCGRCAACTTGGGTGG	67.5	RT domain, *Metaviridae* Group III	-	Sequencing
M32A4_RF01	TGCTGAAGAAGGAGCAACCC	66.6	INT domain, *Metaviridae* Group III	-	Sequencing
M32A3_RF02	GTGGGAGAGCATTTCCTTGG	65.7	INT domain, *Metaviridae* Group III	-	Sequencing
M32A3_LR01	YTGCCAAGCAGCTGTACATG	65.0	LTR region, *Metaviridae* Group III	-	Sequencing
M32A1_RF03	AGCTCCAACCAGCATCACTCC	67.7	INT domain, *Metaviridae* Group III	-	Sequencing
M32A2_RF01	AGAGCCGCRARCTCAACGAG	63.5	INT domain, *Metaviridae* Group III	-	Sequencing
M32A2_RF03	CGTCAGCCTACCACTTCGCAA	69.4	INT domain, *Metaviridae* Group III	-	Sequencing
M32A3_LR02	GTAATATCTGCATTTCGGTGCCACTC	68.9	LTR region, *Metaviridae* Group III	-	Sequencing
M32A4_RF01R	AGTAGAAAGCCCTCTCCACGAGAG	66.9	INT domain, *Metaviridae* Group III	-	Sequencing

Table 3.2 Full-length elements (FLEs) of LTR retrotransposons detected in *Musa* BAC clones using LTR_STRUC data mining software (McCarthy and McDonald, 2003).

MA4_78I12	5'LTR (bp)	3'LTR (bp)	% LTR	Length (bp)	Direct Repeat
MuPseudo1	1739	1776	95.6	9141	ACTTT/ACTTT
MuPseudo2	1225	1225	99.7	9340	ATATT/ATATT
eCoTR-REs	624	624	96	11595	CCAGT/CCAGT
MuMeta1	269	298	87	5712	ATCTA/ATCCA
MuMeta2	2545	2545	98.5	8976	CTTCT/CTTCT
MuMeta3	200	200	98.3	5568	AACA/AACA

MA4_112I10	5'LTR (bp)	3'LTR (bp)	% LTR	Length (bp)	Direct Repeat
MuPseudo3	2272	2273	97.5	10071	CACCT/CACCT

MA4_64C22	5'LTR (bp)	3'LTR (bp)	% LTR	Length (bp)	Direct Repeat
*Meta*Morgane	655	655	97.8	3005	TATGT/TATAT

MA4_54B05	5'LTR (bp)	3'LTR (bp)	% LTR	Length (bp)	Direct Repeat
MuLARD1	441	443	86.8	4548	TTAC/CTAC

MA4_111B14	5'LTR (bp)	3'LTR (bp)	% LTR	Length (bp)	Direct Repeat
MuLARD2	446	446	99.7	4565	GAAC/GAAC

MuG9	5'LTR (bp)	3'LTR (bp)	% LTR	Length (bp)	Direct Repeat
MuLARD3	447	447	93.1	4556	AACC/AACC

MBP_31O07	5'LTR (bp)	3'LTR (bp)	% LTR	Length (bp)	Direct Repeat
*Pseudo*Morgane	1596	1562	93.8	6607	ATGAC/ATGAC

Table 3.3 Repetitive DNA sequence that inserted into the RT gene of the *Metaviridae* Group I element clones. The highlighted clones represent the clones used for annotation in Figure 3.4b.

Metaviridae Group I clones	size (bp)	Homology of the Inserted sequence
383AbTY3-8	383	Radka 3
383AsTY3-2	383	Radka 3
383AsTY3-8	383	Radka 3
383AsTY3-9	383	Radka 3
383B5-1	383	Radka 3
808B5-2	808	Radka 3
383B5-3	383	Radka 3
499B9-1	499	Radka 3
GTY3-3	383	Radka 3
480GTY3-4	480	Radka 3
480GTY3-6	480	Radka 3
383GTY3-14	383	Radka 3
296BuTY3-6	296	Radka 3
380BTY3-5	380	Radka 3
1176GTY3-10	1176	Unknown sequence

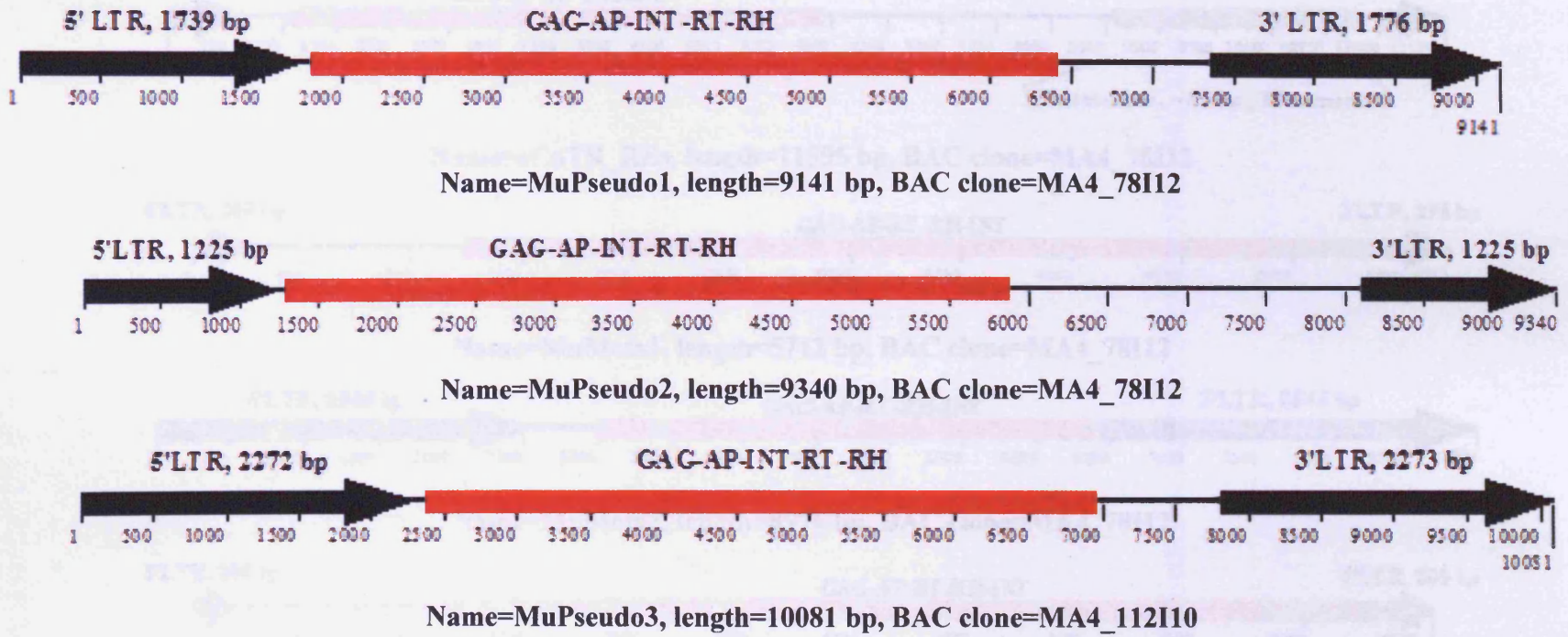


Figure 3.1a: Organization of *Pseudoviridae* elements detected in two BAC clones, MA4_78I12 and MA4_112I10, using LTR_STRUC software (McCarthy and McDonald, 2004). Only one Open Reading Frame (ORF; red) containing GAG protein (GAG), Aspartic Protease (AP), Integrase (INT), Reverse Transcriptase (RT) and RNaseH (RH) was observed in all three *Pseudoviridae* elements. Variable length LTR sequences (black arrow) were detected.

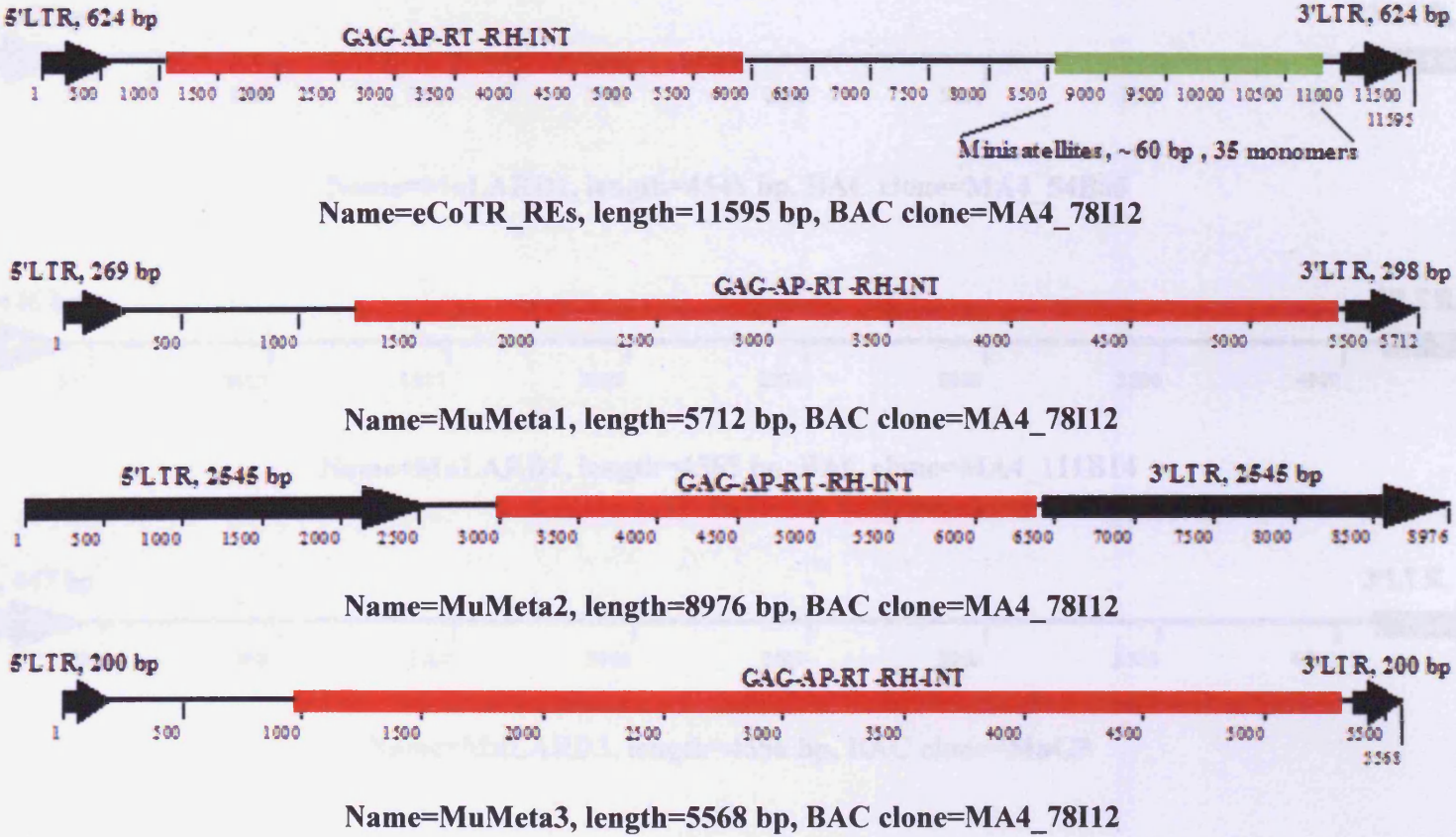
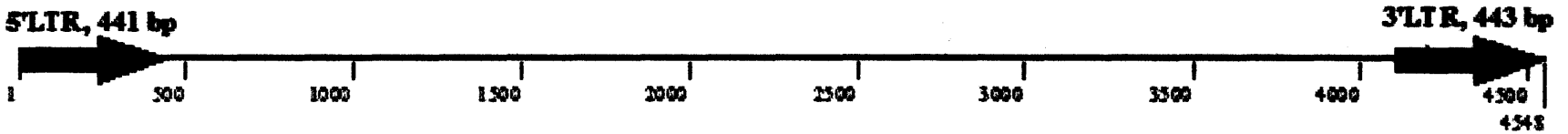
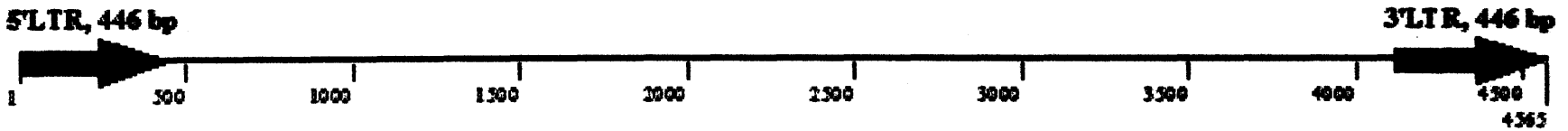


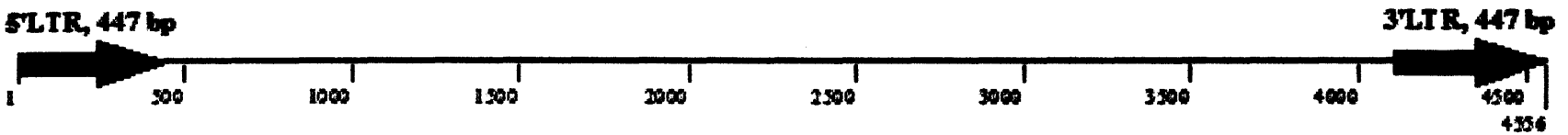
Figure 3.1b: Organization of *Metaviridae* elements detected in BAC clones, MA4_78I12, using LTR_STRUC software (McCarthy and McDonald, 2004). Only one Open Reading Frame (ORF, red) containing GAG protein (GAG), Aspartic Protease (AP), Integrase (INT), Reverse Transcriptase (RT) and RNaseH (RH) was observed in all three *Metaviridae* elements. Note the very short LTRs in 3 of the 4 sequences.



Name=MuLARD1, length=4548 bp, BAC clone=MA4_54B05

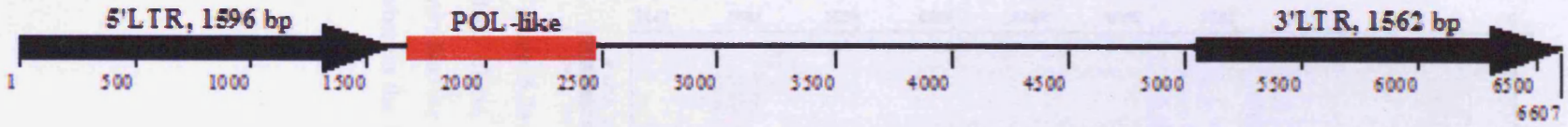


Name=MuLARD2, length=4565 bp, BAC clone=MA4_111B14

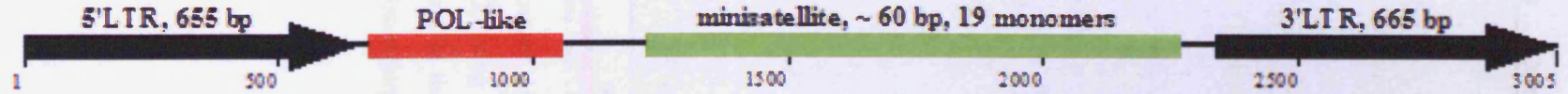


Name=MuLARD3, length=4556 bp, BAC clone=MuG9

Figure 3.1c: Organization of LARD-like elements detected in three BAC clones: MA4_54B05, MA4_111B14 and MuG9, using LTR_STRUC software (McCarthy and McDonald, 2004). No retrotransposon ORF was observed in the internal sequence of LARD-like elements. LTRs are very uniform in length between 441-447 bp.



Name=*Pseudo*Morgane, length=6607 bp, BAC clone=MBP_31O07



Name=*Meta*Morgane, length=3005 bp, BAC clone=MA4_64C22

Figure 3.1d: Organization of *Morgane*-like elements detected in two BAC clones: MBP_31O07 and MA4_64C22, using LTR_STRUC software (McCarthy and McDonald, 2004). Partial retrotransposon polyprotein sequence (POL-like) was observed in the *Morgane*-like elements.

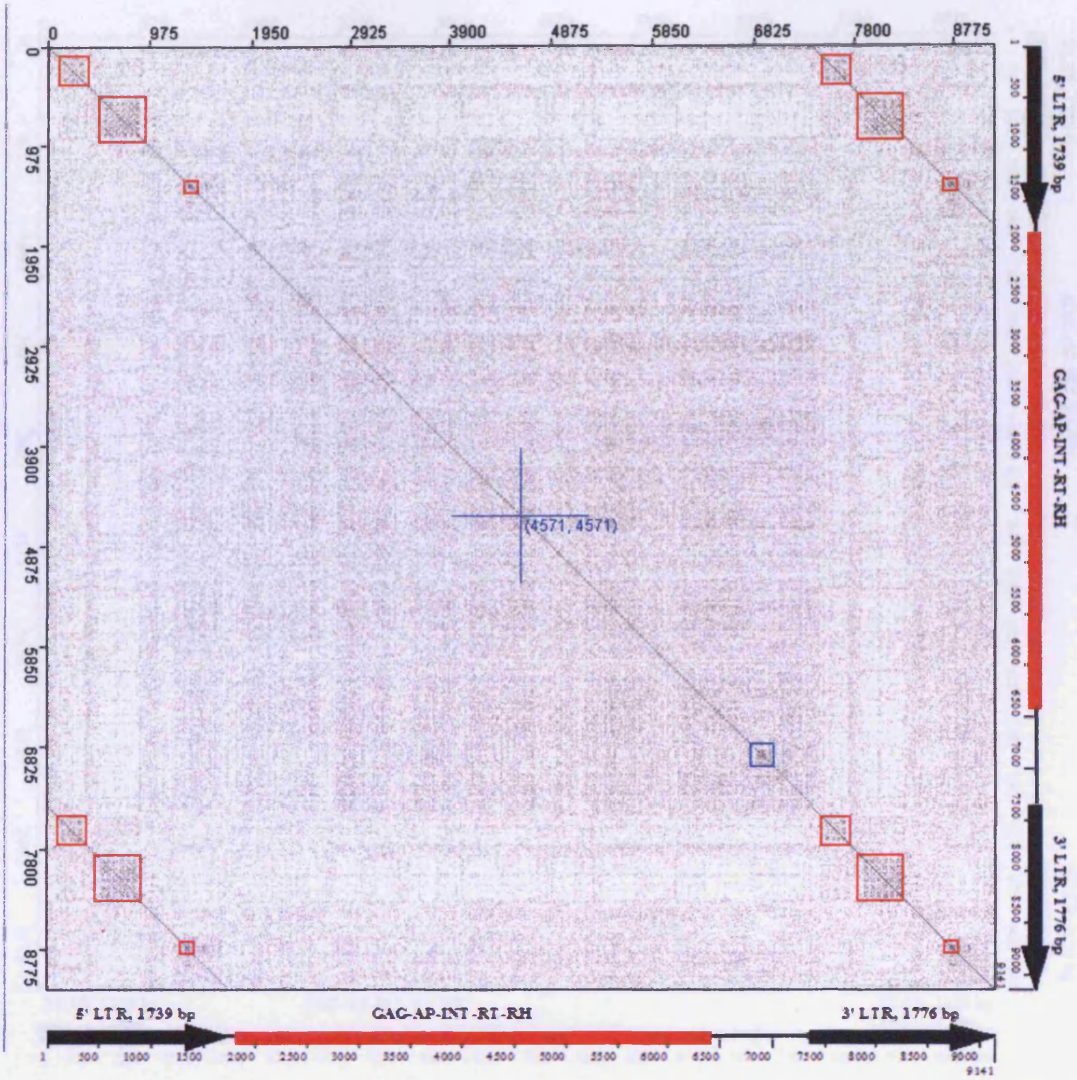


Figure 3.2a: Dotplot analysis of *Pseudoviridae* full-length element, MuPseudo1, detected in BAC clone, MA4_78I12, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the LTR region are represented by the red rectangular boxes whereas the 3' Un-Translated Sequence (3' UTS) is represents by the blue rectangular box.

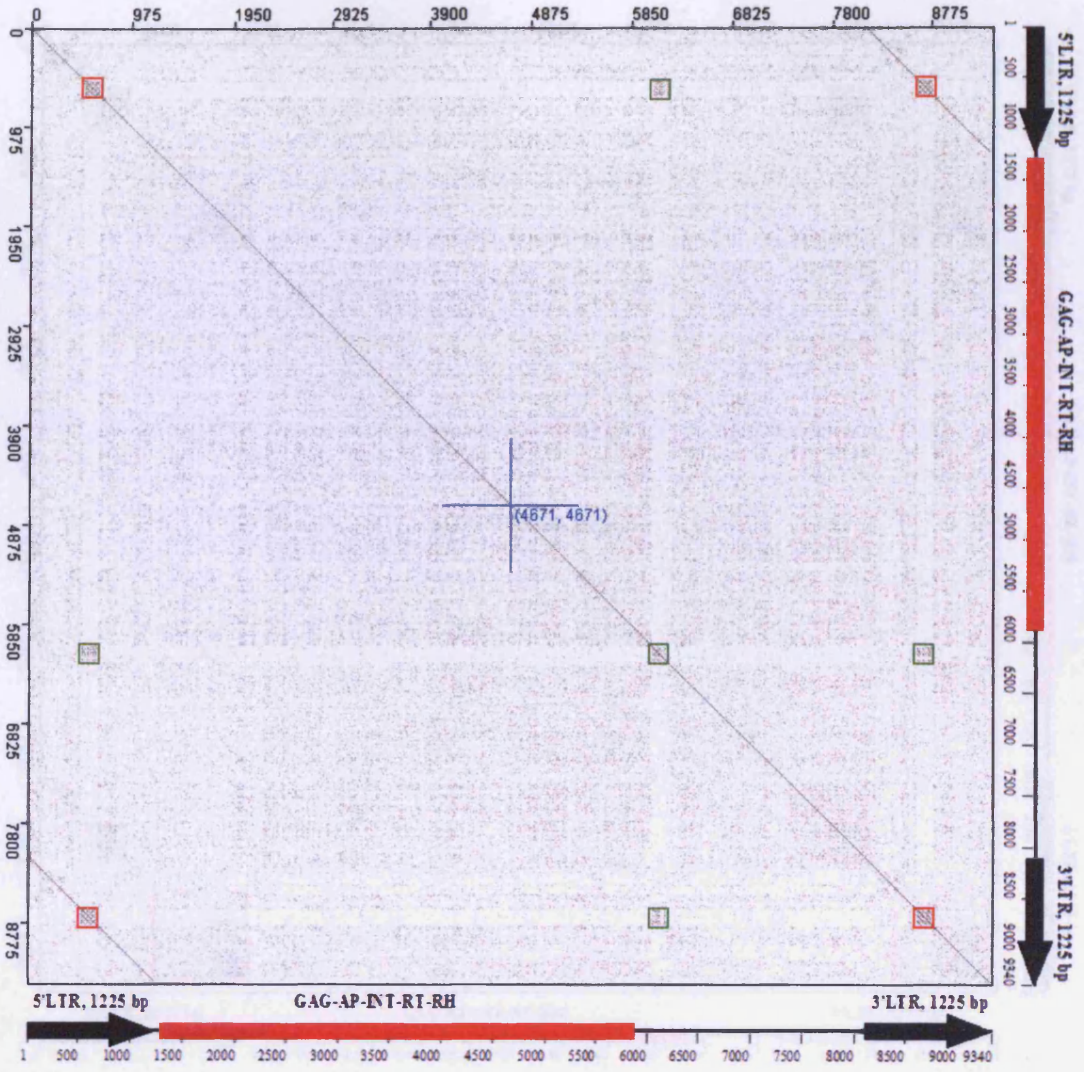


Figure 3.2b: Dotplot analysis of *Pseudoviridae* full-length element, MuPseudo2, detected in BAC clone, MA4_78I12, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the LTR region are represented by the red rectangular boxes. For the same minisatellite repeats found in the LTR region and the 3' Un-Translated Region (3' UTR) are depicted by the green rectangular box.

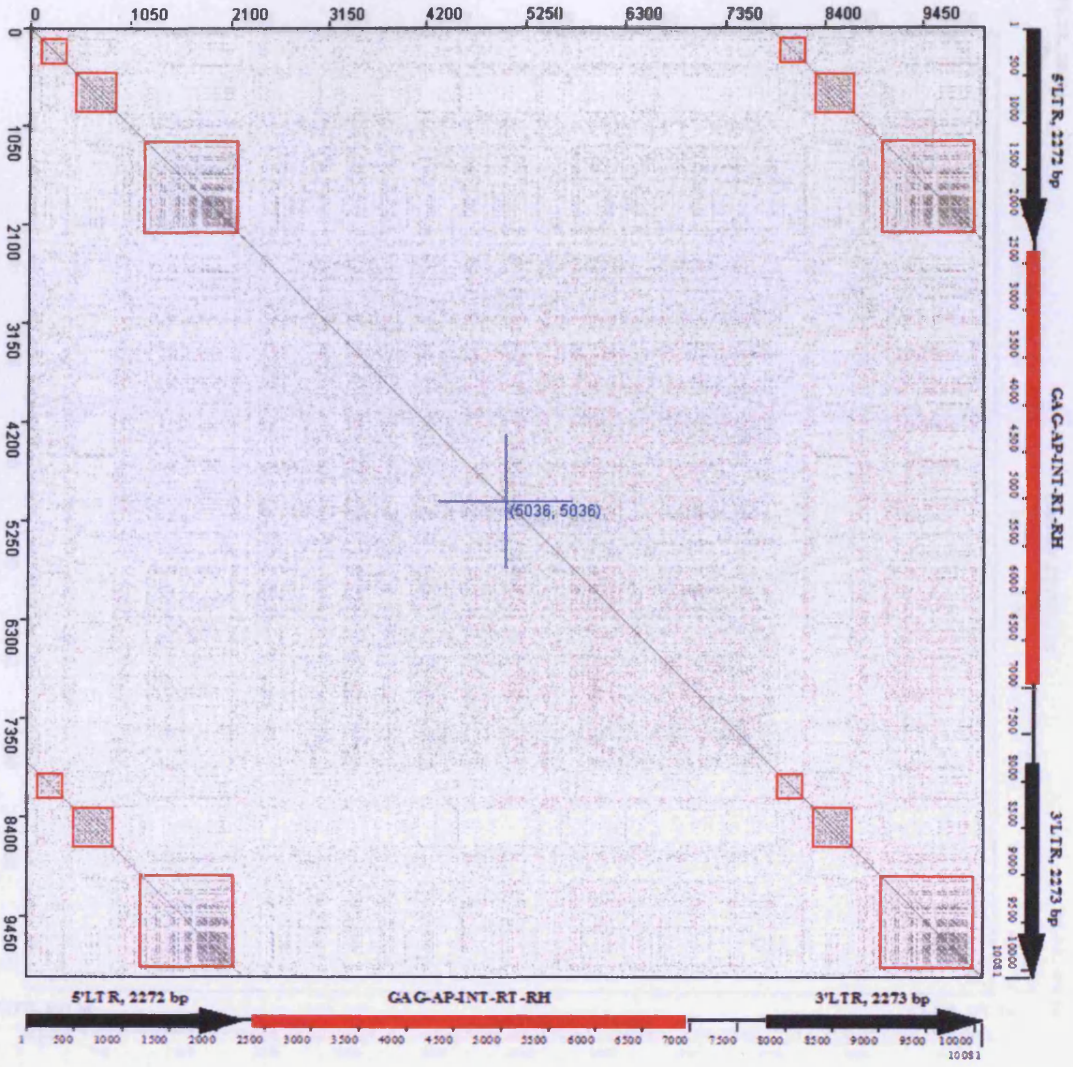


Figure 3.2c: Dotplot analysis of *Pseudoviridae* full-length element, MuPseudo3, detected in BAC clone, MA4_112I10, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the LTR region are represented by the red rectangular boxes.

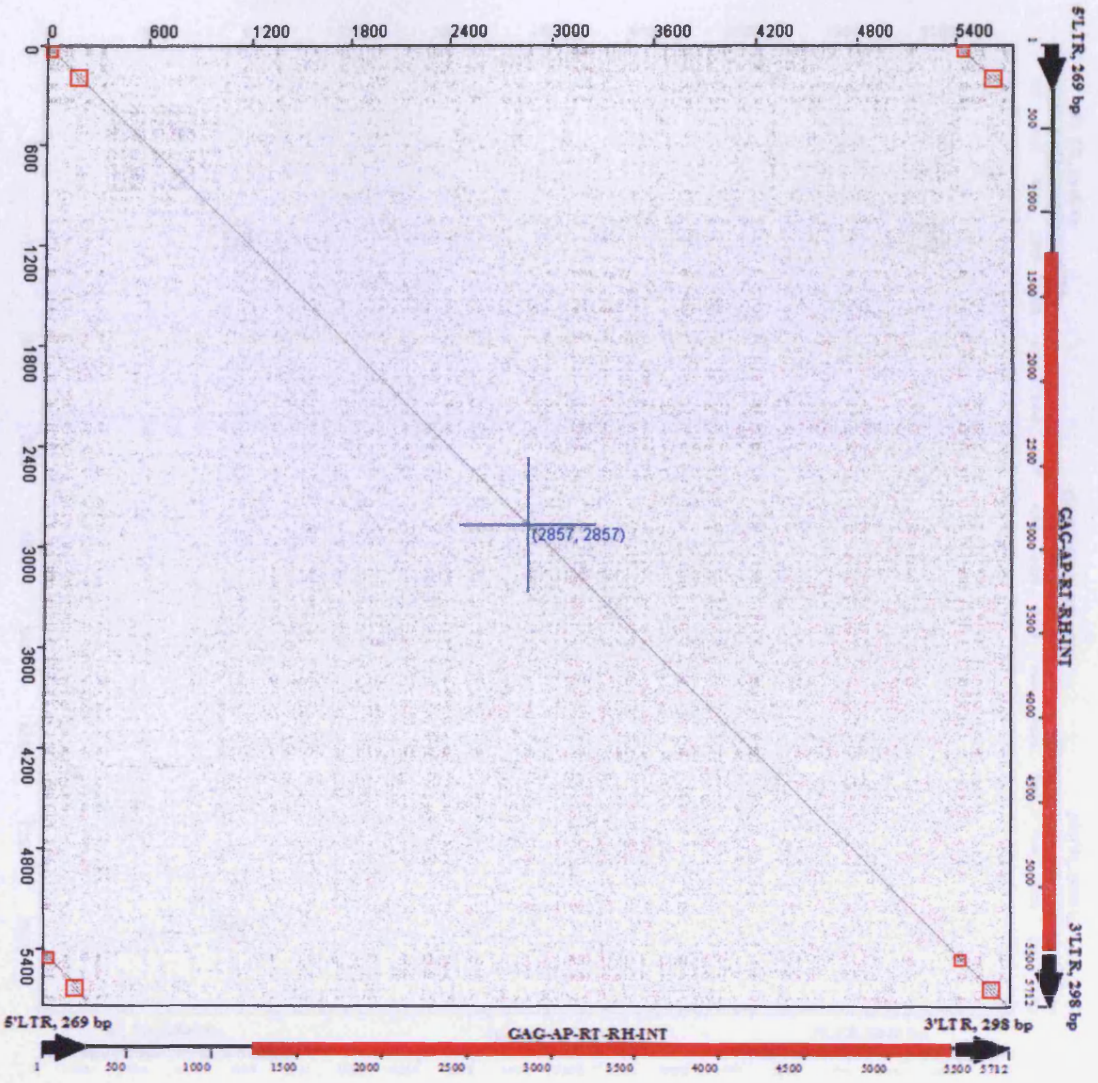


Figure 3.2d: Dotplot analysis of *Metaviridae* full-length element, MuMeta1, detected in BAC clone, MA4_78I12, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the LTR region are represented by the red rectangular boxes.

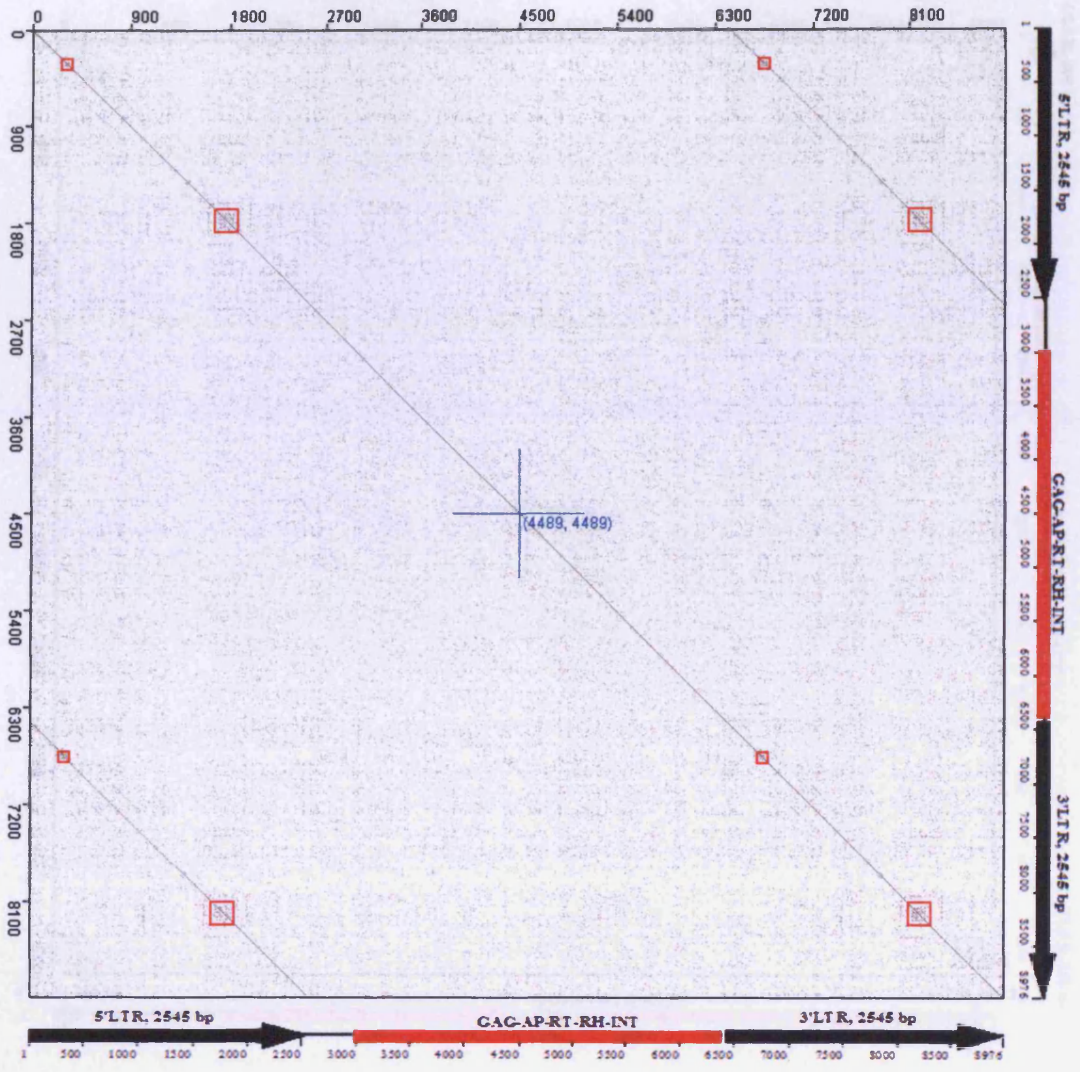


Figure 3.2e: Dotplot analysis of *Metaviridae* full-length element, MuMeta2, detected in BAC clone, MA4_78I12, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the LTR region are represented by the red rectangular boxes.

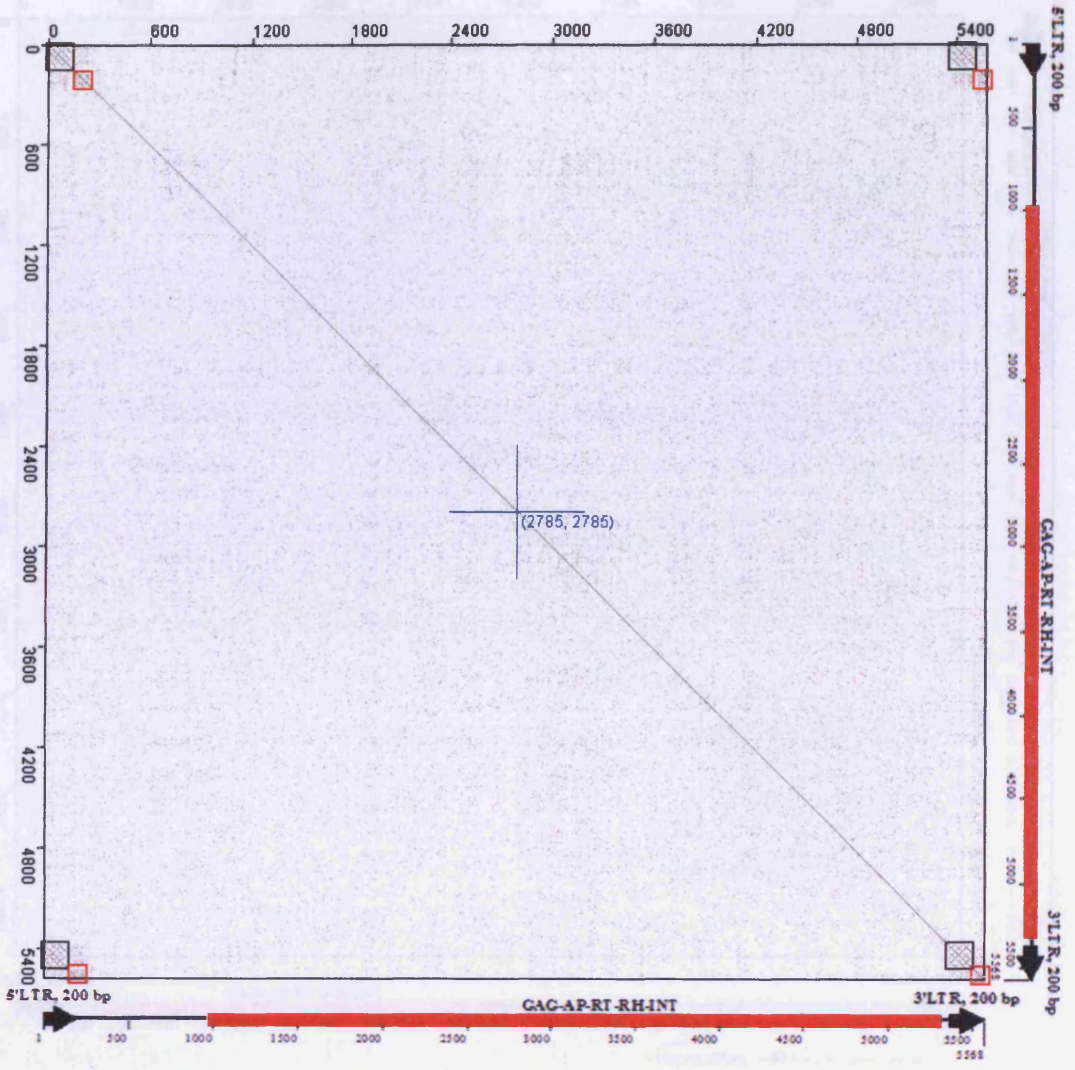


Figure 3.2f: Dotplot analysis of *Metaviridae* full-length element, MuMeta3, detected in BAC clone, MA4_78I12, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the LTR region are represented by the red rectangular boxes whereas the palindrome repeats are represented by the black rectangular boxes.

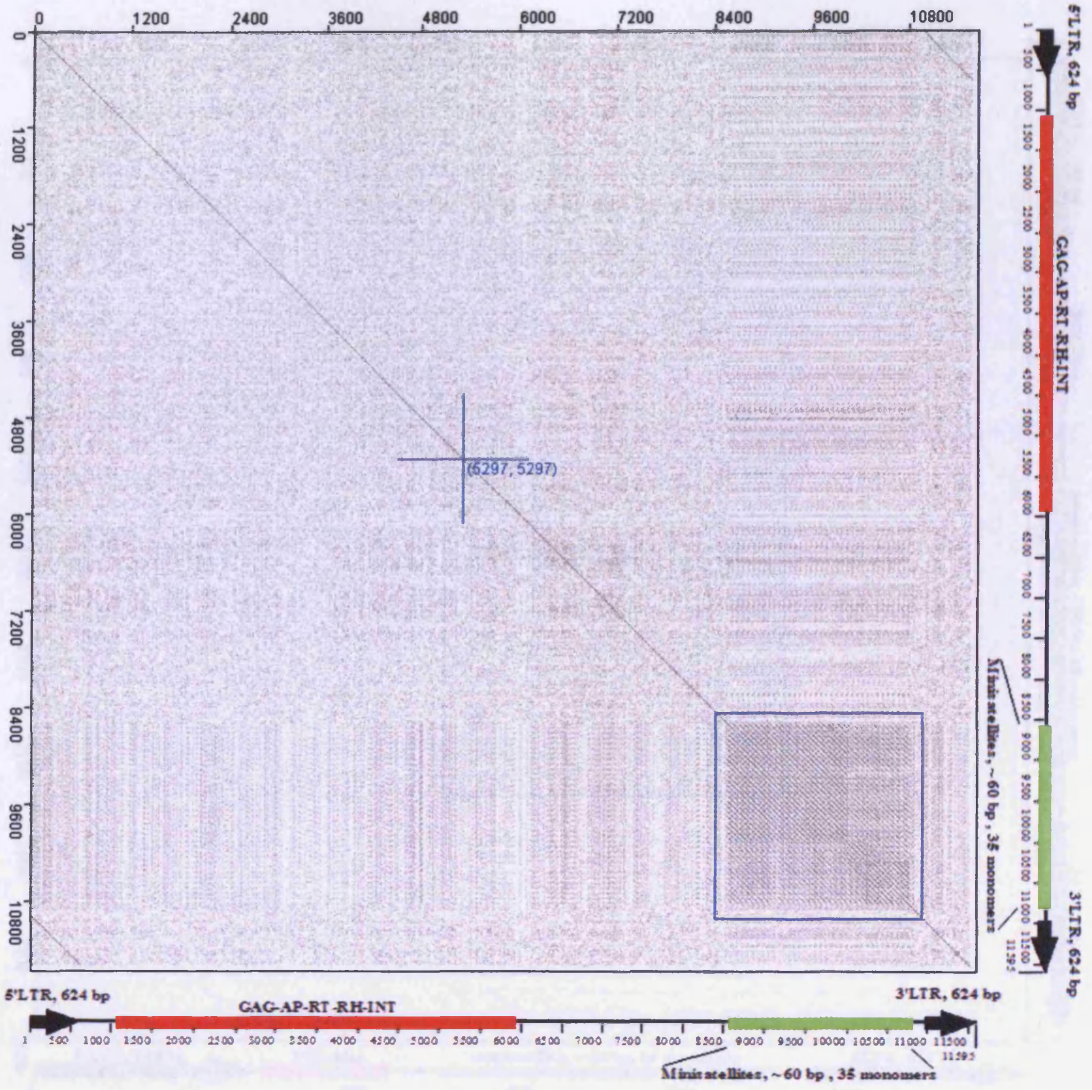


Figure 3.2g: Dotplot analysis of *Metaviridae* full-length element, eCoTR-REs, which contained 35 units of ~ 60 bp minisatellite repeats in its 3'UTS, detected in BAC clone, MA4_78I12, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the 3'UTS region are represented by the blue rectangular boxes.

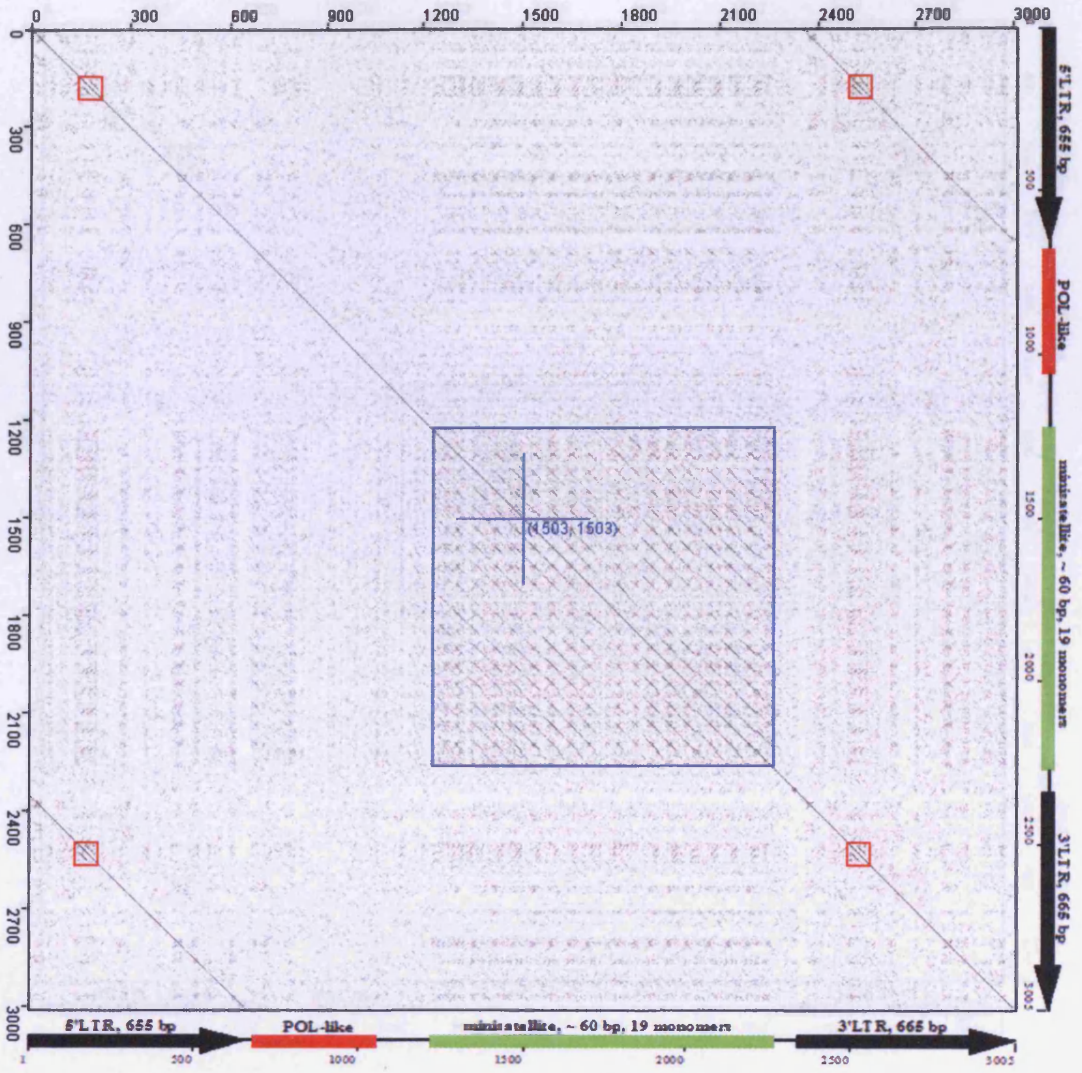


Figure 3.2h: Dotplot analysis of *Morgane*-like element, *Meta*Morgane, which contain 19 units of ~ 60 bp minisatellite repeats in its 3'UTS, detected in BAC clone, MA4_64C22, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the 3'UTS region are represented by the blue rectangular boxes whereas minisatellite repeats found it the LTR region are represented by the red rectangular boxes.

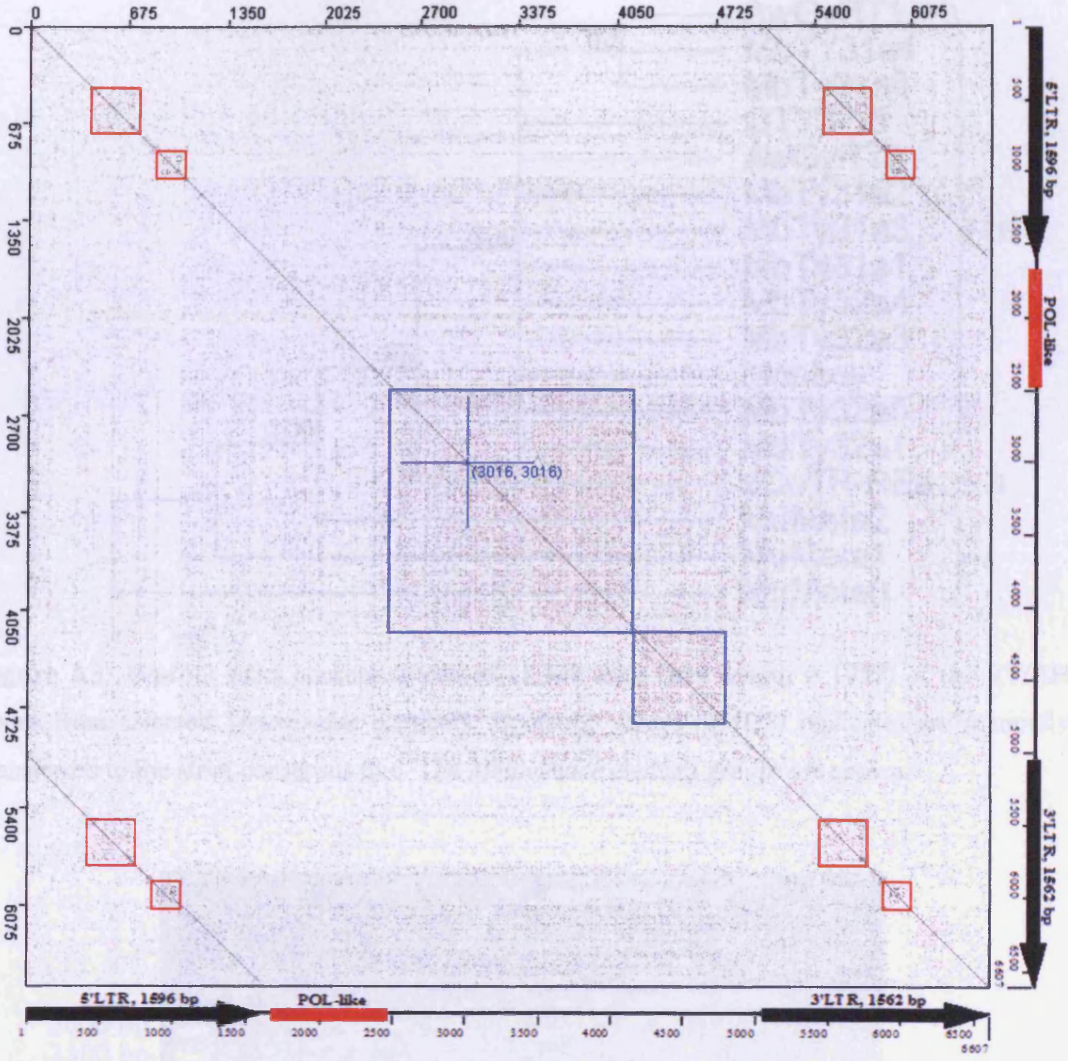


Figure 3.2i: Dotplot analysis of *Morgane*-like element, *Pseudo*Morgane, detected in BAC clone, MBP_31O07, using JDOTTER program (Sonnhammer and Durbin, 1995). The minisatellite repeats found in the LTR region are represented by the red rectangular boxes whereas unknown repeats found in the 3'UTS region are represented by the blue rectangular boxes.

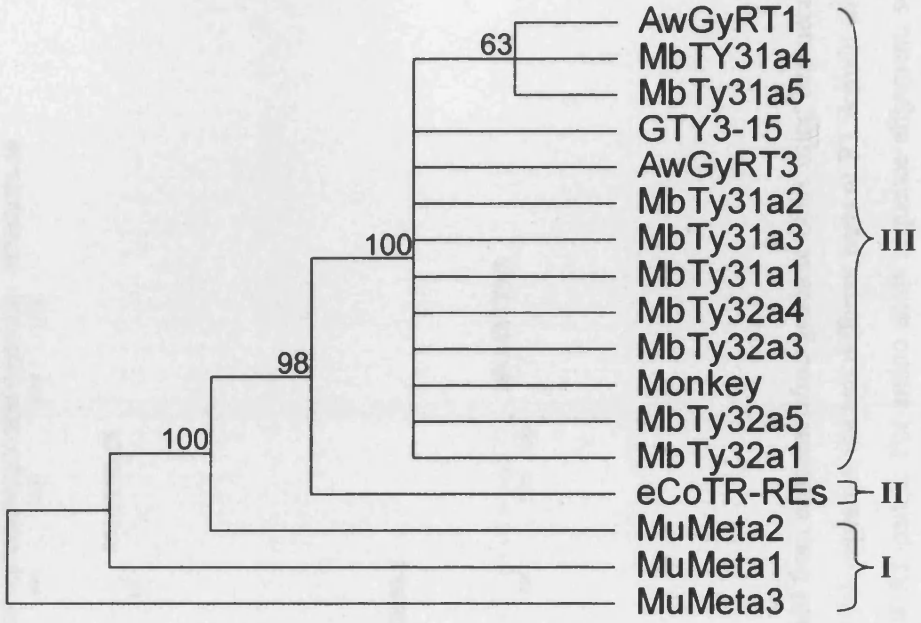


Figure 3.3: Un-root strict consensus tree of 11 MP trees (tree length = 1957) of the RT/RH genes from selected *Metaviridae* elements. Bootstrap values of 1000 replicates are manually transferred to the strict consensus tree. The *Metaviridae* element groups are shown.

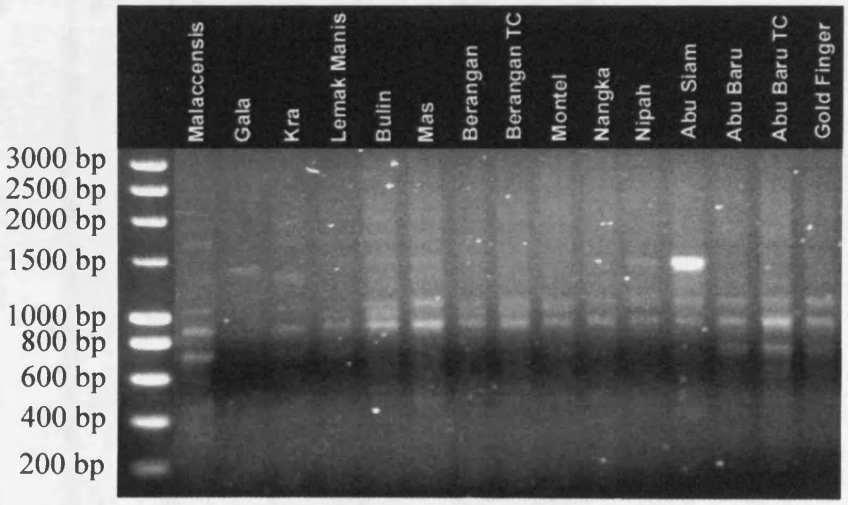


Figure 3.4a: Amplification of *Metaviridae* Group II elements from different *Musa* taxa using degenerate primers, GyRT1 and GyRT4.

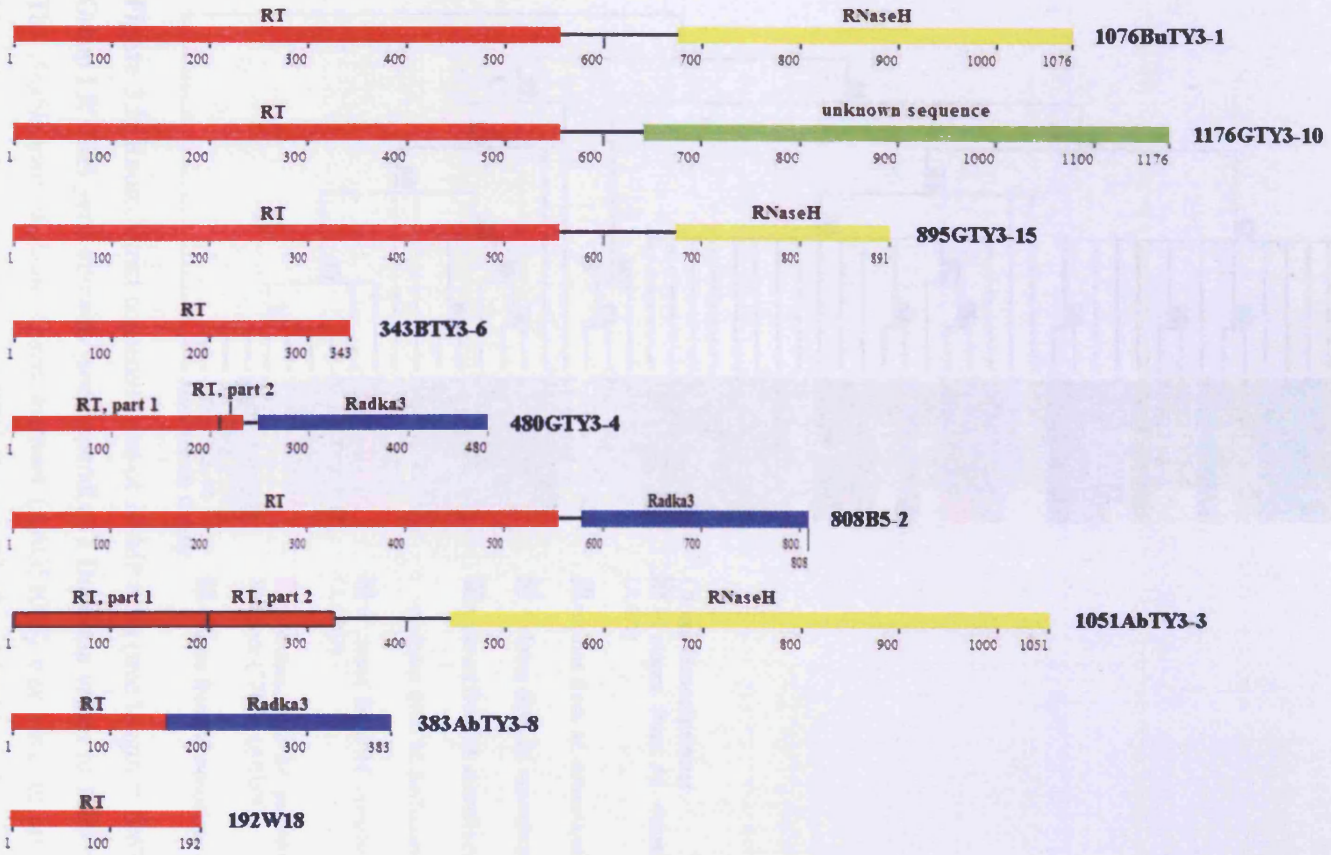


Figure 3.4b: Organization of representative *Metaviridae* Group I element clones amplified from different *Musa* genomic DNA using degenerate primers, GyRT1 and GyRT4. The RT part 1 and 2 in clones, 480GTY3-4 and 1051AbTY3-3, represent the two different parts of RT domain with internal deletion of a 69 amino acid residues at the end of RT part 1 and beginning of RT part 2. For amino acids sequence alignment, see Appendix 1 and for clone description see Figure 3.5.

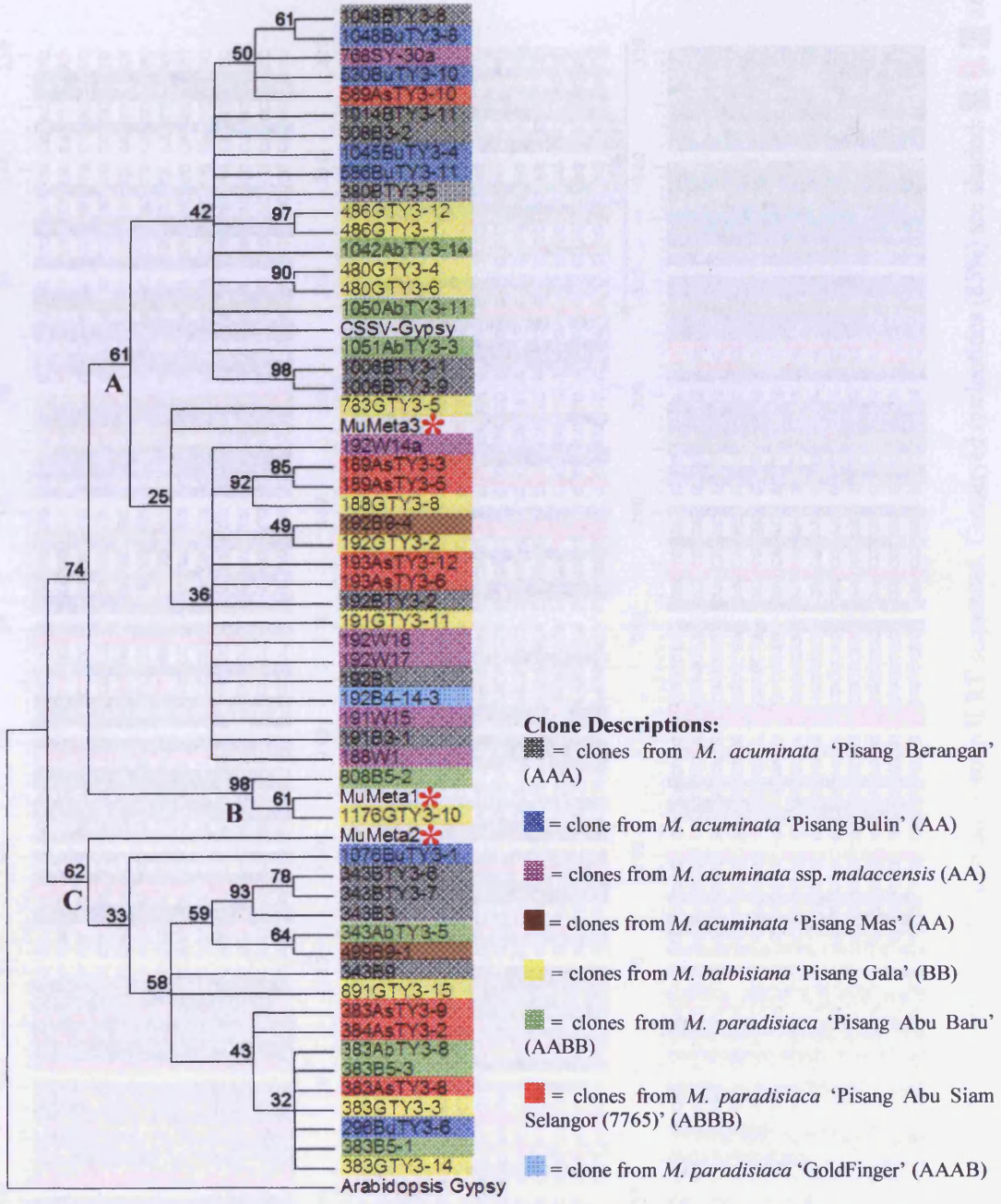


Figure 3.5: Rooted strict consensus tree of 35 MP trees (tree length = 1067) of 54 *Metaviridae* Group I RT/RH gene domains (see Appendix 1). Bootstrap values of 1000 replicates are shown. The *Arabidopsis thaliana Gypsy* element (AAG51046) was used as out-group species. The *Metaviridae* element sequence (CSSV-Gypsy) that found in *Musa Cacao Swollen Shoot Virus* (*CSSV*, AF106947) was retrieved from Genbank database. The three *Metaviridae* Group I FLE sequences from the BAC clones are indicated by the red asterisks (*).

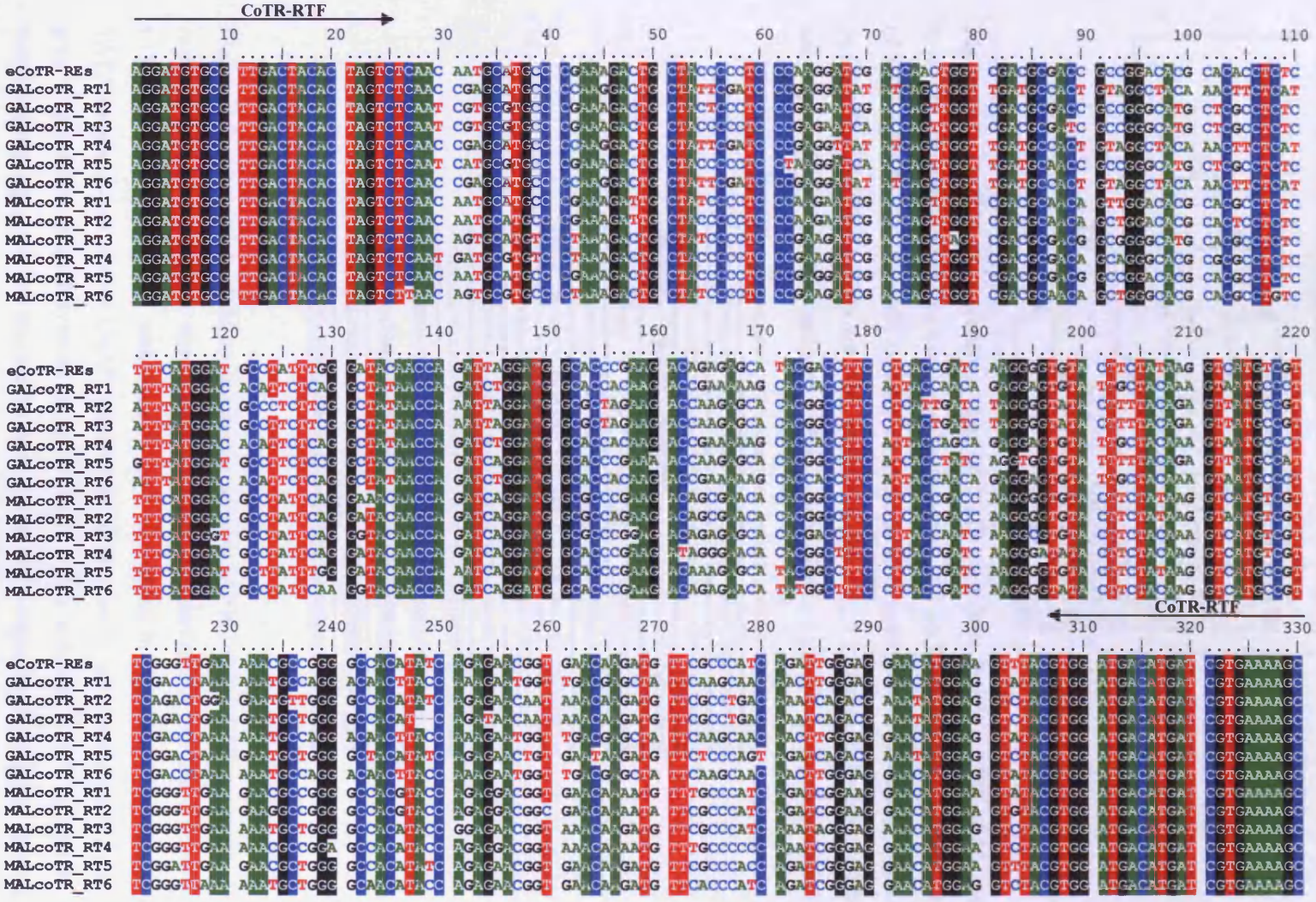


Figure 3.6: Multiple DNA sequence alignment of 13 *Metaviridae* Group II RT sequences. Conserved nucleotides (85%) are shaded: **A**, **T**, **G** and **C**, respectively. Six clones of each species, *M. acuminata* ssp. *malaccensis* (MAL) and *M. balbisiana* 'Pisang Gala' (GAL) were cloned and sequenced. The primer positions are indicated in the alignment.

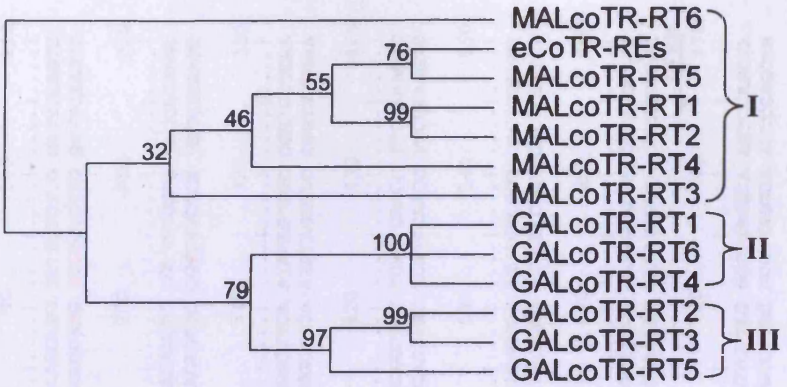


Figure 3.7: Un-root strict consensus tree of 3 MP trees (tree length = 259) of *Metaviridae* Group II elements constructed based on the conserved 330 bp RT sequences. Bootstrap values of 1000 replicates are manually transferred to the strict consensus tree. Three subgroups could be identified; for the *M. acuminata* ssp. *malaccensis* (MAL, A genome) including eCoTR-REs, subgroup I and for the *Musa balbisiana* 'Pisang Gala' (GAL, B genome), subgroup II and III.

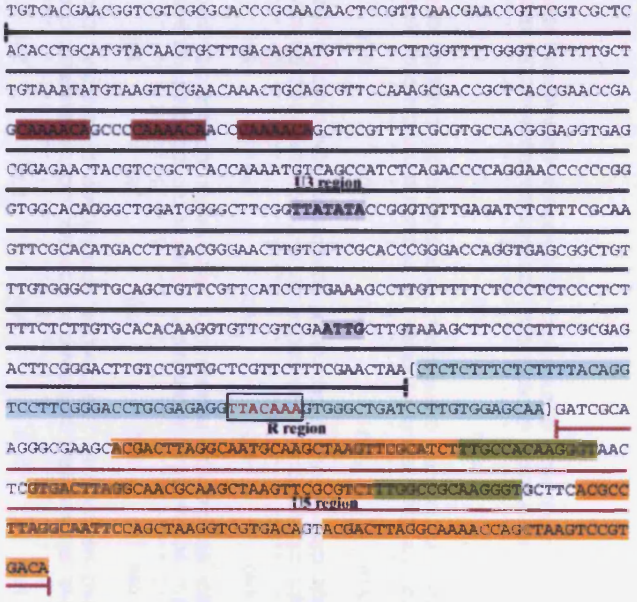


Figure 3.8: Annotation of *Musa acuminata Monkey* (AF143332) 3'LTR sequence. The U3 region is underlined in black whereas the U5 region is underlined in red. The R region of the LTR sequence is identified with brackets and highlighted in light-blue. The nonconsensus TATA sequence and reverse CAAT sequence are delineated by bold letters and highlighted in grey. The putative polyadenylation signal in the R region is boxed. The 7 bp repeats in U3 region are highlighted in red whereas tandem repeats in U5 region are highlighted in orange and green.

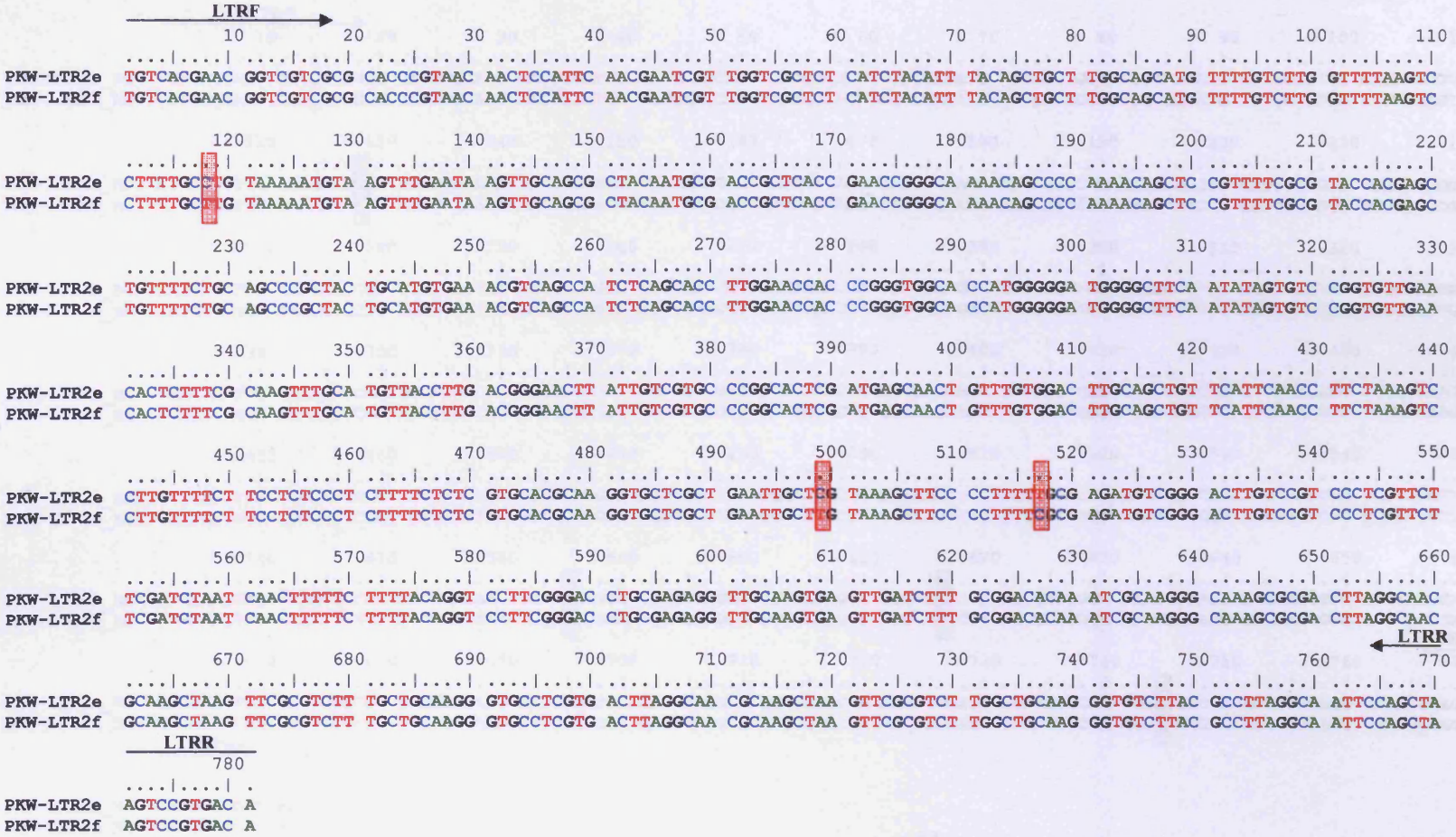


Figure 3.9a: Pairwise alignment of two *Metaviridae* Group III element LTR sequences (PKW-LTR2e & PKW-LTR2f) isolated from *M. balbisiana* 'Pisang Klutuk Wulung'. C↔T transition at three positions were observed and highlighted in red boxes. The primer positions are indicated in the alignment.

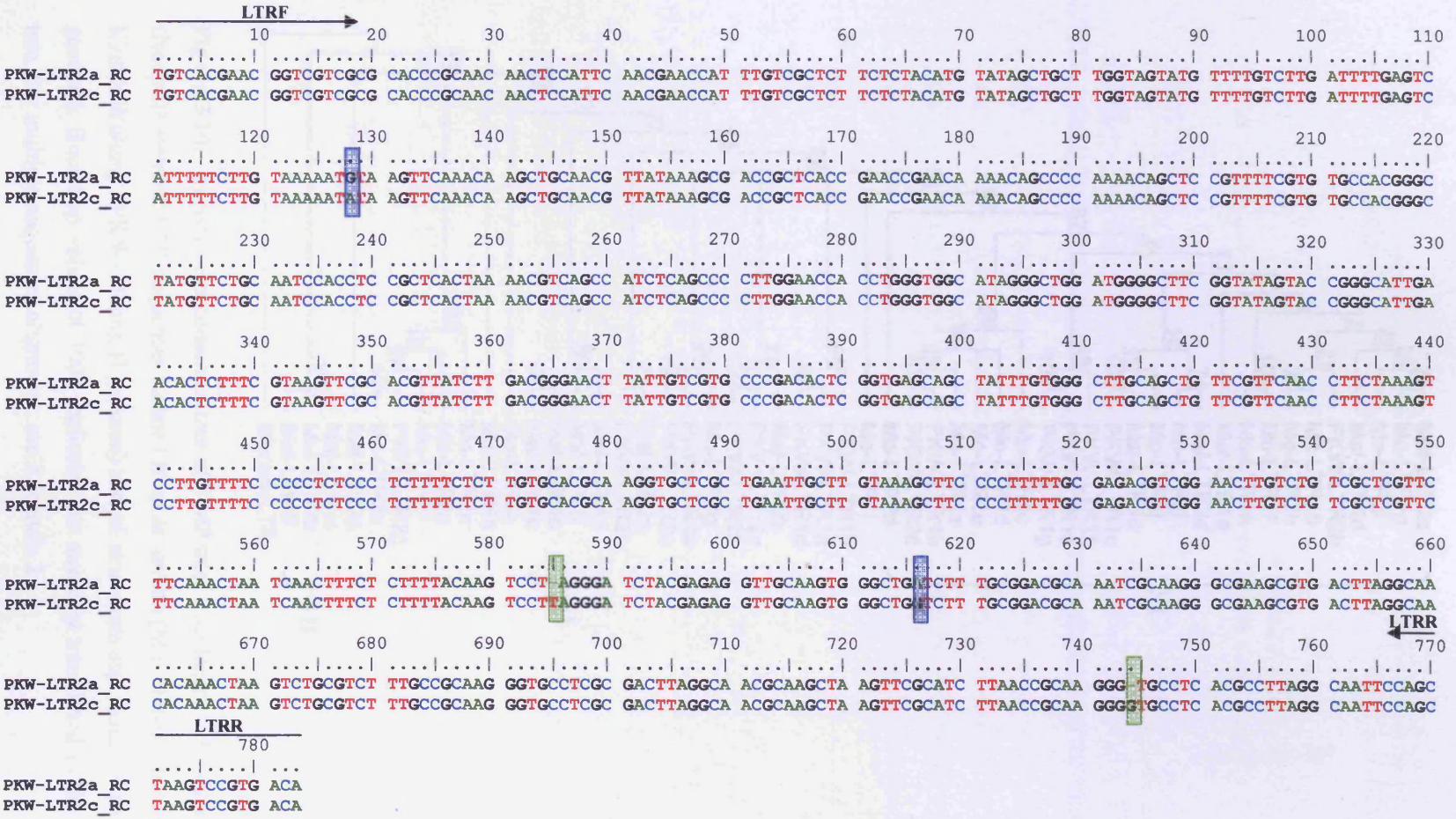


Figure 3.9b: Pairwise alignment of two *Metaviridae* Type III element LTR sequences (PKW-LTR2a & PKW-LTR2c) isolated from *M. balbisiana* 'Pisang Klutuk Wulung'. A↔G transitions and single nucleotide deletion/insertion were highlighted in the blue and green boxes, respectively. The primer positions are indicated in the alignment.

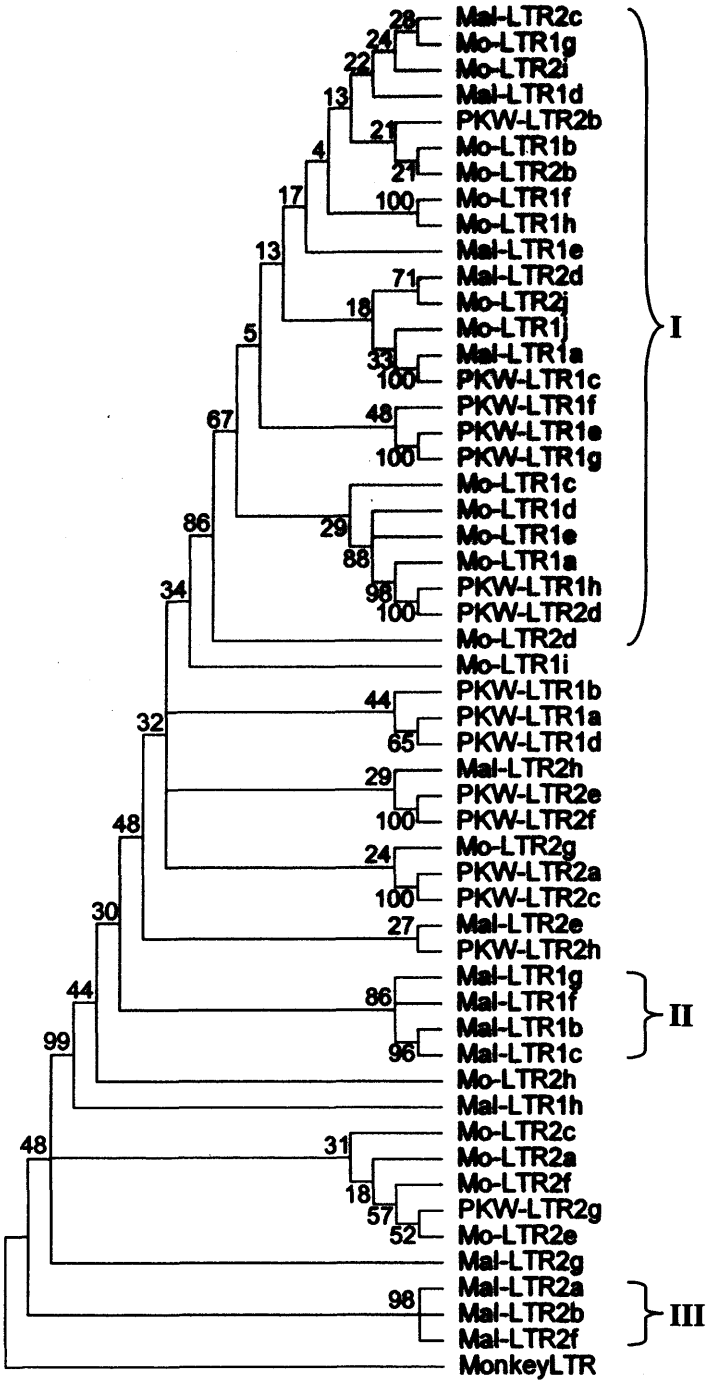


Figure 3.10: Un-root strict consensus tree of 8 MP trees (tree length = 1291) of 52 *Metaviridae* Group III element LTR sequences isolated from *M. ornata* (Mo- clones), *M. balbisiana* 'Pisang Klutuk Wulung' (PKW- clones, B genome) and *M. acuminata* ssp. *malaccensis* (Mal- clones, A genome). Bootstrap values of 1000 replicates are manually transferred to the strict consensus tree. For multiple sequences alignment, see Appendix 2.

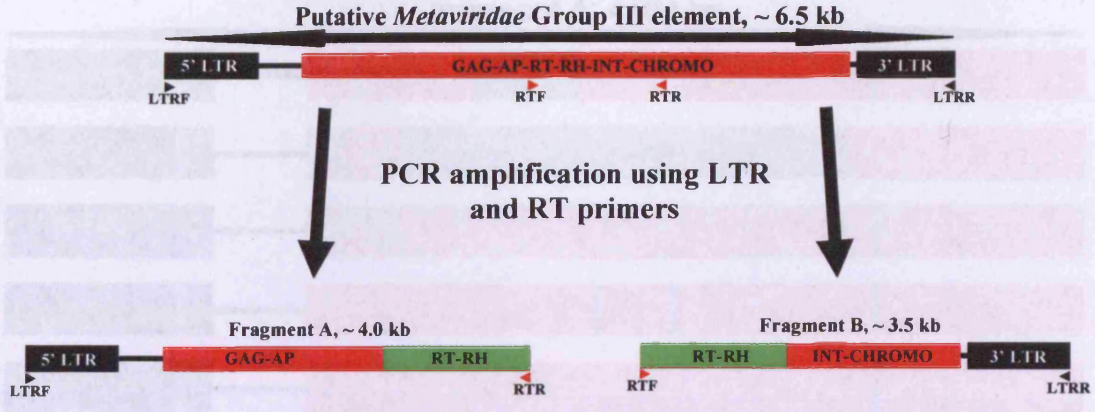


Figure 3.11a: The isolation strategy of the two partial fragments, A and B, of the *Metaviridae* Group III element from total genomic DNA of *Musa balbisiana* 'Pisang Gala'. The LTR and RT primers were designed from the *M. acuminata Monkey* element. The amplified fragments, A and B, which overlap at their RT/RH domains (~ 1.2kb) were then cloned and sequenced using primer walking strategy.

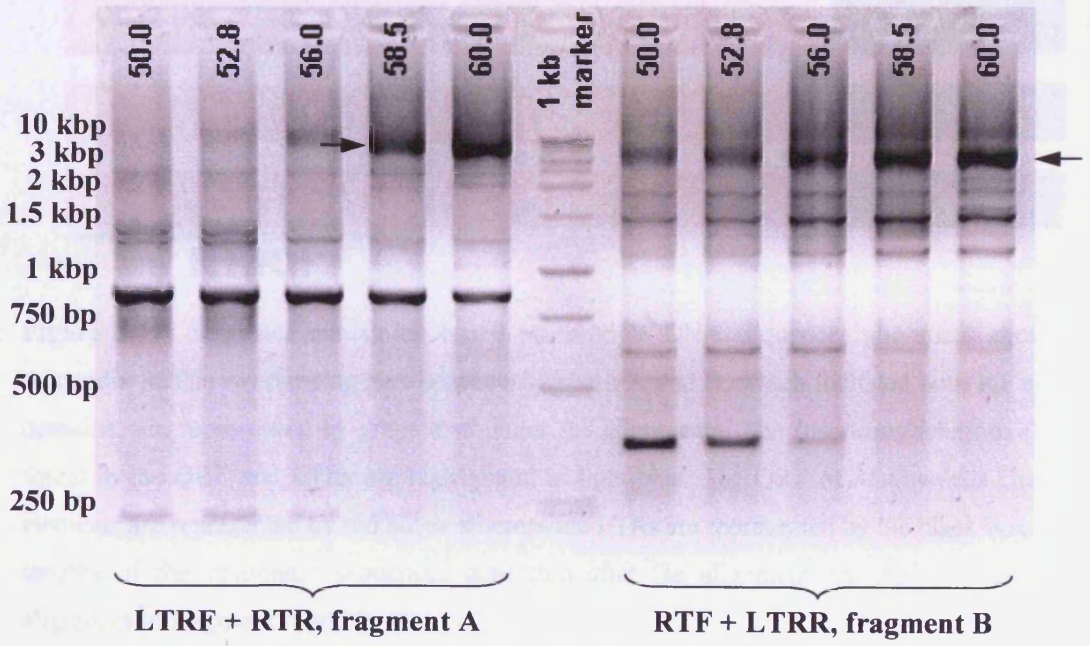


Figure 3.11b: Temperature gradient PCR (from 50°C to 60°C) of *Metaviridae* Group III element amplification using total genomic DNA of *M. balbisiana* 'Pisang Gala'. Arrows represented the expected PCR products size of ~ 4.0 (fragment A) and ~ 3.5 kbp (fragment B). The primer pair sequences are given in Table 3.1.

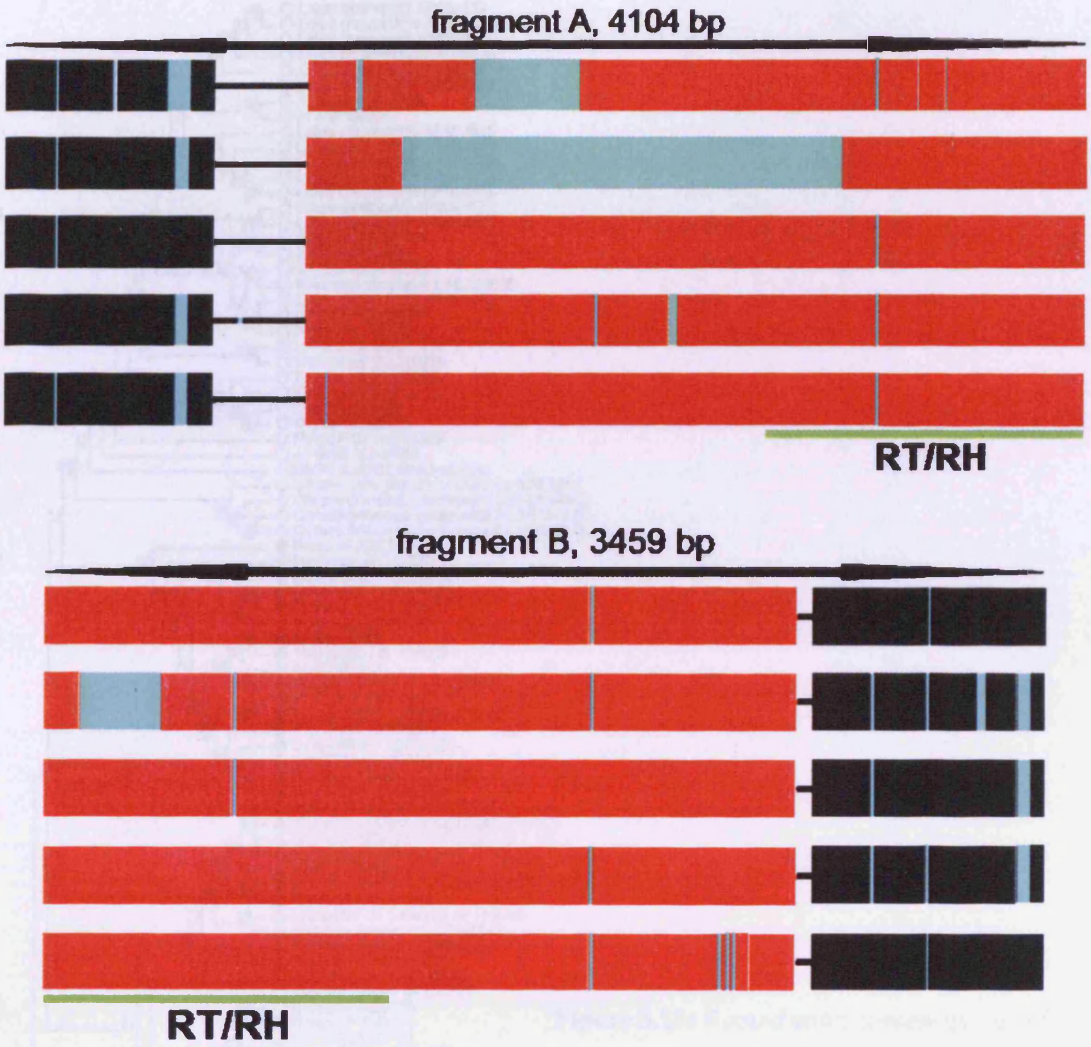


Figure 3.12: Sequence annotation based on multiple DNA sequences alignment showed in Appendix 3. The overlapping part between fragment A and B, which included both RT and RH domains, are represented by green line under the alignments. The insertions/deletions (indels) found in the ORF and LTRs are highlighted in light-blue. The ORF of *Metaviridae* Group III elements are represented by red boxes whereas the LTRs are represented by the black boxes. The lengths of the consensus sequences generated after the alignment are indicated above the alignment of fragment A and B.

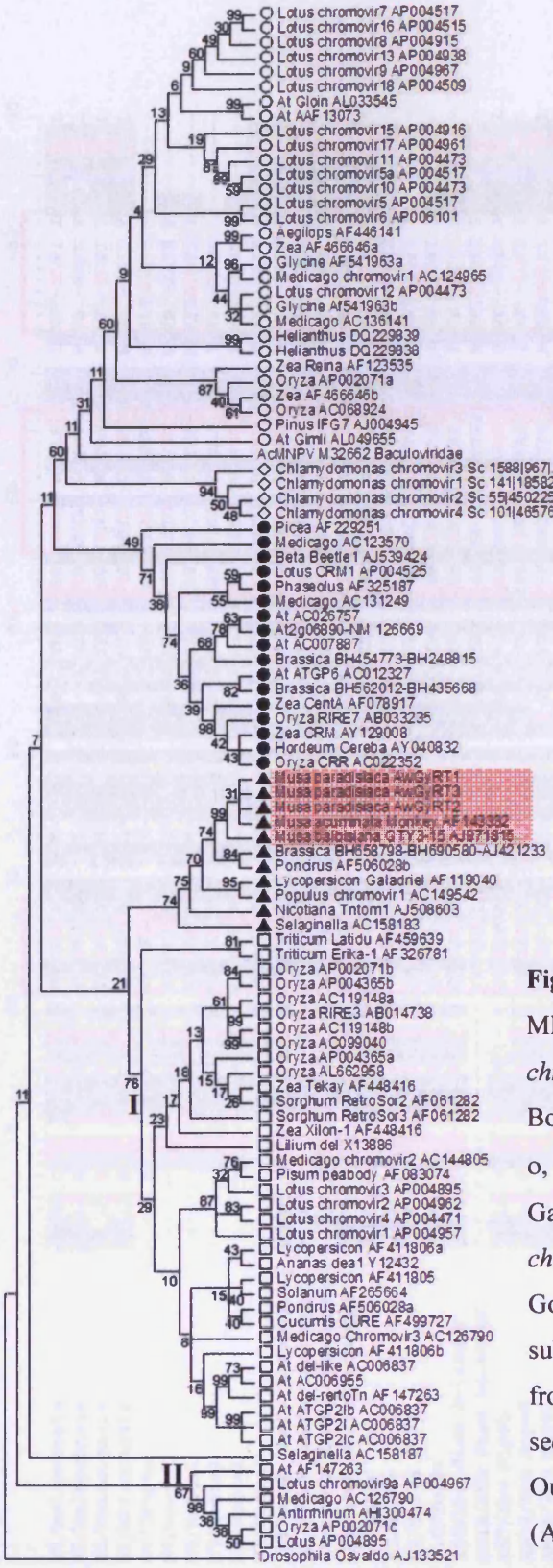


Figure 3.13: Rooted strict consensus tree of 14 MP trees (tree length = 3841) of 106 *chromoviruses* from different organisms Bootstrap values of 1000 replicates are shown. o, ●, ▲, □ and ◊ represent Reina, CRM, Galadriel, Tekay and Chlamyvir groups of *chromoviruses*, respectively (according to Gorinšek *et al.*, 2004). Tekay I and II subgroups are indicated in the tree. Sequences from *Musa* species are shaded in red. For sequence alignment, see Appendix 4. Outgroup taxa = *Drosophila* Osvaldo (AJ133521).

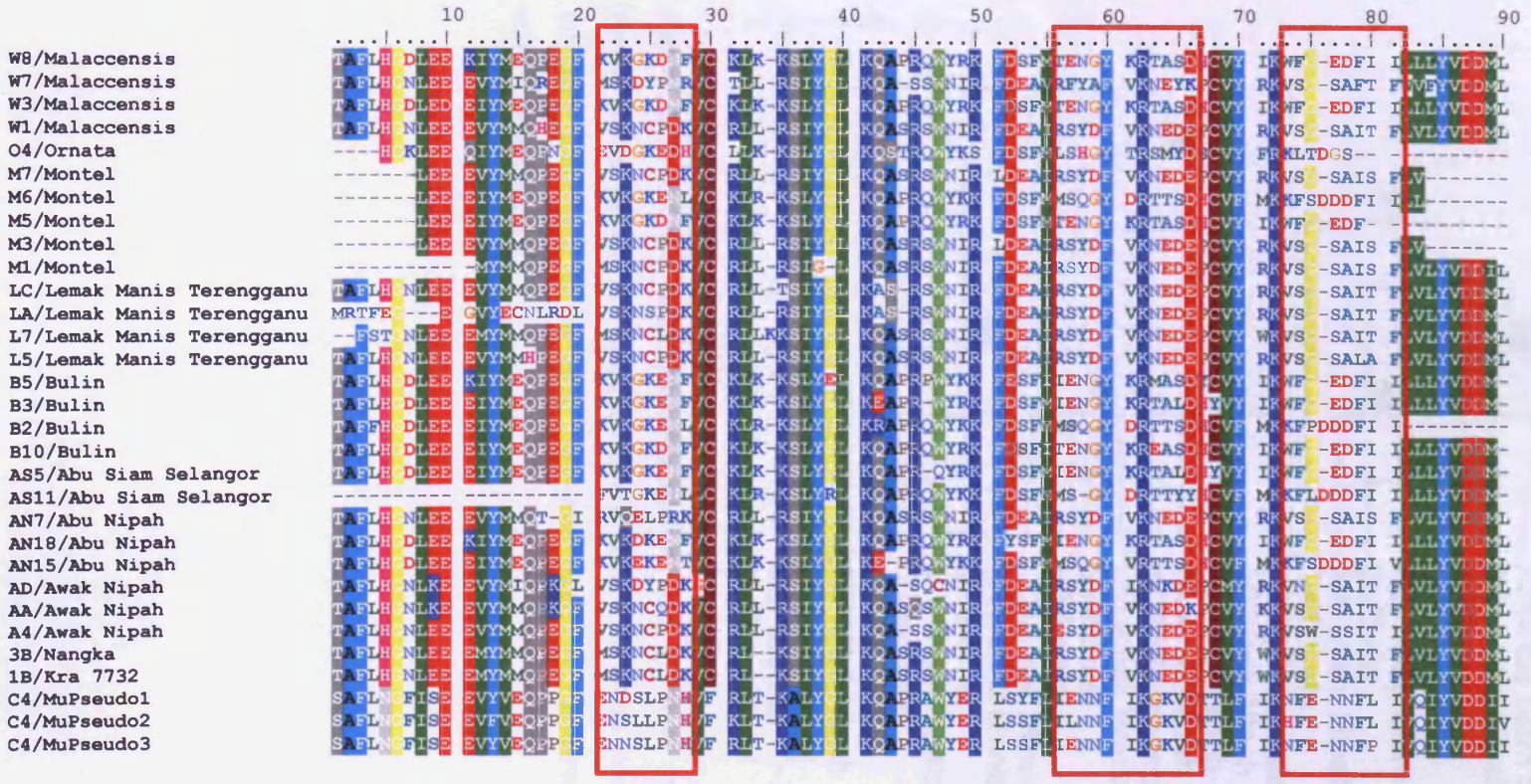


Figure 3.14: Multiple amino acid sequence alignment of *Musa* 28 published *Pseudoviridae* element RT sequences. For clones see Teo *et al.* (2002). Three new *Pseudoviridae* FLEs from *M. acuminata* 'Calcutta 4, detected in BAC clones are C4/MuPseudo1, C4/MuPseudo2 and C4/MuPseudo3. Three variable regions are highlighted in red boxes.

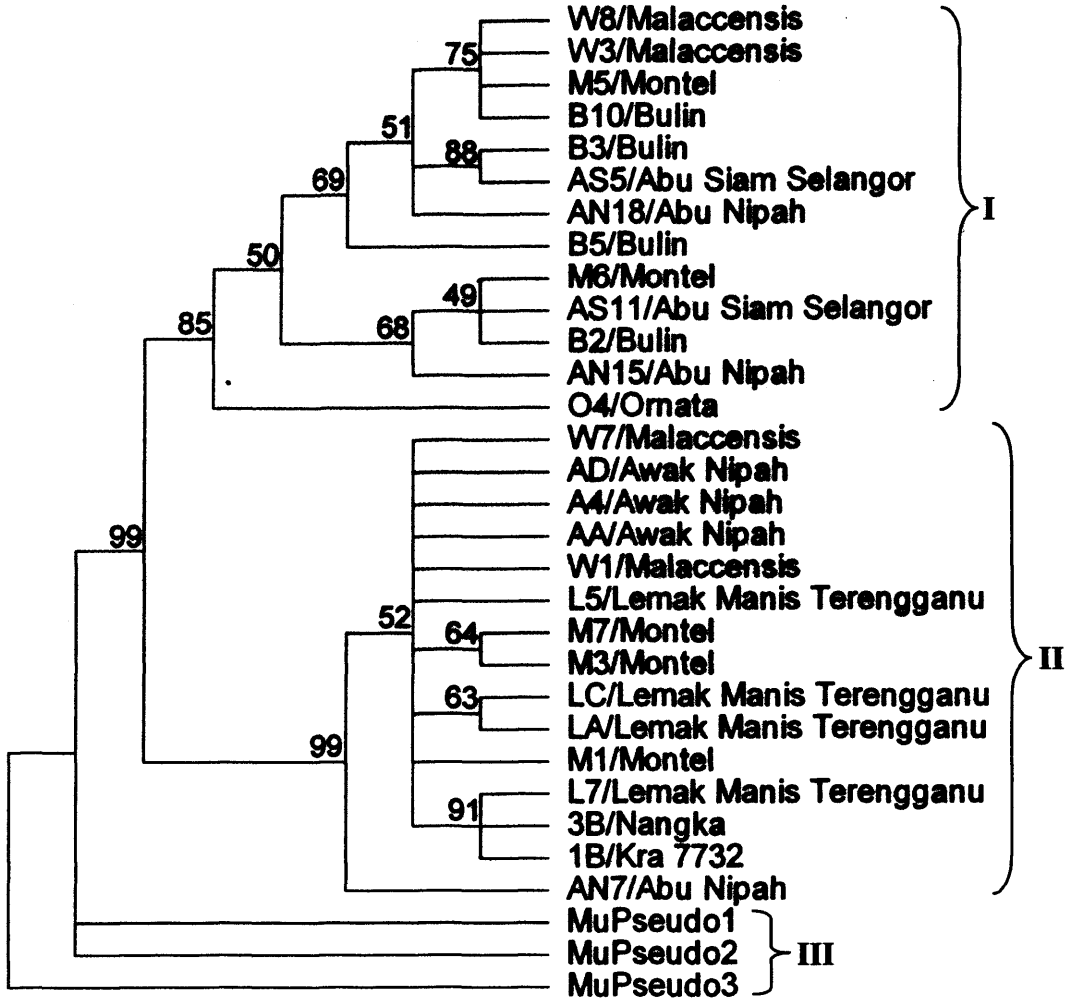


Figure 3.15: Un-root strict consensus tree of 5418 MP trees (tree length = 155) of 31 *Pseudoviridae* element RT genes in *Musa* genomes. Bootstrap values of 1000 replicates are manually transferred to the strict consensus tree. The *Pseudoviridae* element RT gene subgroups are shown.

4.0 Chapter IV: Diversity and organization of tandem DNA repeats

4.1 Summary

Three different classes of tandemly organized repetitive DNA families were isolated from *Musa*: the 5S rDNA, MuTR and CoTR families. Some of 5S rDNA sequences isolated contained a *Pseudoviridae* element RT (PVRT) homologue sequence downstream of the coding region. Different *Musa* taxa were found to have different numbers of 5S rRNA genes with PVRT insertions, A genome *Musa* generally have many more copies than B genome taxa, indeed none at all were detected in *M. balbisiana*, 'Pisang Gala' and 'Butohan 2'. The MuTR family with repeat units of 220bp is present in all *Musa* species tested with small changes in their tandem array patterns. They all contain a 63bp repeat box with homology to the *Metaviridae* LTR region that is postulated to have inserted before the split of the A and B genome taxa. However, since then homogenization and amplification events have given rise to genome-specific MuTR repeat subfamilies. A similar situation is observed for the 60bp minisatellite CoTR repeat family that is inserted, at least in some instances, into a *Metaviridae* element at its 3'UTS region. Phylogenetic analysis separated all the members of the A genome CoTR from the B genome CoTR repeats. In the latter, higher order structures of a 200bp repeat unit were found, consisting of two 60bp units together with some intervening sequences. The fine scale organization and diversity of these three classes of tandemly organized repeats are discussed in detail and evolutionary hypotheses are proposed.

4.2 Introduction

The variety of repetitive DNA sequences that exist in tandemly organized arrays within plant and animal genomes is enormous: ranging from repeats of large, well-defined almost identical units (such as the rRNA and histone genes) to those with complex higher order structures (e.g. telomere associated sequences and centromeric satellite DNAs) and those with individual units consisting of one, two or few base pairs only (e.g. simple sequence repeats)(Streisinger *et al.*, 1966; Heslop-Harrison, 2000a; Schwarzacher, 2003). Tandemly organized repeats undergo variation in the array size due to both recombination and replication processes (Jones and Kafatos, 1982; Moore, 1983). Duplications and deletions of whole units are generated after mispairing at meiotic prophase followed by recombination (unequal crossing over). Mispairing between nucleotides during DNA replication, is also called replication slippage, which occurs particularly in simple sequence repeats which generates variation in copy number (Schmid and Shen, 1985; Walsh, 1987). The variations that can be found in a repeat unit are mutations, base pair substitutions, insertions/deletions and degeneration (Heslop-Harrison, 2000a; Schwarzacher, 2003).

The function of most tandemly organized repeats is still unclear except for the well-defined families such as rRNA and histone genes (Schmidt *et al.*, 1994). While some repeats do have clear function, such as telomeric, centromeric and rDNA, the question still remains of what the role and function if any of the large amounts of repetitive DNA are (Schmidt and Heslop-Harrison, 1998), and speculations have ranged from junk to necessary buffers against mutations, from regulation of gene expression to inert mass that has to be tolerated (Schwarzacher, 2003). Most of tandemly organized repeats are localized in the heterochromatic region of plant and animal chromosomes. Their abundance and distribution patterns are often

chromosome and genome-specific allowing for chromosome identification and tracing of evolutionary events (Schwarzacher, 2003). Different types of tandemly organized repeats have been isolated from many organisms including humans (Okumura *et al.*, 1987); *Drosophila* (Kuhn and Sene, 2005); *Arabidopsis* (Martinez-Zapater *et al.*, 1986; Maluszynska and Heslop-Harrison, 1991; Heslop-Harrison *et al.*, 1999); millet (Kamm *et al.*, 1994); rice (Aragon-Alcaide *et al.*, 1996; Nonomura and Kurata, 1999; Cheng *et al.*, 2002); maize (Ananiev *et al.*, 1998a; b); wheat (Anamthawat-Jonsson and Heslop-Harrison, 1993; Vershinin *et al.*, 1995; Contento *et al.*, 2005; Han *et al.*, 2005) and potato (Tek *et al.*, 2005). Some of these tandemly organized repeats, in particular the centromeric repeats, show a close relationship to LTR retrotransposons which also are localized at the centromeric region of some plant chromosomes (Miller *et al.*, 1998; Cheng *et al.*, 2002; Zhong *et al.*, 2002; Nagaki *et al.*, 2005; Yang *et al.*, 2005; Houben *et al.*, 2007). Only a few reports of tandem repeats in *Musa* have been published (see section 1.3.4).

A variety of different evolutionary mechanisms of sequence amplification are probably operational in the evolution of each tandem repeat DNA family, but the timing and interaction of the various mechanisms with each other might be critical and specific for different repeat families and species evidenced by their varying abundance and chromosomal distribution patterns (Cuadrado and Schwarzacher, 1998). The evolution of different tandem repeat DNA does not always show the characteristics of 'a molecular clock' with constant mutation rates postulating evolutionary waves perhaps occurring during periods of rapid speciation or in response to various stresses (Heslop-Harrison, 2000). It is also evident that homogenization events of tandemly organized arrays occur when continual production of copy number variation happened between those arrays (Smith, 1974,

1976; Tarfor, 1974; Black and Gibson, 1974; Ohta, 1980) and also during different periods of species evolution (Heslop-Harrison, 2000).

In order to understand the organization of different types of *Musa* tandem repetitive DNA and their relationship to the dispersed retrotransposons in *Musa* genomes, three different types of tandemly organized repeats were isolated by PCR amplification, cloned and sequenced. Sequence analysis, annotation and phylogenetic analysis were carried out to study the evolutionary history of these repeats.

4.3 Materials and Methods

4.3.1 Isolation of repetitive DNA sequences

Different primers were designed for the amplification of repeats from genomic DNA by PCR. In all the cases, FastPCR© version 3.1.41 software was used; primer sequences are given in Table 4.1. PCR parameters taking into account the specific annealing temperatures of Table 4.1 followed section 2.2.2. PCR products were purified, cloned and sequenced following section 2.2.4 and 2.2.5. The DNA sequences of the repetitive DNA clones will be submitted to the EMBL database.

4.3.2 Sequence and phylogenetic analysis of repetitive DNA

Multiple sequence alignments for repetitive DNA were carried out using ClustalW software embedded in BioEdit version 7.0.5.3 (Hall, 1999). Furthermore, phylogenetic trees were constructed using MEGA version 3.1 (Kumar *et al.*, 2004) with Kimura's two parameters. Sequence annotations were carried out using JAVA-based graphic alignment tool for comparative sequence analysis, GATAligner and GATAPlotter (Nix and Eisen, 2005).

4.4 Results

4.4.1 Organization and sequence analysis of 5S rRNA genes in *Musa* species

Primers (PTA794For and PTA794REV) were designed from the coding region of pTa794 from wheat (Gerlach and Bedbrook, 1979) (Table 4.1). The expected size of the PCR product using these primers to amplify the complete unit of the 5S rDNA sequence is ~ 400 bp (Table 4.1). However, a band of ~ 600 bp was produced after PCR amplification using genomic DNA of *M. acuminata* 'Pisang Berangan' (AAA) as template (Figure 4.1). High molecular weight smearing was also observed on the gel. The band at ~ 600 bp was excised from the gel and cloned. Three clones were selected and sequenced.

The sizes of the clones were 609 bp for pMa794.1, 606 bp for pMa794.2 and 598 bp for pMa794.3 and show 94 – 96% homology between them. Multiple DNA sequence alignments detected few insertion/deletions (indels) and a small number of base pair changes in the three 5S rDNA clones from *M. acuminata* 'Pisang Berangan' (Figure 4.2). DNA-DNA homology search against the Genbank database showed high homology (85 – 94%) to Radka2 containing the 5S rDNA sequence from *Musa balbisiana* 'Cameroun' (BB) (Valárik *et al.*, 2002), five genomic sequences of *M. acuminata* 'Calcutta 4' (AA) and 5S rRNA genes from other organisms. DNA-protein homology searches of the three clones to the database identified homology to the reverse transcriptase gene of *Pseudoviridae* elements from different organisms. The region, 116 or 134 bp long, which shows homology to *Pseudoviridae* element RT gene (PVRT, Figure 4.3), is inserted in antisense orientation from nt 125 to nt 259 for pMa794.1, from nt 127 to nt 261 for pMa794.2 and from nt 129 to nt 245 for pMa794.3 (Figure 4.2).

Detailed comparative sequence annotation of pMa794.1 and Radka2 clone from *M. balbisiana* 'Cameroun' (BB genome, Valárik *et al.*, 2002) showed that the PVRT homologous region is located downstream of the 5S coding region (Figure 4.3). The conserved active amino acid motif of PVRT gene (5'-DDDFVILLLYVNDTF-3', Xiong and Eickbush, 1990) was observed in all three clones. A short direct duplicated sequence of 34 bp (Dup), which comprises the end of the 5S rRNA gene coding region (5'-TGTTGCACTCCT-3') followed by part of the ITS region (5'-TTTTGCGCCCGAGCGACCAAAA-3'), flanks the PVRT homologous region in all three 5S rDNA clones from *M. acuminata* 'Pisang Berangan' (Figure 4.2 and 4.3). Only one copy of this duplicated sequence found in the Radka2 clone. The 5 bp target site duplication, indicative of *Pseudoviridae* element insertion (CCTCG and CCTCT) was observed at the 5' end of the first and 3' end of the second duplicated sequence (Figure 4.3). On the other hand, duplication and rearrangement of two sequences of the ITS regions of pMa794.1 in Radka 2 was observed together with otherwise divergent sequence stretches (Figure 4.3).

Sequence analysis of the five genomic sequences from *M. acuminata* 'Calcutta 4' showed that they all contain at least parts of the 5S rRNA genes. One particular sequence, MA4-8B23 (Genbank accession number: gi|89746896), contained the complete 5S rRNA gene unit with the coding region of 120 bp and the ITS of 476 bp. DNA sequence comparison of the clone pMa794.1 isolated from *M. acuminata* 'Pisang Berangan' to MA4-8B23 revealed the 5S rRNA coding region (Figure 4.4). Pairwise DNA sequence alignment (Figure 4.4A) and graphical dotplot analysis between pMa794.1 and MA4-8B23 (Figure 4.4B) detected a similar PVRT homologous region with identity of 89.7% in MA4-8B23, albeit with a 45 bp

deletion at the 3' end, and a 51 bp stretch at the end of the ITS region that is missing in pMa794.1 (Figure 4.4).

To determine the presence and frequency of PVRT homologous regions in different *Musa* species, primers (RIAP01F and RIAP01R, Table 4.1) were designed from the 34 bp duplicated sequence flanking the PVRT (Figure 4.3). A ~ 250 bp band should be amplified if there is a similar PVRT insertion into the 5S rRNA genes while the appearance of a PCR band of ~ 460 bp is the product of PCR amplification from two units of 5S rDNA when there is no insertion. The expected ~ 250 bp PCR bands with high intensities were observed in all *M. acuminata* A taxa and *M. balbisiana* 'Lal Velchi' (Figure 4.5) whereas low band intensities were observed in four *M. balbisiana* taxa ('Pisang Tani', 'Pisang Klutuk Wulung', 'Pisang Batu' and 'Honduras'). No expected ~ 250 bp PCR band was observed for *M. balbisiana* 'Pisang Gala' and 'Butohan 2' (Figure 4.5). 460 bp fragments of different intensity were observed in all *Musa* taxa tested, except *M. acuminata* ssp. *zebrina*. PCR bands with size of ~ 200 bp were observed in 6 *Musa* taxa tested probably indicating a different type of insertion or duplication. A sharp band was observed at ~ 600 bp for *M. acuminata* ssp. *errans* (Figure 4.5).

4.4.2 Organization and diversity of tandem repeats, MuTRs, in *Musa* genomes

The second tandem repeat family analysed was the MuTR family that was identified flanking retrotransposons (see Chapter 5). Primers BGyFor and BGyRev (Table 4.1), which had initially been designed to amplify a B-genome specific 275 bp long Inter-Retrotransposon Amplified Polymorphism (IRAP) band, were used to amplify the MuTR repeat from *M. ornata* (Figure 4.6a).

A ladder pattern, indicative of a tandem array was first observed in *M. ornata* using primers BGyFor and BGyRev (Figure 4.6). This tandem repeat family from *Musa* species was named Musa Tandem Repeat (MuTR) and clones were designated specifically to individual species. For instance, clones with MuTR isolated from *M. ornata* were called Musa ornata Tandem Repeat (MoTR). The ~ 220 bp and ~ 450 bp bands of this tandem array from *M. ornata* (Figure 4.6b) were cloned and four positive clones from each band were sequenced. Sequence analysis of the ~ 220 bp (MoTR_1) and ~ 450 bp (MoTR_2) clones showed high homology between the clones and all four MoTR_2 clones are comprised of two full MoTR_1 repeat sequences.

Virtual restriction digestion of all four MoTR_2 clones detected several restriction enzyme sites: *Dde*I (5'-CTNAG-3'), *Sma*I (5'-CCCGGG-3'), *Ava*I (5'-CYCGRG-3'), *Eco*RII and *Bst*NI (5'-CCWGG-3'), *Hpa*II and *Msp*I (5'-CCGG-3'), *Scr*FI (5'-CCNGG-3'), *Fok*I (5'-GGATG-3') (Figure 4.7). Most restriction enzymes generated 220 bp fragments except for *Dde*I which produced 110 bp fragments (except for MoTR_2d). The two different 110 bp fragments defined by *Dde*I were designated MuTRa and MuTRb respectively, and fragments assigned to *M. ornata*, MoTRa and MoTRb respectively (Figure 4.7). The overall average homology between MoTRa and MoTRb in MoTR_1 and MoTR_2 clones is 48.8%. The average G+C content of MoTRa and MoTRb sequences is 69.1% and 54.5%, respectively (Figure 4.7). Blast search analysis of all the MoTRa+b repeats (220 bp) against Genbank DNA database showed that MoTRa+b repeats contained a region of 63 bp with ≥ 90% homology to the LTR region of the *Monkey* retrotransposon sequence (*M. acuminata* 'Grande Naine'; AF143332), which explains the amplification of this repeat by the IRAP primers (BGyFor and BGyRev). This 63 bp

Monkey retrotransposon LTR homology region is designated as 63 bp box. All the restriction enzyme sites identified for MoTRa+b repeats were found in the 63 bp box except for the *Fok*I site where the cutting site is located 4 bp further downstream of this box (Figure 4.7).

For *M. acuminata* ssp. *malaccensis* and *M. balbisiana* 'Pisang Klutuk Wulung', primers MuTR1aF and MuTR1bR were used (Figure 4.6a) to amplify the complete repeat unit of MuTR with its predicted length of 220 bp. These primers (MuTR1aF and MuTR1bR) are located at the 5' end of the MuTRa subrepeat and 3' end of the MuTRb subrepeat, respectively (Figure 4.6a). The monomer (220 bp), dimer (~ 450 bp) and trimer (~ 650 bp) bands observed in *malaccensis* and PKW were cloned and 8 clones from each species were selected for sequencing. High homology was observed in both subrepeats from the three different species, with subrepeat A showing slightly higher variation than subrepeat B. The 63 bp box was also detected in all the Mal_TR and PKW_TR clones. All MuTRa sequences from three different species showed high 86 to 93% DNA-DNA homology between each other. MuTRa sequences from the same species showed higher sequence homology (91-100%). For MuTRb sequences, the same situation was observed where intra-species sequences showed higher homology (87-100%) compared to inter-species sequences and MuTRb sequences from *M. acuminata* ssp. *malaccensis* generally showed higher percentage (90 – 92%) of DNA-DNA homology to sequences from *M. ornata* compared to sequences from *M. balbisiana* 'Pisang Klutuk Wulung' (86 – 90%) except for PKW_TR3c_b2 monomeric sequence (96% homology).

Virtual restriction digestion detected the same restriction enzyme sites in all Mal_TR and PKW_TR clones (Figure 4.8). Interestingly, the *Dde*I restriction enzyme would produce repeated fragments of 220 bp, in Mal_TR and PKW_TR

clones, rather than the 110 bp ones that were observed in MoTR clones (compare Figure 4.7 and 4.8a). The *Dde*I site was found at the MuTRa/b and b/a border in *M. ornata* clones, but only at the MuTRb/a border in *malaccensis* and PKW clones. The *Eco*RII restriction site at the beginning of the MuTRa fragment was missing in some clones and resulted in fragments of 440 bp in clones PKW-TR3b and PKW-TR3f (Figure 4.8b).

Figure 4.9 shows the sequence alignment of 37 MuTRa and 38 MuTRb subrepeats from *malaccensis* (Mal_TRa and Mal_TRb), PKW (PKW_TRa and PKW_TRb) and *M. ornata* (MoTRa and MoTRb). Several deletions were observed in the multiple sequence alignments of the MuTRa subrepeats (Figure 4.9a) whereas only one deletion was observed for MuTRb subrepeat sequence alignment (Figure 4.9b). Regions that were conserved, with fixed nucleotide substitutions (MuTRa and MuTRb) from the same *Musa* species were observed in the multiple sequence alignments (Figure 4.9). For example, two C ↔ T transitions were fixed at nucleotide positions 67 and 68, and a C ↔ A transversion at nucleotide position 58 in the MuTRa subrepeats from the same species (Figure 4.9a). In Figure 4.9b, a fixed C ↔ T transition was observed at nucleotide position 56.

The phylogenetic tree of MuTRa sequences, which was constructed using the Maximum Parsimony (MP) method, distinguished three groups, each consisting of MuTRa sequences from a different species (Figure 4.10). The MuTRa sequences from PKW can be subdivided into three subgroups. For MuTRb sequences, the Maximum Parsimony (MP) tree also showed three groups depending on species (Figure 4.11). However, one PKW_TRb sequence (PKW_TR3c_b2) was grouped together with MuTRb sequences from *malaccensis*. Three MuTRb subgroups were

observed for *M. ornata*. Interestingly, the PKW_TRa and b sequences from the same clone were distributed in different subgroups (Figure 4.10 and 4.11).

4.4.3 Genome-wide amplification of MuTR repeats

In order to study the distribution of MuTR sequences within *Musa* species, three primer pairs originally used to isolate the MuTR sequences (Table 4.1) were used for PCR amplification of total genomic DNA templates from 19 *Musa* species that represent four different sections of the genus *Musa* (*Eumusa*, *Rhodochlamys*, *Callimusa* and *Australimusa*), for accession origin, see Table 2.1.

For the MuTRF + MuTRR primer pair (Figure 4.12a), consistent amplification of the band at ~ 220 bp (MuTRa + MuTRb) across most *Musa* species together with dimer, trimer and other distinctive bands and a high molecular weight smearing was observed except for *M. gracilis* (*Callimusa* section) where no amplification was observed on the gel (Figure 4.12a). The dimer and trimer bands (440 bp and 660 bp) were observed except for *M. violascens* (*Rhodochlamys* section) where only the monomeric band (~ 220 bp) was amplified. An additional band of ~ 180 bp was observed for *Musa* species from four *Musa* section, except for *M. balbisiana* 'Pisang Gala' (*Eumusa* section), *M. violascens* (*Rhodochlamys* section), *M. beccarii* and *M. gracilis* (*Callimusa* section), and *M. textilis* (*Australimusa* section). For *M. acuminata* ssp. *banksii* (*Eumusa* section), additional bands at ~ 300, ~ 500 and ~ 710 bp were observed whereas for *M. lolodensis* (*Australimusa* section), three additional bands at ~ 510, ~ 710, ~ 910 bp were found (Figure 4.12a).

The primer combination of MuTR1aF and MuTR1bR designed from a different region of MuTRa and MuTRb sequences (Table 4.1) amplified the monomer, dimer, trimer and tetramer in *Eumusa* and *Rhodochlamys* sections with no extra bands but a

background smear throughout species from these two sections (Figure 4.12b). Very strong monomeric bands of slightly variable length (around 220 bp) together with an occasional dimer were amplified in *M. violascens*, *M. beccarii*, *M. borneensis*, *M. coccinea* and *M. gracilis* (Figure 4.12b)

For primer combinations of MuTR1aF and MuTR1aR (Table 4.1) which targeted only the MuTRa subrepeat sequences, strong monomeric bands of 110 bp were observed across species from different sections together with the dimer at 330 bp (110 bp + 220 bp) and trimer at 550 bp (110 bp + 440 bp) and high molecular weight smearing in most of the *Musa* species (Figure 4.12c). Notably, only the monomeric band (110 bp) was amplified in *Musa violascens* (*Rhodochlamys* section) (Figure 4.12c).

4.4.4 Sequence diversity of minisatellite, CoTR, in *Metaviridae* element

A long stretch of 35 minisatellite sequences with a monomer size of ~ 60 bp was observed in the *Metaviridae* Group II element eCoTR-REs from BAC clone MA4_78I12 using dot-plot analysis (Figure 3.2g) and is designated as $CoTR_{60}$. Primers designed to target the polypurine tract (PPT) of eCoTR-REs *Metaviridae* element and the minisatellite sequence (Table 4.1; Figure 4.13) were used to amplify this region in *M. acuminata* ssp. *malaccensis* (AA), *M. acuminata* 'Calcutta 4' (AA), *M. acuminata* ssp. *banksii* (AA) and *M. balbisiana* 'Pisang Gala' (BB). Temperature gradient PCR using this primer pair (MTY1TRF + MTY1TRR01) showed different tandem array patterns between the A and B genome *Musa* taxa together with high molecular weight smearing (Figure 4.14). In *malaccensis*, the expected 60 bp ladder pattern starting with 220 bp monomeric band (see Figure 4.13 for explanation of PCR fragment sites) was observed, whereas in *banksii*, additional bands were visible

at 310 bp and the higher multimers were missing. Extra bands about 20 bp longer than the expected ladder pattern are visible in 'Calcutta 4'. For 'Pisang Gala', a higher order ladder banding pattern with 200 bp increments starting from monomer band at 200 bp was observed at 50°C and many more bands together with a strong band at 150 bp were visible at 45°C. Selected clear and sharp PCR bands from *malaccensis* (AA) and 'Pisang Gala' (BB) were cloned and three positive clones sequenced and designated as MalcoTR and GalcoTR, respectively.

The length and sequence of the MalcoTR monomers from *malaccensis* showed 70 – 93% homology to the $CoTR_{60}$ repeats in eCoTR-REs *Metaviridae* elements in consistency with the ladder pattern after PCR (Figure 4.14). However, a higher order tandemly organized repeat is observed in the GalcoTR clones from 'Pisang Gala' with the repeat unit length ranging from 198 – 201 bp, again consistent with Figure 4.14. This sequence is designated as $CoTR_{200}$. Pairwise alignment of CoTR clone, Mal_coTR5b to *Metaviridae* Group II element, eCoTR-REs, showed 70 – 79% homology in the $CoTR_{60}$ repeats regions, whereas the 3' Un-Translated Sequence (3'UTS) is highly variable (Figure 4.15a). DNA homology search analysis of the whole sequence showed that 3 out of 12 MalcoTR clones (Mal_coTR2b, Mal_coTR4a and Mal_coTR5a) showed higher homology (85-91%) to the *Metaviridae* Group II element, eCoTR-REs compared to other clones isolated from the same *Musa* species. Multiple DNA analysis detected high homology in the 3' UTS region between these three clones to eCoTR-REs element (Figure 4.15b). Region that are conserved in the A and B genome $CoTR_{60}$ repeat units were observed in the multiple DNA sequence alignment together with several insertion/deletions (indels) (Figure 4.16).

Maximum Parsimony (MP) tree separated the *malaccensis* $CoTR_{60}$ repeats (MalcoTR) and the eCoTR-REs from BAC MA4-78I12 from those isolated from *M. balbisiana* 'Pisang Gala' (GalcoTR sequences, Figure 4.17). Two subgroups of the B genome Gal $CoTR_{60}$ repeats. Interestingly, the first, third and fifth minisatellite subrepeats from the three clones of GalcoTR were grouped together to form the subgroup I whereas the second and fourth minisatellite subrepeats were grouped together to form the subgroup II (Figure 4.18) due to 200 bp higher order repeat. Sequence annotation and comparison of $CoTR_{60}$ from *malaccensis* and $CoTR_{200}$ showed that two $CoTR_{60}$ repeats are present in $CoTR_{200}$ (Figure 4.18) with DNA-DNA sequence homology between 39 – 75%. The two $CoTR_{60}$ repeats in the $CoTR_{200}$ of 'Pisang Gala' are separated from each other by 22 nucleotides. The 5' end of first $CoTR_{60}$ repeat unit and 3' end of second $CoTR_{60}$ repeat unit in the ~ 200 bp sequence are flanked by two different sequences of 36 – 39 bp and 12 bp, respectively (Figure 4.18).

4.5 Discussion

4.5.1 Organization, diversity and evolution of the *Musa* 5S ribosomal RNA gene cluster

Amplification of the 5S rDNA sequence from *M. acuminata* 'Pisang Berangan' using the wheat 5S rDNA primer pair resulted in PCR products with a larger band size (~ 600 bp) than the cloned wheat 5S rDNA sequence (Figure 4.1). Sequence analysis detected retrotransposon-related sequences in all three 5S rDNA clones isolated from *M. acuminata* 'Pisang Berangan' (AAA) which expanded the size of the 5S rDNA subunit from ~ 400 bp to ~ 600 bp (Figure 4.3). Valárik *et al.* (2002) isolated several different types of repetitive DNA from *Musa* species. One of these clones, Radka2, isolated from *M. balbisiana* 'Cameroun' (BB) showed homology to

the 5S rRNA genes from various organisms. The Radka2 sequence was then compared with the 5S rDNA sequences isolated from 'Pisang Berangan', a *Pseudoviridae* element RT gene (PVRT) homologous region was observed in the latter sequences (Figure 4.3).

Insertion of retrotransposons into genes and ribosomal RNA genes as observed here in *Musa* has also been report for different organisms. Kapitonov and Jurka (2003) reported a novel class of short interspersed repetitive elements (SINEs) called SINE3, which were derived from zebrafish 5S rRNA genes. They proposed that both SINEs and 5S rRNA genes use the Type I RNA polymerase III (pol III) promoter for their transcription due to the three highly conserved functional sites which constitute the pol III internal promoter of the 5S rRNA genes. SINEs contained internal promoters for pol III (Singer, 1982; Okada and Ohshima, 1995; Schmid, 1998). The unusual feature of the pol III promoters is that their control regions are located downstream of the transcription start site (Paule and White, 2000). As a result, a retrotransposed cDNA copy of a SINE transcribed by pol III preserves the internal promoter. The 3'-tail of SINE3 closely resembles the CR1-like non-LTR retrotransposon, whereas the 5' end of SINE3 is ~ 75% homologous to the 5S rRNA gene. The finding of different 5S rDNA sequences with *Pseudoviridae* element insertions in *M. acuminata* ssp. *malaccensis* and 'Calcutta 4', and their organization (Figure 4.2, 4.3 and 4.4), together with the PCR screening results (Figure 4.5), indicate that these sequences might use a similar mechanism to that described by Kapitonov and Jurka (2003) for their amplification in *Musa*, at least in *M. acuminata*.

Kojima and Fujiwara (2003) reported an insertion of a non-LTR retrotransposon family, Mutsu elements, in 20 nucleotides downstream of the 5' terminus of the 5S rRNA gene which destroyed the 5S rRNA genes in some teleost

species. Burke *et al.* (2003) identified a small (100 bp) region in the 28S rRNA gene which serves as the target site for non-LTR retrotransposon insertion in both arthropods and nematodes. The R1 and R2 non-LTR retrotransposons in arthropods have been extensively studied since the early characterization of the rRNA genes of various *Drosophila* species (Long and Dawid, 1980). R1 and R2 elements are widely distributed in all lineages of arthropods where they frequently insert into 10 – 30% of the rDNA units and have completely adapted to life in the rDNA locus (Eickbush, 2002). The insertion of R1 and R2 elements into 28S rRNA genes has given rise to the 28S rRNA gene that can no longer make functional rRNA (Eickbush and Eickbush, 2007). Eickbush and Eickbush (2007) concluded that the R1 and R2 elements are highly stable in the rDNA locus even though the rDNA locus is generally highly adapted to rid itself of variation, as they did not find any evidence for the horizontal transfer that is responsible for the broad distribution of R1 and R2 elements in arthropod lineages. This led to the explanation that the R1 and R2 elements are highly successful parasites in the rDNA locus and that these elements might be involved in the regulation of rRNA synthesis or provide mechanisms to initiate recombination.

The location of the retrotransposon sequence immediately downstream of the 5S rRNA coding region in antisense orientation (Figure 4.3), provides some indication that the retrotransposon may be involved in the regulation of 5S rRNA gene expression. Presence of the expected PCR band (~ 250 bp) in *M. acuminata* (AA) and *M. balbisiana* (BB) subspecies (Figure 4.5) using primers targeting the retrotransposon insertion region, suggests that this retrotransposon insertion event probably occurred before the divergence of the *Musa* A and B genomes. No expected PCR band (~ 250 bp) in two *M. balbisiana*, 'Pisang Gala' and 'Butohan 2' (Figure

4.5) provides evidence that most of the 5S rRNA genes in these varieties contain either no retrotransposon fragment insertion or none that can be amplified using the PCR strategy used here. The presence of the expected band (~ 250 bp) with low intensities indicates that the retrotransposon inserted region is only present in some 5S rRNA gene copies in *M. balbisiana* except for 'Lal Velchi' (Figure 4.5). Generally, A genome species have higher copy numbers of units with retrotransposon insertion, while no higher band in *M. acuminata* ssp. *zebrina* indicates that almost all 5S rRNA genes have the retrotransposon fragment inserted.

Duplication of the 34 bp short direct repeats and 5 bp target site duplication (TSD) (CCTCG/CCTCT) in all three *Musa acuminata* 'Pisang Berangan' (AAA) 5S rDNA clones suggests that this retrotransposon region was generated by a retrotransposon insertion event (Figure 4.2). Furthermore, the observation of partial *Pseudoviridae* element RT gene in pMa794.1 and MA4-8B23 sequences (Figure 4.4), with a 45 bp deletion and loss of the 5 bp TSD in the latter sequence suggests that unequal exchange between sister chromatids or homologous chromosomes may have played a role in the deletion of major components of the *Pseudoviridae* element sequence from these two sequences. The higher *Pseudoviridae* element inserted bands (~ 250 bp, Figure 4.5) in *Musa* A genome species and the gradual decrease of band intensity at ~ 460 bp (resulted from amplification of two 5S rRNA units with no retrotransposon insertion), provides some evidence for partial homogenization and amplification of the 5S rRNA genes that contained the *Pseudoviridae* element inserted sequences (Figure 4.5).

Finally an evolutionary hypothesis is proposed, based on the discussion point described in this chapter (Figure 4.19). In the hypothesis, a *Pseudoviridae* element was inserted into the ancestor genome of the *Musa* species in an antisense orientation

(Figure 4.19). Unequal exchange of the 5S rRNA genes between sister chromatids or homologous chromosomes gave rise to the incomplete deletion of the inserted sequence and resulting in the new 5S rRNA variants. The partial homogenization and amplification of the new 5S rRNA variants led to the presence of both the original and new 5S rRNA genes in the A and B genomes (Figure 4.19).

4.5.2 Diversity, organization and evolution of MuTR repeats in *Musa* genomes

Twenty five tandemly organized repetitive DNA clones were isolated from three *Musa* species: *M. acuminata* ssp. *malaccensis* (Mal_TR), *M. balbisiana* 'Pisang Klutuk Wulung' (PKW_TR) and *M. ornata* (MoTR), and sequence comparison between these three species showed high inter-species homology (86-93%) and placed all the MuTR clones into one big tandem repeat DNA family, called the MuTR family. This family consist of two subrepeats (MuTRa and MuTRb) with only 46.2% DNA sequence homology between them (Figure 4.7). DNA database search analysis resulted in the identification of a 63 bp homologous region (> 90%) in the LTR of the *Metaviridae* Group III element (genus *Chromovirus*), in all the MuTR repeats. This suggests that a *Metaviridae* Group III element insertion occurred during the evolution of the MuTR repeats. A close relationship between LTR retrotransposons and tandem repeat DNAs has been demonstrated in several studies. Some of the studies showed that LTR retrotransposons are inserted into tandemly repeated DNA (Hisatomi *et al.*, 1997; Kumekawa *et al.*, 2001) and the 5S rDNA repeat in *Musa* found here (Figure 4.3), whereas in other studies, some of the tandem repeat DNA shares high sequence homology to part of the retrotransposon and it has been postulated to be derived from retrotransposons (Cheng and Murata, 2003; Yang *et al.*, 2005; Tek *et al.*, 2005, Li and Leung, 2006).

The restriction digestion site of the *Dde*I enzyme found in the 63 bp box (Figure 4.8) allowed the identification of a GC rich 220 bp repeat unit in the MuTR clones isolated from *malaccensis* and PKW. For most of the MoTR clones isolated from *M. ornata*, new *Dde*I sites were observed at the end of each MoTRa subrepeat (Figure 4.7), which generated the 110 bp subrepeat units, MoTRa and MoTRb (Figure 4.7). This suggests that a mutation occurred at the end of the MoTRa subrepeats introducing the new *Dde*I site. The same *Dde*I site was also found in the all the *Metaviridae* Group III element LTR clones isolated from three *Musa* species. This provides further indication for the acquisition of new *Dde*I site in MuTR repeats from *M. ornata* at the end of the MoTRa subrepeats and generated two 110 repeat units (See Chapter 3 for *Metaviridae* Group III element LTR clones).

The presence of other enzyme restriction sites (*Ava*I, *Bst*NI, *Eco*RII, *Hpa*II, *Msp*I, *Scr*FI and *Sma*I), which generated 220 bp repeats in all the MuTR repeats, were only observed at the 63 bp box of the MuTR repeats, together with the presence of the same restriction sites in the LTR clones from three *Musa* species provided some evidence that these restriction sites might be introduced into MuTR repeats by the insertion of *Metaviridae* Group III elements (Figure 4.7 & 4.8). The observation of the *Fok*I restriction enzyme sites at the 3' end of the MuTRa subrepeats (Figure 4.7 & 4.8) which generated 220 bp repeats might provide some indication that this site is present in the MuTR repeats before the insertion of the *Metaviridae* Group III elements. All the evidence discussed above suggests that the MuTR repeats with 220 bp monomeric sequences are the older MuTR repeats compared to 110 bp repeats (MuTRa and MuTRb). The presence of MoTR repeats with 220 bp sequences indicated that the 220 bp repeats in *M. ornata* is partially homogenized by the new 110 bp repeats.

The higher sequence variability observed in the MuTRa subrepeats (Figure 4.9a) compared to the MuTRb subrepeats with only one deletion (indel) (Figure 4.9b) suggests a faster homogenization rate occurs in the MuTRb subrepeats compared to the MuTRa subrepeats, and this led to the different MP tree topology shown in Figures 4.10 and 4.11. Fixed nucleotide substitutions observed in Figure 4.9 divided the MuTR repeats into three groups (Figure 4.10 & 4.11) and higher intra-species MuTR sequence homology (87-100%) is observed in MuTR repeat family compared to inter-species MuTR sequence homology (86-93%).

The homogenization and amplification of tandem repeat DNA in one species allow the generation of genome or species-specific repeats and variants (Flavell, 1982; Rayburn and Gill, 1987; Anamthawat-Jónsson and Heslop-Harrison, 1993; Vershinin *et al.*, 1994; Nagaki *et al.*, 1995, 1998; Kishii *et al.*, 1999; Li *et al.*, 2000; Kishii and Tsujimoto, 2002; Heslop-Harrison, 2003; Heslop-Harrison *et al.*, 2003; Rudd *et al.*, 2006) and to some extent, chromosome-specific tandem repeat DNA (Warburton *et al.*, 1996; Warburton and Willard, 1996). On the other hand, Contento *et al.* (2005) showed that the 120 bp repeat in *Triticeae* has maintained its variation and the diversity seen today among species was postulated to be the same as in the ancestral genome. Another type of the tandem repeat DNA evolution model has been proposed by Salser *et al.* (1976), where the ancestor genome of the related species shared a "library" of conserved tandem repeat sequences. The amplification of these tandem repeat DNA resulted in a major tandem repeat DNA of the particular species. However, this hypothesis still remains unconfirmed due to the extreme diversity and complexity of the tandem repeat DNA profiles. The high sequence homologies from one species due to the homogenization and amplification mechanism made it difficult to detect genome-, species- or chromosome-specific tandem repeat DNA by just

looking at the sequence variability itself (Meštrović *et al.*, 1998). In my study, phylogenetic analysis shown in Figure 4.10 and 4.11 allow the identification of species-specific MuTR repeats suggesting a concerted mode of evolution occurred in the MuTR repeat family.

Amplification of MuTR repeats with three different primer pairs (Table 4.1) using 19 *Musa* species which belong to four different *Musa* sections (*Eumusa*, *Rhodochlamys*, *Callimusa* and *Australimusa*) indicated that MuTR repeats are present in all the *Musa* taxa tested (Figure 4.12). The high molecular weight smearing observed in Figure 4.12 suggesting that MuTR repeats are present in high copy number in the *Musa* species tested. The loss of the tandem array pattern in some of the *Callimusa* species tested particularly, *M. violascens* and *M. gracillis*, might be due to the high sequence variability at the MuTR primer sites (Figure 4.12). The observation of a tandem array in *M. acuminata* ssp. *banksii* (AA) with ~ 280 bp monomeric size and in most of the *Musa* species tested (~ 180 bp monomeric size) (Figure 4.12a), together with the appearance of the MuTR dimer bands with slightly smaller size in Figure 4.12b (~ 210 bp) and 4.13c (~ 310 bp) suggests that this could be the result of internal deletions that occurred in the MuTR repeats. This is supported by the observation of the deleted regions found in MuTRa subrepeat alignment in Figure 4.9a. In plants, over 160 families of tandem repeat DNA have been described (Maca *et al.*, 2002) and often show inconsistent distribution and great differences in their abundance among closely related species (Schmidt and Heslop-Harrison, 1994; Vershinin *et al.*, 1994; Kubis *et al.*, 1997; Macas *et al.*, 2000; Frello *et al.*, 2004; Tek *et al.*, 2005). Many tandem repeats undergo evolution by rounds of sequence divergence and amplification (Flavell *et al.*, 1980) and subsequently repeat variants, subfamilies and families are generated and many of these repeats become

species- or chromosome-specific (Heslop-Harrison *et al.*, 1997; Friesen *et al.*, 2001). However, some tandem DNA repeats are nearly universal within the taxa tested, for example, the 2D8 tandem repeat families in potato are present in almost all species in the genus *Solanum* (Stupar *et al.*, 2002), the pBuM satellite DNA family in *Drosophila* species (Kuhn and Sene, 2005) and the 120 bp repeat in many *Triticeae* (Contento *et al.*, 2005). A similar situation is observed for MuTR tandem repeat DNA family using PCR primers designed from the conserved region of the MuTR repeat units, where the MuTR repeat family is present in all the *Musa* taxa tested (Figure 4.12).

An evolutionary hypothesis is proposed for the *Musa* MuTR repeats based on the results and discussions described in this chapter (Figure 4.20). In the hypothesis, the MuTR repeats in the ancestral *Musa* genome, with a *Fok*I restriction enzyme site located at the 3' end of each repeat unit, contained no *Metaviridae* Group III element insertion. During the evolution of the *Musa* ancestor genome, which gave rise to the current *Musa* species, a *Metaviridae* Group III element inserted into the MuTR repeats and this spanned the inserted genome region (Figure 4.20). Unequal exchange of MuTR repeats occurred between sister chromatids or homologous chromosomes, resulting in the incomplete deletion of the inserted retrotransposon sequence, leaving behind the 63 bp box with 8 different restriction enzyme sites (*Ava*I, *Bst*NI, *Dde*I, EcoRII, *Hpa*II, *Msp*I, *Scr*FI and *Sma*I), and this generated the new MuTR variants (only the *Dde*I site is shown in Figure 4.20). The homogenization and amplification of the new MuTR variants replaced the original MuTR repeats and led to the presence of MuTR repeats all containing the 63 bp box in the A and B genomes. Further mutation in the A and B genome MuTR repeats by the acquisition of a new *Dde*I site at the end of the MuTRa repeat, together with

partial homogenization and amplification processes, gave rise to the current MuTR variants in the *M. ornata* genome (MoTR) with 110 bp and 220 monomeric sizes (Figure 4.20).

4.5.3 Diversity and organization of the CoTR repeats

Thirty five copies of a ~ 60 bp minisatellite repeat, $CoTR_{60}$, were observed in the full-length *Metaviridae* Group III element, eCoTR-Res, near the 3' LTR (Figure 3.2g). The generally high G+C% and sequence conservation at the polypurine tract (PPT) of the *Metaviridae* Group III element allowed amplification of the region between the CoTR repeats and the PPT. Different banding patterns observed in Figure 4.14 indicate that different length and higher order organization of minisatellite repeats are present in the *Musa* A and B genomes.

Fifteen clones of the CoTR repeats from *M. acuminata* ssp. *malaccensis* (MalcoTR) and *M. balbisiana* 'Pisang Gala' (GalcoTR) have been isolated and DNA sequence analysis detected the same $CoTR_{60}$ repeats as found in eCoTR-REs element. Interestingly, the repeat unit of the GalcoTR clones from 'Pisang Gala' is ~ 200 bp rather than ~ 60 bp and further sequence analysis detected two $CoTR_{60}$ sequences incorporating the $CoTR_{60}$ into a higher order structure of the ~ 200 bp repeat unit (Figure 4.18). Further sequence analysis of the $CoTR_{60}$ repeat unit in GalcoTR clones detected high order organization of this repeat unit (Figure 4.18), where all the first $CoTR_{60}$ repeat units in different $CoTR_{200}$ repeat are grouped together to form the Subgroup I of the B genome CoTR repeats in the MP tree (Figure 4.17 & 4.18) whereas all the second $CoTR_{60}$ repeat units formed the Subgroup II (Figure 4.17 & 4.18). This new type of tandemly organized repeat, $CoTR_{200}$, is presented in the *M. balbisiana* 'Pisang Gala' genome only. Interestingly, no specific higher order was

detected for the 35 repeat units in the BAC (eCoTR-REs) giving each unit autonomy to mutate.

For Mal_coTR clones, two different groups of the $CoTR_{60}$ sequence were observed where the first group showed low homology in the 3' Un-Translated Sequence (3' UTS) to eCoTR-REs with several insertion/deletions (indels) (Figure 4.15a), whereas the second group showed higher homology in this region (Figure 4.15b). This suggests that there are at least three different *Metaviridae* Group III elements that carried two different types of tandemly organized repeats ($CoTR_{60}$ & ~ $CoTR_{200}$). PCR bands of similar size in all three *M. acuminata* subspecies were also observed in 'Pisang Gala' but with much lower band intensities (Figure 4.14), suggesting that $CoTR_{60}$ repeats might be present in the *M. balbisiana* 'Pisang Gala' genome at lower copy numbers. Fixed nucleotide substitution in the $CoTR_{60}$ repeats from the same species and higher sequence variability in the inter-species $CoTR_{60}$ repeats (Figure 4.16) allowed the classification of the A and B genome $CoTR_{60}$ repeats (Figure 4.17). This suggests homogenization between the $CoTR_{60}$ repeat units from the same species.

Transposable elements, both retrotransposons and DNA transposons, have been shown to contain tandemly organized repeats of various lengths in their internal domains in different organisms (Garrett *et al.*, 1989; Noma and Ohtsubo, 2000; Chao *et al.*, 2003; Vicient *et al.*, 2005) including *Musa* (see Figure 3.2). The 6,896 bp *TEOS1* element, identified in *Oryza sativa* by Chao *et al.* (2003) by repeat-mining strategy, is composed of two arrays of highly homologous tandem repeats that are separated by 3 spacer sequences. These two tandem arrays were located between the PBS and PPT of the *TEOS1* element, where the 5' array contains 10 tandemly organized copies of 90 bp elements and the 3' array contains 12 copies of the same

element (Chao *et al.*, 2003). DNA sequence analysis of these tandem repeats revealed that they were highly homologous to the tandem repeats near the telomeres of *Oryza sativa* chromosome 2 and the rice panicle EST (Chao *et al.*, 2003). In *Arabidopsis thaliana*, the DNA transposon, *Tnat1*, had tandem repeats comprised of several repeat units of a 60 bp sequence in its internal region and the number of repeats are different among *Tnat1* members whereas another DNA transposon, *Tnat2*, contained a tandem repeat of ~ 240 bp repeat unit length in its internal region (Noma and Ohtsubo, 2000). In the *Xenopus laevis* genome, two families of non-LTR retrotransposon, Tx1d and Tx2d, consisted of two types of 400 bp internal tandem repeats (PTR-1 and PTR-2) and these two tandem repeats are flanked by unique sequences, LCF and RCF (Garrett *et al.*, 1989). In this study, two different types of tandemly organized repeats, $CoTR_{60}$ and $CoTR_{200}$, are identified in *Musa* species and both are located downstream of the eCoTR-REs (*Metaviridae* Group II element) ORF and upstream of its PPT sequences.

Table 4.1 PCR primers used in Chapter 4.

Primer Name	Sequence (5' to 3')	T_m (°c)	Binding site of the primers	PCR product size	Use or the primers
MuTRFor	tgcttcgctckctgtccgc	69.3	MuTRa repeat	220 bp repeat	MuTR isolation
MuTRRev	asgagcrgcaccttccctgc	68.9	MuTRa repeat		MuTR isolation
MuTR1aR	tgagcggacgagcggcgagcggcgg	89.7	MuTRa repeat	220 bp repeat	MuTR isolation
MuTR1bR	tgagaaggctgacgttttgctcc	70.5	MuTRb repeat		MuTR isolation
MuTR1aF	gcaccctggaacccccggtggca	86.4	MuTRa repeat	110 bp MuTRa repeat with MuTR1aR primer	MuTR isolation
MTY1TRF	cccctatcatattcccccctgaaagg	74.4	Part of 3' LTR and PPT	220 bp repeat	$CoTR_{60}$ isolation
MTY1TRR01	cgycgcchvagaccrcaccyghgcggtaacc	80.1	$CoTR_{60}$ repeat unit		$CoTR_{60}$ isolation
PTA794FOR	gatccggtgctttagtgctgg	67.8	5S rRNA gene	400 bp based on wheat 5S rDNA sequence (primers are designed to give the complete unit)	5S rDNA isolation (contain ITS + gene)
PTA794REV	catcagaactccgaagttaagc	65.8	5S rRNA gene		5S rDNA isolation (contain ITS + gene)
BGyFor	gtggccgatagcactgtgcca	72.6	*gypsy*-IRAP B genomic band	275 bp	IRAP
BGyRev	agttggtgcaaggtgcttcgc	70.6	*gyspy*-IRAP B genomic band		IRAP
RIAP01F	ctcgtgttgcactccttttgc	71	PVRT in *Musa* 5S rDNA	247 bp	*Pseudoviridae* insertion study
RIAP01R	ccgacgggagaggttttgg	72	PVRT in *Musa* 5S rDNA		*Pseudoviridae* insertion study

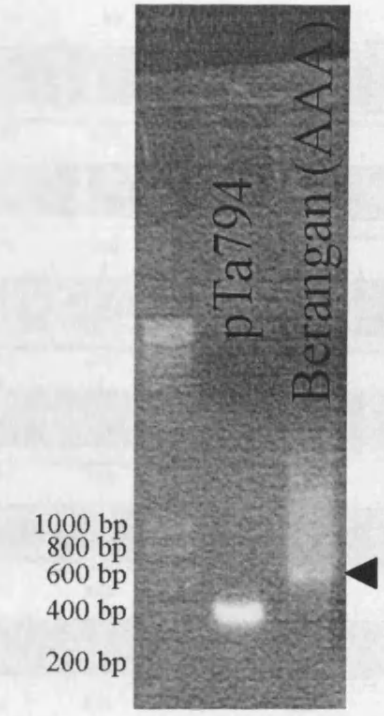


Figure 4.1: PCR amplification of the 5S rDNA sequence from *M. acuminata* 'Pisang Berangan' (AAA) using primers PTA794FOR and PTA794REV (designed from wheat 5S rDNA clone, pTa794; see Table 4.1). The pTa794 plasmid is used as a positive control to check the efficiency of the PCR amplification. 1% (w/v) agarose gel was used and the PCR band at ~ 600 bp (arrowhead) was excised from the gel, cloned and sequenced.

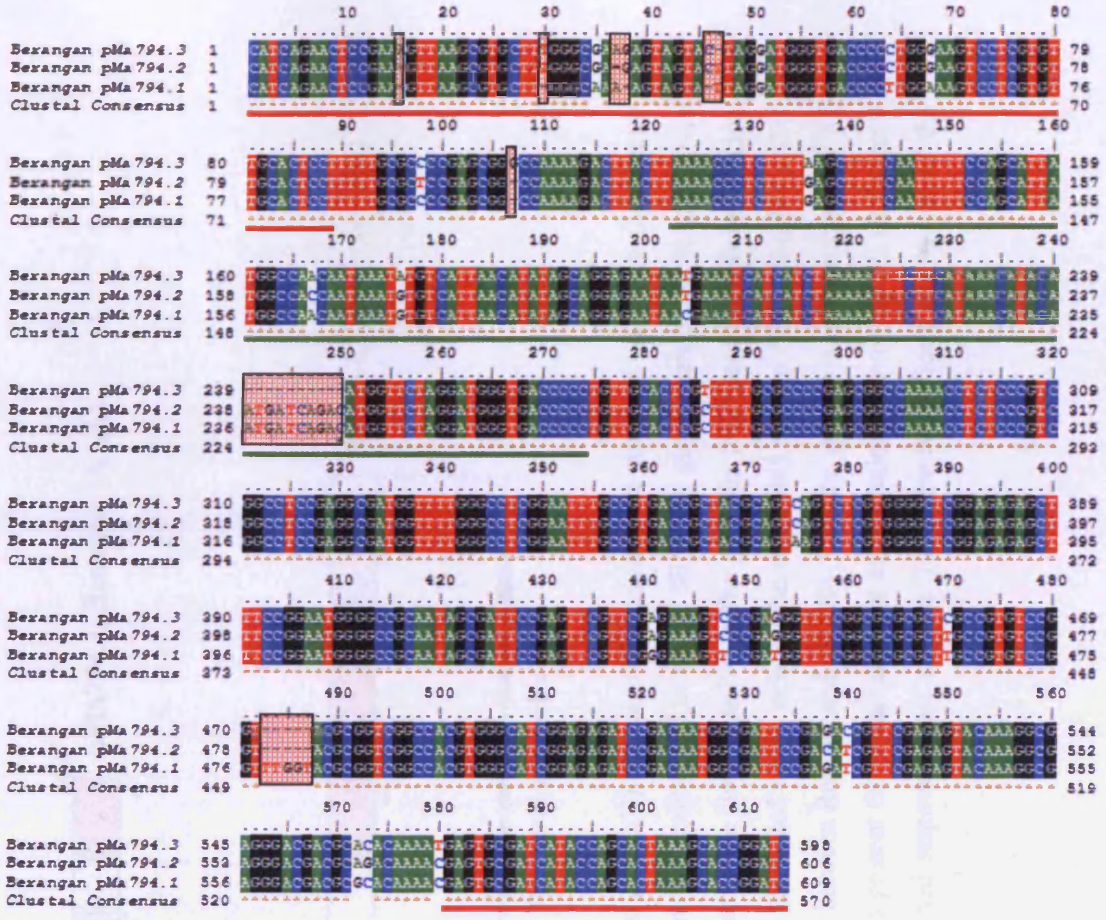


Figure 4.2: DNA sequence alignment of the 5S rDNA clones isolated from *M. acuminata* 'Pisang Berangan' (AAA). The insertion/deletions are indicated by red. The coding region of the 5S rRNA gene and the PVRT homologous region are represented by a red or green bar, respectively.

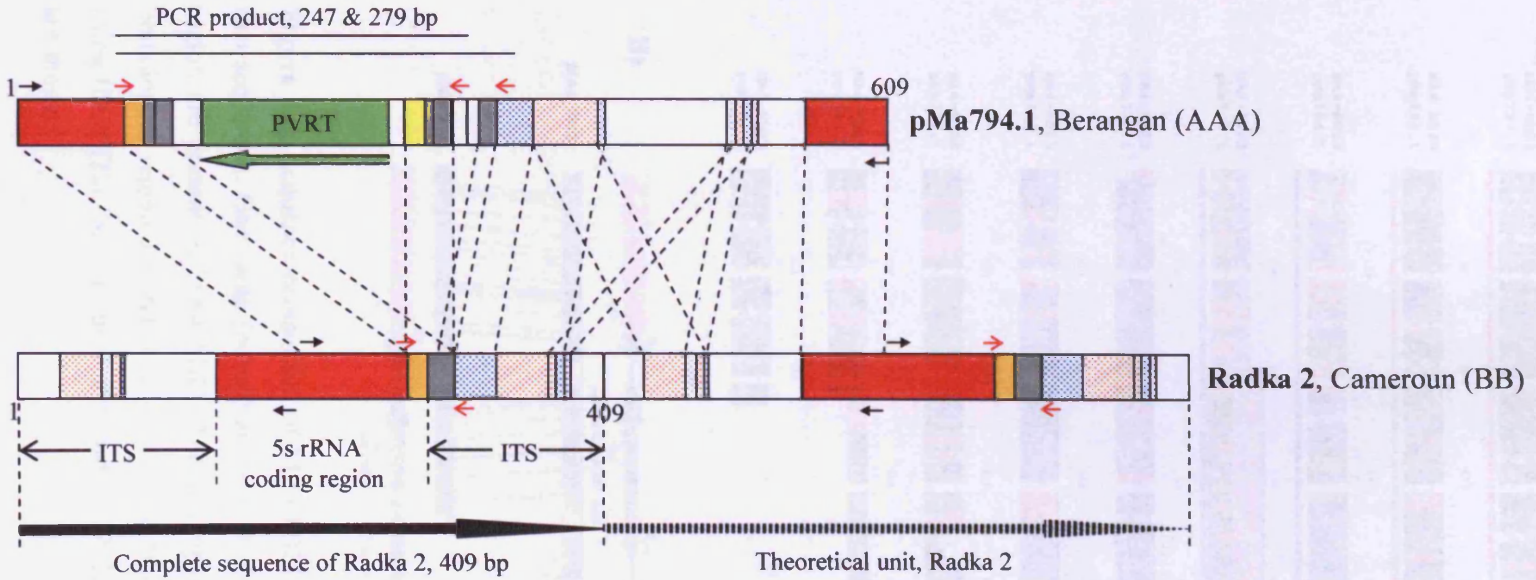


Figure 4.3: Sequence annotation of pMa794.1 and Radka 2 (Valárik *et al.*, 2002). Partial additional units are shown in tandem to allow better visualization of the structure and homology. The red boxes represent the coding region of the 5S rRNA gene whereas the yellow boxes represent the 3' end of the 5S rRNA coding region which is duplicated in the clone pMa794.1. The gray box indicates the duplicated region (Dup) of the Inter-Genic Spacer (ITS) of Radka 2 in pMa794.1 and the dotted boxes indicate the duplicated regions of the pMa794.1 ITS region in Radka 2. The white boxes represent non-homologous regions between Radka 2 and pMa794.1 and the green box represents the PVRT sequence inserted into pMa794.1, in antisense orientation. The arrows present the primer sites for amplification of complete 5S rRNA gene (black arrows) and for amplification of *Pseudoviridae* element inserted sequences (red arrows). Large arrows indicate orientation of the sequences.

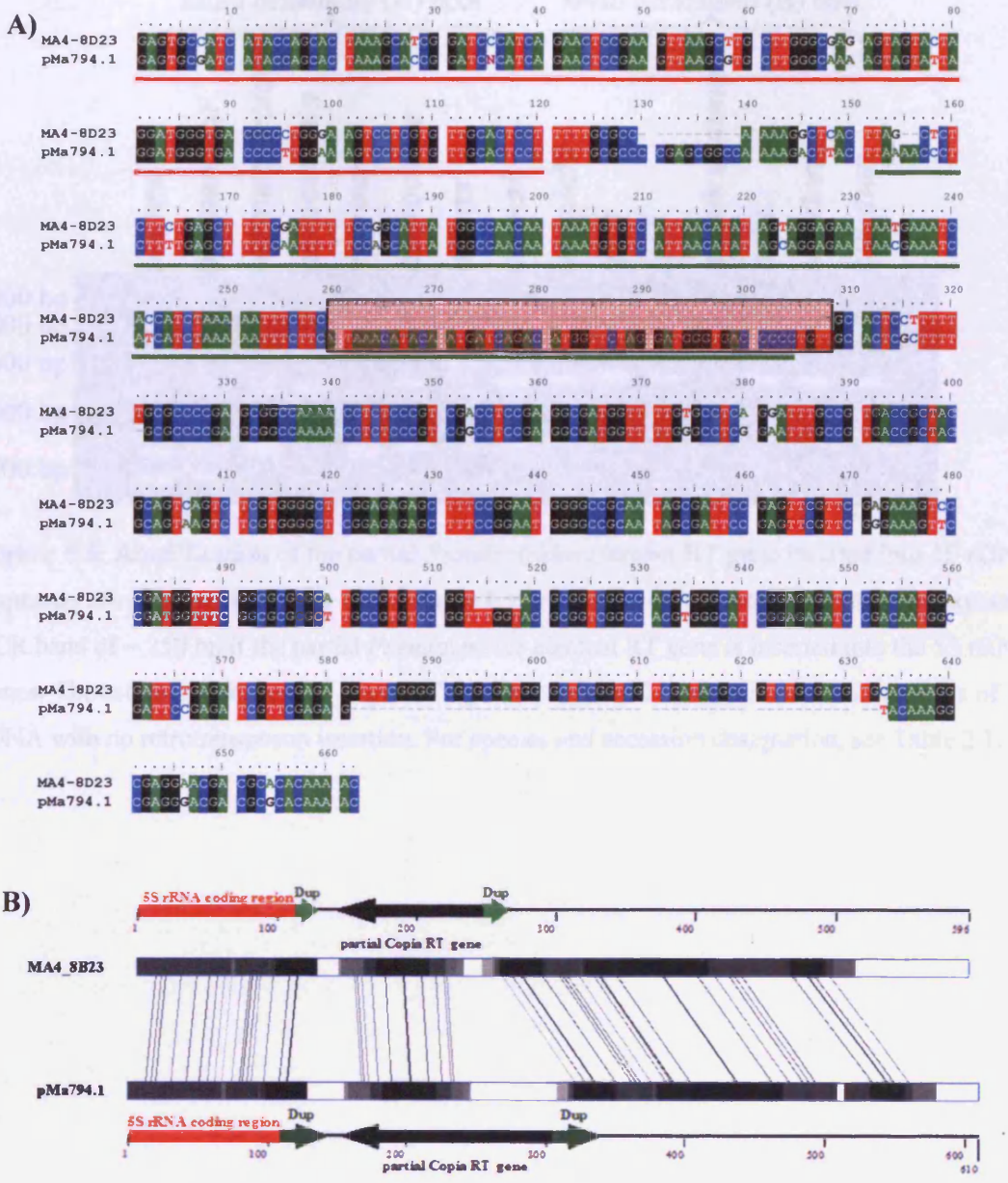


Figure 4.4: Sequence comparison of MA4-8B23 and pMa794.1. **A** Pairwise alignment of the two sequences. One large (boxed) and several small deletions within the PVRT homologous region are present in MA4_8B23. The coding region of the 5S rRNA gene and the PVRT homologous region are underlined in red and green, respectively. **B** Graphical dotplot analysis using JDOTTER (Sonnhammer and Durbin, 1995). The 5S rRNA coding region and Dup region are shown.

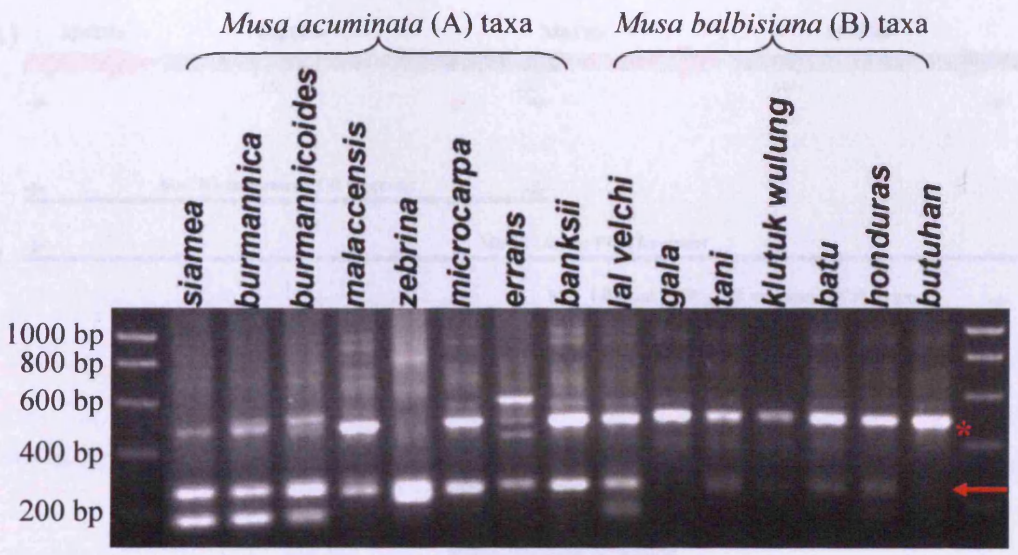


Figure 4.5: Amplification of the partial *Pseudoviridae* element RT gene inserted into 5S rDNA sequence using internal primers (See Figure 4.3 and Table 4.1). Arrow indicates the expected PCR band of ~ 250 bp if the partial *Pseudoviridae* element RT gene is inserted into the 5S rRNA genes. The asterisk indicates the larger PCR product of ~ 460 bp amplified from two units of 5S rDNA with no retrotransposon insertion. For species and accession designation, see Table 2.1.

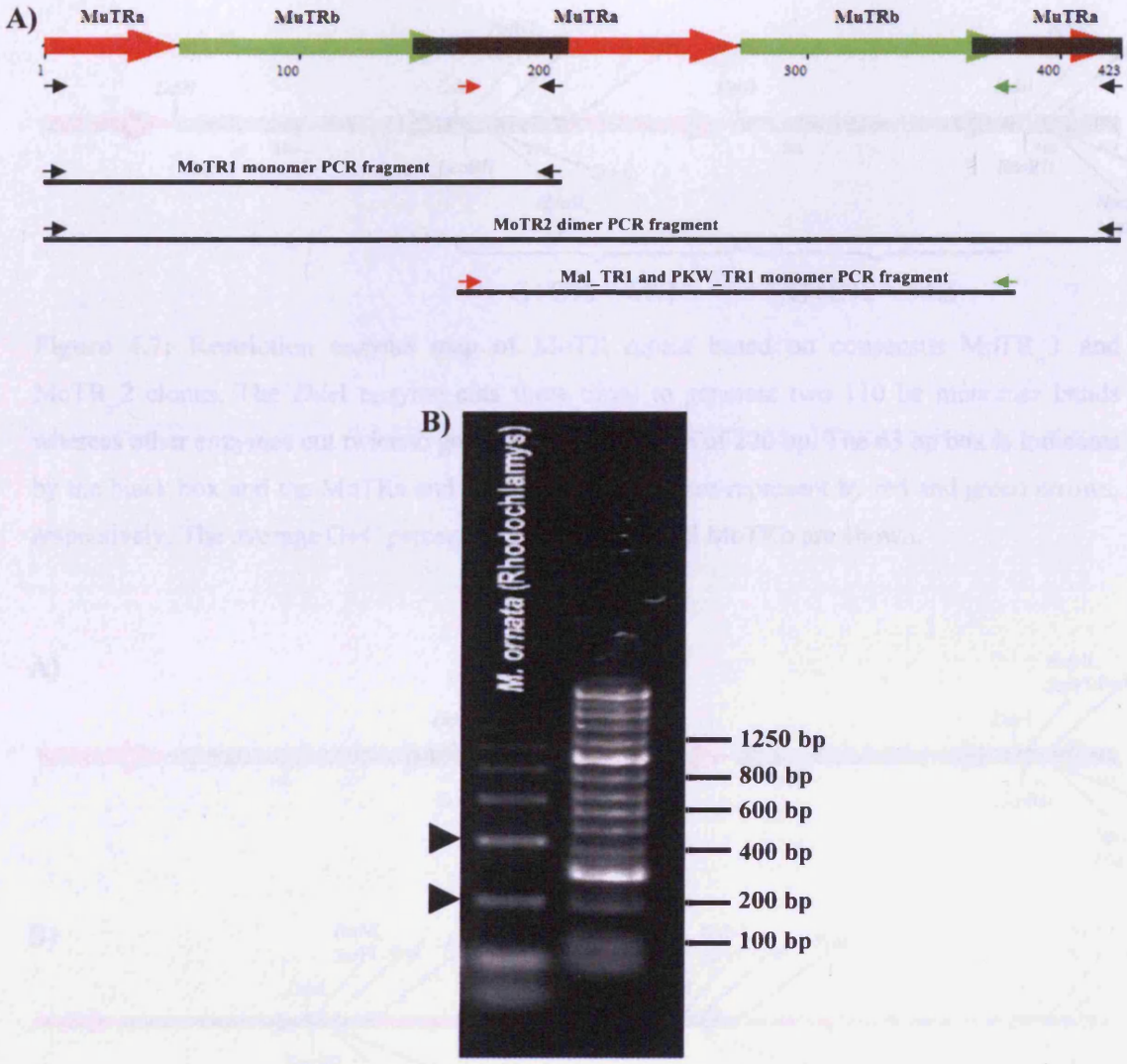


Figure 4.6: PCR scheme of MuTR repeat isolation. **A** PCR primer design strategy to amplify MuTR repeats from three *Musa* species. For MoTR_1 and MoTR_2 clones isolated from *M. ornata*, IRAP internal primers BGyFor and BGyRev (black arrows) were used. For Mal_TR (*M. acuminata* ssp. *Malaccensis*) and PKW_TR (*M. balbisiana* 'Pisang Klutuk Wulung') clones, MuTR specific primers MuTR1aF and MuTR1bR (red and green arrows) were used. Only monomer PCR fragment for Mal_TR and PKW_TR PCR fragment are shown in this figure. The 63 bp box is indicates by the black box and the MoTRa and MoTRb subrepeats are represented by large red and green arrows, respectively. **B** PCR amplification for MuTRs repeats from *M. ornata* using IRAP internal primers, BGyFor and BGyRev. For primer sequences and PCR conditions, see Table 4.1. PCR artifact and primer-dimer bands below 100 bp were observed. The monomer and dimer of MoTR repeats were excised from the gel and cloned (arrowheads).

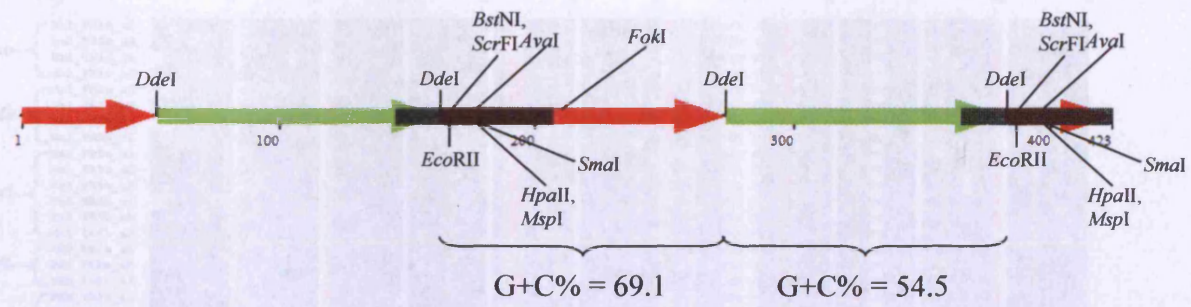


Figure 4.7: Restriction enzyme map of MoTR repeat based on consensus MoTR_1 and MoTR_2 clones. The *Dde*I enzyme cuts three times to generate two 110 bp monomer bands whereas other enzymes cut twice to generate monomer size of 220 bp. The 63 bp box is indicates by the black box and the MoTRa and MoTRb subrepeats are represent by red and green arrows, respectively. The average G+C percentages for MoTRa and MoTRb are shown.

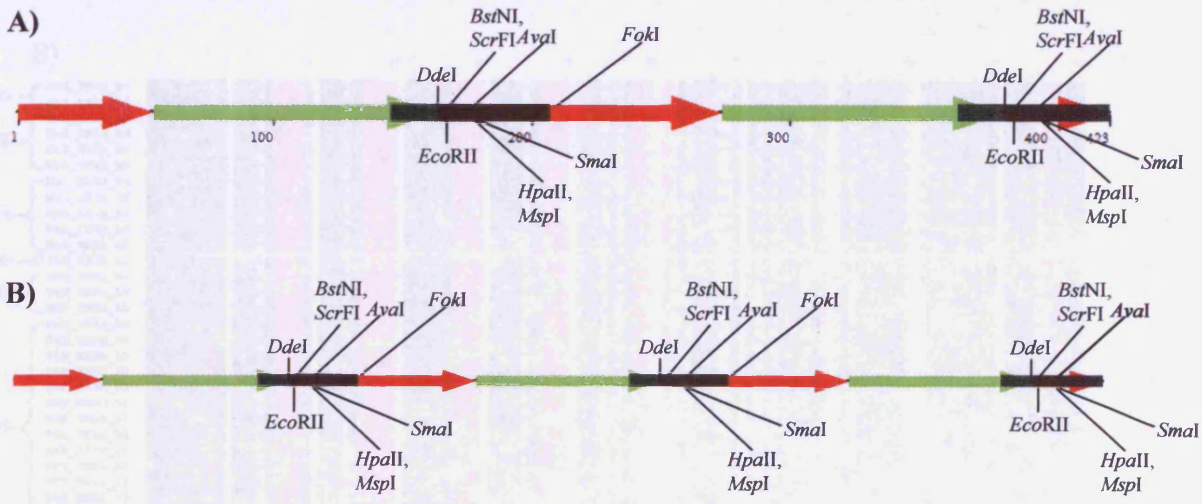


Figure 4.8: Restriction enzyme map of Mal_TR and PKW_TR repeat based on their consensus sequences. **A** For monomer and dimer clones, all enzymes cut twice to generate monomer size of 220 bp. **B** PKW-TR3b and PKW-TR3f clones' restriction enzyme map. The *Eco*RII generated the restriction band of 440 bp in length. The 63 bp box is indicates by the black box and the MuTRa and MuTRb subrepeats are represented by red and green arrows, respectively.

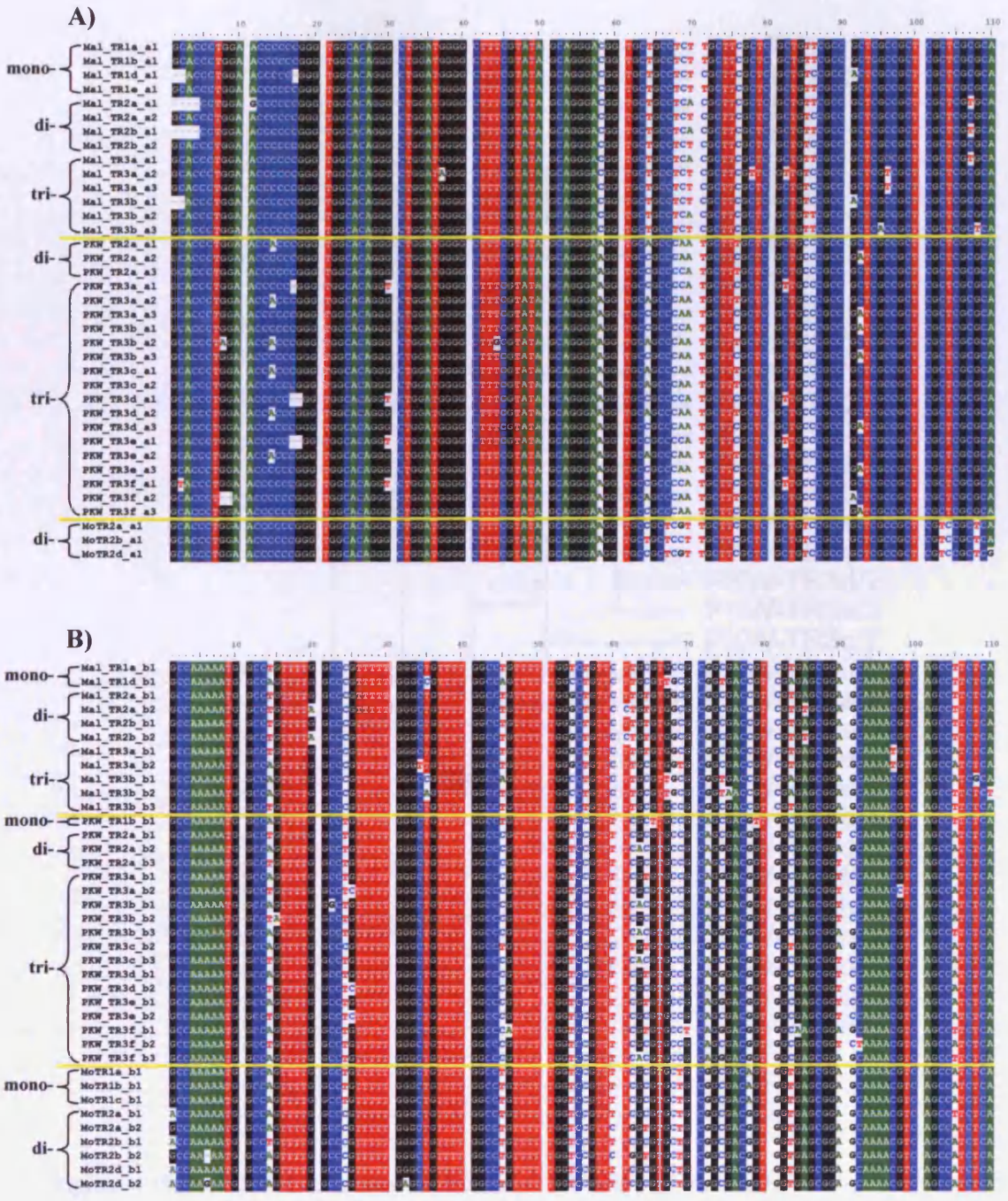


Figure 4.9: Multiple DNA sequence alignment of the MuTR monomer subrepeats: **A** MuTRa and **B** MuTRb. The conserved nucleotides of 80% DNA homology are shaded. Fixed nucleotide substitutions are indicated by the un-shaded nucleotides. Subrepeats isolated as mono-, di- or trimers are grouped together.

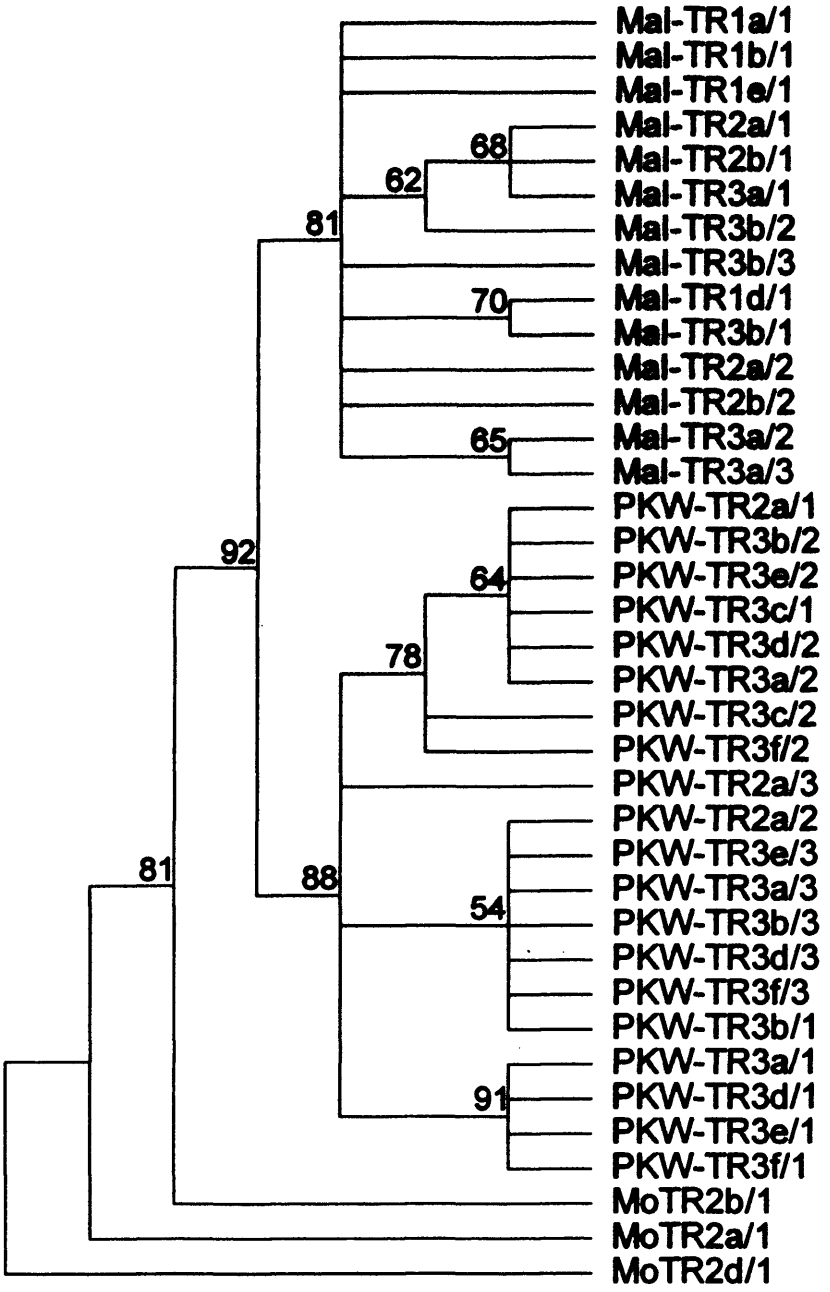


Figure 4.10: Un-root strict consensus tree of 11570 MP trees (tree length = 36) of 37 MuTRa sequences isolated from three *Musa* species: *M. acuminata* ssp. *malaccensis* (Mal_TR, A genome), *M. balbisiana* 'Pisang Klutuk Wulung' (PKW_TR, B genome) and *M. ornata* (MoTR). Bootstrap values of 1000 replicates are shown. Sequences from each species form a separate group. All three MoTRa sequences formed the third group of the MP tree.

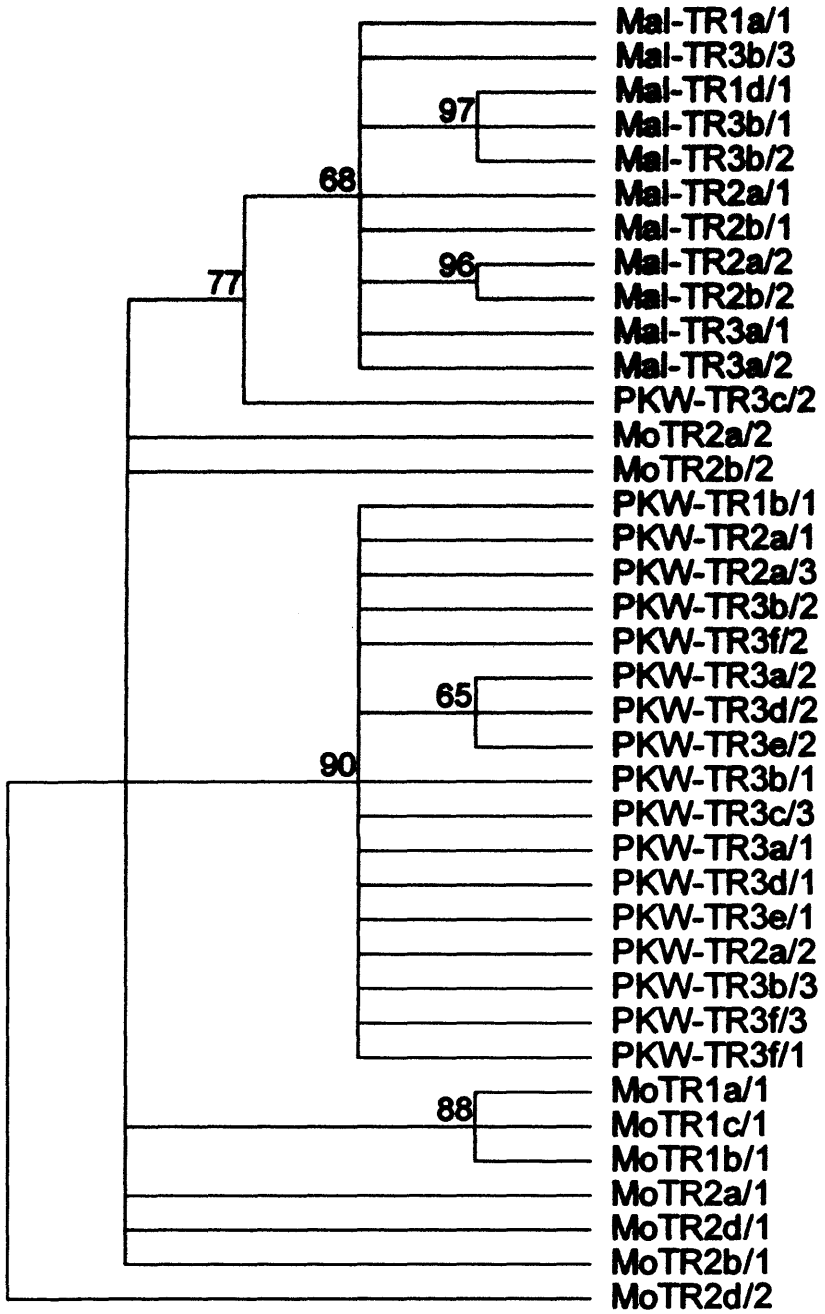


Figure 4.11: Un-root strict consensus tree of 1410 MP trees (tree length = 68) of 38 MuTRb sequences isolated from three *Musa* species: *M. acuminata* ssp. *malaccensis* (Mal_TR, A genome), *M. balbisiana* 'Pisang Klutuk Wulung' (PKW_TR, B genome) and *M. ornata* (MoTR). Bootstrap values of 1000 replicates are shown. Sequences from each species form a separate group.

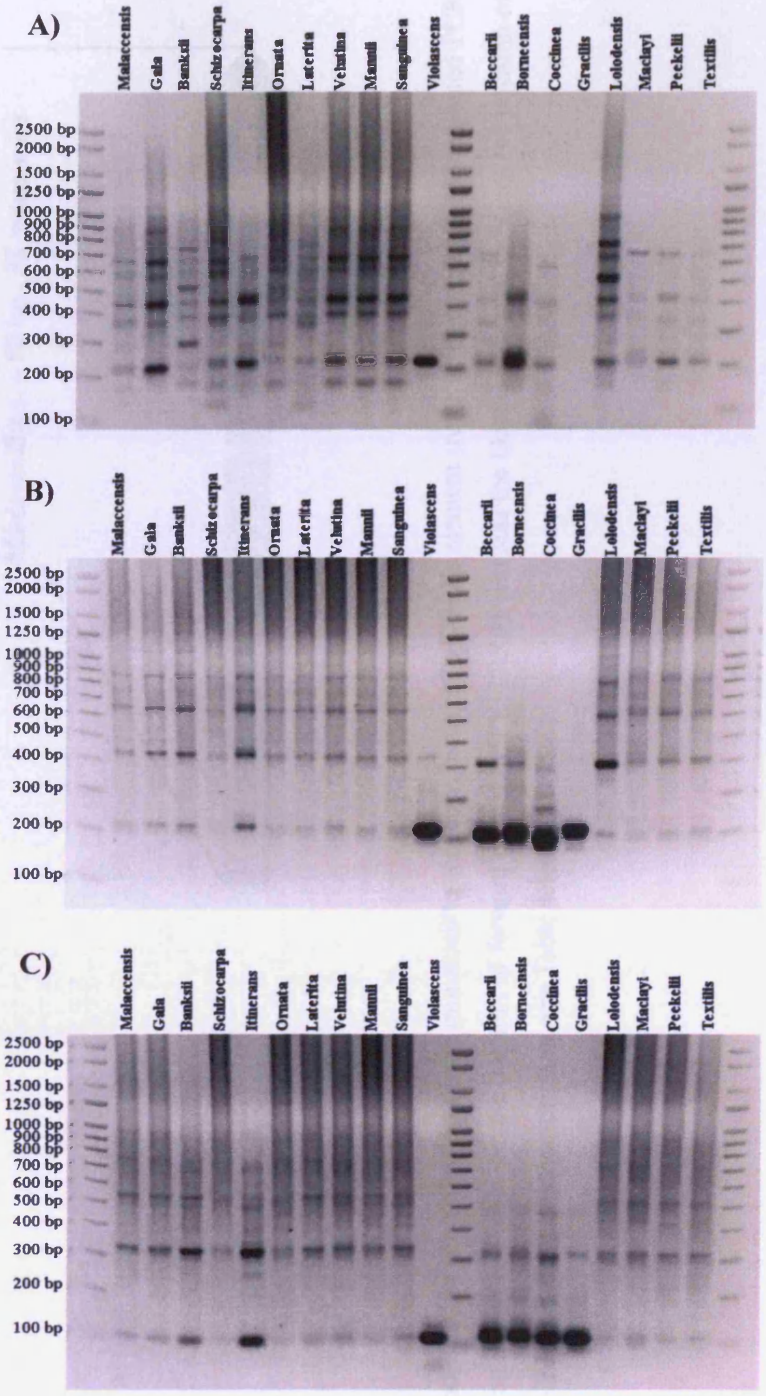


Figure 4.12: Genome wide amplification of MuTR repeats using three primer pairs. A) MuTRF + MuTRR, B) MuTR1aF + MuTR1bR C) MuTR1aF + MuTR1aR. For primer sequences, see table 4.1 and Figure 4.7a and *Musa* species description, see Table 2.1

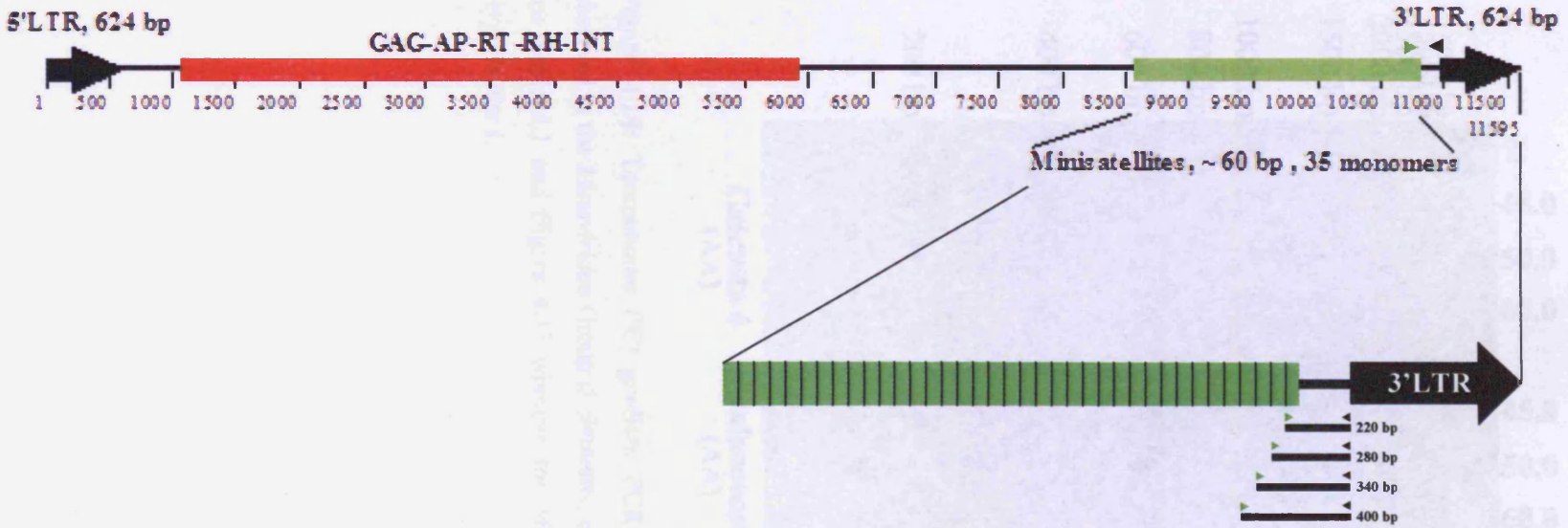


Figure 4.13: Primer position for amplification of the minisatellite sequence in the eCoTR-REs element (refer to Figure 3.1b) and expected PCR products size. The green arrowhead indicates the position of forward primer, MTY1TRR01 whereas the black arrowhead indicates the position of the reverse primer, MTY1TRF. For primer description, see Table 4.1.

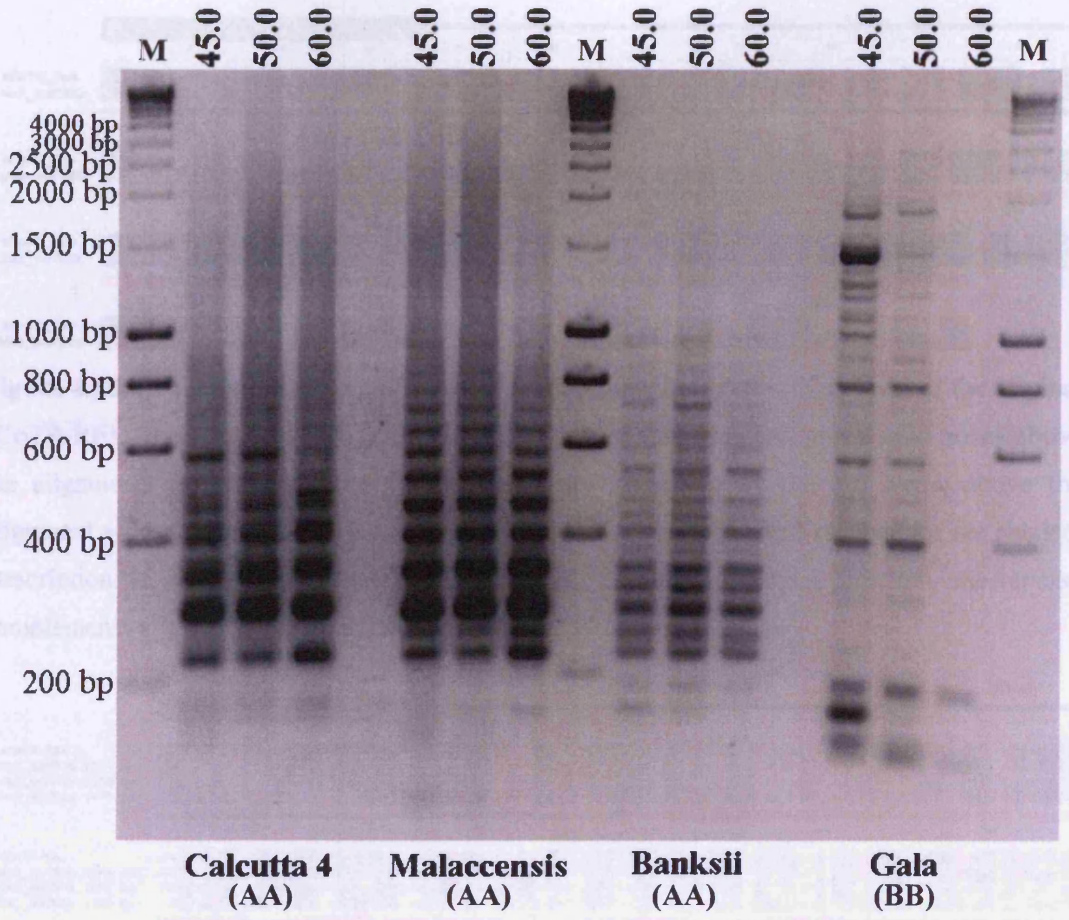


Figure 4.14: Temperature (°C) gradient PCR amplification of the minisatellite sequence observed in the *Metaviridae* Group II elements, eCoTR-REs. For primer sequence and position, see Table 4.1 and Figure 4.13 whereas for *Musa* species description, see Table 2.1. M = Hyperladder I.

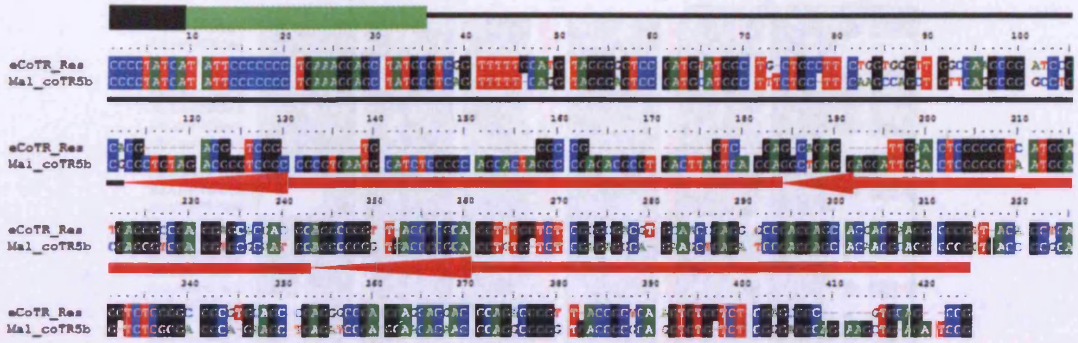


Figure 4.15a: Pairwise alignment of selected CoTR clone, Mal_coTR5b and part of the original eCoTR-REs. The LTR and PPT region are indicating by the black and green solid boxes above the alignment, respectively. The $CoTR_{60}$ units are indicating by the red arrow above the alignment whereas the 3' UTS is indicates by the lines. The conserved nucleotides are shaded. Description of sequence is showed in Figure 4.18. Due to sequencing direction, the reverse complements to Figure 4.13 and 4.18 are shown.

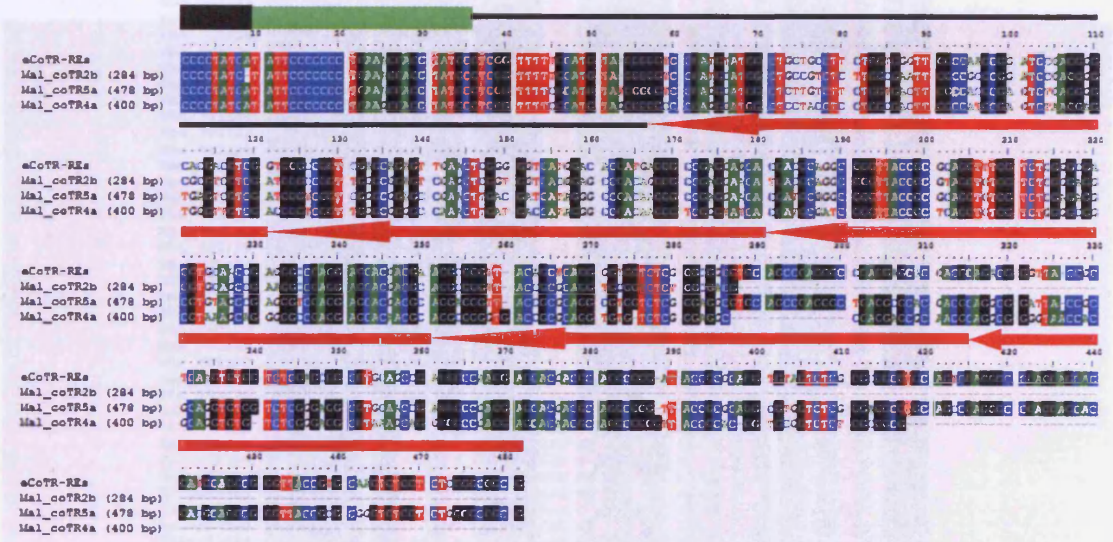


Figure 4.15b: DNA alignment of selected Mal_CoTR clones (Mal_coTR2b, Mal_coTR4a and Mal_coTR5a) and part of the original eCoTR-REs. The LTR and PPT region are indicating by the black and green solid boxes above the alignment. The $CoTR_{60}$ units and 3' UTS are indicated by the red arrow and line above the alignment. Conserved nucleotides are shaded. The high homology in the 3'UTS sequence makes these clones more homologous to eCoTR-REs. Description of sequence is showed in Figure 4.18. Due to sequencing direction, the reverse complements to Figure 4.13 and 4.18 are shown.

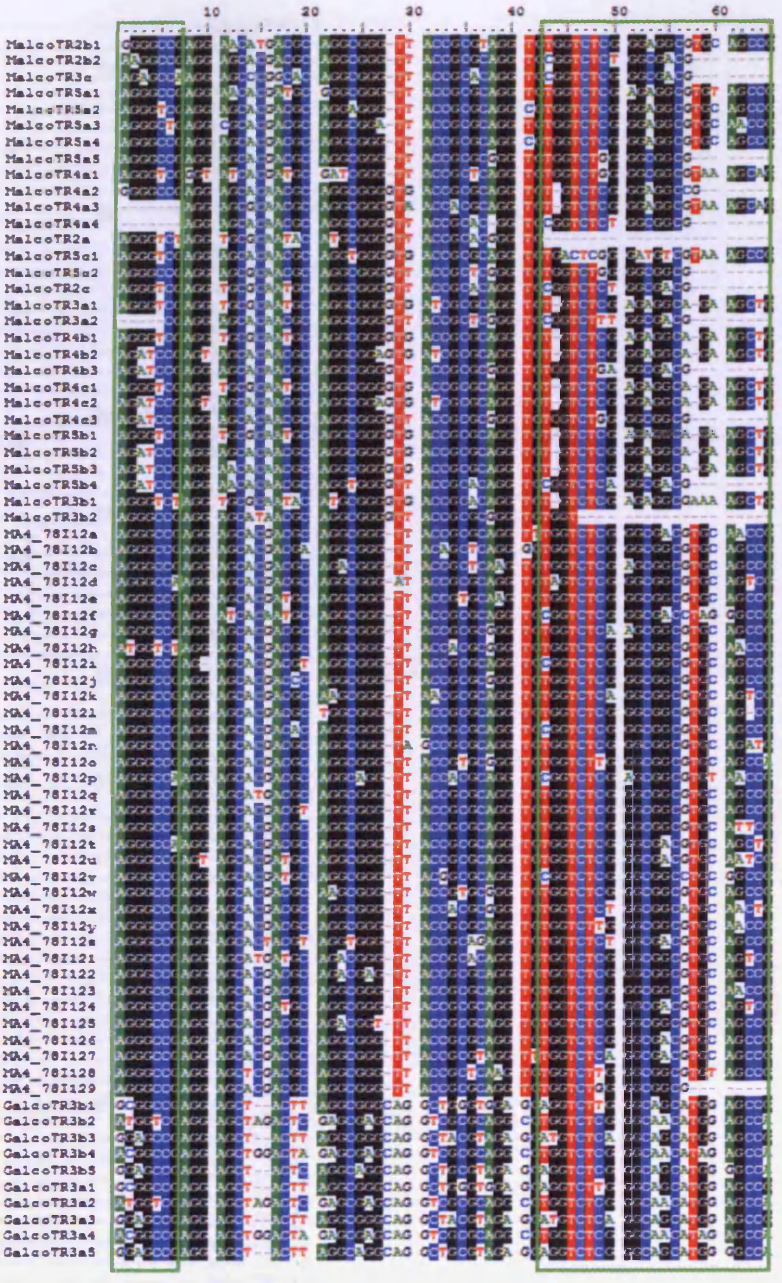


Figure 4.16: Multiple DNA sequence alignment of *Musa* $CoTR_{60}$ repeats isolated from *M. acuminata* ssp. *Malaccensis* (MalcoTR) and *M. balbisiana* 'Pisang Gala' (GalcoTR), together with the 35 $CoTR_{60}$ repeats retrieved from eCoTR-REs (within BAC MA4_78I12). Conserved nucleotides are shaded and regions with insertions/deletions are indicated by green boxes. Description of sequence is showed in Figure 4.18.

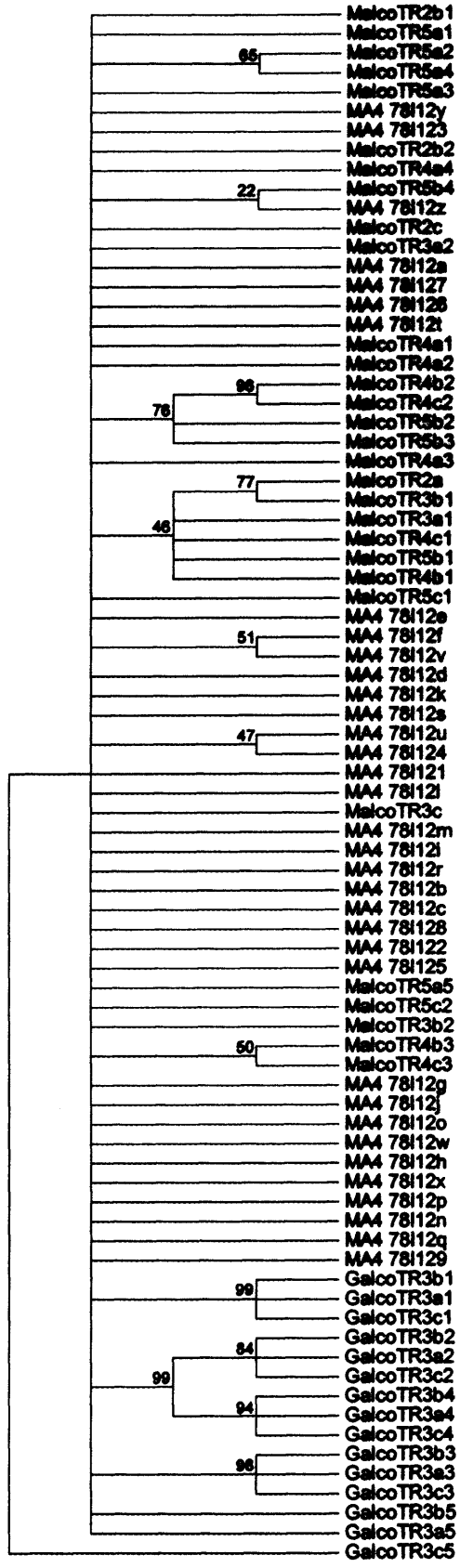


Figure 4.17: Un-root strict consensus tree of 486 MP trees (tree length = 233) of *Musa* CoTR repeats. Description of sequence is showed in Figure 4.18.

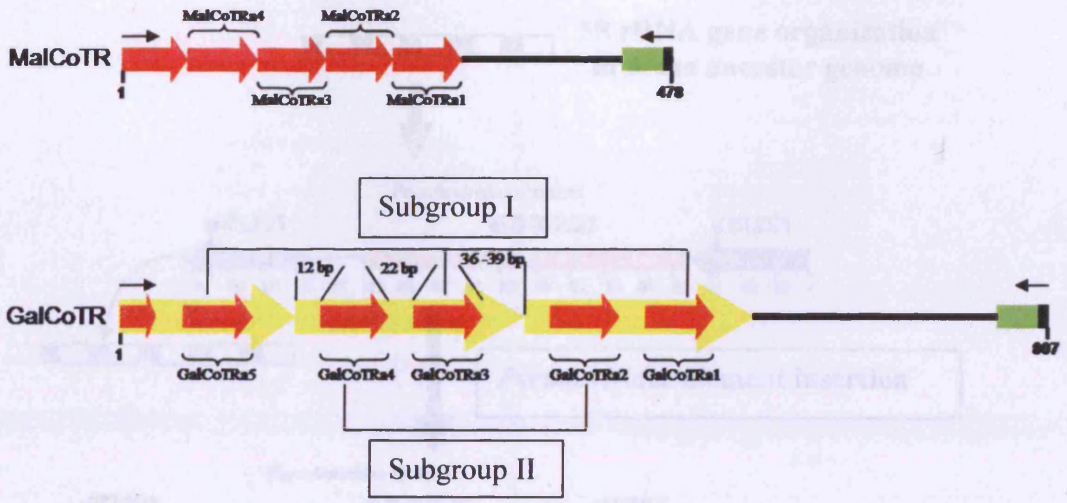


Figure 4.18: Sequence annotation of MalCoTR and GalCoTR clones and organization of the GalcoTR repeats based one the MP tree grouping parameters in Figure 4.17. The red arrows indicate the ~ 60 bp minisatellite repeat $CoTR_{60}$ found in *Metaviridae* Group II elements whereas the yellow arrows represent ~ 200 bp tandem repeat $CoTR_{200}$ which was observed in *M. balbisiana* 'Pisang Gala', with two $CoTR_{60}$ minisatellite sequences in each $CoTR_{200}$ repeat. $CoTR_{60}$ repeat units and primer sites are shown.

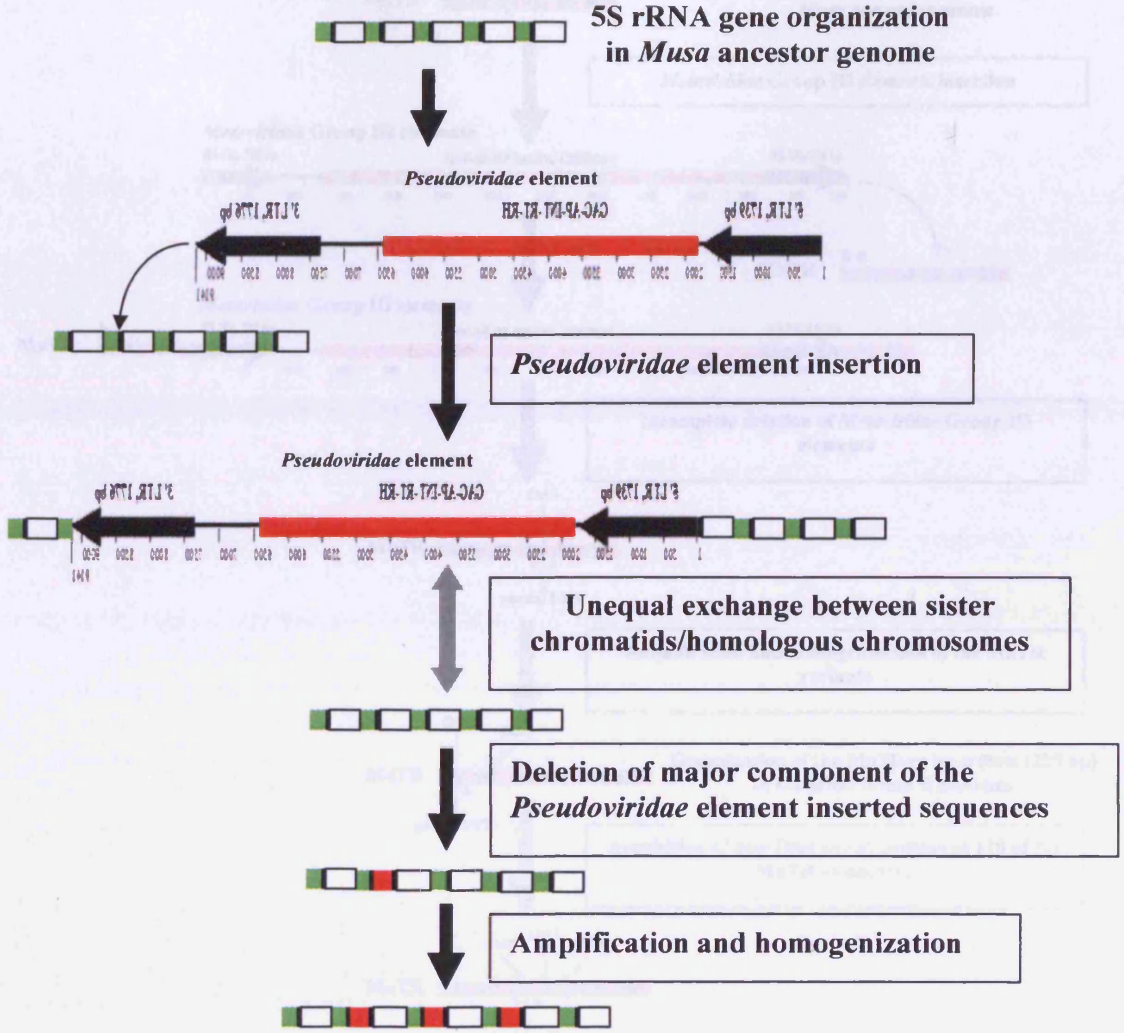


Figure 4.19: The proposed evolution hypothesis of *Musa* 5S rRNA gene. *Pseudoviridae* element is inserted into 3' end of 5S rRNA coding region (green boxes) in anti-sense orientation. The black boxes represent the duplicated region of the short direct repeat whereas the white boxes represent the ITS region. The scale bar = 500 bp per cm.

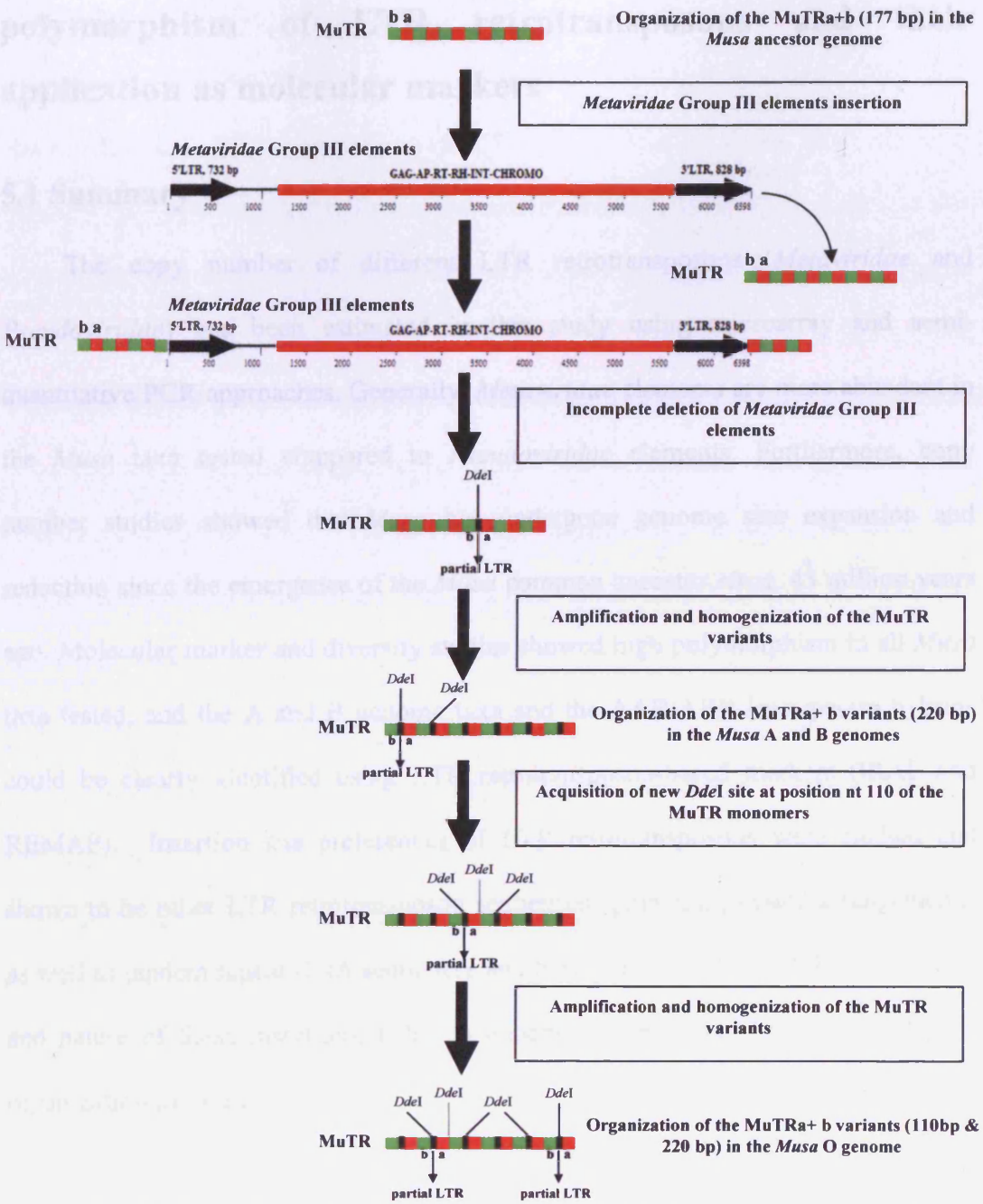


Figure 4.20: The proposed evolution hypothesis of the MuTRs in the *Musa* genomes. The MuTRs (177 bp) in the *Musa* ancestor genome contain no *Metaviridae* Group III element insertion. The insertion and then followed by imperfect deletion of *Metaviridae* Group III element, together with the homogenization process, generated variants of MuTRs (220 bp) with partial LTR sequence of *Metaviridae* Group III elements in the *Musa* A and B genomes. The acquisition of new *Dde*I site in MuTRs variants of the *Musa* O genome, together with partially homogenization process, resulting in the new MuTRs variants of 110 bp and 220 bp.

5.0 Chapter V: Copy number and insertion site polymorphism of LTR retrotransposons and their application as molecular markers

5.1 Summary

The copy number of different LTR retrotransposons (*Metaviridae* and *Pseudoviridae*) has been estimated in this study using microarray and semi-quantitative PCR approaches. Generally, *Metaviridae* elements are more abundant in the *Musa* taxa tested compared to *Pseudoviridae* elements. Furthermore, copy number studies showed that *Musa* has undergone genome size expansion and reduction since the emergence of the *Musa* common ancestor *circa.* 43 million years ago. Molecular marker and diversity studies showed high polymorphism in all *Musa* taxa tested, and the A and B genome taxa and the AAB/ABB interspecies hybrids could be clearly identified using LTR retrotransposon-based markers (IRAP and REMAP). Insertion site preferences of LTR retrotransposons were studied and shown to be other LTR retrotransposon sequences, generating nested arrangements, as well as tandem repeat DNA sequences and host genes. Analysis of the frequencies and nature of these insertions, helps to understand repetitive DNA and host gene organization in *Musa*.

5.2 Introduction

More than 100 LTR retrotransposon families, with individual lengths ranging from ~ 2 kbp to ~ 18 kbp, have been fully characterized in various plant species (see Table 1.1 and references within) and their copy numbers range from a few to several thousand depending on the family. Whilst the content and order of the host genes of

closely related plant species are highly homologous (e.g. Hulbert *et al.*, 1990; Moore *et al.*, 1995; Bennetzen, 1998), most retrotransposon families are species specific (e.g. Avramova *et al.*, 1996; Katsiotis *et al.*, 1997; 2000; SanMiguel and Bennetzen, 1998; Heslop-Harrison, 2000; Friesen *et al.*, 2001, Alix *et al.*, 2004) and their copy numbers are highly variable (SanMiguel and Bennetzen, 1998). This indicates that LTR retrotransposons play an important role in the genome size variation of plant species due to copy number variation (Kumar, 1996, Kumar and Bennetzen, 1999; Vitte and Panaud, 2005).

In maize, the gene contents and orders are very similar to that of sorghum (Hulbert *et al.*, 1990; Moore *et al.*, 1995; Bennetzen, 1998). However, different LTR retrotransposon families in maize have been amplified with copy numbers greater than 10,000 each, which gave rise to a total of approximately 300,000 copies for all maize retrotransposons (SanMiguel and Bennetzen, 1998). Moreover, the retrotransposons that are located near the maize *adh*1 gene are all absent from the orthologous *adh* region in sorghum (Avramova *et al.*, 1996; SanMiguel and Bennetzen, 1998). A similar scenario was observed for rice, where the genome of *Oryza australiensis* is doubled due to the amplification and accumulation of three recently amplified LTR retrotransposon families (*RIRE1*, *Kangourou* and *Wallabi*) to more than 90,000 copies over the last 3 million years (Peigu *et al.*, 2006). Ma and Bennetzen (2004) reported that the nuclear genome sizes of *Oryza sativa* 'Indica' and 'Japonica', have increased more than 2 and 6%, respectively, mainly due to the amplification and accumulation of LTR retrotransposons. Hawkins *et al.* (2006) showed that the *Gossypium* species have undergone 3-fold increments of their genome size due to lineage-specific proliferation and accumulation of LTR retrotransposons since their origin; *Pseudoviridae* elements have differentially

accumulated in the species with the smallest genome whereas *Metaviridae* elements have amplified in the lineages with larger genomes. This evidence suggested an 'increase-only' model for the evolution of genome size as proposed by Bennetzen and Kellogg (1997), although there is also the possibility that LTR retrotransposons can be eliminated by various mechanisms, such as unequal recombination between LTR sequences. Thus, solo LTRs are generated or other type of deletion occurs, which will result in the reduction of the nuclear genome size (Shepherd *et al.*, 1984; Nicolas *et al.*, 1995; SanMiguel *et al.*, 1996; Petrov, 1997; Chen *et al.*, 1998; Vicient *et al.*, 1999; Kirik *et al.*, 2000; Petrov *et al.*, 2000; Devos *et al.*, 2002). However, the lack of observed mechanisms for the 'decrease' model of genome size evolution and the dispersed nature of LTR retrotransposons within host genes, make it unlikely that large deletions without deleterious effects on the survival of the host would occur (Kumar and Bennetzen, 1999). As a result, this has led to the favour of the 'increase-only' model of genome size evolution in plant species.

LTR retrotransposon abundance, dispersed organization, ubiquity, high populations heterogeneity, insertional polymorphism at the gene-rich and gene-poor regions of the host genomes and their roles in the nuclear genome size expansion/reduction and genome rearrangement have allowed them to be used as molecular markers for phylogenetic, biodiversity and genetic linkage analyses in many plant species (see section 1.2.7 and references within; Garber *et al.*, 1999; Gribbon *et al.*, 1999; Kenward *et al.*, 1999; Pearce *et al.*, 1999; Wang *et al.*, 1999; Asif and Othman, 2005; Jing *et al.*, 2005; Smýkal, 2006).

In this chapter, two different approaches, microarray and semi-quantitative PCR, were used to study the copy number variation of the *Pseudoviridae* and *Metaviridae* elements. Three different retrotransposon-based markers (IRAP,

REMAP and PCR-RFLP) were used to study the diversity of *Musa* species. Furthermore, several IRAP bands were cloned and sequenced to study the insertion site preferences of LTR retrotransposons. Finally, internal primers were designed to screen the *Musa* accessions for these insertion sites.

5.3 Materials and Methods

5.3.1 Semi-quantitative PCR

5-fold or 10-fold serial dilutions of the plasmid DNA (standard) and total genomic DNA, 10-fold, of *Musa acuminata* ssp. *malaccensis* (AA) and *M. balbisiana* 'Pisang Gala' (BB) were prepared based on DNA concentrations measured using a UV spectrophotometer. Primer sequences and melting temperatures are described in Table 5.1. PCR conditions and parameters described in section 2.2.2 were followed in detail except the number of PCR cycles was set to 25. All the PCR products (12.5 µl) were run on 1% agarose gels before taking the gel photographs with different UV light intensities. Bands with similar intensities between the plasmid standard and total genomic DNA were selected for calculating copy numbers using the equation 5.1:

[5.1] $$\text{Copy number per haploid genome} = \frac{[M] \times [N]}{[L] \times [D]}$$

M = minimum concentration of nucleic acid detected (g/ml)

N = Avogadro's number (6.022×10^{23} molecules x mole)

L = length of dsDNA in kbp (total length of plasmid + insert, or *1C value)

D = conversion factor from 1 kbp of ds DNA to Daltons, 6.6×10^5 g/mole/kbp

*1C value for *M. acuminata* ssp. *malaccensis* = 587 Mbp and for *M. balbisiana* 'Pisang Gala' = 547 Mbp (Asif *et al.*, 2001).

The copy number of LTR retrotransposons per haploid genome was then obtained by dividing the estimated plasmid standard copy number by the estimated copy number of the haploid genome.

5.3.2 Microarray hybridization and data analysis

Small microarrays were prepared by spotting 24 *Pseudoviridae* clones (Teo *et al.*, 2002) and 29 *Metaviridae* Group I element clones from different *Musa* taxa (Chapter 3) onto aminosilane coated slides as described in section 2.2.13.2. Control DNA (5S rDNA clone pMa794.1; Chapter 4) was distributed throughout the array together with negative controls containing spotting solution without DNA. Arrays were hybridized with sheared and labelled DNA from *M. acuminata* ssp. *malaccensis* (Alexa 594) and *M. balbisiana* 'Pisang Gala' (Alexa 488). The hybridized slides were analyzed using a microarray scanner and the data sets exported to Microsoft® Excel for further analysis. The signal intensity ratio between the LTR retrotransposons and 5S rDNA clone were calculated and used to estimate the copy number of the LTR retrotransposons in both *Musa* species using the equation 5.2,

[5.2] $$\text{Copy number per haploid genome} = (I_{RE}/ I_{Standard}) \times C_{Standard}$$

Where,

I_{RE} = signal intensity of LTR retrotransposon clones

$I_{Standard}$ = signal intensity of standard clone, pMa794.1

$C_{Standard}$ = copy number of standard clone, pMa794.1; 2700 copies

The 5S rRNA gene (Radka2) has been isolated from *M. balbisiana* 'Cameroun' ($BB_{Cameroun}$) by Valárik *et al.*, 2002 and its copy number (2700 copies per haploid genome) and chromosomal loci number had been determined using dot blot hybridization and fluorescent *in situ* hybridization.

5.3.3 IRAP and REMAP PCR amplification

Primers were designed from published and my own retrotransposon LTR and RT sequences after multiple sequence alignments using FastPCR© version 3.1.41 beta from Dr. Ruslan Kalendar. The total genomic DNA of 19 varieties of banana was diluted with sterile dH_2O to 25 ng/µl and 2µl (50ng) used as template in a 25 µl reaction mixture containing 1X PCR buffer (Promega), 1.5mM $MgCl_2$, 10 pmol of each primer, 200 µM dNTP mix, 1.5 U *Taq* polymerase (Promega) in 0.2 ml tubes. Amplification was performed using a T-gradient thermocycler (Biometra) with 1 cycle at 95°C, 5 min; followed by 30 cycles of 95°C, 60s; annealing temperature (T_a)(Nair and Teo *et al.*, 2005; Table 5.1), 60s; ramp +0.5°Cs^{-1} to 72°C; 72°C, 2min + 3s; 1 cycle at 72°C, 10 min. After the PCR reaction, the IRAP and REMAP products were analyzed by electrophoresis on 2% (w/v) agarose (Melford Laboratories Ltd) and detected by ethidium bromide staining. The agarose gel was analysed using GENE FLASH (Syngene).

5.3.4 Phylogenetic analysis of IRAP bands

Phylogenetic trees were constructed by a distance-based method. Evolutionary distances were calculated from tables of bands scored for presence or absence using PAUP version 4.0 beta 10 win (Sinauer Associates, Inc. Publisher). A value of 1 indicates the presence and a value of 0, the absence of a band in a particular location. Trees were constructed by Neighbor-Joining (NJ) clustering method (Saitou and Nei, 1987), implemented in PAUP version 4.0 beta 10 win (Sinauer Associates, Inc. Publisher).

5.3.5 Cloning of the IRAP bands

Selected IRAP bands were excised from the gel and cloned into commercial vectors, based on the protocols described in section 2.2.4. Two or three positive clones were sequenced and homology searches were carried out using BlastN and BlastX for protein for DNA sequences and amino acid analysis, respectively. Selected sequences will be submitted to the EMBL database.

5.3.6 *Musa* Retrotransposon Insertion Screening (MRIS)

Outward primers (Table 5.1) were designed from the LTR retrotransposon sequence and its flanking region. PCR was carried out using parameters described in section 2.2.2 and the melting temperature (T_m) in Table 5.1. The PCR products were analyzed on 1.5% agarose gel.

5.3.7 PCR-RFLP analysis of LTR retrotransposon domains

Pseudoviridae element RT primers (Ty1 and Ty2, Flavell *et al.*, 1992) and *Metaviridae* Group III element RT (RTF and RTR) and LTR (LTRF and LTRR) primers were used to amplify the LTR retrotransposon domains from 19 *Musa* and one *Ensete* species. PCR was carried out using parameters described in section 2.2.2. The primer sequences and annealing temperatures were described in Table 3.1 and Flavell *et al.* (1992). The PCR products were digested overnight with selected restriction enzymes based on cutting sites found in the RT and LTR sequences. The digested products were analyzed on 1.5% agarose gel.

5.4 Results

5.4.1 Copy number quantification of LTR retrotransposons

The nuclear genome size of *M. acuminata* ssp. *malaccensis* and *M. balbisiana* 'Pisang Gala' have been calculated by Asif *et al.* (2001) using flow cytometry and estimated to be 587 Mbp/1C for and 547 Mbp/1C, respectively. Two different semi-quantification approaches were used to estimate the copy number of different LTR retrotransposons. Microarray technology was used for copy number quantification of the *Pseudoviridae* elements and *Metaviridae* Group I elements in *M. acuminata* ssp. *malaccensis* and *M. balbisiana* 'Pisang Gala' (Table 5.2). For *Metaviridae* Group II and III elements, a semi-quantitative PCR approach using LTR and RT specific primers was applied (Figure 5.1).

For the microarray approach, the comparison between the hybridization signal intensities of the LTR retrotransposon clones and the standard clone, the 5S rDNA clone pMa794.1 (see section 4.4.1) allowed quantification of LTR retrotransposon copy numbers (Table 5.2). *Metaviridae* elements, more specifically Group I elements, (2200 – 17000 copies) were present in higher copy numbers compared to the *Pseudoviridae* elements (2800-7000) in both *M. acuminata* ssp. *malaccensis* and *M. balbisiana* 'Pisang Gala' genome (Table 5.2). But the *malaccensis* genome overall contained more *Metaviridae* (3000-17000 copies) and *Pseudoviridae* (4000-7000 copies) elements compared to the 'Pisang Gala' genome (*Metaviridae*: 2200-5800 copies; *Pseudoviridae*: 2800-4000 copies). One *Metaviridae* Group I element clone, AsTY3-8, showed a slightly higher copy number in the 'Pisang Gala' genome (4680 copies) compared to the *malaccensis* genome (4310 copies). The members of the subgroup II and IV of the *Metaviridae* Group I elements, based on MP phylogenetic tree classification described in Figure 3.6 (see Chapter 3), were present in higher

copy number compared to other subgroup members (Table 5.2). In particular, four members of the Subgroup II (AbTY3-8, AsTY3-9, B9-1 and SY-30A) and two members of the Subgroup IV (AbTY3-3 and AbTY3-14) were present in more than 10,000 copies in the *malaccensis* genome (Table 5.2).

For the semi-quantitative PCR approach, amplification with retrotransposon-specific primers of serial dilutions of total genomic DNA and plasmid standards was carried out. Comparison of the PCR bands of similar intensities allowed estimation of LTR retrotransposon copy numbers (Figure 5.1). For the *Metaviridae* Group III element RT genes, the PCR band intensities in lane 7 (*M. acuminata* ssp. *malaccensis*) and lane 13 (*M. balbisiana* 'Pisang Gala') are similar to lane 18 (standard, MA_RT1)(Figure 5.1A). By using the equation in materials and methods (section 5.3.1), its copy number is estimated to be 570 copies for the *malaccensis* genome and 530 copies for the 'Pisang Gala' genome (Table 5.3). For *Metaviridae* Group III element LTRs, PCR band in lane 4 (*M. acuminata* ssp. *malaccensis*) showed similar intensities to lane 9 (standard, MbTY32a1)(Figure 5.1B). Using the same calculation, the copy number of their LTRs in the *malaccensis* genome (Figure 5.1B) is estimated to be 1800 copies (Table 5.3), which is ~ 3 times more than the RT gene. For *Metaviridae* Group II element RT genes, the PCR band intensities of *M. acuminata* ssp. *malaccensis* (lane 12, Figre 5.1C) and *M. balbisiana* 'Pisang Gala' (lane 19, Figure 5.1C) were similar to the amplification product of the standard, MalcoTR_RT1 (lane 4, Figure 5.1C). This shows an increase of ~ 10 fold in the *malaccensis* genome (2700 copies) compared to the 'Pisang Gala' genome (250 copies).

5.4.2 Use of LTR retrotransposons as molecular markers

5.4.2.1 IRAP and REMAP primer design strategy

The PCR amplification schemes of the IRAPs and REMAPs are described in Figure 5.2. LTR retrotransposons can be inserted into the host genome in three different orientations, namely 'head-to-head', 'tail-to-tail' and 'head-to-tail' orientation based on their LTR and ORF sequence. The retrotransposon primers were designed facing outwards from the LTR and RT of LTR retrotransposons whereas for the microsatellites, 7 - 8 repeat units were used directly as primers (Table 5.1). This allows the amplification of their flanking region if the LTR retrotransposons are inserted closely enough together ('nested' structure) for PCR to amplify across. Depending on the distribution of the LTR retrotransposons within the genome, this results in a polymorphic banding pattern after PCR amplification. The insertion of LTR retrotransposons into other types of repetitive DNA, in this case, the microsatellite repeats provides another approach to the study of genomic diversity (Figure 5.2).

5.4.2.2 IRAP and REMAP study on *Musa* species

Highly polymorphic banding patterns were observed in *Musa* taxa from four different sections (*Eumusa*, *Rhodochlamys*, *Callimusa* and *Australimusa*) when outward LTR primers from two different LTR retrotransposons (*Metaviridae* Group III element and MuLARD-like elements) were used (Figure 5.3). In general, 4 to 13 polymorphic bands between 150 bp and 2.5 kbp were observed for two *Metaviridae* Group III element LTR primers: *gypsy*-IRAP primer (GyLTRev, Nair and Teo *et al.*, 2005; Table 5.1 and Figure 5.3A) and GyLTRFor (Table 5.1 and Figure 5.3B) whereas 5 to 16 polymorphic bands between 100 bp and 3.5 kbp were observed for

MuLARD-like element LTR primers (Figure 5.3C) with *M. balbisiana* 'Pisang Gala' (BB) containing the lowest number of IRAP bands among the *Musa* taxa tested. Overall, higher numbers of IRAP bands were observed in taxa from *Rhodochlamys*, *Callimusa* and *Australimusa* sections compared to members of the *Eumusa* section (Figure 5.3A-C). Similar but not identical banding patterns were observed in the two different A genome subspecies, *M. acuminata* ssp. *malaccensis* and ssp. *banksii* with all three IRAPs shown (Figure 5.3A-C). The presence of similar banding patterns at between ~ 100 and 300 bp in all the members of the *Callimusa* and *Australimusa* sections except for *M. violascens* (Figure 5.3C) and shared bands in the other IRAP panels (Figure 5.3B & C) indicate the close relationship of these species.

The observation of a ~ 3.5 kb band (Figure 5.3C) which corresponds to the internal sequence size of the MuLARD-like element indicates that full-length MuLARD-like elements are present in all the *Musa* species tested except *M. sanguinea* (section *Rhodochlamys*). To confirm this, a ~ 3.5 kbp band was partially sequenced from *M. acuminata* ssp. *malaccensis* and a high homology to MuLARD-like elements was indeed observed and confirmed that these ~ 3.5 kbp bands correspond to the internal domain of the MuLARD-like element found in BAC clones (see Chapter 3 for the MuLARD-like description).

Highly polymorphic banding patterns were observed for REMAPs using the outward *Pseudoviridae* RT primer (MusaTY2R) and microsatellite GA repeat primer (BT_GAC) (Figure 5.3D). The number of amplified bands is generally higher than IRAP bands with lowest band number (7 REMAP bands) observed in *M. laterita* (section *Rhodochlamys*). The band size is also bigger than IRAP bands ranging from 400 bp to 2 kbp (Figure 5.3D). Three members of the *Rhodochlamys* section also shared similar REMAP banding patterns (*M. laterita*, *M. sanguinea* and *M. mannii*).

Members of the *Callimusa* and *Australimusa* sections had more REMAP bands than members of the two other sections (*Eumusa* and *Rhodochlamys*) (Figure 5.3D). Again, the two *Musa* A genome subspecies, *M. acuminata* ssp. *malaccensis* and ssp. *banksii* shared many bands (Figure 5.3D).

5.4.2.3 IRAP study on *Musa* intra- and inter-species hybrids

A single *Musa gypsy*-IRAP primer, GyLTRev (Table 5.1) was used to screen 36 banana samples collected from the field station of the Banana Research Station, Kerala Agricultural University, Kerala, India. The *M. balbisiana* 'Pisang Gala' (BB) from Malaysia was used as a control species for B genome identification in this study. The results of this germplasm screening have been published (Nair and Teo *et al.*, 2005).

The 36 cultivars (Table 2.1) showed multiple polymorphic bands (Fig 5.4) and a specific band (~350 bp) was observed in all the diploid (Figure 5.4A) and triploid (Figure 5.4B & C) cultivars containing a B genome, with band intensity increasing in cultivars with two B genomes. This band was designated as *gypsy*-IRAP B specific band ($B_{gypsy\text{-}IRAP}$). However in Figure 5.4B, the $B_{gypsy\text{-}IRAP}$ band intensity is low in three cultivars: 'Velli padati', 'Padati', and 'Monthan'. These cultivars were previously classified as ABB. In Figure 5.4A, two diploid cultivars showed a similar banding pattern: 'Kadali' and 'Cherukadali' and the $B_{gypsy\text{-}IRAP}$ band is observed in both cultivars indicating that these two cultivars are inter-species hybrids. Similar observations were also made in triploid hybrids with no $B_{gypsy\text{-}IRAP}$ band detected in 'Robusta' and 'Dwarf Cavendish', where both of these cultivars had been classified as AAA (Figure 5.4B). The cultivar 'Velli padati' showed a similar $B_{gypsy\text{-}IRAP}$ to that of 'Padati', 'Monthan', 'Chara padati', 'Kumbilla kannan' and 'Etta padati' cultivars

(Figure 5.4B). In Figure 5.4B, the cultivar 'Dooth sagar' showed a high intensity B specific band but based on its morphological characteristic it was thought to be AAB. On the other hand, a $B_{gypsy\text{-}IRAP}$ band with a similar intensity to that of 'Kosta bontha' (ABB) was detected in the cultivar 'Manoranjitham' (AAA) (Figure 5.4C). This indicates that a single *Musa gypsy*-IRAP primer is able to distinguish between intra- and inter-species cultivars tested.

In a previous study (Teo, 2002), a B genome specific band was detected in banana cultivars from the Malaysian germplasm collection (MARDI) using *copia*-IRAP primers published by Kalendar *et al.* (1999) and this band designated as *copia*-IRAP B specific band ($B_{copia\text{-}IRAP}$). A pair of internal primers (BFor and BRev, Nair and Teo *et al.*, 2005; Table 5.1) was designed and used to amplify the $B_{copia\text{-}IRAP}$ band from AAB and ABB inter-species hybrids (Figure 5.5A & B). The $B_{copia\text{-}IRAP}$ band at ~ 420 bp is detected in all the AAB/ABB inter-species hybrids with different band intensities (Figure 5.5A & B). In order to search for further polymorphism in the amplified bands, PCR products were digested with *Alu*I. Remarkably the restriction pattern of all the five ABB cultivars: 'Manorajitham', 'Kosta bontha', 'Kanchi kela' and 'Boothbale' in Figure 5.5C, and 'Dooth sagar' in Figure 5.5D were distinct from the cultivars with AAB genome, giving two fragments around 200bp long, rather than fragments of 300 and 100 bp, in all cases with a conserved shorter fragment (Figure 5.5C & D). This allows the clear identification of AAB and ABB cultivars among all the 36 banana cultivars tested.

5.4.2.4 Variation of IRAP bands in closely related *Musa* cultivars

In this study, seven diploid AA bananas including male parent, 'Pisang Jari Buaya' (AAcv) and female parent, 'Prata Ana' (AAB), of the cultivar 'GoldFinger'

(AAAB) were subjected to IRAP amplification using primer, GyLTRev (Table 5.1). 'GoldFinger' is an improved variety that is resistant to Black Sigatoka and this resistance was inherited from 'Pisang Jari Buaya' (Romero & Sutton, 1997).

The resulting banding patterns of 'Prata Ana' and 'GoldFinger' are very similar (Figure 5.6) indicating that most of the retrotransposon insertion sites in 'Prata Ana' are inherited by 'GoldFinger'. There are two 'Prata Ana' IRAP bands missing in 'GoldFinger'. This suggests that not all the chromosomes of 'Prata Ana' are transferred to 'GoldFinger' and that the A genomes in 'Prata Ana' are heteromorphic. Most of the IRAP bands in 'GoldFinger' were shared among all analyzed AA diploid bananas (Figure 5.6). However, the IRAP banding pattern of 'Pisang Jari Buaya' is more similar to the 'GoldFinger' compared to other AA genome cultivars (Figure 5.6). A dendrogram of scored IRAP bands was constructed using Neighbor-Joining (NJ) method showed that 'Pisang Jari Buaya', 'Prata Ana' and 'GoldFinger' are in the same subgroup (Figure 5.6) as would be expected from 'Prata Ana' and 'Pisang Jari Buaya' as parents of 'GoldFinger'. Hence, analysis of IRAP bands is able to identify the parental origins of hybrid *Musa* taxa.

5.4.2.5 LTR retrotransposon domain diversity in *Musa* genomes

RFLPs of different *Metaviridae* Group III and *Pseudoviridae* element domains were used to study the relationship between LTR retrotransposon diversity and *Musa* genome composition. The RT gene and LTR region of both elements were amplified from 19 *Musa* and one *Ensete* species (Figure 5.7) using retrotransposon specific primers (see Table 3.1). The PCR products were digested with a range of restriction enzymes based on cutting sites found in the DNA sequences (*Hind*III, *ScrF*I, *Rsa*I or *Sma*I) and separated on high percent agarose gels (Figure 5.7).

PCR-RFLP analysis of *Pseudoviridae* RT domain showed that accessions from *Eumusa* (2n=22) and *Rhodochlamys* (2n=22) sections appear to have identical banding patterns of two distinct bands (~ 100 and ~ 120 bp) together with an uncut band at ~ 270 bp (Figure 5.7a). All species in section *Australimusa* (2n=20) showed only one significant band at ~ 150 bp together with uncut band (~270 bp) whereas only the uncut band was observed in all the *Callimusa* (2n=20) species, except for *M. violascens* (Figure 5.7A).

PCR-RFLP analysis of the *Metaviridae* RT domain using three different restriction enzymes (Figure 5.7B & C) showed that accessions within a section seem to have more similar banding patterns than those between sections. However, *M. balbisiana* (BAL), *M. velutina* (VEL) and *M. violascens* (VIO) are different. PCR-RFLP analysis of *Metaviridae* LTRs (Figure 5.7D & E) showed only small differences amongst all *Musa* species tested, but notably *M. balbisiana* (BAL) shows unique combinations of bands and *M. violascens* (VIO) seems to fit better to the *Eumusa* or *Rhodochlamys* section.

5.4.3 *Musa* retrotransposon insertional polymorphism

In order to study which types of 'general' genomic DNA sequences flank LTR retrotransposons, 11 IRAP bands with different sizes from different *Musa* taxa were excised from IRAP gels, cloned and sequenced (Table 5.4). The BlastN analysis against Genbank and local database detected different types of repetitive DNA (tandem repeat DNA and LTR retrotransposon) in all 22 clones (Table 5.4). Internal primers were designed from selected clones to study the insertion site polymorphism in different *Musa* species and cultivars.

One particular IRAP band ($BB_{gypsy\text{-}IRAP}$, see section 5.4.2.3) is only observed in B genome containing *Musa* species and cultivars (Figure 5.4). The clones of this band were sequenced and designated as But-LTR2.2 and But-LTR3.1 (Table 5.4). DNA-DNA homology analysis showed high homology to the MuTR repeats described in Chapter 4 and also part of the *Metaviridae* Group III element LTR region. Further analysis detected complete MuTRb sequences and part of the MuTRa sequence in both clones (Figure 5.8A). The sequence similarity between But-LTR2.2 and But-LTR3.1 is ~ 99%. A single nucleotide insertion/deletion (indel) and two A↔G transitions were detected in the pairwise alignment of the two clones (Figure 5.8B). Internal primers (BGyFor and BGyRev, Table 5.1 and Figure 5.8B) designed from the consensus sequence of But-LTR2.2 and But-LTR3.1 were used to amplify this region in different *Musa* taxa (Figure 5.9). The expected band of 275 bp with different intensities was observed in all *Musa* taxa containing the B genome, together with a low intensity band of different size in *M. schizocarpa* (SS) (Figure 5.9). This confirmed that the $BB_{gypsy\text{-}IRAP}$ band is only present in B genome containing taxa.

Another IRAP band amplified using the *gypsy*-IRAP primer from *M. acuminata* 'Pisang Berangan' (clone B_LTR2.1), showed high homology to the internal domain of *Metaviridae* Group III element (Figure 5.10). This indicated that the two *Metaviridae* Group III elements inserted into each other. In order to testify this finding, the internal primer (MRISFor) and a further outward *Metaviridae* Group III element LTR primer, LTRF01 (see Figure 5.10; sequences given in Table 5.1) were used. After the amplification, the expected multiple banding patterns indicative of 'nested' insertion populations with varying lengths was observed in all the *Musa* cultivars tested together with a ~ 1.1 kbp band in the control DNA, *M. acuminata*

'Pisang Berangan' albeit very weak (Figure 5.11). Interestingly, the expected band at ~ 1.1 kbp was only observed in 12 out of the 26 *Musa* cultivars tested while two bands of ~ 630 and ~ 1600 bp were present in all the *Musa* taxa tested (Figure 5.11A & B). A similar situation was observed in the amplification products of 16 further *Musa* taxa consisting of nine *M. acuminata* and seven *M. balbisiana* accessions (Figure 5.11C). However, the expected band at ~ 1.1 kbp was only observed in one *M. acuminata* ('Calcutta 4') and one *M. balbisiana* ('Lal Velchi') species (Figure 5.11C). The banding pattern of two *M. balbisiana* taxa ('Lal Velchi' and 'Honduras') is different from the remaining *M. balbisiana* taxa tested, with the banding pattern of 'Lal Velchi' resembling the banding pattern observed in *M. acuminata* accessions (Figure 5.11C). Remarkably, two bands of ~ 700 bp and 1300 bp were only observed in some of the *M. acuminata* accessions tested (Figure 5.11C), where the former band is observed in all the island *M. acuminata* accessions (Agutay, Malaccensis, Zebrina, Microcarpa, Banksii and Paliama) and the latter band is observed in four out of 6 island *M. acuminata* accessions (Zebrina, Microcarpa, Banksii and Paliama) (Figure 5.11C).

An insertion of the *Pseudoviridae* element, MuPseudo2 (Figure 5.12A) into Radka10 repeats (Valárik *et al.*, 2002) was observed when analyzing BAC sequences for full-length LTR retrotransposons (see Table 3.2). Part of the Radka10 repeat also showed homology in DNA level to the LTR and Primer Binding Site (PBS) of the *Metaviridae* element, MuMeta3 (Figure 5.12B; see Table 3.2 for MuMeta3 description). In order to analyze the genome wide association of Radka 10 related repeats and LTR retrotransposons, primers were designed from the 3' LTR sequence of MuPseudo2 (MTY13LTRF) and Radka10 repeat (MTY1RISR02) to screen *Musa* species (Figure 5.13). The expected band at size ~ 250 bp was observed in the

control genomic DNA, *M. acuminata* 'Calcutta 4' (Figure 5.13A) and other *Musa* taxa tested except for members of the *Callimusa* section (Figure 5.13B), together with two unexpected bands at size of ~ 210 bp (only observed in members of *Australimusa* section) and ~ 280 bp (observed in *Eumusa* and *Rhodochlamys* sections) (Figure 5.13).

In another analysis, the *Pseudoviridae* element insertions into 5' upstream of the *M. acuminata* class III acidic chitinase gene (AY525367) was detected during the BlastN homology search of *Pseudoviridae* elements against Genbank database (Figure 5.14). Remarkably, primers (MRIS_ChiF and MRIS_ChiR) allowed the detection of *Musa* cultivars which contain *Pseudoviridae* elements inserted into 5' upstream of the *M. acuminata* class III acidic chitinase gene (AY525367) (Figure 5.15) and nineteen out of 32 *Musa* cultivars tested showed amplification of expected band at size of 812 bp (Figure 5.15).

5.5 Discussion

5.5.1 LTR retrotransposon copy number and *Musa* genome size evolution

Two different methodologies were used to quantify the copy number of LTR retrotransposons (*Metaviridae* and *Pseudoviridae*) in *Musa*, mainly focusing on *M. acuminata* ssp. *malaccensis* (A genome) and *M. balbisiana* 'Pisang Gala' (B genome). The first approach by microarray technology used fluorescently labelled total genomic DNA from *malaccensis* and 'Pisang Gala' as probes in hybridization to LTR retrotransposon clones spotted on the microarray slide. The spotted clones containing the RT region showed ~ 40-90% homology to each other (see Chapter 3). Array hybridizations thus give a rough estimate of the presence of sequences 85% or more homologous to the retrotransposon RT domain within the genomes used as

probes. The conserved and ubiquitous 5S rDNA was chosen as control because its copy number has been determined for *Musa* (Valárik *et al.*, 2002). Furthermore, six 5S rDNA chromosome loci have been reported in most *M. balbisiana* species ('Tani', 'Butohan 2', 'Singapuri', and 'M. balbisiana'; Doležel *et al.*, 1994; Lysák *et al.*, 1999; Doležel *et al.*, 2004) except for *M. balbisiana* 'Cameroun' and 'Honduras' which have 4 (Bartoš *et al.*, 2005). This allowed the quantification of the abundance of the individual retrotransposon RT copy number by microarray technology (Table 5.2). The second approach used semi-quantitative PCR with different retrotransposon specific primers (Table 3.1) on the same total genomic DNA used in the microarray analysis. This approach relies on the specificity of the primers and is only able to estimate copy numbers of conserved and less degenerate elements such as the *Metaviridae* Group II and III elements.

The copy numbers of the *Pseudoviridae* elements are generally higher than the *Metaviridae* elements in both *malaccensis* and 'Pisang Gala' genome, except for *Metaviridae* Group I elements (Table 5.2 and 5.3). Interestingly, both elements are more abundant in the *malaccensis* genome compared to the 'Pisang Gala' genome (Table 5.2 and 5.3), with only one exception, clone AsTY3-8 (Table 5.2). it is noteworthy that none of the *Pseudoviridae* and *Metaviridae* RT sequences tested by microarray and PCR were species specific (i.e. showing no hybridization with either genomic DNA). This is in agreement with the phylogenetic trees of these clones (see Chapter 3). Different degrees of copy number increments between the *malaccensis* and 'Pisang Gala' genome were observed ranging from one to about 4 (Table 5.2) and even up to 10 for *Metaviridae* Group II elements (Table 5.3).

Overall, the copy number variation between genomes and retrotransposon families indicates that different amplification rates have operated on different types

of LTR retrotransposons. The *malaccensis* genome (587 Mbp/1C) is generally ~ 7% larger than the 'Pisang Gala' genome (547 Mbp/1C), which suggests that LTR retrotransposons play an important role in their nuclear genome size variation. Similar observations were also reported in other plant species, where LTR retrotransposons contributed to the genome size variation of their host (Bennetzen and Kellogg, 1997; SanMiguel and Bennetzen, 1998; Vicient *et al.*, 1999; Shirasu *et al.*, 2000; Wicker *et al.*, 2001; Ma and Bennetzen, 2004; Guo *et al.*, 2006; Hawkins *et al.*, 2006; Peigu *et al.*, 2006).

High numbers of solo LTRs of the *Metaviridae* Group III elements have been detected in the *malaccensis* genome using semi-quantitative PCR with LTR-specific primers (Figure 5.1B, Table 5.3). If all ~ 570 copies of *Metaviridae* Group III element RTs in the *malaccensis* genome (Figure 5.1A; Table 5.3) come from full-length elements with two LTRs, the LTR copy number should be ~ 1140. However, ~ 1800 copies of LTR were estimated in the *malaccensis* genome (Figure 5.1B; Table 5.3) and this number would increase if the RT domain were truncated, which indicates that about 660 of them could actually be solo LTRs. In another study, the solo LTR of MuLARD-like elements was detected in the *M. acuminata* BAC, clone MuH9 (Aert *et al.*, 2004). Solo LTRs of different LTR retrotransposons were observed in other plant species and have been proposed to form by unequal homologous recombination and illegitimate recombination (Shirasu *et al.*, 2000; Petrov, Lozovskaya and Hartl, 1996; Petrov *et al.*, 2000; Devos, Brown and Bennetzen, 2002; Vitte and Panaud, 2003; Vitte and Panaud, 2005). The partial or complete removal of LTR retrotransposons by these two mechanisms can counterbalance their amplification and result in genome size reduction (Devos, Brown and Bennetzen, 2002; Vitte and Panaud, 2003; Vitte and Panaud, 2005).

Additionally, highly heterogeneous populations of the *Metaviridae* Group I elements were found, where the RT genes are truncated or disrupted by other types of repetitive DNA (see Figure 3.4b), possibly also due to illegitimate recombination. This further supports the existence of processes that eliminate or partially delete full-length LTR retrotransposons and play a role in counterbalancing the amplification of LTR retrotransposons in *Musa*.

5.5.2 *Musa* diversity study using retrotransposon-based markers

In other plant species including *Hordeum* (Kalendar *et al.*, 1999), *Avena* (Yu *et al.*, 2000), *Pisum* (Pearce *et al.*, 2000), *Triticum* and *Aegilops* (Queen *et al.*, 2004) and *Agave* (Bousios *et al.*, 2007), LTR retrotransposon based molecular markers were extensively used to study diversity. In previous studies of *Musa* cultivars, this marker system was used to study the diversity of different cultivars (Teo *et al.*, 2005), where outward LTR primers from barley *Pseudoviridae* elements, *BARE*-1, directly generated highly polymorphic banding patterns in the 16 *Musa* cultivars tested. This resulted in the detection of the *copia*-IRAP B specific band, $B_{copia\text{-}IRAP}$. In this study, in order to increase the specificity of retrotransposon based marker systems, outward LTR primers from *Musa* specific *Metaviridae* Group III elements (Figure 5.3A & B) and LARD-like elements (Figure 5.3C) as well as the microsatellite GA repeats, were used to generate polymorphic banding patterns (Figure 5.3 and 5.4). PCR profiles were indeed clearer with sharp and reproducible bands (Figure 5.3 and 5.4). In addition, a *gypsy*-IRAP B specific band, $B_{gypsy\text{-}IRAP}$, was detected (Figure 5.4 and 5.9). Further, I developed the PCR-RFLP system that allows robust discrimination of B genome containing cultivars (Figure 5.5) and *Musa* sections (Figure 5.7).

Most of the IRAP bands were species specific and generally IRAP bands were seldom shared by all taxa tested (Figure 5.3A - C). This suggests that both MuLARD-like and *Metaviridae* Group III elements are still actively transposed into new genomic locations since the divergence of the *Musa* taxa tested (Figure 5.3A - C). On the other hand, the observation of the IRAP bands that are shared among closely related *Musa* cultivars (Figure 5.4 and 5.6) provide some indication that some LTR retrotransposon insertions are stable and can be inherited (Figure 5.6). Integration of new LTR retrotransposon copies, typically generating 5–12 kb genomic insertions, depends on the LTR retrotransposon families. These new copies are inserted and not transpositionally removed, which facilitates phylogenetic analyses (Shimamura *et al.*, 1997). Accumulation, fixation and incomplete excision of retrotransposon insertions can eventually cause genomic diversification, which can be detected using the retrotransposon-based markers (Rodhe *et al.*, 1995; Rodhe, 1996; Flavell *et al.*, 1998; Kalendar *et al.*, 1999; Pearce *et al.*, 1999; Bousios *et al.*, 2006; Smýkal, 2006). In my study the high homology between IRAP clones (957 bp) from *M. acuminata* 'Pisang Berangan' (AAA) (clone B_LTR1.1) and *M. paradisiaca* 'Mutiara' (AAB) (clone M1B10.1, M1B10.2 and M1B10.3), in terms of clone length, sequence, insertion site and position (Table 5.4) confirms that IRAP bands can be inherited. This is important when parental origin is tested and also to identify cultivars with synonymous names.

Because of the often confusing origin and identity of local crop varieties, characterization of germplasm is extremely difficult based on morphological characters alone. The close genetic relationship among the cultivars as well as somatic mutations contributes further obstacles to the correct identification of the clones (Kahangi *et al.*, 2002). The farming community complicates the matter even

further by conferring local names or dialects. Molecular markers such as microsatellites, RAPDs, and AFLPs have great potential to characterize and identify banana cultivars from Kenya, Brazil and India (Creste *et al.*, 2003; Onguso *et al.*, 2004; Bhat *et al.*, 2004). Here, we applied IRAP markers to a banana germplasm collection from South India that was characterized based on IPGRI descriptors but was reported to contain varieties sometimes known by different names in different regions (Amalraj, 1992; Nair and Teo *et al.*, 2005). IRAP amplification using *gypsy*-LTR primer, GyLTRev, allowed the detection of banana cultivars with synonymous name such as 'Kadali' and 'Cherukadali' (AB); 'Robusta' and 'Dwarf Cavendish' (AAA); 'Etta Padati' and 'Padati' (AAB) and 'Kosta bontha' and 'Manorajitham' (ABB) (Figure 5.4). While IRAPs and REMAPs (Figure 5.3) were able to distinguish synonym species of different *Musa* sections and also cultivars, PCR-RFLP systems are needed to distinguish the B genome containing cultivars (Figure 5.5) and *Musa* sections (Figure 5.7).

Screening for the B genome has become very important as Banana Streak Virus (BSV) has been found to be integrated into the B genome (Harper *et al.*, 1999). While BSV is not normally expressed in diploid Bs, it is actively expressed in the AAB and AAAB interspecies hybrids such as 'Obino I'Ewai' (AAB), 'Mysore' (AAB), 'GoldFinger' (AAAB) (Lockhart and Olszewski, 1993; Lockhart, 1994; Dahal *et al.*, 1998) and other banana genotypes (Harper *et al.*, 1999; 2005; Lockhart and Jones, 1999; Geering *et al.*, 2000; Geering *et al.*, 2005a; b). The screening of the banana germplasm for endogenous BSV has become very important because it is increasingly recognized as causing disease in particular after abiotic stress and tissue culture. This has implications not only in the safety management of germplasm, but also in breeding and vegetative propagation of *Musa* species (Harper and Hull, 1998).

Balint-Kurti *et al.* (2000) suggested that *Metaviridae* LTR retrotransposons (*Monkey*) were introduced into *Musa* prior to the divergence of *M. acuminata* (AA, section *Eumusa*), *M. balbisiana* (BB, section *Eumusa*) and *M. velutina* (section *Rhodochlamys*). They were able to distinguish the A and B genomes using *Metaviridae* LTR retrotransposon probes hybridized on *Hind*III digested genomic DNA from 8 cultivars of banana with *M. velutina* (non A non B genome) as the control, but could not differentiate between AAB and ABB genomes. In this study, the AAB and ABB interspecies *Musa* hybrids could be clearly distinguished using PCR-RFLP patterns generated by the *copia*-IRAP B specific band (Figure 5.5C & D) and this will help banana breeders in the selection of the interspecies hybrids and detection of the B genome integration.

The PCR-RFLP system based on the restriction enzyme sites present in the RT and LTR sequences of the *Pseudoviridae* and *Metaviridae* Group III elements also clearly distinguish *Musa* species within and between sections (Figure 5.7). The presence of similar banding patterns in all members of the *Callimusa* and *Australimusa* sections except for *M. violascens* (section *Callimusa*) (Figure 5.7) suggests that the species from these two sections might have originated from a single common ancestor, $2n=2x=20$. *M. violascens* ($2n=2x=20$, section *Callimusa*) showed a similar PCR-RFLP banding pattern to the *Eumusa* and *Rhodochlamys* accessions (Figure 5.7). *M. violascens* was first discovered in Peninsular Malaysia where a lot of *M. acuminata* ($2n=2x=22$, section *Eumusa*) wild species were observed. This indicates that *M. violascens* might be an intermediate between the two sections.

5.5.3 Insertion site preference of LTR retrotransposon

Sequencing of the IRAP bands amplified by using single outward *Metaviridae* Group III element LTR primer (GyLTRev) allows the detection of the flanking regions between two LTR sequences (Figure 5.2, Table 5.4). Fifteen out of 22 IRAP clones (~ 70%, Table 5.4) showed homology to the internal sequences of three different LTR retrotransposons previously identified in *Musa* (*Metaviridae* Group III element, *Pseudoviridae* element and LARD-like element) (Chapter 3) indicating nested arrangement of LTR retrotransposons within the genome.

A *Metaviridae* Group III element has been shown to insert into *Musa* tandem repeat DNA, MuTR and become part of this tandem repeat DNA (see Figure 4.20). In this chapter, two further *Metaviridae* Group III element insertions were detected in the B specific band, $B_{gypsy\text{-}IRAP}$ (Figure 5.8 - 5.9), which is present in all the B genome containing *Musa* taxa (Figure 5.3A; 5.4 and 5.9, see section 5.5.2). These clones also contain sequences homologous to the MuTR repeats (see Figure 5.8). This suggests that insertion occurred after the divergence of *M. acuminata*, *M. ornata* and *M. balbisiana* from the *Musa* ancestor. One possibility is that in the B genome, two independent insertions of *Metaviridae* Group III elements, in sense and antisense orientation, occurred in the same genomic region that consists of long tandemly organized repeat MuTRs (Figure 5.16). This generated a genomic region that can be detected by a single *Metaviridae* Group III element LTR primer (Figure 5.16). LTR retrotransposons have been shown to insert into tandem repeat DNA in other plant species (Nakajima *et al.*, 1996; Hisatomi *et al.*, 1997; Kumekawa *et al.*, 2001; Cheng *et al.*, 2002; Tek *et al.*, 2005) and sometimes become part of the tandem repeat DNA (Cheng and Murata, 2003).

PCR analysis of one of the IRAP clones, B_LTR2.1 (Figure 5.10; Table 5.4), using an internal primer and outward LTR primer, amplified multiple banding patterns in all the *Musa* taxa tested and together with the IRAP bands (Figure 5.3, 5.4 and 5.6), suggesting a 'nested' structure of the *Metaviridae* Group III elements (Figure 5.11). BAC analysis showed that *Metaviridae* Group II element, eCoTR-REs, were inserted into another *Metaviridae* element from different groups (MuMeta3, *Metaviridae* Group I element) (see Chapter 3). Furthermore, a *Pseudoviridae* element, MuPseudo2, is inserted in the Radka10 repeat (Valárik *et al.*, 2002), which actually comprises the LTR sequence of MuMeta3 and part of the MuMeta3 internal sequence (Figure 5.12) and this insertion is observed in most of the *Musa* species tested except the species belonging to *Callimusa* section (Figure 5.13). The absence of the 245 bp band might be due to high heterogeneity in the primer binding region or this insertion site has been lost during *Musa* evolution. All the results described above suggest that *Metaviridae* Group III elements inserted into LTR retrotransposon sequences from the same family or other families. Similar observations are reported in other plant species (Manninen and Schulman, 1993; Nakajima *et al.*, 1996; SanMiguel *et al.*, 1996; Noma *et al.*, 1997; Vicient and Martinez-Izquierdo, 1997; Kumekawa *et al.*, 1999; Yang *et al.*, 2005).

Detection of the *Pseudoviridae* element insertion in *M. acuminata* Class III acidic chitinase gene (AY525367) (Figure 5.14) and 5S rDNA sequences (see Figure 4.3) provides another perspective to the insertion site preference of the LTR retrotransposons, where they are inserted into the gene-rich region of the *Musa* genome. The observation of this insertion in only 19 out of 32 *Musa* cultivars tested and the absence of the expected bands in two *M. acuminata* ssp. *malaccensis* (AA) accessions (Figure 5.15) suggests that this insertion event occurred after the

divergence of the *M. acuminata* ssp. *malaccensis* from other *M. acuminata* subspecies. In *Musa* species, the fruit ripening mechanism has been intensively studied due to the global interests in banana as worldwide dessert fruit and an important carbohydrate staple crop in the tropics (Marriott, 1980). Several fruit ripening-related proteins have been identified and studied such as those that are involved in ethylene biosynthesis, starch degradation, respiration and cell wall degradation (Medina-Suarez *et al.*, 1997). In another study, the Class III chitinase has been isolated from *M. acuminata* 'Grande Naine' (AAA) and characterized by Peumans *et al.* (2002), where its behaves as fruit-specific vegetative storage protein that accumulates during the early stage of the fruit formation and serves as a source of amino acids for the synthesis of ripening-associated proteins. Differential screening of ripening-related cDNA libraries constructed from the same *Musa* cultivars showed that class III chitinase together with other genes such as starch synthases and jacalin-related lectin, are down-regulated in banana pulp at different ripening stages (Clendennen and May, 1997). Peumans *et al.* (2002) found large quantities of catalytically inactive homologues of class III chitinase that are accumulated in the pulp and peel tissue of the young developing fruits, which is degraded during subsequent ripening. Here, the *Pseudoviridae* element insertion into the 5' upstream of the class III chitinase gene is observed in the same *Musa* cultivar (*M. acuminata* 'Grande Naine', AAA) that was used by Clendennen and May (1997), Medina-Suarez *et al.* (1997) and Peuman *et al.* (2002), suggesting that the inserted *Pseudoviridae* element might play an important role in the expression pattern of the class III chitinase gene during the fruit ripening. In *Drosophila melanogaster*, the LTR retrotransposon *Antonia* that inserted into the intron region of the *Chitinase 3* (*Cht3*) gene is fixed throughout 12 *D. melanogaster* populations and in a *D.*

mauritiana strain and the author suggested that this inserted sequence is of adaptive evolutionary significance for the expression of the heterochromatic genes (McCollum *et al.*, 2002). Many retrotransposons have been shown to preferentially insert within the euchromatic region in plant species, where most of the genes are found (Hirochika *et al.*, 1996a, b; Bhattacharyya *et al.*, 1997; Garber *et al.*, 1999; Parniske and Jones, 1999; Miyao *et al.*, 2003; Witte *et al.*, 2005) and some are reported to alter the expression of the genes in which they are inserted (Marillonnet and Wessler, 1997; Sunako *et al.*, 1999; Nishimura *et al.*, 2000; Selinger and Chandler, 2001; Yao *et al.*, 2001; Zufall and Rausher, 2003; Clark *et al.*, 2006; Lu *et al.*, 2006).

Table 5.1 PCR primers used in Chapter 5.

Primer Name	Sequence (5' to 3')	T_m (°c)	Binding site of the primers	PCR product size	Use of the primers
GyLTRFor	tgtcacggacttagctgg	60	*Monkey* LTR	-	IRAP in Figure 5.3
GyLTRev	cttaggcaaaaccagctaagtccg	67.4	*Monkey* LTR[#]	-	IRAP in Figure 5.3, 5.4, 5.6
BGLFor	gtttatccgtgacattctccc	62.1	5' LTR and PBS of MuLARD elements	4 kbp	IRAP in Figure 5.3
BGLREv	ggcaaaactagctaagtccgt	61.2	3' LTR of MuLARD element		IRAP in Figure 5.3
MusaTY2R	gcatgtcgtcnacatanarc	46	*Pseudoviridae* RT		REMAP in Figure 5.3
BT_GAC	gagagagagagagagagac	48.5	Microsatellite, GA repeat		REMAP in Figure 5.3
BFor	agggttcgaagtataggttcgg	64	*Copia*-IRAP B specific band	400 bp	IRAP in Figure 5.5
BRev	aatgtttaagtagagggcaagag	63.2	*Copia*-IRAP B specific band		IRAP in Figure 5.5
BGyFor	gtggccgatagcactgtgcca	72.6	*gypsy*-IRAP B specific band	275 bp	IRAP in Figure 5.10
BGyRev	agttggtgcaaggtgcttcgc	70.6	*gyspy*-IRAP B specific band		IRAP in Figure 5.10
LTRF01	yrctgaattgcttgtaaagc	57.1	*Monkey* LTR	1110 bp	MRIS in Figure 5.12
MRISF	acattgtggtrtcaragcggg	65.6	PBS of *Monkey*		MRIS in Figure 5.12
MTY13LTRF	tgagattcaacaatattaggagttgg	62.9	MuPseudo2 LTR	245 bp	MRIS in Figure 5.14
MTY1RISR02	agtgttagaacccttgcag	56.8	3' Flanking region of MuPseudo2		MRIS in Figure 5.14
MRIS_ChiF	tctcggttggttccgacagc	69.8	MuPseudo6 LTR	812 bp	MRIS in Figure 5.16
MRIS_ChiR	tctctggagtttggcccatgc	70.6	Chitinase coding region		MRIS in Figure 5.16

* IRAP = Inter-Retrotransposon Amplified Polymorphism; REMAP = Retrotransposon Microsatellite Amplified Polymorphism; MRIS = *Musa* Retrotransposon Insertion Screening [#] sequence published in Nair and Teo *et al.* (2005)

Table 5.2: *Pseudoviridae* and *Metaviridae* Group I element copy number determination using microarray hybridization.

Pseudoviridae	Gala (BB)	*Malaccensis* (AA)	A/B ratio	Subgroup
AN15	2450	5640	2.30	I
AN18	2480	5630	2.27	I
As5	2570	5880	2.28	I
As11	2310	4720	2.04	I
B2	2300	5400	2.35	I
B3	2640	5760	2.18	I
B5	2670	5570	2.08	I
B10	2790	5480	1.96	I
M5	2320	5390	2.32	I
M6	2530	4940	1.95	I
O4	3160	4070	1.29	I
W3	3020	6820	2.26	I
W8	2410	5530	2.29	I
AN7	2430	6060	2.49	II
A4	2700	6350	2.35	II
1B	2860	5450	1.90	II
3B	2880	5440	1.89	II
L5	3260	5250	1.61	II
L7	3460	6660	1.93	II
M1	2700	5250	1.94	II
M3	3000	6340	2.11	II
M7	2980	6010	2.02	II
W1	3910	7120	1.82	II
W7	2360	5400	2.28	II
mean	2760	5670	2.08	

Control	Gala (BB)	*Malaccensis* (AA)	A/B ratio
pMa794.1	2700	5690	2.11

Metaviridae	Gala (BB)	*Malaccensis* (AA)	A/B ratio	Subgroup
AbTY3-5	3450	5710	1.66	II
AbTY3-8	4800	11850	2.47	II
AsTY3-2	3370	8130	2.41	II
AsTY3-8	4680	4310	0.92	II
AsTY3-9	5800	12590	2.17	II
B3 TY3	2490	6010	2.42	II
B9 TY3	2710	4330	1.60	II
B5-1	3100	4290	1.38	II
B5-3	3660	5200	1.42	II
B9-1	3710	10160	2.74	II
SY-30A	3030	12370	4.08	II
AsTY3-6	3110	3940	1.26	III
AsTY3-12	2460	6250	2.54	III
B1 TY3	2210	4510	2.04	III
B3-1	3280	4720	1.44	III
B4-14-3	2460	3780	1.54	III
W1 TY3	2660	3060	1.15	III
W14a TY3	2770	7570	2.73	III
W15 TY3	2280	5270	2.31	III
W17 TY3	2350	5160	2.20	III
W18 TY3	2310	6990	3.03	III
AbTY3-3	5260	12090	2.30	IV
AbTY3-11	4030	5960	1.48	IV
AbTY3-14	5060	17090	3.38	IV
AsTY3-10	2820	8370	2.96	IV
B3-2	3230	6620	2.05	IV
AsTY3-3	2880	5370	1.86	V
AsTY3-5	3000	4870	1.62	V
B9-4	2740	6250	2.28	V
mean	3300	6990	2.12	

Table 5.3: *Metaviridae* Group II and III element copy number based band intensity observed after semi-quantitative PCR.

LTR retrotransposon, domain	Plasmid standard, size	'Pisang Gala' (BB)	*malaccensis* (AA)	A/B ratio
Metaviridae Group III element, RT gene	MA_RT1, 4.11 kbp	530	570	1.07
Metaviridae Group III element, LTR	MbTy32a1, 6.446 kbp	NT	1810	-
Metaviridae Group II element, RT gene	MalcoTR_RT1, 4.345 kbp	250	2700	10.72

*NT = not tested

Table 5.4: Description of the identified sequences flanking the *Metaviridae* Group III LTR retrotransposons.

Clones	Length (bp)	Primer used	*Musa* taxa	Description	Position in clones
B-LTR7.1	432	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 131
					404 – 432
				Unknown sequence	132 – 403
B-LTR6.1	432	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 131
					404 – 432
				Unknown sequence	132 – 403
B-LTR5.1	485	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 133
					456 - 485
				Unknown sequence	134 – 456
B-LTR3.1	485	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 133
					456 – 485
				Unknown sequence	134 – 456
B-LTR3.8	681	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 29
					653 – 681
				Monkey flanking sequence	30 – 541
				Unknown sequence	542 – 652
B-LTR3.4	681	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 29
					653 – 681
				Monkey flanking sequence	30 – 541
				Unknown sequence	542 – 652
B-LTR2.1	831	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 29
					803 – 831
				Metaviridae Group III element internal sequence	30 – 802
B-LTR4.1	565	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 29
					537 – 565
				MuLARD-like internal sequence	30 – 536
B-LTR1.1	957	GyLTRev	*M. acuminata* 'Pisang Berangan'	*Monkey* LTR	1 – 29
					898 – 957
				MuLARD-like internal sequence	30 – 897

M1B14.3	1209	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR Unknown sequence	1 – 29 1181 – 1209 30 – 1180
M1B14.2	830	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR *Pseudoviridae* LTR & RNaseH	1 – 29 802 – 830 30 – 801
M1B9.2	832	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR *Pseudoviridae* LTR & RNaseH	1 – 29 804 – 832 30 – 803
M1B9.3	833	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR *Pseudoviridae* LTR & RNaseH	1 – 29 805 – 832 30 – 804
M1B3.1	263	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR *Metaviridae* Group III element internal sequence	1 – 131 235 – 263 132 – 234
M1B3.2	263	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR *Metaviridae* Group III element internal sequence	1 – 131 235 – 263 132 – 234
M1B3.3	263	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR *Metaviridae* Group III element internal sequence	1 – 131 235 – 263 132 – 234
M1B14.1	1283	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR MuLARD-like internal sequence	1 – 29 1255 – 1283 30 – 1254
M1B10.1	957	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR MuLARD-like internal sequence	1 – 29 898 – 957 30 – 897

M1B10.2	957	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR MuLARD-like internal sequence	1 – 29 898 – 957 30 – 897
M1B10.3	957	GyLTRev	*M. paradisiaca* 'Mutiara' (AAB)	*Monkey* LTR MuLARD-like internal sequence	1 – 29 898 – 957 30 – 897
But-LTR2.2	333	GyLTRev	*M. balbisiana* 'Butohan 2'	*Monkey* LTR Complete MuTRb and partial sequence of MuTRa	1 – 130 305 – 333 131 – 304
But-LTR3.1	332	GyLTRev	*M. balbisiana* 'Butohan 2'	*Monkey* LTR Complete MuTRb and partial sequence of MuTRa	1 – 130 304 – 332 131 – 303

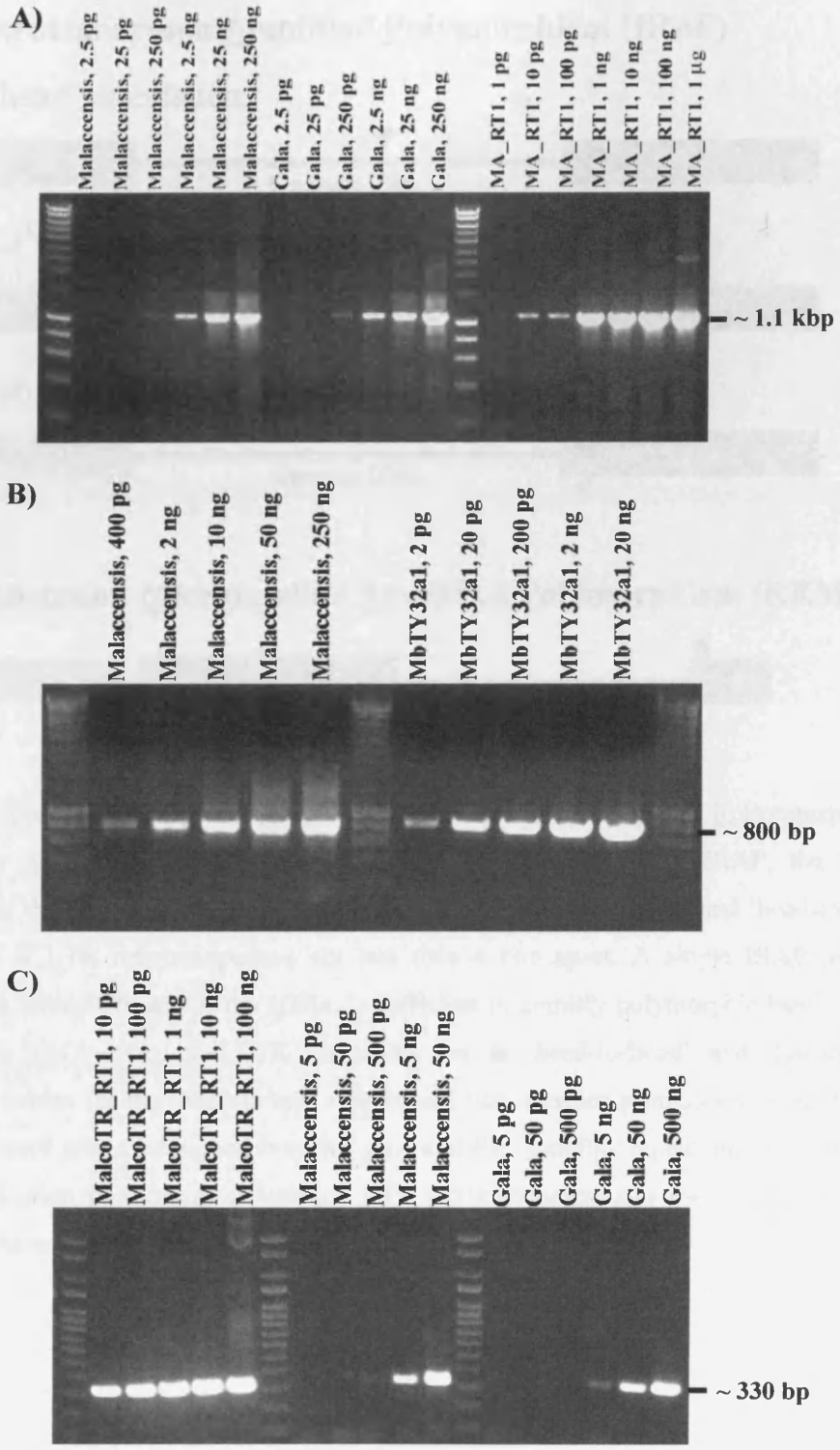


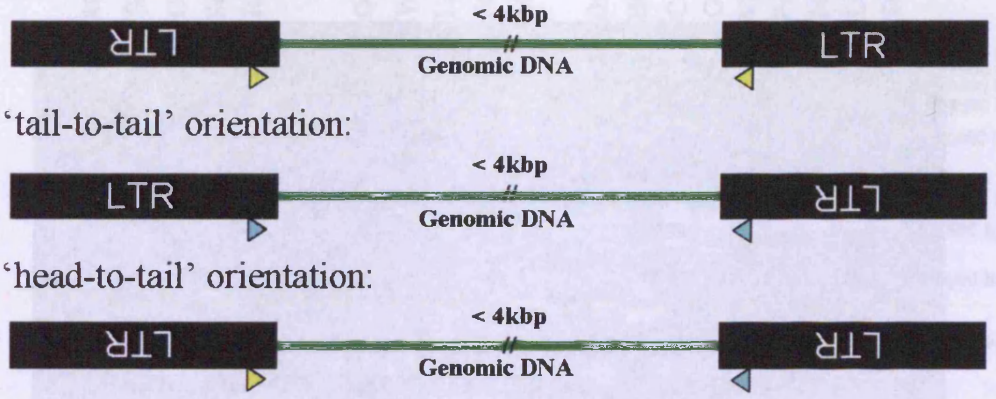
Figure 5.1: Quantification of LTR retrotransposon copy number using semi-quantitative PCR. Plasmid DNA was used as standard. **A** *Metaviridae* Group III element RT gene, plasmid used = MA_RT1, dilution = 10 fold; **B** *Metaviridae* Group III element LTR, plasmid used = MbTY32a1, dilution = 5 fold; **C** *Metaviridae* Group II element RT gene, plasmid used = MalcoTR_RT1, dilution = 10 fold. For calculation and equation, see Materials and Methods.

Inter-Retrotransposon Amplified Polymorphism (IRAP)

'head-to-head' orientation:

'tail-to-tail' orientation:

'head-to-tail' orientation:



Retrotransposon Microsatellite Amplified Polymorphism (REMAP)

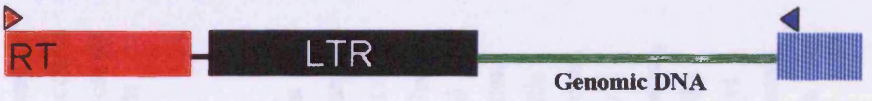


Figure 5.2: Inter-Retrotransposon Amplified Polymorphism (IRAP) and Retrotransposon Microsatellite Amplified Polymorphism (REMAP) PCR schemes. For IRAP, the three orientations of the LTR sequences are shown: 'head-to-head', 'tail-to-tail' and 'head-to-tail' are exploited if LTR retrotransposons are less than 4 kbp apart. A single IRAP primer binding at the immediate end of the LTRs, is sufficient to amplify polymorphic bands from total genomic DNA when the LTR sequences are in 'head-to-head' and 'tail-to-tail' orientation whereas for the 'head-to-tail' orientation, two outward primers are needed. For REMAP, outward primers designed from RT gene and microsatellite repeat units (green box) allow amplification of the region between LTR retrotransposon and microsatellites. The genomic DNAs are shown (green solid lines).

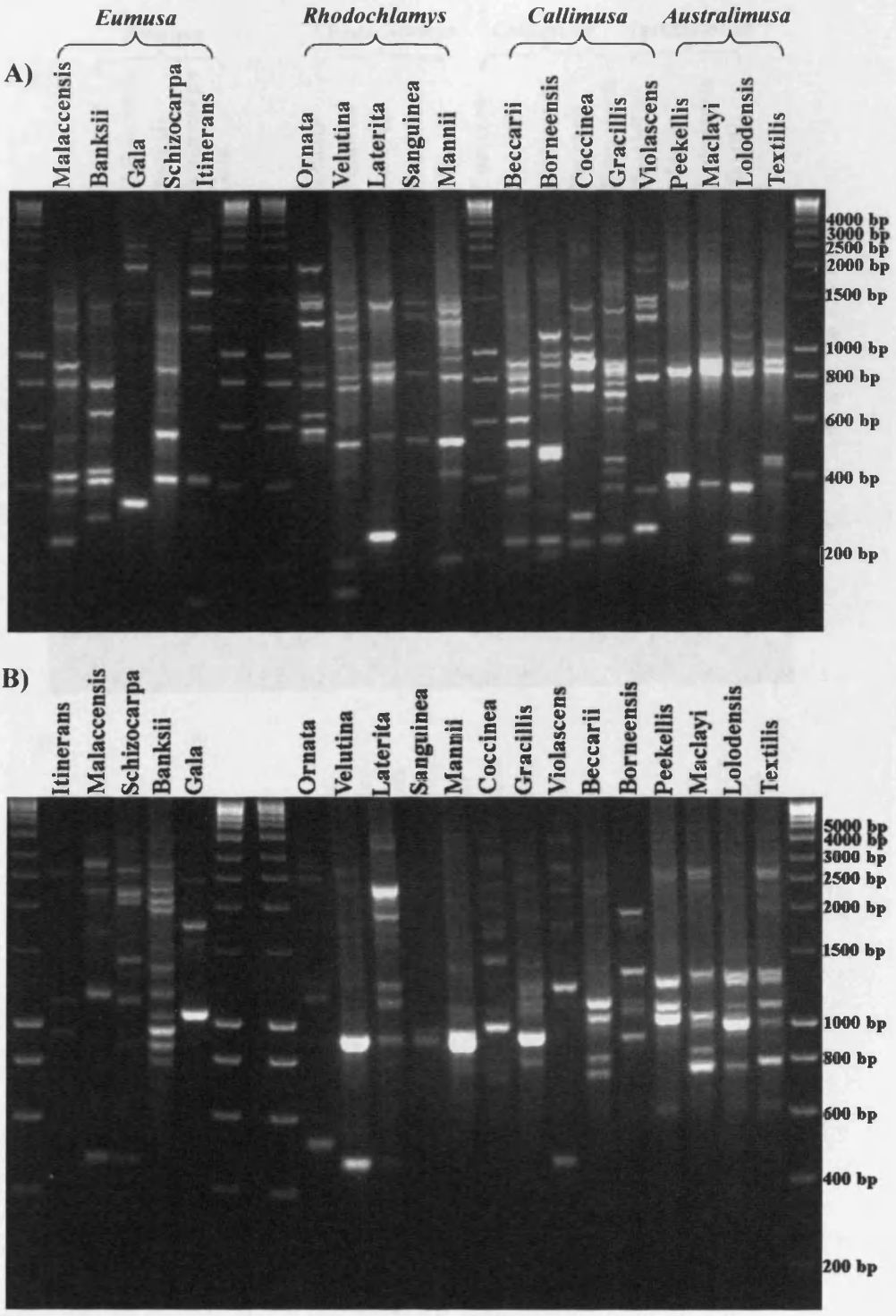


Figure 5.3: IRAP and REMAP polymorphic banding pattern. For IRAP, **A** single *gypsy*-IRAP primer, GyLTRev (Nair and Teo *et al.*, 2005), **B** single primer GyLTRFor. For *Musa* species description see table 2.1 and for primer sequences, see table 5.1. Percentage of agarose gel used = 2%.

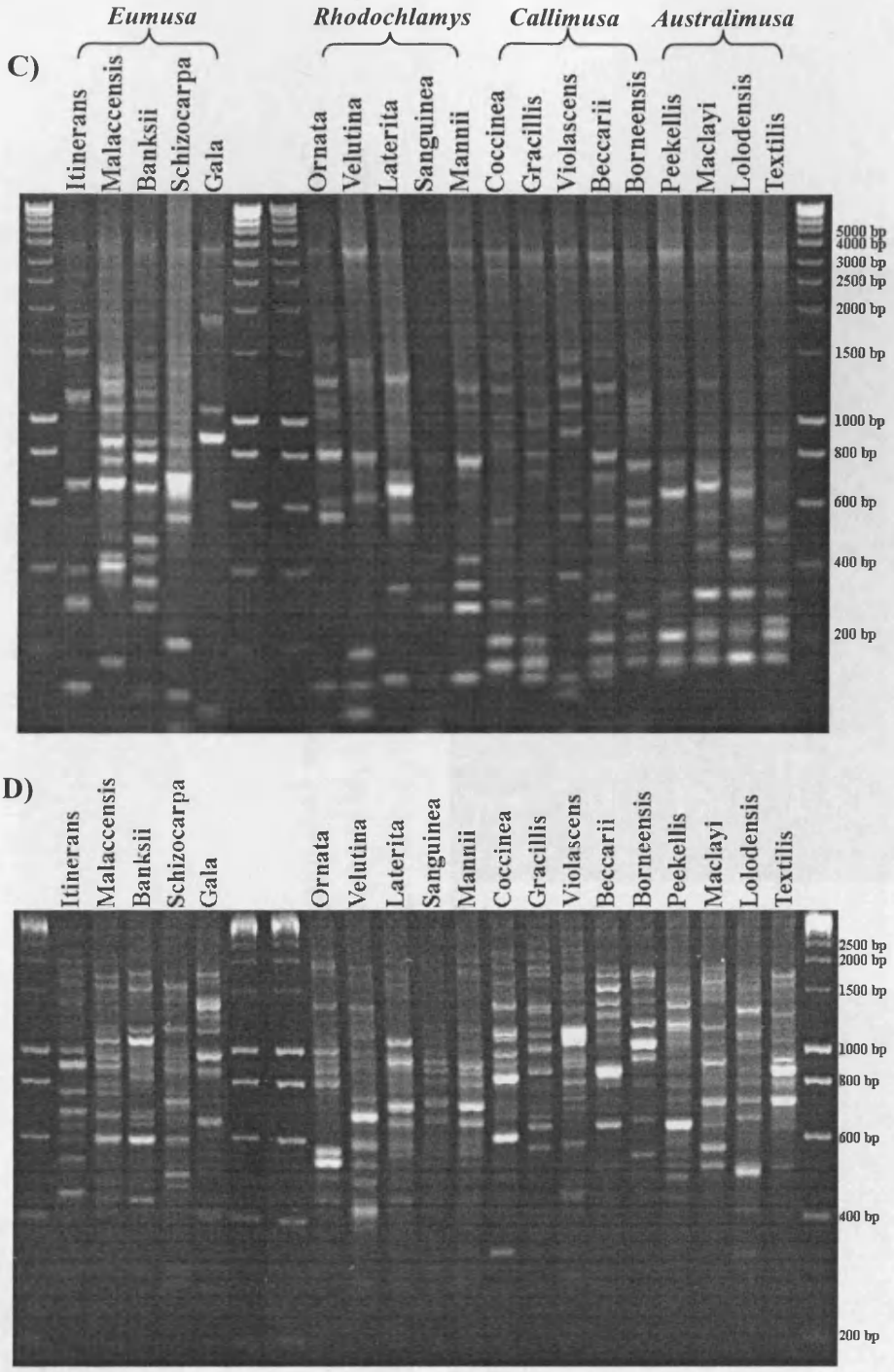


Figure 5.3 continued: IRAP and REMAP polymorphic banding pattern. For IRAP, **C** BGLFor & BGLRev whereas for REMAP, **D** MusaTY2R & BT_GAC were used. For *Musa* species description see table 2.1 and for primer sequences, see table 5.1. Percentage of agarose gel used = 2%.

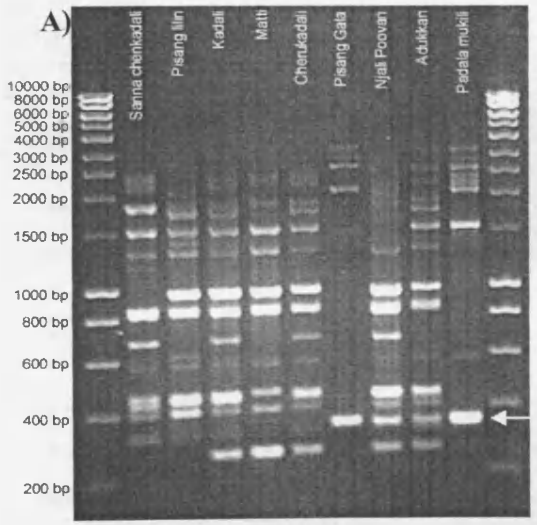
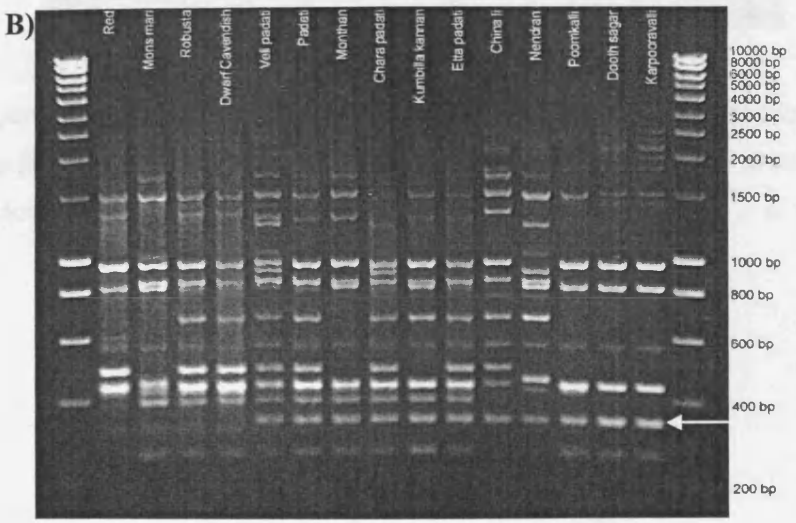
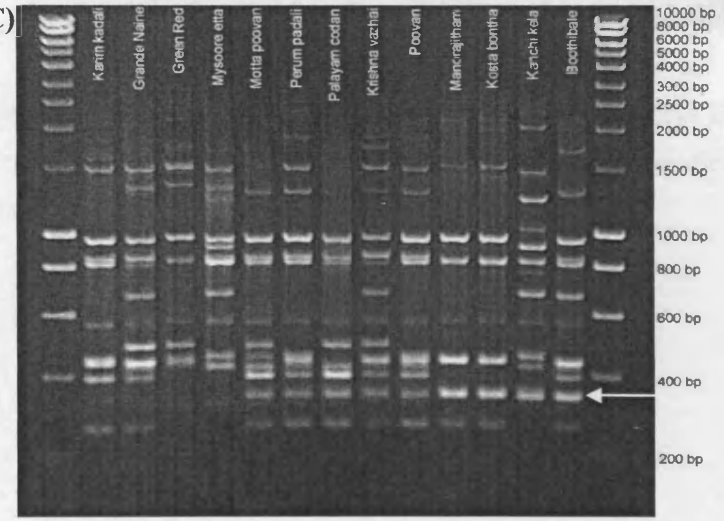


Figure 5.4: PCR with *gypsy*-IRAP primer (GyLTRev) showing the ~350bp B specific band, $BB_{gypsy\text{-}IRAP}$ (arrows). **A** Diploid *Musa* cultivars; **B-C** Triploid *Musa* cultivars. These results are published in Nair and Teo *et al.* (2005).





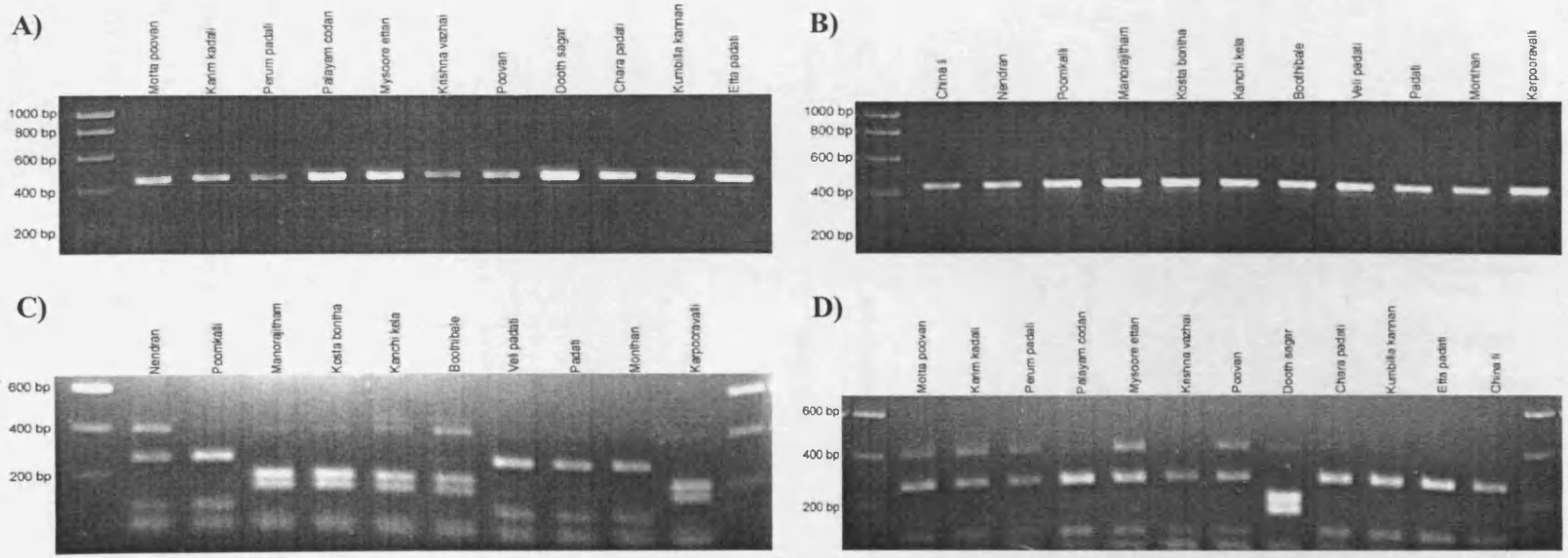


Figure 5.5: PCR amplification of *copia*-IRAP B specific band and *Alu*I enzyme digestion of the PCR products. **A-B** PCR product of *copia*-IRAP B specific band, $BB_{copia\text{-}IRAP}$, using internal primers BFor and BRev; **C-D** *AluI* restriction enzyme polymorphism in the *copia*-IRAP B specific bands of AAB/ABB genomes. These results are published in Nair and Teo *et al.* (2005) except **Figure 5.5A** and **B**.

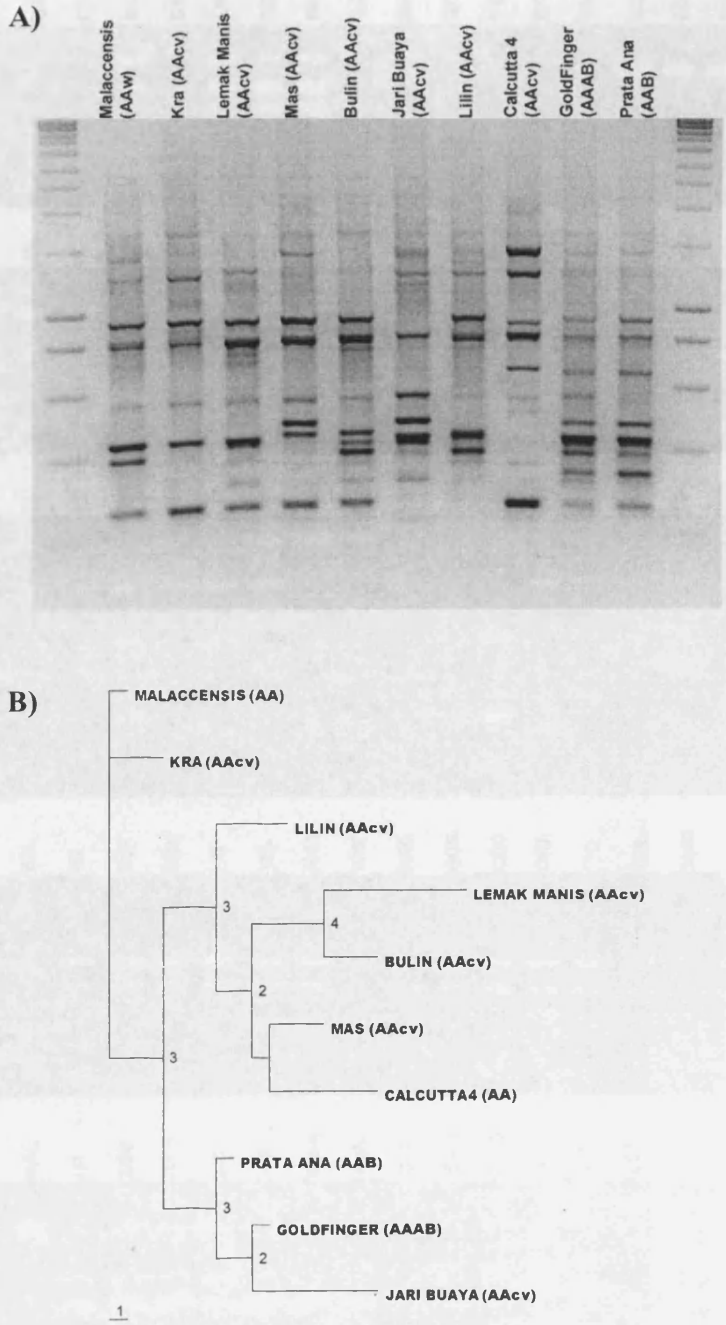


Figure 5.6: IRAP banding pattern on closely related *Musa* cultivars and NJ dendrogram of scored bands. **A** DNA polymorphism pattern generated by *gypsy*-IRAP primer, GyLTRev (Nair and Teo *et al.*, 2005) in selected cultivars. **B** Dendrogram based on scored IRAP bands.

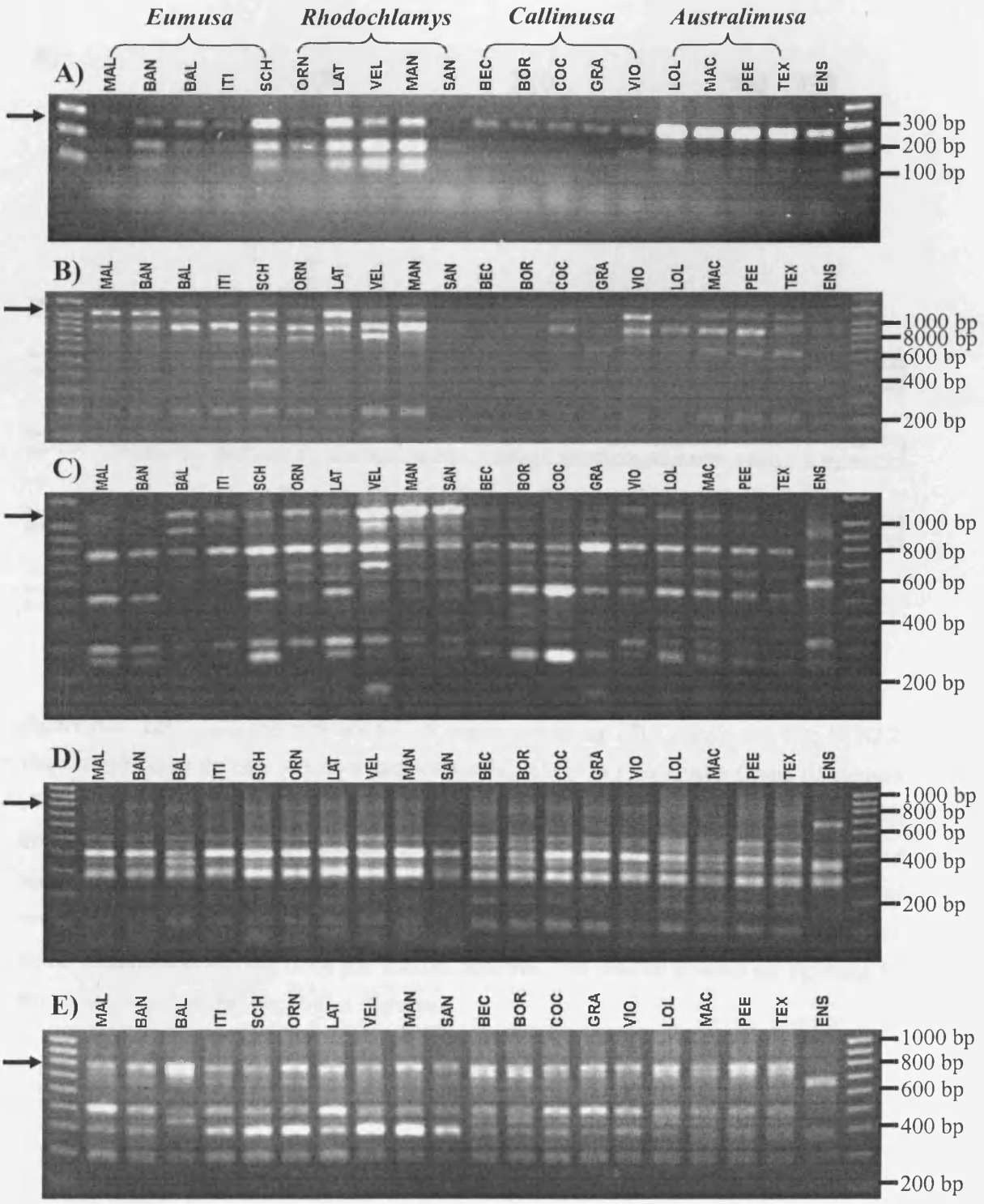


Figure 5.7: PCR-RFLP analysis of LTR retrotransposons. **A** *pseudoviridae* RT domain, *Hind*III; **B** *metaviridae* RT domain, *ScrF*I; **C** *metaviridae* RT domain, *Rsa*I; **D** *metaviridae* LTR, *ScrF*I and **E** *metaviridae* LTR, *Sma*I. Arrows show the uncut bands in each panels. *Musa* taxa and abbreviations given in Table 2.1.

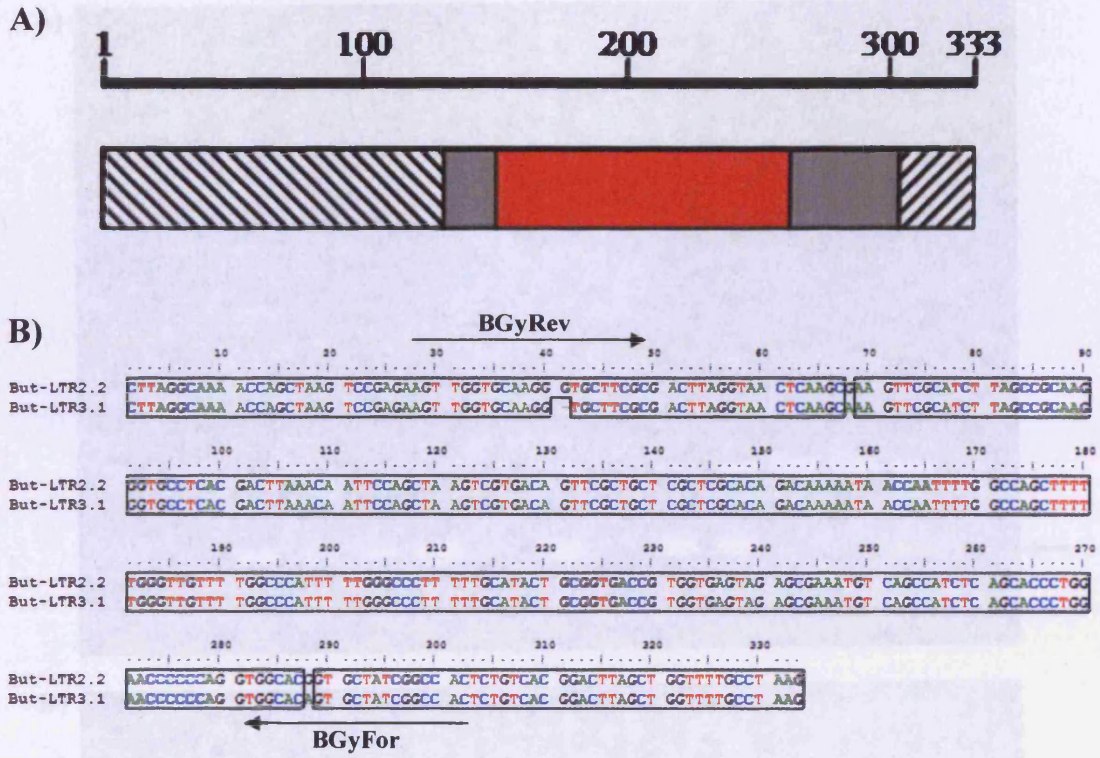


Figure 5.8: $BB_{gypsy\text{-}IRAP}$ sequence analysis. **A** Annotation of the $BB_{gypsy\text{-}IRAP}$ clone, But_LTR2.2. The slashed boxes indicate sequence that showed homology to *Metaviridae* Group III element LTR region. The red box represents sequence that showed homology to MuTRb repeat whereas the gray boxes represent sequence that showed homology to part of the MuTRa repeat. For description of the MuTRa and MuTRb, see MuTR section in Chapter 4. **B** Pairwise alignment of two $BB_{gypsy\text{-}IRAP}$ clones, But_LTR2.2 and But_LTR3.1. Homologous regions are boxed. Two A↔G transitions at position nt 68 and 288 are observed. The internal primers are indicated by arrows above and below the pairwise alignment.

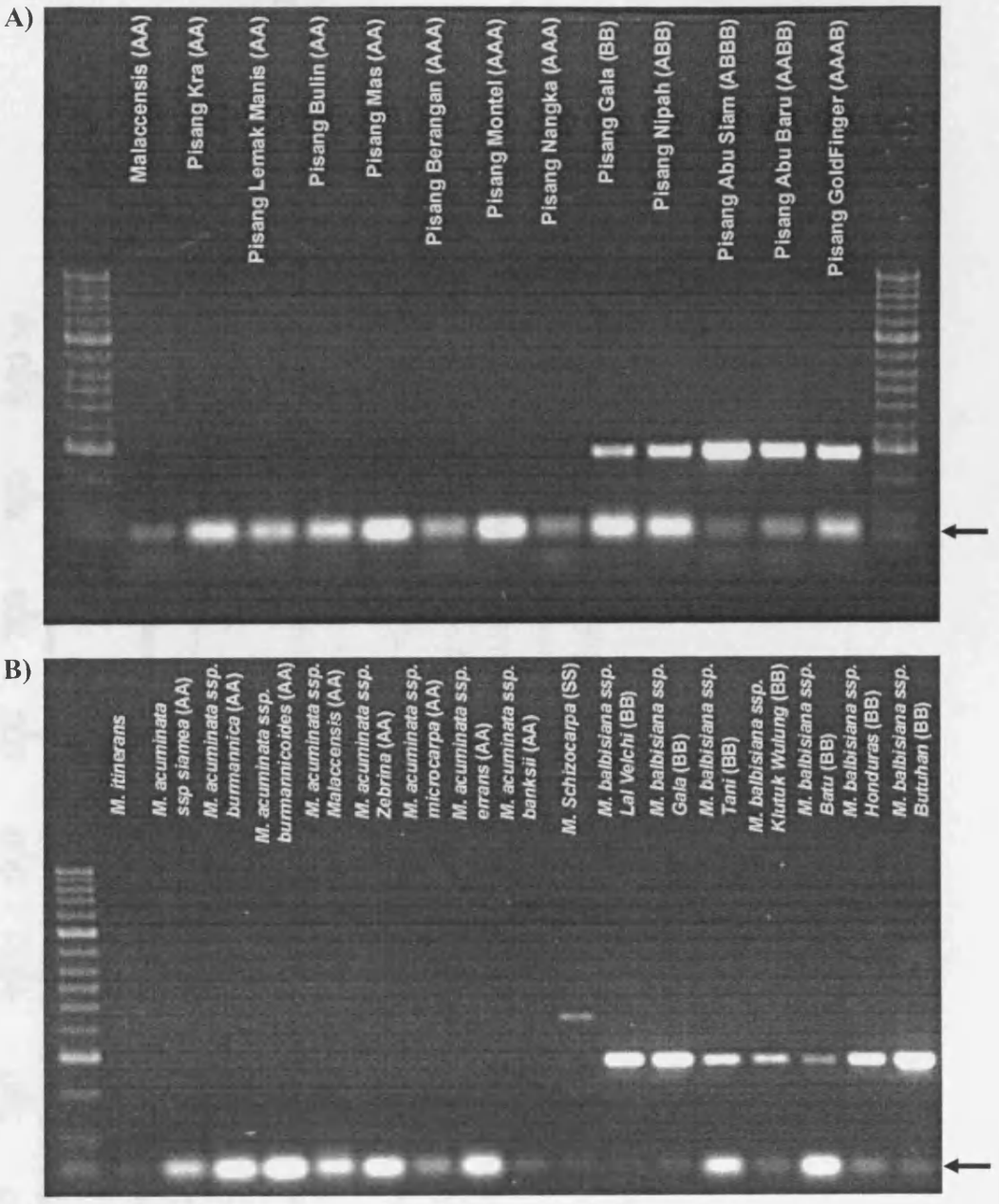


Figure 5.9: Amplification of $BB_{gypsy\text{-}IRAP}$ bands from different *Musa* cultivars (**A**) and species (**B**) using internal primers. BGyFor and BGyRev, designed from clone But_LTR2.1 (Figure 5.9; Table 5.1). Arrows show the PCR primer dimmers that are particularly formed when no primer binding sites are present in the template DNA.

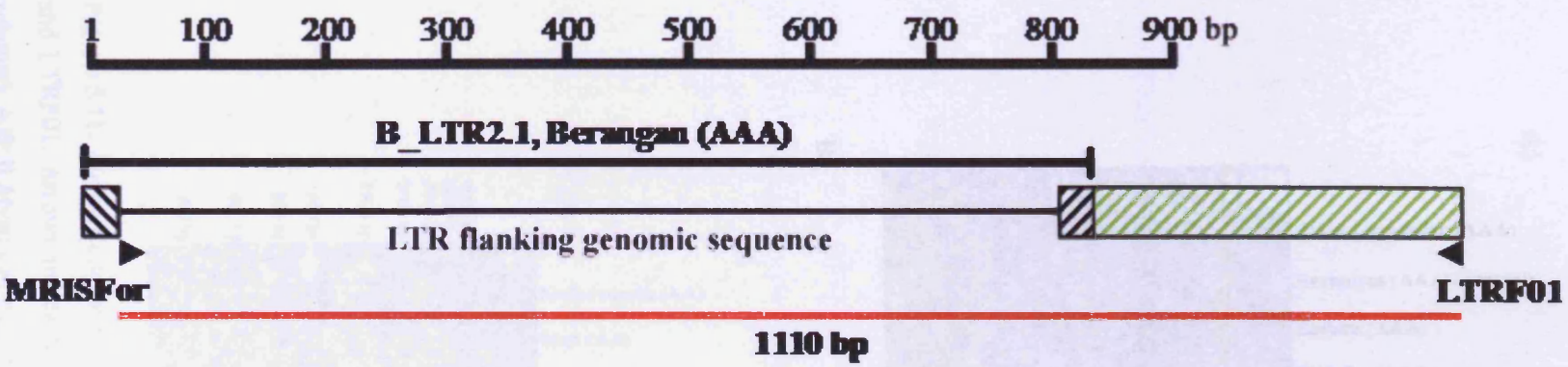


Figure 5.10: PCR scheme for the *Metaviridae* Group III element insertion study. The black slashed boxes indicate two outward facing LTR sequences in IRAP clone B_LTR2.1 (Table 5.4) that contained about 800 bp flanking sequence with homology to *Metaviridae* Group III element internal domain (the solid line). Additional LTR sequence (green slashed box) was added to the end of B_LTR2.1 for better understanding. Primers, MRISFor and LTRF01 (sequence given in Table 5.1), amplify a 1110 bp band if there is a *Metaviridae* Group III element inserted into another *Metaviridae* Group III element. Amplification products of larger or smaller size are expected if the LTR flanking sequences are of variable length but still contain the MRISFor primer sequence.

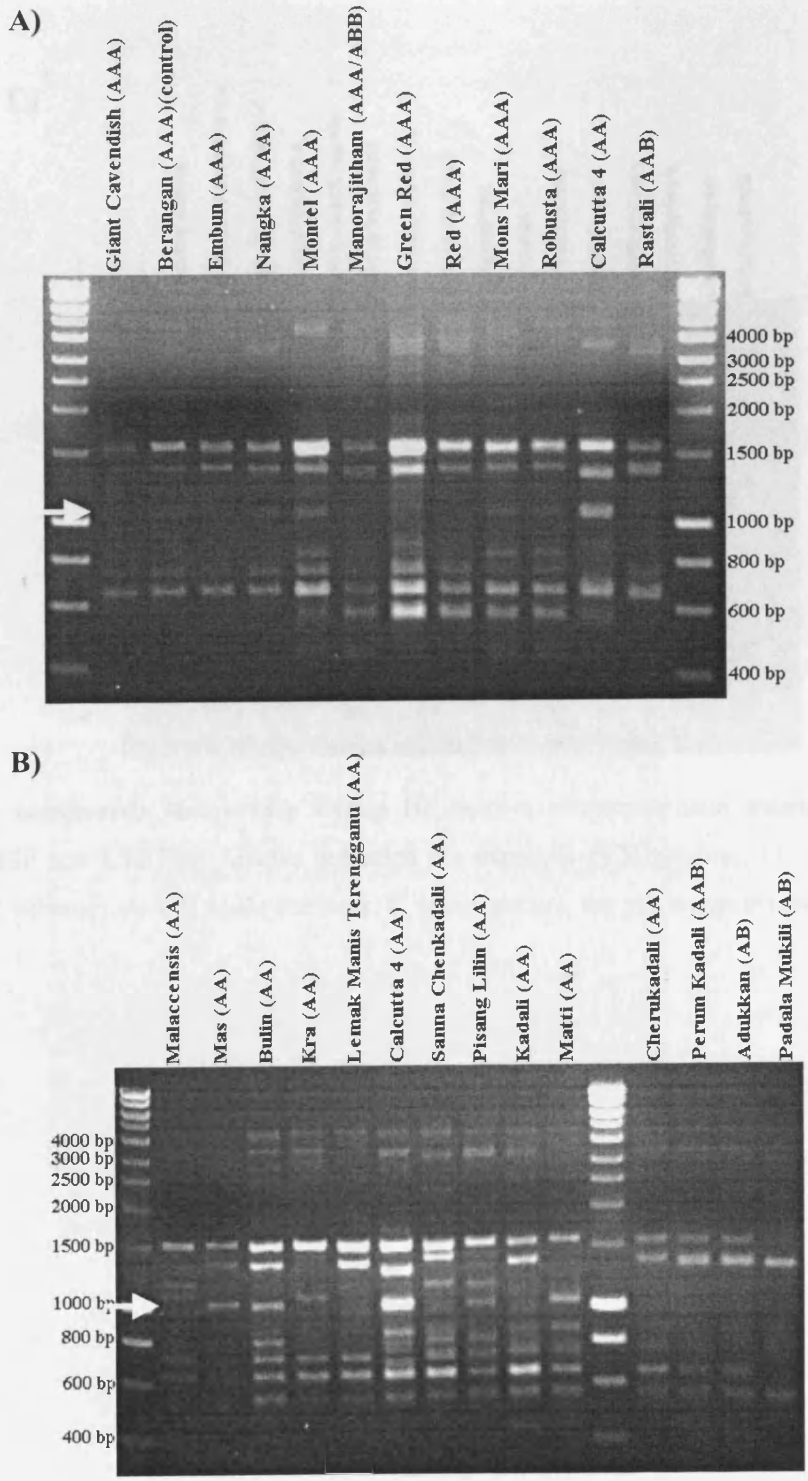


Figure 5.11: *Metaviridae* Group III element retrotransposon insertion study using primers MRISF and LTRF01. Arrows indicated the expected PCR product, 1110 bp (see Figure 5.10 for PCR scheme). **A** & **B** *Musa* cultivars, **C** *Musa* species, the gel image is inversed.

C)

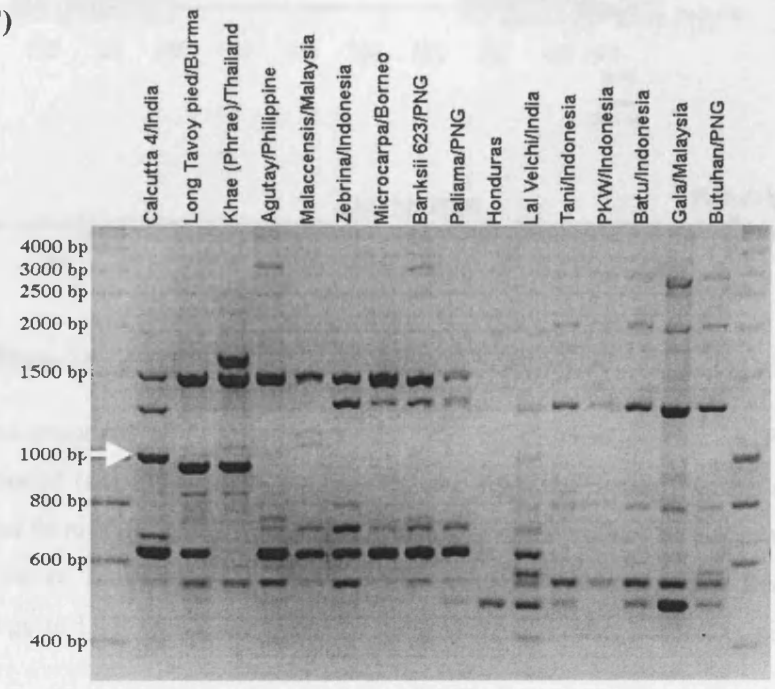


Figure 5.11 continued: *Metaviridae* Group III element retrotransposon insertion study using primers MRISF and LTRF01. Arrows indicated the expected PCR product, 1110 bp (see Figure 5.10 for PCR scheme). **A** & **B** *Musa* cultivars, **C** *Musa* species, the gel image is inversed.

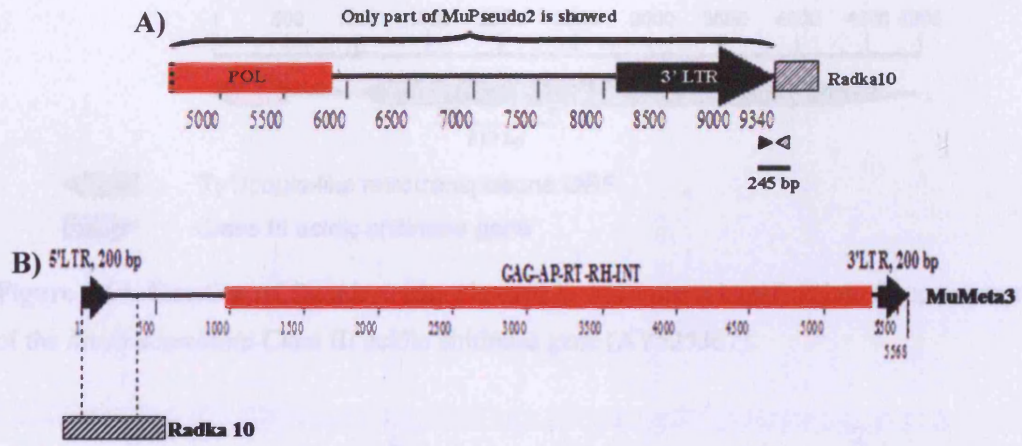


Figure 5.12: Insertion of LTR retrotransposons into other *Musa* repeats. **A** *Pseudoviridae* element MuPseudo2 (see Figure 3.1A) is inserted into Radka10 repeat (Valárik *et al.*, 2002). Primers designed from MuPseudo2 3'LTR and Radka10 repeats produce a PCR band of size 245 bp. **B** Annotation of MuMeta3 (see Figure 3.1B) to Radka 10. Part of the Radka10 repeat showed homology to LTR and PBS regions of the MuMeta3. Primer sequences given in Table 5.1.

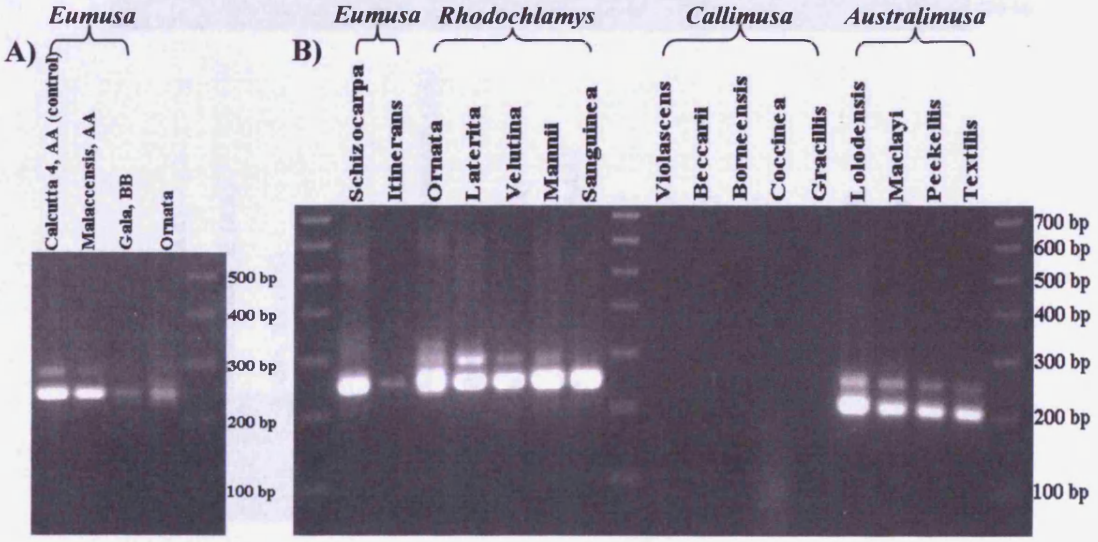


Figure 5.13: *Pseudoviridae* element insertion study using outward LTR primer, MTY13LTRF and Radka10 primer, MTY1RISR02. The expect size of the PCR amplification if there was a *Pseudoviridae* element insertion is 245 bp.

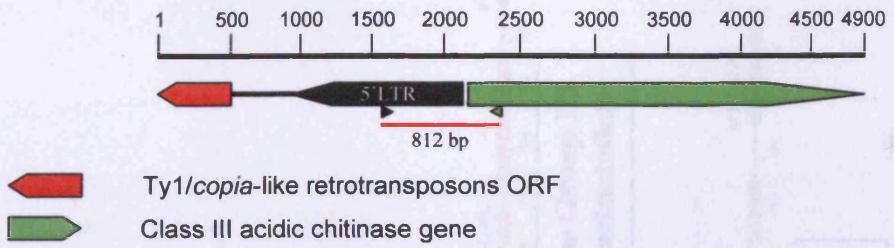


Figure 5.14: Insertion of *Pseudoviridae* element in antisense orientation into 5' upstream region of the *Musa acuminata* Class III acidic chitinase gene (AY525367).

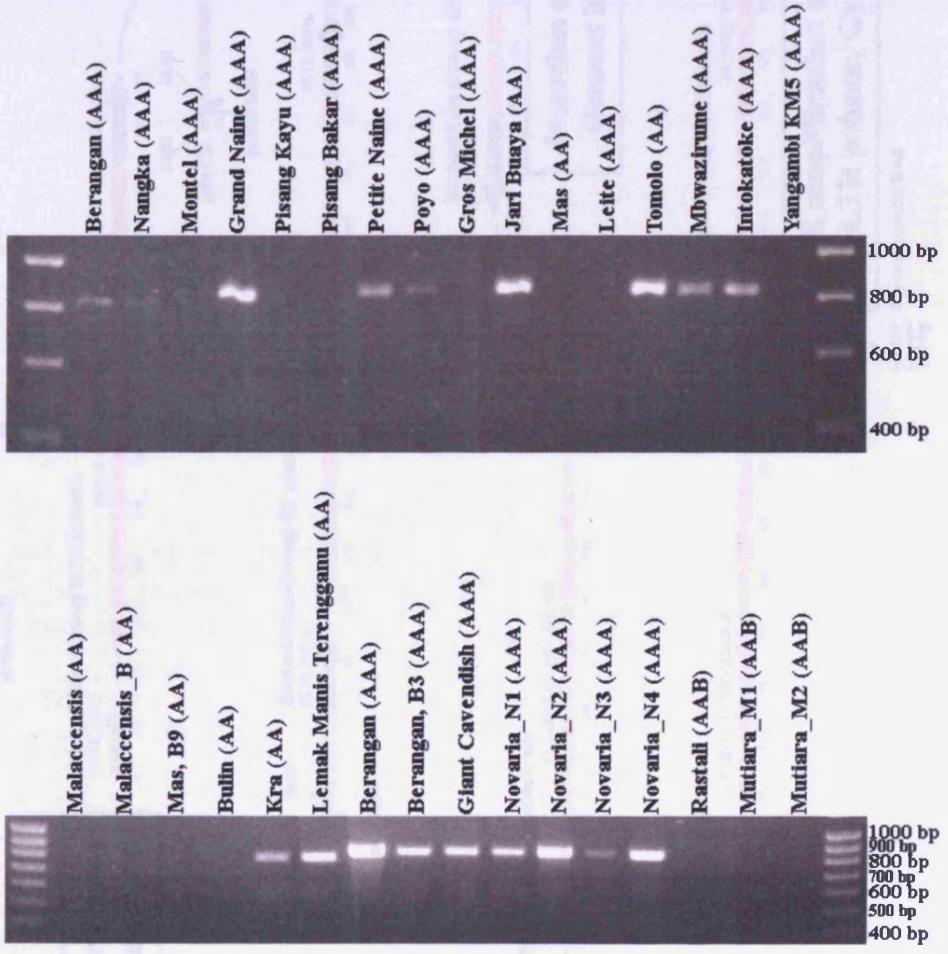


Figure 5.15: PCR screening of the *Pseudoviridae* element insertion into 5' upstream of the chitinase III gene. Primer targeted *Pseudoviridae* element sequence, MRIS_ChiF and chitinase III coding region, MRIS_ChiR, were used to amplify a 812 bp band if there is a *Pseudoviridae* element inserted into 5' upstream of the chitinase III gene.

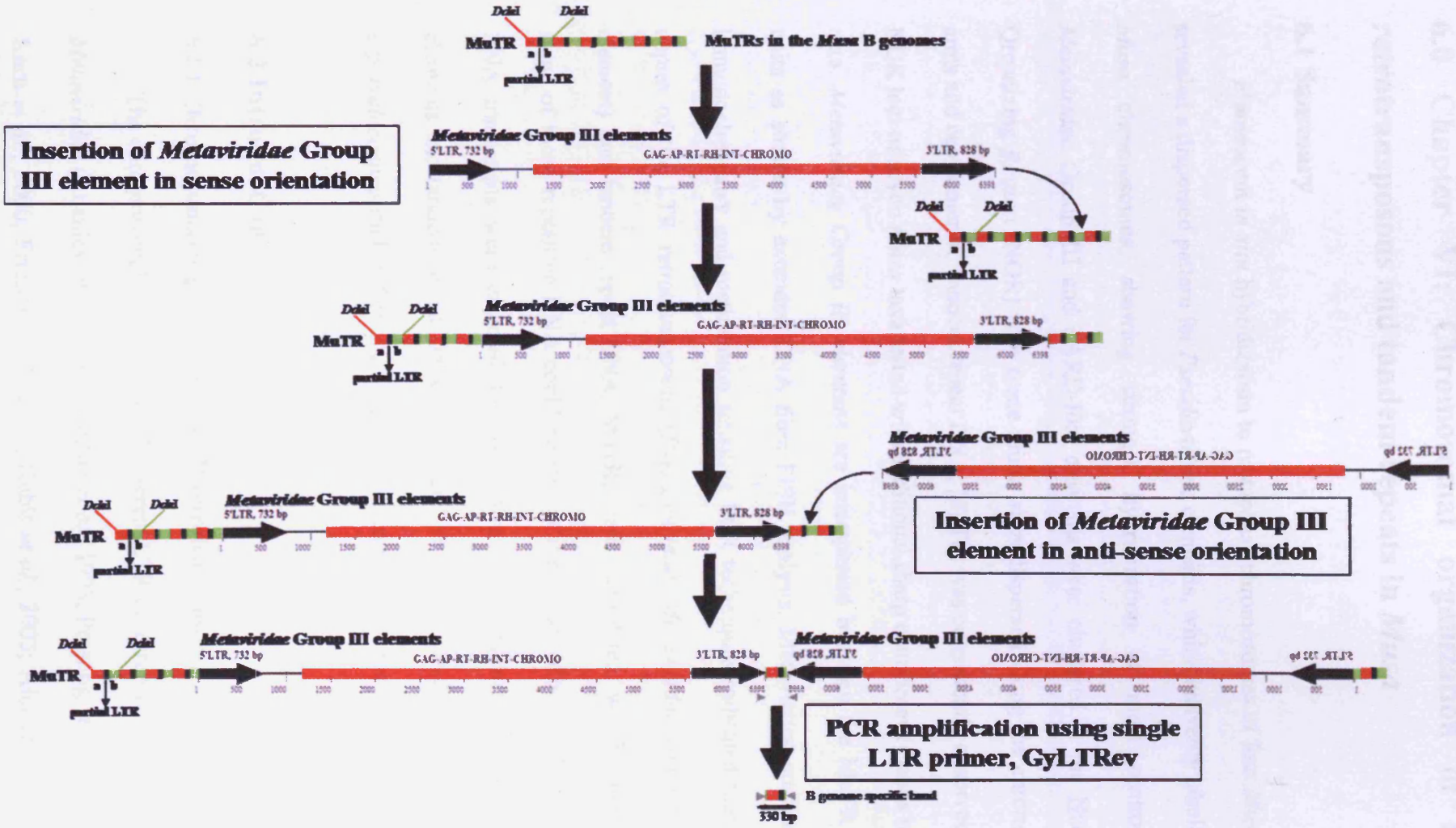


Figure 5.16: The *Metaviridae* Group III element insertion into MuTR repeats created the *gypsy*-IRAP B specific band (330 bp).

6.0 Chapter VI: Chromosomal organization of LTR retrotransposons and tandem repeats in *Musa*

6.1 Summary

Fluorescent *in situ* hybridization to metaphase chromosomes of four *Musa* taxa revealed a dispersed pattern for *Pseudoviridae* elements, which unevenly labelled the *Musa* chromosomes, showing stronger hybridization at most centromeres. *Metaviridae* Group III and LARD-like elements were clustered at the Nucleolar Organizing Region (NOR) with some minor sites dispersed along the chromosome arms and centromeres. Tandem repeat DNA, MuTR, was consistently observed at the NOR loci of seven *Musa* taxa tested with additional sharp centromeric bands in some taxa. *Metaviridae* Group III elements are interspersed between the MuTR repeat units as shown by extended DNA fibre FISH analysis. Methylation studies using immunochemistry and methylation sensitive PCR techniques indicated that not all copies of the LTR retrotransposons (*Pseudoviridae*, *Metaviridae* and LARD-like element) and tandem repeat DNA, MuTR, were methylated, which suggests that some of these repetitive DNA could be still actively transcribed in *Musa*. Indeed, RNA transcripts were detected for some *Metaviridae* Group II and Group III elements and tandem repeat DNA (CoTR and MuTR) isolated from *M. acuminata* ssp. *malaccensis* male inflorescences, and showed high sequence heterogeneity.

6.2 Introduction

6.2.1 Chromosomal organization of LTR retrotransposons

The chromosomal organization of different LTR retrotransposons, mainly from *Metaviridae* (Ananiev *et al.*, 1998b; Miller *et al.*, 1998; Presting *et al.*, 1998; Balint-Kurti *et al.*, 2000; Friesen *et al.*, 2001; Kubis *et al.*, 2003; Alix *et al.*, 2005) and

Pseudoviridae (Katsiotis *et al.*, 1996; Pearce *et al.*, 1996a, b; Brandes *et al.*, 1997; Galasso *et al.*, 1997; Kumar *et al.*, 1997; Pearce *et al.*, 1997; Kuipers *et al.*, 1998; Castilho *et al.*, 2000; Frello and Heslop-Harrison, 2000; Friesen *et al.*, 2001; Vershinin *et al.*, 2002; Alix *et al.*, 2005) families, have been studied in detail in many different plant species using Fluorescent *In Situ* Hybridization (FISH) techniques on metaphase chromosomes.

In situ hybridization data showed that *Pseudoviridae* elements displayed more dispersed distribution patterns throughout the euchromatic and heterochromatic regions of the host chromosomes and sometime co-localized with the 5S rRNA gene loci (Pearce *et al.*, 1996a, b; Brandes *et al.*, 1997; Heslop-Harrison *et al.*, 1997; Castilho *et al.*, 2000; Alix *et al.*, 2005). A different chromosomal organization was observed for *Metaviridae* elements, which had a much more restricted distribution patterns depending on the plant species studied. In certain plant species, *Metaviridae* elements were found mainly in the centromeric regions (Ananiev *et al.*, 1998b; Miller *et al.*, 1998; Presting *et al.*, 1998), whereas in other species, these elements were observed in heterochromatic regions such as Nucleolar Organizing Regions (NOR) with some elements dispersed along the chromosome arms (Balint-Kurti *et al.*, 2000). Furthermore, *Metaviridae* elements that clustered at the sub-terminal regions of the chromosomes were reported by Kubis *et al.* (2003) in oil palm, *Elaeis guineensis*.

6.2.2 DNA methylation and epigenetic silencing of transposable elements

DNA methylation has been shown to play an important role in epigenetic regulation in many organisms (Bird, 2002). Cytosines in the symmetric CpG and CpNpG contexts are methylated in most eukaryotes (Li, 2002) and the methylation is

inherited epigenetically following DNA replication by DNA methyltransferase (Okano *et al.*, 1998; Hsieh, 1999; Lyko *et al.*, 1999; Okano *et al.*, 1999; Martienssen & Colot, 2001). DNA methylation has been reported to affect the expression of many developmental genes (Kakutani, 1996; Kakutani *et al.*, 1997; Soppe *et al.*, 2000; Kankel *et al.*, 2003; Saze *et al.*, 2003). Methylated cytosines may recruit specific chromatin components and initiate heterochromatin formation by histone deacetylation and methylation. On the other hand, modified histones may recruit DNA methyltransferases and mediate cytosine methylation (Tamaru & Selker, 2001; Aufsatz *et al.*, 2004; Matzke & Birchler, 2005). In animals, cytosine methylation is prevalent in symmetrical CpG dinucleotides whereas in plants, it is more often observed in symmetrical CpNpG contexts as well as in non-symmetrical CpHpH (H=C, A, T) (Gruenbaum *et al.*, 1981; Meyer *et al.*, 1994; Grafi *et al.*, 2007) and most methylated cytosines are found in heterochromatic regions enriched with repetitive DNAs such as rDNAs, tandem repeat DNAs and transposable elements (Bender, 2004, Grafi *et al.*, 2007).

Epigenetic silencing of transposable elements by DNA methylation has been studied in different organisms (Lyko *et al.*, 2000; Burgers *et al.*, 2002; Kato *et al.*, 2004; Verdel *et al.*, 2004). Additional to the DNA methylation, transposable elements also can be epigenetically controlled by histone modifications such as deacetylation, histone H3 Lysine-9 methylation (H3mK9) (Rea *et al.*, 2000; Gendrel *et al.*, 2002; Johnson *et al.*, 2002; Schotta *et al.*, 2002; Fuchs *et al.*, 2006), transcriptional and post-transcriptional gene silencing pathways such as RNA interference (RNAi) and RNA-directed DNA methylation (RdDM) (Aufsatz *et al.*, 2002; Mathieu and Bender, 2004; Matzke & Birchler, 2005; Nolan *et al.*, 2005; Huettel *et al.*, 2006; Huettel *et al.*, 2007) and they are all inter-related (Martienssen

& Colot, 2001; Selker, 2002; Sleutels and Barlow, 2002; Lippman *et al.*, 2003; Tran *et al.*, 2005; Huettel *et al.*, 2007).

In *Dictyostelium*, Kuhlmann *et al.* (2005) found that disruption of DnmA (a DNA methyltransferases of the Dnmt2 family) resulted in the loss of methylation and an increase of transcription and transposition of the retrotransposon Skipper, whereas the expression of retrotransposon DIRS-1 was not affected by loss of DnmA. In *Neurospora crassa*, *de novo* cytosine methylation of a LINE-like retrotransposon, *Tad*, inhibits both its expression and transposition and this effect can be reversed by the use of 5-azacytidine (Zhou *et al.*, 2001).

In mammals, the distribution of 5-methylcytosine undergoes radical changes during early development (Monk *et al.*, 1987), but it is prominent in repetitive sequences and absent from CpG islands (Yoder *et al.*, 1997; Bird, 2002). Methyl-CpG-binding domain (MBD) protein in mammals and plants recruit histone deacetylase (HDAC) and histone H3 lysine-9 methyltransferase (HMT) activity (Nan *et al.*, 1998a; b; Saito *et al.*, 2002; Fuks *et al.*, 2003a; b; Unoki *et al.*, 2004; Zemach *et al.*, 2005; Johnson *et al.*, 2007) and Dnmt1 interact directly with HDACs (Fuks *et al.*, 2000).

In plants, silent retrotransposons are methylated (Bennetzen *et al.*, 1994; Flavell, 1994; Martienssen, 1998) and can be reactivated in methylation-defective mutants (Miura *et al.*, 2001; Singer *et al.*, 2001). Methylation is unusual in plant genes, and is restricted to the 5' – and 3' – flanking regions in the few cases in which it has been studied in detail (Walbot & Warren, 1990; Patterson *et al.*, 1993). In *Antirrhinum majus*, Hashida *et al.* (2003) found that higher temperature resulted in hypermethylation and lower temperature resulted in reduced methylation and this induced a remarkable change of the methylation status unique to the retrotransposon

Tam3 sequences in the genome. This methylation effect is reversible within a single generation in response to the temperature changes.

In this chapter, FISH and fibre-FISH were carried out to study the chromosomal and long-range organizations of LTR retrotransposon and tandem repeat DNA. PCR-based and immunocytochemistry methods were used to study the methylation pattern and variation of LTR retrotransposons and tandem repeat DNA. Finally, the RNA transcripts of LTR retrotransposons and tandem repeat DNA were isolated from *M. acuminata* ssp. *malaccensis* male inflorescence.

6.3 Materials and Methods

6.3.1 Fluorescent *in situ* hybridization and immunostaining with anti 5-methylcytosine antibody on metaphase spreads and extended DNA fibres

The protocol for FISH on metaphase spreads and extended DNA fibres were described in sections 2.2.10 and 2.2.11. FISH experiments were carried out on six *Musa* species (*M. schizocarpa*, $2n=2x=22$; *M. acuminata* ssp. *banksii*, $2n=2x=22$; *M. balbisiana* 'Butohan 2', $2n=2x=22$; *M. ornata*, $2n=2x=22$; *M. textilis*, $2n=2x=20$; *M. beccarii* 'Hottana', $2n=2x=18$) and three cultivars (*M. paradisiaca* 'Nendran', $2n=3x=33$; 'Pisang Awak', $2n=4x=44$ and *M. acuminata* 'Pisang Mas', $2n=2x=22$). Metaphase spreads were counter-stained with DAPI to check the number of metaphases, and the integrity and morphology of the chromosomes. The LTR retrotransposons (*Metaviridae*, *Pseudoviridae* and LARD-like element) and tandem repeat DNA, MuTR, were labeled with biotin-16- or digoxigenin-11-dUTP according to the labeling protocol described in section 2.2.9.1. The description of the probes is provided in Table 6.1.

For immunochemistry studies, *in situ* hybridization was carried out according to the protocol described in sections 2.2.10 and 2.2.11 and followed by the immunostaining with anti 5-methylcytosine antibody using the protocol in section 2.2.12. FISH probe description is provided in Table 6.1. Two replicates of each repetitive probe were carried out in these experiments with 8 slides per experiment and 10 to 20 metaphases per slide analyzed dependent on the metaphase index.

6.3.2 Methylation-sensitive PCR

1 μg of the total genomic DNA of *Musa* species (*M. acuminata* ssp. *malaccensis*, *M. balbisiana* and *M. ornata*) and cultivars (*M. paradisiaca* 'Pisang Awak', 'Pisang Rastali' and 'Pisang Mutiara', and *M. acuminata* 'Red' and 'Green Red') were digested overnight with methylation sensitive, ScrFI (CpNpG) and HpaII (CpG) or non-sensitive restriction enzymes, BstNI (CpA/TpG) and MspI (CpG) according to the enzyme buffers and digestion conditions provided by the manufacturer (New England Biolabs®). After the digestion period, 50 ng of digested total genomic DNA was used for PCR amplification with specific primers. PCR condition and parameters, gel electrophoresis and gel image visualization as in sections 2.2.2 and 2.2.3. The methylation-sensitive PCR scheme is shown in Figure 6.1.

6.3.3 Total RNA isolation and transcript analysis

The total RNA isolation of *M. acuminata* ssp. *malaccensis* male inflorescence and mRNA enrichment protocols were described in sections 2.2.6 and 2.2.7. The RT-PCR conditions and parameters described in section 2.2.8 were followed. The primer sequences and description used for RT-PCR are provided in Table 3.1 (CoTR-RTF +

CoTR-RTR for *Metaviridae* Group II element and RTF + RTR for *Metaviridae* Group III element) and Table 4.1 (MuTR1aF + MuTR1bR for MuTR repeat, and MTY1TRF + MTY1TRR01 for CoTR repeat). The sequence and phylogenetic analyses that were used for the genomic sequence study were also followed for the RNA transcripts of LTR retrotransposons and tandem repeats (as described in section 2.2.5).

6.4 Results

6.4.1 Chromosomal localization of LTR retrotransposons

Twenty eight genomic *Pseudoviridae* element clones were isolated from 10 different banana varieties in a previous study (Teo *et al.*, 2000). The clone M1 isolated from *Musa acuminata* 'Montel' (AAA) was used as a FISH probe.

FISH showed that clone M1 was dispersed throughout the *Musa* chromosomes with higher signal intensity near the centromeric regions in *M. schizocarpa* and *M. ornata* (Figure 6.2). The FISH signal intensities varied between the chromosomes within the same species (Figure 6.2C-D). Interestingly, significant signal intensity differences were observed in the two *M. paradisiaca* cultivars (Figure 6.2E & F), where some of the chromosomes showed low or almost no *in situ* signals, while others had large amounts of signal (arrows; Figure 6.2E and F). High background signals were observed in all the *Musa* taxa tested (Figure 6.2C – F).

For *Metaviridae* Group III elements, two overlapping clones, MbTy31a3 and MbTy32a4 spanning the full length element (see Figure 3.11 and 3.12) isolated from *M. balbisiana* 'Pisang Gala', were used as probes. They showed strong *in situ* hybridization signal at the NOR with weak hybridization sites dispersed throughout the chromosome arms together with high background signal in *M. balbisiana*

'Butohan 2' ($2n$=$2x$=22, Figure 6.3A – C) and *M. ornata* ($2n$=$2x$=22, Figure 6.3D – F). No significant difference in the distribution pattern of the hybridization sites was observed between the two *Musa* species tested (Figure 6.3C & F).

FISH using LARD-like element clone W-GL1 isolated from *M. acuminata* ssp. *malaccensis*, showed that the element was mainly localized at the NORs of *M. balbisiana* 'Butohan 2' (Figure 6.4A – C) and *M. acuminata* 'Pisang Mas' (Figure 6.4D – F). Weak hybridization sites dispersed along the chromosome arms were observed in both species with no significant difference, in signal distribution pattern (Figure 6.4C & F).

6.4.2 Chromosomal localization of tandem repeats, MuTR

FISH analysis using MuTR clone MoTR1a isolated from *M. ornata* (see Chapter 4) showed that the number of chromosome loci of MuTR repeats varied between *Musa* species (Figure 6.5A – G). However, the two strongest hybridization sites at the NOR were consistent throughout all species tested. Most *Musa* species contain one pair of NOR loci, except *M. ornata* ($2n$=$2x$=22) and *Musa beccarii* 'Hottana' ($2n$=$2x$=18) with four and six NOR sites, respectively (Osuji *et al.*, 1998; Lysák *et al.*, 1999; Doležel *et al.*, 2004; Bartoš *et al.*, 2005). These additional NORs also show FISH signal with MuTR (Figure 6.5D & F).

In *M. balbisiana* 'Butohan 2', the MuTR repeat showed only the two strong sites at the NOR (Figure 6.5A). In two *M. acuminata* taxa tested, additional weaker signals were observed in the centromeric regions; in 'Pisang Mas', only one centromeric site, occasionally none but never two (Figure 6.5B) and in *banksii*, 8 to 10 centromeric MuTR sites were found in most metaphase cells (Figure 6.5G). For *M. textilis*, two strong MuTR sites at the NOR loci were observed together with two

weaker signals at centromeres (Figure 6.5C). In *M. ornata*, two strong signals and two weaker signals of MuTR repeats were observed at the NORs (Figure 6.5D). *M. schizocarpa* from section *Eumusa* contained the highest number of the MuTR loci of all seven species tested, with the centromeric regions of all 22 chromosomes showing hybridization signals at different intensities (Figure 6.5E). *M. beccarii* 'Hottana' ($2n=2x=18$) showed six hybridization signals at all the NOR loci (Figure 6.5F). In some metaphase and pro-metaphase chromosomes (e.g. Figure 6.5A & B) where the NOR region was not condensed fully, the FISH signals form a trail (arrows), connecting the chromosomal satellite to the corresponding arm.

6.4.3 Organization of LTR retrotransposons and tandem repeats on extended DNA fibres

The MuTR clone, MoTR1a, and the overlapping clones of *Metaviridae* Group III elements (MbTy31a3 and MbTy32a4) were used as probes to hybridize to extended DNA fibres of *M. balbisiana* 'Butohan 2'. This showed that MuTR repeats are organized in long continuous stretches of several kb that indicate large tandemly organized arrays (Figure 6.6A, C – G) however several gaps in between the larger repeat stretches were observed. Detailed analysis of about 80 fibres (from 4 experiments) found two tandem array organizations for the MuTR repeat arrays: firstly a more compacted form such as in Figure 6.6C & G where gaps only appeared after a stretch of signal of 7 – 8 green dots (one dot = 1 kb). The second tandem array organization was more loose, with larger and more frequent gaps (Figure 6.6D – F). In all the DNA fibres analyzed, the compact forms of MuTR repeat arrays were more frequently observed compared to the loose form. *Metaviridae* Group III elements showed a dispersed pattern and were only found in

few locations on the same DNA fibre and interspersed in between some but not all the MuTR repeats (Figure 6.6C – G). No long continuous stretches of *Metaviridae* Group III elements were observed in any of the DNA fibres analyzed.

6.4.4 Distribution of cytosine methylation in metaphase spreads and extended DNA fibres

6.4.4.1 Metaphase spreads

The combination of *in situ* hybridization and immunochemical techniques was developed for *Musa* to determine the extent of cytosine methylation and its distribution. In this combined technique, the FISH hybridization step was carried out first and the FISH detection step was combined with immunostaining using anti 5-methylcytosine antibodies (see section 2.2.12). Previously, the two methods had been performed consecutively requiring photographing and relocation of metaphases after the 2nd procedure (see Neves *et al.*, 1997and Contento, 2005). The combined method shortens the experimental time and effort and also allows direct comparison of FISH and cytosine methylation signals.

All the metaphases of diploid *M. balbisiana* 'Butohan 2' (BB) analysed by immunostaining showed unevenly distributed cytosine methylation with significant differences observed between chromosomes and also between different chromosome regions (Figure 6.7 – 6.10). Small variations in the cytosine methylation distribution patterns were observed between the metaphases analysed, where cytosine methylation tended to be more uniformly distributed along more extended pro-metaphase chromosomes (Figure 6.8 & 6.9) compared to later metaphase chromosome status (Figure 6.7 & 6.10).

In general, ten to eleven chromosomes per metaphase showed higher levels of cytosine methylation, mainly concentrated at the centromeric regions and six to seven chromosomes showed a weaker but more uniform distribution along the entire chromosome (clearly shown in Figure 6.7 & 6.10). Normally, two to four chromosomes showed relatively low level of cytosine methylation (Figure 6.7 – 6.10). Often, cytosine methylation differences were observed in *M. balbisiana* 'Butohan 2' NOR loci; with one locus more heavily methylated than the other (Figure 6.7 – 6.10).

The FISH signal of *Metaviridae* Group III elements labeled the chromosome arms, two NOR loci (arrows) and most centromeres (Figure 6.7C) as described before (section 6.4.1, Figure 6.3A-C). Most of the *Metaviridae* Group III elements were located in methylated areas of the *M. balbisiana* 'Butohan 2' genome, but others were located in low methylated regions, such as the NOR loci (Figure 6.7D).

Variation in FISH signals of the LARD-like elements that located at the NOR loci of *M. balbisiana* 'Butohan 2' genome were observed with one of the locus containing more LARD-like elements (Figure 6.8C, see also Figure 6.3). The NOR locus with stronger LARD-like element signals contained a slightly lower level of cytosine methylation than the other NOR locus (Figure 6.8). The minor dispersed sites of LARD-like elements at the centromeres and the chromosome arms were in general heavily methylated (Figure 6.8D).

For *Pseudoviridae* elements, different methylation status was observed where the majority of elements were located in heavily methylated areas, around the centromeres; although there is no direct correlation of methylation signal and *Pseudoviridae* FISH signal (Figure 6.9C & D). Immunostaining FISH analysis of MuTR repeats showed that these repeats were located at the two NOR loci of *M.*

balbisiana 'Butohan 2', one signal was considerably stronger and extended over a large area in most metaphases (example in Figure 6.10C). The NOR locus that contained more abundant MuTR repeats seemed to be more heavily methylated than the other locus (Figure 6.10D).

6.4.4.2 Extended DNA fibers

The combined *in situ* hybridization and immunochemistry technique was further applied to extended DNA fibres of *M. balbisiana* 'Butohan 2' to detect individual cytosine methylation events of MuTR repeat at kilobase resolution.

Different levels of cytosine methylation were detected on these extended DNA fibres, some contained low levels of cytosine methylation (Figure 6.11) and others were heavily methylated (Figure 6.12). Several gaps without cytosine methylation signal were observed on the heavily methylated extended DNA fibres (bottom images in Figure 6.12). The MuTR FISH signals were located in low methylation regions of the extended DNA fibres (red dots in the bottom images of Figure 6.11) with other individual methylated repeat dots being more often next to rather than overlapping with the methylation antibody (yellow dots in the bottom images of Figure 6.11). All MuTR repeats found on the heavily methylated extended DNA fibres were covered by anti-5-methylcytosine antibody (yellow dots in the bottom images of Figure 6.12).

The much dispersed nature of the LTR retrotransposon with very few dots on each DNA fibre (see Figure 6.6, red signals) has limited the application of the same technique on extended DNA fibre scale.

6.4.5 Cytosine methylation study by methylation sensitive-PCR, MS-PCR

DNA methylation status of different domains, RT and LTR, of *Metaviridae* Group III elements were studied using methylation sensitive-PCR (MS-PCR). Total genomic DNA of *Musa paradisiaca* 'Pisang Awak' (ABB) was digested with methylation sensitive restriction enzymes, *ScrF*I (CpNpG) and *Hpa*II (CpG), and methylation non-sensitive restriction enzymes, *BstN*I (CpA/TpG) and *Msp*I (CpG), before PCR amplification with domain-specific primers (see methylation-sensitive PCR scheme in Figure 6.1).

For the RT domain, no difference in band intensities was observed when *Hpa*II and *Msp*I digested DNA was used as templates for PCR amplification (Figure 6.13A, lane 5 & 6) indicating that no CpG sites are present in the amplified fragments. However, decreases in PCR band intensities for *BstN*I digested genomic DNA (Figure 6.13A, lane 3) were observed compared to the uncut (Figure 6.13A, lane 2) and *ScrF*I digested DNA (Figure 6.13A, lane 4). This indicates that CpNpG sites are present in some fragments and are methylated. A similar situation was observed for the LTR region shown in Figure 6.13B, where only BstNI digested DNA showed decreased PCR band intensities (Figure 6.13B, compare lane 2 with lane 1 and 3).

MS-PCR of MuTR repeats presents a more complicated picture (Figure 6.13C). Identical patterns of uncut (Figure 6.13C, lane 2) and *Hpa*II cut DNA (Figure 6.13C, lane 5) with a strong monomer band of 220 bp and several weaker bands forming a ladder pattern were found, and higher molecular weight smears were observed after PCR amplification whereas for the *Msp*I cut DNA, two strong bands were observed with high molecular weight smears (Figure 6.13C, lane 6). Different banding patterns were observed for *ScrF*I and *BstN*I lanes (Figure 6.13C) compared to the uncut DNA lane. For the *ScrF*I lane, a strong monomeric 220 bp band was observed together

with other weak bands and less high molecular weight smears and bands more than 450 bp were absent in this lane (Figure 6.13C, lane 3). For the *BstN*I lane, a strong 110 bp band was observed together with a weaker monomeric band and a ~ 330 bp band (Figure 6.13C). Furthermore, less high molecular weight smear was observed in this lane (Figure 6.13C, lane 4). A complex methylation pattern is thus predicted for MuTR (see discussion). For MuTR repeats methylation sensitive PCR showing a identical higher molecular weight pattern after PCR amplification using uncut and *Hpa*II cut DNA as template (Figure 6.13C) together with reduction of the band size to the monomer and dimer (200bp and 400bp) but maintenance of some higher bands in the *Msp*I lane indicates that CpG sites are heavily methylated, but not present in all repeat units (Figure 6.13C, lane 6). The *BstN*I lane indicates that most units contain a CpNpG site and the band smaller than the repeat units has probably resulted from a secondary priming site (Figure 6.13C, lane 4). Methylation sensitive *ScrF*I digestion produced a strong monomeric band with less high molecular weight smear as in uncut DNA indicated that only a few CpNpG sites are methylated hindering digestion (Figure 6.13C, lane 3).

The methylation status of LTR retrotransposon flanking regions on different *Musa* genomes (Figure 6.14) was studied using outward LTR primers with restriction enzyme digested DNA as templates. No significant differences in the PCR banding pattern of *Metaviridae* Group III element flanking region was observed for uncut, *ScrF*I-cut, *Hpa*II-cut and *Msp*I-cut DNA in all three *Musa* species tested (Figure 6.14A) whereas several bands were lost in the *BstN*I-cut DNA (Figure 6.14A, lane 3, 8 and 13) indicating CpNpG methylation and that no CpG sites are present in the amplified fragments. A similar situation was observed for LARD-like element flanking regions, where only *BstN*I-cut DNA showed differences in the PCR banding

pattern between uncut and other cut DNA (Figure 6.14B), which also indicates that no CpG sites are present in the amplified fragments and some of the fragments are methylated at their CpNpG sites.

Methylation sensitive PCR was carried out on two types of somaclonal variations (disease resistance and fruit color type variation) (Figure 6.15). *M. paradisiaca* 'Pisang Mutiara' is the somaclonal variant of *M. paradisiaca* 'Pisang Rastali', which survived in the *Fusarium* 'hotspot' field study (Asif and Othman, 2005), whereas *M. acuminata* 'Green Red' is the somaclonal variant of *M. acuminata* 'Red'. A loss of bands or presence of weaker bands was observed in *BstN*I lane of *M. paradisiaca* 'Pisang Mutiara' (Figure 6.15A, lane 7 and 10) compared with the *BstN*I lane of *M. paradisiaca* 'Pisang Rastali' (Figure 6.15A, lane 4). No significant difference was observed for the *ScrF*I lane of *M. paradisiaca* 'Pisang Rastali' and 'Pisang Mutiara' (Figure 6.15A, lane 3, 6 and 9). For fruit colour somaclonal variants, the *BstN*I lanes showed no variation in their banding patterns (Figure 6.15B, lane 4 and 9) and only small variations was observed in the *Hpa*II and *Msp*I cut banding patterns (Figure 6.15B, compare *Hpa*II and *Msp*I lanes).

6.4.6 Transcription of LTR retrotransposons and tandem repeats

The enriched mRNA of *M. acuminata* ssp. *malaccensis* male inflorescence was used as a template for reverse transcriptase-PCR (RT-PCR) to isolate RNA transcripts of the LTR retrotransposon RT domains (*Metaviridae* Group II and Group III elements) and tandem repeats (MuTR and CoTR). Total genomic DNA from the same material was used as a control in this study. The cDNA PCR products of LTR retrotransposons (*Metaviridae* Group II and III elements) and tandem repeat DNA (MuTR and CoTR) were excised from the gel, cloned and sequenced.

For *Metaviridae* Group II (Figure 6.16, lane 2 & 3) and Group III elements (Figure 6.16, lane 4 & 5), extra bands of lower molecular size (~ 210 bp and ~ 350 bp, respectively) were observed together with the bands of the expected size, ~ 330 bp and ~ 1100 bp, in RT-PCR products (Figure 6.16, lane 3 for *Metaviridae* Group II element and lane 5 for *Metaviridae* Group III element). High molecular weight smearing was observed for genomic PCR products of the *Metaviridae* Group III element (Figure 6.16, lane 4). For CoTR repeats, a ladder pattern was observed in PCR products from total genomic DNA (Figure 6.16, lane 6) whereas only three bands of different size together with high molecular weight smearing were observed in RT-PCR products (Figure 6.16, lane 7). Genomic PCR and RT-PCR of MuTR repeats (Figure 6.16, lane 8 & 9, respectively) generated similar ladder patterns with monomer sizes of ~ 220 bp. The tetramer and above PCR products were not visible on the gel for genomic PCR (Figure 6.16, lane 8), which might be due to the lower DNA template (25 ng) and PCR cycles used for MuTR amplification. The trimer band of MuTR repeat is not visible in the RT-PCR product (Figure 6.16, lane 9) compared to the genomic PCR product (Figure 6.16, lane 8).

All the RNA transcript clones of *Metaviridae* Group II elements were clustered together with the genomic copies from *M. acuminata* ssp. *malaccensis* in the rooted Maximum Parsimony (MP) tree (Figure 6.17). The genomic copies of *Metaviridae* Group II elements from *M. balbisiana* 'Pisang Gala' can be subdivided into putative active (clone GALcoTR-RT1, 4 & 6) and non-active (clone GALcoTR-RT2, 3 & 5) subgroups based on the presence or absence of stop codons and frameshifts found in these sequences (see Appendix 5). The *Metaviridae* element from rice was used as out-group to root the MP tree. For *Metaviridae* Group III elements, most of the RNA transcript clones formed one group (Group I, Figure 6.18), which represented the

first active elements. Two of the RNA transcript clones (rGyRT4 & 16) were grouped together with genomic copies from *M. balbisiana* 'Pisang Gala' (MbTy31a and MbTy32a), which represented the active group II (Figure 6.18). Three non-active elements (MbTy32a1, MbTy32a5 & *Monkey*) were observed in MP tree based on RT domain sequences (Figure 6.18).

For CoTR repeats, the RNA transcript clones could be divided into two groups, active Group I and II (Figure 6.19). One of the genomic clones, Mal_coTR3c, was grouped together with the second active group, Group II (Figure 6.19). The majority of genomic clones from *M. acuminata* ssp. *malaccensis* were divided into two groups, non-active Group A and B (Figure 6.19). The genomic clones from *M. balbisiana* 'Pisang Gala' were used as an out-group to root the MP tree (Figure 6.19). For MuTR repeats, the combined MuTRa and MuTRb fragments (110 bp for each fragment) were used for phylogenetic analysis and a more complicated situation was observed, where the RNA transcript fragments were interspersed with the genomic fragments (Figure 6.20). The genomic fragments from *M. ornata* were used as out-group to root the MP tree (Figure 6.20).

6.5 Discussion

6.5.1 Genomic organization of LTR retrotransposons and tandem repeat DNA in *Musa* genomes

6.5.1.1 *Pseudoviridae* elements

In situ hybridization on four *Musa* chromosomes using the RT region of *Pseudoviridae* elements as probe, showed a dispersed distribution with stronger signals near centromeres and weaker to no signals along the euchromatin (Figure 6.2 and 6.8C). This suggested that *Pseudoviridae* elements are preferably concentrated in

the centromeric heterochromatic region of *Musa* chromosomes that contain lower numbers of functional genes. The data from genomic organization studies on other plant species using *in situ* hybridization with *Pseudoviridae* element probes on metaphase chromosomes have revealed that they are generally dispersed throughout the entire euchromatin, either evenly or unevenly, depending on the plant species studied and also the element used (Brandes *et al.*, 1997; Heslop-Harrison *et al.*, 1997; Moore *et al.*, 1991; Pearce *et al.*, 1997; Pearce *et al.*, 1996a, b). For example, the paracentromeric heterochromatic region in *Arabidopsis thaliana* and *Cicer arietinum* and other small genome plants (Schmidt *et al.*, 1995; Brandes *et al.*, 1997; Heslop-Harrison *et al.*, 1997). However, more often these elements were found to be present in lower copy number or sometimes absent in certain regions of the host chromosomes (e.g. centromeres, interstitial and terminal heterochromatic regions and ribosomal DNA sites), although there were some exceptions, where certain elements are preferentially abundant at the terminal heterochromatic regions i.e. in *Allium cepa* (Pearce *et al.*, 1996b).

In the two diploid species (*M. schizocarpa*, Figure 6.2C and *M. ornata*, Figure 6.2D) chromosomes showed similar FISH signal distribution with only small variation in signal strength, while in the two *Musa* cultivars (*M. paradisiaca* 'Nendran', AAB; Figure 6.2E and *M. paradisiaca* 'Pisang Awak', ABBB; Figure 6.2F) notable differences between chromosomes were observed. Due to the similar size of chromosomes, it is not possible to distinguish chromosome of A or B genome origin and to argue that chromosomes from one genome contain more *Pseudoviridae* sequences than the other. A more likely explanation is that *Pseudoviridae* elements are preferentially concentrated and amplified in certain *Musa* chromosomes within a genome i.e. in the B genome (see Figure 6.9C). This is supports by the preferential

localization of *Pseudoviridae* elements at the centromeric heterochromatic regions reported in the small genome plants (Schmidt *et al.*, 1995; Brandes *et al.*, 1997; Heslop-Harrison *et al.*, 1997) and being the case for *Musa* (see above), and is also evidenced by the size variation of the centromeric heterochromatic blocks observed after DAPI staining (Figure 6.9A).

6.5.1.2 *Metaviridae* Group III elements

Two *Metaviridae* Group III element clones, MbTy31a3 and MbTy32a4, that overlapped at the RT/RNaseH domain to form the *Musa* specific full-length element (see Figure 3.11 & 3.12), were used to study the genomic organization of *Metaviridae* Group III elements. FISH analysis revealed strong *in situ* hybridization signals at the NOR loci of *Musa* chromosomes with other weaker dispersed signals on most chromosomes, particularly at, the centromeres (Figure 6.3 and 6.7C). This pattern is in contrast to *Pseudoviridae* elements which can not be detected at NOR loci of *Musa* chromosomes (Figure 6.2). However, both *Metaviridae* Group III elements and *Pseudoviridae* elements were found at the centromeric regions, suggesting that these two elements might occupy similar locations at the heterochromatic regions of the centromeres. In *M. acuminata* 'Grand Naine' (AAA) genome, Balint-Kurti *et al.* (2000) also found that copies of the *Monkey* retrotransposon were concentrated in the NORs and co-localized with rRNA genes by FISH, while other copies of *Monkey* appeared to be dispersed throughout the genome. The high background signal observed in many preparations (see Figure 6.3) is probably the product of the probe hybridizing to remaining RNA copies of *Metaviridae* Group III elements. The partial RNA transcript of RT domain of these

elements isolated from the male inflorescence mRNA of *M. acuminata* ssp. *malaccensis* described in this chapter, section 6.4.6, supported this observation.

Several groups have identified retrotransposon sequences in the centromeric regions of various cereal species (Ananiev *et al.*, 1998a, b; Miller *et al.*, 1998; Presting *et al.*, 1998). Probes with sequence homology to *Metaviridae* elements were isolated from a BAC clone of a sorghum centromere, and when used as a FISH probe, hybridized exclusively to centromeric regions of sorghum chromosomes and those of all *Poaceae* tested (Miller *et al.*, 1998). A barley homologue of this sorghum *Metaviridae* element integrase sequence was used to clone a barley retrotransposon (*cereba*) that was subsequently found to be almost exclusively localized within the centromeric regions of several cereal species, including barley, wheat and rye (Presting *et al.*, 1998). The presence of *Metaviridae* element sequences in the centromeric regions of all investigated *Poaceae* species suggested that these retrotransposons are either ancient insertions that amplified before the divergence of the *Poaceae* or have arrived and/or amplified independently and become preferentially located in centromeric regions, in all of those species.

6.5.1.3 *Musa* LARD-like elements

The distribution pattern of the *Musa* LARD-like element was similar to that of the *Metaviridae* Group III elements, where the stronger *in situ* signals were observed on the NOR loci of all *Musa* taxa tested with weaker signals at the centromeres and euchromatic regions (Figure 6.4 and 6.8). However, the *in situ* hybridization signals of LARD-like elements on the NOR loci varied significantly, with one NOR having a stronger signal than the other in both the A and B genome species analyzed. The observation of FISH signals of the LARD-like elements and *Metaviridae* Group III

elements on the NOR loci indicate that these two elements are co-localized at the NORs of *Musa* chromosomes.

In barley, FISH analysis using LTR probes of barley LARD showed that these elements are not evenly distributed, the chromosome arms where uniformly labeled except for the telomeres, NORs and centromeres (Kalendar *et al.*, 2004). This is differs with what was observed for LARD-like elements in *Musa*, where they are preferentially inserted into the NORs and centromeric regions of *Musa* chromosomes (Figure 6.4C & F; 6.8C) and indicates that LARD elements can have different insertion site preferences in different taxonomic groups.

6.5.1.4 MuTR repeats

The tandem repeat, MuTR, was first isolated from *M. ornata* due to the insertion of *Metaviridae* Group III elements into this repeat family. Subsequently, MuTR repeats from *M. acuminata* ssp. *malaccensis* and *M. balbisiana* 'Pisang Klutuk Wulung' were isolated using primers designed from the conserved region and showed high homology MuTR repeats from *M. ornata* (see Chapter 4).

The MuTR repeat consistently hybridized to the NOR loci of all the species tested with small variation in signal intensities (Figure 6.5; 6.10C and D). However, the *in situ* signals on the centromeres varied depending on the *Musa* species. For instance, the MuTR repeat was not detected at any of the centromeric regions in *M. balbisiana* 'Butohan 2' (BB) chromosomes (Figure 6.5A; 6.9C), whereas in *M. schizocarpa* (SS), the MuTR repeat probe hybridized to all the centromeres (Figure 6.5E). Furthermore, the two *M. acuminata* (AA) genomes tested showed different number of hybridization sites on their centromeres, with 8 to 10 sites on the *banksii* genome (Figure 6.5G) and only one site in the 'Pisang Mas' genome (Figure 6.5B).

Different types of tandem repeat DNA have been shown to occupy different regions of the eukaryotic genome: centromeres, subtelomeric and intercalary heterochromatin, and the rDNA regions (Heslop-Harrison *et al.*, 1999; Nonomura and Kurata, 1999; Kumekawa *et al.*, 2000; Gindullis *et al.*, 2001; Kishii *et al.*, 2001; Nonomura and Kurata, 2001; Cheng *et al.*, 2002; Murata, 2002; Heslop-Harrison *et al.*, 2003; Schwarzacher, 2003; Alkhimova *et al.*, 2004; Nagaki *et al.*, 2003a, b; 2004; Contento *et al.*, 2005) and sometimes, they co-exist with other type of repetitive DNA such as LTR retrotransposons, especially at the centromeric regions (see above, Pellissier *et al.*, 1996; Sun *et al.*, 1997; Gindullis *et al.*, 2001; Kumekawa *et al.*, 2001; Zhong *et al.*, 2002; Nagaki *et al.*, 2003a; 2004).

A ~ 63 bp *Metaviridae* Group III element LTR-homologue sequence was identified within the MuTR repeats (see Figure 4.6 to 4.8) and there is a small possibility that the MuTR probes cross-hybridized with *Metaviridae* Group III elements at the NOR due to the high homology with the ~ 63 bp sequence (Figure 4.6 to 4.8). However, this possibility is unlikely as the MuTR *in situ* signal is much stronger and more uniform on metaphases and the fibre-FISH results showed that most of the *Metaviridae* Group III element signals were interspersed in between the MuTR repeat signals with only limited signal overlap (Figure 6.5).

The presence of this ~ 63 bp homologous box together with the genomic organization of *Metaviridae* Group III elements that are abundant at the NOR loci and fewer copies on centromeres (Figure 6.3), suggests that the retrotransposons might have played an important role in the transposition of MuTR repeats from the NOR loci to some AA and all SS genome centromeres (Figure 6.5B – G) or other process might involved such as the rolling cycle mechanism proposed by Okumura *et al.* (1987). The close relationship between centromeric tandem repeat DNA and LTR

retrotransposons, in particular *Metaviridae* elements, has been studied in other plant species. Tandem repeats and *Metaviridae* elements were not only found adjacent to each other at the centromeres, but some of tandem centromeric repeats are derived from different parts of an LTR retrotransposon (Aragon-Alcaide *et al.*, 1996; Jiang *et al.*, 1996; Ananiev *et al.*, 1998a; Miller *et al.*, 1998; Presting *et al.*, 1998; Langdon *et al.*, 2000; Cheng and Murata, 2003) or as in the MuTR repeat described here parts of retrotransposon sequences are incorporated in higher order repeat structures.

On the *M. balbisiana* 'Butohan 2' DNA fibres, the *Metaviridae* Group III elements displayed a much dispersed pattern with only few FISH dots per fibre which corresponds to the dot like appearance of the FISH signal on metaphase chromosome (see Figure 6.3 and 6.8). The MuTR repeats on the other hand showed long stretches of hybridization signal, that is the characteristic of tandemly organized repeats (Figure 6.6) and again corresponds to their very strong and compact FISH signal on metaphase chromosomes (Figure 6.5). The distribution patterns of MuTR repeat on extended DNA fibres (Figure 6.6) indicated two different types, compact and loose genomic organization, even though they occupied the same region of the *Musa* chromosome; the NOR loci (Figure 6.3 & 6.5). The different forms of MuTR repeat arrays, compact (Figure 6.6C & G) and loose (Figure 6.6D – F), might be due to the different extension rates of the DNA fibres during the preparation step although this is unlikely when comparing absolute density. From extended early prophase chromosomes it is clear that some copies of MuTR are within the active part of the NOR and these might be interspersed with 45S rDNA copies. The NOR region in *Musa* is complex and contains, as shown here by FISH, many different types of repetitive DNA. LTR retrotransposons (*Metaviridae* Group III and LARD-

like elements), and tandem repeat DNA (MuTR and 45S rRNA gene). A model of their organization is presented in Chapter 7.

6.5.2 Cytosine methylation of LTR retrotransposons and MuTR repeats

The cytosine methylation levels and distribution on *M. balbisiana* 'Butohan 2' chromosomes observed in this study varied slightly depending on the metaphase spread and the chromosome. In general, immunostaining with antibodies against 5-methylcytosine showed that the centromeric regions of about half of the chromosomes are heavily methylated while the distal euchromatin and remaining centromeric regions are generally free from methylation (Figure 6.7 – 6.10). Cytosine methylation level variations are also observed on the NOR loci; with one locus more heavily methylated than the other (Figure 6.7 – 6.10). This wide-range cytosine methylation distribution on individual chromosomes may be due to the technical factors (presence of cytoplasm or the chromosome was not totally flat). However, this variation was observed on all the metaphase spreads analysed from several experiments that were carried out at different times and so can be regarded as real.

Immunochemical studies on *A. cepa* and *V. faba*, two species with large genomes, showed unevenly distributed cytosine methylation along the chromosomes (Castiglione *et al.*, 1995; Frediani *et al.*, 1996). Another study by Siroky *et al.* (1998) showed that in *Melandrium album* (small genome with sex chromosomes), was the autosomes were evenly methylated, whilst sex chromosomes in both male and female metaphases were unevenly distributed. In contrast to the above studies, the methylation pattern in diploid *Triticeae* species (genome sizes slightly smaller than *A. cepa*, *V. faba* and *M. album*), was uniform methylated the euchromatin with no

differences between chromosomes (Contento, 2005). In polyploid *Triticeae*, uneven methylation distribution was observed (Contento, 2005). Among these studied species (*A. cepa*, *V. faba*, *M. album* and diploid *Triticeae* species), *Musa* with the smallest genome size (500 – 600 Mbp/1C) showed cytosine methylation distribution pattern similar to the species with large genomes (*A. cepa*, *V. faba* and diploid *Triticeae*) compared to *M. album*.

In situ hybridization of LTR retrotransposons (*Metaviridae* Group III element, Figure 6.7; LARD-like element, Figure 6.8; and *Pseudoviridae* element, Figure 6.9) together with the anti 5-methylcytosine antibody immunostaining data showed different methylation status of LTR retrotransposons. The majority of these elements was located at the heavily methylated centromeric region although some elements were found on the low methylated centromeres. For the *Metaviridae* Group III elements located at the NOR loci (Figure 6.7D), variation in the cytosine methylation status provided some indication that actively transcribed elements might still exist in the *Musa* genome because DNA methylation has been shown to play an important role in silencing of LTR retrotransposons in different organisms (Hirochika *et al.*, 2000; Lyko *et al.*, 2000; Zhou *et al.*, 2001; Burgers *et al.*, 2002; Hashida *et al.*, 2003; Lippman *et al.*, 2003; Kato *et al.*, 2004; Liu *et al.*, 2004; Verdel *et al.*, 2004; Kuhlmann *et al.*, 2005). The methylation sensitive PCR results also reflect a similar situation where not all the RT and LTR domains are methylated (Figure 6.12A & B, respectively). This is supported by the isolation of partial RT domain mRNA from male inflorescences of *M. acuminata* ssp. *malaccensis* (Figure 6.16).

A similar situation was observed for MuTR repeat, located on the NOR loci with different levels of cytosine methylation (Figure 6.10). This variation is also observed in the extended DNA fibre FISH, where low methylated (Figure 6.11) and

heavily methylated (Figure 6.12) DNA fibres of *M. balbisiana* 'Butohan 2' were observed. MuTR repeats located on the heavily methylated DNA fibres were methylated (Figure 6.12) whereas most of the repeats on the low methylation region were not (Figure 6.11). The methylation sensitive PCR results also showed that not all MuTR repeats are methylated (Figure 6.13C), this supports the immunochemical studies on metaphases (Figure 6.10D) and extended DNA fibres (Figure 6.11 & 6.12). This provides the prerequisite for the transcription of the MuTR repeats, which is supported by the observation of RNA transcripts of those repeats in *M. acuminata* ssp. *malaccensis* (Figure 6.16).

6.5.3 Types of cytosine methylation of the LTR retrotransposon, its flanking region and MuTR repeats

Cytosine methylation is prevalent in symmetrical CpG contexts in animals whereas the symmetrical CpNpG contexts as well as in non-symmetrical CpHpH (H=C, A, T) are more often observed in plants (Gruenbaum *et al.*, 1981; Meyer *et al.*, 1994; Grafi *et al.*, 2007).

In general, *Metaviridae* RT and LTR sequences showed low levels of methylation due to the lack of symmetrical cytosines. The sequences analyzed by methylation-sensitive PCR did not contain CpG sites (see Chapter 3). Hence, neither methylation sensitive nor non-sensitive restriction enzymes can cut the RT and LTR sequences, and this result in the PCR fragments with similar intensities (see Figure 6.13A and B). CpNpG sites were rare but when present were methylated (Figure 6.13A and B). Flanking sequences of *Metaviridae* Group III and MuLARD elements showed higher levels of methylation at CpNpG sites but again lacked CpG sites (Figure 6.14).

MuTR repeat methylation sensitive PCR showed that CpG sites where found are heavily methylated (Figure 6.13C). Most repeat units contain a CpNpG site but only a few are methylated hindering digestion (Figure 6.13C). Although less common in plants, CpG methylation has been shown to be important for heterochromatin formation. Tanaka *et al.* (2005) reported that the centromere protein B (CENP-B), which specifically binds to the centromeric 17 bp CENP-B box DNA, was preferentially bound to the unmethylated CENP-B box DNA and the affinity of this binding was reduced to the nonspecific DNA binding level by the methylation of the CpG sites in CENP-B box DNA. A similar situation might be expected for the MuTR repeats which are located at the centromeres of some *Musa* species tested (Figure 6.5B –G).

Teo *et al.* (2005) and further work described in Chapter 5 showed that IRAP markers can detect new retrotransposon insertion in tissue cultured bananas. Furthermore, Asif and Othman (2005) showed that IRAP markers can distinguish the *M. paradisiaca* 'Pisang Mutiara' from its mother plant, *M. paradisiaca* 'Pisang Rastali' and detect new insertion sites. My study has shown that mutations occurred in *M. paradisiaca* 'Pisang Mutiara', which introduce new CpNpG sites (Figure 6.15A, compare lane 4, 7 and 10) that are methylated (Figure 6.15A, compare lane 6, 7, 9 and 10). This indicates that not only DNA methylation but also DNA sequence mutations which increase the available cytosines for methylation are involved in retrotransposon regulation in disease resistance type somaclonal variants. No difference in the *BstN*I cut banding pattern was observed in fruit colour somaclonal variants, *M. acuminata* 'Green Red' and its mother plant, *M. acuminata* 'Red' (compare Figure 6.15B, lane 4 and 9), indicating that neither methylation nor mutation of the retrotransposon are involved in this type of somaclonal variation. Liu

et al. (2004) showed that rice endogenous retrotransposon, *Tos17*, is activated by tissue culture and this process is accompanied by cytosine demethylation and also causes heritable alteration in the methylation pattern of retrotransposon flanking regions. Kubis *et al.* (2003) showed that in oil palm, genome-wide reduction of DNA methylation was observed during tissue culture but was restored to near-normal levels in regenerated trees. They also showed that different levels of sequence methylation were observed between ortet palms (parent) and regenerated palm trees with mantled phenotype (degenerated fruits) but no differences in the genomic organization of the different classes of transposable DNA elements were observed in those trees.

6.5.4 Diversity of RNA transcripts of LTR retrotransposon and tandem repeat

High DNA sequence homology at the RT domains of *Metaviridae* Group II and III elements has led to the prediction that some of these elements are still actively transcribed in *Musa* (see Chapter 3). Indeed, the isolation of RNA transcripts from *Metaviridae* Group II and III elements (Figure 6.16) supports this prediction. Furthermore, the Maximum Parsimony (MP) tree of *Metaviridae* Group II elements (Figure 6.17) and Group III elements (Figure 6.18) shows that there are active and inactive groups of these elements in *Musa*. The interspersion between RNA transcripts and genomic copies of *Metaviridae* Group II elements (Figure 6.17) indicates that these genomic copies are still active in the *M. acuminata* ssp. *malaccensis* genome. Two RNA transcript (rGyRT4 & 16) of *Metaviridae* Group III elements of *M. acuminata* ssp. *malaccensis* (AA) are grouped together with the genomic copies (MbTy31a and MbTY32a, Figure 6.18) isolated from *M. balbisiana*

'Pisang Gala' (BB) suggesting that this group of retrotransposons is actively transcribed in the A and B genomes.

Many genomic studies have shown that plant genomes contain medium to high copy number of *Metaviridae* elements, which are ubiquitous and highly heterogeneous component of their host genomes (SanMiguel and Bennetzen, 1998; Balint-Kurti, 2000; Yang *et al.*, 2005; Peigu *et al.*, 2006). The highly heterogeneous genomic populations of *Metaviridae* elements may give rise to a RNA population with high heterogeneity if they are actively transcribed and this will subsequently increase the heterogeneity of *Metaviridae* element genomic population in the host genome. In *Musa* species, no identical RNA sequence of *Metaviridae* Group II and III elements were found (Figure 6.17 & 6.18) suggesting a high heterogeneity of RNA transcript population, which will generate high heterogeneous genomic copies if they successfully integrate into the *Musa* genome.

The isolation of RNA transcripts of CoTR repeats that are inserted into a Metaviridae Group II elements suggests that these repeats are transcribed together with the *Metaviridae* Group II element in the *M. acuminata* ssp. *malaccensis* (Figure 6.16, lane 6 & 7). However, the number of repeat units that are found in the RNA transcripts of this element are much lower, which results in the amplification of two strong bands of different sizes (Figure 6.16, lane 7) compared to the ladder pattern observed in the genomic DNA (Figure 6.16, lane 6). The observation of CoTR genomic copy grouped together with RNA transcripts in the active Group II (Figure 6.19) suggests that this genomic copy, Mal-coTR3c, might be still actively transcribed in the *M. acuminata* ssp. *malaccensis*. The remaining RNA transcripts and genomic clones form separate branches on the MP tree suggesting that some groups of CoTR repeats are active while others are not (Figure 6.19). The

amplification of RNA transcripts of MuTR repeats (Figure 6.16) indicates that this repeat family is transcribed in *Musa*. The Mal-TR genomic clones are grouped together with the RNA transcripts (Figure 6.20) suggesting that these genomic clones might be the counterparts of the isolated RNA transcripts.

Tandem repeat or satellite DNA has been shown to be actively transcribed *in vivo* and *in vitro* in other organisms, such as in *Notophthalmus viridescens* (Epstein and Pabón-Peña, 1991) and *Dolichopoda* cave cricket species (Rojas *et al.*, 2000). Some of these RNA transcripts do self-cleave *in vitro*, which is correlated with the efficient *in vivo* processing of long primary transcripts into monomer-sized RNA (Rojas *et al.*, 2000). In *Schistosoma mansoni*, the RNA transcripts of the Smα satellite DNA, which contained a hammerhead ribozyme domain, which is expressed as a long multimeric precursor RNA that can be cleaved *in vitro* and *in vivo* into unit-length fragments (Ferbeyre *et al.*, 1998). For *Musa* species, the RNA transcripts of the MuTR repeat are also transcribed in multimeric RNA form, ladder pattern in RT-PCR product (Figure 6.16, lane 9). However, further analysis is needed to study whether this multimeric RNA form can be cleaved, either *in vitro* or *in vivo*, to generate the unit-length fragments.

Table 6.1 Description of the *in situ* hybridization probes used in this chapter

Probe Name	Description	Length (bp)	Application
M1	*Pseudoviridae* element, RT domain	309*	FISH and immunochemistry studies
MbTy31a3	*Metaviridae* Group III element, containing 5' LTR, AP, GAG and RT, RNaseH domains	3000#	FISH, fiber-FISH, immunochemistry studies
MbTy32a4	*Metaviridae* Group III element, containing RT, RNaseH, INT, CHROMO and 3' LTR domains	4000#	FISH, fiber-FISH, immunochemistry studies
W-GL1	*Musa* LARD-like element, internal domain	~ 3500#	FISH and immunochemistry studies
MoTR1a	MuTR repeat	220+	FISH, fiber-FISH, immunochemistry studies

* sequence published in Teo *et al.* (2002)
# sequence described in Chapter 3
+ sequence described in Chapter 4

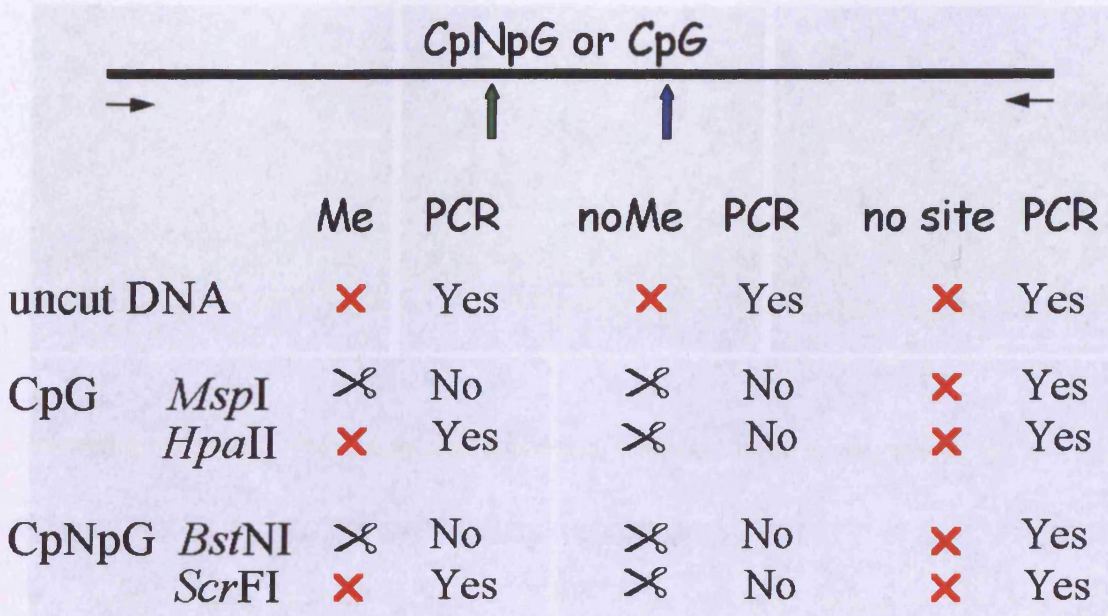


Figure 6.1: Methylation-sensitive PCR scheme. Genomic DNA containing CpNpG or CpG sites (green and blue arrows) are shown in black solid line. Primers are indicated by black arrows. In the uncut genomic DNA where all the CpNpG or CpG sites are not digested (X), PCR products with expected band sizes are generated. In the CpNpG and CpG enzyme-digested genomic DNA, PCR products only generated in the template DNA with no cutting sites (no site) or with methylated sites (Me). For genomic DNA that containing non-methylated sites (noMe), the DNA is digested (✂) by either methylation-sensitive (*Hpa*II, *Scr*FI) or non-sensitive (*Msp*I, *BstN*I) and no PCR products are generated.

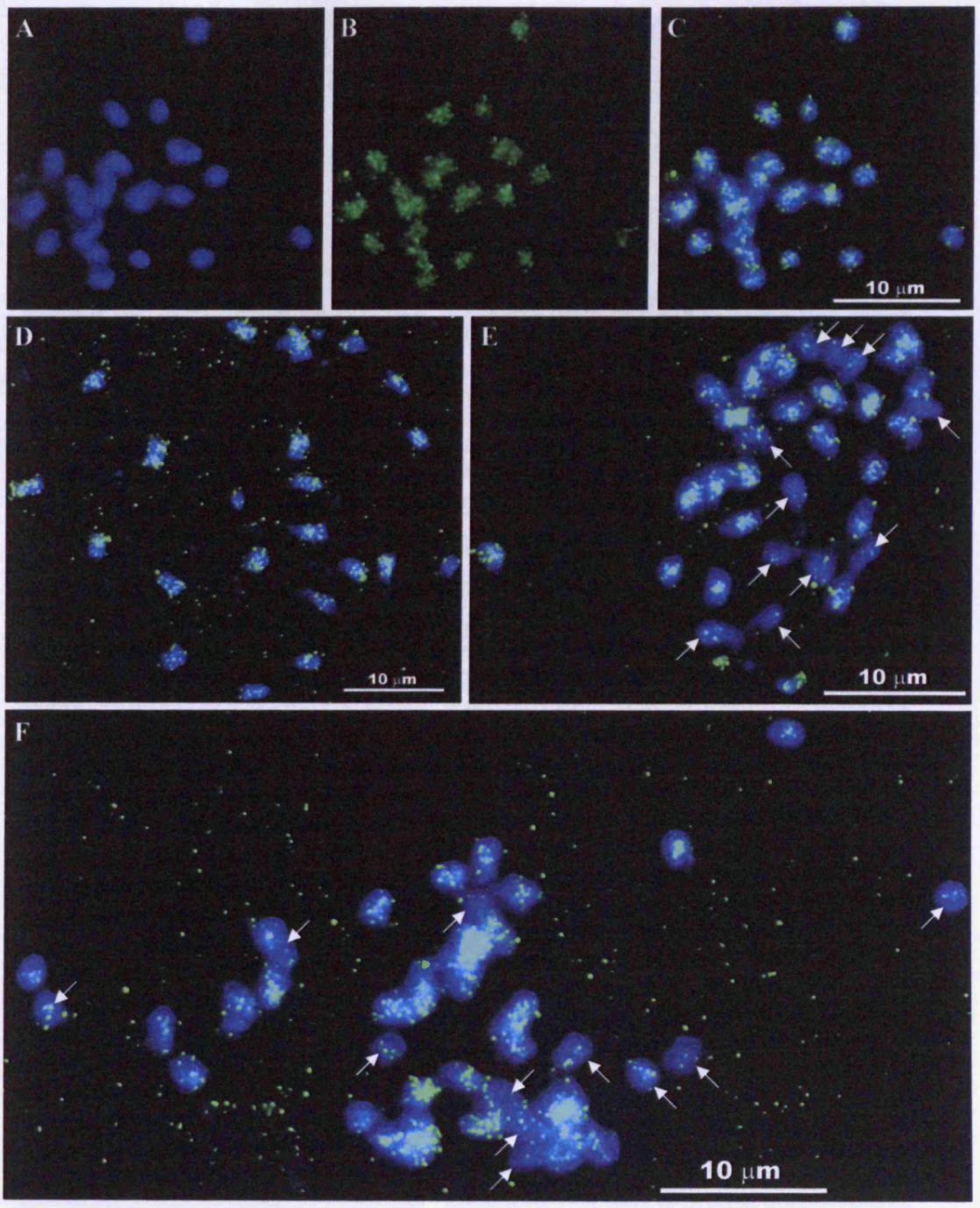


Figure 6.2: *In situ* hybridization of the *Pseudoviridae* element, clone M1 (green signal), to root tip chromosome preparations of *Musa* species counterstained with DAPI (blue): **A-C** *M. schizocarpa* (SS), **D** *M. ornata* and *M. paradisiaca* varieties: **E** 'Nendran' (AAB), **F** 'Pisang Awak' (ABBB). **A** DAPI signal on *M. schizocarpa*; **B** *in situ* signal on *M. schizocarpa*; **C-F** overlay of DAPI and *in situ* hybridization signals. Arrows show the chromosomes containing low amounts of *Pseudoviridae* elements.

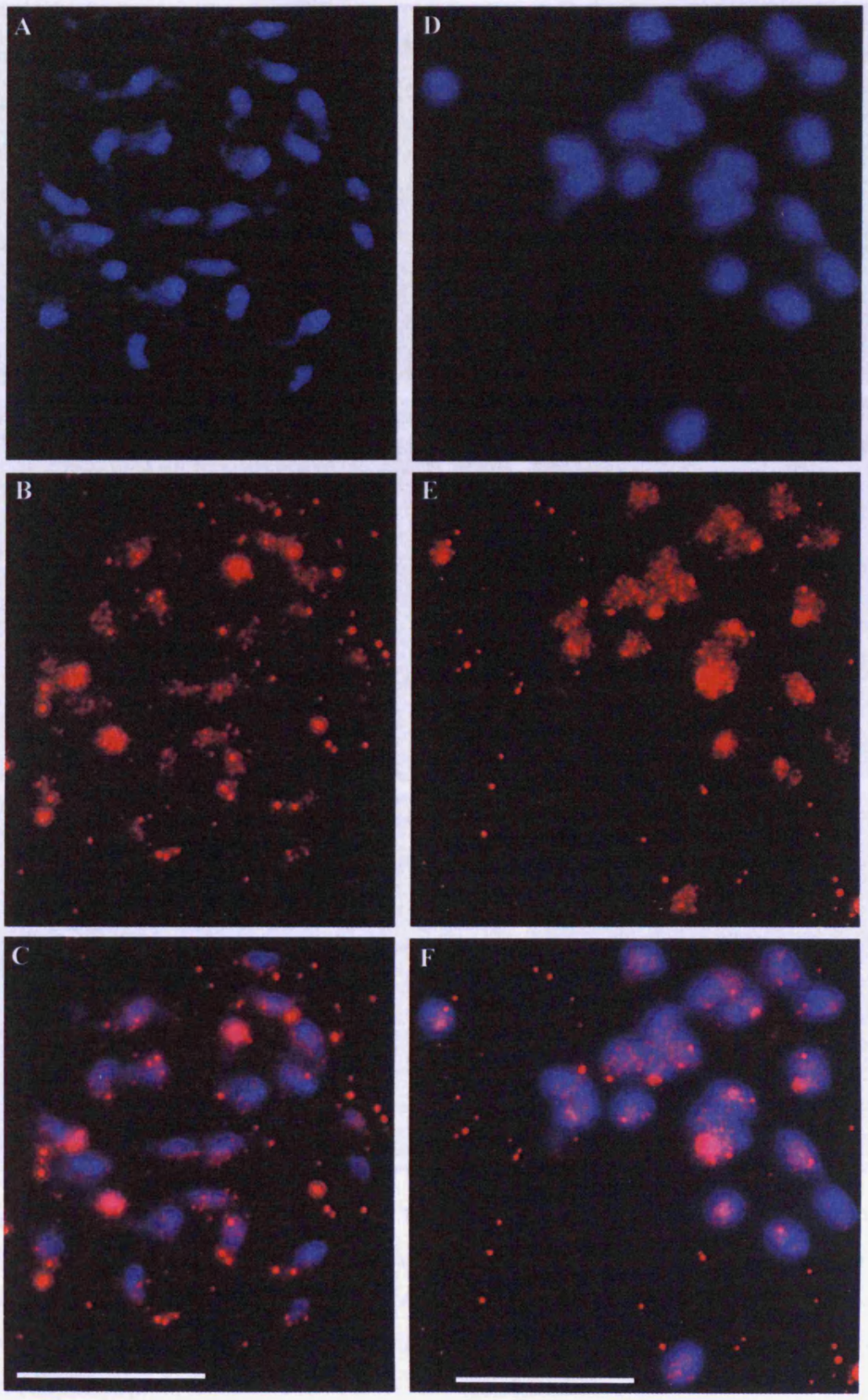


Figure 6.3: *In situ* hybridization of *Metaviridae* Group III element (combined probe MbTy31a3 + MbTy32a4; red signal) to root tip chromosome preparations of *Musa*: *M. balbisiana* 'Butohan 2', (BB) (**A-C**); *M. ornata* (**D-F**). Chromosomes were counter-stained with DAPI (blue). Bar = 10 μm.

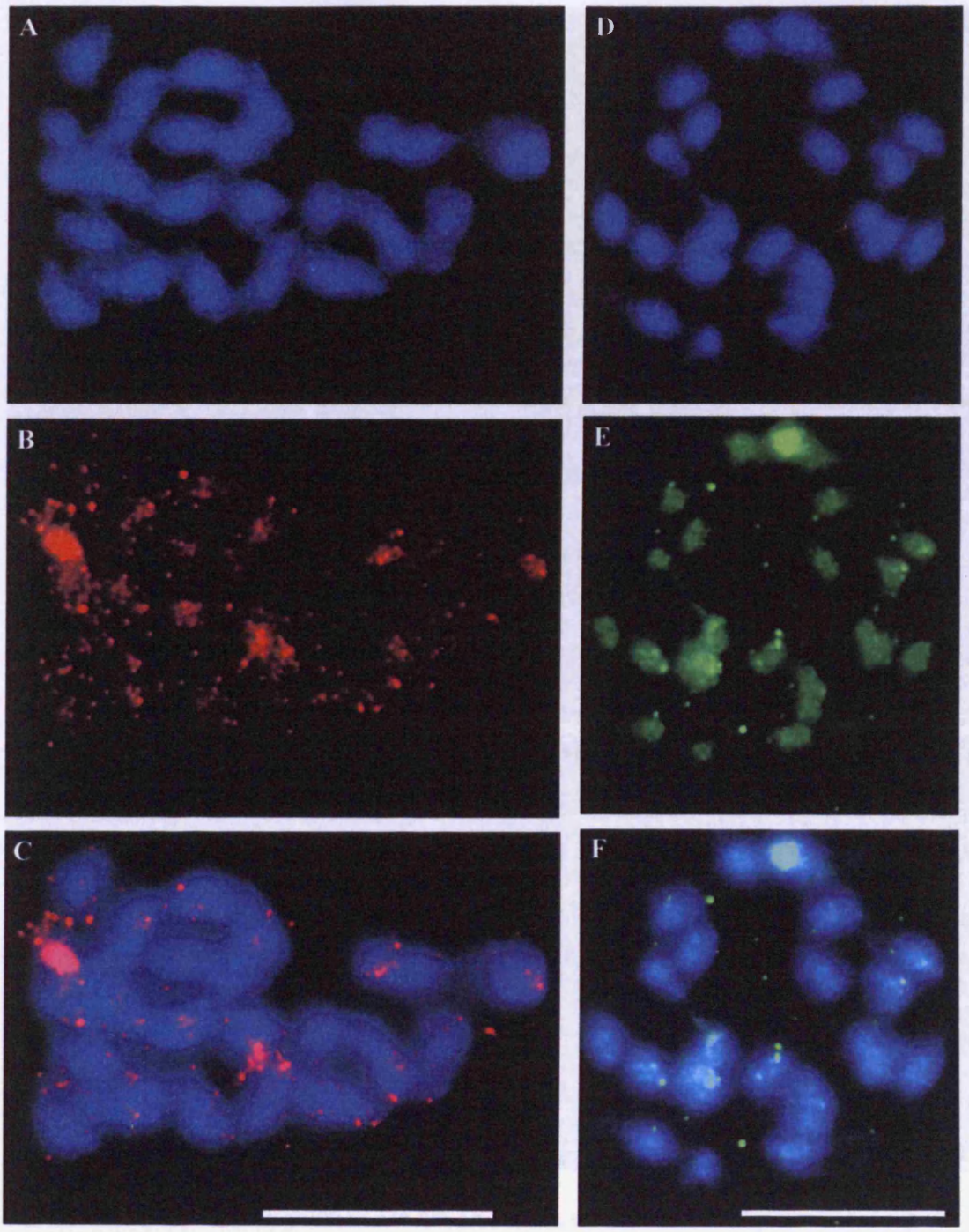


Figure 6.4: *In situ* hybridization of the LARD-like element, clone W-GL1, to root tip chromosome preparations of two *Musa* species. **A-C** *M. balbisiana* 'Butohan 2' (BB) (red signal), **D-F** *M. acuminata* 'Pisang Mas' (AA) (green signal). Chromosomes were counter-stained with DAPI (blue; **A** and **D**). **B, E** *in situ* signal. **C, F** overlay of DAPI and *in situ* hybridization signals. Bar = 10 μm.

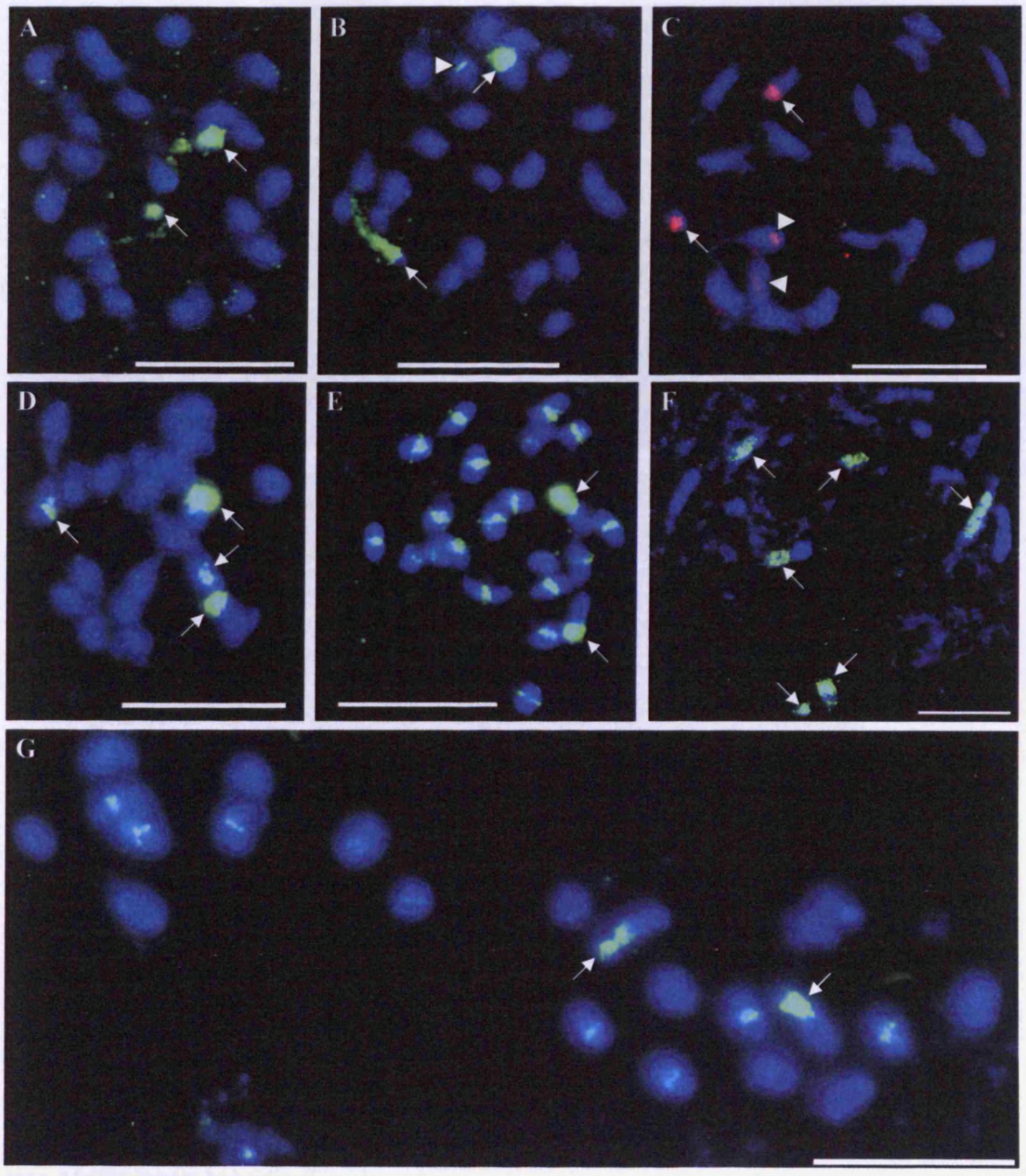


Figure 6.5: *In situ* hybridization of the MuTR repeat, clone MoTR1a, (green or red signal) to root tip chromosome preparations of seven *Musa* species. **A** *M. balbisiana* 'Butohan 2' (BB), **B** *M. acuminata* 'Pisang Mas' (AA), **C** *M. textilis* (TT), **D** *M. ornata* **E** *M. schizocarpa* (SS), **F** *M. beccarii* 'Hottana', **G** *M. acuminata* subspecies *Banksii* (AA). Arrows show the *in situ* signals at NOR loci whereas arrowheads show some of the extra (1-2) sites in **B** and **D**. Bar = 10 μm.

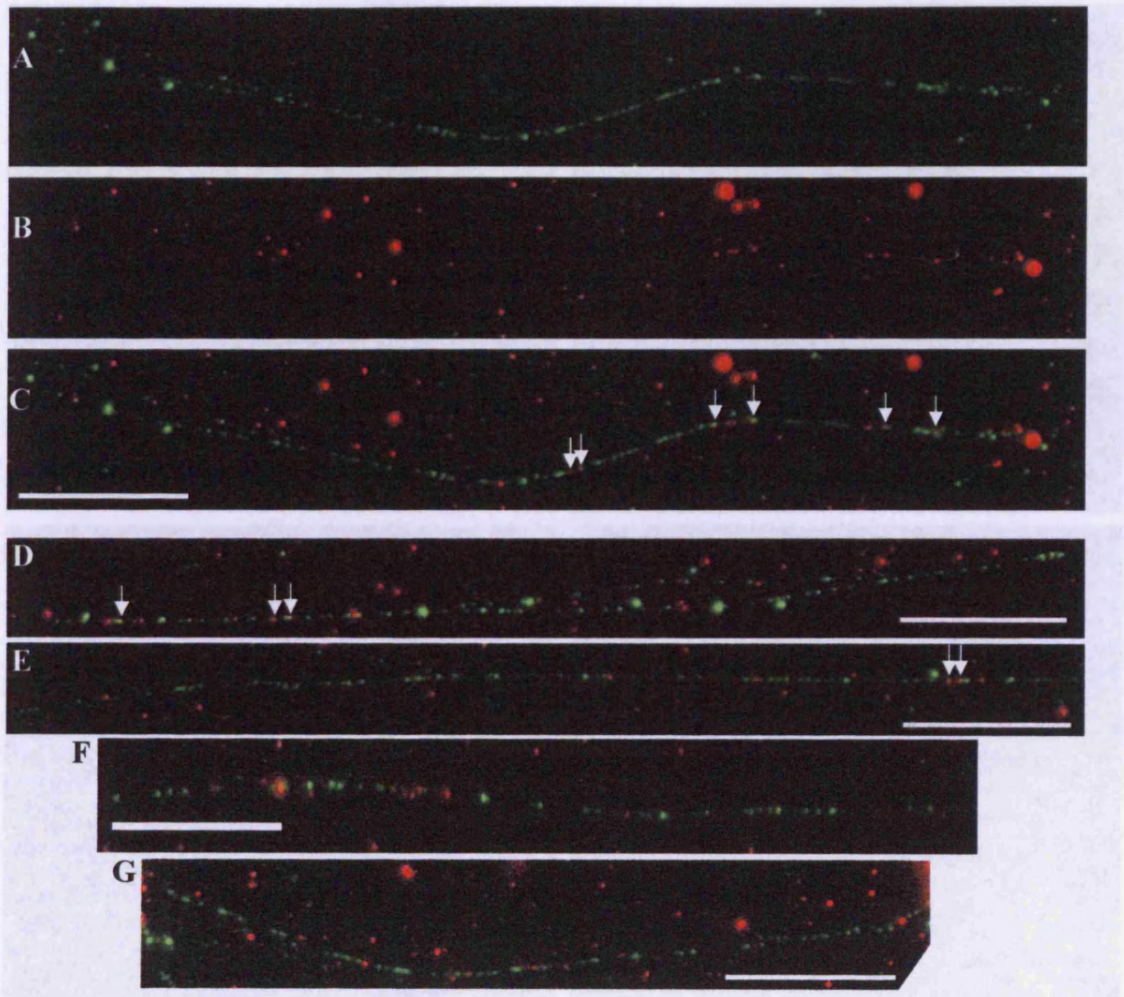


Figure 6.6: *In situ* hybridization of MuTR repeats, clone MoTR1a, (green signal) and *Metaviridae* Group III element (combined probe MbTy31a3 + mbTy32a4, red signal) on extended DNA fibers of *M. balbisiana* 'Butohan 2' (BB). Extended DNA fiber FISH of **A)** MuTR repeats, **B)** *Metaviridae* Group III element and **C - G)** overlay of MuTR repeats and *Metaviridae* Group III element. The MuTR repeats show long stretches of hybridization sites along the extended DNA fibers with several gaps and the *Metaviridae* Group III element probe is interspersed within the long MuTR repeat hybridization sites. Arrows showed colocalization sites of MuTR repeats and *Metaviridae* Group III elements. Bar = 28.7 kbp/10 μm based on the extension degree of the DNA fibers prepared from *Arabidopsis thaliana* (Jackson *et al.*, 1998).

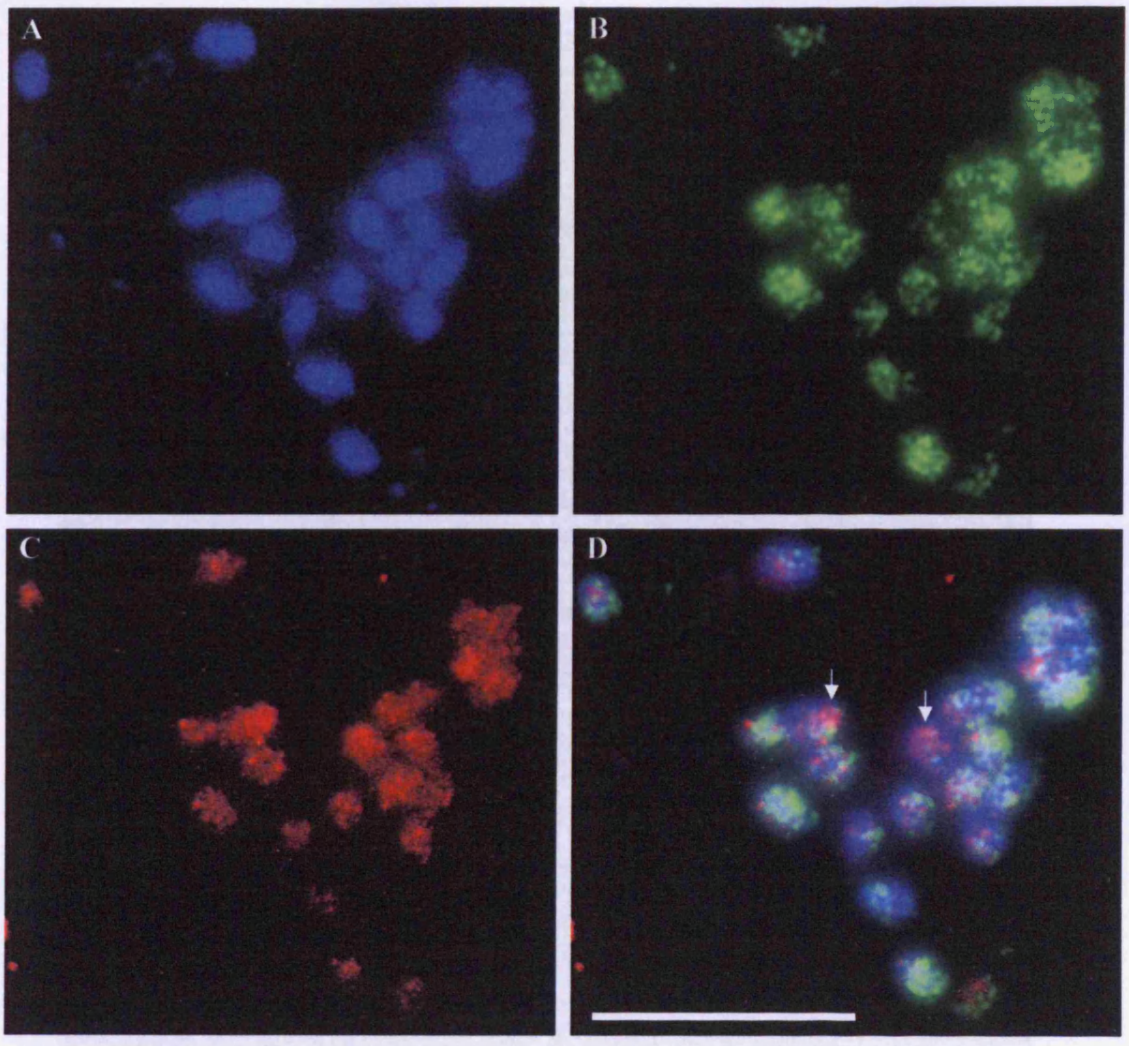


Figure 6.7: *In situ* hybridization of the *Metaviridae* Group III elements (combined probe MbTy31a3 + mbTy32a4, red signal) and immunostaining of *M. balbisiana* 'Butohan 2' (BB) metaphase chromosomes with anti-5-methylcytosine antibody. The *Metaviridae* Group III element was labelled with biotin-dUTP and detected with Alexa 594-streptavidin (red signal). The chromosomes were also incubated with anti-5-methylcytosine and detected with secondary antibody conjugated to FITC (green signal). **A** DAPI signal, **B** methylcytosine signal, **C** *In situ* hybridization signal, **D** Overlay plate of **A-C**. Arrows showed the *in situ* signals of *Metaviridae* Group III elements at the NOR loci. Bar = 10 μm.

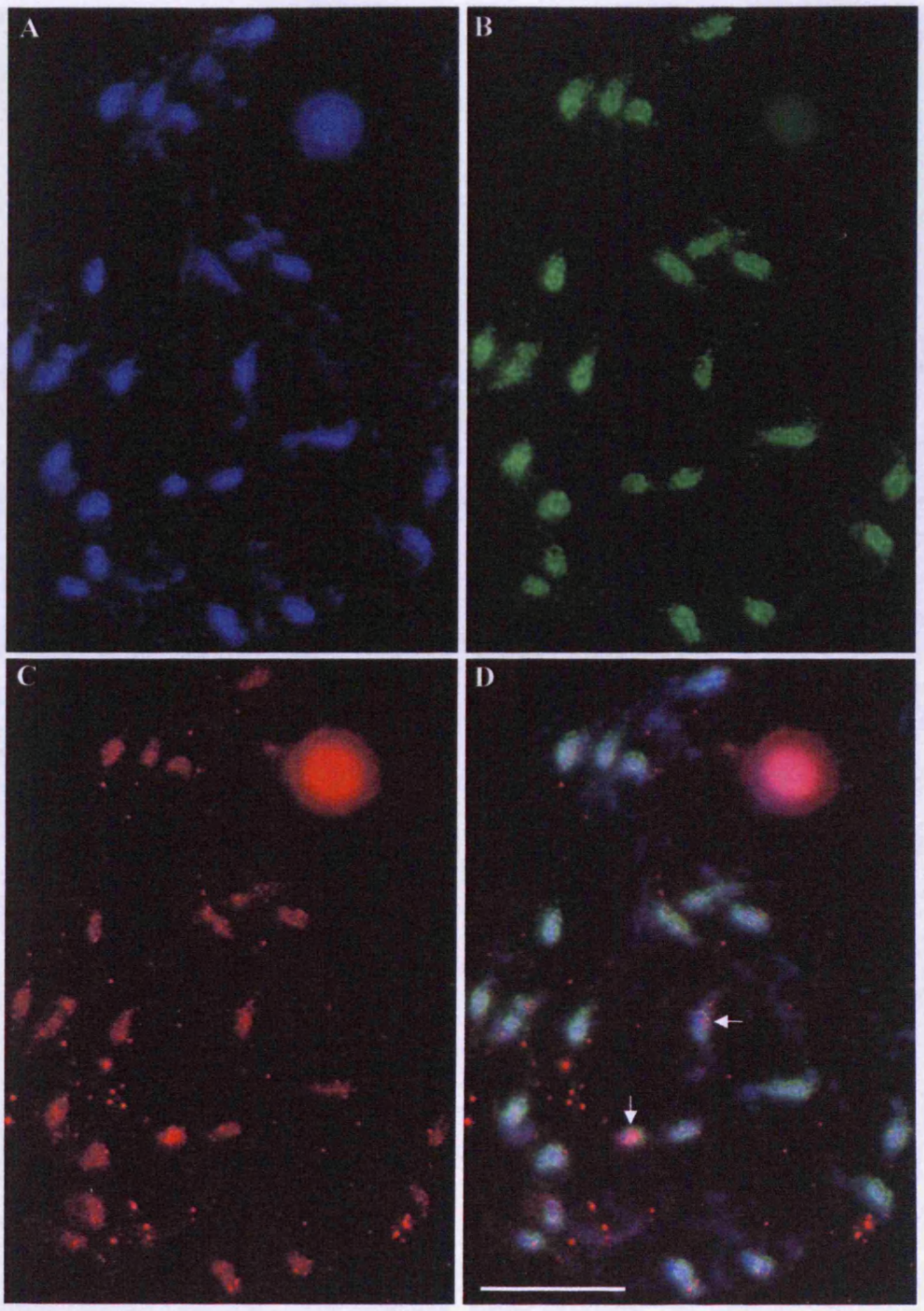


Figure 6.8: *In situ* hybridization of the LARD-like elements, clone W-GL1 and immunostaining of *M. balbisiana* 'Butohan 2' (BB) metaphase chromosomes with anti-5-methylcytosine antibody. The LARD-like element was labelled with biotin-dUTP and detected with Alexa 594-streptavidin (red signal). The chromosomes were also incubated with anti-5-methylcytosine and detected with secondary antibody conjugated to FITC (green signal). **A** DAPI signal, **B** methylcytosine signal, **C** *In situ* hybridization signal, **D** Overlay plate of **A-C**. Arrows showed the *in situ* signals of LARD-like elements at the NOR loci. The large red circle was caused by an out of focus dye precipitate. Bar = 10 μm.

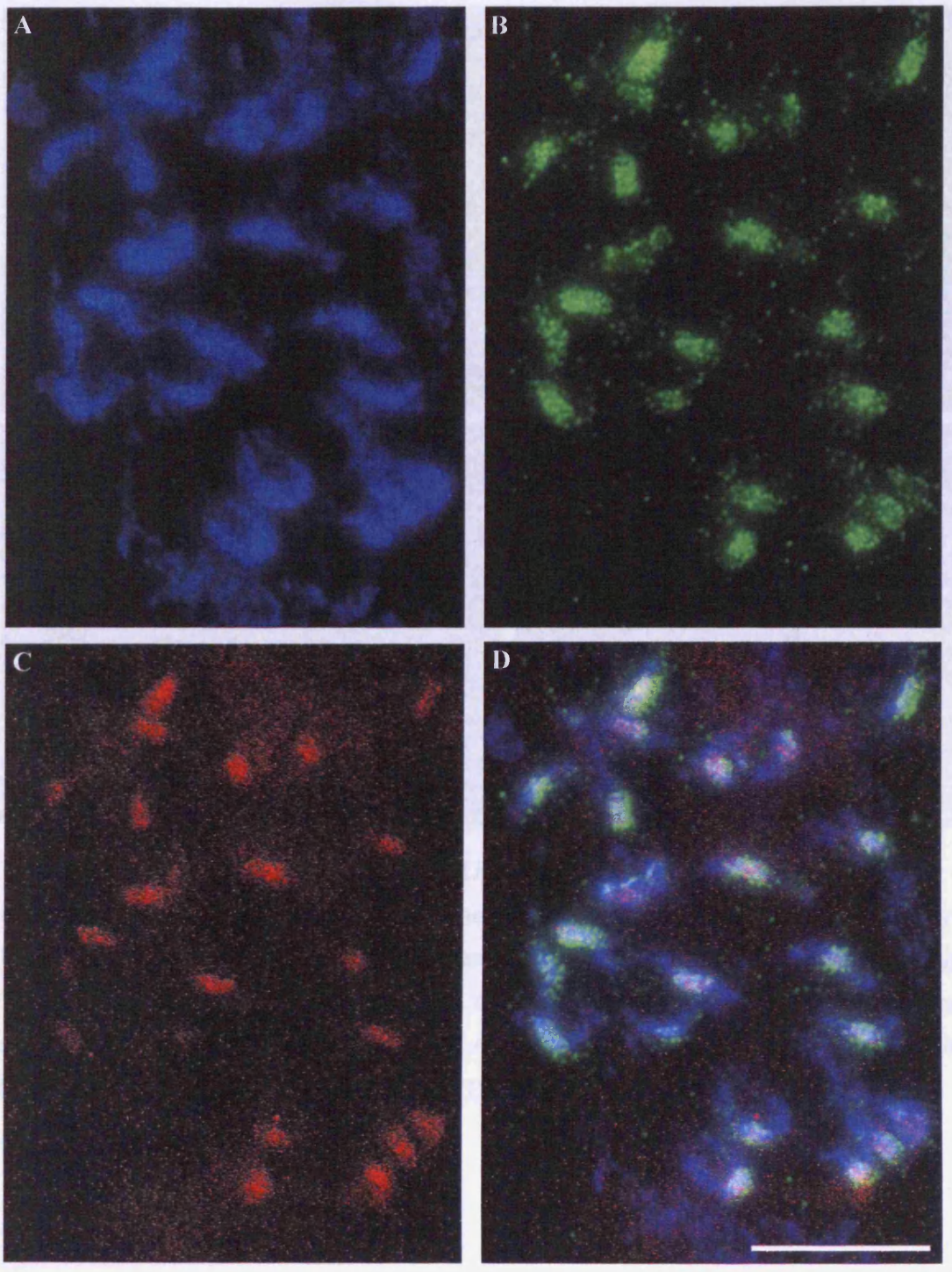


Figure 6.9: *In situ* hybridization of the *Pseudoviridae* elements, clone M1 and immunostaining of *M. balbisiana* 'Butohan 2' (BB) metaphase chromosomes with anti-5-methylcytosine antibody. The *Pseudoviridae* element was labelled with biotin-dUTP and detected with Alexa 594-streptavidin (red signal). The chromosomes were also incubated with anti-5-methylcytosine and detected with secondary antibody conjugated to FITC (green signal). **A** DAPI signal, **B** methylcytosine signal, **C** *In situ* hybridization signal, **D** Overlay plate of **A-C**. Bar = 10 μm.

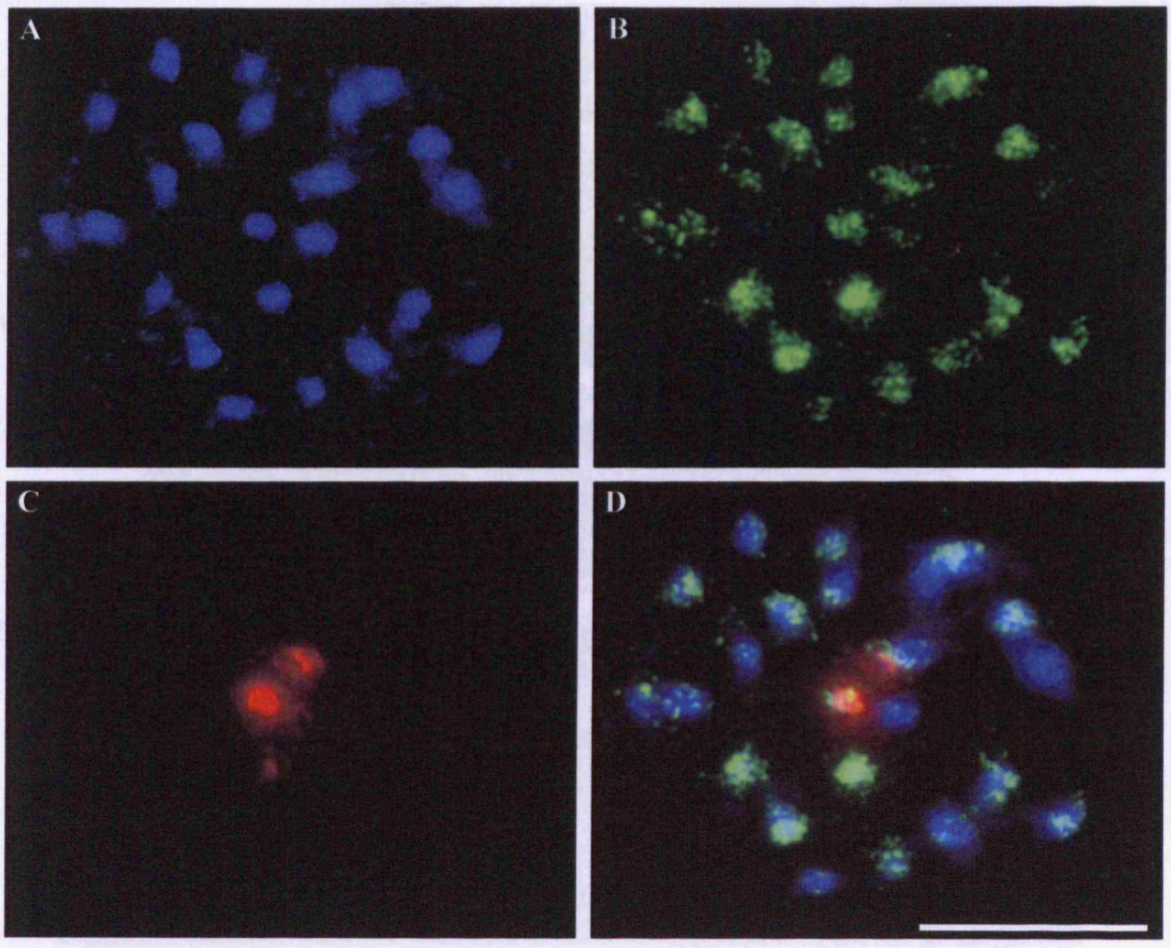


Figure 6.10: *In situ* hybridization of the MuTR repeat, clone MoTR1a and immunostaining of *M. balbisiana* 'Butohan 2' (BB) metaphase chromosomes with anti-5-methylcytosine antibody. The MuTR repeat was labelled with biotin-dUTP and detected with Alexa 594-streptavidin (red signal). The chromosomes were also incubated with anti-5-methylcytosine and detected with secondary antibody conjugated to FITC (green signal). **A** DAPI signal, **B** methylcytosine signal, **C** *In situ* hybridization signal, **D** Overlay plate of **A-C**. Bar = 10 μm.

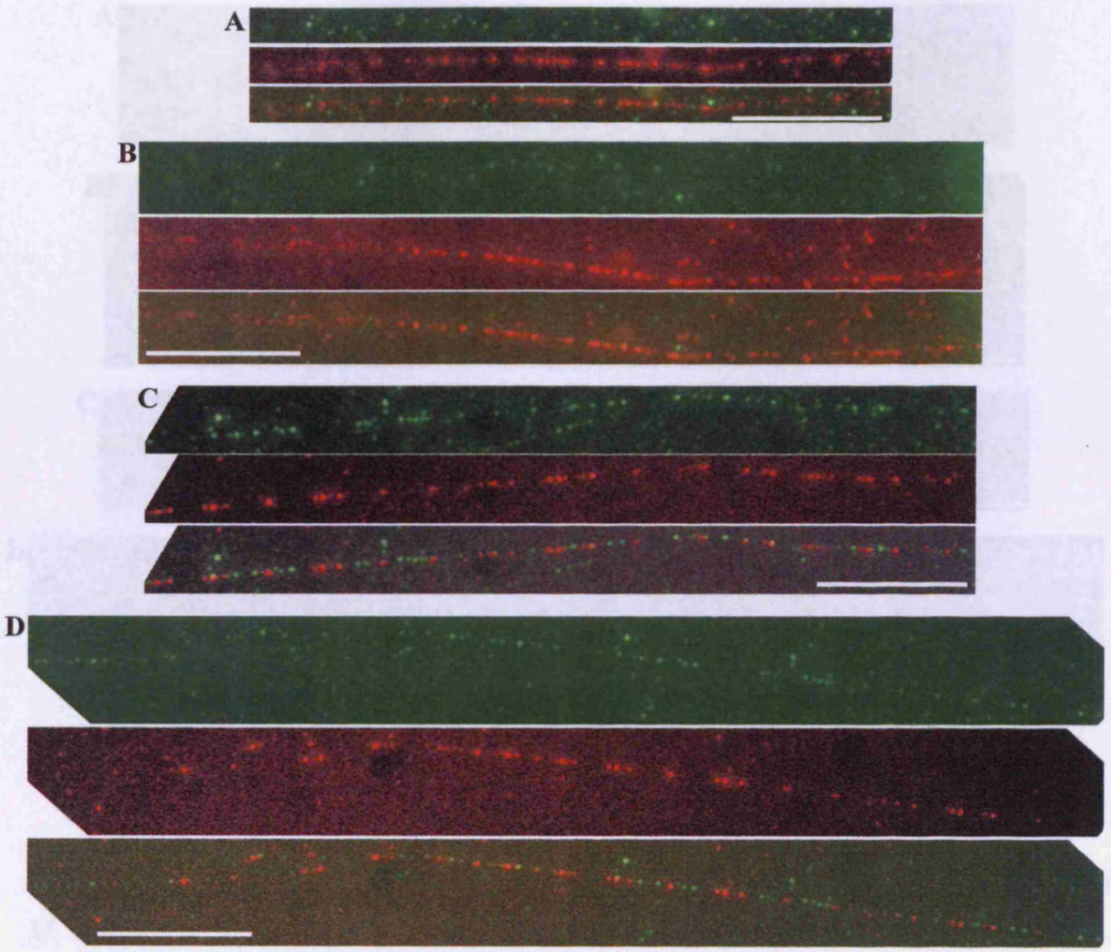


Figure 6.11: *In situ* hybridization of MuTR repeats, clone MoTR1a and immunostaining of *M. balbisiana* 'Butohan 2' (BB) low methylated extended DNA fibers with anti-5-methylcytosine antibody. For each fiber in **A, B, C, D** three pictures are shown. **Top** immunostaining signals of anti-5-methylcytosine antibody (green signal); **Middle** *in situ* hybridization signals of MuTR repeats (red signal); **Bottom** overlay plates of MuTR repeats and anti-5-methylcytosine antibody. Bar = 28.7 kbp/10 μm based on the extension degree of the DNA fibers prepared from *Arabidopsis thaliana* (Jackson *et al.*, 1998).

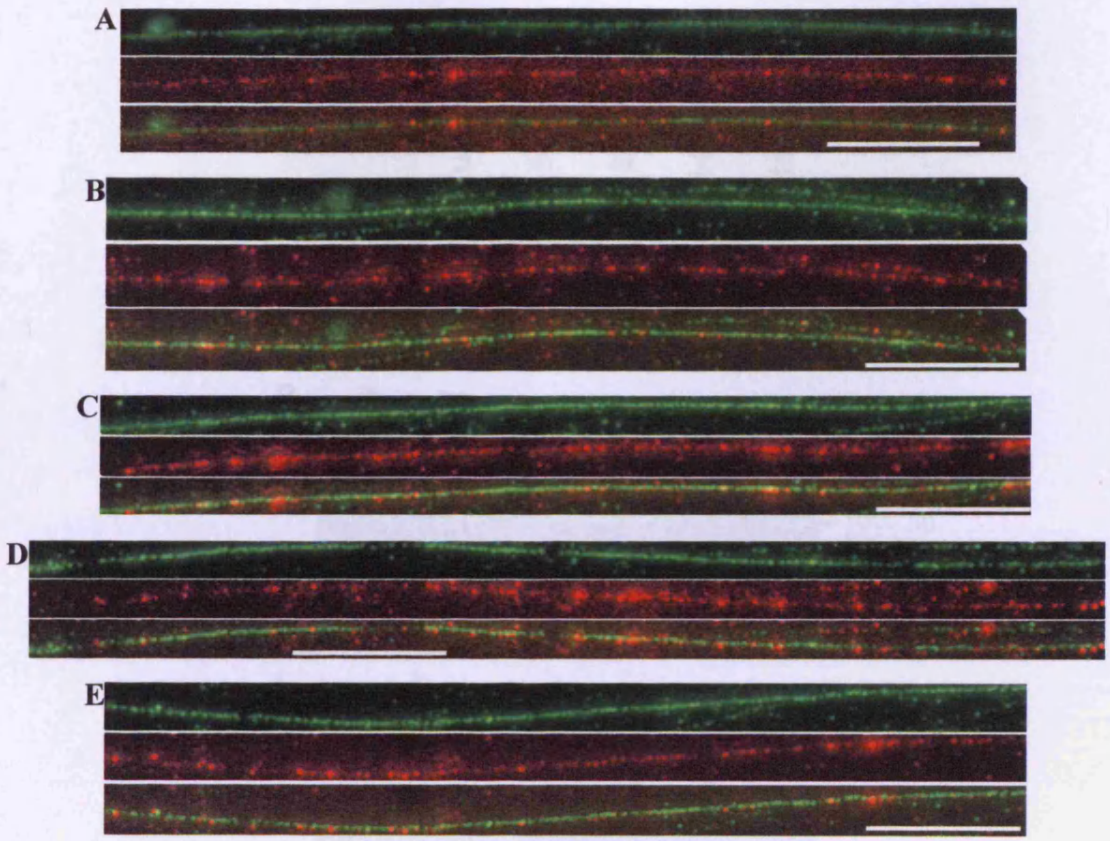


Figure 6.12: *In situ* hybridization of MuTR repeats clone MoTR1a and immunostaining of *M. balbisiana* 'Butohan 2' (BB) heavily methylated extended DNA fibers with anti-5-methylcytosine antibody. The MuTR repeat was labelled with biotin-dUTP and detected with Alexa 594-streptavidin (red signal). The extended DNA fiber was incubated with anti-5-methylcytosine and detected with secondary antibody conjugated to FITC (green signal). For each fiber in **A, B, C, D, E** three pictures are shown. **Top** immunostaining signals of anti-5-methylcytosine antibody (green signal); **Middle** *in situ* hybridization signals of MuTR repeats (red signal); **Bottom** overlay plates of MuTR repeats and anti-5-methylcytosine antibody. Bar = 28.7 kbp/10 μm based on the extension degree of the DNA fibers prepared from *Arabidopsis thaliana* (Jackson *et al.*, 1998).

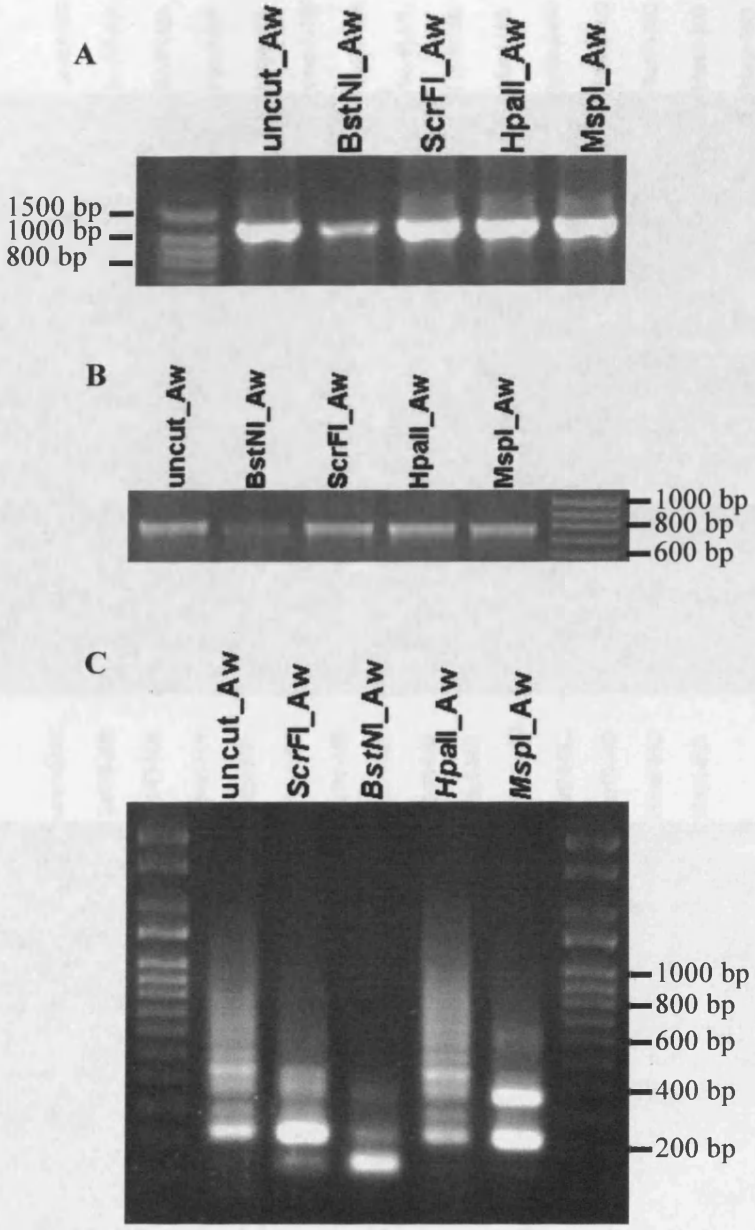


Figure 6.13: Methylation sensitive PCR analysis of *Metaviridae* Group III element and MuTR repeat on *M. paradisiaca* 'Pisang Awak' (ABB). **A** *Metaviridae* Group III element RT domain, primers: RTF + RTR; **B** *Metaviridae* Group III element LTRs, primers: LTRF + LTRR; **C** MuTR repeat, primers: MuTR1aF + MuTR1bR. Primer sequences are given in Chapter 3 for **A** and **B**, and in Table 4.1 for **C**.

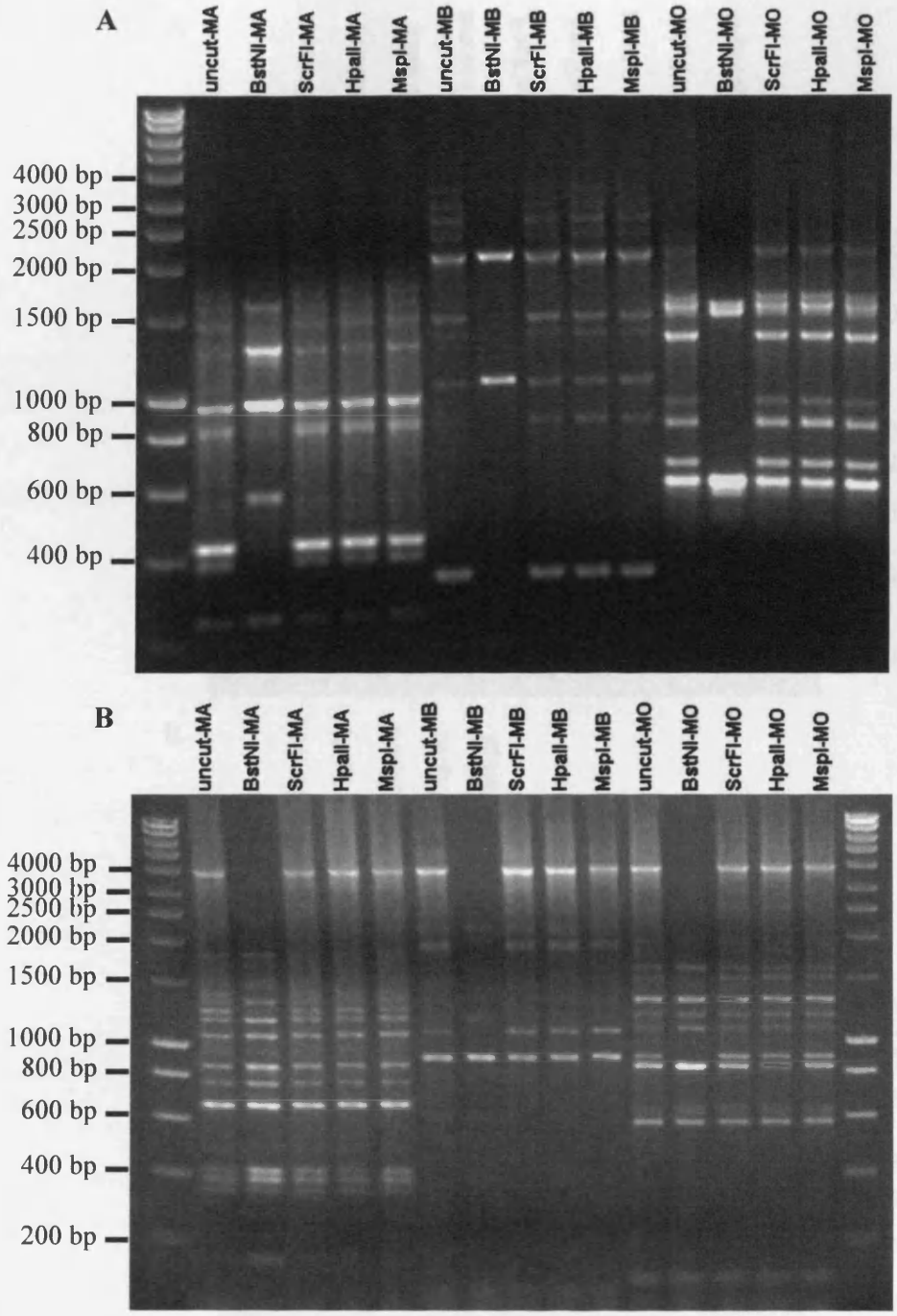


Figure 6.14: Genomewide methylation study of LTR retrotransposon flanking regions using outward LTR primers. **A** *Metaviridae* Group III elements, primer: GyLTRev (Nair and Teo *et al.*, 2005); **B** MuLARD elements, primers: BGLFor + BGLRev. MA = *M. acuminata* ssp. *malaccensis*, MB = *M. balbisiana*, and MB = *M. ornata*. Primer sequences are given in Table 5.1.

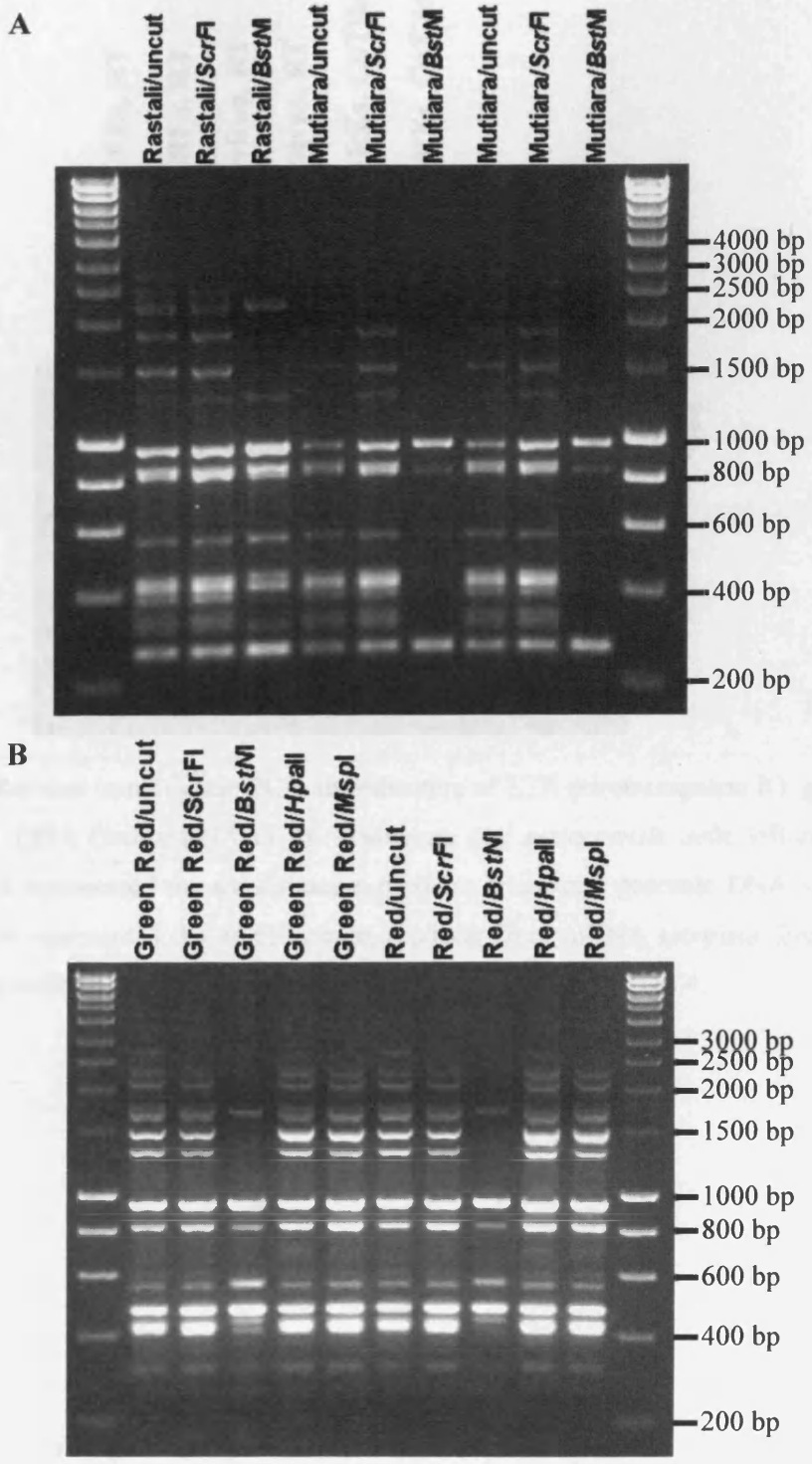


Figure 6.15: Methylation sensitive PCR analysis of *Metaviridae* Group III elements flanking regions on somaclonal variants. **A** disease resistance type, *M. paradisiaca* 'Pisang Rastali' (mother plant) and 'Pisang Mutiara'; **B** fruit color type, *M. acuminata* 'Green Red' and 'Red' (mother plant). Primers used = GyLTRev (Nair and Teo *et al.*, 2005; in Table 5.1)

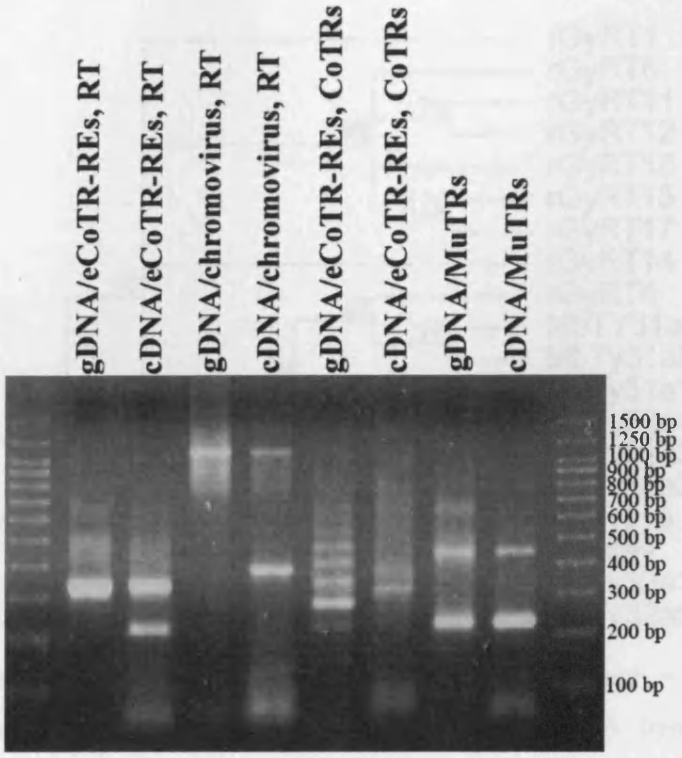


Figure 6.16: Reverse transcriptase-PCR amplification of LTR retrotransposon RT gene and tandem repeat DNA from mRNA of *M. acuminata* ssp. *malaccensis* male inflorescence. Line 2, 4, 6, 8 represented the amplification products from total genomic DNA template. Line 3, 5, 7, 9 represented the amplification products from mRNA template. For primer sequence and condition, see PCR amplification chapters in Chapter 3 and 4.

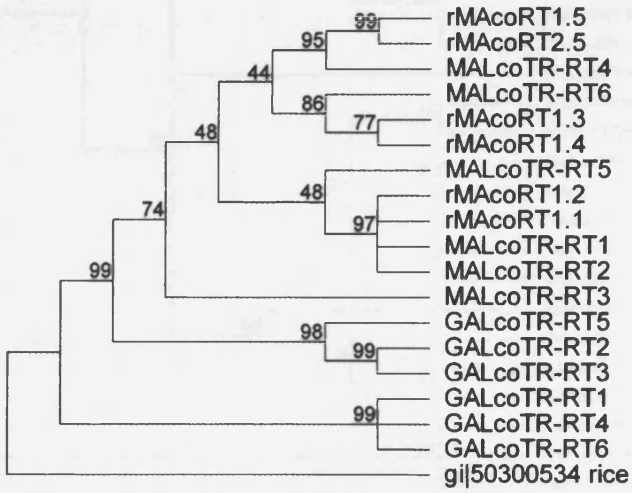


Figure 6.17: Rooted strict consensus tree of 18 MP tree (tree length = 399), after 1000 bootstrap replicates, of 6 *Metaviridae* Group II RT gene RNA transcripts from *M. acuminata* ssp. *malaccensis* male inflorescence mRNA (rMAcoRT) and 6 genomic clones from *M. acuminata* ssp. *malaccensis* (MALcoTR-RT) and *M. balbisiana* 'Pisang Gala' (GALcoTR-RT). The *Oryza sativa Metaviridae* element (gi|50300534) was used as out-group species.

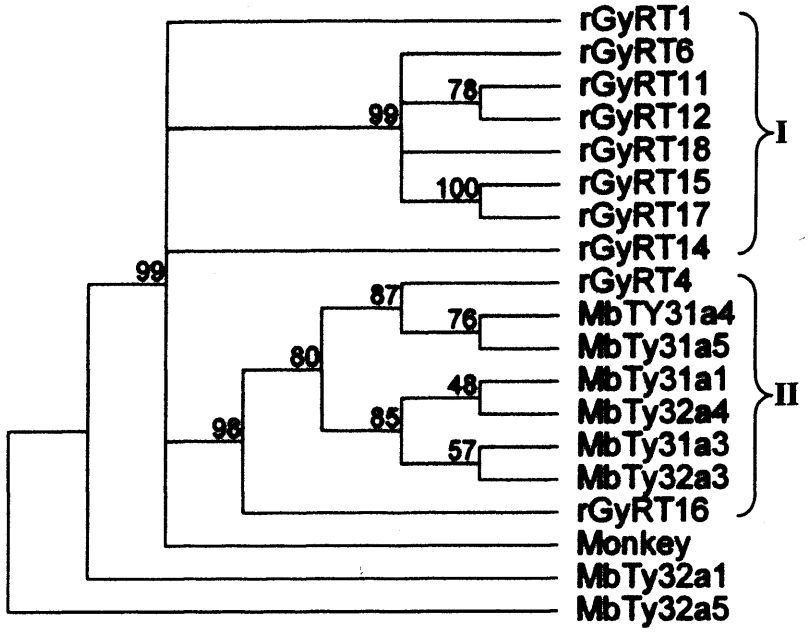


Figure 6.18: Un-root strict consensus tree of 8 MP trees (tree length = 805), after 1000 bootstrap replicates, of 9 *Metaviridae* Group III RT gene RNA transcripts from *M. acuminata* ssp. *malaccensis* male inflorescence mRNA (rGyRT) and genomic clones, *Monkey* retrotransposon, from *M. acuminata* 'Grand Nain' (1 clone) and *M. balbisiana* 'Pisang Gala' (MbTy31a & MbTy32a; 8 copies).

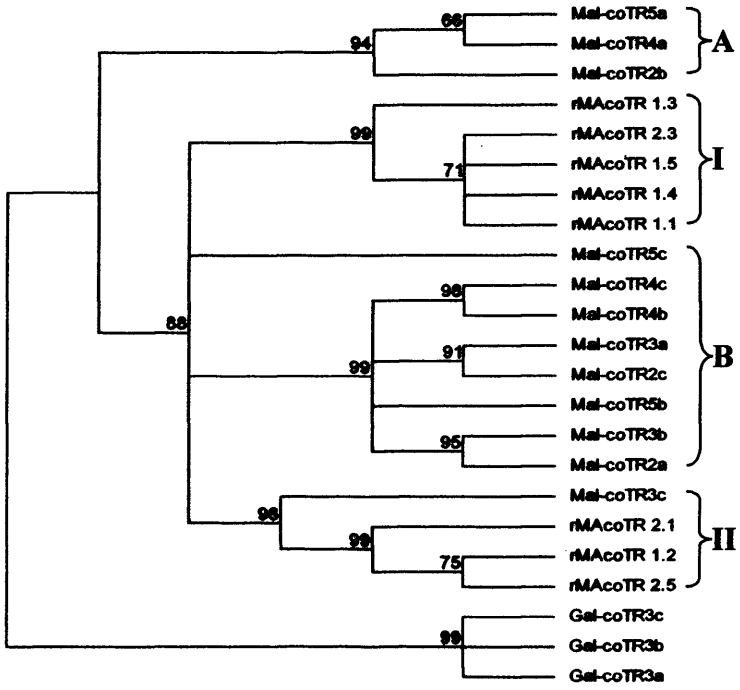


Figure 6.19: Rooted strict consensus tree of 583 MP trees (tree length = 361), after 1000 bootstrap replicates, of 8 CoTR RNA transcripts from *M. acuminata* ssp. *malaccensis* male inflorescence mRNA (rMAcoTR) and 12 genomic clones. Three genomic clones from *M. balbisiana* 'Pisang Gala' were used as out-group.

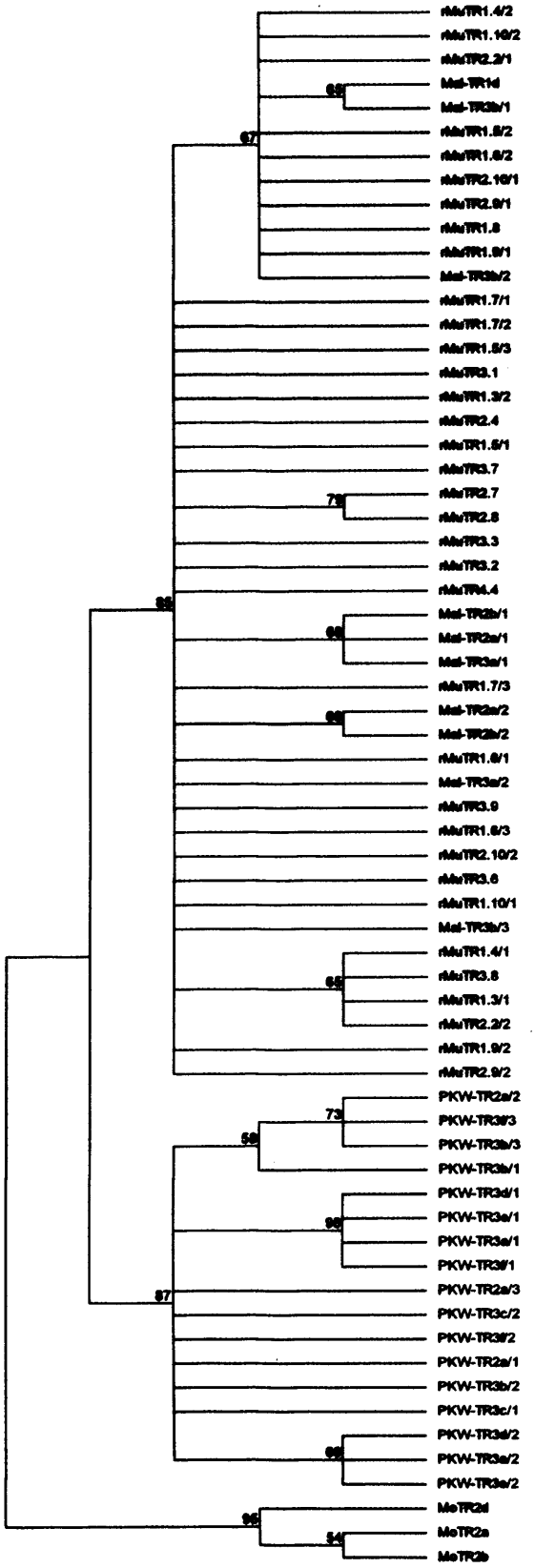


Figure 6.20: Rooted strict consensus tree of 1067 MP trees (tree length = 140), after 1000 bootstrap replicates, of MuTR RNA transcripts from *M. acuminata* ssp. *malaccensis* male inflorescence mRNA (rMuTR) and genomic clones. Genomic clones from *M. ornata* were used as out-group.

7.0 Chapter VII: General Discussion

7.1 LTR retrotransposons

Using molecular cloning, data-mining and cytogenetic techniques, a comprehensive picture of LTR retrotransposons and their behaviour within the *Musa* genome was established (Chapters 3, 5 and 6). Four classes of retrotransposons were identified within *Musa*. They comprise classical elements in the *Pseudoviridae* and *Metaviridae* that both contain the complete set of retrotransposon polyprotein sequences (GAG, AP, INT, RT and RH). Two new *Musa* elements were detected; LARD-like and *Morgane*-like elements that lack retrotransposon ORFs or contain only parts of them.

7.1.1 Comparison between *Pseudoviridae* and *Metaviridae*

The classification of retroelements (see Figure 1.3) puts the classical LTR retrotransposons, *Pseudoviridae* and *Metaviridae*, into two distinct families based on the sequence of the genes in the POL region (see Figure 1.2). Sequence and chromosomal distribution analysis carried out in this study also revealed clear differences between these two families although many similarities were found.

Pseudoviridae and *Metaviridae* elements share similar locations within the chromosomes at the centromeric heterochromatin of the *Musa* chromosomes (see Figures 6.2; 6.3; 6.7 and 6.9; see also Figure 7.3a), but *Pseudoviridae* are absent from the NOR region and show more sites dispersed throughout the euchromatic part of the chromosomes. *Metaviridae* are amplified in the NOR region and together with other repeats form a complex organisation at the NOR (Figure 7.3b).

Copy numbers of *Pseudoviridae* are probably higher within *Musa* genomes than *Metaviridae* elements (see section 5.4.1 and Table 5.2), and both elements show

great DNA diversity (Chapter 3). Although the LTR retrotransposon gene domains are well conserved, their replication is very error-prone due to the lack of proofreading repair activity by RNA polymerase and reverse transcriptase (Preston, 1996; Boutabout *et al.*, 2001). These result in individual copies with high sequence variability, mutations, premature stop codons and/or frameshifts, which were observed in both *Metaviridae* and *Pseudoviridae* (see Chapter 3). This variation together with insertion polymorphism can be exploited for molecular markers to distinguish species, varieties and hybrids within *Musa* (see section 5.5.2 and later section 7.1.4).

The analysis of full-length element (FLE) sequences of both *Metaviridae* and *Pseudoviridae* elements allowed the observations of several interesting features (Figure 3.2). The LTR regions of both elements are extremely variable in length and structure from one element copy to the next, but the two LTRs from a single element are very similar to each other sometimes even identical. These high similarities were interpreted as a sign of recent transposition (Kumar and Bennetzen 1999) and indicating their transposition ability. Minisatellite sequences of variable length and repeat number were observed at different positions within the LTR and also the 3'UTS (colour boxes in Figure 3.2) and are likely to be responsible for the length variations seen.

7.1.2. Diversity of *Metaviridae* elements

Metaviridae elements (Figure 3.3), based on the amino acid sequence homology of RT/RH genes, were subdivided into three groups. These three groups were analysed in detail revealing significant differences between them. Group I and II elements are part of the genus *Metavirus* and do not contain an envelope gene.

Within *Metaviridae* Group I, many degenerate or truncated elements were isolated (Figure 3.4b) that were derived through mutation and recombination events (see section 7.1.3) and might also be due to the nested insertions of LTR retrotransposons into each other, as observed in other plant species (e.g. SanMiguel *et al.*, 1996; Wicker *et al.*, 2001; 2003). These degenerate and nested insertions are unlikely to be active. *Metaviridae* Group II elements include an internal tandem repeat, CoTR, within the 3'UTS (Figure 3.1b) and the copies isolated here showed little variation (Figure 3.6). RNA transcripts of Group II elements were found, but only if they contained less units of the CoTR repeat (Figure 6.16 and 6.17), indicating that they can be active.

A CHROMO domain found in *Metaviridae* FLE sequences identifies the third group of *Metaviridae* (Group III elements, Figure 3.13) and I suggest classifying them as the genus *Chromovirus* (Figure 1.2), based not only on the presence of the CHROMO domain as was suggested by the classification of Marín and Lloréns (2000) but also on account of their RT/RH domain (Figure 3.3). The CHROMO domains are present in various eukaryotic proteins involved in chromatin remodelling and the regulation of gene expression (Eissenberg, 2001). In addition, they may function as protein-interacting, RNA-interacting or DNA-binding modules (Marmorstein, 2001). Malik and Eickbush (1999) suggested that the CHROMO domains could help to target the chromoviruses to specific chromatin regions in particular centromeres. In *Musa*, Group III elements are located at the centromere and also the NOR region (Figures 6.3 and 6.7) and may play an important role in chromatin remodelling of these regions. A further characteristic of Group III elements is that many solo-LTRS were detected in *Musa* species (Figure 5.1). The formation of the solo-LTRs is thought to result from non-reciprocal recombination

between the LTRs of a single LTR retrotransposon (Devos *et al.*, 2002; Vitte and Panaud, 2003).

7.1.3 Other types of LTR retrotransposons

Both a *Pseudoviridae* and *Metaviridae Morgane*-like element were identified (Figures 3.1d; 3.2h and i) showing only remnants of the POL region and several repeats within them. On the other hand, LARD-like elements without functional ORFs were detected in *Musa* BAC clones (Figure 3.1c) and FISH analysis revealed their similar chromosomal organization to *Metaviridae* Group III elements (Figures 6.3 and 6.4). Several mechanisms have been proposed for deriving new types of LTR retrotransposon such as *Morgane* elements (Sabot *et al.*, 2006; Tanskanen *et al.*, 2007) are derived from the classical LTR retrotransposons (Vicient *et al.*, 2005; Vitte and Panaud, 2005; Sabot and Schulman, 2007). Retrotransposons can be subject to four forms of recombination: template-switching during reverse transcription, integrase-catalyzed integration of one element into another, LTR-LTR recombination, allelic recombination and gene conversion between homologous chromosomes (Vicient *et al.*, 2005). Vitte and Panaud (2005) proposed that recombination mechanisms that occur between direct repeats anywhere in the sequence of the classical LTR retrotransposons will lead to the internal deletions of these elements and subsequently, result in the formation of new types of elements. In addition to the recombination mechanism, Sabot and Schulman (2007) have proposed another potential mechanism that involves abnormal template switching between two different LTR retrotransposon RNAs by inter-strand pairing during reverse transcription to generate a chimeric cDNA. This mechanism has been proposed for the generation of *Veju*_L (Sabot *et al.*, 2005) and *BARE-2* (Vicient *et*

al., 2005; Tanskanen *et al.*, 2007), and might also possible for the LARD elements (Kalendar *et al.*, 2004; Figure 3.1c). The detection of putative active *Metaviridae* Group II elements (Figures 6.16 and 6.17) together with the conserved sequence features shared between eCoTR-Res and *Meta*Morgane (Figures 3.1b and d), provide some indication that the template switch mechanism proposed by Sabot and Schulman (2007) might be a more likely explanation for the generation of this *Meta*Morgane element in *Musa* species. The RNA transcripts of the *Metaviridae* Group II elements might be able to form a template-switch complex by inter-strand pairing between two different RNA transcripts and lead to the creation of the *Meta*Morgane elements.

7.1.4 Insertional polymorphism and molecular marker application

The insertion locations of plant LTR retrotransposons have been shown to be highly variable (Waugh *et al.*, 1997; Ellis *et al.*, 1998; Kalendar *et al.*, 1999; Porceddu *et al.*, 2002), since they insert into genes, retrotransposons and other repetitive DNA (Garber *et al.*, 1999; Kumekawa *et al.*, 1999; Parniske and Jones, 1999; Miyao *et al.*, 2003; Witte *et al.*, 2005; Yang *et al.*, 2005). The insertion of LTR retrotransposons into genes might change the expression level of the gene in which they are inserted (Marillonnet and Wessler, 1997; Sunako *et al.*, 1999; Nishimura *et al.*, 2000; Selinger and Chandler, 2001; Yao *et al.*, 2001; Zufall and Rausher, 2003; Clark *et al.*, 2006; Lu *et al.*, 2006). In this study, *Musa Metaviridae* and *Pseudoviridae* LTR retrotransposons were shown to insert into other LTR retrotransposons (Table 5.1; Figure 5.10), 5S rDNA sequences (Figure 4.3), MuTR repeats (Table 5.1; Figure 5.8), Class III acidic chitinase genes (AY525367) (Figure 5.14) and other repetitive DNAs (Figure 3.4b; Figure 5.12). Further analysis of LTR

retrotransposon flanking sequences allowed the screening of *Musa* germplasm using PCR primers designed from retrotransposons and their flanking sequences to detect the insertion events of these elements (Figures 5.9; 5.11; 5.13 and 5.15).

The insertion of LTR retrotransposons is permanent and can eventually cause genomic diversification which facilitates phylogenetic analyses (Shimamura *et al.*, 1997). The highly polymorphic insertion properties of LTR retrotransposons allow the development of retrotransposon-based molecular markers to study the diversity of different plant species (Rodhe *et al.*, 1995; Rodhe, 1996; Flavell *et al.*, 1998; Kalendar *et al.*, 1999; Pearce *et al.*, 1999; Queen *et al.*, 2004; Nair and Teo *et al.*, 2005; Teo *et al.*, 2005; Bousios *et al.*, 2007; Smýkal, 2006, Häkkinen *et al.*, 2007). In *Musa* species, the application of LTR retrotransposon-based molecular markers allows identification and classification of *Musa* genomes. The PCR-RFLP markers that exploit the retrotransposon internal domain sequence polymorphism, clearly classified 19 *Musa* species into four sections (Figure 5.7) and complement classical morphological classifications. IRAP markers which are based on LTR retrotransposon polymorphic insertion patterns, clearly identified the B genome in *Musa* taxa tested (Figures 5.3 to 5.5, 5.9 and Nair and Teo *et al.*, 2005) and showed that these markers were inheritable (Figure 5.6), which will help banana breeders in distinguishing the synonymous cultivars/varieties and inter- or intra-species hybrids.

7.1.5 Evolution of LTR retrotransposons

Previous studies based on the sequence homology and functions of RNaseH domains between non-LTR retrotransposons and LTR retrotransposons strongly suggests that the LTR retrotransposons are evolved from a younger lineage of non-LTR retrotransposons that acquired the GAG and aspartic protease domains from the host and transposase (TP) that subsequently evolved to integrase (INT) from DNA

transposons (Malik and Eickbush, 2001). However, the exact mechanism how non-LTR retrotransposons can acquire the LTRs during the evolution process to form the LTR retrotransposons are still unknown. Here, I support the evolutionary hypothesis of the acquisition of LTRs from non-LTR retrotransposons based on previous published plant studies and my findings of retrotransposon diversity and organisation in *Musa* (Figure 7.1).

LTR retrotransposons that do not encode any functional ORFs have been discovered in wheat and barley (TRIM, Witte *et al.*, 2001; LARD, Kalendar *et al.*, 2004) and were found within *Musa* in my study (Figure 3.1c). The cereal LARD elements have been shown to share high DNA level homology with different *Metaviridae* elements, *Erika*-1 and *Bagy*-1 (Wicker *et al.*, 2001; Kalendar *et al.*, 2004). The Frodo class of plant TRIM elements has been proposed to originate from 5S rRNA gene, mobile RNA and retrotransposons due to the high homology observed between the LTRs of the TRIM elements with the 5S rRNA (Kalendar, 2006). On the other hand, different types of classical LTR and non-LTR retrotransposon have been shown to insert within the internal domain of TRIM (Witte *et al.*, 2001). Furthermore, I have shown that LTR retrotransposons have been inserted into the tandem repeat DNA sequences, the 5S rDNA sequences (Figure 4.3) and the MuTR repeats (Table 5.4, section 7.2) as has also been observed in many other studies (Nakajima *et al.*, 1996; Hisatomi *et al.*, 1997; Kapitonov and Jurka, 2003; Kumekawa *et al.*, 2001; Cheng *et al.*, 2002; Tek *et al.*, 2005; Cheng and Murata, 2003). These evidences suggest that these different types of repetitive DNAs are closely related to each other in *Musa* and are also physically located within the same chromosomal regions of the centromeres and the NOR (see Chapter 6 and Figure 7.3). Hence, non-LTR retrotransposons can transpose into the LARD-like or

TRIM-like elements and even tandem repeat DNA which resulted in the formation of the older lineage of the *Pseudoviridae* elements, *Hemivirus* or *Pseudovirus* (Figure 7.1), classified by Peterson-Burch and Voytas (2002)(see also Figure 1.2). The endonuclease (EN) domain of non-LTR retrotransposons, that cleaves the target DNA to generate a primer for reverse transcription (called target-primed reverse transcription), is postulated to be lost or replaced during the evolution process by transposase (TP) that subsequently evolved into INT (Figure 7.1). Such gradual addition and replacement of domains has been observed in the non-LTR retrotransposons (Malik and Eickbush, 2001).

The genus *Hemivirus* or *Pseudovirus* evolved into the genus *Sirevirus* by acquiring the envelope (ENV) domain from the other viruses (Figure 7.1). Malik *et al.* (2000) proposed that a non-viral ancestor to the genus *Errantivirus* of *Metaviridae* elements acquired the ENV domain from other viruses such as baculoviruses. However, it is well know that *Pseudoviridae* elements and retroviruses are more ancient than the *Metaviridae* elements (Malik and Eickbush, 1999; 2001; Marín and Lloréns, 2000; Gorinšek *et al.*, 2004; Hansen and Heslop-Harrison, 2004). This leads to a possibility that the ENV domain might already been acquired by the genus *Sirevirus* of the *Pseudoviridae* elements long before the emergence of the *Metaviridae* elements (Figure 7.1).

The genus *Sirevirus* subsequently evolved into a retrovirus by switching the position of the INT domain from downstream of the RT/RH domains to the upstream of these domains (green arrow, Figure 7.1). The switching of these gene domains are possibly due to the infectious nature of the retroviruses and they are only found in the animal kingdom. The retrovirus lineage eventually gave rise to the formation of the *Errantivirus* lineage and followed by the *Metavirus* and *Semotivirus* lineages that lost

the ENV domain, based on the phylogenetic results by Malik and Eickbush (1999; 2001), Marín and Lloréns (2000) and Gorinšek *et al.* (2004)(Figure 7.1).

7.1.6 DNA methylation and transcription

In plant genomes, cytosine methylation is most often found in symmetrical CpG and CpNpG contexts as well as in non-symmetrical CpHpH (H=C, A, T) (Gruenbaum *et al.*, 1981; Meyer *et al.*, 1994; Grafi *et al.*, 2007). Moreover, most of these residues are found in the heterochromatin of plants, where abundant repetitive DNAs are detected (Bender, 2004, Grafi *et al.*, 2007). Methylation sensitive PCR of *Metaviridae* Group III element RT and LTR sequences found no symmetrical CpG sites while most were methylated in their symmetrical CpNpG sites, but some were not (Figure 6.13). Immunostaining of *Musa* metaphase chromosomes with anti-5-methylcytosine antibody showed that the majority of LTR retrotransposons (*Metaviridae* Group III element, Figure 6.7; LARD-like element, Figure 6.8; and *Pseudoviridae* element, Figure 6.9) were located in the heavily methylated heterochromatic regions, centromeres and/or NORs, although some elements were found in low methylated genomic regions. The detection and isolation of heterogeneous populations of *Metaviridae* Group II and III element RNA transcripts (Figure 6.16) in *Musa acuminata* and the identification of active and inactive groups within each retrotransposon class (Figures 6.17 and 6.18) could link transcription to the non-methylated genomic clones. DNA methylation studies of LTR retrotransposons in different organisms reported that most of the silenced elements are methylated (Hirochika *et al.*, 2000; Lyko *et al.*, 2000; Zhou *et al.*, 2001; Burgers *et al.*, 2002; Hashida *et al.*, 2003; Lippman *et al.*, 2003; Kato *et al.*, 2004; Liu *et al.*, 2004; Verdel *et al.*, 2004; Kuhlmann *et al.*, 2005).

Studies on *Musa* somaclonal variants (Figure 6.15) provide another perspective to the DNA methylation of LTR retrotransposons. Activation of LTR retrotransposons under stress conditions such as wounding, pathogen attack and the tissue culture process has been observed in plants (Peschke *et al.*, 1987; Grandbastien *et al.*, 1989; Peschke and Phillips, 1991; Hirochika, 1993; Hirochika *et al.*, 1996a, b; Wessler, 1996; Mhiri *et al.*, 1997; Grandbastien, 1997; Liu *et al.*, 2004) and these activations can subsequently generate mutations (Wessler *et al.*, 1995; Hirochika *et al.*, 1996a, b). The tissue culture process has been shown to generate somaclonal variation in plant by inducing global or sequence-specific hypomethylation of the DNA (Jaligot *et al.*, 2000; Matthes *et al.*, 2001; Tregear *et al.*, 2002; Kubis *et al.*, 2003, Hirochika *et al.*, 1996a, b; 2000). In *Musa*, symmetrical CpNpG methylation variation and more importantly DNA sequence mutation to introduce more CpNpG sites (Figure 6.15) were observed at the *Metaviridae* Group III element flanking sequences in the disease resistant somaclonal variants (*M. paradisiaca* 'Pisang Mutiara' (AAB), Figure 6.15A) but not in the fruit colour somaclonal variant (*M. acuminata* 'Green Red' (AAA), Figure 6.15B). This might play an important role in the creation of disease resistant somaclonal variants, although these types of variation could be genotype dependent (Nielsen *et al.*, 2001; Liu *et al.*, 2004).

7.2 Tandemly organized DNA

Three classes of tandem repeats were identified in *Musa* that are all connected with LTR retrotransposons. They either have a LTR retrotransposon inserted within their repeat unit, or are themselves inserted within LTR retrotransposons. LTR retrotransposons and tandem repeats are also found in the same chromosomal region

of *Musa* genomes. This has implications for the amplification and evolution of repeats within *Musa* and models are discussed in the section 7.3.

7.2.1 Diversity and organization of tandem repeats

The MuTR family, which consists of two subrepeats (MuTRa and MuTRb, Figures 4.6 to 4.8), from three *Musa* species (*M. acuminata* ssp. *malaccensis*, *M. balbisiana* 'Pisang Klutuk Wulung' and *M. ornata*) showed higher intra-species DNA sequence homology (87-100%) compared to inter-species homology (86-93%) and were present in all the *Musa* species examined (Figure 4.12). The high intra-species homology enables the identification of species-specific MuTR repeats by phylogenetic analysis (Figures 4.10 and 4.11). DNA sequence analysis identified a 63 bp box in all the MuTR repeats (Figures 4.6A to 4.8) with more than 90% homology to the LTR of the *Metaviridae* Group III element (genus *Chromovirus*). This suggests that a *Metaviridae* Group III element had been inserted into MuTR repeats during its evolution (Figure 4.20). In other plant species, the insertion of LTR retrotransposons into tandem repeat DNA has been reported (Hisatomi *et al.*, 1997; Kumekawa *et al.*, 2001) and it is not uncommon for these inserted sequences to become part of the repeat sequences (Cheng and Murata, 2003; Yang *et al.*, 2005; Tek *et al.*, 2005, Li and Leung, 2006).

Part of a *Pseudoviridae* LTR retrotransposon was found inserted into some copies of the 5S rDNA in *Musa* (Figures 4.3 and 4.19), with a higher incidence in the A genome than the B genome. The location of the retrotransposon sequence immediately downstream of the 5S coding region in antisense orientation could indicate that they are involved in the regulation of RNA expression (see section

4.5.1) similar to pol III promoters (Paule and White 2000) or various repeats inserted within rRNA genes reported (see Kojima and Fuijiwara 2003; Eickbush 2002).

Another type of relationship between tandem repeat DNA and LTR retrotransposons was observed in the *Metaviridae* Group III element, eCoTR-Res (Figure 3.1b). In this element, a long stretch of ~ 60 bp repeated sequence was observed near the 3' LTR (Figures 3.1b and 3.2g). Further studies on CoTR repeats led to the identification of high order organization of these repeats in *Musa acuminata* and *M. balbisiana* genome (Figure 4.18) and species-specific repeats (Figure 4.17). Transposable elements, both retrotransposons and DNA transposons, that contain tandemly organized repeats of various length in their internal domain have been detected in other species (Garrett *et al.*, 1989; Noma and Ohtsubo, 2000; Chao *et al.*, 2003; Vicient *et al.*, 2005). Recently, a family of autonomous *Metaviridae* elements that harbours tandem repeat DNA (*Retand* family) has been reported in *Silene* species and the tandem repeat region of these elements is more amplified in the non-autonomous elements (Kejnovsky *et al.*, 2006). A similar situation is predicted for CoTR repeats in the *Musa* genome, where multiple PCR bands arranged in tandem arrays were observed in the genomic DNA whereas only two bands were detected in the cDNA (Figure 6.16). This suggests that the CoTR repeats found in RNA copies of *Metaviridae* Group II element are more amplified in their genomic copies.

The homogenization and amplification of tandem repeat DNA in one species, which subsequently generated genome or species-specific repeats and variants has been reported in other plant species (Rayburn and Gill, 1986; Anamthawat-Jónsson and Heslop-Harrison, 1993; Vershinin *et al.*, 1994; Nagaki *et al.*, 1995, 1998a, b; Kishii *et al.*, 1999; Li *et al.*, 2000; Kishii and Tsujimoto, 2002; Heslop-Harrison,

2003; Heslop-Harrison *et al.*, 2003; Rudd *et al.*, 2006). In the *Musa* genome, the process of homogenization, amplification, and mutation of MuTR (Figure 4.9) and CoTR (Figure 4.16) repeats has allowed the generation of species-specific repeats/variants, which can be detected by phylogenetic analysis (Figures 4.10 and 4.11 for MuTR repeats and Figure 4.17 for CoTR repeats).

7.2.2 Evolution and amplification of tandem repeats

FISH experiments using MuTR repeats showed their consistent localization at the NOR loci of all the species tested with small variation in signal intensities (Figure 6.5). Additionally MuTR repeats were located at the centromeres of some *Musa* species (Figure 6.5). One possibility is that MuTR repeats originated and amplified in the NOR loci and then jumped to the centromeres during *Musa* evolution by rolling cycle or transposition mechanisms. Rolling cycle mechanisms have been proposed for the replication of tandem repeat DNA by the formation of an extrachromosomal DNA followed by DNA reintegration into the host genome (Okumura *et al.*, 1987; Cohen *et al.*, 1999; 2003; 2005; Krzywinski *et al.*, 2005; Tek *et al.*, 2005) and no RNA transcript is generated during this process. However, RNA transcripts of MuTR repeat were detected in this study (Figure 6.16). In addition, the detection of the 63 bp box in the MuTR repeats (Figures 4.6 to 4.8), and the co-localization of *Metaviridae* Group III elements and MuTR repeats at the NOR loci (Figure 6.3) and also in the extended DNA fibres (Figure 6.6) provide some indication that *Metaviridae* Group III elements might be involved in the transposition and amplification of MuTR repeats. One possibility is that the MuTR repeats at the NOR loci are transcribed by the host RNA polymerase II that is generated by the RNA transcripts with poly(A) tails at their 3' end, which is then followed by the

reverse-transcription of these transcripts with the help of LTR retrotransposons reverse transcriptase, in this case, *Metaviridae* Group III elements to generate the cDNA copies of these repeats (Figure 7.2). This is followed by the insertion of the cDNA copy into the centromeres of *Musa* species, either facilitated by the integrase activity of *Metaviridae* Group III elements or by homologous recombination between the 63 bp box and the LTR of centromeric *Metaviridae* Group III elements (Figure 7.2). Once the MuTR repeats have transposed to the centromeres, they can be further amplified by the rolling cycle mechanism, which has been reported to generate large array sizes over short periods of time (Lo *et al.*, 1999; Krzywinski *et al.*, 2005).

7.2.3 DNA methylation and transcription

Immunochemical analysis showed different levels of cytosine methylation for MuTR repeats located in the NOR loci of *Musa balbisiana* 'Butohan 2' chromosomes (Figure 6.10), where the active part of the NOR contains weakly methylated MuTR repeats, while the distal part of the NOR region and the chromosomal satellite beyond contains heavily methylated MuTR repeats. The extended fibre studies confirms two populations of MuTR repeats, those that form large arrays and are heavily methylated and those that are in shorter arrays and weakly methylated (Figures 6.11 and 6.12). The methylation sensitive PCR showed that some of the MuTR repeats are not methylated at all (Figure 6.13C). Interestingly if MuTR is methylated, it occurs at both symmetrical CpG and CpNpG sites (Figure 6.12C), in contrast to the *Metaviridae* Group III elements (Figures 6.13A and B) which are only methylated at their symmetrical CpNpG contexts.

The observation of MuTR repeats located at the low methylated region of the NOR and the presence of non-methylated repeat units suggests that MuTR might be

actively transcribed in *Musa*. RNA transcripts of MuTR are indeed found (Fig 6.16) and we now consider their possible role. Tandem repeat DNA has been shown to be actively transcribed *in vivo* and *in vitro* in other organisms (Epstein and Pabón-Peña, 1991; Ferbeyre *et al.*, 1998; Rojas *et al.*, 2000; Hall *et al.*, 2002; Volpe *et al.*, 2002; 2003; Usakin *et al.*, 2007) and in plants (Nagaki *et al.*, 2003b; May *et al.*, 2005; Zhang *et al.*, 2005; Lee *et al.*, 2005; 2006; Yan *et al.*, 2006). The fact that repetitive DNA can be transcribed has changed the view that sees heterochromatin as a junkyard of silent non-coding DNA and defective transposons. There is also increased evidence that epigenetic phenomenon not only affect the regulation of genes, but are also a major factor in the prevention of expression (silencing) of repetitive sequences in particular tandem repeats and transposable and retrovirus related elements (Lippman & Martienssen 2004; Staginnus *et al.*, 2007). DNA sequences present in heterochromatin can give rise to small RNAs that, by means of RNA interference, direct the modification of proteins and DNA in heterochromatic repeats and transposable elements. Heterochromatin and the dynamics of chromosome structure have thus emerged as a key factor in epigenetic regulation of gene expression, chromosome behaviour and evolution.

7.3 The *Musa* genome

7.3.1 Chromosomal organization of different repetitive DNA in the *Musa* genome

A major portion of plant genomes is comprised of different types of repetitive DNA, that form distinct clusters, most often co-localized on the same region of the chromosomes, such as the centromeres, subtelomeric and rDNA loci (e.g. Kumar and Bennetzen, 1999; Heslop-Harrison, 2000a; Schwarzacher, 2003). Molecular

cytogenetic and physical mapping have generated models of plant chromosome and repetitive DNA organisation for the large grained cereals (Heslop-Harrison and Bennett, 1990, Schwarzacher, 1996), rice (Cheng *et al.*, 2007), sugar beet (Schmidt and Heslop-Harrison, 1998), *Arabidopsis* (Brandes *et al.*, 1997a; Maluszynska and Heslop-Harrison, 1991; Heslop-Harrison *et al.*, 1999; 2003; Kumekawa *et al.*, 2000; 2001) and tomato (Stack and Anderson, 2001). Here, I propose a model (Figure 7.3) for the chromosome organisation of *Musa* that represents an important sister group to the grasses within the monocots.

This model is based on the chromosomal organization of three LTR retrotransposons (*Pseudoviridae* elements, Figure 6.2; *Metaviridae* Group III elements, Figure 6.3 and LARD-like elements, Figure 6.4) and one tandem repeat DNA (MuTR repeat, Figure 6.5) found in this study, together with the rDNA and *Metaviridae* element loci reported in other *Musa* cytogenetic studies (Osuji *et al.*, 1997; 1998; Balint-Kurti *et al.*, 2000; Valárik *et al.*, 2002; Doleželová *et al.*, 1998; Lysák *et al.*, 1999; Doležel *et al.*, 1994, 2004; Bartoš *et al.*, 2005). The centromeric regions of the *Musa* chromosomes contain different types of repetitive DNA (LTR retrotransposons and MuTR repeats from this study; 5S rDNA sequences and Radka family; Valárik *et al.*, 2002), and both MuTR repeats and 5S rDNA sequences are more abundant in this region and more uniformly distributed compared to other repetitive DNA (Figure 7.3a). LTR retrotransposons are unevenly distributed throughout the chromosome arms and *Pseudoviridae* elements are present in higher copy number compared to the other two LTR retrotransposons (*Metaviridae* Group III elements and LARD-like elements). This is complemented by the quantitative analysis in this study, which showed that the *Musa* genome contains higher copy

numbers of *Pseudoviridae* elements compared to *Metaviridae* elements except for *Metaviridae* Group I elements (Tables 5.2 and 5.3).

In the model of the NOR region (Figure 7.3b), the 45S rDNA sequences show more dispersed distribution patterns on the active and distal part of the NOR locus and the chromosomal satellites based on previous studies (Osuji *et al.*, 1998; Valárik *et al.*, 2002; Doleželová *et al.*, 1998; Doležel *et al.*, 2004; Bartoš *et al.*, 2005), while the MuTR repeats are uniformly distributed in these regions. The more dispersed nature of the 45S rDNA sequences might be due to their low copy number in the *Musa* genome (16000 copies per haploid genome; Valárik *et al.*, 2002). The MuTR repeat loci at the chromosomal satellites are partly overlapped by the *Metaviridae* Group III elements and LARD-like element loci. As stated earlier, the co-localization of repetitive sequences in defined chromosomal regions have significant implications for their function and evolution (see sections 7.1.1 and 7.2.2).

7.3.2 Repetitive DNA and *Musa* genome size evolution

Repetitive DNA, mainly LTR retrotransposons, has been shown to play an important role in increasing the genome size of different plant species (SanMiguel *et al.*, 1996; Bennetzen and Kellogg, 1997; SanMiguel and Bennetzen, 1998; Kumar and Bennetzen, 1999; Vicient *et al.*, 1999; Shirasu *et al.*, 2000; Wicker *et al.*, 2001; Ma *et al.*, 2004; Ma and Bennetzen, 2004; Guo *et et al.*, 2006; Hawkins *et al.*, 2006; Peigu *et al.*, 2006). In this study, genome size increments, due to the amplification of LTR retrotransposons (*Pseudoviridae* and *Metaviridae* element) and MuTR repeats, has been observed in *Musa* species (Tables 5.2 and 5.3). Both *Pseudoviridae* and *Metaviridae* elements generally showed two-fold copy number increments in the *malaccensis* genome (587 Mbp/1C) compared to the 'Pisang Gala' genome (547

Mbp/1C)(Table 5.2), while *Metaviridae* Group II elements showed ~ 10-fold copy number increments in the *malaccensis* genome compared to the 'Pisang Gala' genome (Table 5.3). In addition, the increment in the chromosomal loci of MuTR repeats detected at the centromeres is correlated with the genome size increments of the species tested (Figure 6.5). Other repetitive DNA such as Radka family have also been shown to be present in higher copy number in the A genome (Valárik *et al.*, 2002).

On the other hand, the elimination of the LTR retrotransposons by various mechanisms, such as unequal homologous recombination between LTR sequences to generate the solo LTRs and the partial or total deletion of LTR retrotransposon through illegitimate recombination mechanisms, resulting in genome size reduction has been reported (Petrov, Lozovskaya and Hartl, 1996; Vicient *et al.*, 1999; Petrov *et al.*, 2000; Shirasu *et al.*, 2000; Devos, Brown and Bennetzen, 2002; Vitte and Panaud, 2003; Ma *et al.*, 2004; Vitte and Panaud, 2005; Guo *et et al.*, 2006). In *Musa*, solo-LTR formation of the *Metaviridae* Group III element has been observed in *malaccensis*. Furthermore, highly heterogeneous populations of the *Metaviridae* Group I elements with RT genes disrupted by other types of repetitive DNA are detected in *Musa* genomes (Figure 3.4b), and the illegitimate recombination mechanisms has been proposed for this type of LTR retrotransposon population in plant genomes (Vitte and Panaud, 2005).

All the evidence suggests that *Musa* species have undergone both genome size expansion by amplification of LTR retrotransposons and MuTR repeats, and reduction events by elimination of LTR retrotransposons since the appearance of the *Musa* common ancestor *circa*. 43 million years ago.

7.4 Application of repetitive DNA studies on *Musa* diversity and breeding program

Based on all the evidences presented in this work, it is concluded that repetitive DNA including both LTR retrotransposons (*Metaviridae*, *Pseudoviridae*, LARD-like and *Morgane*-like) and tandem repeat DNAs (MuTR and CoTR) are highly heterogeneous in *Musa* genome. The copy number variation of these repetitive DNAs in this work clearly showed their involvement in the expansion/reduction of the *Musa* genome during the evolution process which subsequently gave rise to the present-day *Musa* taxa with different genome sizes.

The application of LTR retrotransposons as molecular markers in this work allows the clear distinction of *Musa* sections, species and close-related varieties/cultivars. Furthermore, these retrotransposon-based markers are able to help *Musa* breeders in the screening of economic importance bananas in the early stage of *Musa* breeding program. This shows strong potential for the application of this marker system in *Musa* diversity, *ex situ* conservation and breeding studies.

Molecular cytogenetic and immunochemical aspects in this work showed distinct chromosomal localization of different repetitive DNA in *Musa* genomes and also the relationship of these repetitive DNAs when they occupy the similar chromosomal locations such as the centromeres and NOR. This provides the initial step for future cytogenetic studies such as the chromosomal evolution and mapping, and also the chromatin profiling of *Musa*.

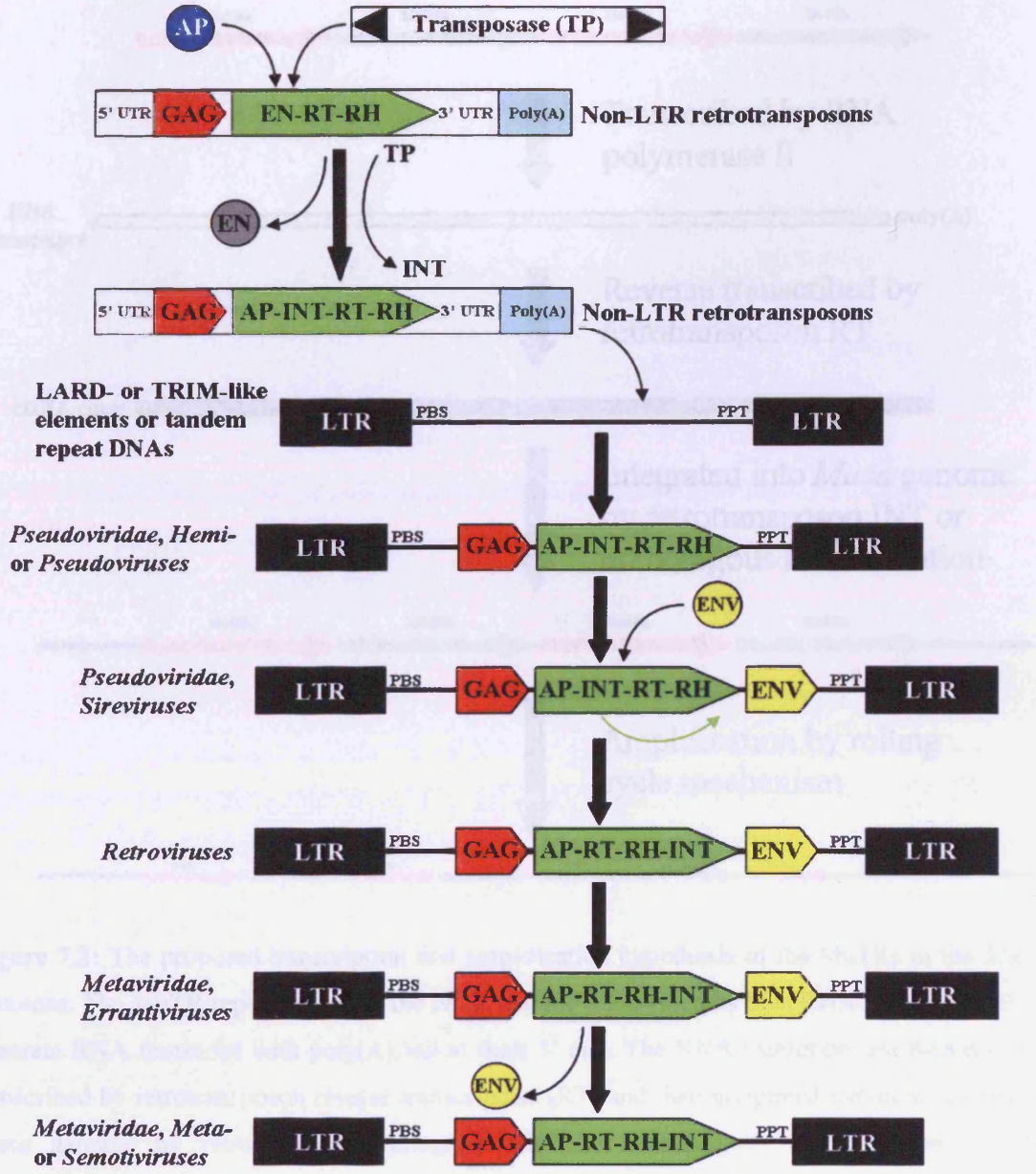


Figure 7.1: The LTR retrotransposons evolution hypothesis based on evidences from other studies (Doolittle *et al.*, 1989; Malik and Eickbush, 1999; 2001; Marín and Lloréns, 2000; Witte *et al.*, 2001; Peterson-Burch and Voytas, 2002; Gorinšek *et al.*, 2004; Hansen and Heslop-Harrison, 2004; Kalendar *et al.*, 2004) and in this study. During the evolution process of non-LTR retrotransposons, these elements acquired the aspartic protease (AP) from their host and also transposases (TP) that subsequently evolved to integrase (INT) from the DNA transposons, while the endonuclease (EN) are removed or replaced (Malik and Eickbush, 2001). New non-LTR retrotransposons transposed into LARD-like or TRIM-like elements or tandem repeat DNA (Witte *et al.*, 2001; Kapitonov and Jurka, 2003; Kalendar *et al.*, 2004) and gave rise to the *Pseudoviridae* elements that further evolved to the retroviruses and *Metaviridae* elements (Malik and Eickbush, 2004).

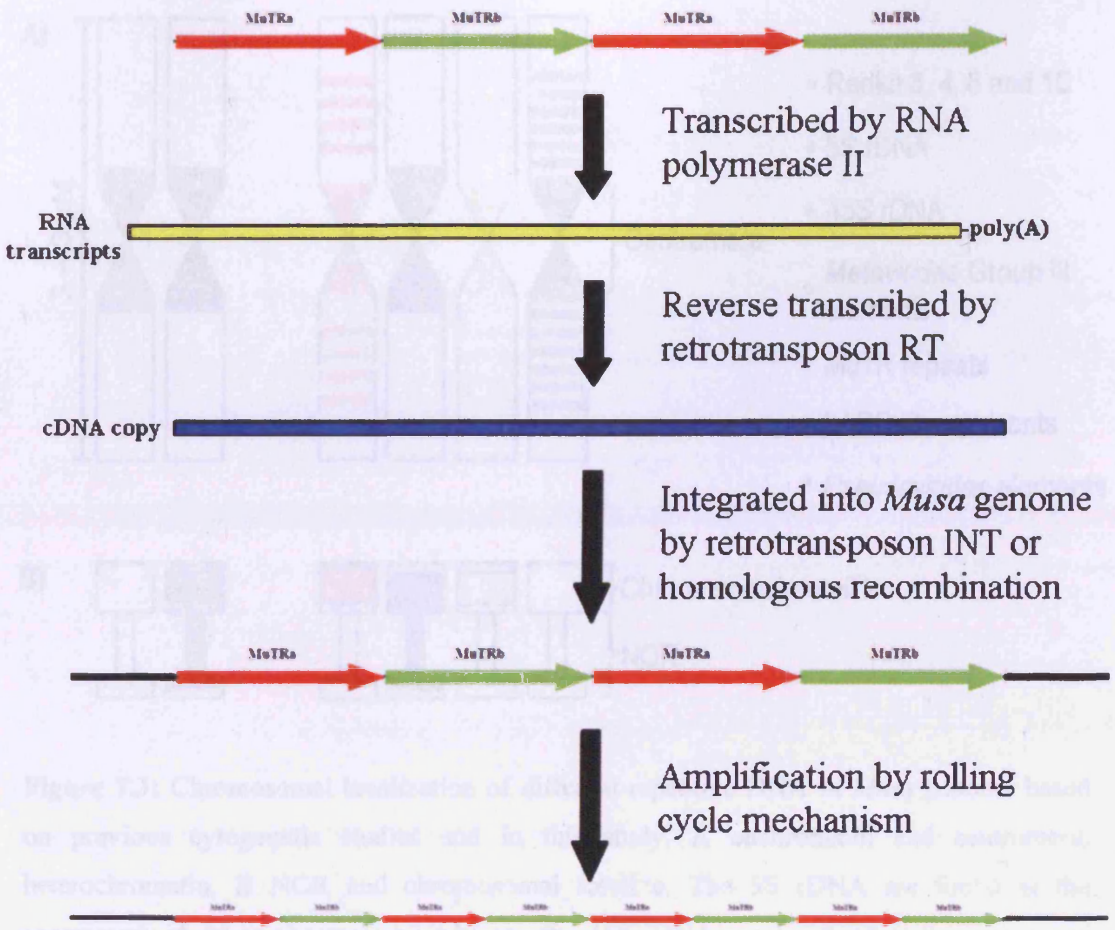


Figure 7.2: The proposed transcription and amplification hypothesis of the MuTRs in the *Musa* genomes. The MuTR repeats found at the NOR loci are transcribed by host RNA polymerase II to generate RNA transcript with poly(A) tail at their 3' end. The RNA transcripts are then reverse transcribed by retrotransposon reverse transcriptase (RT) and then integrated into centromeres of *Musa* genome by retrotransposon integrase (INT) or homologous recombination. Newly integrated MuTR repeats are subsequently amplified by the rolling cycle mechanism to generate large array of these repeats.

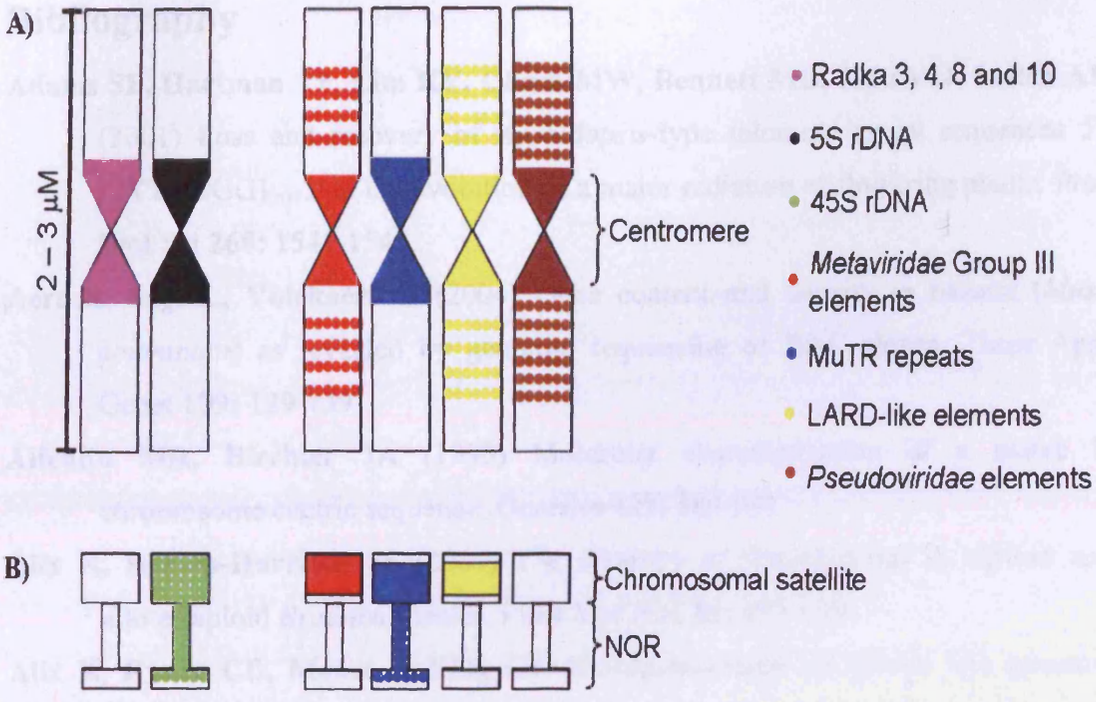


Figure 7.3: Chromosomal localization of different repetitive DNA in *Musa* genome based on previous cytogenetic studies and in this study. **A** euchromatin and centromeric heterochromatin, **B** NOR and chromosomal satellite. The 5S rDNA are found at the centromere of *Musa* chromosome whereas the 45S rDNA are localized at the active and distal part of NOR loci and also at the chromosomal satellite (Osuji *et al.*, 1997; 1998; Valárik *et al.*, 2002; Doležel *et al.*, 2004a; b; Bartoš *et al.*, 2005). Both *Metaviridae* Group III elements (Figure 6.2) and LARD-like elements (Figure 6.3) are localized at the chromosomal satellite, centromeric region and also dispersed along chromosome arm with latter elements are more abundant on the chromosome arm. *Pseudoviridae* elements (Figure 6.1) are not detected at the active region of NOR loci and chromosomal satellite but present in higher copy number at the centromere and chromosome arm compare to *Metaviridae* Group III elements and LARD-like elements. The MuTR repeats are found along the active and distal part of the NOR loci, the chromosomal satellite (Figure 6.4) and the centromeres of some *Musa* species (Figure 6.4B, C, E, G).

Bibliography

Adams SP, Hartman TP, Lim KY, Chase MW, Bennett MD, Leitch IJ, Leitch AR (2001) Loss and recovery of *Arabidopsis*-type telomere repeat sequences 5'-(TTTAGGG)$_{(n)}$-3' in the evolution of a major radiation of flowering plants. Proc Biol Sci **268:** 1541-1546

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

Alfenito MR, Birchler JA (1993) Molecular characterization of a maize B chromosome centric sequence. Genetics **135:** 589-597

Alix K, Heslop-Harrison JS (2004) The diversity of retroelements in diploid and allotetraploid *Brassica* species. Plant Mol Biol **54:** 895-909

Alix K, Ryder CD, Moore J, King GJ, Heslop-Harrison JS (2005) The genomic organization of retrotransposons in *Brassica oleracea*. Plant Mol Biol **59:** 839-851

Alkhimova OG, Mazurok NA, Potapova TA, Zakian SM, Heslop-Harrison JS, Vershinin AV (2004) Diverse patterns of the tandem repeats organization in rye chromosomes. Chromosoma **113:** 42-52

Allshire RC, Gosden JR, Cross SH, Cranston G, Rout D, Sugawara N, Szostak JW, Fantes PA, Hastie ND (1988) Telomeric repeat from *T. thermophila* cross hybridizes with human telomeres. Nature **332:** 656-659

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res **25:** 3389-3402

Amalraj VA (1992) Collecting banana germplasm in Southern India. Plant Genet Resour Newsl **88-89:** 64-66

Amon P, Ivanov I (2003) Genomic DNA labeling for hybridization with DNA arrays. Biotechniques **34:** 700-702, 704

Anamthawat-Jónsson K, Heslop-Harrison JS (1993) Isolation and characterization of genome-specific DNA sequences in *Triticeae* species. Mol Gen Genet **240:** 151-158

Ananiev EV, Phillips RL, Rines HW (1998a) A knob-associated tandem repeat in maize capable of forming fold-back DNA segments: are chromosome knobs megatransposons? Proc Natl Acad Sci USA **95:** 10785-10790

Ananiev EV, Phillips RL, Rines HW (1998b) Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc Natl Acad Sci USA **95:** 13073-13078

Antonius-Klemola K, Kalendar R, Schulman AH (2006) TRIM retrotransposons occur in apple and are polymorphic between varieties but not sports. Theor Appl Genet **112:** 999-1008

Aragon-Alcaide L, Miller T, Schwarzacher T, Reader S, Moore G (1996) A cereal centromeric sequence. Chromosoma **105:** 261-268

Arrighi FE, Manley Mandel M, Bergendahl J, Hsu TC (1970) Buoyant densities of DNA of mammals Biochem Gent **4:** 367-376

Ashikawa I, Kurata N, Nagamura Y, Minobe Y (1994) Cloning and mapping of telomere-associated sequences from rice. DNA Res **1:** 67-76

Asif MJ, Mak C, Othman RY (2001) Characterization of indigenous *Musa* species based on flow cytometric analysis of ploidy and nuclear DNA content. Caryologia **54:** 161-168

Asif MJ, Othman YR (2005) Characterization of Fusarium Wilt–Resistant and Fusarium Wilt–Susceptible Somaclones of Banana Cultivar Rastali (*Musa* AAB) by Random Amplified Polymorphic DNA and Retrotransposon Markers. Plant Mol Biol Rep **23:** 241-249

Aufsatz W, Mette MF, Matzke AJ, Matzke M (2004) The role of MET1 in RNA-directed de novo and maintenance methylation of CG dinucleotides. Plant Mol Biol **54:** 793-804

Aufsatz W, Mette MF, van der Winden J, Matzke M, Matzke AJ (2002) HDA6, a putative histone deacetylase needed to enhance DNA methylation induced by double-stranded RNA. EMBO J **21:** 6832-6841

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1995) Short Protocols in Molecular Biology, Ed 3rd. John Wiley & Sons, Inc., New York

Avramova Z, Tikhonov A, SanMiguel P, Jin YK, Liu C, Woo SS, Wing RA, Bennetzen JL (1996) Gene identification in a complex chromosomal continuum by local genomic cross-referencing. Plant J **10:** 1163-1168

Bae YA, Moon SY, Kong Y, Cho SY, Rhyu MG (2001) CsRn1, a novel active retrotransposon in a parasitic trematode, *Clonorchis sinensis*, discloses a new phylogenetic clade of Ty3/*gypsy*-like LTR retrotransposons. Mol Biol Evol **18:** 1474-1483

Balint-Kurti PJ, Clendennen SK, Dolezelova M, Valarik M, Dolezel J, Beetham PR, May GD (2000) Identification and chromosomal localization of the *monkey* retrotransposon in *Musa* sp. Mol Gen Genet **263:** 908-915

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

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

Beebee T, Rowe G (2004) Functional (coding) DNA. *In* T Beebee, G Rowe, eds, An introduction to molecular ecology. Oxford University press, Oxford

Beguiristain T, Grandbastien MA, Puigdomenech P, Casacuberta JM (2001) Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiol **127:** 212-221

Bender J (2004) Chromatin-based silencing mechanisms. Curr Opin Plant Biol **7:** 521-526

Bennetzen JL (1996) The contributions of retroelements to plant genome organization, function and evolution. Trends Microbiol **4:** 347-353

Bennetzen JL (1998) The structure and evolution of angiosperm nuclear genomes. Curr Opin Plant Biol **1:** 103-108

Bennetzen JL, Kellogg EA (1997) Do Plants Have a One-Way Ticket to Genomic Obesity? Plant Cell **9:** 1509-1514

Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P (1994) Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome **37:** 565-576

Bhat KV, Amaravathi Y, Gautam PL, Velayudhan KC (2004) AFLP characterization and genetic diversity analysis of Indian banana and Plantain cultivars (*Musa* spp.). Plant Genetic Resources: Characterization and Utilization **2:** 121-130

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

Bhat KV, Jarret RL, Liu Z-W (1994) RFLP characterization of Indian *Musa* germplasm for clonal identification and classification. Euphytica **80:** 95-103

Bhattacharyya MK, Gonzales RA, Kraft M, Buzzell RI (1997) A *copia*-like retrotransposon Tgmr closely linked to the *Rps1*-k allele that confers race-specific resistance of soybean to *Phytophthora sojae.* Plant Mol Biol **34:** 255-264

Bingham PM, Zachar Z (1989) Suppressible insertion-induced mutations in *Drosophila.* Prog Nucl Acid Res Mol Biol **36:** 87-98

Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev **16:** 6-21

Black JA, Gibson D (1974) Neutral evolution and immunoglobulin diversity. Nature **250:** 327-328

Blume B, Barry CS, Hamilton AJ, Bouzayen M, Grierson D (1997) Identification of transposons-like elements in non-coding regions of tomato ACC oxidase genes. Mol Gen Genet **254:** 297-303

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol **43:** 403-434

Boeke JD, Eickbush TH, Sandmeyer SB, Voytas DF (2004) *Pseudoviridae. In* CM Fauquet, ed, Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. Academic Press, New York, pp 663–672

Bond HE, Flamm WG, Burr HE, Bond SB (1967) Mouse satellite DNA. Further studies on its biological and physical characteristics and its intracellular localization. J Mol Biol **27:** 289

Bousios A, Saldana-Oyarzabal I, Valenzuela-Zapata AG, Wood C, Pearce SR (2007) Isolation and characterization of Ty1-*copia* retrotransposon sequences in the blue agave (*Agave tequilana* Weber var. azul) and their development as SSAP markers for phylogenetic analysis. Plant Sci **172:** 291-298

Boutabout M, Wilhelm M, Wilhelm FX (2001) DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1. Nucl Acids Res **29:** 2217-2222

Bowen NJ, McDonald JF (1999) Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements. Genome Res **9:** 924-935

Boyko E, Kalendar R, Korzun V, Fellers J, Korol A, Schulman AH, Gill BS (2002) A high-density cytogenetic map of the *Aegilops tauschii* genome incorporating

retrotransposons and defense-related genes: insights into cereal chromosome structure and function. Plant Mol Biol **48:** 767-790

Brandes A, Thompson H, Dean C, Heslop-Harrison JS (1997a) Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. Chromosome Res **5:** 238-246

Brandes A, HeslopHarrison JS, Kamm A, Kubis S, Doudrick RL, Schmidt T (1997b) Comparative analysis of the chromosomal and genomic organization of Ty1-*copia*-like retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant Mol Biol **33:** 11-21

Bureau TE, White SE, Wessler SR (1994) Transduction of a Cellular Gene by a Plant Retroelement. Cell **77:** 479-480

Burgers WA, Fuks F, Kouzarides T (2002) DNA methyltransferases get connected to chromatin. Trends Genet **18:** 275-277

Burke WD, Singh D, Eickbush TH (2003) R5 retrotransposons insert into a family of infrequently transcribed 28S rRNA genes of planaria. Mol Biol Evol **20:** 1260-1270

Camirand A, Brisson N (1990) The complete nucleotide sequence of the Tst1 retrotransposon of potato. Nucl Acids Res **18:** 4929

Carreel F, Fauré S, Gonzales de Leon D, Lagoda PJL, Perrier X, Bakry F, Tezenas du Montcel H, Lanaud C, Horry JP (1994) Evaluation de la diversité génétique chez les bananiers diploïdes (*Musa* sp.). Genet Sel Evol **26:** 125-130

Casacuberta JM, Vernhettes S, Audeon C, Grandbastien MA (1997) Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. Genetica **100:** 109-117

Casacuberta JM, Vernhettes S, Grandbastien MA (1995) Sequence variability within the tobacco retrotransposon Tnt1 population. EMBO J **14:** 2670-2678

Castiglione MR, Giraldi E, Frediani M (1995) The DNA methylation pattern of *Allium cepa* metaphase chromosomes. Biol Zent bl **114:** 57-66

Castilho A, Heslop-Harrison JS (1995) Physical mapping of 5S and 18S-25S rDNA and repetitive DNA sequences in *Aegilops umbellulata*. Genome **38:** 91-96

Castilho A, Vershinin A, Heslop-Harrison JS (2000) Repetitive DNA and the chromosomes in the genome of oil palm (*Elaeis guineensis*). Ann Bot **85:** 837-844

Chao Y, Su C, Chow T, Lee P, Chung C, Huang J, Liu S, Hsing YC (2003) TEOS1, a Novel Transposable Element Family from *Oryza Sativa*. Bot Bull Acad Sin **44:** 1-11

Chavanne F, Zhang DX, Liaud MF, Cerff R (1998) Structure and evolution of *Cyclops*: a novel giant retrotransposon of the Ty3/*gypsy* family highly amplified in pea and other legume species. Plant Mol Biol **37:** 363-375

Cheesman EE (1932) Genetical and cytological studies in *Musa*. I. Certain hybrids of the Gros Michel banana; II. Hybrids of the Mysore banana. J Genet **26:** 219-312

Chen C, Gentles AJ, Jurka J, Karlin S (2002) Genes, pseudogenes, and Alu sequence organization across human chromosomes 21 and 22. Proc Natl Acad Sci USA **99:** 2930-2935

Chen M, SanMiguel P, Bennetzen JL (1998) Sequence organization and conservation in sh2/a1-homologous regions of sorghum and rice. Genetics **148:** 435-443

Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell **14:** 1691-1704

Cheng Z, Buell CR, Wing RA, Gu M, Jiang J (2007) Toward a cytological characterization of the rice genome. Genome Res **11:** 2133-2141

Cheng Z, Stupar RM, Gu M, Jiang J (2001) A tandemly repeated DNA sequence is associated with both knob-like heterochromatin and a highly decondensed structure in the meiotic pachytene chromosomes of rice. Chromosoma **110:** 24-31

Cheng ZJ, Murata M (2003) A centromeric tandem repeat family originating from a part of Ty3/*gypsy*-retroelement in wheat and its relatives. Genetics **164:** 665-672

Clark LA, Wahl JM, Rees CA, Murphy KE (2006) Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. Proc Natl Acad Sci USA **103:** 1376-1381

Clendennen SK, May GD (1997) Differential gene expression in ripening banana fruit. Plant Physiol **115:** 463-469

Cohen S, Agmon N, Yacobi K, Mislovati M, Segal D (2005) Evidence for rolling circle replication of tandem genes in *Drosophila*. Nucl Acids Res **33:** 4519-4526

Cohen S, Menut S, Mechali M (1999) Regulated formation of extrachromosomal circular DNA molecules during development in *Xenopus laevis*. Mol Cell Biol **19:** 6682-6689

Cohen S, Yacobi K, Segal D (2003) Extrachromosomal circular DNA of tandemly repeated genomic sequences in *Drosophila*. Genome Res **13:** 1133-1145

Cohn M, McEachern MJ, Blackburn EH (1998) Telomeric sequence diversity within the genus *Saccharomyces*. Curr Genet **33:** 83-91

Contento A (2005) Genome interactions in polyploid plant species. University of Leicester, Leicester

Contento A, Heslop-Harrison JS, Schwarzacher T (2005) Diversity of a major repetitive DNA sequence in diploid and polyploid *Triticeae*. Cytogenet Genome Res **109:** 34-42

Copeland CS, Brindley PJ, Heyers O, Michael SF, Johnston DA, Williams DL, Ivens AC, Kalinna BH (2003) Boudicca, a retrovirus-like long terminal repeat retrotransposon from the genome of the human blood fluke *Schistosoma mansoni*. J Virol **77:** 6153-6166

Copenhaver GP, Nickel K, Kuromori T, Benito MI, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD, McCombie WR, Martienssen RA, Marra M, Preuss D (1999) Genetic definition and sequence analysis of *Arabidopsis* centromeres. Science **286:** 2468-2474

Corneo G, Ginelli E, Polli E (1967) A satellite DNA isolated from human tissues. J Mol Biol **23:** 619

Corneo G, Ginelli E, Soave C, Bernardi G (1968) Isolation and characterization of mouse and guinea pig satellite deoxyribonucleic acids. . Biochemistry **7:** 4373

Costa AP, Scortecci KC, Hashimoto RY, Araujo PG, Grandbastien MA, Van Sluys MA (1999) *Retrolycl-1*, a member of the *tnt1* retrotransposon super-family in the *Lycopersicon peruvianum* genome. Genetica **107:** 65-72

Cox AV, Bennett ST, Parokonny AS, Kenton A, Callimassia MA, Bennett MD (1993) Comparison of Plant Telomere Locations using a PCR-generated Synthetic Probe. Ann Bot **72:** 239-247

Craig NL (1997) Target site selection in transposition. Annu Rev Biochem **66:** 437-474

Creste S, Benatti TR, Orsi MR, Risterucci A-M, Figueira A (2006) Isolation and characterization of microsatellite loci from a commercial cultivar of *Musa acuminata*. Mol Ecol Notes **6:** 303-306

Creste S, Neto AT, Figueira A (2001) Detection of Single Sequence Repeat Polymorphisms in Denaturing Polyacrylamide Sequencing Gels by Silver Staining. Plant Mol. Biol. Report **19:** 299-306

Creste S, Neto AT, Silva SO, Figueira A (2003a) Genetic characterization of banana cultivars (*Musa* spp.) from Brazil using microsatellite markers. Euphytica **132:** 259-268

Creste S, Neto AT, Vencovsky R, Silva SO, Figueira A (2003b) Genetic diversity of *Musa* diploid and triploid accessions from the Brazilian banana breeding program estimated by microsatellite markers. Genet Resour Crop Evol **51:** 723-733

Crouch HK, Crouch JH, Jarret RL, Cregan PB, Ortiz R (1998) Segregation at microsatellite loci in haploid and diploid gametes of *Musa*. Crop Sci **38:** 211-217

Crouch JH, Vuylsteke D, Ortiz R (1998) Perspectives on the application of biotechnology to assist the genetic enhancement of plantain and banana (*Musa* spp.). Electron J Biotechnol **1:** 1-12

Csink AK, Henikoff S (1998) Something from nothing: the evolution and utility of satellite repeats. Trends Genet **14:** 200-204

Cuadrado A, Schwarzacher T (1998) The chromosomal organization of simple sequence repeats in wheat and rye genomes. Chromosoma **107:** 587-594

D'Hont A, Paget-Goy A, Escoute J, F. C (2000) The interspecific genome structure of cultivated banana, *Musa* spp. revealed by genomic DNA in situ hybridization. Theor Appl Genet **100:** 177-183

d'Adda di Fagagna F, Hande MP, Wei-Min Tong W-M, Lansdorp PM, Wang Z-Q, Jackson SP (1999) Functions of poly(ADP-ribose) polymerase in controlling telomere length and chromosomal stability. Nat Genet **23:** 76-80

Dahal G, d'A Hughes J, Thottappilly G, Lockhart BEL (1998) Effect of temperature on symptom expression and reliability of banana streak Badnavirus detection in naturally infected plantain and banana (*Musa* spp.). Plant Dis **82:** 16-21

Daniells JW, Jenny C, Karamura DA, Tomekpe K, Arnaud E, Sharrock S (2001) Musalogue: A catalogue of *Musa* germplasm. Diversity in the genus *Musa*. INIBAP, Montpellier, France

Demey JR, Gamez E, Molina S, Infante D (2004) Comparative study of the discriminating capacity of AFLP and ISTR markers for genetic analysis of *Agave fourcroydes*. Plant Mol Biol Report **22:** 29-35

Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res **12:** 1075-1079

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

Doležel J, Valárik M, Vrana J, Lysak MA, Hribova E, Bartos J, Gasmanova N, Doleželová M, Safar J, Simkova H (2004) Molecular cytogenetics and cytometry of banana (*Musa* spp.). *In* MS Jain, R Swennen, eds, Banana Improvement: Cellular, Molecular Biology, and Induced Mutations. Science Publishers, Inc., Enfield, pp 229 – 244

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 Plantarum **41:** 497 – 505

Dong F, Miller JT, Jackson SA, Wang GL, Ronald PC, Jiang J (1998) Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc Natl Acad Sci USA **95:** 8135-8140

Doolittle RF, Feng D-F, Johnson MS, McClure MA (1989) Origins and evolutionary relationships of retroviruses. Quart Rev Biol **64:** 1-30

Drouin G, Dover GA (1987) A plant processed pseudogene. Nature **328:** 557– 558

Eickbush TH (1992) Transposing without ends: the non-LTR retrotransposable elements. New Biol **4:** 430-440

Eickbush TH (2002) Repair by retrotransposition. Nat Genet **31:** 126-127

Eickbush TH, Eickbush DG (2007) Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics **175:** 477-485

Eickbush TH, Malik HS (2002) Origins and evolution of retrotransposons. *In* NL Craig, R Craigie, M Gellert, AL Lambowitz, eds, Mobile DNA II ASM Press Washington, DC, pp 1111-1144

Eissenberg JC (2001) Molecular biology of the chromo domain: an ancient chromatin module comes of age. Gene **275:** 19–29

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Polymorphism of insertion sites of Ty1-*copia* class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet **260:** 9-19

Epstein LM, Pabón-Peña LM (1991) Alternative modes of self-cleavage by newt satellite 2 transcripts. Nucl Acids Res **19:** 1699-1705

Espino RC, Pimentel RB (1990) Electrophoresic analysis of selected isozymes in BB-cultivars of Philippine bananas. *In* RL Jarret, ed, Indentification of Genetic Diversity in the Genus *Musa*. INIBAP, Montpellier, France, pp 36-40

Fajkus J, Kralovics R, Kovarik R, Fajkusova L (1995) The telomeric sequence is directly attached to the HRS60 subtelomeric tandem repeat in tobacco chromosomes. FEBS lett **364:** 33-35

FAOStat (2007) FAOSTAT http://faostat.fao.org/. Accessed September 2007

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

Fay MF, Chase MW (1996) Resurrection of *Themidaceae* for the *Brodiaea* alliance and recircumscription of *Alliaceae*, *Amaryllidaceae* and *Agapanthoideae*. Taxon **45:** 441-451

Ferbeyre G, Smith JM, Cedergren R (1998) *Schistosome* satellite DNA encodes active hammerhead ribozymes. Mol Cell Biol **18:** 3880-3888

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992) Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucl Acids Res **20:** 3639-3644

Flavell AJ, Knox MR, Pearce SR, Ellis TH (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J **16:** 643-650

Flavell AJ, Pearce SR, Heslop-Harrison P, Kumar A (1997) The evolution of Ty1-*copia* group retrotransposons in eukaryote genomes. Genetica **100:** 185-195

Flavell AJ, Pearce SR, Kumar A (1994) Plant transposable elements and the genome. Curr Opin Genet Dev **4:** 838-844

Flavell AJ, Smith DB, Kumar A (1992) Extreme heterogeneity of Ty1-*copia* group retrotransposons in plants. Mol Gen Genet **231:** 233-242

Flavell RB (1982) Sequence amplification, deletion and rearrangement: major sources of variation during species divergence. *In* GA Dover, RB Flavell, eds, Genome Evolution. Academic Press, London, UK, pp 301-323

Flavell RB (1994) Inactivation of gene expression in plants as a consequence of specific sequence duplication. Proc Natl Acad Sci USA **91:** 3490-3496

Flavell RB, Bedbrook J, Jones J, O'Dell M, Gerlach WL, Dyer TA, Thompson RD (1980) Molecular events in the evolution of cereal chromosomes. *In* DR Davies, DA Hopwood, eds, The 4th John Innes Symposium "The Plant Genome". The John Innes Charity, Norwich, UK, pp 15-30

Flavell RB, O'Dell M, Thompson WF (1988) Regulation of cytosine methylation in ribosomal DNA and nucleolus organizer expression in wheat. J Mol Biol **204:** 523-534

Frame IG, Cutfield JF, Poulter RT (2001) New BEL-like LTR-retrotransposons in *Fugu rubripes*, *Caenorhabditis elegans*, and *Drosophila melanogaster*. Gene **263:** 219-230

Francki MG (2001) Identification of Bilby, a diverged centromeric Ty1-*copia* retrotransposon family from cereal rye (*Secale cereale* L.). Genome **44:** 266-274

Fransz PF, Armstrong S, de Jong JH, Parnell LD, van Drunen C, Dean C, Zabel P, Bisseling T, Jones GH (2000) Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. Cell **100:** 367-376

Frediani M, Giraldi E, Castiglione MR (1996) Distribution of 5-methylcytosine-rich regions in the metaphase chromosomes of *Vicia faba*. Chromosome Res **4:** 141-146

Frello S, Heslop-Harrison JS (2000) Repetitive DNA sequences in *Crocus vernus* Hill (*Iridaceae*): the genomic organization and distribution of dispersed elements in the genus *Crocus* and its allies. Genome **43:** 902-909

Frello S, Orgaard M, Jacobsen N, Heslop-Harrison JS (2004) The genomic organization and evolutionary distribution of a tandemly repeated DNA sequence family in the genus *Crocus* (*Iridaceae*). Hereditas **141:** 81-88

Friesen N, Brandes A, Heslop-Harrison JS (2001) Diversity, origin, and distribution of retrotransposons (*gypsy* and *copia*) in conifers. Mol Biol Evol **18:** 1176-1188

Fry K, Salser W (1977) Nucleotide sequences of HS- satellite DNA from kangaroo rat *Dipodomys ordi* and characterization of similar sequences in other rodents. Cell **12:** 1069-1084

Fuchs J, Demidov D, Houben A, Schubert I (2006) Chromosomal histone modification patterns--from conservation to diversity. Trends Plant Sci **11:** 199-208

Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T (2000) DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet **24:** 88-91

Fuks F, Hurd PJ, Deplus R, Kouzarides T (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. Nucl Acids Res **31:** 2305-2312

Fuks F, Hurd PJ, Wolf D, Nan X, Bird AP, Kouzarides T (2003) The methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. J Biol Chem **278:** 4035-4040

Gai X, Voytas DF (1998) A single amino acid change in the yeast retrotransposon Ty5 abolishes targeting to silent chromatin. Mol Cell **1:** 1051-1055

Galasso I, Harrison GE, Pignone D, Brandes A, HeslopHarrison JS (1997) The distribution and organization of Ty1-*copia*-like retrotransposable elements in the genome of *Vigna unguiculata* (L) Walp (cowpea) and its relatives. Ann Bot **80:** 327-333

Ganal MW, Lapitan NL, Tanksley SD (1991) Macrostructure of the tomato telomeres. Plant Cell **3:** 87-94

Garber K, Bilic I, Pusch O, Tohme J, Bachmair A, Schweizer D, Jantsch V (1999) The Tpv2 family of retrotransposons of *Phaseolus vulgaris*: structure, integration characteristics, and use for genotype classification. Plant Mol Biol **39:** 797-807

Garrett JE, Knutzon DS, Carroll D (1989) Composite transposable elements in the *Xenopus laevis* genome. Mol Cell Biol **9:** 3018-3027

Gawel NJ, Jarret RL (1991) Cytoplasmic genetic diversity in bananas and plantains. Euphytica **52:** 19-23

Gawel NJ, Jarret RL, Whittemore AP (1992) Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa*. Theor Appl Genet **84:** 286-290

Ge XJ, Liu MH, Wang WK, Schaal BA, Chiang TY (2005) Population structure of wild bananas, *Musa balbisiana*, in China determined by SSR fingerprinting and cpDNA PCR-RFLP. Mol Ecol **14:** 933-944

Geering AD, Olszewski NE, Harper G, Lockhart BE, Hull R, Thomas JE (2005) Banana contains a diverse array of endogenous badnaviruses. J Gen Virol **86:** 511-520

Geering AD, Pooggin MM, Olszewski NE, Lockhart BE, Thomas JE (2005) Characterisation of Banana streak Mysore virus and evidence that its DNA is integrated in the B genome of cultivated *Musa*. Arch Virol **150:** 787-796

Geering ADW, McMichael LA, Dietzgen RG, Thomas JE (2000) Genetic diversity among Banana streak virus isolates from Australia. Phytopathology **90:** 921-927

Gendrel AV, Lippman Z, Yordan C, Colot V, Martienssen RA (2002) Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene DDM1. Science **297:** 1871-1873

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucl Acids Res **7:** 1869-1885

Gilley D, Blackburn EH (1996) Specific RNA residue interactions required for enzymatic functions of *Tetrahymena* telomerase. Mol Cell Biol **16:** 66-75

Gilley D, Lee MS, Blackburn EH (1995) Altering specific telomerase RNA template residues affects active site function. Genes Dev **9:** 2214-2226

Gindullis F, Desel C, Galasso I, Schmidt T (2001) The large-scale organization of the centromeric region in *Beta* species. Genome Res **11:** 253-265

Goodwin TJ, Poulter RT (2001) The DIRS1 group of retrotransposons. Mol Biol Evol **18:** 2067-2082

Goodwin TJ, Poulter RT (2004) A new group of tyrosine recombinase-encoding retrotransposons. Mol Biol Evol **21:** 746-759

Gorinšek B, Gorinšek F, Kordiš D (2004) Evolutionary genomics of *Chromoviruses* in eukaryotes. Mol. Biol. Evol. **21:** 781-798

Gowen S (1995) Bananas and Plantains. Chapman and Hall, London

Grafi G, Zemach A, Pitto L (2007) Methyl-CpG-binding domain (MBD) proteins in plants. Biochim Biophys Acta **1769:** 287-294

Grandbastien MA (1992) Retroelements in higher-plants. Trends Genet **8:** 103-108

Grandbastien MA, Lucas H, Morel JB, Mhiri C, Vernhettes S, Casacuberta JM (1997) The expression of the tobacco Tnt1 retrotransposon is linked to plant defense responses. Genetica **100:** 241-252

Grandbastien MA, Spielmann A, Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant-cell genetics. Nature **337:** 376-380

Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A, Flavell AJ (1999) Phylogeny and transpositional activity of Tyl-*copia* group retrotransposons in cereal genomes. Mol Gen Genet **261:** 883-891

Gruenbaum Y, Naveh-Many T, Cedar H, Razin A (1981) Sequence specificity of methylation in higher plant DNA. Nature **292:** 860-862

Guo X, Xu G, Zhang Y, Wen X, Hu W, Fan L (2006) Incongruent evolution of chromosomal size in rice. Genet Mol Res **5:** 373-389

Häkkinen M (2001) *Musa laterita*: An Ornamental Banana. Fruit Gardener **33:** 6-7

Häkkinen M (2003) *Musa campestris* Beccari varieties in Northern Borneo. Philippine Agricultural Scientist **86:** 424-435

Häkkinen M (2004a) *Musa campestris* Becc. (*Musaceae*) varieties in northern Borneo. Folia Malaysiana **5:** 81-100

Häkkinen M (2004b) *Musa voonii*, a new *Musa* species from Northern Borneo and discussion of the section *Callimusa* in Borneo. Acta Phytotax Geobot **54:** 79-88

Häkkinen M (2005) *Musa azizii*, a new *Musa* species (*Musaceae*) from Northern Borneo. Acta Phytotax Geobot **56:** 27-31

Häkkinen M (2006a) *Musa barioensis*, a new *Musa* species (*Musaceae*) from Northern Borneo. Acta Phytotax Geobot **57:** 55-60

Häkkinen M (2006b) *Musa lawitiensis* Nasution & Supard. (*Musaceae*) and its intraspecific taxa in Borneo. Adansonia **28:** 55-65

Häkkinen M, De Langhe E (2001) *Musa acuminata* in Northern Borneo. *In* INIBAP report. INIBAP, Montpellier, France, p 23

Häkkinen M, Meekiong K (2004) A new species of *Musa* from Borneo. Syst Biodiv **2:** 169-173

Häkkinen M, Meekiong K (2005) *Musa borneensis* Becc. (*Musaceae*) and its intraspecific taxa in Borneo. Acta Phytotax Geobot **56:** 213-230

Häkkinen M, Sharrock S (2002) Diversity in the genus *Musa* - Focus on *Rodochlamys*. *In* Annual Report 2001. INIBAP, Montpellier, France, pp pp 16-23

Häkkinen M, Suleiman M, Gisil J (2005a) *Musa beccarii* Simmonds (*Musaceae*) varieties in Sabah, Northern Borneo. Acta Phytotax Geobot **56:** 137-142

Häkkinen M, Teo CH, Othman YR (2007) Genome Constitution for *Musa beccarii* (*Musaceae*) Varieties. Acta Phytotax Sin **45:** 69-74

Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI (2002) Establishment and maintenance of a heterochromatin domain. Science **297:** 2232-2237

Hall TA (1999) BioEdit: a user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser **41:** 95-98

Han F, Fedak G, Guo W, Liu B (2005) Rapid and repeatable elimination of a parental genome-specific DNA repeat (pGc1R-1a) in newly synthesized wheat allopolyploids. Genetics **170:** 1239-1245

Hansen CN, Heslop-Harrison JS (2004) Sequences and phylogenies of plant pararetroviruses, viruses and transposable elements. Adv Bot Res **41:** 165-193

Harper G, Hart D, Moult S, Hull R, Geering A, Thomas J (2005) The diversity of Banana streak virus isolates in Uganda. Arch Virol **150:** 2407-2420

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

Harrington JJ, Van Bokkelen G, Mays RW, Gustashaw K, Willard HF (1997) Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. Nat Genet **15:** 345-355

Harrison CJ, Langdale JA (2006) A step by step guide to phylogeny reconstruction. Plant J **45:** 561-572

Harrison GE, Heslop-Harrison JS (1995) Centromeric repetitive DNA in the genus *Brassica*. Theor Appl Genet **90:** 157-165

Hashida SN, Kitamura K, Mikami T, Kishima Y (2003) Temperature shift coordinately changes the activity and the methylation state of transposon Tam3 in *Antirrhinum majus*. Plant Physiol **132:** 1207-1216

Havecker ER, Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. Genome Biol **5:** 225

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. Genome Res **16:** 1252-1261

Heitz E (1928) Das heterochromatin der Moose. Jahrb Wiss Bot **69:** 762-818

Henderson ER, Blackburn EH (1989) An overhanging 3' terminus is a conserved feature of telomeres. Mol Cell Biol **9:** 345-348

Henikoff S, Comai L (1998) A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. Genetics **149:** 307-318

Hennig W (1999) Heterochromatin. Chromosoma **108:** 1-9

Henning KA, Novotny EA, Compton ST, Guan XY, Liu PP, Ashlock MA (1999) Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. Proc Natl Acad Sci USA **96:** 592-597

Heslop-Harrison JS (2000a) RNA, genes, genomes and chromosomes: repetitive DNA sequences in plants. Chromosomes Today **13:** 45-56

Heslop-Harrison JS (2000b) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. Plant Cell **12:** 617-636

Heslop-Harrison JS (2003) Planning for remodelling: nuclear architecture, chromatin and chromosomes. Trends Plant Sci **8:** 195-197

Heslop-Harrison JS, Bennett MD (1990) Nuclear architecture in plants. Trend Genet **6:** 401-405

Heslop-Harrison JS, Brandes A, Schwarzacher T (2003) Tandemly repeated DNA sequences and centromeric chromosomal regions of *Arabidopsis* species. Chromosome Res **11:** 241-253

Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin AV, Alkhimova EG, Kamm A, Doudrick RL, Schwarzacher T, Katsiotis A, Kubis S, Kumar A, Pearce SR, Flavell AJ, Harrison GE (1997) The chromosomal distributions of Ty1-*copia* group retrotransposable elements in higher plants and their implications for genome evolution. Genetica **100:** 197-204

Heslop-Harrison JS, Murata M, Ogura Y, Schwarzacher T, Motoyoshi F (1999) Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. Plant Cell **11:** 31-42

Higashiyama T, Noutoshi Y, Akiba M, Yamada T (1995) Telomere and LINE-like elements at the termini of the *Chlorella* chromosome I. Nucl Acids Symp Ser **34:** 71-72

Hill P, Burford D, Martin DMA, Flavell AJ (2005) Retrotransposon populations of *Vicia* species with varying genome size. Mol Genet Genomics **273:** 371-381

Hirochika H (1993) Activation of tobacco retrotransposons during tissue-culture. EMBO J **12:** 2521-2528

Hirochika H, Fukuchi A, Kikuchi F (1992) Retrotransposon families in rice. Mol Gen Genet **233:** 209-216

Hirochika H, Hirochika R (1993) Ty1-*copia* group retrotransposons as ubiquitous components of plant genomes. Japanese J Genet **68:** 35-46

Hirochika H, Okamoto H, Kakutani T (2000) Silencing of retrotransposons in *Arabidopsis* and reactivation by the ddm1 mutation. Plant Cell **12:** 357-368

Hirochika H, Otsuki H, Yoshikawa M, Otsuki Y, Sugimoto K, Takeda S (1996) Autonomous transposition of the tobacco retrotransposon Tto1 in rice. Plant Cell **8:** 725-734

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

Hisatomi Y, Hanada K, Iida S (1997) The retrotransposon RTip1 is integrated into a novel type of minisatellite, MiniSip1, in the genome of the common morning glory and carries another new type of minisatellite, MiniSip2. Theor Appl Genet **95:** 1049-1056

Houben A, Schroeder-Reiter E, Nagaki K, Nasuda S, Wanner G, Murata M, Endo TR (2007) CENH3 interacts with the centromeric retrotransposon cereba and GC-rich satellites and locates to centromeric substructures in barley. Chromosoma **116:** 275-283

Hohn T, Richert-Pöggeler KR, Staginnus C, Harper G, Schwarzacher T, Teo CH, Teycheney P-Y, Iskra-Caruana ML, Hull R (2007) Evolution of integrated plant viruses. . *In* M Rossinck, ed, Plant virus evolution. Springer-Verlag, Berlin, in press

Howell EC, Newbury J, Swennen RL, Withers LA, Ford-Lloyd BV (1994) The use of RAPD for identifying and classifying *Musa* germ plasm. Genome Biol **37:** 328-332

Hsieh CL (1999) In vivo activity of murine *de novo* methyltransferases, Dnmt3a and Dnmt3b. Mol Cell Biol **19:** 8211-8218

Hu W, Das OP, Messing J (1995) Zeon-1, a member of a new maize retrotransposon family. Mol Gen Genet **248:** 471-480

Hudakova S, Michalek W, Presting GG, ten Hoopen R, dos Santos K, Jasencakova Z, Schubert I (2001) Sequence organization of barley centromeres. Nucl Acids Res **29:** 5029-5035

Huettel B, Kanno T, Daxinger L, Aufsatz W, Matzke AJ, Matzke M (2006) Endogenous targets of RNA-directed DNA methylation and Pol IV in *Arabidopsis*. EMBO J **25:** 2828-2836

Huettel B, Kanno T, Daxinger L, Bucher E, van der Winden J, Matzke AJ, Matzke M (2007) RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. Biochim Biophys Acta **1769:** 358-374

Hulbert SH, Richter TE, Axtell JD, Bennetzen JL (1990) Genetic mapping and characterization of sorghum and related crops by means of maize DNA probes. Proc Natl Acad Sci USA **87:** 4251-4255

Hull R (1999) Classification of reverse transcribing elements: a discussion document. Arch Virol **144:** 209-214

Hull R (2001) Classifying reverse transcribing elements: a proposal and a challenge to the ICTV. Arch Virol **146:** 2255-2261

ICTVdB (2006) Index to ICTVdB virus descriptions. *In* C Büchen-Osmond, ed, *ICTVdB - The Universal Virus Database*, version 4. ICTVdB Management, Mailman School of Public Health, Columbia University, New York

Ikeno M, Grimes B, Okazaki T, Nakano M, Saitoh K, Hoshino H, McGill NI, Cooke H, Masumoto H (1998) Construction of YAC-based mammalian artificial chromosomes. Nat Biotechnol **16:** 431-439

Inukai T (2004) Role of transposable elements in the propagation of minisatellites in the rice genome. Mol Genet Genomics **271:** 220-227

Isobe M, Hashimoto K (1994) The chromosome count of nine taxa in *Musa* and its allied genus *Musella*. Bull Hiroshima Bot Gard **15:** 7-11

Jackson SA, Wang ML, Goodman HM, Jiang J (1998) Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. Genome **41:** 566-572

Jaligot E, Rival A, Beulé T, Dussert S, Verdeil J-L (2000) Somaclonal variation in oil palm (*Elaeis guineensis* Jacq.): the DNA methylation hypothesis. Plant Cell Reports **19:** 684-690

Jarret RL, Bhat KV, Cregan P, Ortiz R, Vuylsteke D (1994) Isolation of microsatellite DNA markers in *Musa*. InfoMusa **3:** 3-4

Jarret RL, Gawel NJ, Whittemore A (1992) RFLP-based phylogeny of *Musa* species in Papua New Guinea. Theor Appl Genet **84**

Jenny C, Carreel F, Bakry F (1997) Revision on banana taxonomy: 'Klue Tiparot' (*Musa* sp) reclassified as a triploid. Fruits **52:** 83-91

Jiang J, Nasuda S, Dong F, Scherrer CW, Woo SS, Wing RA, Gill BS, Ward DC (1996) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. Proc Natl Acad Sci USA **93:** 14210-14213

Jiang JM, Birchler JA, Parrott WA, Dawe RK (2003) A molecular view of plant centromeres. Trends Plant Sci **8:** 570-575

Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler SR (2002a) Dasheng: A recently amplified nonautonomous long terminal repeat element that is a major component of pericentromeric regions in rice. Genetics **161:** 1293-1305

Jiang N, Jordan IK, Wessler SR (2002b) Dasheng and RIRE2. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. Plant Physiol **130:** 1697-1705

Jin W, Lamb JC, Vega JM, Dawe RK, Birchler JA, Jiang J (2005) Molecular and functional dissection of the maize B chromosome centromere. Plant Cell **17:** 1412-1423

Jin W, Melo JR, Nagaki K, Talbert PB, Henikoff S, Dawe RK, Jiang J (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. Plant Cell **16:** 571-581

Jin YK, Bennetzen JL (1989) Structure and coding properties of Bs1, a maize retrovirus-like transposon. Proc Natl Acad Sci USA **86:** 6235-6239

Jin YK, Bennetzen JL (1994) Integration and nonrandom mutation of a plasma membrane proton ATPase gene fragment within the Bs1 retroelement of maize. Plant Cell **6:** 1177-1186

Jing R, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis TH, Flavell AJ (2005) Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in *pisum* species. Genetics **171:** 741-752

Johns MA, Mottinger J, Freeling M (1985) A low copy number, *copia*-like transposon in maize. EMBO J **4:** 1093-1101

Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. Curr Biol **12:** 1360-1367

Johnson LM, Bostick M, Zhang X, Kraft E, Henderson I, Callis J, Jacobsen SE (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. Curr Biol **17:** 379-384

Jones CW, Kafatos FC (1982) Accepted mutations in a gene family: evolutionary diversification of duplicated DNA. J Mol Evol **19:** 87-103

Jordan EG, Martini G, Bennett MD, Flavell RB (1982) Nucleolar fusion in wheat. J Cell Sci. **56:** 485-495

Kaemmer D, Fischer D, Jarret RL, Baurens F-C, Grapin A, Dambier D, Noyer JL, Lanaud C, Kahl G, Lagoda PJL (1997) Molecular breeding in the genus *Musa*: a strong case for STMS marker technology. Euphytica **96:** 49-63

Kahangi EM (2002) RAPD profiling of some banana (*Musa* spp) varieties selected by small scale farmers in Kennya. *In* JM Kenya Wesonga, T Losenge, CK Ndung'u, K Ngamau, JBM Njoroge, FK Ombwara, SG Agong, eds, Proceedings of the Second Horticultural seminar on Sustainable Horticultural Production in the Tropics. Jomo Kenyatta University of Agriculture and Technology (JKUAT), Juja, Kenya.

Kakutani T (1997) Genetic characterization of late-flowering traits induced by DNA hypomethylation mutation in *Arabidopsis thaliana*. Plant J **12:** 1447-1451

Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, Richards EJ (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. Proc Natl Acad Sci USA **93:** 12406-12411

Kalendar R (2006) A new Frodo class of plant TRIM LTR-retroelements has origin from 5S rRNA gene, mobile RNA and retrotransposons. *In* A Ciudad Autónoma de Buenos Aires, ed, Segundas Jornadas Iberoamericanas de Bioinformática, Regente Palace Hotel, Buenos Aires

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman A (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet **98:** 704-711

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA **97:** 6603-6607

Kalendar R, Vicient CM, Peleg O, Anamthawat-Jonsson K, Bolshoy A, Schulman AH (2004) Large retrotransposon derivatives: abundant, conserved but

nonautonomous retroelements of barley and related genomes. Genetics **166:** 1437-1450

Kamm A, Doudrick RL, Heslop-Harrison JS, Schmidt T (1996) The genomic and physical organization of Ty1-*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

Kamm A, Schmidt T, Heslop-Harrison JS (1994) Molecular and physical organization of highly repetitive, undermethylated DNA from *Pennisetum glaucum*. Mol Gen Genet **244:** 420-425

Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddeloh JA, Riddle NC, Verbsky ML, Richards EJ (2003) *Arabidopsis* MET1 cytosine methyltransferase mutants. Genetics **163:** 1109-1122

Kapitonov VV, Jurka J (1999) Molecular paleontology of transposable elements from *Arabidopsis thaliana*. Genetica **107:** 27-37

Kapitonov VV, Jurka J (2003) A novel class of SINE elements derived from 5S rRNA. Mol Biol Evol **20:** 694-702

Kaszas E, Birchler JA (1996) Misdivision analysis of centromere structure in maize. EMBO J **15:** 5246-5255

Kaszas E, Birchler JA (1998) Meiotic transmission rates correlate with physical features of rearranged centromeres in maize. Genetics **150:** 1683-1692

Kato M, Takashima K, Kakutani T (2004) Epigenetic control of CACTA transposon mobility in *Arabidopsis thaliana*. Genetics **168:** 961-969

Katsiotis A, Hagidimitriou M, Heslop-Harrison JS (1997) The close relationship between the A and B genomes in *Avena* L. determined by molecular cytogenetic analysis of total genomic, tandemly and dispersed repetitive DNA sequences. Ann Bot **79:** 103-109

Katsiotis A, Loukas M, Heslop-Harrison JS (2000) Repetitive DNA, genome and species relationships in *Avena* and *Arrhenatherum* (*Poaceae*). Ann Bot **86:** 1135-1142

Katsiotis A, Schmidt T, HeslopHarrison JS (1996) Chromosomal and genomic organization of Ty1-*copia*-like retrotransposon sequences in the genus *Avena*. Genome **39:** 410-417

Kejnovsky E, Kubat Z, Macas J, Hobza R, Mracek J, Vyskot B (2006) Retand: a novel family of *gypsy*-like retrotransposons harboring an amplified tandem repeat. Mol Genet Genomics **276:** 254-263

Kentner EK, Arnold ML, Wessler SR (2003) Characterization of high-copy-number retrotransposons from the large genomes of the Louisiana iris species and their use as molecular markers. Genetics **164:** 685-697

Kenward KD, Bai D, Ban MR, Brandle JE (1999) Isolation and characterization of Tnd-1, a retrotransposon marker linked to black root rot resistance in tobacco. Theor Appl Genet **98:** 387-395

Kim AI, Lyubomirskaya NV, Belyaeva ES, Shostack NG, Ilyin YV (1994) The introduction of a transpositionally active copy of retrotransposon *GYPSY* into the stable strain of *Drosophila melanogaster* causes genetic instability. Mol Gen Genet **242:** 472-477

Kimura Y, Tosa Y, Shimada S, Sogo R, Kusaba M, Sunaga T, Betsuyaku S, Eto Y, Nakayashiki H, Mayama S (2001) OARE-1, a Ty1-*copia* retrotransposon in oat activated by abiotic and biotic stresses. Plant Cell Physiol **42:** 1345-1354

Kirchner J, Connolly CM, Sandmeyer SB (1995) Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retrovirus-like element. Science **267:** 1488-1491

Kirik A, Salomon S, Puchta H (2000) Species-specific double-strand break repair and genome evolution in plants. EMBO J **19:** 5562-5566

Kishii M, Nagaki K, Tsujimoto H (2001) A tandem repetitive sequence located in the centromeric region of common wheat (*Triticum aestivum*) chromosomes. Chromosome Res **9:** 417-428

Kishii M, Nagaki K, Tsujimoto H, Sasakuma T (1999) Exclusive localization of tandem repetitive sequences in subtelomeric heterochromatin regions of *Leymus racemosus* (*Poaceae*, *Triticeae*). Chromosome Res **7:** 519-529

Kishii M, Tsujimoto H (2002) Genus-specific localization of the TaiI family of tandem-repetitive sequences in either the centromeric or subtelomeric regions in *Triticeae* species (*Poaceae*) and its evolution in wheat. Genome **45:** 946-955

Kit S (1961) Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. . J Mol Biol **3:** 711

Klobutcher LA, Swanton MT, Donini P, Prescott DM (1981) All gene-sized DNA molecules in four species of Hypotrichs have the same terminal sequence and an unusual 3' terminus. Proc Natl Acad Sci USA **78:** 3015-3019

Kojima KK, Fujiwara H (2003) Evolution of target specificity in R1 clade non-LTR retrotransposons. Mol Biol Evol **20:** 351-361

Kojima KK, Fujiwara H (2004) Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. Mol Biol Evol **21:** 207-217

Konieczny A, Voytas DF, Cummings MP, Ausubel FM (1991) A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics **127:** 801-809

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

Krzywinski J, Sangare D, Besansky NJ (2005) Satellite DNA from the Y chromosome of the malaria vector *Anopheles gambiae*. Genetics **169:** 185-196

Kubis S, Heslop-Harrison JS, Schmidt T (1997) A family of differentially amplified repetitive DNA sequences in the genus *Beta* reveals genetic variation in *Beta vulgaris* subspecies and cultivars. J Mol Evol **44:** 310-320

Kubis SE, Castilho AM, Vershinin AV, Heslop-Harrison JS (2003) Retroelements, transposons and methylation status in the genome of oil palm (*Elaeis guineensis*) and the relationship to somaclonal variation. Plant Mol Biol **52:** 69-79

Kubis SE, Heslop-Harrison JS, Desel C, Schmidt T (1998) The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. Plant Mol Biol **36:** 821-831

Kuhlmann M, Borisova BE, Kaller M, Larsson P, Stach D, Na J, Eichinger L, Lyko F, Ambros V, Soderbom F, Hammann C, Nellen W (2005) Silencing of retrotransposons in *Dictyostelium* by DNA methylation and RNAi. Nucl Acids Res **33:** 6405-6417

Kuhn GC, Sene FM (2005) Evolutionary turnover of two pBuM satellite DNA subfamilies in the *Drosophila buzzatii* species cluster (*repleta* group): from alpha to alpha/beta arrays. Gene **349:** 77-85

Kuipers AG, Heslop-Harrison JS, Jacobsen E (1998) Characterisation and physical localisation of Ty1-*copia*-like retrotransposons in four *Alstroemeria* species. Genome **41:** 357-367

Kumar A (1996) The adventures of the Ty1-*copia* group of retrotransposons in plants. Trends Genet **12:** 41-43

Kumar A (1998) The evolution of plant retroviruses: moving to green pasture. Trends Plant Sci. **3:** 371– 374

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Annu Rev Genet **33:** 479-532

Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ (1997) The Ty1-*copia* group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. Genetica **100:** 205-217

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform **5:** 150-163

Kumekawa N, Hosouchi T, Tsuruoka H, Kotani H (2000) The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 5. DNA Res **7:** 315-321

Kumekawa N, Hosouchi T, Tsuruoka H, Kotani H (2001) The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 4. DNA Res **8:** 285-290

Kumekawa N, Ohtsubo H, Horiuchi T, Ohtsubo E (1999) Identification and characterization of novel retrotransposons of the *gypsy* type in rice. Mol Gen Genet **260:** 593-602

Labrador M, Corces VG (1997) Transposable element-host interactions: Regulation of insertion and excision. Annu Rev Genet **31:** 381-404

Lall IP, Maneesha, Upadhyaya KC (2002) *Panzee*, a *copia*-like retrotransposon from grain legume pigeonpea (*Cajanus cajan* L.). Mol Genet Genomics **267:** 271-280

Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. . Annu Rev Biochem **62:** 587-622

Lanaud C, Tezenas du Montcel H, Jolivot MP, Glaszmann JC, Gonzalez de Leon D (1992) Variation of ribosomal gene spacer length among wild and cultivated bananas. Heredity **68:** 147-156

Langdon T, Seago C, Mende M, Leggett M, Thomas H, Forster JW, Thomas H, Jones RN, Jenkins G (2000) Retrotransposon evolution in diverse plant genomes. Genetics **156:** 313-325

Laten HM, Majumdar A, Gaucher EA (1998) SIRE-1, a *copia*/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc Natl Acad Sci USA **95:** 6897-6902

Leblanc P, Desset S, Giorgi F, Taddei AR, Fausto AM, Mazzini M, Dastugue B, Vaury C (2000) Life cycle of an endogenous retrovirus, ZAM, in *Drosophila melanogaster*. J Virol **74:** 10658-10669

Lee D, Ellis THN, Turner L, Hellens RP, Cleary WG (1990) A *copia*-like element in *Pisum* demonstrates the uses of dispersed repeated sequences in genetic-analysis. Plant Mol Biol **15:** 707-722

Lee HR, Neumann P, Macas J, Jiang J (2006) Transcription and evolutionary dynamics of the centromeric satellite repeat CentO in rice. Mol Biol Evol **23:** 2505-2520

Lee HR, Zhang W, Langdon T, Jin W, Yan H, Cheng Z, Jiang J (2005) Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. Proc Natl Acad Sci USA **102:** 11793-11798

Li E (2002) Chromatin modification and epigenetic reprogramming in mammalian development. Nat Rev Genet **3:** 662-673

Li J, Leung FC (2006) A CR1 element is embedded in a novel tandem repeat (HinfI repeat) within the chicken genome. Genome **49:** 97-103

Li YC, Lee C, Sanoudou D, Hseu TH, Li SY, Lin CC (2000) Interstitial colocalization of two cervid satellite DNAs involved in the genesis of the Indian muntjac karyotype. Chromosome Res **8:** 363-373

Lim JK, Simmons JM (1994) Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. BioEssays **16:** 269-273

Lin JY, Jacobus BH, SanMiguel P, Walling JG, Yuan Y, Shoemaker RC, Young ND, Jackson SA (2005) Pericentromeric regions of soybean (*Glycine max* L. Merr.) chromosomes consist of retroelements and tandemly repeated DNA and are structurally and evolutionarily labile. Genetics **170:** 1221-1230

Lippman Z, Martienssen R (2004) The role of RNA interference in heterochromatic silencing. Nature **431:** 364-370

Lippman Z, May B, Yordan C, Singer T, Martienssen R (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. PLoS Biol **1:** E67

Liu ZL, Han FP, Tan M, Shan XH, Dong YZ, Wang XZ, Fedak G, Hao S, Liu B (2004) Activation of a rice endogenous retrotransposon Tos17 in tissue culture is accompanied by cytosine demethylation and causes heritable alteration in methylation pattern of flanking genomic regions. Theor Appl Genet **109:** 200-209

Lo AW, Liao GC, Rocchi M, Choo KH (1999) Extreme reduction of chromosome-specific alpha-satellite array is unusually common in human chromosome 21. Genome Res **9:** 895-908

Lockhart BEL (1994) Banana streak virus. *In* RC Ploetz, GA Zentmyer, WT Nishijima, KG Rohrbach, HD Ohr, eds, Compendium of Tropical Fruit Diseases. American Phytopathology Society Press, St Paul, Minnesota, pp 19-20

Lockhart BEL, Jones DR (1999) Banana streak. *In* DR Jones, ed, Diseases of banana, abacá, and enset. CABI Publishing, Wallingford, pp 263-274

Lockhart BEL, Olszewski NE (1993) Serological and genomic heterogeneity of banana streak badnavirus: implications for virus detection in *Musa* germplasm. *In* J Ganry, ed, Breeding Banana and Plantain for Resistance to Diseases and Pests. CIRAD, Montpellier, pp 105-113

Loguercio LL, Wilkins TA (1998) Structural analysis of a hmg-coA-reductase pseudogene: insights into evolutionary processes affecting the hmgr gene family in allotetraploid cotton (*Gossypium hirsutum* L.). Curr Genet **34:** 241-249

Loh JP, Kiew R, Set O, Gan LH, Gan YY (2000) Amplified fragment length polymorphism fingerprinting of 16 banana cultivars (*Musa* cvs.). Mol Phylogenet Evol **17:** 360-366

Long EO, Dawid IB (1980) Alternative pathways in the processing of ribosomal RNA precursor in *Drosophila melanogaster*. J Mol Biol **138:** 873-878

Lopez-Leon MD, Cabrero J, Camacho JP (1999) Unusually high amount of inactive ribosomal DNA in the grasshopper *Stauroderus scalaris*. Chromosome Res **7:** 83-88

Lu S, Van Eck J, Zhou X, Lopez AB, O'Halloran DM, Cosman KM, Conlin BJ, Paolillo DJ, Garvin DF, Vrebalov J, Kochian LV, Kupper H, Earle ED, Cao J, Li L (2006) The cauliflower Or gene encodes a DnaJ cysteine-rich domain-containing protein that mediates high levels of beta-carotene accumulation. Plant Cell **18:** 3594-3605

Lyko F, Ramsahoye BH, Kashevsky H, Tudor M, Mastrangelo MA, Orr-Weaver TL, Jaenisch R (1999) Mammalian (cytosine-5) methyltransferases cause genomic DNA methylation and lethality in *Drosophila*. Nat Genet **23:** 363-366

Lyko F, Whittaker AJ, Orr-Weaver TL, Jaenisch R (2000) The putative *Drosophila* methyltransferase gene dDnmt2 is contained in a transposon-like element and is expressed specifically in ovaries. Mech Dev **95:** 215-217

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

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci USA **101:** 12404-12410

Ma JX, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res **14:** 860-869

Macas J, Meszaros T, Nouzova M (2002) PlantSat: a specialized database for plant satellite repeats. Bioinformatics **18:** 28-35

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

Makunin IV, Pokholkova GV, Kholodilov NG, Zakharkin SO, Bonaccorsi S, Dimitri P, Zhimulev IF (1999) A novel simple satellite DNA is colocalized with the Stalker retrotransposon in *Drosophila melanogaster* heterochromatin. Mol Gen Genet **261:** 381-387

Malik HS, Eickbush TH (1999) Modular evolution of the integrase domain in the Ty3/*Gypsy* class of LTR retrotransposons. J Virol **73:** 5186-5190

Malik HS, Eickbush TH (2001) Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res **11:** 1187-1197

Malik HS, Henikoff S, Eickbush TH (2000) Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res **10:** 1307-1318

Maluszynska J, Heslop-Harrison JS (1991) Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. Plant J. **1:** 159-166

Manninen I, Schulman AH (1993) Bare-1, a *copia*-like retroelement in barley (*Hordeum vulgare* L). Plant Mol Biol **22:** 829-846

Manninen O, Kalendar R, Robinson J, Schulman AH (2000) Application of BARE-1 retrotransposon markers to the mapping of a major resistance gene for net blotch in barley. Mol Gen Genet **264:** 325-334

Mao L, Begum D, Goff SA, Wing RA (2001) Sequence and analysis of the tomato *JOINTLESS* locus. Plant Physiol **126:** 1331-1340

Marillonnet S, Wessler SR (1997) Retrotransposon insertion into the maize waxy gene results in tissue-specific RNA processing. Plant Cell **9:** 967-978

Marillonnet S, Wessler SR (1998) Extreme structural heterogeneity among the members of a maize retrotransposon family. Genetics **150:** 1245-1256

Marín I, Lloréns C (2000) Ty3/*Gypsy* retrotransposons: Description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. Mol Biol Evol **17:** 1040-1049

Marmorstein R (2001) Protein modules that manipulate histone tails for chromatin regulation. Nat Rev Mol Cell Biol **2:** 422–432

Marriott J (1980) Bananas--physiology and biochemistry of storage and ripening for optimum quality. Crit Rev Food Sci Nutr **13:** 41-88

Martienssen R (1998) Transposons, DNA methylation and gene control. Trends Genet **14:** 263-264

Martienssen RA, Colot V (2001) DNA methylation and epigenetic inheritance in plants and filamentous fungi. Science **293:** 1070-1074

Martinez-Izquierdo JA, Garcia-Martinez J, Vicient CM (1997) What makes Grande1 retrotransposon different? Genetica **100:** 15-28

Martinez-Zapater JM, Estelle MA, Somerville CR (1986) A highly repeated DNA sequence in *Arabidopsis thaliana*. Mol Gen Genet **204:** 417-423

Martini G, Flavell RB (1985) The control of nucleolus volume in wheat: A genetic study at three developmental stages. Heredity **54:** 111-120

Mathieu O, Bender J (2004) RNA-directed DNA methylation. J Cell Sci **117:** 4881-4888

Matthes M, Singh R, Cheah S-C, Karp A (2001) Variation in oil palm (*Elaeis guineensis* Jacq.) tissue culture-derived regenerants revealed by AFLPs with methylation-sensitive enzymes. Theor Appl Genet **102:** 971-979

Matyunina LV, Jordan IK, McDonald JF (1996) Naturally occurring variation in *copia* expression is due to both element (cis) and host (trans) regulatory variation. Proc Natl Acad Sci USA **93:** 7097-7102

Matzke MA, Birchler JA (2005) RNAi-mediated pathways in the nucleus. Nat Rev Genet **6:** 24-35

May BP, Lippman ZB, Fang Y, Spector DL, Martienssen RA (2005) Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. PLoS Genet **1:** e79

McCarthy EM, Liu J, Lizhi G, McDonald JF (2002) Long terminal repeat retrotransposons of *Oryza sativa*. Genome Biol **3:** 0053.1-0053.11

McCarthy EM, McDonald JF (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics **19:** 362-367

McClure MA (1991) Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol Biol Evol **8:** 835-856

McCollum AM, Ganko EW, Barrass PA, Rodriguez JM, McDonald JF (2002) Evidence for the adaptive significance of an LTR retrotransposon sequence in a *Drosophila* heterochromatic gene. BMC Evol Biol **2:** 5

McEachern MJ, Krauskopf A, Blackburn EH (2000) Telomeres and their control. Annu Rev Genet **34:** 331-358

Medina-Suarez R, Manning K, Fletcher J, Aked J, Bird CR, Seymour GB (1997) Gene expression in the pulp of ripening bananas. Two-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis of in vitro translation products and cDNA cloning of 25 different ripening-related mRNAs. Plant Physiol **115:** 453-461

Meštrović N, Plohl M, Mravinac B, Ugarkovic D (1998) Evolution of satellite DNAs from the genus *Palorus* - experimental evidence for the "library" hypothesis. Mol Biol Evol **15:** 1062-1068

Meunier B, Tian G-L, Macadre C, Slonimski PP, Lazowska J (1990) Group II introns transpose in yeast mitochondria. *In* E Quagliariello, S Papa, F Palmieri, C Saccone, eds, Structure Function and Biogenesis of Energy Transfer Systems. Elsevier Scientific Publishing Company, Amsterdam, pp 169–174

Meyers BC, Tingey SV, Morgante M (2001) Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res **11:** 1660-1676

Meyer P, Niedenhof I, ten Lohuis M (1994) Evidence for cytosine methylation of non-symmetrical sequences in transgenic *Petunia hybrida*. EMBO J **13:** 2084-2088

Meyne J, Baker RJ, Hobart HH, Hsu TC, Ryder OA, Ward OG, Wiley JE, Wurster-Hill DH, Yates TL, Moyzis RK (1990) Distribution of non-telomeric sites of the (TTAGGG)n telomeric sequence in vertebrate chromosomes. Chromosoma **99:** 3-10

Meyne J, Ratliff RL, Moyzis RK (1989) Conservation of the human telomere sequence (TTAGGG)n among vertebrates. Proc Natl Acad Sci USA **86:** 7049-7053

Mhiri C, Morel JB, Vernhettes S, Casacuberta JM, Lucas H, Grandbastien MA (1997) The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol Biol **33:** 257-266

Michel F, Lang BF (1985) Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nat Genet **316:** 641-643

Miller JT, Dong F, Jackson SA, Song J, Jiang J (1998) Retrotransposon-related DNA sequences in the centromeres of grass chromosomes. Genetics **150:** 1615-1623

Miura A, Yonebayashi S, Watanabe K, Toyama T, Shimada H, Kakutani T (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. Nature **411:** 212-214

Miyao A, Tanaka K, Murata K, Sawaki H, Takeda S, Abe K, Shinozuka Y, Onosato K, Hirochika H (2003) Target site specificity of the Tos17 retrotransposon shows a preference for insertion within genes and against insertion in retrotransposon-rich regions of the genome. Plant Cell **15:** 1771-1780

Monk M, Boubelik M, Lehnert S (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. Development **99:** 371-382

Moore G, Lucas H, Batty N, Flavell R (1991) A family of retrotransposons and associated genomic variation in wheat. Genomics **10:** 461-468

Moore GP (1983) slipped-strand mispairing and evolution of introns. Trends Biochem Sci **8:** 411-414

Moore SP, Powers M, Garfinkel DJ (1995) Substrate-specificity of Ty1 integrase. J Virol **69:** 4683-4692

Mroczek RJ, Dawe RK (2003) Distribution of retroelements in centromeres and neocentromeres of maize. Genetics **165:** 809-819

Mukai Y, Endo TR, Gill BS (1991) Physical mapping of the 18S.26S rRNA multigene family in common wheat: Identification of a new locus. Chromosoma **100:** 71-78

Murata M (2002) Telomeres and centromeres in plants. Curr Genet **3:** 527-538

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J (2004) Sequencing of a rice centromere uncovers active genes. Nat Genet **36:** 138-145

Nagaki K, Neumann P, Zhang D, Ouyang S, Buell CR, Cheng Z, Jiang J (2005) Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol Biol Evol **22:** 845-855

Nagaki K, Song J, Stupar RM, Parokonny AS, Yuan Q, Ouyang S, Liu J, Hsiao J, Jones KM, Dawe RK, Buell CR, Jiang J (2003a) Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. Genetics **163:** 759-770

Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J (2003b) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. Genetics **163:** 1221-1225

Nagaki K, Tsujimoto H, Isono K, Sasakuma T (1995) Molecular characterization of a tandem repeat, Afa family, and its distribution among *Triticeae*. Genome **38:** 479-486

Nagaki K, Tsujimoto H, Sasakuma T (1998a) A novel repetitive sequence of sugar cane, SCEN family, locating on centromeric regions. Chromosome Res **6:** 295-302

Nagaki K, Tsujimoto H, Sasakuma T (1998b) H genome specific repetitive sequence, pEt2, of *Elymus trachycaulus* in part of Afa family of *Triticeae*. Genome **41:** 134-136

Nair AS, Teo CH, Schwarzacher T, Heslop-Harrison JS, (2005) Genome classification of banana cultivars from South India using IRAP markers. . Euphytica **144:** 285-290

Nakajima R, Noma K, Ohtsubo H, Ohtsubo E (1996) Identification and characterization of two tandem repeat sequences (TrsB and TrsC) and a retrotransposon (RIRE1) as genome-general sequences in rice. Genes Genet Syst **71:** 373-382

Nan X, Cross S, Bird A (1998) Gene silencing by methyl-CpG-binding proteins. Novartis Found Symp **214:** 6-16

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature **393:** 386-389

Neves N, Silva M, Heslop-Harrison JS, Viegas W (1997) Nucleolar dominance in *triticales*: control by unlinked genes. Chromosome Res **5:** 125-131

Nicolas AL, Munz PL, Young CS (1995) A modified single-strand annealing model best explains the joining of DNA double-strand breaks mammalian cells and cell extracts. Nucl Acids Res **23:** 1036-1043

Nielsen ML, Hermansen TD, Aleksenko A (2001) A family of DNA repeats in *Aspergillus nidulans* has assimilated degenerated retrotransposons, Mol Genet Genomics **26:** 883–887

Nishimura A, Morita M, Nishimura Y, Sugino Y (1990) A rapid and highly efficient method for preparation of competent *Escherichia coli* cells. Nucl Acids Res **18:** 6169

Nishimura M, Hayashi N, Jwa NS, Lau GW, Hamer JE, Hasebe A (2000) Insertion of the LINE retrotransposon MGL causes a conidiophore pattern mutation in *Magnaporthe grisea*. Mol Plant Microbe Interact **13:** 892-894

Nix DA, Eisen MB (2005) GATA: a graphic alignment tool for comparative sequence analysis. BMC Bioinformatics **6:** 9

Nolan T, Braccini L, Azzalin G, De Toni A, Macino G, Cogoni C (2005) The post-transcriptional gene silencing machinery functions independently of DNA methylation to repress a LINE1-like retrotransposon in *Neurospora crassa*. Nucl Acids Res **33:** 1564-1573

Noma K, Nakajima R, Ohtsubo H, Ohtsubo E (1997) RIRE1, a retrotransposon from wild rice *Oryza australiensis*. Genes Genet Syst **72:** 131-140

Noma K, Ohtsubo E (2000) Tnat1 and Tnat2 from *Arabidopsis thaliana*: novel transposable elements with tandem repeat sequences. DNA Res **7:** 1-7

Nonomura K, Kurata N (2001) The centromere composition of multiple repetitive sequences on rice chromosome 5. Chromosoma **110:** 284-291

Nonomura KI, Kurata N (1999) Organization of the 1.9-kb repeat unit RCE1 in the centromeric region of rice chromosomes. Mol Gen Genet **261:** 1-10

Nsabimana A, van Staden J (2006) Ploidy investigation of bananas (*Musa* spp.) from the National Banana Germplasm Collection at Rubona–Rwanda by flow cytometry. South African J Bot **72:** 302-305

Ohta T (1980) Evolution and variation of multigene families. Springer-Verlag, New York

Ohtsubo H, Kumekawa N, Ohtsubo E (1999) RIRE2, a novel *gypsy*-type retrotransposon from rice. Genes Genet Syst **74:** 83-91

Okada N, Ohshima K (1995) Evolution of tRNA-derived SINEs. *In* RJ Maraia, ed, The impact of short interspersed elements (SINEs) on the host genome. R.G Landes, Austin, Texas, pp 61-79

Okano M, Bell DW, Haber DA, Li E (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. Cell **99:** 247-257

Okano M, Xie S, Li E (1998) Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. Nucl Acids Res **26:** 2536-2540

Okumura K, Kiyama R, Oishi M (1987) Sequence analyses of extrachromosomal *Sau3A* and related family DNA: analysis of recombination in the excision event. Nucl Acids Res **15:** 7477-7489

Onguso JM, Kahangi EM, Ndiritu DW, Mizutani F (2004) Genetic characterization of cultivated bananas and plantains in Kenya by RAPD markers. Sci Hortic **99:** 9-20

Oriero CE, Odunola OA, Lokko Y, Ingelbrecht I (2006) Analysis of B-genome derived simple sequence repeat (SSR) markers in *Musa* spp. African J Biotech **5:** 126-128

Ortiz R (1997) Occurrence and Inheritance of 2*n* Pollen in *Musa*. Ann Bot **79:** 449-453

Oselebe HO, Tenkouano A, Pillay M (2006a) Ploidy variation of *Musa* hybrids from crosses. African J Biotech **5:** 1048-1053

Oselebe HO, Tenkouano A, Pillay M, Obi IU, Uguru MI (2006b) Ploidy and genome segregation in *Musa* breeding populations assessed by flow cytometry and randomly amplified polymorphic DNA markers. J Am Soc Hortic Sci **131:** 780-786

Osuji JO, Crouch J, Harrison G, Heslop-Harrison JS (1998) Molecular cytogenetics of *Musa* species, cultivars and hybrids: location of 18S-5.8S-25S and 5S rDNA and telomere-like sequences. Ann Bot **82:** 243-248

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

Osuji JO, Okoli BE, Ortiz R (1996) A improved procedure for mitotic studies of the *Eumusa* section of the genus *Musa* L. (*Musaceae*). InforMusa **5:** 12-14

Page BT, Wanous MK, Birchler JA (2001) Characterization of a maize chromosome 4 centromeric sequence: evidence for an evolutionary relationship with the B chromosome centromere. Genetics **159:** 291-302

Panstruga R, Buschges R, Piffanelli P, Schulze-Lefert P (1998) A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucl Acids Res **26:** 1056-1062

Parniske M, Jones JD (1999) Recombination between diverged clusters of the tomato Cf-9 plant disease resistance gene family. Proc Natl Acad Sci USA **96:** 5850-5855

Patterson GI, Thorpe CJ, Chandler VL (1993) Paramutation, an allelic interaction, is associated with a stable and heritable reduction of transcription of the maize b regulatory gene. Genetics **135:** 881-894

Paule MR, White RJ (2000) Survey and summary: transcription by RNA polymerases I and III. Nucl Acids Res **28:** 1283-1298

Pearce SR, Harrison G, HeslopHarrison PJS, Flavell AJ, Kumar A (1997) Characterization and genomic organization of Ty1-*copia* group retrotransposons in rye (*Secale cereale*). Genome **40:** 617-625

Pearce SR, Harrison G, Li D, Heslop-Harrison J, Kumar A, Flavell AJ (1996a) The Ty1-*copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet **250:** 305-315

Pearce SR, Knox M, Ellis TH, Flavell AJ, Kumar A (2000) Pea Ty1-*copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in *Pisum*. Mol Gen Genet **263:** 898-907

Pearce SR, Pich U, Harrison G, Flavell AJ, Heslop-Harrison JS, Schubert I, Kumar A (1996b) The Ty1-*copia* group retrotransposons of *Allium cepa* are

distributed throughout the chromosomes but are enriched in the terminal heterochromatin. Chromosome Res **4:** 357-364

Pearce SR, Stuart-Rogers C, Knox MR, Kumar A, Ellis TH, Flavell AJ (1999) Rapid isolation of plant Tyl-*copia* group retrotransposon LTR sequences for molecular marker studies. Plant J **19:** 711-717

Pelissier T, Tutois S, Deragon JM, Tourmente S, Genestier S, Picard G (1995) Athila, a new retroelement from *Arabidopsis thaliana*. Plant Mol Biol **29:** 441-452

Pelissier T, Tutois S, Tourmente S, Deragon JM, Picard G (1996) DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in Athila retroelement sequences. Genetica **97:** 141-151

Pelisson A, Mejlumian L, Robert V, Terzian C, Bucheton A (2002) *Drosophila* germline invasion by the endogenous retrovirus *gypsy*: involvement of the viral env gene. Insect Biochem Mol Biol **32:** 1249-1256

Pelsy F, Merdinoglu D (2002) Complete sequence of *Tvv1*, a family of *Ty1 copia*-like retrotransposons of *Vitis vinifera* L., reconstituted by chromosome walking. Theor Appl Genet **105:** 614-621

Peschke VM, Phillips RL (1991) Activation of the maize transposable element Suppressor-mutator (Spm) in tissue culture. Theor Appl Genet **81:** 90-97

Peschke VM, Phillips RL, Gengenbach BG (1987) Discovery of transposable element activity among progeny of tissue culture-derived maize plants. Science **238:** 804-807

Peterson-Burch BD, Voytas DF (2002) Genes of the *Pseudoviridae* (Tyl/*copia* retrotransposons). Mol Biol Evol **19:** 1832-1845

Petrov D (1997) Slow but Steady: Reduction of Genome Size through Biased Mutation. Plant Cell **9:** 1900-1901

Petrov DA, Lozovskaya ER, Hartl DL (1996) High intrinsic rate of DNA loss in *Drosophila*. Nature **384:** 346-349

Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL (2000) Evidence for DNA loss as a determinant of genome size. Science **287:** 1060-1062

Peumans WJ, Proost P, Swennen RL, Van Damme EJ (2002) The abundant class III chitinase homolog in young developing banana fruits behaves as a transient vegetative storage protein and most probably serves as an important supply of

amino acids for the synthesis of ripening-associated proteins. Plant Physiol **130:** 1063-1072

Pich U, Fuchs J, Schubert I (1996) How do *Alliaceae* stabilize their chromosome ends in the absence of TTTAGGG sequences? Chromosome Res **4:** 207-213

Pich U, Schubert I (1998) Terminal heterochromatin and alternative telomeric sequences in *Allium cepa*. Chromosome Res **6:** 315-321

Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, Panaud O (2006) Doubling genome size without polyploidization: Dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. Genome Res **16:** 1262-1269

Pillay M, Ogundiwin E, Tenkouano A, Dolezel J (2006) Ploidy and genome composition of Musa germplasm at the International Institute of Tropical Agriculture (IITA). African J Biotech **5:** 1224-1232

Porceddu A, Albertini E, Barcaccia G, Marconi G, Bertoli FB, Veronesi F (2002) Development of S-SAP markers based on an LTR-like sequence from *Medicago sativa* L. Mol Genet Genomics **267:** 107-114

Pouteau S, Huttner E, Grandbastien MA, Caboche M (1991) Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. EMBO J **10:** 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

Preston BD (1996) Error-prone retrotransposition: rime of the ancient mutators. Proc Natl Acad Sci USA **93:** 7427-7431

Provan J, Thomas WTB, Forster BP, Powell W (1999) *Copia*-SSR: A simple marker technique which can be used on total genomic DNA. Genome **42:** 363-366

Purseglove JW (1972) Tropical Crops Monocotyledons Longman, London

Purugganan MD, Wessler SR (1994) Molecular evolution of *magellan*, a maize Ty3/*gypsy*-like retrotransposon. Proc Natl Acad Sci USA **91:** 11674-11678

Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2004) Retrotransposon-based molecular markers for linkage and genetic diversity analysis in wheat. Mol Genet Genomics **271:** 91-97

Raina SN, Mukai Y, Kawaguchi K, Goel S, Jain A (2001) Physical mapping of 18S–5·8S–26S and 5S ribosomal RNA gene families in three important vetches

(*Vicia* species) and their allied taxa constituting three species complexes. Theor Appl Genet **103:** 839-845

Ramsay L, Macaulay M, Cardle L, Morgante M, degli Ivanissevich S, Maestri E, Powell W, Waugh R (1999) Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. Plant J **17:** 415-425

Rayburn AL, Gill BS (1986) Molecular analysis of the D-genome of the *Triticeae*. Theor Appl Genet **73:** 385-388

Rea S, Eisenhaber F, O'Carroll D, Strahl BD, Sun ZW, Schmid M, Opravil S, Mechtler K, Ponting CP, Allis CD, Jenuwein T (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature **406:** 593-599

Richards EJ, Chao S, Vongs A, Yang J (1992) Characterization of *Arabidopsis thaliana* telomeres isolated in yeast. Nucl Acids Res **20:** 4039-4046

Richards EJ, Goodman HM, Ausubel FM (1991) The centromere region of *Arabidopsis thaliana* chromosome 1 contains telomere-similar sequences. Nucl Acids Res **19:** 3351-3357

Richardson DL (1961) Note on the reversion of the dwarf banana "cocos" to Gros michel. Tropical Agriculture **38:** 35-37

Rodhe W, Kullaya A, Rodriguez J, Ritter E (1995) Genome analysis of cocos nucifera L. by PCR amplification of spacer sequences separating a subset of *copia*-like EcoRI repetitive elements. J Genet Breed **49:** 179-186

Rohde W (1996) Inverse sequence-tagged repeat (ISTR) analysis, a novel and universal PCR-based technique for genome analysis in the plant and animal kingdom. J Genet Breed **50:** 249-261

Rojas AA, Vazquez-Tello A, Ferbeyre G, Venanzetti F, Bachmann L, Paquin B, Sbordoni V, Cedergren R (2000) Hammerhead-mediated processing of satellite pDo500 family transcripts from *Dolichopoda* cave crickets. Nucl Acids Res **28:** 4037-4043

Round EK, Flowers SK, Richards EJ (1997) *Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure. Genome Res **7:** 1045-1053

Roux NS, Strosse H, RToloza A, Panis B, Doležel J (2004) Detecting ploidy level instability of banana embryogenic cell suspension cultures by flow cytometric. .

In MS Jain, R Swennen, eds, Banana Improvement: Cellular, Molecular Biology, and Induced Mutations. Science Publishers, Inc., Enfield, pp pp 251-261

Rudd MK, Wray GA, Willard HF (2006) The evolutionary dynamics of alpha-satellite. Genome Res **16:** 88-96

Sabot F, Schulman AH (2007) Template switching can create complex LTR retrotransposon insertions in *Triticeae* genomes. BMC Genomics **8:** 247

Sabot F, Sourdille P, Bernard M (2005) Advent of a new retrotransposon structure: the long form of the Veju elements. Genetica **125:** 325-332

Sabot F, Sourdille P, Chantret N, Bernard M (2006) Morgane, a new LTR retrotransposon group, and its subfamilies in wheats. Genetica **128:** 439-447

Saito Y, Kanai Y, Sakamoto M, Saito H, Ishii H, Hirohashi S (2002) Overexpression of a splice variant of DNA methyltransferase 3b, DNMT3b4, associated with DNA hypomethylation on pericentromeric satellite regions during human hepatocarcinogenesis. Proc Natl Acad Sci USA **99:** 10060-10065

Salser W, Bowen S, Browne D, el-Adli F, Fedoroff N, Fry K, Heindell H, Paddock G, Poon R, Wallace B, Whitcome P (1976) Investigation of the organization of mammalian chromosomes at the DNA sequence level. Fed Proc **35:** 23-35

Sandoval JA, Cote FX, Escoute J (1996) Chromosome number variations in micropropagated true-to-type and off-type banana plants (*Musa* AAA Grande Naine cv.). In Vitro Cell Dev Biol Plant **32:** 14-17

SanMiguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot **82:** 37-44

SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet **20:** 43-45

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, MelakeBerhan 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

Saunders VA, Houben A (2001) The pericentromeric heterochromatin of the grass *Zingeria biebersteiniana* (2n = 4) is composed of Zbcen1-type tandem repeats that are intermingled with accumulated dispersedly organized sequences. Genome **44:** 955-961

Saze H, Mittelsten Scheid O, Paszkowski J (2003) Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. Nat Genet **34:** 65-69

Schmid CW (1998) Does SINE evolution preclude Alu function? Nucl Acids Res **26:** 4541-4550

Schmid CW, Shen C-KJ (1985) The evolution of interspersed repetitive DNA sequences in mammals and other vertebrates. *In* RJ MacIntyre, ed, Molecular evolutionary genetics. Plenum Press, New York, pp 323-358

Schmidt T, Heslop-Harrison JS (1994) Variability and evolution of highly repeated DNA sequences in the genus *Beta*. Genome **36:** 1074-1079

Schmidt T, Heslop-Harrison JS (1998) Genomes, genes and junk: the large scale organization of plant chromosomes. Trends Plant Sci **3:** 195-199

Schmidt T, Kubis S, Heslop-Harrison JS (1995) Analysis and chromosomal localization of retrotransposons in sugar beet (*Beta vulgaris* L.): LINEs and Ty1-*copia*-like elements as major components of the genome. Chromosome Res **3:** 335-345

Schotta G, Ebert A, Krauss V, Fischer A, Hoffmann J, Rea S, Jenuwein T, Dorn R, Reuter G (2002) Central role of *Drosophila* SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J **21:** 1121-1131

Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001) Genomic and genetic definition of a functional human centromere. Science **294:** 109-115

Schwarzacher T (1996) The physical organization of *Triticeae* chromosomes. Symp Soc Exp Biol **50:** 71-75

Schwarzacher T (2003) Meiosis, recombination and chromosomes: a review of gene isolation and fluorescent in situ hybridization data in plants. J Exp Bot **54:** 11-23

Schwarzacher T, Heslop-Harrison JS (2000) Practical *in situ* Hybridization. BIO Scientific Publisher Limited, Oxford

Selinger DA, Chandler VL (2001) B-Bolivia, an allele of the maize b1 gene with variable expression, contains a high copy retrotransposon-related sequence immediately upstream. Plant Physiol **125:** 1363-1379

Selker EU, Freitag M, Kothe GO, Margolin BS, Rountree MR, Allis CD, Tamaru H (2002) Induction and maintenance of nonsymmetrical DNA methylation in *Neurospora*. Proc Natl Acad Sci USA **99:** 16485-16490

Shepherd NS, Schwarz-Sommer Z, Blumberg vel Spalve J, Gupta M, Wienand U, Saedler H (1984) Similarity of the Cin1 repetitive family of *Zea mays* to eukaryotic transposable elements. Nature **307:** 185-187

Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N (1997) Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature **388:** 666-670

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res **10:** 908-915

Silayoi B, Chomchalow N (1987) Cytotaxonomic andmorphological studies of Thai banana cultivars. *In* GJ Persley, E De Langhe, eds, Proc. Banana and Plantain Breeding Strategies. . ACICR Proc., Canberra, pp 88-98

Simmonds NW (1954) Isolation in *Musa*, Sections *Eumusa* and *Rhodochlamys*. Evolution **8:** 65-74

Simmonds NW (1962) The Evolution of the Bananas. Longman, London

Simmonds NW (1976) Bananas. *In* N.W.Simmonds, ed, Evolution of Crop Plants. Longman, London and New York, pp 211–215

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

Singer MF (1982) SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. Cell **28:** 433-434

Singer T, Yordan C, Martienssen RA (2001) Robertson's Mutator transposons in *A. thaliana* are regulated by the chromatin-remodeling gene Decrease in DNA Methylation (DDM1). Genes Dev **15:** 591-602

Siroky J, Castiglione MR, Vyskot B (1998) DNA methylation patterns of *Melandrium album* chromosomes. Chromosome Res **6:** 441-446

Sleutels F, Barlow DP (2002) The origins of genomic imprinting in mammals. Adv Genet **46:** 119-163

Smith GP (1974) Unequal crossover and the evolution of multigene families. Cold Spring Harb Symp Quant Biol **38:** 507-513

Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. Science **191:** 528-535

Smýkal P (2006) Development of an efficient retrotransposon-based fingerprinting method for rapid pea variety identification. J Appl Genet **47:** 221-230

Smyth DR (1993) Plant retrotransposons. *In* DPS Verman, ed, Control of Plant Gene Expression. CRC Press, Inc., Boca Raton, Florida, pp 1-15

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

Song SU, Gerasimova T, Kurkulos M, Boeke JD, Corces VG (1994) An env-like protein encoded by a *Drosophila* retroelement: evidence that *gypsy* is an infectious retrovirus. Genes Dev **8:** 2046-2057

Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167:** GC1-10

Soppe WJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, Peeters AJ (2000) The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. Mol Cell **6:** 791-802

Southern EM (1970) Base sequence and evolution of Guinea pig-satellite DNA. Nature **227:** 794-798

Stack SM, Anderson LK (2001) A model for chromosome structure during the mitotic and meiotic cell cycles. Chromosome Res **9:**175-198

Staginnus C, Gregor W, Mette F, Teo CH, Borroto-Fernandez EG, Machado MLDC, Matzke MA, Schwarzacher T (2007) Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. BMC Plant Biol **7:** 24

Stover RH, Simmonds NW (1987) Bananas. Longman Sci and Techn, Harlow

Streisinger G, Okada Y, Emrich J, Newton J, Tsugita A, Terzaghi E, Inouye M (1966) Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. Cold Spring Harb Symp Quant Biol **31:** 77-84

Stupar RM, Song J, Tek AL, Cheng Z, Dong F, Jiang J (2002) Highly condensed potato pericentromeric heterochromatin contains rDNA-related tandem repeats. Genetics **162:** 1435-1444

Sun X, Wahlstrom J, Karpen G (1997) Molecular structure of a functional *Drosophila* centromere. Cell **91:** 1007-1019

Sunako T, Sakuraba W, Senda M, Akada S, Ishikawa R, Niizeki M, Harada T (1999) An allele of the ripening-specific 1-aminocyclopropane-1-carboxylic acid

synthase gene (ACS1) in apple fruit with a long storage life. Plant Physiol **119:** 1297-1304

Sỳkorova E, Leitch AR, Fajkus J (2006) Asparagales telomerases which synthesize the human type of telomeres. Plant Mol Biol **60:** 633-646

Sỳkorova E, Lim KY, Chase MW, Knapp S, Leitch IJ, Leitch AR, Fajkus J (2003) The absence of *Arabidopsis*-type telomeres in *Cestrum* and closely related genera *Vestia* and *Sessea* (*Solanaceae*): first evidence from eudicots. Plant J **34:** 283-291

Syomin BV, Ilyin YV (2004) The functional motifs that are revealed in the *gypsy* Gag amino acid sequence. Dokl Biochem Biophys **398:** 291-293

Takeda S, Sugimoto K, Otsuki H, Hirochika H (1999) A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. Plant J **18:** 383-393

Taketa S, Ando H, Takeda K, von Bothmer R (1999) Detection of *Hordeum marinum* genome in three polyploid *Hordeum* species and cytotypes by genomic in situ hybridization. Hereditas **130:** 185-188

Talbert PB, Masuelli R, Tyagi AP, Comai L, Henikoff S (2002) Centromeric localization and adaptive evolution of an *Arabidopsis* histone H3 variant. Plant Cell **14:** 1053-1066

Tamaru H, Selker EU (2001) A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature **414:** 277-283

Tanaka Y, Kurumizaka H, Yokoyama S (2005) CpG methylation of the CENP-B box reduces human CENP-B binding. FEBS J **272:** 282-289

Tanskanen JA, Sabot F, Vicient C, Schulman AH (2007) Life without GAG: the BARE-2 retrotransposon as a parasite's parasite. Gene **390:** 166-174

Tarfor KD (1974) Unequal mitotic sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancy. Cold Spring Harb Symp Quant Biol **38:** 491-500

Tek AL, Jiang J (2004) The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. Chromosoma **113:** 77-83

Tek AL, Song J, Macas J, Jiang J (2005) Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. Genetics **170:** 1231-1238

Temin HM (1980) Origin of retroviruses from cellular moveable genetic elements. Cell **21:** 599-600

Teo CH (2002) Classification of the Malaysian banana varieties using retrotransposon markers. University Putra Malaysia, Serdang

Teo CH, Tan SH, Ho CL, Faridah QZ, Othman YR, Heslop-Harrison JS, Kalendar R, Schulman AH (2005) Genome constitution and classification using retrotransposon-based markers in the orphan crop banana. J. Plant Biol. **48:** 96-105

Teo CH, Tan SH, Othman YR, Schwarzacher T (2002) The cloning of Ty1-*copia*-like retrotransposons from 10 varieties of banana (*Musa* Sp.). J Biochem Mol Biol Biophys **6:** 193-201

Terol J, Castillo MC, Bargues M, Perez-Alonso M, de Frutos R (2001) Structural and evolutionary analysis of the *copia*-like elements in the *Arabidopsis thaliana* genome. Mol Biol Evol **18:** 882-892

Tran RK, Zilberman D, de Bustos C, Ditt RF, Henikoff JG, Lindroth AM, Delrow J, Boyle T, Kwong S, Bryson TD, Jacobsen SE, Henikoff S (2005) Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*. Genome Biol **6:** R90

Tregear JW, Morcillo F, Richaud F, Berger A, Singh R, Cheah SC, Hartmann C, Rival A, Duval Y (2002) Characterization of a defensin gene expressed in oil palm inflorescences: induction during tissue culture and possible association with epigenetic somaclonal variation events. J Exp Bot **53:** 1387-1396

Turcich MP, BokhariRiza A, Hamilton DA, He CP, Messier W, Stewart CB, Mascarenhas JP (1996) PREM-2, a *copia*-type retroelement in maize is expressed preferentially in early microspores. Sex Plant Reprod **9:** 65-74

Turcich MP, Mascarenhas JP (1994) PREM-1, a putative maize retroelement has LTR (long terminal repeat) sequences that are preferentially transcribed in pollen. Sex Plant Reprod **7:** 2-11

Ude G, Pillay M, Nwakanma D, Tenkouano A (2002a) Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. Theor Appl Genet **104:** 1239-1245

Ude G, Pillay M, Nwakanma D, Tenkouano A (2002b) Genetic Diversity in *Musa acuminata* Colla and *Musa balbisiana* Colla and some of their natural hybrids using AFLP Markers. Theor Appl Genet **104:** 1246-1252

Unoki M, Nishidate T, Nakamura Y (2004) ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. Oncogene **23:** 7601-7610

Usakin L, Abad J, Vagin VV, de Pablos B, Villasante A, Gvozdev VA (2007) Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in *Drosophila melanogaster* ovaries. Genetics **176:** 1343-1349

Valárik M, Simkova H, Hribova E, Safar J, Dolezelova M, Dolezel J (2002) Isolation, characterization and chromosome localization of repetitive DNA sequences in bananas (*Musa* spp.). Chromosome Res **10:** 89-100

Valmayor RV (1991) Bananas and plantains in the Philippines. *In* RV Valmayor, ed, Banana and Plantains R & D in Asia and the Pacific. INIBAP, Montpellier, France, pp 87-120

Valmayor RV (2001) *Musa exotica*. Philippine Agricultural Scientist **84:** 325-331

Valmayor RV, Danh L, Häkkinen M (2004) Rediscovery of *Musa splendida* A. Chevalier and description of two new species (*Musa viridis* and *Musa lutea*). Philippine Agricultural Scientist **87:** 110-118

Valmayor RV, Jones DR, Subijanto, Polprasid P, Jamaluddin SH (1990) Bananas and plantains in Southeast Asia. INIBAP, Montpellier, France

van Leeuwen H, Monfort A, Zhang HB, Puigdomenech P (2003) Identification and characterisation of a melon genomic region containing a resistance gene cluster from a constructed BAC library. Microcolinearity between *Cucumis melo* and *Arabidopsis thaliana*. Plant Mol Biol **51:** 703-718

Varagona MJ, Purugganan M, Wessler SR (1992) Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. Plant Cell **4:** 811-820

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D (2004) RNAi-mediated targeting of heterochromatin by the RITS complex. Science **303:** 672-676

Vernhettes S, Grandbastien MA, Casacuberta JM (1998) The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. Mol Biol Evol **15:** 827-836

Verriès C, Bes C, This P, Tesniere C (2000) Cloning and charaterization of Vine-1, an LTR-retrotransposon-like element in *Vitis vinifera* L and other *Vitis* species. Genome **43:** 366-376

Vershinin AV, Druka A, Alkhimova AG, Kleinhofs A, Heslop-Harrison JS (2002) LINEs and *gypsy*-like retrotransposons in Hordeum species. Plant Mol Biol **49:** 1-14

Vershinin AV, Schwarzacher T, Heslop-Harrison JS (1995) The large-scale genomic organization of repetitive DNA families at the telomeres of rye chromosomes. Plant Cell **7:** 1823-1833

Vershinin AV, Svitashev S, Gummesson P-O, Salomon B, R. RvB, Bryngelsson T (1994) Characterization of a family of tandemly repeated DNA sequences in *Triticeae*. Theor Appl Genet **89:** 217-225

Vicient CM, Kalendar R, Anamthawat-Jonsson K, Suoniemi A, Schulman AH (1999) Structure, functionality, and evolution of the BARE-1 retrotransposon of barley. Genetica **107:** 53-63

Vicient CM, Kalendar R, Schulman AH (2001) Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Res **11:** 2041-2049

Vicient CM, Kalendar R, Schulman AH (2005) Variability, recombination, and mosaic evolution of the barley BARE-1 retrotransposon. J Mol Evol **61:** 275-291

Vicient CM, Martinez-Izquierdo JA (1997) Discovery of a Zdel transposable element in *Zea* species as a consequence of a retrotransposon insertion. Gene **184:** 257-261

Vitte C, Bennetzen JL (2006) Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc Natl Acad Sci USA **103:** 17638-17643

Vitte C, Panaud O (2003) Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. Mol Biol Evol **20:** 528-540

Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. Cytogenet Genome Res **110:** 91-107

Volpe T, Schramke V, Hamilton GL, White SA, Teng G, Martienssen RA, Allshire RC (2003) RNA interference is required for normal centromere function in fission yeast. Chromosome Res **11:** 137-146

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SI, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science **297:** 1833-1837

Voytas DF, Ausubel FM (1988) A *copia*-like transposable element family in *Arabidopsis thaliana*. Nature **336:** 242-244

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

Voytas DF, Konieczny A, Cummings MP, Ausubel FM (1990) The structure, distribution and evolution of the Ta1 retrotransposable element family of *Arabidopsis thaliana*. Genetics **126:** 713-721

Walbot V, Warren C (1990) DNA methylation in the Alcohol dehydrogenase-1 gene of maize. Plant Mol Biol **15:** 121-125

Walsh JB (1987) Persistence of tandem arrays: Implications for satellite and simple-sequence DNAs. Genetics **115:** 555-567

Wang S, Zhang Q, Maughan PJ, Saghai Maroof MA (1997) *Copia*-like retrotransposons in rice: sequence heterogeneity, species distribution and chromosomal locations. Plant Mol Biol **33:** 1051-1058

Wang SP, Liu N, Peng KM, Zhang QF (1999) The distribution and copy number of *copia*-like retrotransposons in rice (*Oryza sativa* L.) and their implications in the organization and evolution of the rice genome. Proc Natl Acad Sci USA **96:** 6824-6828

Wang Y, Tang X, Cheng Z, Mueller L, Giovannoni J, Tanksley SD (2006) Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. Genetics **172:** 2529-2540

Warburton P, Willard H (1996) Evolution of centromeric alpha satellite DNA: molecular organization within and between human and primate chromosomes. *In* M Jackson, T Strachan, G Dover, eds, Genome Evolution. BIOS Scientific, Guildford, UK, pp 121-145

Warburton PE, Haaf T, Gosden J, Lawson D, Willard HF (1996) Characterization of a chromosome-specific chimpanzee alpha satellite subset: evolutionary relationship to subsets on human chromosomes. Genomics **33:** 220-228

Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BB, Powell W (1997) Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol Gen Genet **253:** 687-694

Wellinger RJ, Wolf AJ, Zakian VA (1993) *Saccharomyces* telomeres acquire single-strand TG1-3 tails late in S phase. Cell **72:** 51-60

Wessler SR (1996) Turned on by stress. Plant retrotransposons. Curr Biol **6:** 959-961

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev **5:** 814-821

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J **26:** 307-316

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. Plant Cell **15:** 1186-1197

Williamson VM (1983) Transposable elements in yeast. Int Rev Cytol **83:** 1-25

Witte CP, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA **98:** 13778-13783

Witte CP, Tiller S, Isidore E, Davies HV, Taylor MA (2005) Analysis of two alleles of the urease gene from potato: polymorphisms, expression, and extensive alternative splicing of the corresponding mRNA. J Exp Bot **56:** 91-99

Wong C, Kiew R, Argent G, Set O, Lee SK, Gan YY (2002) Assessment of the validity of the sections in *Musa* (musaceae) using AFLP. Ann Bot **90:** 231-238

Wong LH, Choo KH (2004) Evolutionary dynamics of transposable elements at the centromere. Trends Genet **20:** 611-616

Wright DA, Ke N, Smalle J, Hauge BM, Goodman HM, Voytas DF (1996) Multiple non-LTR retrotransposons in the genome of *Arabidopsis thaliana*. Genetics **142:** 569-578

Wright DA, Voytas DF (1998) Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/*gypsy* retrotransposons that encode envelope-like proteins. Genetics **149:** 703-715

Wright DA, Voytas DF (2002) Athila4 of *Arabidopsis* and Calypso of soybean define a lineage of endogenous plant retroviruses. Genome Res **12:** 122-131

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J **9:** 3353-3362

Yan H, Ito H, Nobuta K, Ouyang S, Jin W, Tian S, Lu C, Venu RC, Wang GL, Green PJ, Wing RA, Buell CR, Meyers BC, Jiang J (2006) Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. Plant Cell **18:** 2123-2133

Yang TJ, Lee S, Chang SB, Yu Y, de Jong H, Wing RA (2005) In-depth sequence analysis of the tomato chromosome 12 centromeric region: identification of a large CAA block and characterization of pericentromere retrotranposons. Chromosoma **114:** 103-117

Yano ST, Panbehi B, Das A, Laten HM (2005) Diaspora, a large family of Ty3-*gypsy* retrotransposons in *Glycine max*, is an envelope-less member of an endogenous plant retrovirus lineage. BMC Evol Biol **5:** 30

Yao J, Dong Y, Morris BA (2001) Parthenocarpic apple fruit production conferred by transposon insertion mutations in a MADS-box transcription factor. Proc Natl Acad Sci USA **98:** 1306-1311

Yoder JA, Walsh CP, Bestor TH (1997) Cytosine methylation and the ecology of intragenomic parasites. Trends Genet **13:** 335-340

Yu GX, Wise RP (2000) An anchored AFLP- and retrotransposon-based map of diploid *Avena*. Genome **43:** 736-749

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R,

Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science **296:** 79-92

Zemach A, Li Y, Wayburn B, Ben-Meir H, Kiss V, Avivi Y, Kalchenko V, Jacobsen SE, Grafi G (2005) DDM1 binds *Arabidopsis* methyl-CpG binding domain proteins and affects their subnuclear localization. Plant Cell **17:** 1549-1558

Zhang W, Yi C, Bao W, Liu B, Cui J, Yu H, Cao X, Gu M, Liu M, Cheng Z (2005) The transcribed 165-bp CentO satellite is the major functional centromeric element in the wild rice species *Oryza punctata*. Plant Physiol **139:** 306-315

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell **14:** 2825-2836

Zhou Y, Cambareri EB, Kinsey JA (2001) DNA methylation inhibits expression and transposition of the *Neurospora* Tad retrotransposon. Mol Genet Genomics **265:** 748-754

Zou S, Voytas DF (1997) Silent chromatin determines target preference of the *Saccharomyces* retrotransposon Ty5. Proc Natl Acad Sci USA **94:** 7412-7416

Zufall RA, Rausher MD (2003) The genetic basis of a flower color polymorphism in the common morning glory (*Ipomoea purpurea*). J Hered **94:** 442-448