Genome organisation, evolution and biodiversity in *Musa*: Application to stress-related gene discovery and plant breeding

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Azhar Mohamad
Department of Biology
University of Leicester

December 2006

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 June 2003 to June 2006.

Signed.....

Azhar Mohamad, December 2006

Genome organisation, evolution and biodiversity in *Musa*: Application to stress-related gene discovery and plant breeding

By

Azhar Mohamad

Abstract

Bananas, genus *Musa,* are giant perennial herbs with a small genome. They occur as diploid, triploid and tetraploid plants throughout the tropics with a centre of diversity in South-East Asia. Abiotic and biotic stresses are major factors affecting banana production, but genetic diversity within the gene-pool of stress responsive genes has not been characterized at the molecular level. The identification, isolation and test of expression of both structural and signalling genes are important for breeding purposes. This thesis shows how anonymous markers are used to assay genetic polymorphisms, and then describes the analysis of major biotic-stress related genomic motifs from diverse *Musa* accessions, before identifying Bacterial Artificial Chromosome (BAC) clones carrying biotic and abiotic stress-related DNA motifs. Microsatellite (SSR) and retroelement-related sequences are abundant and can be exploited as anonymous genetic markers in *Musa*. SSR primers designed within sequenced BAC clones identified polymorphisms in *Musa*. LTR-retrotransposon fragments, including *Pseudoviridae* (Ty1-*copia*-like) and *Metaviridae* (Ty3-*gypsy*-like) families were present as different families in BACs with no correspondence to particular *Musa* genomes, but are useful for understanding genome evolution. Primers designed from genomic and EST databases were exploited to characterize sequences containing Nucleotide Binding Sites (NBS) and Leucine-Rich Repeat (LRR) motifs (associated with disease resistance genes), and genes associated with tolerance to heat (HSP, 70HSP), salinity (STP) and drought (DRFP) stress. Neither NBS nor LRR sequences are conserved with respect to genome, indicating that R-genes are specific and exist independently. BAC libraries allow tagging of conserved domains of NBS, LRR, retroelement and SSR motifs, giving understanding of the genomic context and control of R-genes. Reliable characterization of these domains in *Musa* is possible via PCR-based screening. The thesis gives a broad insight into genome organisation, evolution and diversity of major classes of R-genes, enabling progress towards gene discovery and exploitation for plant breeding.

Dedication

I would like earnestly dedicate to my lovely and caring wife Che Hasmah Che Cob, my parents even at a distance Mohamad Yusof and Che Eshah Che Had. And my brothers and sisters whose hands always rise for me for their spiritual inspiration in motivate me to the highest ideals of life in writing the thesis. Also to my children, Amirah, Afiqah and Ahmad Ammar who give me cheers and happiness during the hard time in finishing my course study.

Acknowledgments

I would like to express my sincere gratitude to my supervisor Prof. Pat Heslop Harrison for his inspiring supervision throughout this study, helpful advice and guidance in my writing, technical experiments in molecular and radioactive works, constructive criticism, encouragement, motivation, moral support and understanding in all aspects toward finishing the study.

To Dr. Trude Schwarzacher for giving me her moral support, encouragement, helping and introducing me to molecular cytology works. Dr. Sybile Ellen Kubis and Dr. Colin Ferris deserved to be mentioned for their helpful conversations, motivation and optimistic comments during assessment sessions. Also Dr. Eamonn Mallon the post-graduate tutor who is simple and reliable for conversations.

To Dr. John Bailey for helping me in chromosome isolation, Jean Liggins, Maurice Andrews, Judy Horner for supplementary things in Laboratory and greenhouse, Steve Ison for computer stuff, Sue Ogden and Lesley Barnett for ordering laboratory stuffs and fellows in laboratory who helped me throughout the study. And also clerical staff, Penny Butler, Penny Chamberlain and Elaine Rawdin who were willing to assist at any occasion in helping me with University documents.

To my directors and colleagues at Malaysian Institute for Nuclear Technology Research (MINT), thank you for your supports and encouragements: Dr. Daud Mohamad (Director General), Dr. Nahrul Khair Alang Md. Rashid, Dr. Norimah Yusof, Dr. Rusli Ibrahim, Dr. Khairuddin Abdul Rahim, Dr. Nashriyah Mat, Dr. Mat Rasol Awang, Dr. Sobri Hussein, Dr. Mohd Nazir Basiran, Mahmood Dollah, Norpaiza M. Hasssan, Zainon Othman, Mek Zah, Dr. Khairulzaman Dahlan, Rapieh Aminnuddin, Abd. Rahim Harun, Tuan Sembak Tuan Endot, Rabiah Abu Hassan, Zahariah Hamid and Zakaria Taib. Also, staffs from Public Service Department of Malaysia, Mohd Khidir, Aszhari Mohamad and Hamidah Ayub for their efforts in managing my scholarship and fees.

Last but not least, to all my friends Saadiah Jamli, Teo Chee How, Emmanuel Otwe, Ulrike Anhalt, Husni Ibrahim, Haniza M. Zain, Said Hafidh and Alexandra Contento, for all the laugh, cheerful company, motivation and helping me in finishing my work. To all members of Malay Student Society of Leicester (MASLEICS) I say a big thank you for the moral support during my stay in Leicester.

Finally, a sincere appreciation to Government of Malaysia for sponsorship, the Generation Challenge Program of CGIAR for consumable support and the Global *Musa* Genomics Consortium for access to genomic resources.

Azhar Mohamad

Abbreviations

BAC	Bacterial Artificial Chromosome
AFLP	Amplified Fragment Length Polymorphism
bp, kb, Mbp	base pair, kilo base, Mega base pair
COS	Conserved Orthologue Synteny
DNA	deoxyribonucleic acid
EST	Express Sequence Tags
FITC	Fluorescein Isothiocyanate
IPTG	isopropyl-β-thiogalactosid
l, ml, µl	liter, milliliter, microliter
NBS	Nucleotide Binding Sites
LB	Luria-Bertani
LTR	long terminal repeat
LRR	Leucine Rich Repeat
°C	degree celsius
PCR	Polymerase Chain Reaction
UV	Ultra Violet
RAPD	Random Amplified Polymorphic DNA
RGA	Resistance Gene Analog
RI	Ribonuclease Inhibitor
SH	Southern Hybridization
SNP	single nucleotiode polymorphism
SSR	Single Sequence Repear
sp	species
V	volt
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosid
YAC	Yeast Artificial Chromosome

Table of Contents

2.0 Materials and Methods

CHAPTER 1

1.0 Introduction

1.1 Banana Breeding

1.1.1 Importance

Bananas and plantains, in the genus *Musa*, are the among the world's most important crops, belonging to one of the fewer than 15 species with a global annual production of more than 100 Mt. *Musa* does not only provide an important source of food, but also contributes economically and culturally both in the producer countries and in those temperate countries reliant on its importation. Bananas are both a major fresh fruit species, and a starchy staple energy crop. They are available throughout the year. Much effort has been expended in the domestication of wild accessions and selection among cultivars to generate variation relevant both for breeding purposes and to improve the product for the consumer.

The taxonomy of *Musa*, which comprises about 50 species, has never been fully resolved, and remains a subject of debate (Bartos *et al.*, 2005). Neither the range of diversity, both among the neither wild and cultivated species, nor species hybridity have been clearly described. Cultivated banana can be broadly classified as dessert (for eating raw) and plantain (which requires cooking) types, although there is no clear botanical distinction between these two types. They are free flowering, with a range of maturity period from 9-15 months after planting. In an intensively managed plantation, plants are replaced every 6 to 10 production cycles (giving the opportunity for a relatively rapid dissemination of improved germplasm), although plants can continue to regenerate and produce crops almost indefinitely.

In many countries, ranging from the United Kingdom to Uganda and India, bananas are the major single fresh fruit in the market. There is also a significant commercial interest in bananas for processing into purees, juices, chips or confectionary ingredients. Banana improvement is difficult because of high sterility in the parthenocarpic varieties and the large physical size of the plants (with consequences on breeding timescale, large plantation areas and difficulties in experimental replication).

These factors have limited the extent of the genetic knowledge particularly with respect to unimproved germplasm, and led to a lack of experience in the manipulation of genes which control important plant traits; the genetics of all the other major crops are better known, and specifically, no genetic maps or mapping populations, required for genetic studies, have been developed in *Musa*.

Banana fruits are of dietary importance, as they have a high nutritional value, including fibre, vitamins and minerals. The fruit can be eaten fresh or cooked; the mode of cooking depends on region and culture. For example, in Costa Rica ripening banana flesh is used to make a thick syrup, while in Africa, beer or wine is brewed from ripe fruit. The East African highland banana is harvested unripe and eaten boiled and mashed. Additionally, various parts of the banana plant are used as a source of traditional medicine, for animal feed, for roofing, or various handicrafts, and the species contributes to many cultural and religion functions. Most recently, bananas have been genetically manipulated to express vaccines (Mason *et al.*, 2002).

In Malaysia, some 25,600 ha. has been planted to banana (Jamil, 1998; FAOSTAT, 2004), yielding 0.53Mt of fruit per year. Most of the production is consumed locally by smallholders, so that only 12% is available for export, mainly to Singapore and Brunei (Jamil, 1998). The local diploid bananas popular as a dessert include Mas and Lemak Manis. Other important cultivars are the triploid types such as Rastali, Masak Hijau and Embun. The most common cooking bananas are Raja, Nangka, Awak, Abu and Tanduk (Siti Hawa, 1998).

Banana was one of the first crops to be domesticated. The selection from wild species of edible banana and plantain has been carried out since the prehistoric period, with systematic breeding being applied more recently. But so far breeding programmes have enjoyed only limited success in producing commercially accepted varieties with resistance to disease, high yield and suitable quality (Buddenhagen, 1987; Bakry *et al.*, 2001). Lately, some varieties bred by FHIA and the mutant variety Novaria have shown potential for commercial acceptance.

However, badly characterized genetic variability and the sterility in *Musa* remain as major obstacles for the improvement of banana and plantain via conventional breeding. Sterility is not limited to the triploid types (Agarwal, 1987; Vuylsteke, 1993), but is also present in diploid ones (Raman *et al.,* 1970), where parthenocarpy (the development of fruit without pollination) can occur.

1.1.2 Genomes and taxonomy

Bananas belong to the family Musaceae, which includes a number of giant perennial herbaceous plants. The two main genera within the *Musaceae* are *Ensete* and *Musa*. The genus *Ensete* is composed of monocarpic herbs, none of which produce edible fruits. It is cultivated in Southern Ethiopia as a major staple, with starch prepared from the pseudostem, known as kocho, which is used to make bread. Only two of eight species have any economic importance (*E. ventricosum* and *E. edule*) and these are used, respectively, as a food and a fibre crop. The pseudostem is harvested before maturity or just at flowering (Birmeta *et al.*, 2004). In Africa, *Ensete* is an important crop for food and is also utilised for animal forage, fibre production, as a source of construction materials and as an antimicrobial (Holscher and Schneider, 1998).

Musa species have a relatively small genome of 500-600 Mbp (Lysak *et al.*, 1999), compared to 430Mbp of the model monocotyledon rice. Diploid, triploid and tetraploid forms are all known. The name *Musa paradisiaca* L was introduced in 1753 (Linnaeus, 1753). *Musa* nomenclature has become complicated, with the introduction of new accessions, cultivars and species. According to Karamura (1998), *Musa* x *paradisiaca* is applicable to all hybrids between *M. acuminata* and *M. balbisiana.* In Linnaeus' time, samples were too limited to give a clear picture of diversity and this led to an oversimplified classification. A clear differentiation exists between dessert and plantain types in Africa and Latin America, but cultivars are generally more confused in Southeast Asia.

Either Malaysia (Simmonds, 1962) or Indonesia (Horry *et al.,* 1997) is thought to be the centre of diversity of *M. acuminata,* while *M. balbisiana* originates from India, Myanmar, Thailand and The Philippines. The distribution and spread of *M. acuminata* and *M. balbisiana* began as a result of human selection of seedless, parthenocarpic individuals, which are immediately identifiable and easily propagated. As the centre of *Musa* diversity, Southeast Asia has many cultivars that share more or less similar characteristics but which carry numerous names and synonyms. In addition, different languages, spelling and regional dialects make it difficult to identify specific cultivars.

Remarkably, the highest extent of synonymy occurs in the dessert banana. For example, the diploid cultivar Mas Sucrier is also known as. Mas (Malaysia), Mas (Indonesia), Kluai Khai (Thailand) and Chuoing Trung (Vietnam). But in Malaysia itself, Mas has many other names e.g. Mas Besar, Mas Kampung. Banana taxonomists in Southeast Asia have identified 68 cultivars with synonyms and 81 with unique names (Valmayor *et al.,* 2000). This has revealed that *M. acuminata* are more common than *M. balbisiana* or their hybrids. Furthermore, a number of dual-purpose cooking/dessert banana types have been recognized.

There are four main sections within the genus *Musa,* namely *Australimusa, Callimusa, Rhodochlamys* and *Eumusa*. The chromosome number of the former two is based on 10, whereas in the latter two it is based on 11. The section *Australimusa* is poorly understood, consisting of banana namely Fe'i and includes *M. textilis* (T-genome), which are important for its fibre properties (*M. maclayi, M. peekelii, M. jakeyi* and *M. lolodensis*). The sections *Callimusa* and *Rhodochlamys* are related to ornamental bananas. Currently four wild *Rhodochalamys* species are commonly recognized, i.e. *M. laterita, M. ornata, M. sanguinea* and *M. velutina.* In *Callimusa* only three wild species have been characterized i.e. *M. coccinea, M. gracilis* and *M. violascens. Eumusa* is the most important section, containing the edible and cultivated bananas *M. acuminata* (A-genome), *M. balbisiana* (B-genome) and *M. schizocarpa* (S-genome).

The identification and classification of the S and T genomes are quite recent, as the type species are found mostly in Papua New Guinea (Shepherd *et al.,* 1984; Arnaud *et al.,* 1997). However, these species have contributed to the origin of some New Guinea cultivars (Carreel, 1994), as confirmed by genomic *in situ* hybridization (D'Hont *et al.,* 2000).

The natural hybridization which occurs between wild species has allowed the generation of a spectrum of triploid hybrids (such as AAA, AAB and ABB). The ploidy level, the chromosome composition, the genomic structure, and the phylogenetic relationships between hybrid cultivars are all important determinants of the breeding potential of wild types. Where fertility is an issue, it is also important to understand the extent and type of chromosomal translocations and other rearrangements. Based on flow cytometery analysis, *M. balbisiana* has the smallest genome size (1.03pg/2C, 2n=2x=22) followed by *M. acuminata* (1.11pg/2C, 2n=2x=22), *M. schizocarpa* (1.18pg/2C, 2n=2x=22) and *M. textilis* (1.27pg/2C, 2n=2x=20) (Dolezel *et al.,* 1994; D'hont *et al.,* 2000). *M. acuminata* subspecies vary in genome size, ranging between 1.20pg/2C and 1.30pg/2C, and the analysis becomes uncertain when hybrids are involved, as DNA content can vary widely even between cultivars carrying similar genomes (Kamate *et al.,* 2001). It is clear that *M. acuminata* exists as a heterogeneous mixture of 'A' genomes of different origin. While genome size may vary in newly obtained hybrids due to aneuploidy, established hybrid cultivars are characterized by their euploid chromosome number and they maintain a stable genome size (Argawal 1987; D'hont *et al.,* 2000). The determination of ploidy level is most conveniently done by flow cytometry, which can clearly discriminate between the diploid, triploid and tetraploid individuals.

Banana ploidy is conventionally described by a combination of the letters A, B, S or T to indicate the level and constitution with respect to the donor diploids. In cultivated varieties, evolution has led to vegetative parthenocarpy (seedless fruits) and female sterility.

The hybrid cultivars between and within *M. acuminata*, *M. balbisiana* and *M. schizocarpa* have generated both homogenomic and heterogenomic types i.e. AAA, ABB, AB (ITC 0990, Vunapope, cooking), AB (ITC 0987, Auko, cooking). Although crosses between *M. schizocarpa* and *M. acuminata* (AS) are rare, their hybrids have been found in Papua New Guinea, and are suitable for both dessert and cooking purposes (e.g., ITC 0604 (diploid, Japaraka), ITC 0791 (diploid, Kokor, cooking/desert), ITC 0822 (diploid, Tonton kepa, dessert), ITC 0877 (diploid, Gebi, cooking).

Bananas are parthenocarpic, in which a mass of edible pulp develops without requiring any pollination. Few (if any) seeds are produced, and most often the seeds are non-viable The development of parthenocarpy has been reported to be controlled by the three dominant genes *P1, P2* and *P3* (Dodds and Simmons, 1948; Simmonds, 1953). A segregation for these genes has been found in progeny of crosses between the cultivated variety Calcutta-4 (resistant to Black Sigatoka) used as a male parent and wild type females (Ortiz and Vuylsteke, 1992). A cross between a wild type *M. acuminata* and cultivated banana to transfer resistance to Fusarium wilt disease has shown that the resistance trait is not always transmitted to the progeny, and so probably involves multiple gene control (Rowe, 1984; Shepherd, 1990; Rowe and Rosales, 1993).

Hybrid bananas have great potential to improve the crop, and new breeding strategies exploiting information from genomic studies are likely to lead to more rapid progress. Both wild and edible banana are useful as a starting material for breeding, although natural and synthetic tetraploids and even triploids can also be used as parents to produce seeds (e.g. Gros Michel (AAA), in which immature seeds can be isolated at a very low frequency).

1.1.3 Production constraints

The consistent supply of high quality fruit is important to both the fresh and processed banana industries. Thus, the National Agriculture Policy of Malaysia has identified bananas as one of the 15 crops to be prioritised for development (Kamarudin, 1996).

The major issues in the banana industry are field management, pest and disease control, and a lack of understanding of banana genetics and diversity, so the focus of banana research in the Seventh Malaysian Plan (1996-2000) is to integrate a package of technology for pest and disease control, harvesting and handling during transportation. These goals are important to achieve the aims of the Third Agricultural Policy 1998-2010, which emphasises the need for high quality and high value banana production.

The extensive use of pesticides, fungicides and herbicides, which is costly, can cause environmental damage and has adverse health affects on farmers and smallholders, has become unacceptable to the public and is therefore now a political issue. Typical commercial crop regimes use 40-50 pesticide applications, accounting for 40% of the production cost. Furthermore, the heavy use of chemicals also affects yield and generates abnormality not only in the plant but also in the pathogens. Research progress in molecular biology and biotechnology, including *in vitro* techniques and cryopreservation, marker systems, genetic engineering, BAC technologies, and genetic and physical mapping are expected to accelerate the breeding of new and improved banana varieties.

1.1.4 Germplasm utilisation

The International Network for the Improvement of Banana and Plantain (INIBAP) was created in 1985 as part of IPGRI (International Plant Genetic Resources Institute), a CGIAR Institute, and now part of the 'Commodities for Livelihoods' programme. Its aim is to support the sustainable production of banana and plantain for domestic consumption and for local and export markets. Based in Montpellier, France, with regional offices all over the world, its main task is to conserve and managing diversity, and use diversity for the genetic improvement of bananas. The first regional network (The Latin America and the Caribbean Network) was established in 1987. Later, this network was separated into the Plantain and Banana Research and Development Network for Latin America and the Caribbean (*MUSA*LAC) under the auspices of the Foro Regional de Investigación y Desarrollo Tecnológico Agropecuario para América Latina y el Caribe (FORAGRO).

As the network expanded, other networks emerged in different regions and started to make linkages. The Asia and Pacific Regional INIBAP Network (ASPNET) was established in 1991, and was later known as the Banana Asia Pacific Network (BAPNET). As many researchers gathered and exchanged information, communication and collaborations, the Regional Information and Documentation Network (RISBAP) was set up in 1996 at the BAPNET regional base in The Philippines. Furthermore, the Banana Research Network for East and Southern Africa (BARNESA) was established in 1997 under the auspices of the Association for Strengthening Agricultural Research in East and Central Africa (ASARECA). In West and Central Africa, a Banana Research Network for West and Central Africa (*MUSA*CO) was established with the cooperation of the International Institute of Tropical Agriculture (IITA) and the Centre Africain de Récherches sur Bananiers et Plantains (CARBAP) in Cameroon.

Most banana-growing countries have significant germplasm collections (e.g. MARDI in Malaysia, where some 600 accessions are maintained). In conserving and managing the diversity of banana, INIBAP implemented conservation strategies including *in vitro* methods, field gene-banks and on-farm conservation. The INIBAP gene-bank, known as INIBAP Transit Centre (ITC), has been established at the Catholic University, Leuven (KULeuven) in Belgium. *Musa* accessions have been maintained using *in vitro* methods (tissue culture, and increasingly through cryopreservation), and so far, more than 1100 accessions labelled with an ITC code, and including wild species as well as local, commercial and improved varieties, have been documented. Where appropriate, ITC-registered germplasm clearly labelled with an ITC accession number and name was used in the present study. As an insurance, duplicate genebanks have been established at the Taiwan Banana Research Institute (TBRI) and the Centro Agronómico Tropical de Investigación y Enseñanza (CATIE), Costa Rica. The application of biotechnologies, such as molecular markers and molecular cytogenetics, have made physical and genetic mapping feasible. Access to all this information has been facilitated by the *Musa* Germplasm Information System (MGIS), which provides a standardized mechanism for the collection, storage and management of key information relating to the origin and characteristics of *Musa* germplasm (http://mgis.grinfo.net).

As the banana is the 'fruit for people', the *Guidelines for the Safe Movement of Musa Germplasm* were revised by FAO, IPGRI and INIBAP in 1996 to develop more sensitive and reliable tests to detect infection and permit both safer and faster transfer of germplasm throughout the world. Virus indexing centres were established in France (Centre de Coopération Internationale en Recherche Agronomique pour le Développement, CIRAD), Australia (Queensland Department of Primary Industries, QDPI) and South Africa (Plant Protection and Research Institute, PPRI). Under a Material Transfer Agreement, updated in 2003, this guarantees that the germplasm is clean and remains available within the public domain. The discovery of the integration of the pararetrovirus Banana Streak Virus (BSV) into the nuclear genome of the B genome has added a level of complexity to the distribution of 'virus-free' germplasm, and approaches to handling this development are under discussion (Geering *et.al.*, 2005; PARADIGM reports, http://paradigm.cirad.fr/projectinformation/index projectinformation.html).

In broadening the genetic base of materials available to banana and plantain for breeders, the INIBAP coordinated programme PRO*MUSA* (Global Program for *Musa* Improvement) aims to accelerate the overall impact of *Musa* improvement research at a global level. The materials and information are important to help understand devastating diseases, the demand for new varieties, and the requirement for stress resistance (abiotic and biotic) in global world banana production.

As biotechnology is beginning to provide reliable tools to understand the diversity of the *Musa* genome, INIBAP launched the Global *Musa* Genomics Consortium in 2001. This led to the establishment of the *Musa* Genomics Resources Centre (MGRC) at the Institute for Experimental Botany (IEB) in the Czech Republic. The IEB role is to provide BAC libraries, repetitive DNA clones and cytogenetic markers for the members of 29 public-funded institutions in 15 countries and other interested researchers. Recently, the 'Generation Challenge Program' was initiated to support the impact and implement the research outcomes of the Global *Musa* Genomics Consortium. The International *Musa* Testing Program was set up to evaluate and provide technical guidelines to farmers so that elite *Musa* varieties produced by breeding programs could be sustained.

1.1.5 Progress in conventional breeding

Since 1984, the Centro National De Pesquisa de Mandioca e Fruticultura (CNPMF) and the Fundación Hondureña de Investigación Agrícola (FHIA) have initiated an improvement programme for diploid banana. This is the most successful banana breeding programme up to now, and has produced a number of interesting new varieties, and served to disseminate breeding technology. Multiple crossing of wild type bananas from Papua New Guinea, Java, Malaysia and The Philippines gave rise to the diploid SH-2095, which became a parent of many FHIA varieties. FHIA hybrids are highly variable with respect to a range of agronomic characters and disease resistance. FHIA-18 (AAAB) is resistant to black leaf streak disease, *Fusarium* wilt and the nematode *Radopholus similis*, but has a mean bunch weight of only 28.5 kg (Rowe 1998).

Consequently, additional selections were made, leading to the identification of the hybrids FHIA-01 (AAAB), FHIA-03 (AABB) and FHIA-23 (AAAA), which are also resistant to *Fusarium* wilt and tolerant to the banana weevil (Nowakunda *et al.,* 2000), but have high yields. Some of these FHIA hybrids were evaluated and tested in Peru, but poor disease management and planting systems negated some of their good characteristics. Furthermore, the unique flavour and texture of the fruit of these new hybrids were not accepted by local people (Krauss *et al.,* 2001). The availability of genetic resources has played an important role in generating new varieties through conventional breeding, but the process is time consuming and laborious, since thousands of crosses and selections are required to identify an individual having a combination of with favourable characteristics (Oselebe *et al.,* 2006).

In all, 14 improved hybrids were produced from the banana and plantain program at IITA (Vuylsteke *et al.* 1993). All the hybrids tested produced taller plants, with progeny from *M. balbisiana* (TMB x 1378) exceeding 4m in height. Other characters found in progenies were late flowering with inconsistent yield, and various levels of resistance to black sigatoka diseases (Crouch and Ortiz, 1996). Till now, no new hybrid with a suitable combination of characteristics for widespread release and adoption has been identified from the IITA programme.

These results show the importance and impact of the wild banana in providing the genetic resources needed for the production of new varieties. Conventional breeding has led breeders to improve and assimilate both cultivated and wild type bananas for better quality and quantity production, but there remains a major degree of uncertainty in breeding outcomes. The selection of suitable candidates as parents remains problematical. As a centre of banana diversity, the Asian region is likely to contribute novel genetic traits in the future, and thus, a better knowledge of methods for conservation, selection, documentation and evaluation are required.

At the Malaysian Agriculture Research and Development Institute (MARDI), over 200 accessions representing both *M. acuminata* and *M. balbisiana* groups have been documented and described using the recommended IBPGR (predecessor organization of IPGRI) descriptors for banana (1996). However, only 5% are commercially utilised (Siti Hawa, 1998). As *Musa* has such a complicated genetic system, neither natural nor conventional breeding can fully access the genetic variability available. As expected, the inheritance of agronomic characteristics and disease resistance in certain *Musa* cultivars and species has indicated the occurrence of dominance among expressed genes. Biotechnological approaches, molecular techniques and bioinformatics will be useful to support and enhance the breeding programs.

1.2 Genome analysis in *Musa*

The genome consists of single copy genes, regulatory sequences and repetitive DNA. Gene sequences are transcribed to mRNA before being translated into proteins. Fragments of *in vitro* transcribed mRNA are referred to as expressed sequences tags (ESTs), and these can be used to identify functionality as part of the processes of gene discovery (Schmitt *et al.,* 1999; Ewing and Green, 2000; Susko and Roger, 2004; Lee *et al.,* 2005), specific gene expression (Bonaldo *et al.,* 1996; Audic, 1997) and splicing (Gupta *et al.*, 2004). Essential steps for genome analysis require gene assignment and marker systems prior to genetic and physical mapping. The resultant markers are important for the determination of chromosome location and intra-chromosomal linkage.

1.2.1 *Musa* as a genome model

The *Musa* genome size of 500-600 Mb places it between the large genome crop species such as maize (2400 Mb) and barley (4900 Mb) and the small genome species rice (430 Mb). The genome sequence of rice has been completed, there are major ongoing sequencing programmes for sorghum (750 Mb), and large amounts of sequence data exist for maize and barley. Currently, the primary model plant for genome and expression comparison is *Arabidopsis thaliana* (130 Mb) a dicotyledonous species, but rice has the advantage of being simultaneously both a crop and a model species. Based on its genome size, the banana genome has the potential to be a model plant representing the monocotyledonous crop species as an outgroup from rice, with opportunities for the study of diversity and the evolution of its reproductive system, vegetative propagation, polyploidization and parthenocarpy.

Musa species form a plant group consisting of both partially and highly fertile wild type species, expressing a range of both abiotic and biotic stress resistance. Thus, *Musa* is an excellent candidate to understand plant and pathogen co-evolution, which can then be understood at the genomic level. The existence of polyploid types in *Musa* gives an opportunity to study the evolution of polyploidy, and because different ploidies are present, whole genomes can be brought together in the sterile hybrids and exploited for trait improvement. Mass production and handling of the resultant new plant types can utilise biotechnology such as tissue culture, cryopreservation and/or somatic embryogenesis, which requires less maintenance and field area.

Parthenocarpic cultivars are not widely used directly for breeding but the character exists in many species and normally is present in all varieties with good fruit size, yield and quality (Varoquaux *et al.,* 2000). Thus, the stable parthenocarpic characteristic is essential in new varieties to maintain fruit quality and high yield (Pandolfini *et al.,* 2002). Parthenocarpy in combination with sterility is relatively rare in monocotyledonous plants, but as it exists naturally in *Musa,* it offers a good comparison between other dicotyledenous fruit crop species such as tomato, strawberry and melon, which have unstable parthenocarpy.

1.2.2 Molecular markers for genome analysis

DNA molecular markers rely on sequence polymorphism in the genomic DNA, which can be defined by either a restriction enzyme-based or by an amplification-based technique. Ideally markers should be economical (appliance and development cost), independent of each other, robust, frequently occurring and polymorphic. The experimental requirements are that they should be easy to assay, rapid, show high reproducibility, be exchangeable between laboratories and have the potential to be automated. Co-dominant inheritance is preferable to dominant, since it allows the heterozygous to be distinguished from the homozygous state (Piepho and Koch, 2000). According to Kosman and Leonard (2005), there is no suitable method to measure dissimilarity with dominant markers unless the plant material is highly inbred and therefore largely homozygous.

Restriction fragment length polymorphisms (RFLPs) were the first molecular marker system utilised for mapping the human genome (Botstein *et al.*, 1980). These markers are co-dominant and are independent of the growth stage of the plant or the environment, but their use is limited by the high quantity of DNA needed, and often by a lack of polymorphism. RFLP has been used to study the taxonomy and phylogeny of *Musa* sp, but is too labour-intensive for routine breeding applications (Gawel *et al.*, 1992; Jenny *et al.*, 1997; Crouch *et al.*, 1998). More recently, markers based on PCR have been developed, including Random amplified Polymorphic DNA (RAPD) (Williams *at al.*, 1990), ISSR (Zietkiewicz *et al.*, 1994), AFLP (Vos *et al.*, 1995), SSR, IRAP/REMAP, SNP and CAPs. RAPD have been used to classify *Musa* germplasm, but many primers are required, since there is a strong tendency for primers to be genome-specific (Howell *et al.*,1994; Pillay *et al.*, 2000). AFLP markers are useful in separating the different sections of the *Musa* genus, and have been informative in understanding the genetic diversity present in collections of *Musa* (Ude *et al.*, 2002, 2003; Wong *et. al.* 2002). SNP analysis produces a high level of polymorphism. Although these markers are codominant, their development is demanding, since it relies on the generation of sequence data (Gupta *et al.*, 2001). Nonetheless, the value of SNP in selecting cultivars of soybean for improved resistance to soybean cyst nematode (SCN) has justified the effort necessary to generate these markers (Young, 1999).

It is important but can be difficult to select specific approaches or techniques in utilising markers as each one has both advantages and disadvantages. However, the selection critically depends on the objective of the analysis, population structure, globally diversity of the species, available facilities, timeframe, potential and overall budget in the research proposed.

The most commonly used PCR-based markers are RAPD, AFLP and SSR, along with amplification of genes (EST) or other defined sequences with flanking primers. The former two require no prior sequence information, but are dominant markers, whereas SSR markers are codominant, but require sequence data for their development. Dominant markers such as RAPD and AFLP were similar in their ability to estimate intra-specific genetic diversity in plants, provided that a sufficient number of primers / primer combinations were employed (Virk *et al.*, 2000; Garcia *et al.*, 2001; Uptmoor *et al.,* 2003; Nybom, 2004).

Efficient markers can support conventional breeding by tagging specific alleles of key genes, allowing for rapid screening, germplasm characterisation, genotype identification, phylogenetic analysis and selection (Rafalski *et al.*, 1993; Crouch *et al.,* 1998; Gupta *et al.*, 1999). Molecular markers are particularly useful for gene identification and mapping. Once a gene for a particular trait has been found, and the coding sequence cloned, the latter can be used as a probe to locate homologs in other species. Comparative genome analysis has shown that gene order is often conserved between related species (Devos and Gale, 2000; Paterson *et al.,* 2000; Tanksley, 2004; Mueller *et. al.,* 2005). Currently, CIRAD is constructing a genetic map in *Musa* based on RFLPs, AFLPs and SSRs. The map aims to locate genes of interest in relation to linked markers, and once established, can help define genes of agronomic interest, such as those encoding stress tolerances. As the aggressiveness, sensitivity and availability of pathogens can vary, disease screening is difficult whether in field or in the greenhouse. At best, it is a time consuming process, and requires intensive and expensive labour for plant maintenance, data collection and analysis. Linked markers can be exploited to stack independent resistance and tolerance genes, and this is seen as a viable strategy to overcome some of the constraints imposed by pests and disease.

1.3 Cloning and DNA analysis

1.3.1 Clone libraries

Cloning is a procedure for multiplying discrete fragments of DNA in a bacterial (or sometimes eukaryotic) cell and thus usually allows incorporation of fragments of DNA from one organism in another, giving a large number of copies from a single sequence fragment. In order to represent a whole genome as a collection of small fragments, many clones are required, and the resulting set of bacterial or yeast cells is known as a clone library. A large number of clones are needed to increase the statistical likelihood that any given sequence is present in the library. Several types of library have been defined, including genomic (genomic DNA clones) and cDNA (complementary DNA, transcribed from mRNA).

Making a library requires the purification of template (genomic DNA or cDNA), the ligation of the fragments into a vector and the transformation of the recombinant vector into a host cell. A genomic library will include genic sequence, along with promoter regions, introns and exons, as well as non-genic sequence; whereas a cDNA library consists of non-intron coding sequence. Genes not expressed in the source tissue will not be represented in the cDNA library. The analysis of genomic clones enables the sequencing and analysis of the nature and location of promoter, intron and/or exon elements, and homology comparisons between genic and non-coding regions allows insights into gene evolution and diversification. The hypothetical or *in silico* translation of DNA sequence can allow inferences to be made regarding gene function and expression.

Small clones, typically in maintained in plasmids (larger ones are carried in bacteriophages or cosmids), are used widely for mapping, marker development or cloning mRNA to give EST libraries. Very large insert clones (upwards of 100 kb) are propagated in BACs (Bacterial Artificial Chromosomes) or YACs (Yeast Artificial Chromosomes), and these are used to study genomic organization.

The procedures depend on the ability of vectors including DNA inserts to propagate and maintain their life cycles in the selected host, either bacteria or virus. In this procedure, foreign DNA is inserted into the vector by ligating the DNA into a complementary site in the vector by the action of DNA ligase enzyme. Then, the vector carrying insert is transformed to either bacteria or virus by previously treated them with $CaCl_2$ to increase pores on their membrane cell wall. The insert size of foreign DNA gives an impact on transformation efficiency i.e. less molecule size produced high percentage of chances to be transformed.

1.3.2 Cloning vectors and transformation efficiency

The major requirement for a cloning vector is its ability to accept foreign DNA without affecting the life cycle of the host cell. Plasmids can tolerate inserts of <10 kb, bacteriophages 9-20 kb, cosmids 33-47 kb, BACs 75-125 kb and YACs 100-1000 kb.

1.3.3 Development of Bacterial Artificial Chromosomes (BACs)

The BAC cloning system was developed as part of the Human Genome Project in early 1990s, since the sequencing strategy required large, stable and non-chaemeric inserts. The BAC system uses an F-factor-based vector which allows a complex genomic DNA insert to be maintained in the *E. coli* host with a high level of structural stability. Both agarase and temperature sensitive alkaline phosphatase enzymes are important elements in the system. The former is used to recover high molecular weight DNA from Pulsed-Field Gel Electrophoretic separations, with minimal risk of DNA damage. A standard enzyme for DNA dephosphorylation is obtained from *E.coli* alkaline phosphatase which hard to remove from the reaction resulted a risk in DNA loss (Shizuya *et al.,* 1992; Shizuya *et al.,* 2001). Alternatively, elecroelution can be used to yield intact visible DNA and it is amiable to large fragment cloning with BAC vectors (Strong *et al.,* 1997). Figure 1 shows a schematic overview of the BAC construction method.

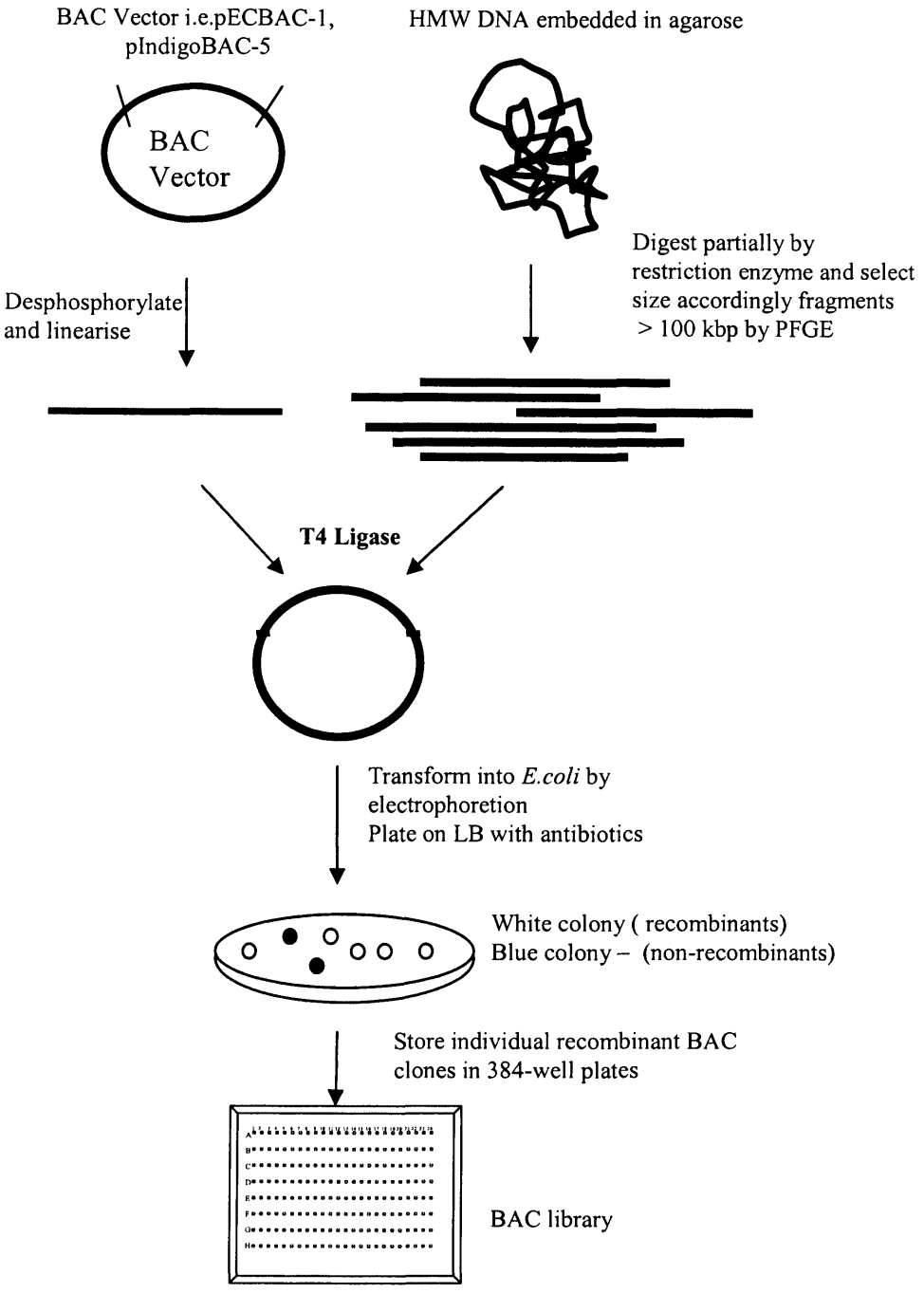


Figure 1.1. A general scheme for the construction of a BAC library.

The purification of genomic DNA is usually designed to minimize the presence of plastid and mitochondrial DNA, although most genomic libraries do inevitably include a small proportion of such organellar DNA; some libraries are contaminated with pathogen DNA (particularly fungi and viruses), which happened to be present in or on the source tissue.

As BAC libraries have now been constructed from many species with different genome sizes, comparisons between related chromosomal regions have become possible. The advantage of BACs lies in their low level of chimerism, high transformation efficiency, good stability in the bacterial host and easy handling procedures (Frengen *et al.*, 1999). Some disadvantages of single copy clones in BACs include a low level of DNA recovery and reduced purity of the DNA with respect to host DNA. To overcome this, Wild *et al.* (2002) have developed a novel BAC vector in which a high copy number is present, driven by the activation of the OriV origin of replication.

1.3.4 Importance of BAC clones for genome analysis.

BAC libraries have applications ranging from genome sequencing to gene discovery, and have helped to broaden the understanding of species diversification and evolution, through the elucidation of genome organisation, and the characterization of relationships between species. The large insert DNA present in BAC libraries enables both genome sequencing and physical mapping. The digestion of BAC with restriction enzymes enables overlapping sequences within pairs of BAC clones to be identified to generate contiguous arrays of BACs ('contigs'). Information from BAC sequencing can be utilised for the positioning of target genes, such as those encoding abiotic and biotic resistances. Repetitive DNA can also be exploited as a source of PCR-based markers.

The application of BAC libraries has accelerated genomics research related to human (Kim *et al.,* 1996; Asakawa *et al.,* 1997), fungi (Diaz Parez *et al.,* 1996), plant pathogens (Whisson *et al.,* 2001) and plant species (Woo *et al.,* 1994; Tao *et al.*, 1994; Lijavetzky *et al.*, 1999; Moullet *et al.*, 1999; Fu and Dooner 2000; Vilanova *et al.*, 2003; Hong *et al.*, 2004). BAC libraries can be used to mark the physical location of genes, for gene cloning and to provide a foundation for the sequencing of the entire genome (Zhang *et al.*, 1994).

BAC inserts are useful as probes for FISH (fluorescent *in situ* hybridization), for chromosome mapping and QTL analysis. They have been used for mapping in a number of important plant crops, including rice (Tao *et al.,* 1994; Wang *et al.,* 1995), sorghum (Woo *et al.,* 1994), wheat (Lijavetzky *et al.,* 1999; Moullet *et al.*, 1999), maize (Fu and Dooner 2000), chilli pepper (Yoo *et al.,* 2001), apricot (Vilanova *et al.,* 2003), ornamental crops (Kaufmann *et al.,* 2003) and melon (Leeuwen *et al.,* 2003).

Genomic distribution of ribosomal DNA in *Musa* has been widely explored using FISH, but karyotyping markers are still scarce (Osuji *et al.,* 1997; D'Hont *et al.,* 2000; Valarik *et al.*, 2002, Bartos *et al.,* 2005). The construction of a BAC library from both *M. acuminata* (Vilarinhos *et al.,* 2003) and *M. balbisiana* (Safar *et al.,* 2004) has produced an opportunity to extend molecular cytogenetics to these species.

BACs libraries are particularly valuable for characterizing specific regions of the genome, repetitive sequences such as microsatellites, retroelements, and gene clusters such as disease resistance and growth regulation genes. Currently five BAC libraries are available from *M. acuminata* and *M. balbisiana* (Table 1.1).

Table 1.1. *Musa* BAC library.

Name	Accession	No of clone	Coverage	References
C4BAM	*M. acuminata* 'Calcutta-4 '	17280 clones	3x genome coverage	James *et. al.*, (*unpublished)
MA4	*M. acuminata* 'Calcutta-4'	55296 clones	9x genome coverage	Vilarinhos *et al.*, 2003
TG BIBAC	*M. acuminata* 'Tuu Gia' (Diploid AA)	30700 clones	5x genome coverage	Ortiz-Vazquez *et al.*,2005
MBP	*M. balbisiana* 'Klutuk wulung' (Diploid BB)	36864 clones	9x genome coverage	Safar *et al.*, 2004
MAC	*M. acuminata* 'Grande Naine' (Triploid AAA)	55296 clones	4.5x genome coverage	Kaemmer *et al.*, 2002

*INCO-DC project IC18-CT97-0192

1.4 Repetitive DNA in plant genomes

Repetitive DNA is sequence present in multiple copies without any clear biological function. The major repetitive DNA sequences are divided into two classes - tandem repeat non-coding sequences and dispersed sequences. The former include telomeric repeats, satellites, minisatellites and microsatellites, whereas the latter are primarily composed of transposable elements (TEs) (Kumar and Bennetzen 1999; Heslop-Harrison, 2000). According to Jurka *et al.*, (2005), an understanding of genome structure and evolution is incomplete without a detailed study of these repetitive sequences.

1.4.1 Microsatellite (SSR)

Simple sequence repeats (SSR) include both microsatellites and minisatellites; the former have a motif length of 1-6 bp, and the latter >15 bp. The satellite repeat motif is typically 100 bp. SSRs are present in all eukaryotes, mostly in the form of di-, tri- or tetra repeats, and are common in the centromeric and sub-telomeric region of all chromosomes. SSRs can show high levels of polymorphism with respect to the number of repeat units and their function is unknown. They are particularly useful as a source of molecular markers (Heslop-Harrison, 2000).

The genomic distribution of SSRs is probably non-random and may have arisen as a result of chromatin organization, the regulation of gene activity, recombination, DNA replication, and the cell cycle (Schmidt and Heslop Harrison, 1996; Li *et al.*, 2002). They can be located in transcribed regions, including within protein-coding genes, although the repeat number is typically relatively small in these regions (Tautz, 1989; Kantety *et al.*, 2002; Thiel *et al.*, 2003). Since SSR sequences are short, they are particularly amenable to amplification by PCR.

1.4.2 Transposable elements

In eukaryotes, the two distinct classes of TE are retroelements (Class I) and DNA transposons (Class II). However, several elements remain unclassified, as their mechanism of transposition is still unclear, and these include families of interspersed elements with terminal inverted repeats known as inverted-repeat transposable elements (MITEs). The first MITEs were described in the grass genome (Bureau and Wessler, 1992).

Retroelements and DNA transposons are distinguished from one another by their mechanism of transposition. The latter can excise and reintegrate elsewhere in the genome, whereas the former replicates via an RNA intermediate, leaving the original copy, and forming a new duplicate copy which is integrated elsewhere in the genome. Both retroelements and transposons comprise autonomous and non-autonomous elements.

The autonomous elements include the gene sequences essential for propagation in the host genome, whereas non-autonomous elements require enzymatic activiy which has to be supplied *in trans* by autonomous elements (Lewin, 2000, Moran and Morish, 2005). Helitrons are a particular class of autonomous eukaryotic elements which have an open reading frame (ORF) encoding protein(s) with nuclease/ligase domains and using rolling circle replication (Kapinotonov and Jurka, 2001; Feschotte and Wessler, 2001).

Integration, excision and replication of TEs can increase or decrease genome size, but it is unclear whether the number of TE families is a major factor in genome expansion (Kidwell, 2002). In the *Arabidopsis* genome sequence, 50% of TEs were DNA transposons and 22% retroelements, with 28% remaining unclassified (Le *et al.*, 2000). Retroelements are ubiquitous in plant and animal genomes. Commonly, they have two ORFs, the first of which encodes a structural protein core known as gag (Hansen *et al.,* 1992), and the second (the *pol* gene) has protease, reverse transcriptase (RT), RNase H and integrase domains, which are required for replication and integration into the host chromosome (Bingham and Zachar, 1989; Varmus and Brown, 1989). The RT is the only coding region that includes completely conserved regions between all retroelements and therefore is useful for phylogenetic analysis.

Occurrence and abundance of retroelements in plant species represents an obstacle for genome sequencing, as they produce a high frequency of intergenic and/or repeat regions (Brandes *et al.*, 1997). RT is an ancient and widespread enzyme among viruses, prokaryotes and eukaryotes. (Baltimore, 1970; Temin and Mizutani, 1970; Xiong and Erickbush, 1990). Interestingly the retroelements encoded their own RT for propagation (Inouye and Inouye, 1991) which gives conserved region in their family (Xiong and Eickbush, 1990) and it is predicted that they might be having a common and ancient origin (Eickbush, 1994).

A particular group of retroelements includes retroviruses, pararetroviruses and the abundant long terminal repeat retrotransposons (LTR-retrotransposons), which are non-infective. The LTR retrotransposons are divided into the *Pseudoviridae* (Ty1-*copia*) and *Metaviridae* (Ty3-*Gypsy*). Both Ty1-copia and Ty3-gypsy correspond to the elements first described in *Saccharomyces cerevisiae* and *Drosophila melanogaster* (Xiong and Eickbush, 1990), and are named after the yeast Ty1 and Ty3 (Clare and Farabaugh, 1985; Hansen *et al.,* 1992), and the *Drosophila copia* and *gypsy* (Mount and Rubin, 1985; Marlov *et al.,* 1986) elements. The non-viral retroelements include the LINEs (long interspersed nuclear elements) and SINEs (short interspersed nuclear elements) and group-II mitochondrial introns (Hansen and Heslop-Harrison, 2004). A classification of the retroelements and related viruses is shown in Figure 1.2.

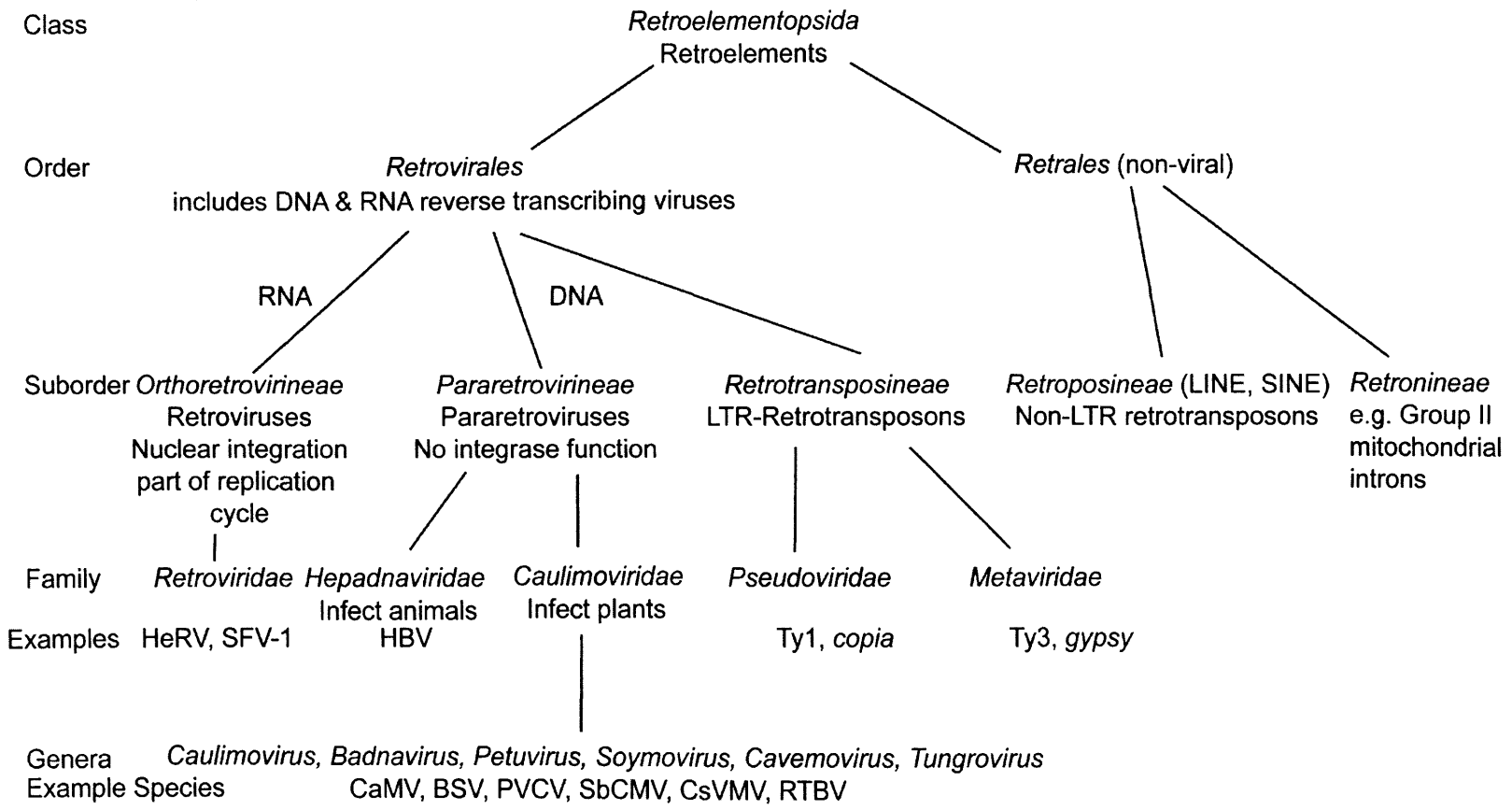


Figure 1.2. Diagram from Hansen and Heslop-Harrison, (2004) shows relationship of retroelements and related viruses.

The difference between the *Pseudoviridae* and the *Metaviridae* lies in the orientation of their RT and integrase domains within the *pol* gene. *LINEs* are rather simple structures with similar properties, but have an endonuclease function rather than the integrase present in both Ty1-*Copia* and Ty3-*Gypsy* elements (Figure 1.3). The most abundant TEs in plants are the LTR-retrotransposons, including both the *Pseudoviridae* (copia) and the *Metaviridae* (gypsy) (Feng *et al.*, 2002, Sasaki *et al.*, 2002), and the non LTR-retrotransposon LINEs.

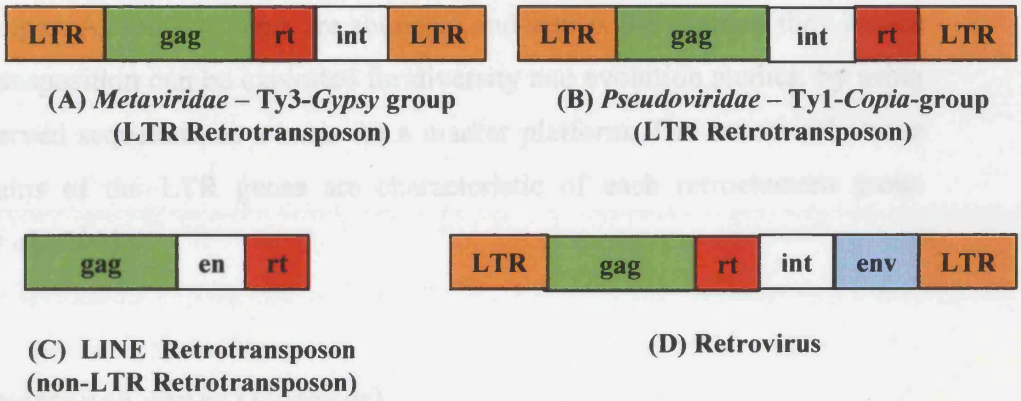


Figure 1.3 Orientation features of different transposable elements (A) *Gypsy-like*, (B) *Copia-like*, (C) *LINE* and (D) Retrovirus. A difference between *Metaviridae* and *Pseudoviridae* is the position of rt and int. The *Envelope* gene is a feature in retroviruses, and some *Metaviridae*, for binding of virus particles to their cellular receptors enabling virus to enter and start a replication cycle. gag (group specific antigen), en (endonuclease), rt (reverse transcriptase), int (integrase), LTR (long terminal repeat), env (envelope gene).

Gao *et al.* (2004) divided the LTR retrotransposons into three further groups- full-length elements, solo LTRs and fragmented elements. The first includes both autonomous and non-autonomous elements and contains all the characteristic features of LTR retrotransposons. Solo LTRs are related to recombination products between the flanking LTRs, and fragmented elements are partial or truncated sequences of LTR retrotransposons. The latter group is the most common retrotransposon in many plant species.

The insertion, splicing and replication of sequence fragments due to retroelement activity have altered gene regulation, expression and sequence amplification. The extent of insertional activity can be correlated with changes in environment (related to stress) of the host (Grandbastien 1998; Kalendar *et al.*, 2000). Through evolution, questions arise on timeframe of the activities and stability of the changes based on the stress conditions exposed to the plants remains and/or changes. As retroelements are abundant and active, the changes they induce by their transposition can be exploited for diversity and evolution studies, by using their conserved sequences as a basis for a marker platform. The conserved amino acid domains of the LTR genes are characteristic of each retroelement group (Friesen *et al.*, 2001).

1.5 Resistance genes (R-genes)

1.5.1 History

A single gene (R) can confer resistance to biotic stress. The first proven Mendelian inheritance of a disease resistance was demonstrated in wheat's reaction to infection by stripe rust, *Puccinia striiformis* (Biffen, 1905). The genes in plant species related to pathogen, host and environment responses are the major factor that contributes towards resistance and hence survival. Resistance gene (R-genes) is not only beneficial in understanding the mechanism, evolution and diversity of disease resistance, but is also important in developing practical applications for plant improvement. The systematic naming of resistance genes, the determination of their inheritance and segregation, and the selection of crop plants with desirable alleles have been carried out by plant breeders in sexually propagated crops from at least the 1950s (Allard, 1960).

Cloned R-genes, which are keys to the response to viral, bacterial or fungal pathogens, share DNA sequences and protein motifs, and thus confer resistance even though different mechanisms of activation may be involved (Kanazin *et al.,* 1996). Successful breeding for resistance should benefit from an understanding of the molecular polymorphism in these genes. Conventional phenotypic testing for resistance and/or tolerance in germplasm can be a costly and demanding process, as it often requires large scale screening of semi-mature plants in the field (Pink, 2002). Particular progress has been made in understanding the mechanisms behind the function of R-genes, avirulence (avr) genes, and their signalling components (e.g. LRR) and recognition domains (e.g.. NBS) (McDowel *et al.,* 2003, Jones and Dangl, 2006). In the model plant species *A. thaliana*, at least 150 R-genes are distributed unevenly throughout the genome, and approximately 200 genes contain an NBS-LRR related domain, representing 1% of the full gene complement (*Arabidopsis* Genome Initiative, 2000).

Pto (resistance to *Pseudomonas syringae)* was the first R-gene gene to be isolated from tomato *(Lycopersicon esculantum)* (Martin *et al.,* 1993). The interaction between *Pto* and the bacterial pathogen *avrPto* provides confirmation of the gene-for-gene hypothesis (Flor, 1956). The hypothesis referred to every gene for resistance in host is matched with a gene for avirulence in pathogen. Already stated In *A. thaliana, Rpm1* and *Rps2* confer resistance to *P. syringae* strains which express non-homologous *avr* genes (Mindrinos *et al.*, 1994; Grant *et al.*, 1995). R-genes can differ in their organisation: *Pto* does not include any NBS-LRR domain, but these domains are present in *Rps2* and *Rpm1*. *Cf-9* and *Cf-4* (resistance to *Cladosporium fulvum*) were isolated from *L. pimpinellifolium* (Jones *et al.,* 1994) and *L. hirsutum* (Thomas *et al.,* 1997), and are homologs with extracellular LRR domains, without any cytoplasmic signalling domain.

Many other R-genes have been obtained related to NBS-LRR domains from *A. thaliana, Nicotiana, Oryza,* flax, pepper and lettuce. The R-genes are grouped according to their nature, the recognisable motifs, and (N-terminal to the NBS) presence of TIR domains (homology to *Drosophila* Toll and mammalian Interleukin-1 receptor) and non-TIR domains (sometimes a form of leucine zipper-LZ).

However, the LRR may also bind to trans-membrane protein (TM) or kinase group proteins. Currently, seven predicted orientations have been postulated (Hammond-Kosack and Jones, 1997; Meyers *et al.,* 1999; Pan *et al.*, 2000a; Pan *et al.,* 2000b; Dangl and Jones 2001; Hulbert *et al.,* 2001; Bai *et al.,* 2002; Martin *et al.*, 2003), as illustrated in Figure 1.4.

1.5.2 Abiotic and biotic stress in plant

The term 'abiotic' relates to physical or inorganic factors, whereas 'biotic' refers to living organisms (including viruses). Biotic stress resistance is equivalent to disease resistance, while abiotic tolerance typically integrates many components, which act simultaneously and interact with one another. The imposition of abiotic stress can affect the quality and integrity of the biotic stress defensive mechanism as it involves common cellular pathways. The common abiotic stresses affecting plants are cold (chilling and frost), heat (high temperature), salinity (salt), drought (water deficit and/or water excess), excessive radiation (high intensity of ultra-violet or visible light) and chemicals (including pollutant, nutrient deprivation, oxidative, heavy metals, pesticides, aerosols).

Both abiotic and biotic factors can generate stress in plants. Understanding signal transduction pathways are necessary for plant survival to elaborate mechanisms in perceiving external signals and to manifest adaptive responses with proper physiological and morphological (Bohnert *et al.,* 1995). According to Yang *et al.,* (1997), both extracellular stimuli and the activation of defence responses require a complex interplay of signalling cascades, controlled by protein phosphorylation.

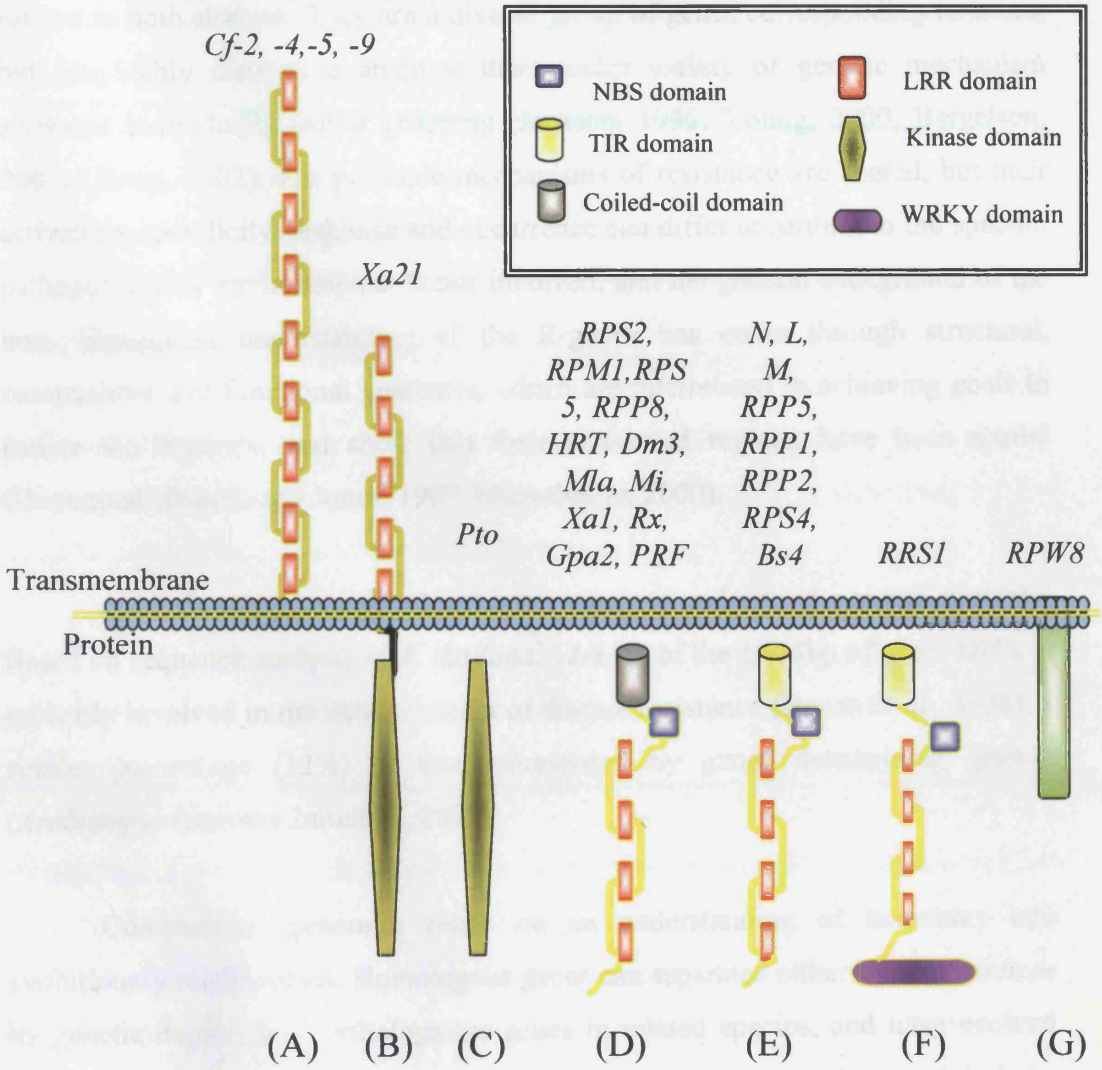


Figure 1.4 Schematic of conserved predicted structures for genetically defined plant resistance proteins that can be classed into seven major sub-families (A-G). Transmembrane proteins are existed in (A) and (B) connected to extracellular LRR. (expanded to include new classes after Hammond-Kosack, 1997; Pan *et al.*, 2000a).

1.5.3 Conservation of resistance genes (R-genes)

Resistance is critical for an organism to survive both abiotic and biotic stress. In plant species, the R-genes encode protein-binding proteins, which are related to both stresses. They are a diverse group of genes corresponding to stress, but are highly distinctive amongst them under variety of genetic mechanism activated individually and/or grouping (Jackson, 1999; Young, 2000; Bergelson, 2001; Chang, 2002). The principle mechanisms of resistance are shared, but their activation, specificity, response and occurrence can differ according to the specific pathogen and/or environmental factor involved, and the genetic background of the host. Significant understanding of the R-genes has come through structural, comparative and functional genomics, which are interrelated in achieving goals to isolate the R-genes, and show that their conserved regions have been shared (Hammond-Kosack and Jones, 1997; Michelmore, 2000).

Analysis at the sequence level is necessary to understand genome structure. Based on sequence analysis in *A. thaliana*, 12-14% of the 1.9Mbp of genic DNA is probably involved in the determination of disease resistance (Bevan *et al.*, 1998), a similar percentage (12%) to that represented by genes determining growth (*Arabidopsis* Genome Initiative, 2000).

Comparative genomics relies on an understanding of taxonomy and evolutionary relationships. Homologous genes can separated either by speciation or by genetic duplication. Orthologs are genes in related species, and have evolved from a common ancestral sequence. Paralogs are gene copies produced by duplication, and thus represent sequences derived from a common ancestral gene which duplicated within an organism and then diverged.

By identifying conserved domains and motifs, related either indirectly or directly to the R-genes, degenerate primers can be designed, and the intervening sequences can then be amplified and cloned, producing large numbers of candidate genes from diverse species. These can lead to the identification of many novel resistance genes (Leister *et al.*, 1996; Meyers *et al.*, 1999; Pan *et al.*, 2000a). Candidate R-gene sequence is simple to obtain, but validation of function requires a study of gene expression.

Functional genomics aims to describe the expression of genes *in vivo*. However, a difficulty arises when the sequences exist in many protein families such as NBS, which involve at the beginning of activation signal resistance pathways (Baker *et al.*, 1997; Meyers *et al.,* 1999) but cannot be expressed. A short motif of LRR exists in many repeat numbers, which can stand alone, but performing resistance genes such as Cf-4, Cf-9 integrate with NBS performing NBS-LRR encoded genes for disease or pest resistance.

As the NBS-LRR alone cannot be expressed, the sequences might be important for marker systems, identifying and determine of NBS-LRR clusters amongst resistance genes, which are corresponding to their specificity and evolution. Implementation and understanding the functional genomics in relation to the R-genes motifs may lead to a complete R-genes isolation expressed by the identified plant at as early tissue culture stage.

1.5.4 Nucleotide binding site (NBS) and Leucine-rich repeat (LRR) domains

The majority of cloned plant disease resistance genes (R-genes) encode proteins with an LRR (leucine rich repeats) region and a nucleotide binding site (NBS) domain, forming the NBS-LRR complex. In plants, R-gene homologs are abundant, and it is believed that the NBS participates in signal transduction, while the LRR domain is primarily responsible for elicitor recognition. Thus, the NBS-LRR complex works as a confirmation medium towards isolation the R-genes and may act as the primer of a defence response cascade, leading to cell death.

Resistance specificities and families of R-gene homologs are frequently found in clusters in plants. In *A. thaliana,* soybean, and potato, they have been shown to be genetically linked to resistance specificities (Kanazin *et al.,* 1996; Leister *et al.,* 1996; Botella *et al.*, 1997). The *N* gene of tobacco is related to the NBS-LRR gene family (Whitham *et al.*, 1994), and some *Xa21* homologs are physically linked to one another as tandem repeats in rice (Song *et al.,* 1997). Multiple homologs of the *L6* gene reside at the *M* locus in flax, one of which is the *M* gene (Anderson *et al.,* 1997).

The *Pto* and *Fen* genes of tomato discriminate between the products of the bacterial *avrPto* gene and the insecticide fenthion, and belong to a gene family of serine/threonine protein kinases (Loh and Martin, 1995), members of which are physically linked to one another (Salmeron *et al.,* 1996). Two tandemly repeated copies of *Cf-2* are physically linked to an additional gene *Hcr2s* (Dixon *et al.,* 1996).

A basic mechanism of cell to cell communication uses the secretion of ligands, which bind to cell receptors, and process catalytic activities in protein kinases (Jones and Dangl 2006). However, protein kinase families are large and beyond the scope of the present study. The first plant receptor like kinase (RLK) was discovered in *A. thaliana* (Walker and Zhang, 1990), and was used to demonstrate that plant cells are able to perceive external signals at the plasma membrane.

Extensive study of *A. thaliana* RLK revealed that more than 400 genes encode putative receptor like kinases (PRK) (Tichtinsky *et al.,* 2003) and these are classified on the basis of their extracellular domains, with leucine-rich repeats (LRR) amongst the largest group (Shiu and Bleeckeer, 2002). The PRKs have been defined as proteins which contain an extracellular domain, a single-pass transmembrane domain and a cytoplasmic serine/threonine (ser/thr) protein kinase domain (Shiu and Bleeckeer, 2001a, 2001b).

In animals, the LRR motifs are the same, but since they do not possess a cytoplasmic kinase, the signal is transmitted through the activation of co-receptors (Dievard and Clark, 2004). As the kinase families are enormous, the components of ser/thr kinase activity in plants and animals can be classified in terms of LRR-RLK and LRR-receptors respectively. Thus, the genes encode ligands, receptors and signalling components and their sequence can be expected to provide information pertaining to the diversity and specificity of their mechanisms.

LRR motifs comprising approximately 24 amino acids have been found in the primary structure of a large number of proteins, involved in biologically important processes such as hormone reception enzyme inhibition and cell adhesion (Buchanan and Gay, 1996, Kajava, 1998; Kobe and Kajava, 2001). Kobe and Deisenhofer (1993) described the first LRR protein from a Ribonuclease Inhibitor (RI) consisting of structural units of β-strands and α-helices. These structures are arranged in parallel in a curved `horseshoe` shape with the β-strand and α-helix lining the inner and outer circumference respectively. Thus, LRR proteins are present in many structural protein families (Kobe and Kajava, 2000) and based on their length and consensus sequence, at least six families of LRR protein have been identified (Kajava, 1998).

1.6 Aims

The preceding review has attempted to provide a background to the evolution, phylogeny, breeding constraints and importance of gaining an understanding of the diversity present in *Musa* spp. The isolation and characterization of key DNA sequences are necessary to understand the large-scale evolution, organisation and diversity in the *Musa* genome. To isolate, characterize and understand the homology of these sequences, a combined approach, involving molecular biology, some cytogenetics and bioinformatics, has been employed. The objectives were:

i) To determine aspects of diversity in the repetitive DNA present in the *Musa* genomes and BAC clones, using PCR-based techniques based on degenerate and specific oligonucleotide primers targeting retroelements and Simple Sequence Repeats (SSRs).

ii) To design degenerate oligonucleotide primers associated with R-gene motifs, specifically the Nucleotide Binding Site (NBS) and Leucine Rich Repeats (LRR) domains.

iii) To clone, sequence and analyse, both at the nucleic acid and the protein levels, the amplified NBS and LRR domains.

iv) To understand the nature and diversity of the NBS and LRR domains and their evolutionary and functional relationships using the sequence resources in the MIPS-Sputnik EST database, and genomic clones from diploid A, B, S, T genome and hybrid *Musa* accessions.

v) To design specific oligonucleotide primers targeting selected EST sequences for abiotic stress and to test these against *Musa* genomic DNA.

vi) To screen BAC libraries of *M. acuminata* and *M. balbisiana* by PCR-based techniques.

The development and implementation of DNA-based molecular markers for the assessment of diversity and evolution in *Musa* spp. provides molecular descriptors of the species, and a means to measure biodiversity. The distribution of repetitive DNA elements and their mobility, and the presence and nature of NBS and LRR signalling motifs, might provide markers which can enable selection and characterization of *Musa* BAC libraries to understand the gene locations for future studies towards gene discovery in plant breeding program.

CHAPTER 2

2.0 Materials and Methods

2.1 Materials

2.1.1 Plant materials

Musa germplasm from the INIBAP Transit Centre (ITC), Leuven, Belgium, was obtained in the form of small tissue culture plants, and was grown in the greenhouse at Leicester University. The collection was composed of 19 samples, representing various *Musa* species and accessions, all of which were 'FAO designated' (i.e. public property with no restrictions on use) and considered to be virus free. IPTJ, a wild type of *M. acuminata* 'Malaccensis' from the University of Malaya, was grown at the Botanical Garden of Leicester University. A total of 48 DNA samples of *Musa* varieties and two cultivated triploid samples of Malaysian commercial banana were obtained from, respectively, the Centre de coopération internationale en recherche agronomique pour le développement (CIRAD) Montpellier, France and the Malaysian Institute for Nuclear Technology Research (MINT) Bangi, Malaysia. Table 2.1 shows all the banana samples used in this study.

Genomic DNAs of the samples grown in the greenhouse were isolated from young leaves following a modified CTAB method (Gawel *et al.*, 1991). A 1/10 volume solution of 10 mg/ml RNase was added to eliminate RNA. The quality of both the original and the diluted DNA was assessed by agarose gel electrophoresis, measured against HyperLadder I, and the concentration was estimated with a Diode array spectrophotometer model Hewlett Packard at 260 nm. The DNA concentration in μg/ml was calculated as OD_{260} x 50 x Dilution factor.

2.1.2 Bacterial Artificial Chromosomes (BACs)

All individual BAC clone samples were obtained from the *Musa* Genome Resources Centre (MGRC), based at Institute of Experimental Botany (IEB) Czech Republic in stab cultures (2.5 ml screw-top tubes filled with Luria Bertani agar). The BAC clones were re-grown in liquid LB, and plated to obtain single colonies before proceeding to the isolation of DNA (refer 2.2.3).

Table 2.1. Banana samples for the project study. (See introduction for notes on taxonomy.)

No	Accession	ITC Code	Genome
1	Awak	ITC0213	ABB
2	Berangan	ITC1287	AAA
3	Butuhan	ITC1074	BT
4	*Musa acuminata* 'Calcutta-4'	ITC0249	AA
5	Giant Cavendish	ITC0346	AAA
6	Jari Buaya	ITC0312	AA
7	*Musa balbisiana* 'Klutuk Wulung'	ITC1063	BB
8	Mas	ITC0653	AA
9	Nangka	ITC0004	AAA
10	Obino L'Ewai	ITC0109	AAB
11	*Musa ornate*	ITC0370	
12	Radjah	ITC0243	AAB
13	*Musa textilis*	ITC0539	TT
14	*Musa velutina*	ITC0638	
15**	Mutiara	MAL	AAB
16**	Rastali	MAL	AAB
17*	*Musa balbisiana* 'Batu'	ITC1156	BB
18*	Blugoe	ITC0767	ABB
19*	*Musa acuminata errans* 'Agutay'	ITC1028	AA
20*	*Musa acuminata siamea* Khae	ITC0660	AA
21*	*Musa acuminata banksii* Paliama	ITC0766	AA
22*	*Musa balbisiana* Tani	ITC1120	BB
23*	*Musa balbisiana* Honduras	ITC0247	BB
24*	Saba	ITC1138	ABB
25*	Yawa2	ITC1238	ABBT
26***	*Musa acuminata malaccensis* (IPTJ)	MAL	AA
27	*Musa schizocarpa*	ITC0852	SS
28*	Tiparot	ITC0652	ABB
29*	Pelipita	ITC0472	(ABB)♦
30*	Ambon	ITC1064	AAA
31*	Mysore	ITC1441	AAB
33*	Lujugira Mbwazirume	ITC0084	AAA
32*	Gaba-gaba	ITC0307	AA
33*	*Musa acuminata malaccensis*	ITC0250	AA
34*	Pahang	ITC0609	AA
35*	Madu	ITC0258	AA
36*	Borneo	ITC0253	AA

INIBAP; *CIRAD, **MINT, ** Botanical Garden Leicester University
♦2n=8A+25B; D'Hont et al. 2000

Table 2.1.(Continued). Banana samples for the project study.

No	Accession	ITC Code	Genome
37*	Lilin	ITC0001	AA
38	*Musa schizocarpa*	ITC0599	SS
39	*Musa schizocarpa*	ITC1002	SS
40	*Musa textilis*	ITC1072	TT
41	*Musa balbisiana*	ITC0545	BB
42*	Nadan	ITC0852	AAB
43*	Pome/Prata Foconah	ITC0649	AAB
44*	Pome/Prata Prata Ana	ITC0962	AAB
45*	Orishele	ITC1325	AAB
46*	Dole	ITC0767	ABB
47*	Cavendish Grande Naine		AAA
48*	Orotaya Pisang Kayu	ITC0420	AAA
49*	Figue Pomme	ITC0769	AAB
50*	Mothan	ITC0046	ABB
51*	*Musa acuminata burmannica*	ITC 0283	AA
52*	Safet Vaichi	ITC 0245	ABcv
53*	Cavendish Petite Naine	ITC 0654	AAA
54*	Cavendish Poyo		AAA
55*	Popoulou	ITC 0335	AAB
56*	Peyan		ABB
57*	Gros Michel		AAA
58*	Wompa	ITC 1152	AS
59*	Kunnan	ITC 1034	ABcv
60*	Raja Bulu	ITC 0843	AAB
61*	Rio	ITC 0277	AAA
62*	Ney mannan	ITC 0020	ABB
63*	*Musa acuminata zebrine*	ITC 1177	AA
64*	Tomolo	ITC 1187	AAcv
65*	*Musa balbisiana* Lal vinci		BB
66*	Awak	ITC 06559	ABB
67*	Lujugira Itokatoke	ITC0082	AAA
68*	Ibota Yangambi	ITC 1123	AAA
69*	Red Yade	ITC 1140	AAB
70*	Nendra padathi	ITC 0243	AAB

INIBAP; *CIRAD, **MINT, ** Botanical Garden, Leicester University.
♦2n=8A+25B; D'Hont *et al.*, 2000.

2.1.3 BAC filters

BAC filters were obtained from MGRC, IEB, Czech Republic. Details of the BAC libraries are shown in Table 2.2. Each membrane was divided into six fields, each consisting of 384 squares. The squares represent the row and column identification of the BAC. Within each square are 16 positions where eight clones are spotted in duplicate, thereby allowing the pattern of the spotted clones to generate the plate address of the BAC (Figure 2.1). The library name and filter number was used to orient the membrane. Decoded plate numbers (e.g. F1, F2 and F3) was employed where multiple filters from the same library were used. Each 22x22 cm membrane contains up to 18,432 clones (384 x 48).

Table 2.2. The origin of BAC libraries and the BAC filters used for the project.

Library Name:	**MA4 BAC library**	**PKW BAC Library**	**C4BAM BAC Library**
Library type:	BAC	BAC	BAC
Accession:	*M. acuminata* 'Calcutta-4'	*M. balbisiana* 'Klutuk Wulung'	*M. acuminata* 'Calcutta-4'
Plant material:	Greenhouse-grown leaves	Greenhouse-grown leaves	Field-grown cigar leaves
Vector:	pIndigoBAC-5	pIndigoBAC-5	pECBAC-1
Host strain:	*E.coli* DH 10B	*E.coli* DH 10B	*E. coli* DH 10B
Restriction site:	*Hind*III	*Hind*III	*Bam*HI
Insert size (kbp):	100	135	110
Ligation:	2	3	1
Coverage:	9	9	3
Number of clones:	55296	36864	17280
Number of plates:	144	96	45
Number/type of filters:	3 (Gridding pattern 4x4) 4 (Gridding pattern 3x3)	2 (Gridding pattern 4x4) 3 (Gridding pattern 3x3)	1 (Gridding pattern 4x4) 2 (Gridding pattern 3x3)
Constructor(s):	Piffanelli P. Vilarinhos A.	Piffanelli P. Safar J.	James A.Tao Q.
date of birth (DD-MM-YYYY):	01-07-2001	01-03-2002	01-03-1999
Place of birth:	CIRAD	CIRAD	A&M
Specifications:	Live bacterial clones in 384-well plates	Live bacterial clones in 384-well plates	Live bacterial clones in 384-well plates
Curator(s):	Piffanelli P. Sekerova M.	Piffanelli P.Sekerova M.	Sekerova M. Zhang H.
Storage place(s):	CIRAD IEB	CIRAD IEB	A&M IEB
Construction method:	Nuclei were purified by centrifugation and embedded in agarose plugs. Megabase DNA was partially digested with *Hind*III, subjected to two rounds of size selection by PFGE, electro-eluted and ligated to the linearized and dephosphorylated vector.	Nuclei were purified by centrifugation and/or flow cytometry and embedded in agarose plugs. Megabase DNA was partially digested with *Hind*III, subjected to two rounds of size selection by PFGE, electro-eluted and ligated to the linearized and dephosphorylated vector	Nuclei were purified by centrifugation and embedded in agarose plugs. Megabase DNA was partially digested with *Bam*HI

From www.musagenomics.org redirected to
http://bioinfo.inibap.org/statusdb/display.php?page=librarydetails&id=MBP

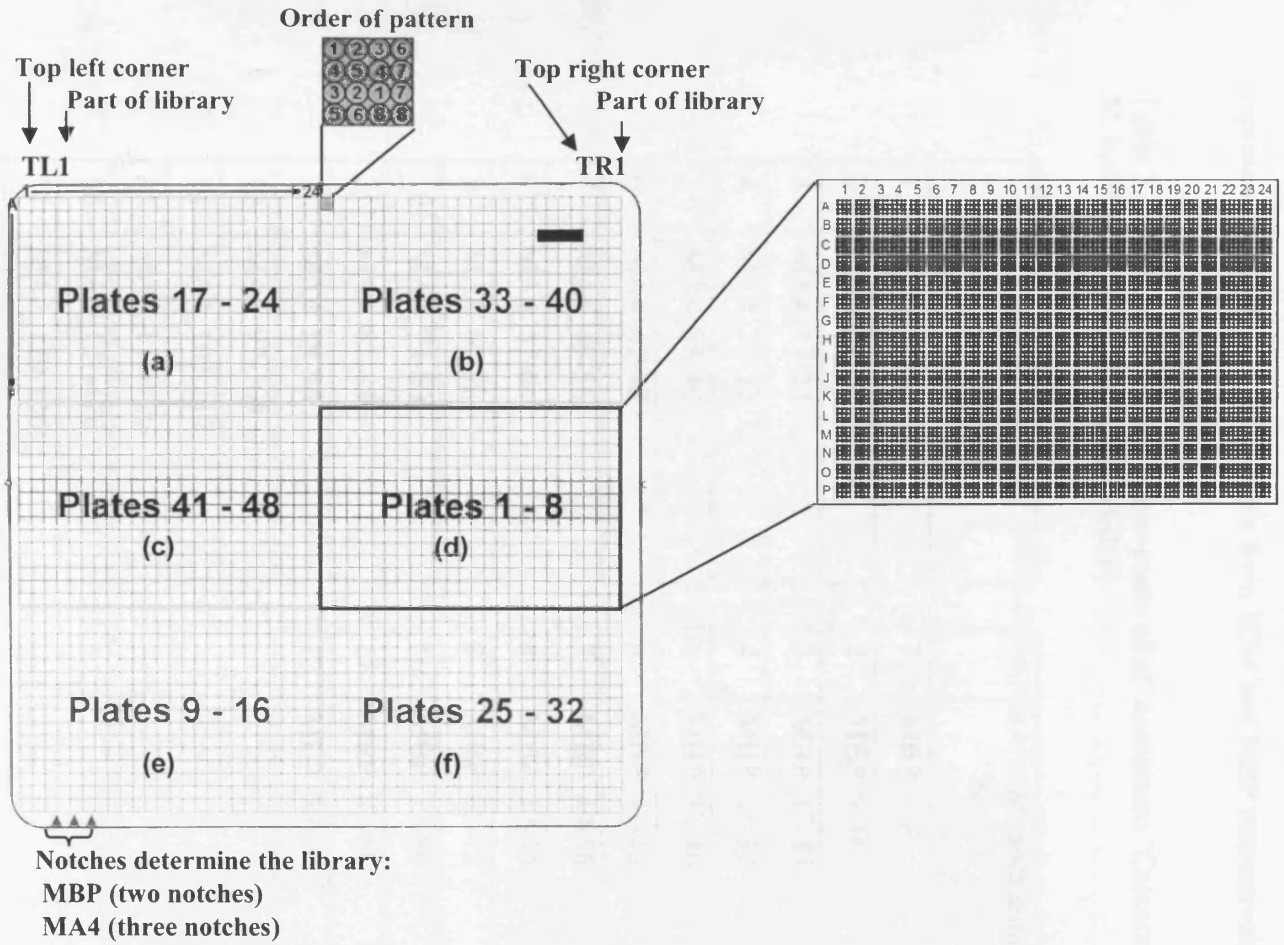


Figure 2.1. Arrangement of the high-density BAC filters from the *Musa* libraries used for colony screening by hybridization. Each 22 x 22 cm membrane has two spots from each of 384 x 48 (= 18432) BAC colonies from the MA4 and MBP libraries. The six fields (a-f) each have BAC clones from eight 384-well plates, with each colony double-spotted (4x4 format) in the 'order of pattern' from the wells of the plates corresponding to each position. (Source: GMCC, Dolezel)

2.1.4 BAC pools

The DNA pools from the BAC libraries of MA4 (*M. acuminata* 'Calcutta-4' libraries) and MBP (*M. balbisiana* 'Klutuk Wulung') were obtained from MGRC, IEB, Czech Republic. The MA4 and MBP libraries consist of 144 and 96 plates respectively (Table 2.3). The plates were arranged in sets of eight, representing 18 and 12 super-pools from MA4 and MBP respectively (Figure 2.2).

Table 2.3. *Musa* BAC library super-pools of *M. acuminata* 'Calcutta-4' (MA4) and *M. balbisiana* 'Klutuk Wulung' (MBP)

	M. acuminata 'Calcutta-4' (MA4)
1	MA4_1-8
2	MA4_9-16
3	MA4_17-24
4	MA4_25-32
5	MA4_33-40
6	MA4_41-48
7	MA4_49-56
8	MA4_57-64
9	MA4_65-72
10	MA4_73-80
11	MA4_81-88
12	MA4_89-96
13	MA4_97-104
14	MA4_105-112
15	MA4_113-120
16	MA4_121-128
17	MA4_129-136
18	MA4_137-144

	M. balbisiana 'Klutuk Wulung' (MBP)
1	MBP_1-8
2	MBP_9-16
3	MBP_17-24
4	MBP_25-32
5	MBP_33-40
6	MBP_41-48
7	MBP_49-56
8	MBP_57-64
9	MBP_65-72
10	MBP_73-80
11	MBP_81-88
12	MBP_89-96

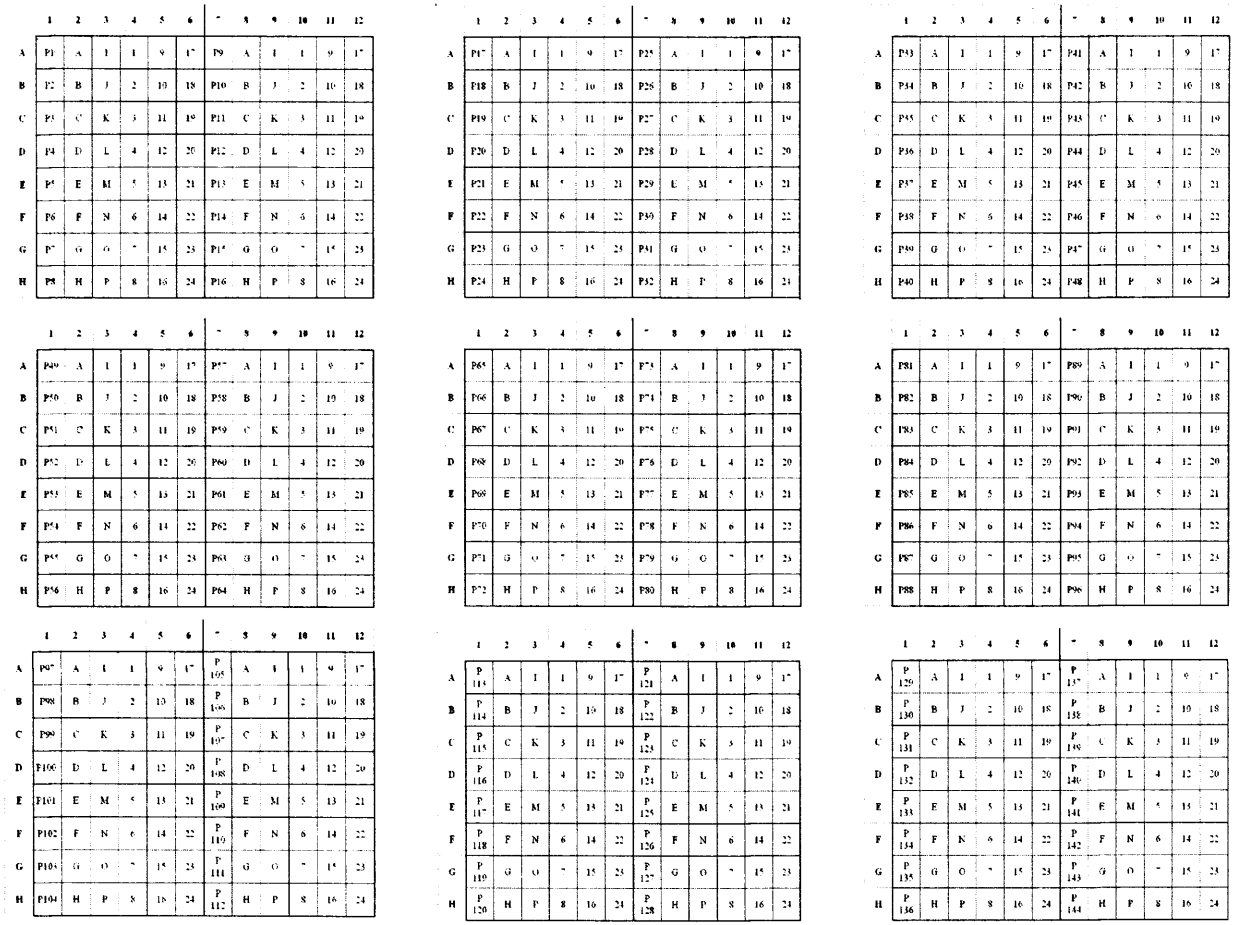


Figure 2.2 Plate layout of DNA BAC pools for MA4 library (P1 to P144) and MBP library (P1 to P96). These plates originated from *M. acuminata* 'Calcutta-4' (diploid A genome) and *M. balbisiana* 'Klutuk Wulung' (diploid B genome) respectively. Each column 1– 6 and 7–12 of BAC pools originated from sets of 8 plates. Columns 1 and 7 were obtained from plate pools (each from one 384-well plate); columns 2-3 and 8-9 were assigned as 'row pools'; columns 4-6 and 10-12 were assigned as 'column pools' (both were prepared from sets of 8 plates).

2.1.5 Abiotic stress plasmids

DNA from six plasmids containing ESTs (cloned mRNA) was obtained from CIRAD, France, and from EMBRAPA, Brazil (Table 2.5), and was used as hybridization probes and for the design of PCR primers to identify BAC clones from *M. acuminata* 'Calcutta-4' and *M. balbisiana* 'Klutuk Wulung'.

Table 2.5 Plasmids used as hybridization probes and for the design of PCR primers to identify sequences related to genes for abiotic stress from *Musa* BACs filters and by PCR-based screening.

TRAIT	DESCRIPTION	CLONE
Drought	Drought Responsive Family Protein	MA4LIMFES014A_G03
Salt	Salt Tolerance Protein (Member of the Constans Zinc Finger Family)	MACVLIMFLS011D_A07
Cold	Low Temperature Induced Protein*	MA4LIMFES003C_G06
Heat	Class-1 LMW Heat Shock Protein	MUC4LH1002_F07
	70 kDa Heat Shock Protein	MA4LIMFES014A_E03

*This follows the automated annotation from the EST using the Brazilian 'Genoma' analysis program, but the LTIP homology to EMBL AY656247 17.7 kDa low temperature induced protein OsLti17.7 is below the most widely accepted thresholds (see www.tigr.org).

Buffers:

10x TE buffer (Tris-EDTA buffer)

100mM Tris (tris-hydroxymethylamino-methane)-HCl, pH8
10mM EDTA (ethylene-diamine-tetra-acetic acid), pH8

TAE buffer	**Gel-loading dye buffer 6x:**
40 mM Tris-acetate	0.25% Bromophenol blue,
1mM EDTA	0.25% Xylene cyanol FF
	30% Glycerol

Medium:

Luria Bertani (LB) agar	Luria Bertani (LB) agar
Bacto-Tryptone 1% (w/v) Yeast Extract 0.5% (w/v) 86mM NaCl; 20mM glucose	Bacto-agar 1.5% (w/v) 20mM glucose

2.2 DNA extraction

2.2.1 Gel electrophoresis

DNA fragments were electrophoretically separated in 1-2% agarose gel (a polysaccharide consisting of a linear polymer and D-galactose and 3,6-anhydro L-galactose) using 1xTAE buffer at 5-8 V/cm for 1-2 hours as the separation capability of each molecule depends on the molecular size of the DNA fragments. DNA fragments were visualised under UV light by the addition to the gel mixture of 0.1 μg/ml ethidium bromide.

2.2.2 Genomic banana DNA isolation

Total genomic DNA was extracted from fresh leaves using the CTAB method (Gawel *et al.,* 1991) with minor modifications. Between one and two g of fresh young leaves (ideally the unexpanded cigar leaves at the top of the plant), without any obvious necrotic lesions, were snap frozen and ground in liquid nitrogen to prevent enzymatic degradation and the release of phenolic compounds from wounded tissue. Finely ground leaves were transferred to a preheated 50ml Falcon tube containing 15 ml CTAB extraction buffer (CTAB Cetyltrimethylammonium bromide 2% (w/v), 100 mM Tris-HCl, pH8; 1.4 M NaCl; 20 mM EDTA, pH8) supplemented with 30 μl β-mercaptoethanol and incubated at 60°C for an hour by gently shaking in a water-bath. One volume of chloroform-isoamylalcohol 24:1 (v/v) was added, the solution was mixed by gentle inversion at room temperature for 5 minutes, and centrifuged at 5000 rpm for 10 minutes at room temperature, resulting in a separation into three layers (aqueous at the top, leaf debris, organic solvent). The aqueous layer was transferred to a new 50 ml Falcon tube prior to the addition of 0.6 volumes of cold (-20°C) isopropanol. The mixture was gently inverted and incubated on ice for 30 minutes. Precipitated DNA was spooled and transferred to new microcentrifuge tube, which was centrifuged at 13000 rpm at RT for1 minutes to allow the removal of the aqueous waste. To wash the DNA, 100 μl of 70% ethanol was added and the tube was briefly centrifuged. The pelleted DNA was dried at 37°C for 10 minutes and re-suspended in 1xTE buffer (250-500 μl), by incubation overnight at 4°C.

10 μg/ml RNAse was added to the dissolved DNA, and incubated for 30 minutes at 37°C, after which the DNA was re-precipitated by the addition of 1/10 volume sodium acetate 3 M (pH6.8) and 2 volumes of 70% ethanol. The mixture was incubated on ice for 30 minutes prior to pelletting by centrifugation at 13000 rpm at RT for 1 minute. The pelleted DNA was dried at 37°C for 10 minutes and re-suspended in 250 μl 1xTE buffer, as above. The OD_{260} of the DNA was measured by spectrophotometry in a 1:200 (DNA:water) dilution to assess concentration, and the DNA quality was checked by 1% agarose gel electrophoresis.

2.2.3 BAC plasmid DNA isolation

BAC clones were inoculated into 10 ml sterile (autoclaved 15 minutes at 121°C under 1.2 psi pressure) Luria Bertani medium (LB) containing 5 g/l yeast extract, 10 g/l tryptone and 10 g/l NaCl, supplemented after autoclaving with 13.6 μg/ml chloramphenicol. The cultures were incubated overnight at 37°C prior to streaking onto solidified agar medium (15 g/l agar in LB). The plates required an overnight incubation at 37°C to obtain single colony BAC clones. Single colonies were introduced into 10ml sterile LB medium and incubated overnight at 37°C. Plasmid DNAs were isolated using a modified alkaline lysis method as described by Sambrook *et al.,* (1989). Plasmid DNA was quantified by measurement of OD_{260} and DNA quality was checked by agarose gel electrophoresis. RNA was removed by the addition to the isolated plasmid DNA of 1/10 volume with 10 mg/ml RNase. The quality of the plasmid DNA preparations was checked by 1% agarose electrophoresis in 1xTAE buffer, and its quantity was assessed spectrophotometrically from 1:200 (DNA:Water) dilutions.

2.3 Primer design

PCR primers were designed from protein or nucleotide alignments, which allowed for the design of degenerate and specific primers. The degenerate primers were useful for amplifying homologous genes from different species and were used in a candidate gene approach as described in Chapters 3, 4, and 5. Specific primers were applicable for the amplification of homologous sequences within the same and/or closely related species.

Degenerate primers can be occupied by one ore more possible nucleotides using International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes in representing targeted proteins. Degeneracy was calculated by multiplying a different numbers of nucleotide sequences that it represents (described by IUPAC ambiguity codes). For example, M= 2 residues, V=3 residues, N= 4 residues; for sequence GCMTNAT and GCVTNAT have degeneracy of 8 (2x4=8) and 12 (3x4=12) respectively. The IUPAC ambiguity codes are shown in Table 2.5.

Table 2.5. IUPAC ambiguity codes.

Code	Description
M	AC
R	AG
W	AT
S	CG
Y	CT
K	GT
V	ACG
H	ACT
D	AGT
B	CGT
N	ACGT

Specific oligonucleotide primers were designed for SSR, LRR and abiotic stress domains using Primer3 software (Rozen *et al.*, 2000; http://www-genome.wi.mit.edu/cgi-bin/primer/ primer3_www.cgi) aiming to generate the longest possible amplicons. DNA sequence alignments for the regions of interest are described in Chapters 3, 5, and 6. Based on DNA sequence obtained from GenBank, specific primers were designed considering melting temperature, significant product length relative to the particular region of interest, GC content, 3' stability, predicted secondary structure, primer-dimer formation between primers and primer length. Degenerate NBS and LRR primer design was based on protein alignments obtained from the EMBL database, as described in Chapters 4 and 5, respectively.

Two common procedures were applied in the design of degenerate primers: the first was the computation of the sequence alignment, and the second the identification of homology and conserved regions in the sequences. Factors considered included annealing temperature, the appropriate GC content (between 40-50%), sticky ends (for cloning proposes) and the distance between adjacent conserved regions, level of degeneracy and primer length. Specific primers relating to genes for abiotic stress tolerance were based on *Musa* EST clones and the *Musa* 3' EST database donated to the Global *Musa* Genomics Consortium (GMGC) by Syngenta and maintained at MIPS (Munich Information Centre for Protein Sequences, Munich, Germany). Use of the Syngenta *Musa* 3' EST database is acknowledged. All primers were synthesized by Sigma.

2.4 Polymerase Chain Reaction (PCR)

2.4.1 Optimization

The amount of genomic DNA and primer pairs used for PCRs are described in Chapters 3, 4, 5 and 6. PCRs were performed in a TGradient Thermocycler (Biometra) in a 25 µl reaction volume containing 1xPCR buffer (Promega), 2.5 mM $MgCl_2$, 0.23 µM dNTPs (Bioline), 0.4 µM of each degenerate primer, 0.2U *Taq* DNA Polymerase (Promega) and 50-100 ng/µl template genomic DNA.

Cycling conditions were: 94°C for 5 minutes (initial denaturation) followed 30 cycles of 94°C for 30 seconds (denaturation), the annealing temperature dependent on primer combinations for 30-50 seconds, 72°C for 60 seconds (elongation). And a final elongation step of 72°C for 1 minute. PCR products were separated by 1 to 2% agarose gel electrophoresis, depending on the expected size of the PCR products.

2.4.2 Purification of PCR-amplification fragments

PCR products from 50 µl reactions were separated by 1.2% agarose gel electrophoresis. The relevant bands were excised and purified using Qiagen Gel extraction kits following the manufacturer's protocol. An attempt was made to sequence the fragments directly from the purified fragments, but this procedure was unreliable as the PCR product was often heterogeneous, especially when degenerate primers were used. Instead, all purified fragments were cloned in pGEM-T Easy vector prior to sequencing.

2.5 Cloning

2.5.1 Competent cells

Competent cells (DH5α) were multiplied by overnight incubation at 37°C in 10ml antibiotic free LB medium. The next day, the preparations were plated out and single colonies were isolated after an overnight incubation at 37°C on antibiotic free LB-agar plates. A single colony was picked, inoculated in 10 ml antibiotic free LB medium and re-grown overnight at 37°C in an orbital incubator (shaking at 220 rpm/minute). Each 5ml overnight culture was then inoculated in a 250 ml conical flask containing 50 ml antibiotic free SOB medium supplemented with 10mM $MgSO_4.7H_2O$ and 0.2% glucose and incubated at 37°C on an orbital incubator (shaking at 220 rpm/minute) for 90-100 minutes until the OD_{600} reached value of 0.6. The culture was then centrifuged at 3000 rpm for 10 minutes at RT to pellet the bacteria, which were then gently re-suspended in 25 ml ice-cold 50 mM $CaCl_2$ in a 50 ml Falcon tube, leaving the cells on ice for 30 minutes before re-pelletting by centrifugation at 3000 rpm for 10 minutes at 4°C.

The cells were then gently re-suspended in 2.1 ml ice-cold 50 mM $CaCl_2$ on ice and 900 μl ice-cold 50% glycerol (final concentration 15%) was added. Aliquots of 100 μl of cells were placed into a microcentrifuge tube on ice, flash-frozen in liquid nitrogen and stored at -80°C.

2.5.2 Ligation and transformation

PCR fragments were purified using a Qiagen Gel extraction kit following the manufacturer's protocol. Ligation was performed in a 15 µl volume using the pGEM-T Easy Vector System I kit (Promega) consisting of 7.5 μl 2xRapid Ligation Buffer (60 mM Tris-HCl, pH7.8; 20 mM $MgCl_2$; 20 mM DTT; 2 mM ATP; 10% PEG); 0.9 µl pGEM-T Easy vector; 1.2 μl T4 DNA Ligase; 5.4 μl DNA in an overnight incubation at 4°C.

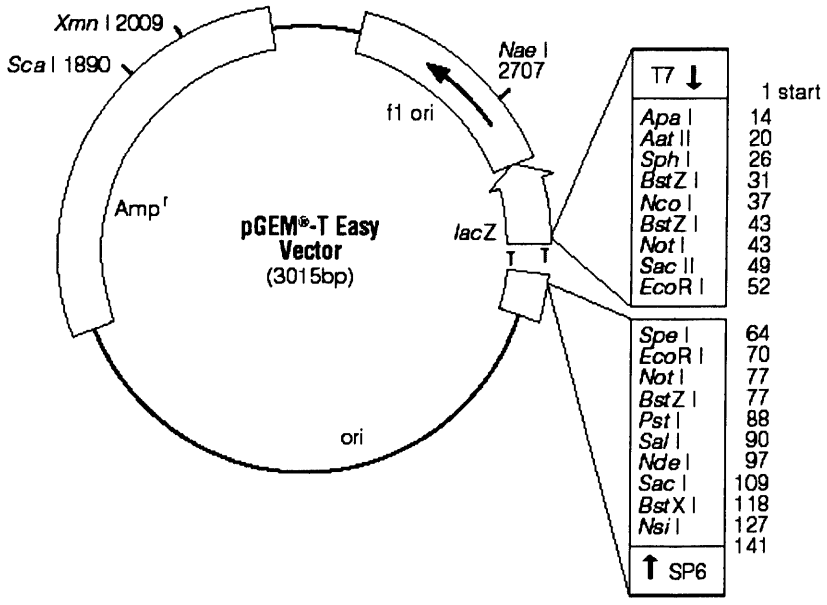


Figure 2.3. Digram of pGEM-T Easy vector with multiple restriction sites for multiple cloning region allowing realise of the insert by digestion using single restriction enzyme. Both ends of the plasmid has a 3'-terminal thymidine to increase ligation efficiency of PCR fragments blunt ends by preventing vector from recircularisation and creating compatible overhangs for PCR products.

Transformation was effected by adding 15 μl ligation mixture to 100 μl thawed competent cell (DH5α) in a 1.8 ml microcentrifuge tube, holding the reaction on ice for 30 minutes before heat-shocking at 42°C for 60 seconds. The transformed cells were then incubated on ice for another 10 minutes. Preheated antibiotic-free LB medium was then added to the transformed cells, which were incubated at 37°C for 90-100 minutes on an orbital incubator shaker at 220 rpm/min. After the incubation period, the 200 μl cell culture was plated on LB medium agar supplemented with ampicillin (100 μg/ml), X-gal 5-bromo-4-chloro-3-indolyl-β-D-galactosidase (40 μg/ml) and IPTG isopropyl-β–D–thiogalactopyranoside (0.5 mM), and incubated overnight at 37°C.

2.5.3 Recombinant cell screening

Recombinant clone selection was based on screening for white colonies, which contain a pGEM-T Easy vector with an insert. Each white colony was picked and grown in 10ml LB medium with ampicillin (40 μg/ml) overnight at 37°C in an orbital incubator (shaking at 250 rpm/minute). Confirmation of the recombinant status of the clone was effected with PCR, primed with the universal M13 forward and reverse primers which anneal either side of the multiple cloning region. The 15 μl PCR mixture contained 1xPCR buffer (Bioline) 5 μM of both, M13 primers, 0.1U *Taq* DNA Polymerase (Bioline) and 0.5 μl LB medium culture containing a putative recombinant clone. The PCR cycling conditions were 94°C for 5 minutes, 30 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds, followed by 72°C for 7 minutes. The PCR product was assessed by 1% agarose gel electrophoresis in 1xTAE buffer.

Recombinant plasmid DNA isolation was carried out from minipreps of positive clones using the Qiagen Miniprep-Kit, according to the manufacturer's protocol. The quality of the DNA was checked by 1% agarose gel electrophoresis in 1xTAE buffer. Finally, selected clones were characterized by restriction enzyme analysis in a 15 μl mixture containing 0.1U *Eco*RI, 100 ng/μl plasmid and 1xReaction Buffer, which allowed the size of the insert to be established.

2.5.4 Clone storage

For storage, 500 µl overnight cultures of selected clones were mixed with 500 µl sterilized 50% glycerol and kept in -80°C.

Solution and Buffers:

Preparing 2M $MgCl_2$	Preparing **50mM $CaCl_2$**	Preparing **250mM KCl**
($MgCl_2.6H_2O$; MW : 203.31)	($CaCl_2.6H_2O$; MW : 6219.08)	(MW: 74.56)
Dissolve 10.17g $MgCl_2$ in 20ml dH_2O, top up to 50ml	Dissolve 1.1g $CaCl_2$ in 80ml dH_2O, top up to 100ml	Dissolve 1.86g KCl in 80ml dH_2O, top up to 100ml

SOB medium :

Dissolve in 95 ml dH_2O:

- 2.0 g Tryptone
- 0.5 g Yeast Extract
- 0.05 g NaCl

Add 1 ml 250 mM KCl, adjust to pH7.0 with NaOH, top up to 100 ml and aliquots 50 ml to 250 ml flask. **Add 0.5 ml sterile 2 M $MgCl_2$** to each flask before use.

LB medium :

Dissolve in 95 ml dH_2O:

- 1.0 g Tryptone
- 0.5 g Yeast Extract
- 1.0 g NaCl

Adjust to pH7.0 with NaOH, top up to 100 ml, aliquots 10 ml to 30 ml vial and autoclave
*For plate preparation, add 1.5 g agar, autoclave in 250 ml Schott Bottle

50% Glycerol (20 ml)

- Dilute 10ml Glycerol in 10ml dH_2O, mix well and autoclave

2.6 Clone sequencing

2.6.1 DNA sequencing

The DNA concentration of selected clones was established by diluting 1µl DNA plasmid in 9 µl dH_2O, and separating this sample by 1% agarose gel electrophoresis in 1xTAE buffer. Clones were verified by restriction enzyme analysis, as described above. Double stranded plasmid DNA was sequenced commercially (JIC, Norwich) using M13F and M13R primers according to the requirements of an ABI3700 capillary sequencing system.

2.6.2 Sequence analysis: alignment and sequence trees construction

Sequence data were received in the ABI format and were analyzed by Chromas software. Both DNA and deduced amino acid sequences were aligned and sequence trees were constructed in ***CLC Free Workbench 2.2.5.*** Default settings were applied to both multiple alignments (full alignment, gap open 10, gap extension 1, ends gap as any other) and sequence trees, based on pairwise distances of the UPGMA (Unweighted Pair Group Method Using Arithmetic Averages) clustering algorithm (Michener and Sokal, 1957; Sneath and Sokal, 1973).

Bootstrap was used to evaluate reliability of the constructed tree (Felsenstein, 1985). Estimates were made by allocating and positioning a non-member sequence at the lowest average dissimilarity for clustering. The reliability of the tree was evaluated and supported with 1000 bootstrap replicates. Low bootstrap values correspond to a significant grouping of samples. Branch lengths are proportional to the average amino acid substitutions per site.

2.7 DNA for probe labelling

DNA probes were prepared from clones containing plasmids with relevant inserts (i.e. retroelements, NBS and LRR). The clones were digested with *Eco*RI overnight at 37°C before separation by 1% agarose gel electrophoresis to purify the plasmid. The plasmids were extracted from the agarose gel using the Qiagen extraction kit according to the manufacturer's protocol and quantified by 1% agarose gel electrophoresis.

2.8 Southern hybridization

2.8.1 Restriction enzyme digestion

Genomic DNA samples were digested with the restrictions enzymes *Hae*III, *Bam*HI, *Hind*III, *Eco*RV, *Eco*RI and *Sau*3A. Pilot tests were carried out to optimize DNA and enzyme concentrations. Each digestion required 6-8 µg of genomic DNA per 25 µl reaction, and restrictions were obtained from an overnight incubation at 37°C. Digested samples were separated by 1% agarose gel electrophoresis in 1xTAE buffer for 2-3 hours at 5V/cm.

2.8.2 Southern blotting

To transfer the separated DNA fragments from the agarose gel to a nylon membrane, the gel was first depurinated in 0.25 N HCl for 20 minutes, denatured in NaOH 0.5 M, NaCl 1.5 M for 20-30 minutes, and neutralized in Tris-HCl 0.5 M pH7.5, NaCl 1.5 M for 30 minutes. The gel was finally rinsed twice with dH_2O.

The transfer apparatus consisted of a tray filled with 10xSSC, and a platform to support the gel and the filter paper (Whatman 3MM) wicks. The gel was placed upside-down on the platform, followed by a wetted nylon membrane (Hybond N+, Amersham), a further three pieces of filter paper wetted with 10x SSC, and a stack of paper towels. The whole was stabilised with a weighted glass plate. Gel edges and the tray were sealed with cling film. Overnight transfer (for at least 15 hours) was required to complete the process. Following transfer, the membranes were rinsed in 2x SSC, air dried and exposed to ultra violet (UV) light to promote DNA cross-linking and stored at room temperature before use.

Buffers:

Buffer 1 : Depurination (1000 ml)	**Buffer II : Denaturation (500 ml)**	**Buffer III: Neutralisation (1000 ml)**
21.6 ml of 11.6M HCl [0.25 M] Make up to 1000 ml, autoclave, store RT	43.9 g NaCl [1.5 M] 10 g NaOH [0.5 M] Make up to 500 ml, autoclave, store RT	87.7 g NaCl [1.5 M] 500 ml of 1M Tris-HCl, pH 7.4 [0.5 M] Make up to 1000 ml, autoclave, store RT

2.8.3 Membrane hybridization

Pre-hybridisation of the prepared membrane was carried out by 3-4 hour rotation in a Hybaid hybridization tube at 60°C in a Hybaid hybridization oven. The pre-hybridization mixture consisted of formamide 50% (v/v), 5xSSC, blocking reagent (Roche) 2% (w/v), salmon sperm DNA (1 μg/μl), 100 mM EDTA and SDS (sodium dodecyl sulphate) 0.02% (w/v). The volume of hybridization solution required was 5 ml per 100 cm^2 membrane. Overnight hybridization with labelled probe was performed in the same tube at 60°C with the same formulation as for pre-hybridization plus dextran sulphate 50% (w/v).

After hybridization, the hybridization solution was stored at -20°C for possible reuse. The membranes were rinsed at RT and washed twice each with 12.5 ml 2xSSC and 0.1% SDS at 60°C for 30 minutes in rotating hybridization oven and washed twice. After the final rinse, the membranes were dried with tissue paper, wrapped in Saran-wrap and placed in an auto-radiographic cassette, along with a sheet of X-ray film (FUJI Medical X-Ray Film) on each side of the membrane.

The cassette was stored at -20°C and the exposure time required depended on signal strength, varying from overnight (> 500 counts/second) to a week (<100 counts/second) or more. The exposed X-Ray film was processed in a photographic developing machine, scanned with an EPSON 1600 Pro scanner and the image imported into Adobe Photoshop 6.0. The membranes were re-used following a stripping procedure in which they were briefly rinsed in sterile water, washed twice in 0.2 M NaOH/SDS 0.1% (w/v) for 15 minutes at 37°C, and finally rinsed in 2xSSC for 5 minutes before wrapping in Saran-wrap. When not immediately re-used, the membranes could be stored at 4°C for lengthy periods.

2.8.4 Radioactive probe labelling

Probes were labelled with the Random Primers Labelling System (Invitrogen Life Technology), following the manufacturer's protocol. Approximately 50-100 ng plasmid DNA was dissolved in 10 μl sterilized dH_2O and denatured in boiling water for 5 minutes followed by immediate cooling on ice. The labelling mixture consisted of 2 µl dATP, 2 µl dGTP, 2 µl dTTP, 20 μl 2x5 Random Primer Buffer Mix and 13µl dH_2O made up to a total volume of 49 μl. Pipette tips containing dried ^{32}P-dCTP were used to mix and at the same time dissolve the ^{32}P in the tip. After the addition of 1μl Klenow Fragment (3U/µl) to the reaction, the mixture was centrifuged and then incubated at 37°C for 30 to 60 minutes. The probe was denatured at 100°C for 10 minutes before adding to the hybridization mixture.

2.8.5 Scoring signal from membrane filters

Digested genomic DNAs were scored as banding patterns on the X-ray film. For BAC filters, the scoring was based on the square-grid orientation and confirmed by replication of the BAC clones on the square (order of pattern) as shown in Figure 2.1.

Buffers:

20x SSC (saline sodium citrate or standard saline citrate)

175.3 g NaCl
88.2 g $Na_3C_6H_5O_7.2dH_20$
(Top up to 1000 ml dH_20, autoclave, store RT)

SDS (Sodium dodecyl sulphate)

10%: 10 g in 100 dH_20

Dextran Sulfate (50%)

5 g in 10 ml in dH_20 (filter sterilized)

2.9 *In situ* hybridization

2.9.1 Root preparation and chromosome spreads

Root tips were collected and pre-treated in 0.05% 8-hydroxyquinoline for 1 hour at RT and metaphase spreads of fixed root tips were prepared according to the protocol described by Dolezel *et al.* (1998). Fluorescent *in situ* hybridization (FISH) and genomic *in situ* hybridization (GISH) were carried according to Schwarzacher and Heslop Harrison (2000) with minor modifications.

2.9.2 Probe labelling

The *Pseudoviridae* fragment was amplified from MA4_2A3 by Ty1 and Ty2 primers, cloned into pGEM-T Easy vector, and labelled as 2BAC_A38. 2BAC_A38 and a 5S rDNA insert present in pTa794 (Gerlach and Dyer 1980) were each labelled with both DIG (digoxigenin 11-dUTP) and BIO (Biotin 16-dUTP) by PCR labelling using universal M13 primers. BIO and DIG were synthesized by Life Technologies. PCRs were performed in a TGradient Thermocycler (Biometra) in 30 μl volumes containing 1xPCR buffer (Bioline), 1.5 mM $MgCl_2$, 0.23 μM dNTPs (Bioline), 0.33 μM M13 primers, 0.2U *Taq* DNA Polymerase (Promega) and 10-15 ng/μl template DNA. Cycling conditions were: 94°C for 5 minutes (initial denaturation), followed 35 cycles of 94°C for 30 seconds (denaturation), 45°C for 30 seconds (annealing), and 72°C for 60 seconds (elongation), with a final elongation of 72°C for 1 minutes.

To remove unincorporated nucleotides, *Taq* DNA polymerase (from PCR labelling) and salts, 1/10 volume of 3 M sodium acetate (pH 6.8) and 2 volumes of 96% ethanol was added to the reaction, and held overnight -20°C. The mixture was then centrifuged at 12000 rpm for 30 minutes at 4°C, washed in 70% EtOH, and centrifuged at the same speed for 1 minute at 4°C. The supernatant was discarded and the pellet was air dried at 37°C for 10 minutes before dissolving in TE and storing at -20°C.

Dot blots were performed to check the efficiency of the probe labelling. A nylon membrane was wetted in buffer 1 (100 mM Tris-HCl, pH 7.5; 15 mM NaCl) for 5 minutes and dried between filter papers. Labelled probe (0.5 μl) was spotted onto the membrane, which was then air-dried for 5 minutes. The membrane was then immersed in buffer 1 for 1 minute, incubated with agitation in buffer 2 (blocking reagent (Roche) 0.5% (w/v) in buffer 1) for 30 minutes at RT followed by antibody solution (anti-biotin-alkaline phosphatase (Roche) or anti-digoxigenin-alkaline phosphatase (Boehringer Mannheim) conjugated antibody diluted to 0.75U/ml in buffer 1) for another 30 minutes at 37°C. Membranes were washed by immersion in buffer 1 with agitation for 5-10 minutes at RT prior to equilibration with buffer 3 (100 mM Tris-HCl, pH9.5; 100 mM NaCl; 50 mM $MgCl_2$) for 2 minutes.

Finally, the membrane was incubated in buffer 3 containing NBT (4-nitroblue tetrazolium chloride at 75 mg/ml) and BCIP (5-bromo-4-chloro-2-indolylphosphate at 50 mg/ml) and detection reagents (Life Technologies) at room temperature in the dark. Well labelled probes produced a strong dark-grey signal.

2.9.3 Slide preparations

Metaphase spreads for *in situ* hybridization were prepared by a dropping technique. Fresh, healthy roots were harvested from greenhouse grown plants, and were pre-treated by soaking in root buffer (RB) for 5 minutes before being treated with 8-hydroxyquinoline (0.05% w/v) for 4 hours at room temperature, and fixed in 3:1 96% ethanol:acetic acid for 24 hours at 4°C. The roots were then transferred to 70% ethanol and stored at -20°C.

The fixed root tips were cut into two pieces in the Suspension Buffer (SB) with a sharp blade, and treated with 2% cellulase and 2% pectinase in a microcentrifuge tube for 1 hour at 30°C to suspend the cells. The suspended cells were filtered (nylon filter, sieve size 150 μm) into a clean microcentrifuge tube and incubated for 5 minutes at room temperature followed by centrifuging at 1000rpm for 5 minutes. The supernatant was discarded and 200 μl SB was added to re-suspense the pellet. The process was repeated twice and followed by the addition of 200 μl 70% ethanol. Finally, after the third supernatant was discarded, 100 μl of 70% ethanol was added and the suspended cells were stored at -20°C.

5 μl of the stored suspension cells were dropped onto an ice-cold slide, followed by 5 μl of freshly prepared fixative. The slide was dried briefly, and then soaked in absolute ethanol for 10 seconds and left to dry at room temperature. Metaphase chromosomes were observed under a phase contrast light microscope at a 400 magnification.

2.9.4 Fluorescent *In situ* Hybridization (FISH)

Slides were pre treated by adding 200 μl RNAse (100 μg/ml RNAse in 2xSSC), covering with a plastic cover slip and incubating in a moist chamber for 1h at 37°C. The slides were then washed in 2xSSC for 2 minutes and this step was repeated for another 10 minutes, followed by 0.01 M HCl for 2 minutes on an orbital shaker at 37°C.

To reduce the amount of cytoplasm in the cells, 200 µl of pepsin (5 µg/ml in 0.01M HCl) was added to each slide, which was covered with plastic coverslip. The slides were incubated for 10 minutes at 37°C followed by rinsing with dH_2O for 1 minute and washing twice in 2xSSC for 5 minutes on an orbital shaker at 37°C. Finally, the slides were treated with 4% paraformaldehyde and washed twice with 2xSSC for 5 minutes at RT prior to dehydration in a graded ethanol series of 70%, 90% and 96% each for 2 minutes followed by air drying.

The FISH mixture was made up to a 40 µl volume, and consisted of 50–100 ng/slide DNA probe, formamide 50% (v/v), 2xSSC, dextran sulphate 10% (v/v), SDS 0.125% (w/v), 0.125mM EDTA and 1µg salmon sperm DNA. The formamide and EDTA enhances binding between probe and target sequence (chromosomes) at 37°C whereas salmon sperm DNA prevents hybridization at non-specific sites.

The hybridization mixture was denatured at 70°C for 10 minutes and immediately cooled on ice for 10 minutes prior to spreading over the chromosome preparation. Denaturation of probe and chromosomes in hybridization affected by treatment at 70–80°C for 6 minutes followed by overnight incubation in ThermoHybaid model HyPo-20 at 37°C. Post hybridization washes consisted of an immersion in 2xSSC at 40°C to remove the plastic cover slip, a rinsing in 2xSSC at 42°C for 1 minute, and two stringent washes with formamide 20% (v/v) in 0.1x SSC at 42°C for 5 minutes with agitation. The final wash was in 0.1xSSC at 42°C for 5 minutes and 2xSSC at 42°C for 3 minutes. Each washing step involved agitation and was repeated twice. For signal detection, the slides were agitated for 5 minutes in detection buffer (4xSSC, Tween20 0.1%(v/v)), and 250 µl 5% Bovine Serum Albumin (w/v) (BSA; BSA blocks non specific sites hybridized to the chromosomes) prepared in detection buffer was then applied, followed by an incubation in a moist box at 37°C for 30 minutes before the addition of the detection solution (consists of anti-digoxigenin-antibody and/or streptavidin conjugated to flourochromes to detect DNA probe).

The detection solution was prepared by adding 1-6 μg/ml conjugated anti-Digoxigenin (2-3 μl FITC antiDig [200 μg/ml stock]) and conjugated avidin or streptavidin (1 μl Alexa streptavidin [200 μg/ml] or 1 μl Cy3streptavidin [1 mg/ml]) to 200 μl blocking solution supplemented with 5% BSA (w/). On each slide, 50 μl of solution was applied, covered by plastic cover slip and incubated for 60 minutes at 37°C in the dark. The slides were then washed with detection buffer at 40°C first for 2 minutes and then twice more for 8 minutes each with agitation on an orbital shaker. Counterstaining was effected by the addition of 100 μl DAPI (4', 6-diamidino-2-phenylindole) (4 μg/ml in McIlvaine's buffer) (Sigma), and the slides were covered with a plastic cover slip and incubated in the dark for 10 minutes at room temperature. The slides were then rinsed in detection buffer before mounting in anti-fade solution (prevent fluorescently stained molecules from fading during observation under fluorescent microscope) (AF1, Agar Scientific). Finished slides were kept in cold, dry and dark conditions for at least for 24 hours prior to analysis.

2.10 Photography and imaging

Slides prepared for FISH were kept in the dark at 4°C prior to image capture, to enhance fluorescent dyes and ensure that the antifade has been fully absorbed. The slides were analyzed with an Axioplan 2 epifluorescence microscope with single band pass filters (Zeiss) equipped with a cooled CCD camera (Optronics, model S97790). Images were overlaid and analyzed with Adobe Photoshop 6.0 software.

CHAPTER 3

3.0 Diversity of retroelements and microsatellite sequences in the banana (*Musa* spp.) genome and large-insert clones BACs genome

3.1 Introduction

The work presented in this chapter aims to understand specialized aspects of genomic diversity and large scale genome organization in *Musa*, particularly with respect to two major repetitive genomic components, retroelements and microsatellites (or simple sequence repeats, SSRs). Currently, many interesting and important genes have been isolated through BAC libraries in many plant species and have been used for sequencing, genome mapping and amplification by PCR for searching well-defined genomic regions for phenotypically significant genes of interest in plant breeding. This might be an advantage for study of dispersion in BAC clones which carries part of the genomic DNA and might be expected to give important information to understand and develop marker systems and making further assessment in plant genome evolution.

Retroelements, or type I transposable elements, are abundant ubiquitous DNA elements which form RNA transcripts, encode reverse transcriptase (RT) and other protein constituents. They copy themselves and reinsert into the genome giving a dispersed distribution, and sometimes lie adjacent to or within genes (Hansen and Heslop-Harrison, 2004). The RT sequence is the only coding region that has conserved regions between all retroelements and is therefore useful for sequence tree analysis. However, repeats in the coding sequence from retroelements exist without any clear biological functions but might be important as marker systems and targeting genes for gene structure and evolution. Retroelement mapping is important to understand the structural organization of the *Musa* genome. Localization of *Pseudoviridae* and *Metaviridae* elements in other plant species is well studied and it is known that some *Pseudoviridae* and *Metaviridae* retrotransposons are not widely dispersed in the genome but distributed in centromeres and telomeres of the chromosome respectively. According to Heslop-Harrison *et al.* (1997) *Pseudoviridae* retrotransposons are mainly distributed evenly in plant chromosomes, but differ between species.

Fluorescent *in situ* hybridization using the retroelement RT sequences as probe enabled aspects of genome organization in the retrotransposon to be analyzed in *Musa* spp. However cytogenetic mapping, based on cytological studies is hard to perform as *Musa* chromosomes are relatively small in size and 'vanish' in the cytoplasm in the cells limiting the study at the cytogenetic levels.

As reported by San Miguel *et al.* (1998), there is evidence that the retroelements contribute to gene regulation and cellular stress responses through rapid turnover and re-amplification of the elements within species. The *Pseudoviridae* (Ty1-*copia* like) and *Metaviridae* (Ty3-*gypsy* like) primers were designed based on RT conserved region by Flavell *et al.* (1992) and Kubis *et al.* (1998) respectively to amplify the retrotransposons in many species.

DNA microsatellites (SSR), which are also abundant, exist as non-coding sequences and are distributed throughout the genome. The SSR sequences can be flanked and amplified by PCR-based techniques using a pair of primers flanking to the repeat sequences. Conservation and specificity of the sequences within individual plants show a co-dominant mode of inheritance by variation in the number of the repeats produced at different sizes of PCR products, creating polymorphism amongst the species or accessions within a species.

The polymorphisms and co-dominant inheritance allows discrimination between homozygous and heterozygous alleles which are useful for individual and/or population fingerprinting, genetic mapping and diversity studies. As complete sequences of *Musa* BAC clones are accessible from Aert *et al.* (2004) (and since 2006, following completion of the work described here, additional sequences from the Genbank/EMBL database and www.musagenomics.org), repeat sequences can be searched for SSR primers.

One of the main strategies in *Musa* breeding is based on making crosses between a partially fertile triploid banana and a wild and/or cultivated diploid donor which has special desired characteristics such as resistance and/or high yield. However, this approach requires better understanding and knowledge of genetic diversity of both wild and cultivated donor as the procedures consume time and are extremely laborious to accomplish.

In this work, genomic DNA from *Musa* was used to amplify retroelements of LTR-retrotransposons (*Pseudoviridae* and *Metaviridae*) and SSR sequences by using degenerate and specific oligonucleotide primers respectively. Bacterial artificial chromosome (BAC) clones derived from *M. acuminata* and *M. balbisiana* were chosen arbitrarily and incorporated in this study to understand organisation and dispersion of the retroelements and SSRs in *Musa* spp. Previously, many works have been reported for retroelements related to *Pseudoviridae* and *Metaviridae* in other plant species; my aim was to understand diversity and organization of these retroelements in *Musa* species and diverse accessions through PCR-based techniques to amplify, clone, sequence and construct sequence trees. Many primers for SSR have been introduced and my aim was to utilize sequences in two BAC clones published by Aert *et al.* (2004) namely MuH9 and MuG9 by design SSR primers flanking the SSR tandem repeats in the BAC sequences and understand the use of BAC-derived SSR markers for *Musa* spp.

3.2 Material and methods

3.2.1 DNA isolation

Plasmid isolation for 10 BAC clones provided by MGRC (Table 3.1) and genomic DNA isolation of *Musa sp.* were described in Chapter 2. A total of 56 genomic DNA samples were used for PCR amplification with *Pseudoviridae* and *Metaviridae* primers.

Table 3.1. List of 10 BAC clones originated from *M. acuminata* 'Culcutta-4' and *M. balbisiana* 'Klutuk Wulung' used for this study.

No	Clone Name (*M. acuminata*) 'Culcutta-4'	No	Clone Name (*M. balbisiana*) 'Klutuk Wulung'
1	MA4_2A02	1	MBP_1B01
2	MA4_2A03	2	MBP_1B02
3	MA4_2A04	3	MBP_1B03
4	MA4_2A05	4	MBP_1B04
5	MA4_2A06	5	MBP_1B05

3.2.2 **Retroelement Primers**

Degenerate oligonucleotide primers were used in amplification of both *Pseudoviridae* and *Metaviridae* retroelement groups. For *Pseudoviridae* upstream primer Ty1-1 encoding the amino acid sequence 'TAFLHG' and downstream Ty1-2 *copia* primer encoding 'YVDDML' were originated from Flavell *et al.* (1992) and amplified band at approximately 270 bp. In *Metaviridae*, a combination of primers Gyrt1 encoding 'LSGYHQI' and Gyrt4 encoding 'YAKLSKC' upstream and downstream respectively (Table 3.2) followed Friesen *et al.* (2001) and Kubis *et al,* (1998) resulted in a PCR product at approximately 300bp. The same primers were used by Vershinin *et al.* (2002) and Kubis *et al.* (2003) on *Hordeum* and *Elaeis guineensis* respectively amplified PCR product at approximately 420 bp.

Table 3.2. Retroelement primers used varies in T_a (annealing temperature) and T_m (melting temperature) given is based on label from the company (SIGMA) product.

Primer Name	Sequence 5'-3'	T_m (°C)	% Degeneracy	T_a (°C)
Ty1-1 (Forward)	CAN GCN TTY YTN CAY GG	53.3	35	44
Ty1-2 (Reverse)	ARC ATR TCR TCN ACR TA	39.7	30	
Gyrt1 (Forward)	MRN ATG TGY GTN GAY TAY MG	50.7	40	42
Gyrt4 (Reverse)	RCA YTT NSW NAR YTT NGC R	50.5	53	

Nucleotide degeneracies:
R=A+G; Y=C+T; M=A+C; S=G+C; W=A+T; N=A+G+C+T

PCR reactions were performed in 25 µl reaction and cycling condition was described in Chapter 2 (Refer 2.4.1). Annealing temperature used for *Pseudoviridae* primers and *Metaviridae* primers were at 44°C and 42°C respectively. PCR products were separated on 1.5% agarose gel.

3.2.3 Cloning and sequencing

Genomic DNA from wild type of *Musa* i.e. *M. acuminata*, *M. balbisiana*, *M. schizocarpa* and *M. textilis* representing A, B, S and T constitutions were amplified, cloned and sequenced. All PCR products obtained were cloned and screened as described in Chapter 2.

3.2.4 Microsatellite primers

Primers flanking SSRs were designed (described in Chapter 2) based on two BAC sequences of MuG9 (15 primer sets) and MuH9 (20 primer sets) published by Aert *et al.* (2004). Initially, four primers were selected each from BAC MuH9 (MuH9-SSR1, MuH9-SSR5, MuH9-SSR6 and MuH9-SSR12) and BAC MuG9 (MuG9-SSR1, MuG9-SSR2, MuG9-SSR12 and MuG9-SSR14) to test on genomic DNA of 54 *Musa* accessions from wild to hybrid cultivars and/or species (Table 3.3). The selected primers as shown in table 3.4.

PCR reactions were performed in a 25 µl reaction as described in Chapter 2 (Refer to 2.4.1) for 30 cycles. Annealing temperature for MuH9-SSR5, MuH9-SSR6 was 54°C, whereas MuH9-SSR1 and MuG-SSR12 were 51°C and 57°C respectively. PCR products were separated on a 2.0% agarose gel which was sufficient to resolve major polymorphisms.

Table 3.3. Genomic DNA samples used for SSR study.

Genome Constitution	AA	BB	TT	BT	SS	AAA	ABB	AAB	*Musa ornata*	*Musa velutina*
No of sample	10	4	2	1	3	12	7	13	1	1

Table 3.4. Eight of the microsatellite primer pairs designed from BAC clones sequences of (A) MuG9 and (B) MuH9 published by Aert *et al.* (2004).

(A) Primer Designed from BAC MuG9 sequence

Primer Name	**Primer sequence (5' to 3')** (First line: Forward Primer) (Second Line: Reverse Primer)	**Repeat**	**Expected Size (bp)**
MuG9-SSR1	TTT GGC TTC GTG CCT CTC TTT GAT TGT TGT AGA CAA TTG GTT C	$(CT)_8$	150
MuG9-SSR2	TCA TCC GCA TCA CTA GAA CG TGT CTT GGC ATG CAT CTC TC	$(AG)_{15}C(GA)_8$	162
MuG9-SSR12	TGC AAC ACA AGC CCA CTT AC TGT CTT GGC ATG CAT CTC TC	$(GA)_{14}GC(GA)_8$	162
MuG9-SSR14	AAA AAT TCC CTA CAT GTC TTC G TGA TCA TCC ATC CCA CAT CTC	$(ATT)_{26}$	201

(B) Primer Designed from BAC MuH9 sequence

Primer name	**Primer sequence (5' to 3')** (First line: Forward Primer) (Second Line: Reverse Primer)	**Repeat**	**Expected Size (bp)**
MuH9-SSR1	TGC GTT TCC AGT GGA TTA TG AGA CAC TTG GAG AGG GAG AGG	$(CT)_{11}$	155
MuH9-SSR5	CGT TCC CTT CTT TGC CTT TAC CGG TCG AAC ATC TGG AAG AC	$(TCC)_8$	177
MuH9-SSR6	TGG CAA TCT AGT TGG ATT AGG G GTA GTT GGG AGG AGG TGT GG	$(CTT)_3$	206
MuH9-SSR12	AAC TAG GCA GCA AAC CTT CG TTC CCA AAG GCA AAG AAG AG	$(CTTT)_6$	224

3.2.5 Sequence analysis

Sequence alignment and tree construction were described in Chapter 2, section 2.6.2.

3.2.6 *In situ* hybridization

Root harvesting, slide preparation and fluorescent *in situ* hybridization (FISH) including *Pseudoviridae* probe labelling (derived from MA4-2A03 labelled as 2BAC-2A38) and *in situ* procedure were described in Chapter 2. A repeated unit of 5S rDNA-gene derived from wheat known as pta794 (410 bp) was as a control DNA probe (Gerlach and Dyer, 1980).

3.2.7 Southern hybridization

Genomic DNAs of *Musa* accessions were digested with six restriction enzymes i.e. *Eco*RV, *Eco*RI, *Sau*3A, *Hae*III, *Bam*HI and *Hin*dIII, whereas BAC plasmids were digested with eight restriction enzymes i.e. *Apa*I, *Hpa*II, *Msp*I, *Eco*RI, *Sau*3AI, *HaeIII*, *Bam*HI and Hind*III*. Probe labelling (BAC plasmid derived from MA4-2A03 labelled as 2BAC-2A38) for Southern hybridization was described in Chapter 2.

3.3 Results

3.3.1 Retroelements

3.3.1.1 PCR amplification

Pseudoviridae primers produced a dominant single band at approximately 270 bp (Figure 3.1). *Metaviridae* primers produced seven patterns of multiple bands ranging approximately from 400 to 1200 bp; the most dominant bands were approximately at 1050 bp and 1400bp (Figure 3.2).

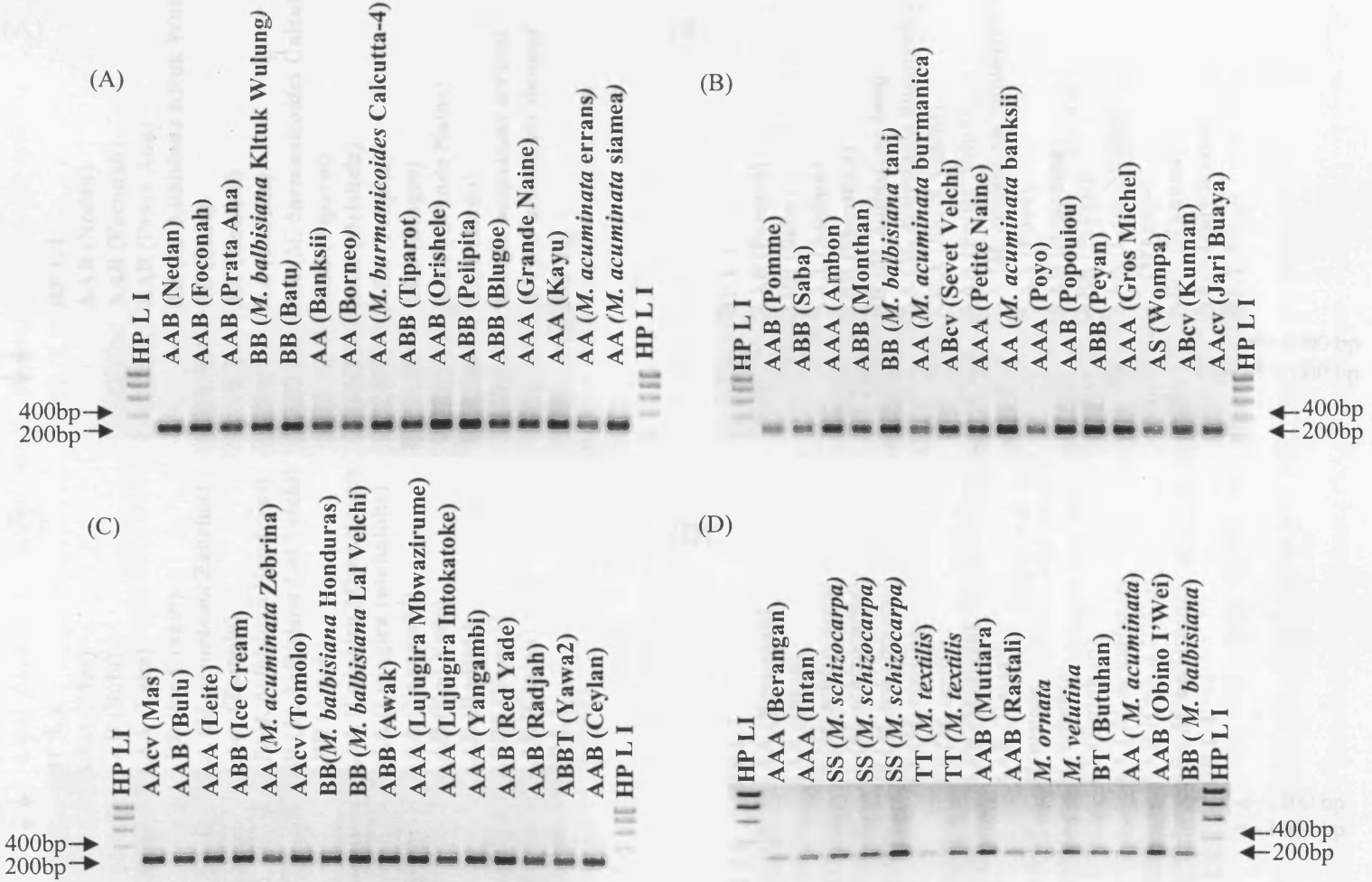


Figure 3.1. Gel image (A-D) of PCR product of retroelements from *Pseudoviridae* group *(Copia-like*) amplified from genomic DNA by using Ty1-1 and Ty1-2 primers with HyperLadder I (HPL I) as marker at both side of the gel. *Copia-like* primers produced a dominant single band at approximately 270bp. four commercially cultivated banana varieties from Malaysia are Berangan, Intan, Rastali and Mutiara.

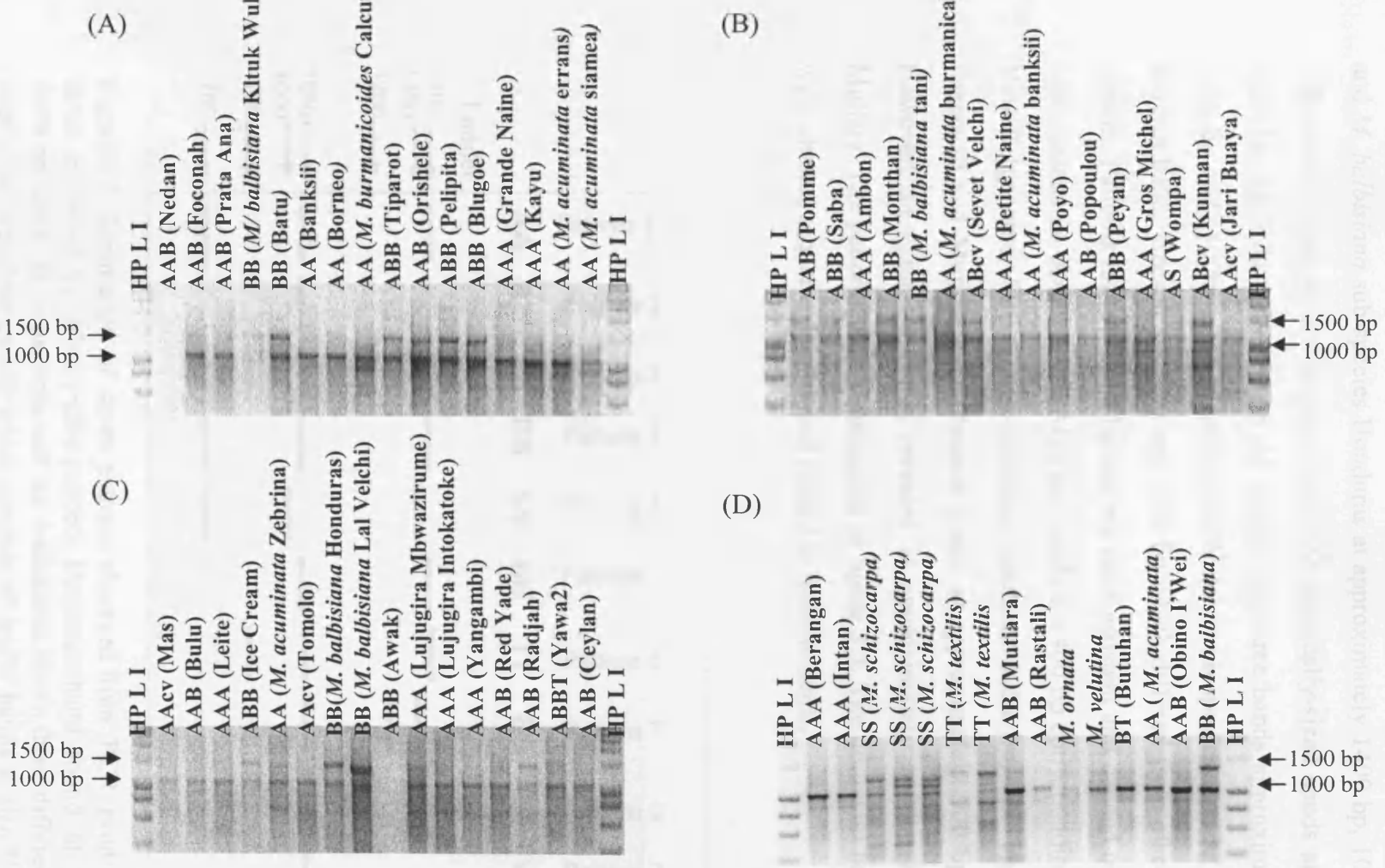


Figure 3.2. Gel image (A-D) of PCR product of retroelements from *Metaviridae* group *(Gypsy-like*) amplified from genomic DNA by using Gyrt1 and Gyrt4 primers with HyperLadder I (HPL I) as marker at both side of the gel. Gyrt-*like* primers produced a dominant band at approximately 1050bp and 1400bp. Four commercially cultivated banana varieties from Malaysia namely Berangan, Intan, Rastali and Mutiara are included.

Both *M. acuminata* and *M. balbisiana* had three different patterns of fragments and one of the patterns was identical as observed in *M. acuminata* sub-species *zebrina* and *M. balbisiana* sub-species Honduras at approximately 1400 bp, 1050 and 700 bp. All samples from *M. schizocarpa* had three identically-sized bands at approximately 1050 bp, 1200 bp and 1350 bp. *M. textilis* had three bands approximately at 900bp, 1050 bp and 1400bp. *M. balbisiana* 'Lal Velchi' sub-species type-3 (sample number 40) had bands sizes at 1000 bp and 1300 bp which differed from other *M. balbisiana* species. Patterns for cultivated banana are not consistent and depend on their genomic constitutions i.e. AAB, AAA, ABB and bands at 1400 bp corresponding to B genome. Yawa2 (genome ABBT) has dominant bands at 1050 bp, 900 bp and 1400 bp. Berangan and Mutiara have identical bands at approximately 900 bp and 1050 bp. However, an extra band was revealed at approximately 1400bp for Rastali and Mutiara. The patterns were summarised in figure 3.3. The dominant bands at 1050 bp were cloned, sequenced and found related to retroelements.

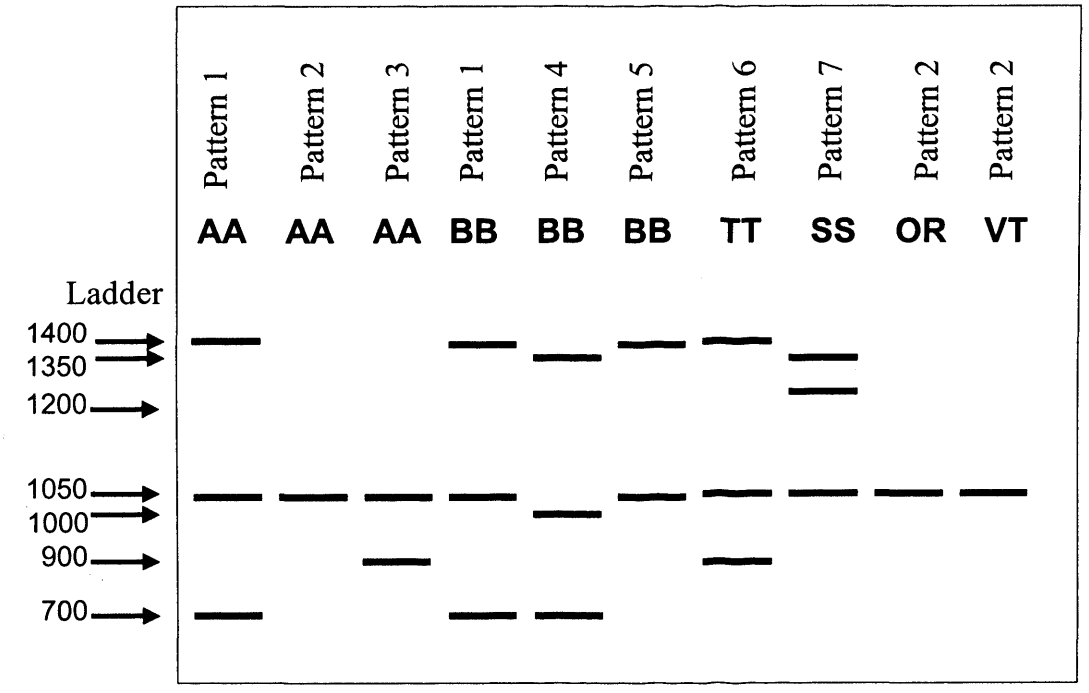


Figure 3.3. Summaries of seven patterns observed from PCR product of genomic DNA amplified by *Metaviridae* primers. Dominant band found at approximately 1050 bp. Both *M. acuminata* and *M. balbisiana* shows three different patterns but sharing in one of the pattern which consists of triple bands at size 700 bp, 1050 bp and 1400 bp. Three bands were identified in *M. textilis* and *M. schizocarpa* identical at 1050bp. *M. ornata* and *M. velutina* have identical single bands at same size at 1050 bp. The marker shows the size of the related bands and not in scales.

The retroelements primers were further used on 10 BAC clones originating from *M. acuminata* (MA4_2A02, MA4_2A03, MA4_2A04, MA4_2A05, and MA4_2A06) and *M. balbisiana* (MBP_1B01, MBP_1B02, MBP_1B03, MBP_1B03, MBP_1B04, MBP_1B05). Single product band amplified from *Pseudoviridae* primers at approximately 270 bp. As expected multiple bands amplified from *Metaviridae* primers with dominant bands at approximately 1050 (Figure 3.4).

The dominant bands of the PCR products obtained from *Metaviridae* and single product from *Pseudoviridae* were selected each from *M. acuminata* for cloning and sequencing. Selected BACs from *Pseudoviridae* amplification were MA4_2A03 and MBP_1B03. Both BAC sequences were confirmed related to *Pseudoviridae* retrotransposons and pooled together with RT fragments amplified from genomic DNA for multiple analysis and tree construction.

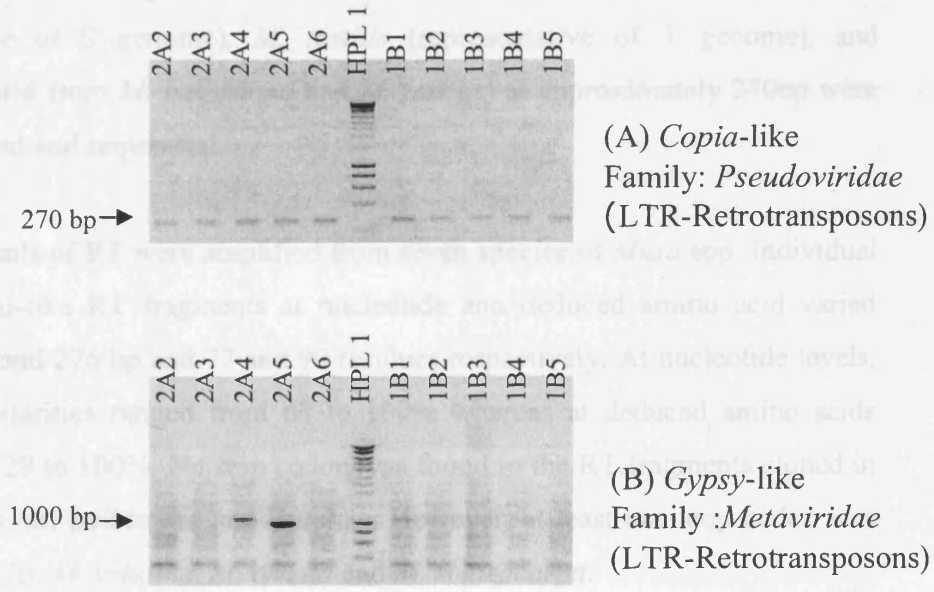


Figure. 3.4. PCR products of 10 BAC clones originated from *M. acuminata* (MA4_2A2, MA4_2A3, MA4_2A4, MA4_2A5, MA4_2A6) and *M. balbisiana* (MBP_1B1, MBP_1B2, MBP_1B3, MBP_1B4, MBP_1B5) with HyperLadder 1 (HPL 1)) as marker in the middle of the gel. Primers from (A) *Pseudoviridae* Ty1-1/Ty1-2 primers, (B) *Metaviridae* Gyrt1/Gyrt4 primers.

For *Metaviridae*, three candidates of BACs were selected for cloning and sequencing, two from *M. acuminata* (MA4_2A04, MA4_2A05) and one from *M. balbisiana* (MBP_1B01). Sequence derived from BAC MA4_2A05 was slightly shorter (approximately 50 bp) as compared to MBP_1B01 and MA4_2A04. RT sequences of *Metaviridae* amplified from BAC clones were not good as compared to genomic DNA. They were confirmed as related to the *Metaviridae* RT *M. acuminata* partial *Metaviridae* retrotransposon clone BTY3-8 (EMBO Accession no.AM040200) with more than 80% similarity at nucleotide level. Clone G42 consists of a *Metaviridae* RT fragment amplified from genomic DNA of *M. acuminata* 'Calcutta-4' which has 90% similarity at nucleotide level and was used as probe for Southern hybridization.

3.3.1.2 Sequence analysis of *Pseudoviridae*

In order to understand how the distribution of these retroelements in *Musa* spp., only *Pseudoviridae* like retroelements was further selected and the dominant bands from *M. acuminata* (Calcutta-4; representative of A genome), *M. balbisiana* (representative of B genome), *M. ornata, M. velutina, M. schizocarpa* (representative of S genome), *M. textilis* (representative of T genome), and Butuhan (hybrid from *M. balbisiana* and *M. textilis*) at approximately 270bp were isolated, cloned and sequenced.

21 fragments of RT were amplified from seven species of *Musa* spp. Individual sizes of *copia*-like RT fragments at nucleotide and deduced amino acid varied between 259 and 276 bp and 77 and 90 residues respectively. At nucleotide levels, fragment similarities ranged from 67 to 100% whereas at deduced amino acids ranging from 29 to 100%. No stop codon was found in the RT fragments cloned in *M. acuminata, M. balbisiana* and Butuhan. However, at least one stop codon was found *M. textilis, M. velutina, M. ornata* and *M. schizocarpa.*

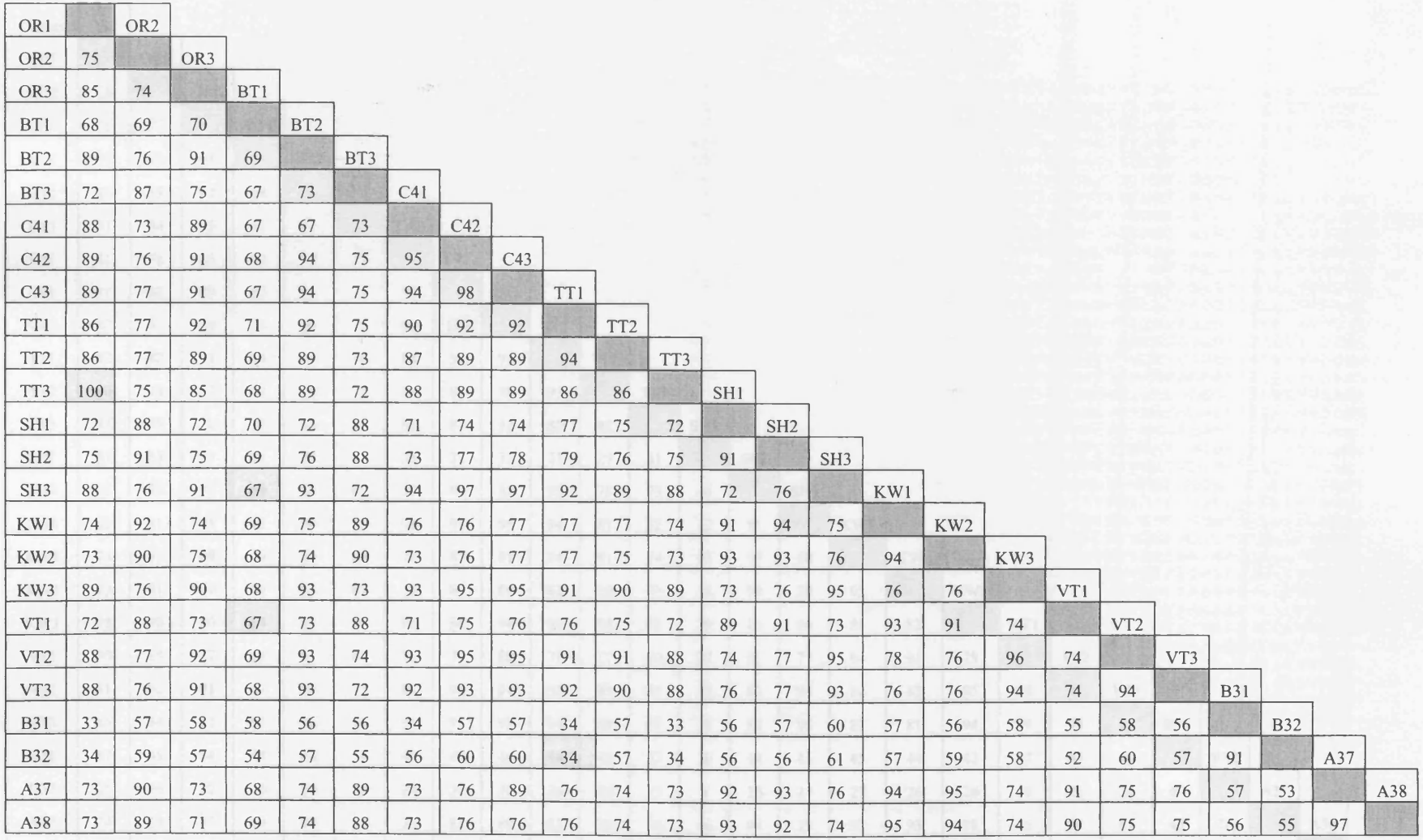
Average similarity of RT fragments at deduced amino acids in each species were 77% (*M. ornata)*, 69 (Butuhan), 93% (*M.* acuminata Calcutta-4), 86% (*M. textilis)*, 52% (*M. schizocarpa)*, 86% (*M. balbisiana)* and 84% (*M. velutina).* Whereas at nucleotide were 78% (*M. ornata)*, 70 (Butuhan), 96% (*M.* acuminata 'Calcutta-4'), 87% (*M. textilis)*, 80% (*M. schizocarpa)*, 82% (*M. balbisiana)* and 81% (*M. velutina).*

The highest and the lowest similarity of each clone in the individual species at deduced amino acid level were *M. acuminata* and *M. schizocarpa.* At nucleotide levels, *M. acuminata* was still the highest but Butuhan was found to be the lowest similarity amongst them. The similarity in percentage at deduced amino acids and nucleotides are shown in Figure 3.5 and Figure 3.6.

All the deduced amino acids sequences obtained from genomics *Musa* sp. (three clones from each species) and BAC clones (two clones from each fragment) were aligned in CLC free workbench version 2.2.5. As references, *M. acuminata* and ten other RT sequences from different species from EMBL/Gene bank were included for reference species (Figure 3.7).

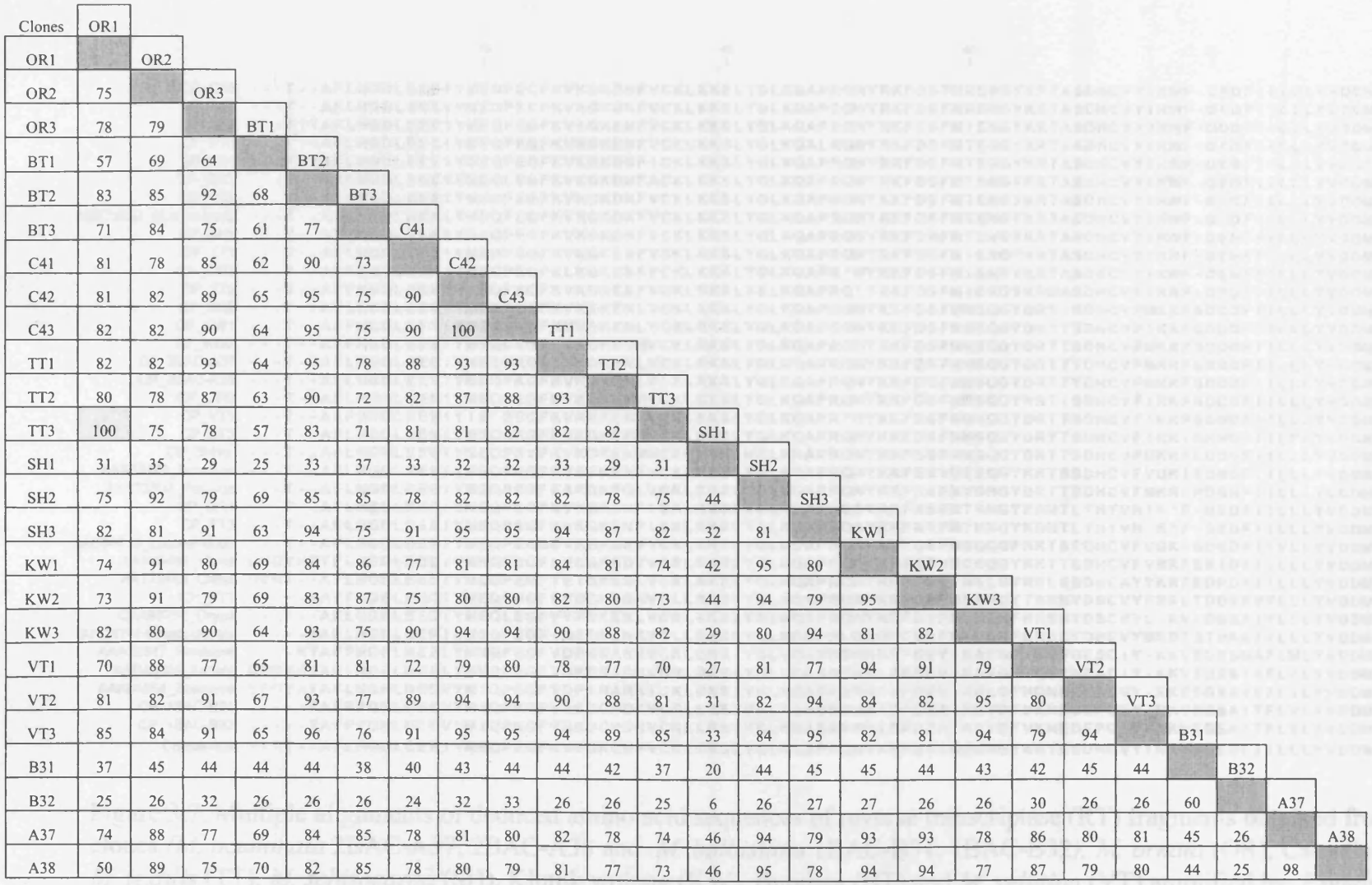
Sequence tree with bootstrap was generated using UPGMA method from the alignment and a *Citrus* element was used as an out group (Figure 3.8).The sequence tree of RT fragments based on deduced amino acid sequences resulted two main clusters but is not strongly supported by the bootstrap.

All RT sequences for reference species except *Populus* were found in different clades. RT sequences from *Secale, Hordeum and Zea mays* species were grouped together with CP_1BAC-B31 and 1BAC-B32. RT fragments from *Oryza* and Butuhan (hybrid from *M. balbisiana* and *M. textilis*) were grouped together with high bootstrap support value. Another three species, *Solanum esculentum* (formerly genus *Lycopersicon*), *Solanum* and *Vigna* were grouped together with their clades clearly separated from the majority of *Musa* RT clones.



	OR1	OR2	OR3	BT1	BT2	BT3	C41	C42	C43	TT1	TT2	TT3	SH1	SH2	SH3	KW1	KW2	KW3	VT1	VT2	VT3	B31	B32	A37	A38
OR1																									
OR2	75																								
OR3	85	74																							
BT1	68	69	70																						
BT2	89	76	91	69																					
BT3	72	87	75	67	73																				
C41	88	73	89	67	67	73																			
C42	89	76	91	68	94	75	95																		
C43	89	77	91	67	94	75	94	98																	
TT1	86	77	92	71	92	75	90	92	92																
TT2	86	77	89	69	89	73	87	89	89	94															
TT3	100	75	85	68	89	72	88	89	89	86	86														
SH1	72	88	72	70	72	88	71	74	74	77	75	72													
SH2	75	91	75	69	76	88	73	77	78	79	76	75	91												
SH3	88	76	91	67	93	72	94	97	97	92	89	88	72	76											
KW1	74	92	74	69	75	89	76	76	77	77	77	74	91	94	75										
KW2	73	90	75	68	74	90	73	76	77	77	75	73	93	93	76	94									
KW3	89	76	90	68	93	73	93	95	95	91	90	89	73	76	95	76	76								
VT1	72	88	73	67	73	88	71	75	76	76	75	72	89	91	73	93	91	74							
VT2	88	77	92	69	93	74	93	95	95	91	91	88	74	77	95	78	76	96	74						
VT3	88	76	91	68	93	72	92	93	93	92	90	88	76	77	93	76	76	94	74	94					
B31	33	57	58	58	56	56	34	57	57	34	57	33	56	57	60	57	56	58	55	58	56				
B32	34	59	57	54	57	55	56	60	60	34	57	34	56	56	61	57	59	58	52	60	57	91			
A37	73	90	73	68	74	89	73	76	89	76	74	73	92	93	76	94	95	74	91	75	76	57	53		
A38	73	89	71	69	74	88	73	76	76	76	74	73	93	92	74	95	94	74	90	75	75	56	55	97	

Figure 3.6. Similarity in percentage at nucleotide sequences of RT *Pseudoviridae* retroelement fragments between individual clones of *Musa sp.*



Clones	OR1	OR2	OR3	BT1	BT2	BT3	C41	C42	C43	TT1	TT2	TT3	SH1	SH2	SH3	KW1	KW2	KW3	VT1	VT2	VT3	B31	B32	A37	A38
OR1																									
OR2	75																								
OR3	78	79																							
BT1	57	69	64																						
BT2	83	85	92	68																					
BT3	71	84	75	61	77																				
C41	81	78	85	62	90	73																			
C42	81	82	89	65	95	75	90																		
C43	82	82	90	64	95	75	90	100																	
TT1	82	82	93	64	95	78	88	93	93																
TT2	80	78	87	63	90	72	82	87	88	93															
TT3	100	75	78	57	83	71	81	81	82	82	82														
SH1	31	35	29	25	33	37	33	32	32	33	29	31													
SH2	75	92	79	69	85	85	78	82	82	82	78	75	44												
SH3	82	81	91	63	94	75	91	95	95	94	87	82	32	81											
KW1	74	91	80	69	84	86	77	81	81	84	81	74	42	95	80										
KW2	73	91	79	69	83	87	75	80	80	82	80	73	44	94	79	95									
KW3	82	80	90	64	93	75	90	94	94	90	88	82	29	80	94	81	82								
VT1	70	88	77	65	81	81	72	79	80	78	77	70	27	81	77	94	91	79							
VT2	81	82	91	67	93	77	89	94	94	90	88	81	31	82	94	84	82	95	80						
VT3	85	84	91	65	96	76	91	95	95	94	89	85	33	84	95	82	81	94	79	94					
B31	37	45	44	44	44	38	40	43	44	44	42	37	20	44	45	45	44	43	42	45	44				
B32	25	26	32	26	26	26	24	32	33	26	26	25	6	26	27	27	26	26	30	26	26	60			
A37	74	88	77	69	84	85	78	81	80	82	78	74	46	94	79	93	93	78	86	80	81	45	26		
A38	50	89	75	70	82	85	78	80	79	81	77	73	46	95	78	94	94	77	87	79	80	44	25	98	

Figure 3.5. Similarity in percentage at deduce amino acids of RT *Pseudoviridae* retroelement fragments between individual clones in *Musa sp.*

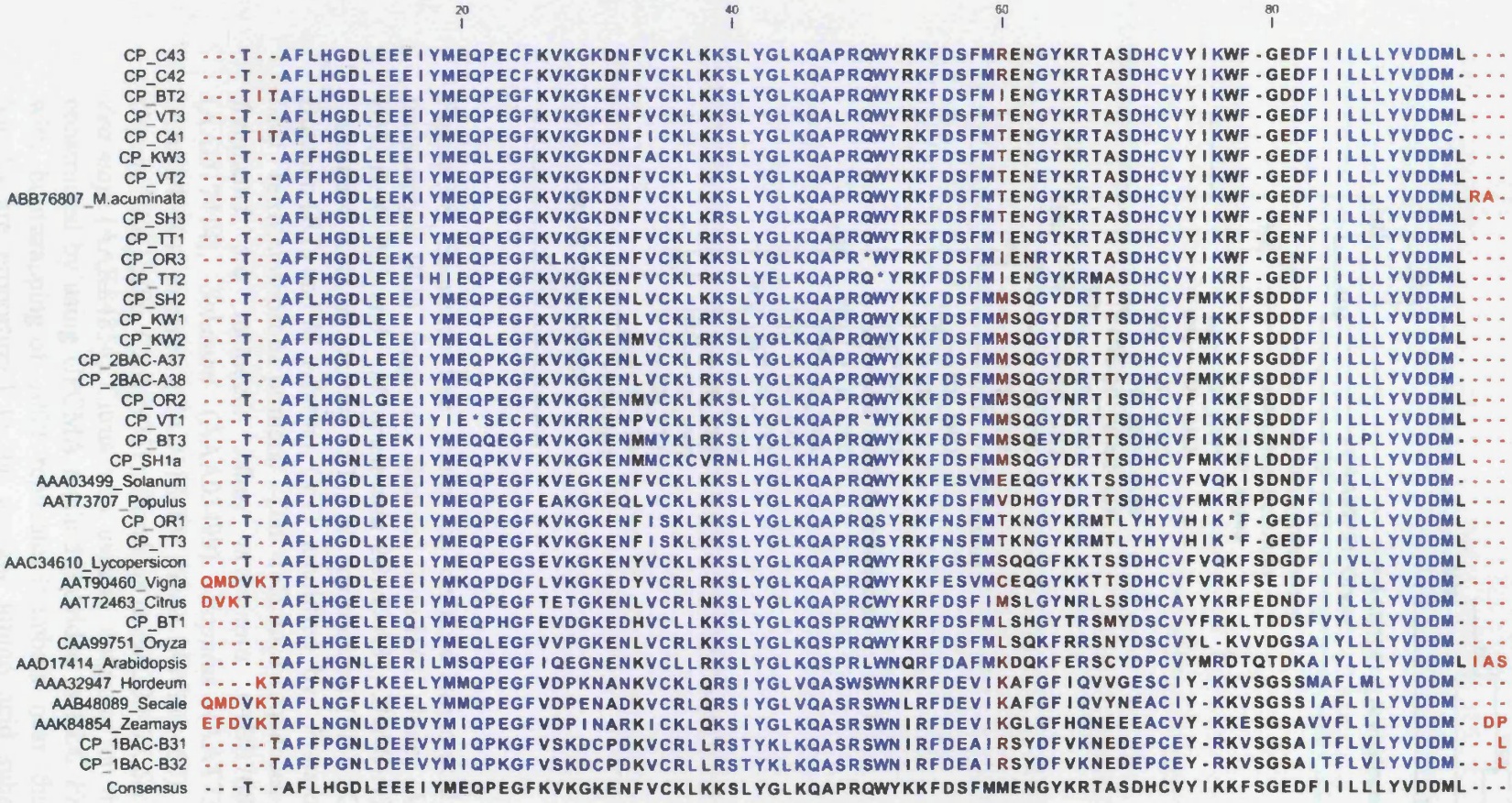


Figure 3.7. Multiple alignments of deduced amino acid sequences of reverse transcriptase (RT) fragments obtained from BAC clones *(M. acuminata* 2BAC-A37, 2BAC-A38 and *M. balbisiana* 1BAC-B31, 1BAC-B32*), M. ornata* (OR), Calcutta-4 (C4), *M. textilis* (TT), *M. schizocarpa* (SH), Klutuk wulung (KW), Butuhan (BT) and *M. velutina* (VT) amplified by *Pseudoviridae (Ty1-copia-like) retrotransposons* primers. Fifteen reference sequences from different RT fragments are included: *M. acuminata* (ABB76807), *Arabidopsis* (AAD17414), *Solanum* (AAA03499), *Populus* (AAT73707), *Lycopersicon* (AAC34610), *Brassica* (CAD11848), *V igna* (AAT90460), *Citrus* (AAT72463), *Oryza* (BAA12896, CAA99751), *Hordeum* (AAA32947), *Secale* (AAB48089) and *Zea mays* (AAK84854). The alignment was provided by ***CLC Free Workbench 2.2.5.***

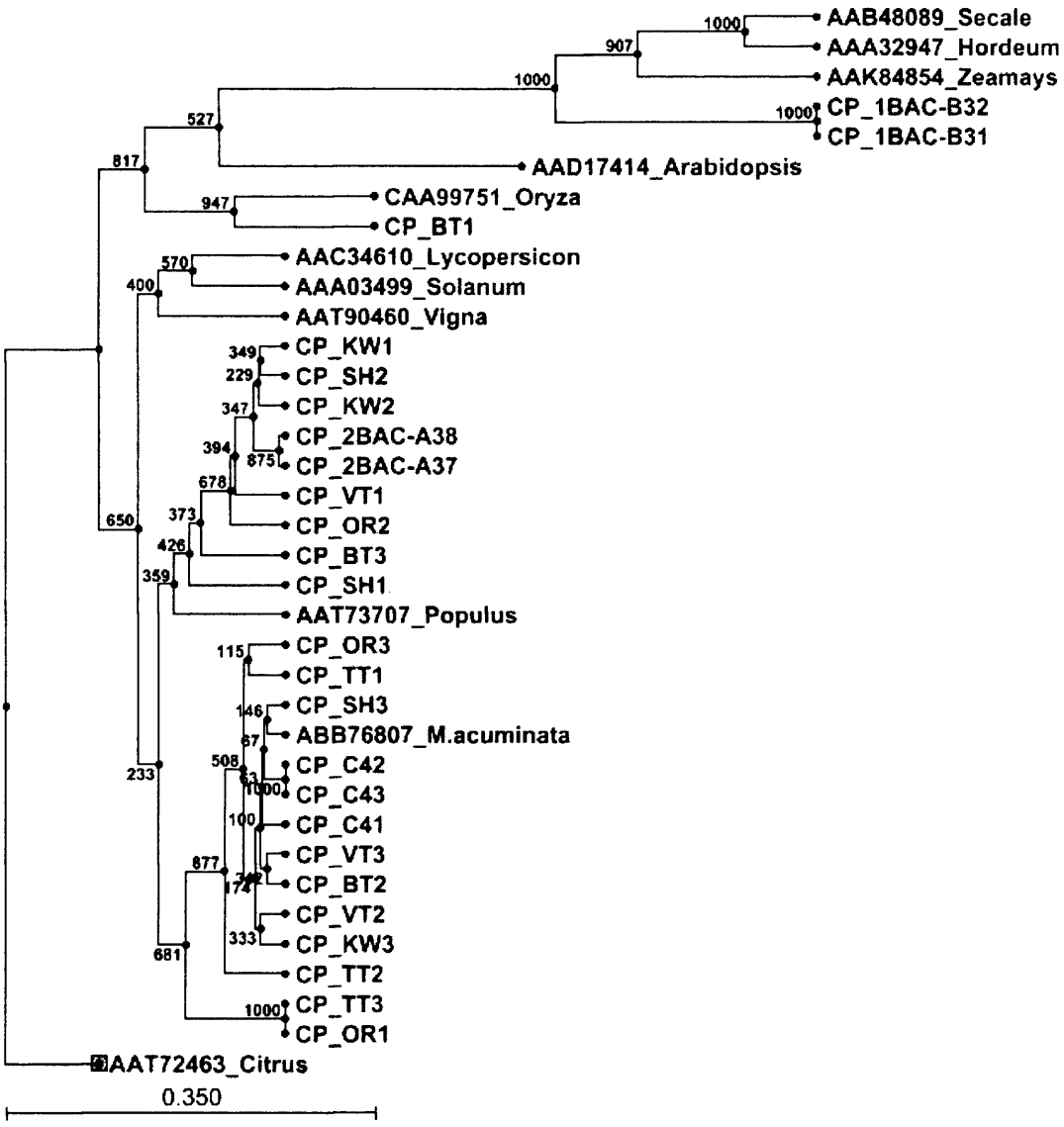


Figure 3.8. Sequence tree model on deduced amino acid sequences of reverse transcriptase (RT) fragments obtained from BAC clones *(M. balbisiana,* 1BAC-B31, BAC-B32 and *M. acuminata* 2BAC-A37, 2BAC-A38*), M. ornata* (OR), Calcutta-4 (C4), *M. textilis* (TT), *M. schizocarpa* (SH), Klutuk wulung (KW), Butuhan (BT) and *M. velutina* (VT) amplified by *Pseudoviridae (Ty1-copia-like)* retrotransposons primers. Ten reference sequences from different RT fragments are included: *Musa acuminata* (ABB76807), *Arabidopsis* (AAD17414), *Solanum* (AAA03499), *Populus* (AAT73707), *Lycopersicon* (AAC34610), *Brassica* (CAD11848), *Vigna* (AAT90460), *Citrus* (AAT72463), *Oryza* (BAA12896, CAA99751), *Hordeum* (AAA32947), *Secale* (AAB48089) and *Zea mays* (AAK84854). Citrus was used as outgroup for this tree. The tree was constructed by using UPGMA method provided in ***CLC Free Workbench 2.2.5*** with bootstrapping of 1000 replicates (numbers near the branches). Branch lengths are proportional to the average amino acid substitutions per site as indicated by the scale.

The sequence tree of RT sequence of *Pseudoviridae* was again constructed based on deduced amino acid sequences amongst the *Musa* RT clones. Four main clusters supported by high bootstrap values were found in *Musa*. RT from MBP_1B03 has a separated cluster. Two main clusters were mixed where no specific species were allocated in one cluster. Interestingly, one clone from Butuhan was allocated strong support by bootstrap in an individual cluster (Figure 3.9).

Nucleotide sequences related to RT were aligned (Figure 3.10) and the unrooted sequence tree clearly showed six clusters of the RT fragments in *Musa* spp. supported by strong bootstrap values (more than 85%). Three clones from Butuhan were found in different clusters. Remarkably RT sequence for clone CP-TT3 and CP-OR1 were 100% identical and grouped in separated clusters from other clones which derived from the same species (Figure 3.11).

3.3.1.3 Genomic organisation of retrotransposons

Genomic organization of *Pseudoviridae* and *Metaviridae* were observed by Southern hybridization. *Musa* genomic DNA from 18 accessions was digested with six different restriction enzymes, blotted on membrane filters and hybridized with ^{32}P radioactive labelled probe from clone 2BAC-A38 (*Pseudoviridae)* and G42 (*Metaviridae*) shown in figure 3.12.

*Eco*RV digestion gave two strong bands to all samples for *Pseudoviridae* probe at 3 kb and 1 kb. However, multiple bands between 0.4 kb to 5 kb were found in all samples with *Metaviridae* probe with polymorphic bands at line 2 (*M. acuminata* 'Calcuta-4'), 4 (*M. balbiasiana* 'Klutuk Wulung'), 5 (*M. balbisiana* 'Tani'), 6 *(M. shizocarpa*), 7 (*M. textilis*) and 8 (*M. ornata*). *Sau*3A digestion resulted in double bands at 5kb for *Pseudoviridae* but *Metaviridae* produced multiple bands at 4 to 10kb in line 1 (Mas), 10 (Giant Cavendish), 11 (Berangan), 12 (Awak), 13 (Pelipita) 14 (Mutiara).

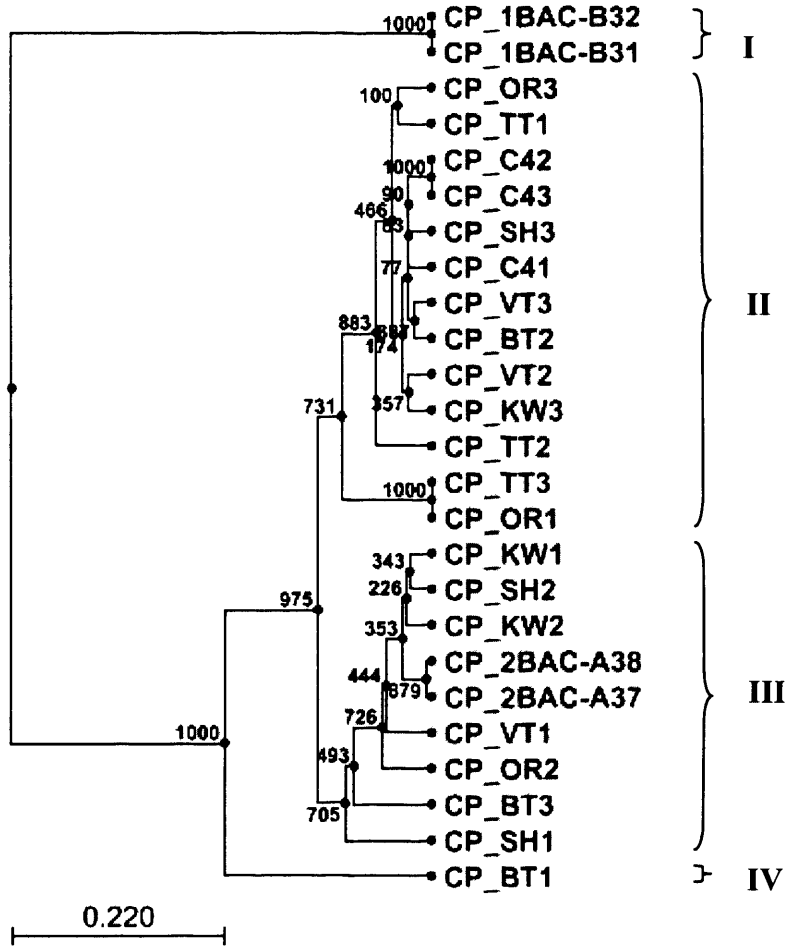


Figure 3.9. sequence tree model on deduced amino acid sequences of reverse transcriptase (RT) fragments obtained from BAC clones *(M. balbisiana,* 1BAC-B31, BAC-B32 and *M. acuminata* 2BAC-A37, 2BAC-A38*), M. ornata* (OR), Calcutta-4 (C4), *M. textilis* (TT), *M. schizocarpa* (SH), Klutuk wulung (KW), Butuhan (BT) and *M. velutina* (VT) amplified by *Pseudoviridae (Ty1-copia-like)* retrotransposons primers. The tree was constructed by using UPGMA method provided in ***CLC Free Workbench 2.2.5*** with bootstrapping of 1000 replicates (numbers near the branches). Branch lengths are proportional to the average amino acid substitutions per site as indicated by the scale.

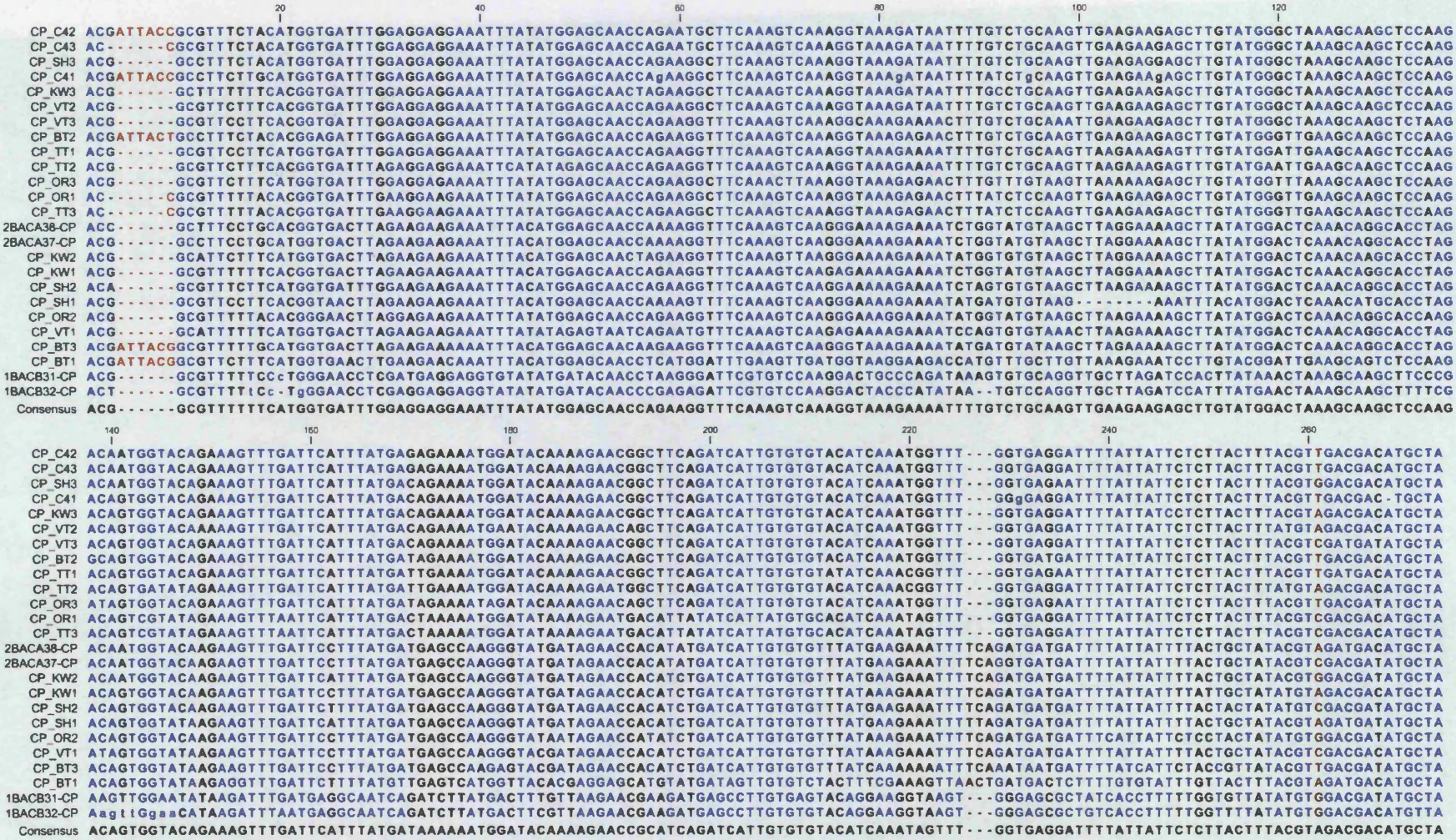


Figure 3.10. Multiple alignments of nucleotide sequences of reverse transcriptase (RT) fragments obtained from BAC clones *(M. acuminata* 2BAC-A37, 2BAC-A38 and *M. balbisiana* 1BAC-B31, 1BAC-B32*), M. ornata* (OR), Calcutta-4 (C4), *M. textilis* (TT), *M. schizocarpa* (SH), Klutuk wulung (KW), Butuhan (BT) and *M. velutina* (VT) amplified by *Pseudoviridae (Ty1-copia-like)* retrotransposons primers. The alignment was provided in ***CLC Free Workbench 2.2.5.***

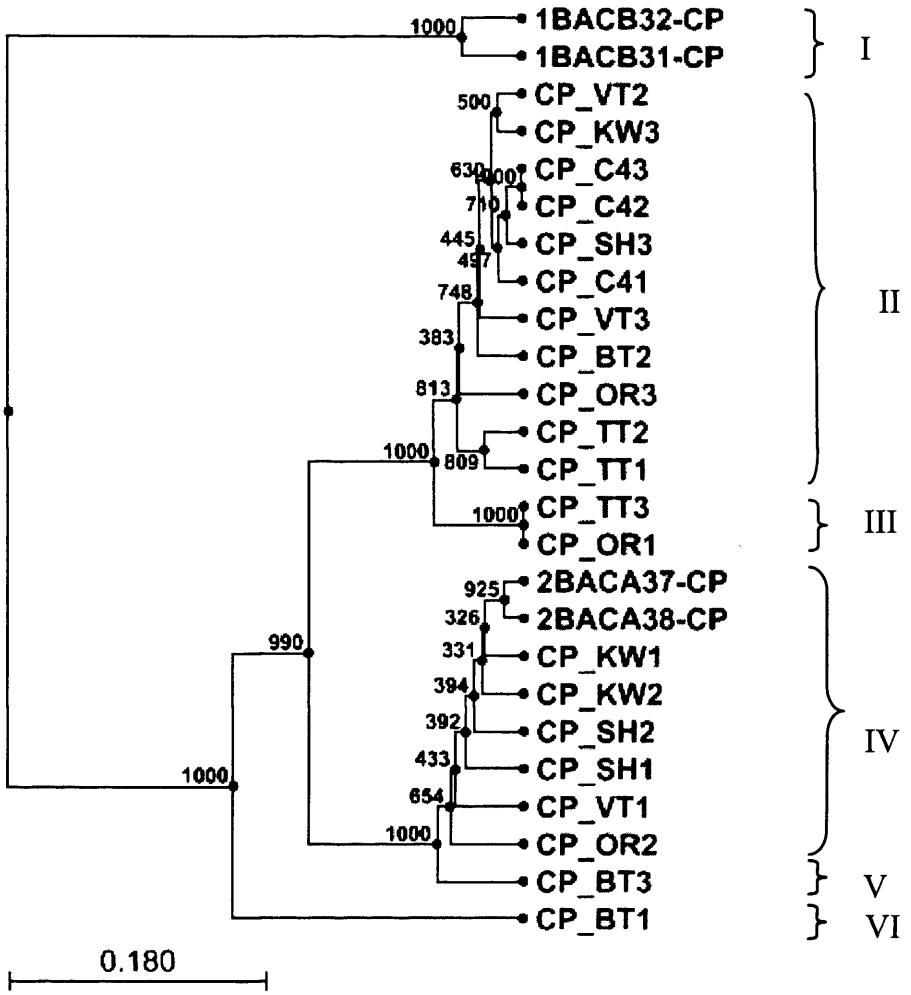


Figure 3.11. Sequence tree model on nucleotide sequences of reverse transcriptase (RT) fragments obtained from BAC clones *(M. balbisiana,* 1BAC-B31, BAC-B32 and *M. acuminata* 2BAC-A37, 2BAC-A38*), M. ornata* (OR), Calcutta-4 (C4), *M. textilis* (TT), *M. schizocarpa* (SH), *M. balbisiana* 'Klutuk wulung' (KW), Butuhan (BT) and *M. velutina* (VT) amplified by *Pseudoviridae (Ty1-copia-like)* retrotransposons primers. Six clusters of RT sequences supported by highly percentage of bootstrap were observed but without any corresponding to genome constitution in *Musa sp.* The tree was constructed by using UPGMA method provided in ***CLC Free Workbench 2.2.5*** with bootstrapping of 1000 replicates (numbers near the branches). Branch lengths are proportional to the average nucleotide substitutions per site as indicated by the scale.

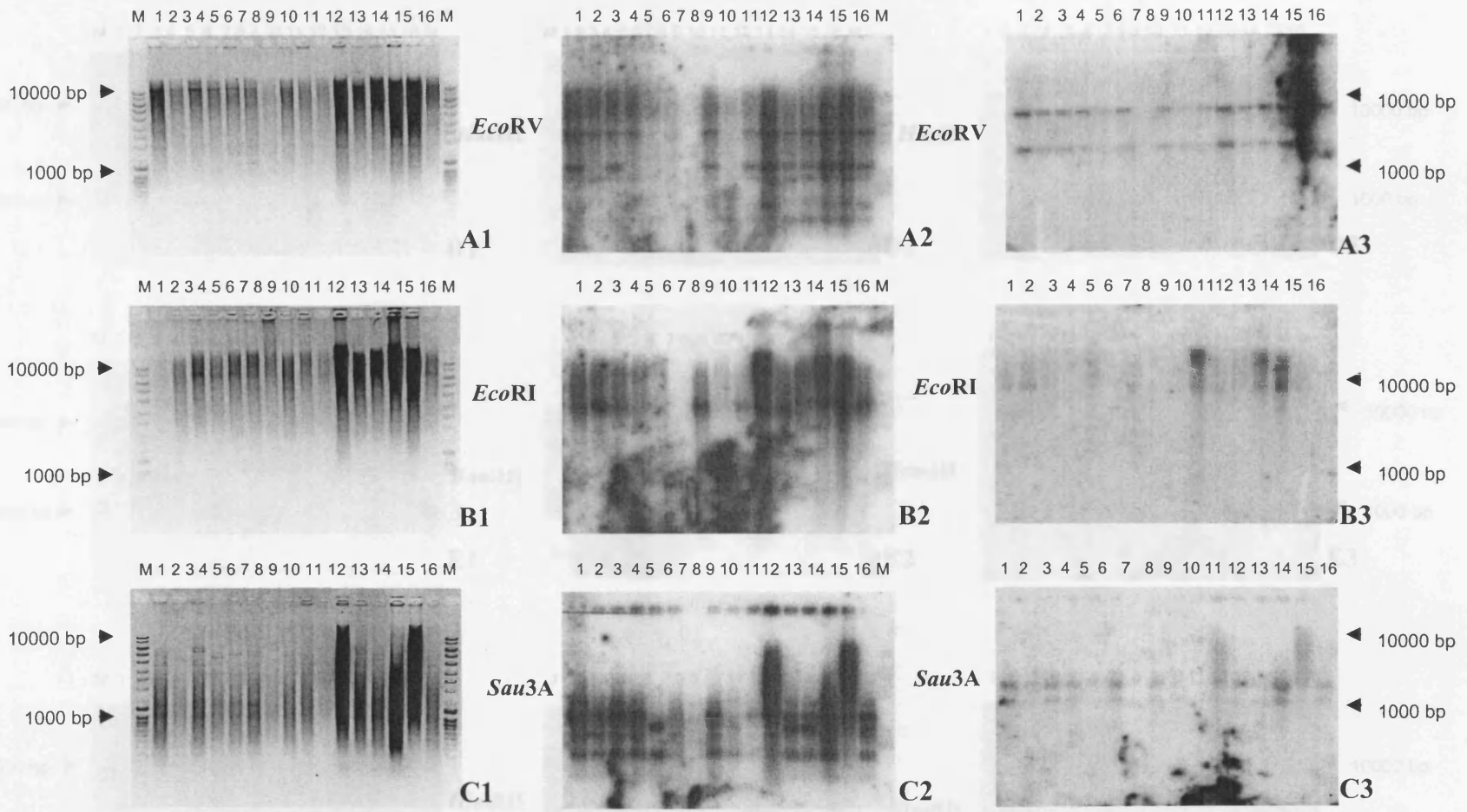


Figure 3.12. Genomic organization of *Musa sp*. analyzed by southern hybridization using *Ty1-Copia-like* and *Ty3-Gypsy-like* probe on 16 genomic DNA digested with 6 different restriction enzymes (A-*EcoRV*, B-*EcoRI*, C- *Sau3A*, D- *Hae*III, E-*BamHI*, F-*HindIII*). Electrophoresis gel of digested genomic DNA stained with ethidium bromide (A1-F1). Membrane filters hybridized with *Ty3-Gypsy*-like and *Ty1-copia-like* probe obtained from clone Gy42 (A2-F2) and A38 (A3-F3) respectively. Sample 1) Mas (AA), 2) *M. acuminata* 'Calcutta-4' (AA), 3) *M. malaccensis* (AA), 4) *M. balbisiana* ' Klutuk Wulung' (BB), 5) *M. balbisiana* 'Tani' (BB), 6) *M. schizocarpa* (SS), 7) *M. textilis* (TT), 8) *M. ornata*, 9) *M. velutina*, 10) Giant cavendish (AAA), 11) Berangan (AAA), 12) Awak (ABB), 13) Pelipita (ABB), 14) Mutiara (ABB), 15) Rastali (ABB), 16) Obino L'Ewai (AAB).

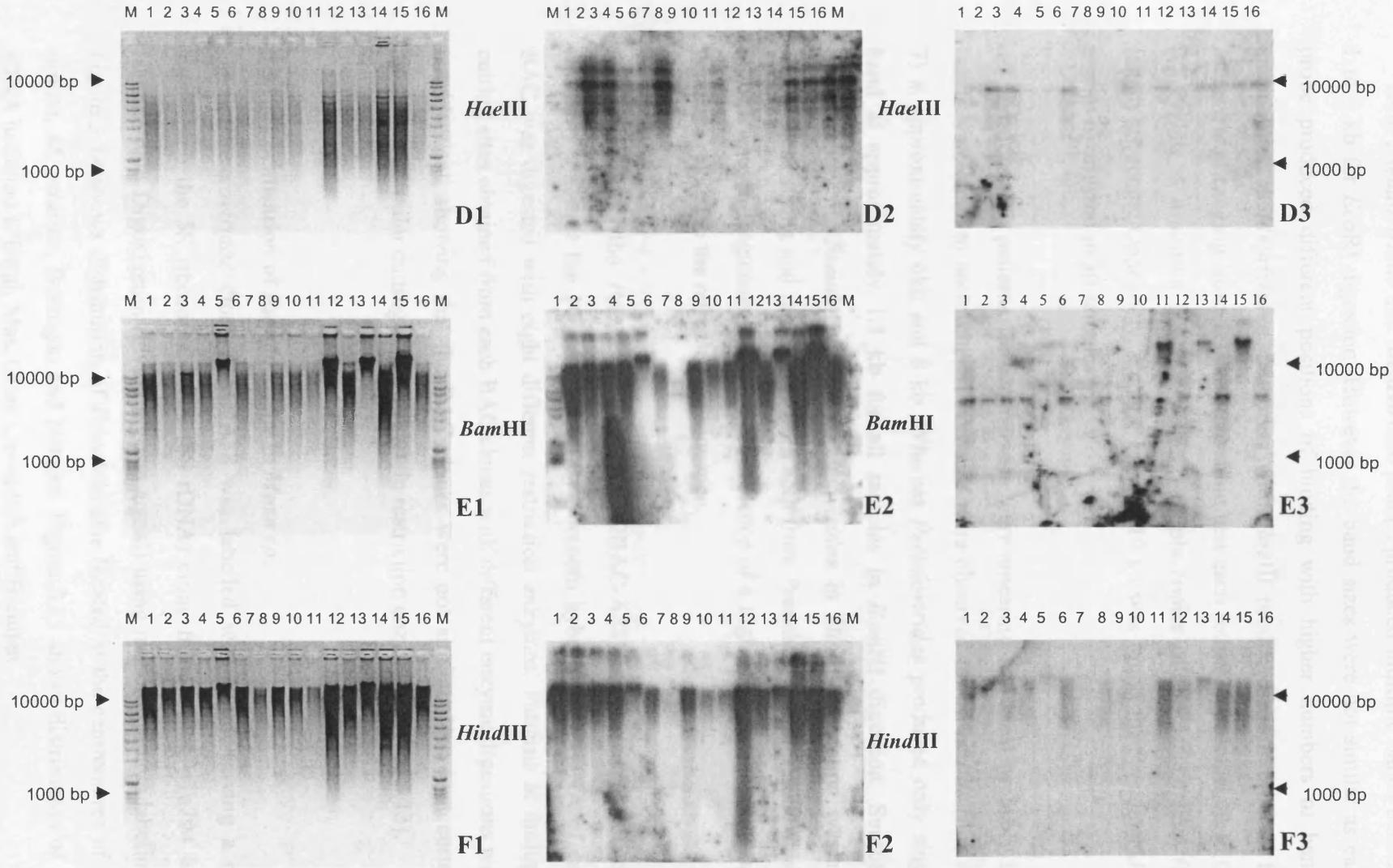


Figure 3.12. Continued.

Both *Pseudoviridae* and *Metaviridae* probes produced multiple bands between 1 to 3 kb for *Eco*RI digestion. However the band sizes were not similar as each probe produced different positions of binding with higher numbers of bands encountered in *Metaviridae*. Digestion with *Hae*III produced discrete bands for *Metaviridae* ranging from 1-10kb suggesting that each retrotransposon flanked the RT region at a similar distance within multiple copies of the *Metaviridae* full length sequence. Oppositely, a single band at 10 kb was found for *Pseudoviridae* probes hybridized in all samples.

Hybridization patterns for *Metaviridae* were smearing and most probably the bands were close to each other. Two bands were observed for *M. textilis* (line no. 7) at approximately 6kb and 8 kb. Whereas *Pseudoviridae* produced only single band at approximately 1.3 kb for all samples in *Bam*HI digestion. Southern hybridization for *Pseudoviridae* and *Metaviridae* in *Hin*dIII digestion was not clear. These strong and faint bands from both from *Pseudoviridae* and *Metaviridae* retrotransposon fragments indicated the presence of a high copy numbers in *Musa* genomes related to the retroelement sequence.

Furthermore, the *Pseudoviridae* probe (2BAC-A38) was used to identify similarity amongst the 10 BAC clones by Southern hybridization. DNA of each BAC was digested with eight different restriction enzymes. Patterns at multiple cutting sites obtained from each BAC clones with different enzyme digestions were not identical, showing that the BAC clones were not similar. However, certain clones have similar cutting size with certain restriction enzymes (Figure 3.13).

3.3.1.4 Localisation of *Pseudoviridae* in *Musa sp.*

Pseudoviridae clone, BAC2A38 was labelled with biotin (giving a red signal) and the 5S ribosomal DNA (5S rDNA) clone from wheat, pTa794 was labelled with Digoxigenin (giving a green signal) using random primers labelling. Figure 3.14 shows distribution of *Pseudoviridae* located at the centromeres of *M. ornata*, *M. velutina*, Berangan and Nangka. Figure 3.15 shows distribution of 5S rDNA in Obino L'Ewai, Mas, Giant Cavendish and Butuhan.

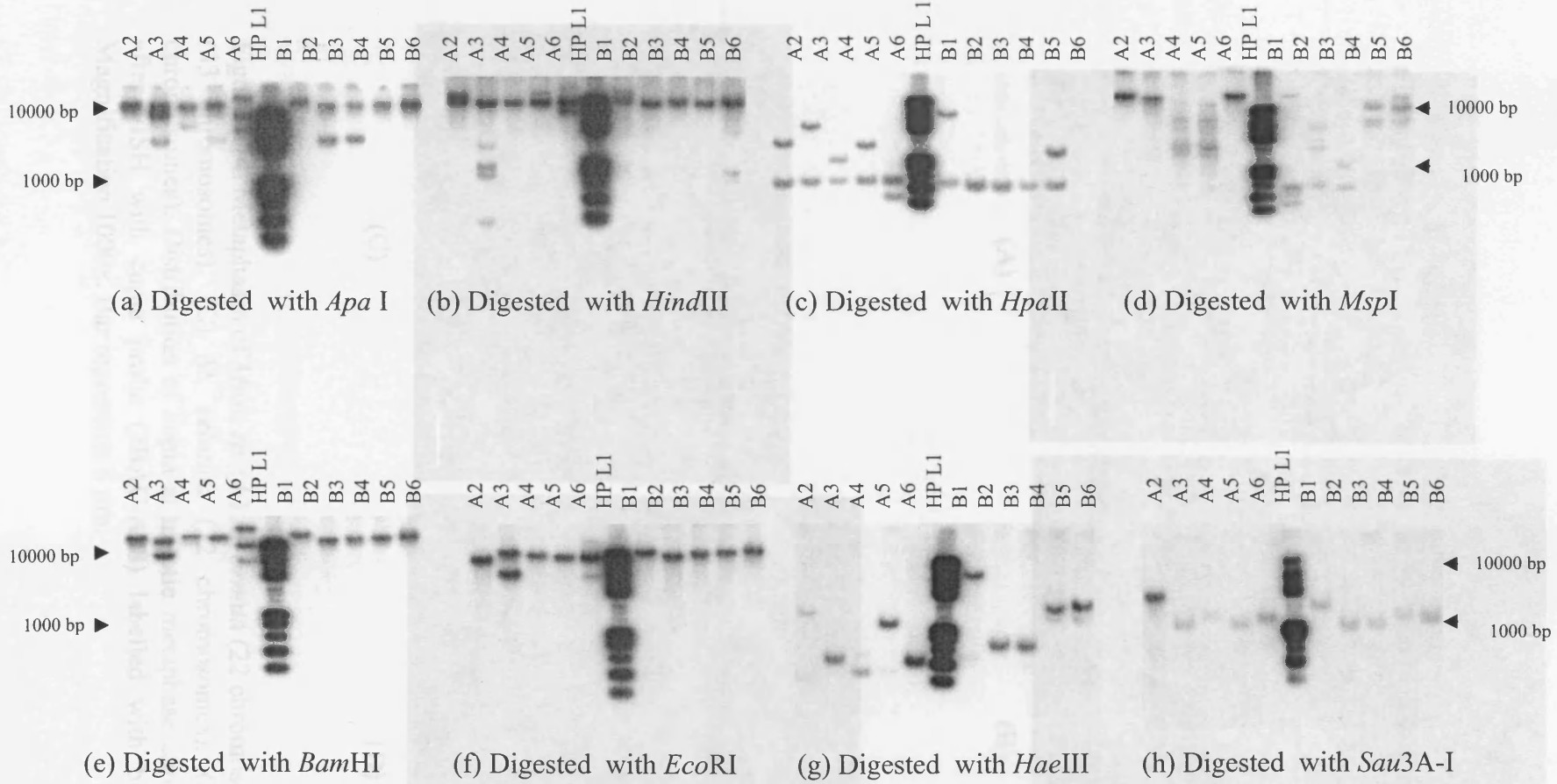


Figure 3.13. Organisation of RT genes in 10 BAC clones originated from MA4 library of *M. acuminata* 'Calcutta-4' (MA4_2A02, MA4_2A03, MA4_2A04, MA4_2A5, MA4_2A06) and MBP library of *M. balbisiana* 'Klutuk Wulung' (MBP_1B01, MBP_1B02, MBP_1B03, MBP_1B04. MBP_1B05) digested with eight different restriction enzymes and hybridized with *Pseudoviridae* fragment from clone 2BAC-A38. HyperLadder 1 (HPL1) was used as marker.

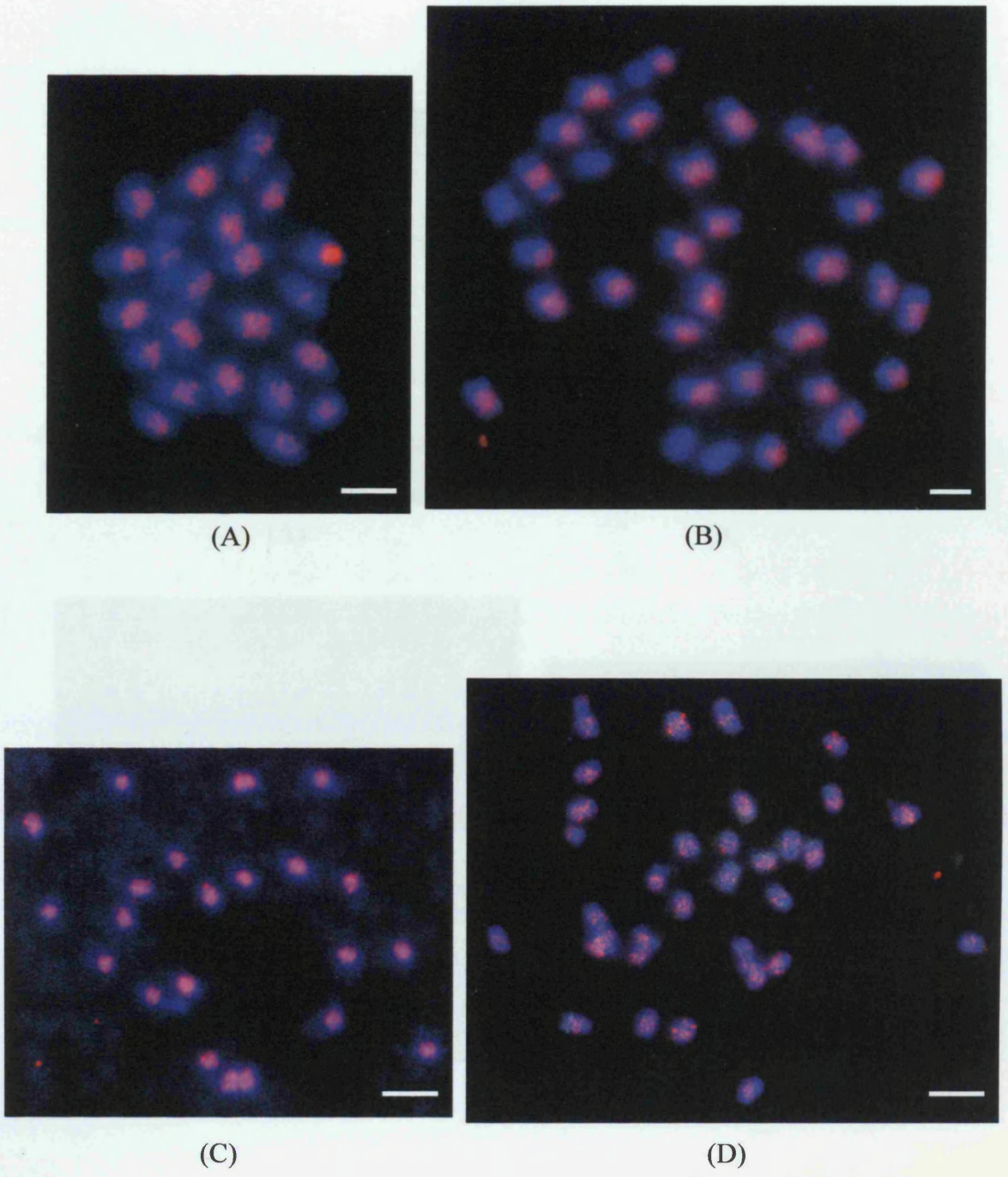


Figure 3.14. Metaphases of *Musa sp.* (A) *M. ornata* (22 chromosomes), (B) Nangka (33 chromosomes), (C) *M. velutina* (22 chromosomes), (D) Berangan (33 chromosomes). Distribution of copia at mitotic metaphase chromosome of *Musa* after FISH with copia probe (2BAC-A38) labelled with biotin (red signal). Magnification 1000x. Bar represents 5 μm.

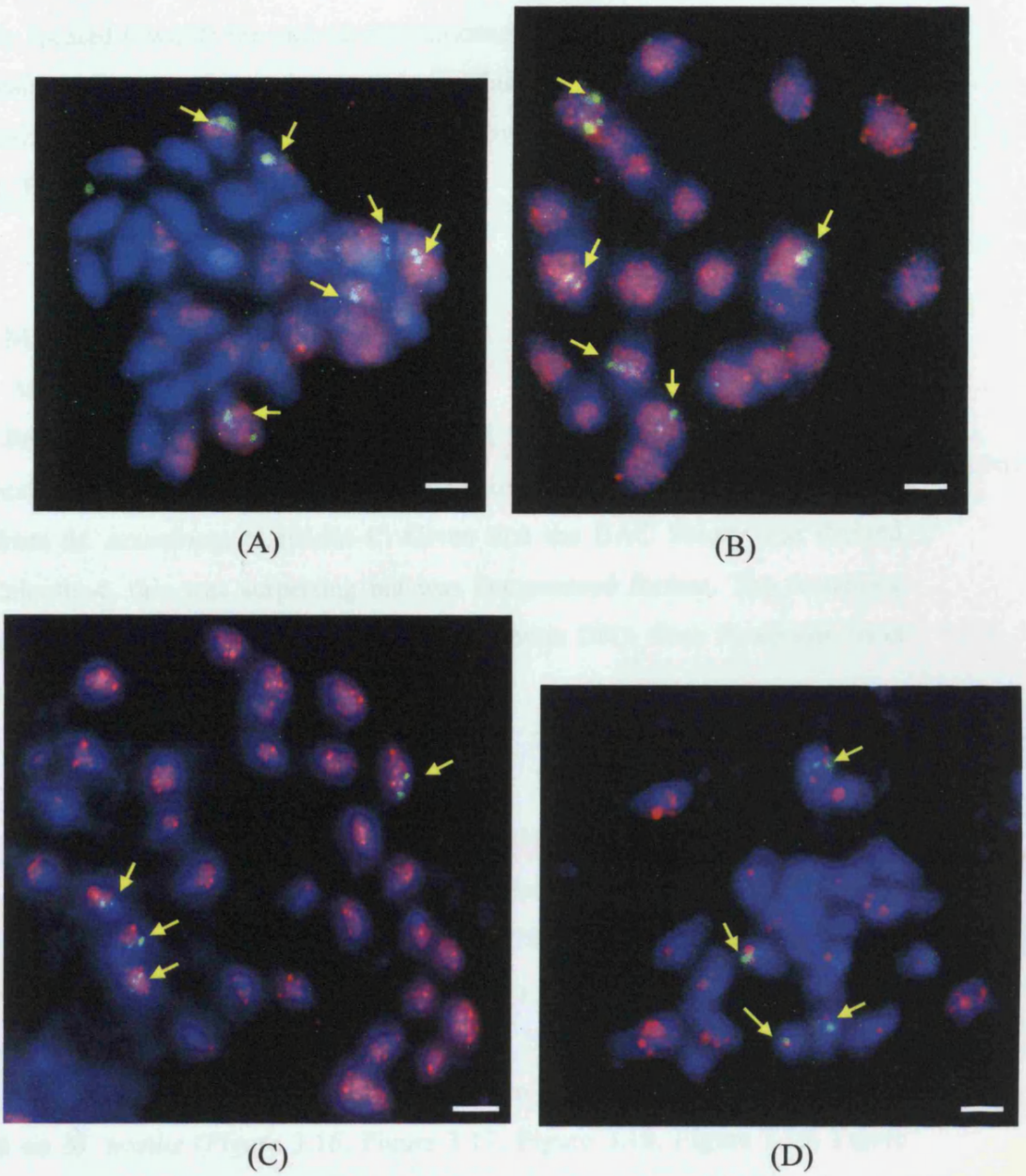


Figure 3.15. Metaphases of *Musa sp.* (A) Obino L'Ewai (33 chromosomes), (B) Mas (22 chromosomes), (C) Giant Cavendish (33 chromosomes) and (D) Butuhan (22 chromosomes). Distribution copia and 5S rDNA on mitotic metaphase chromosome of *Musa* after FISH with labelled probes. Clone 2BAC-A38 was labelled with biotin (red signal) and pta794 was labelled with DIG (green signal) marked by arrows. pta794 is a repeated unit of 5S rDNA-gene derived from wheat with a sequence size at 410 bp (Gerlach and Dyer, 1980) and used as a control DNA probe. Magnification 1000x. Bar represents 5 µm.

The number of pTa794 signals (green) for 5S rDNA was variable and normally located towards the end of chromosome arms (Figure 3.15). There were six signals in Obino L'Ewai, five signals in Mas and four signals in both Giant Cavendish and Butuhan. Distribution of *Pseudoviridae* was less in Butuhan and Obino L'Ewai as compared to other samples.

3.3.2 Microsatellite

3.3.2.1 Microsatellite primers

Primers MuG9-SSR1, MuG9-SSR2 and MuG9-SSR14 did not give any amplification after optimization (gradient annealing temperature) by using genomic DNA from *M. acuminata* 'Calcutta-4'. Given that the BAC library was derived from Calcutta-4, this was surprising but was not pursued further. The remaining working primers were further used to amplify genomic DNA from 56 accessions in *Musa* spp.

3.3.2.2 PCR-based polymorphism on genomic DNA

Primers MuH9-SSR1, MuH9-SSR5, MuH9-SSR6 and MuH9-SSR12 produced bands at approximately 150 bp, 180bp, 200 bp, 220 bp respectively with various polymorphisms visible from single or two alleles in each genomic DNA. However primer MuG9, which was imperfect SSR produced from single to multiple alleles ranging approximately at 150bp to 300 bp. None of these primers worked on *M. textilis* (Figure 3.16, Figure 3.17, Figure 3.18, Figure 3.19, Figure 3.20) suggesting a null allele and at least one of each primer binding site was missing.

Primers MuH9-SSR1, MuH9-SSR5, MuH9-SSR6 and MuH9-SSR12 are flanking the motifs $(CT)_{11}$, $(TCC)_8$, $(CTT)_3$ and $(CTTT)_6$ respectively. Primers MuH9-SSR1 produced a single thick or light band at approximately 150 bp. Although this primer produced only single products, their sizes were not the same. *Musa* accessions with diploid AA varied in their size ranging from approximately 140 to 150 bp. A diploid *M. balbisiana* 'Klutuk Wulung' has a strong band but similar band size as compared to other *M. balbisiana* species.

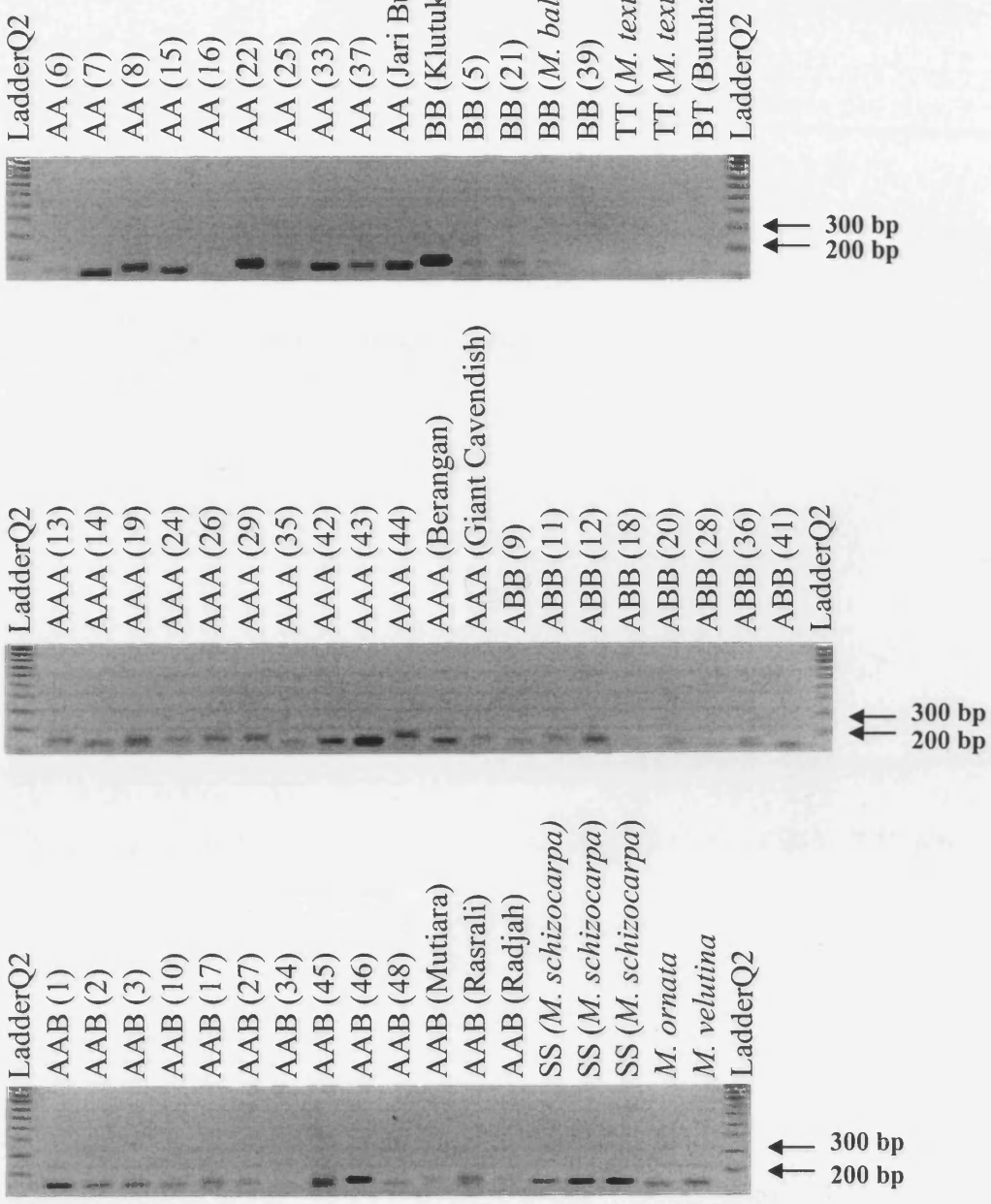


Figure 3.16. PCR products by using microsatellite primers MuH9-SSR1, repeat $(CTT)_{11}$ on genomic DNA of *Musa sp.* YorkBio Q-Step2 Ladder (LadderQ2) were located on both side of the images.

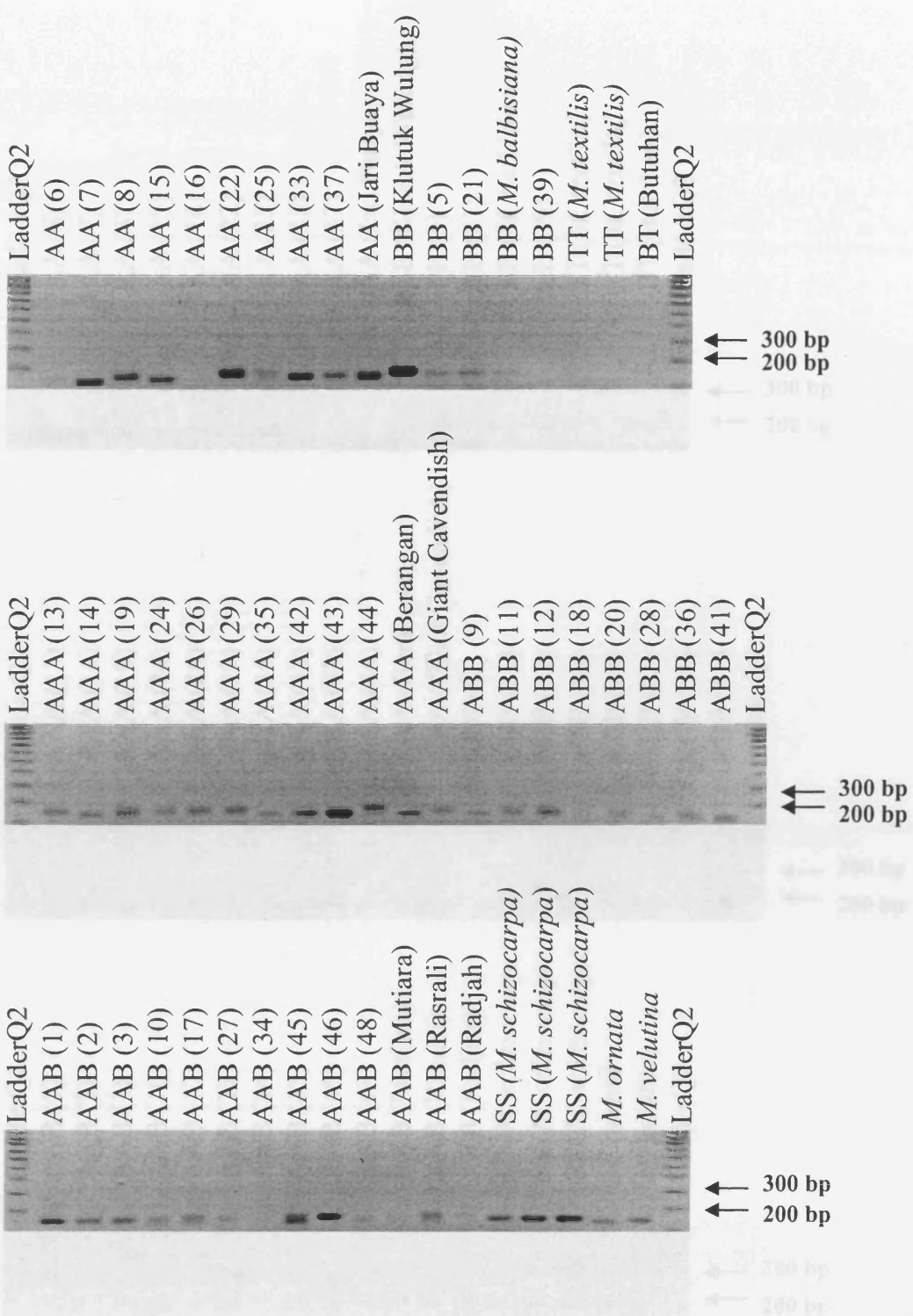


Figure 3.16. PCR products by using microsatellite primers MuH9-SSR1, repeat $(CTT)_{11}$ on genomic DNA of *Musa sp.* YorkBio Q-Step2 Ladder (LadderQ2) were located on both side of the images. See appendix for the numbers on the gel.

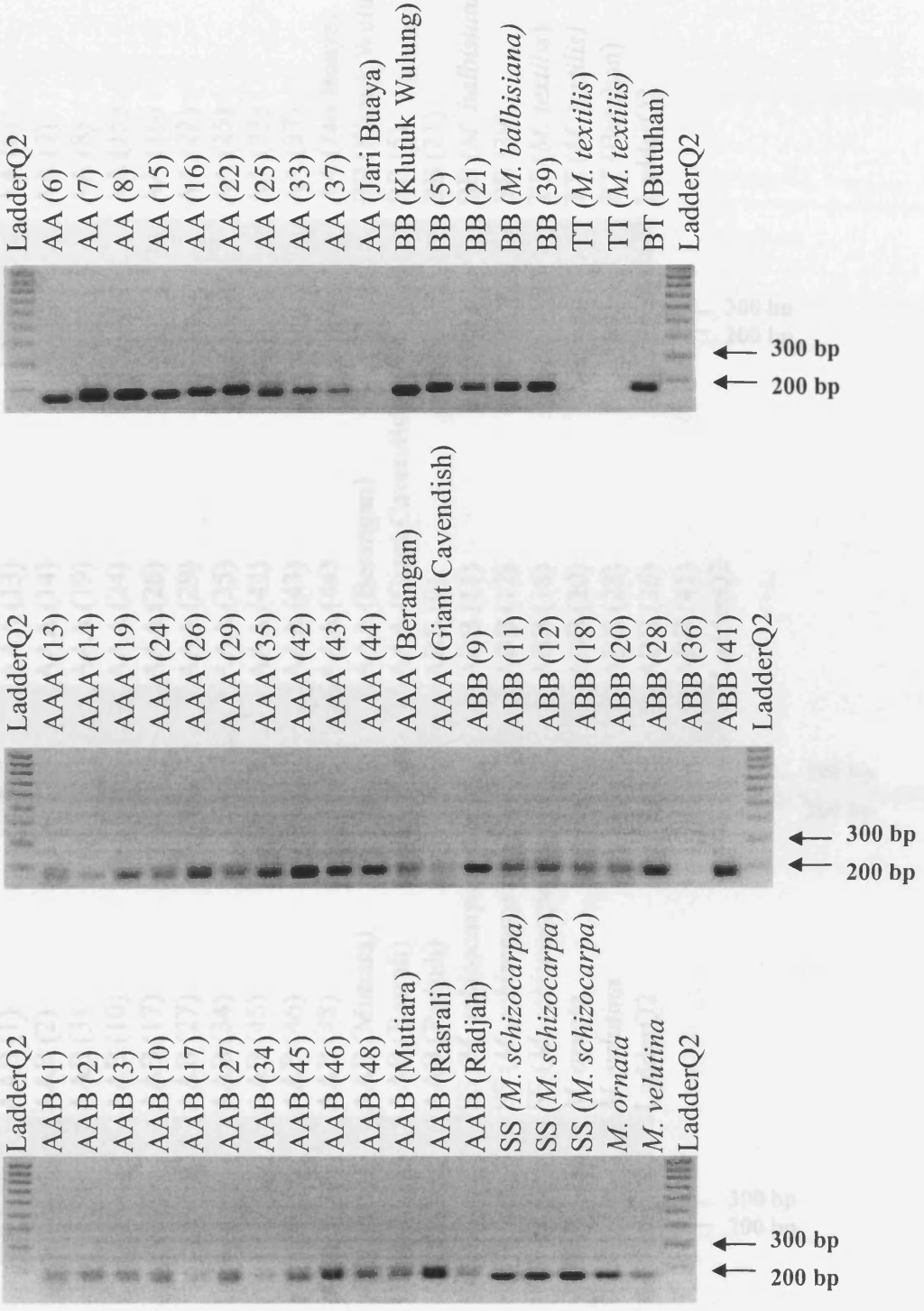


Figure 3.17. PCR products by using microsatellite primers MuH9-SSR5, repeat $(TCC)_8$ on genomic DNA of *Musa sp.* YorkBio Q-Step2 Ladder (LadderQ2) were located on both side of the images. See appendix for the numbers on the gel.

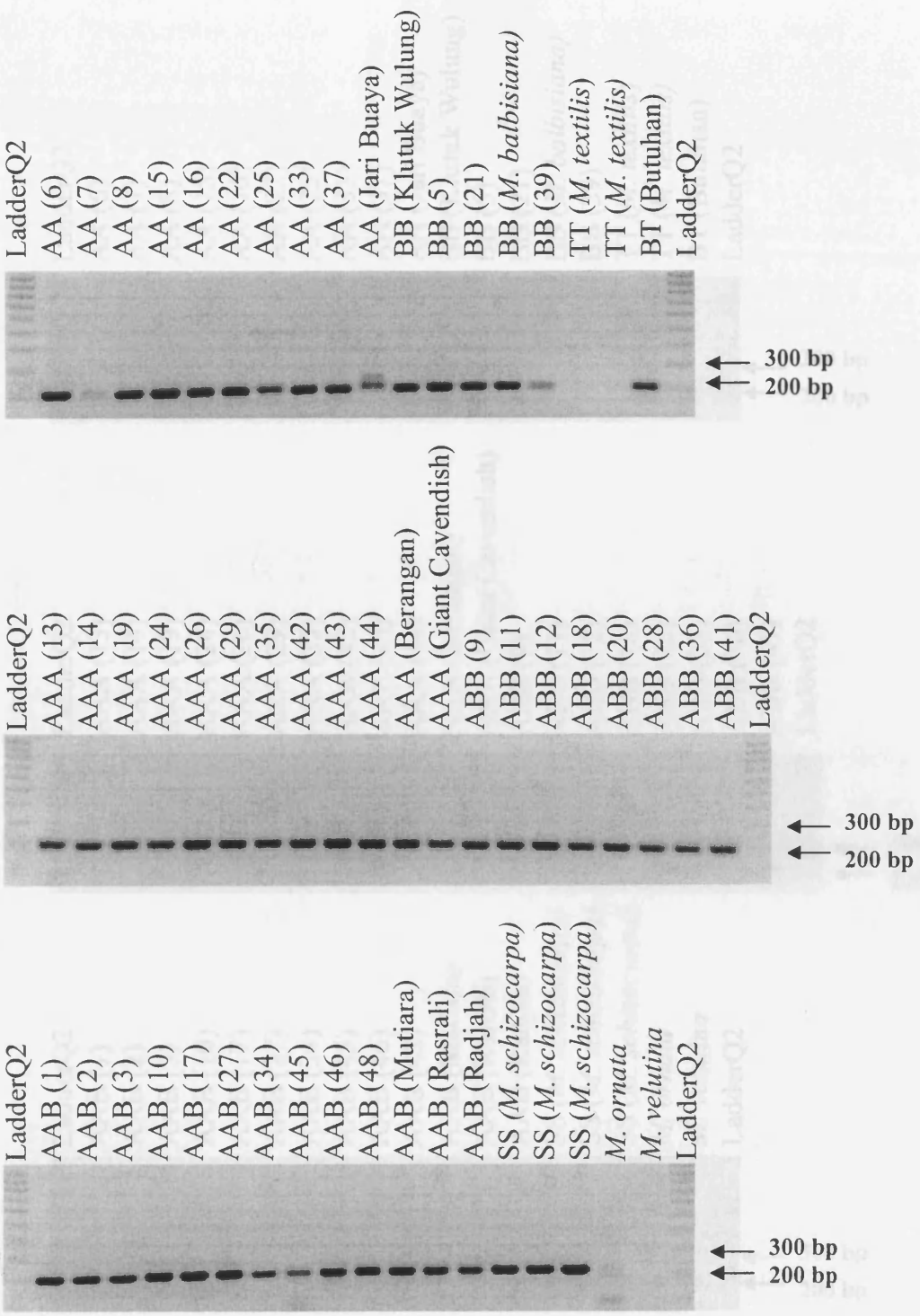


Figure 3.18. PCR products by using microsatellite primers MuH9-SSR6, repeat $(CTT)_3$ on genomic DNA of *Musa sp.* YorkBio Q-Step2 Ladder (LadderQ2) were located on both side of the images. See appendix for the numbers on the gel.

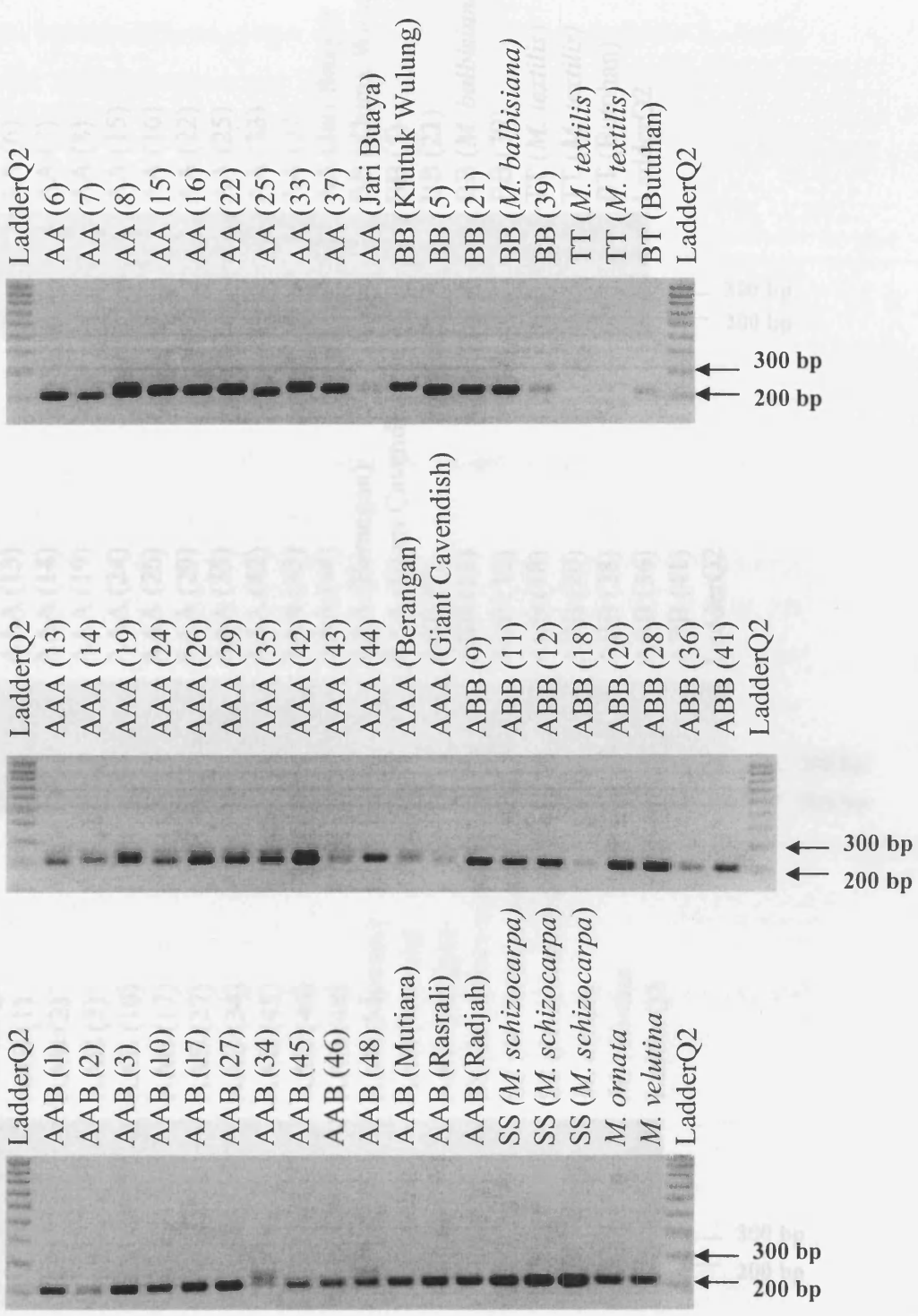


Figure 3.19. PCR products by using microsatellite primers MuH9-SSR12, repeat $(CTTT)_6$ on genomic DNA of *Musa sp.* YorkBio Q-Step2 Ladder (LadderQ2) were located on both side of the images. See appendix for the numbers on the gel.

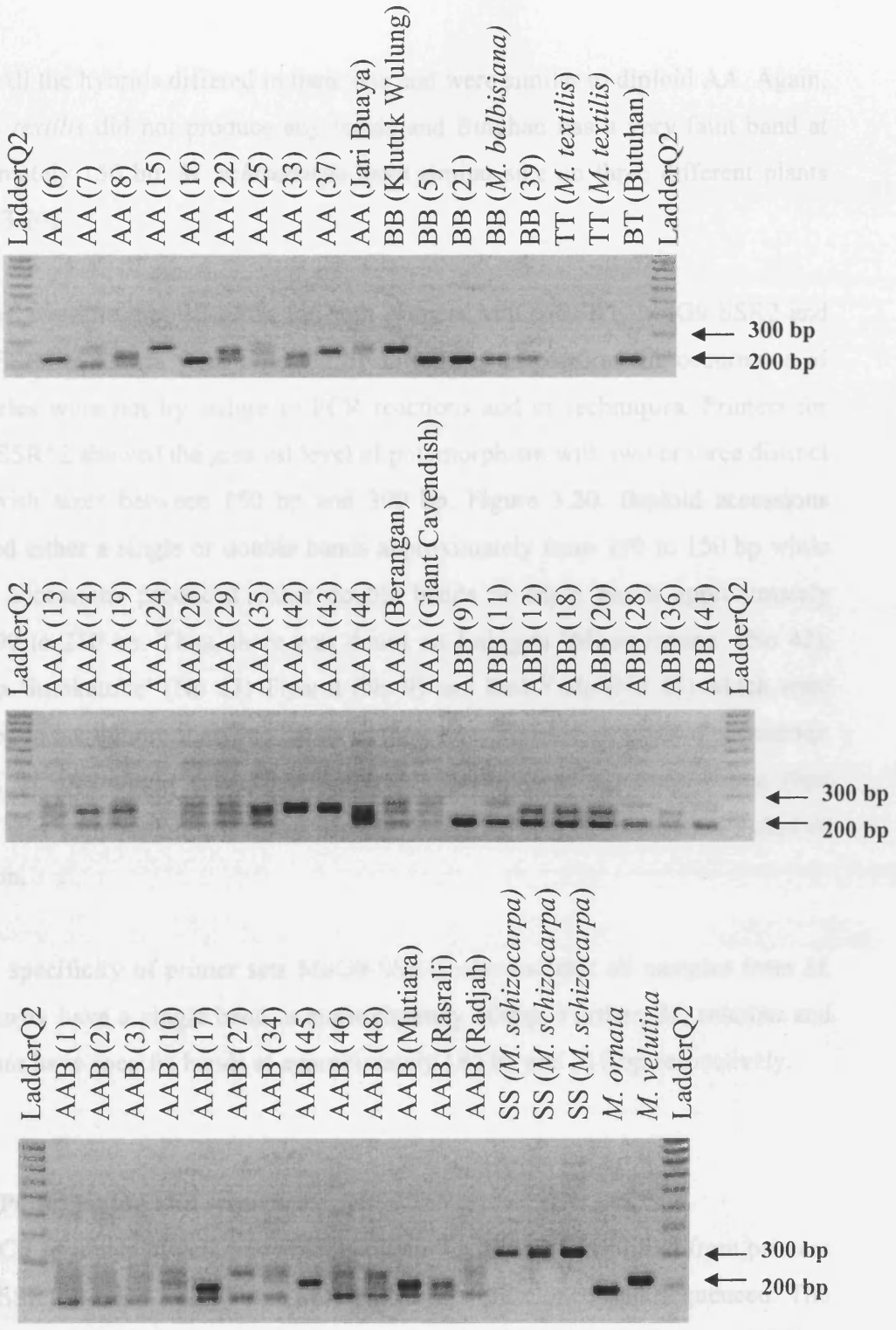


Figure 3.20. PCR products by using microsatellite primers MuG9-SSR12, repeat (GA)14GC(GA) on genomic DNA of *Musa sp.* YorkBio Q-Step2 Ladder (LadderQ2) were located on both side of the images. See appendix for the numbers on the gel.

All the hybrids differed in their size and were similar to diploid AA. Again, both *M. textilis* did not produce any bands and Butuhan has a very faint band at approximately 150 bp. *M. schizocarpa* have similar size on three different plants (Figure 3.20).

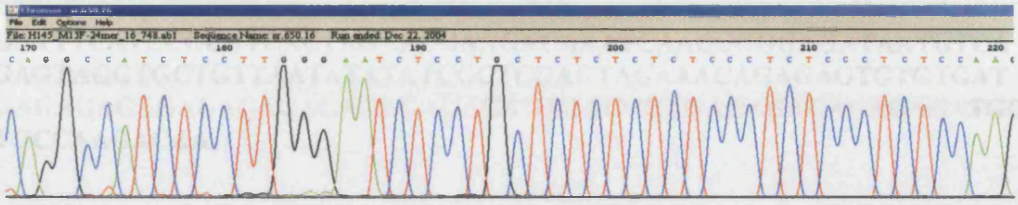
There were no amplification for both primers MuG9-SSR1, MuG9-SSR2 and MuG9-SSR14. Assays were repeated by few times to confirm the occurrence of null alleles were not by failure in PCR reactions and or techniques. Primers for MuG9-SSR12 showed the greatest level of polymorphism with two or three distinct bands with sizes between 150 bp and 300 bp. Figure 3.20. Diploid accessions produced either a single or double bands approximately from 190 to 150 bp while triploid accessions produced either double bands or triple bands approximately from 190 to 250 bp. Thus, there was doubt on Lujugira 'Mbwazirume' (No 42), Lujugira 'Intokatoke' (No 43) Tiparot (No 9) and Red Yade (No 45) which were supposed to have more than one bands as they were claimed as triploid accessions. Further, *M. acuminata* subspecies *banksii* 'Paliama' (no 25) from Papua New Guinea was supposed not to have triploid bands as it was claimed as a diploid accession.

The specificity of primer sets MuG9-SSR12 showed that all samples from *M. schizocarpa* have a single band at approximately 300bp. Further, *M. velutina* and *M. ornata* have specific bands at approximately 180 bp and 210 bp respectively.

3.3.3. PCR cloning and sequence

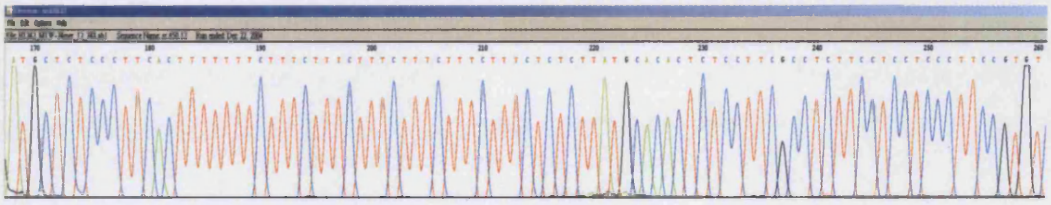
PCR products of genomic *M. acuminata* 'Calcutta-4' amplified from primers MuH9-SSR1, MuH9-SSR12 and MuG9-SSR12 were cloned and sequenced. The sequence confirmed clearly the presence of repeat and $(GA)_{14}GC(GA)_8$ and$(CT)_{11}$, $(CTTT)_6$ respectively (Figure 3.21, Figure 3.22).

TGCGTTTCCAGTGGATTATGATAAGTCTTCGGAGCCGAAGGTTTCTGATGGGATCA
GATCGAAAGGTGGAATCCTGCTGCCACCTCAGGCCATCTCCTTGGGAATCTCCTGC
TCTCTCTCTCCCTCTCCCTCTCCAAGTGTCTA



(A) MuH9-SSR1 (repeat $(CT)_{11}$)

AACTAGGCAGCAAACCTTCGGGTCTGAAATTGTTATCAAATTTTCAAATTTTTGAC
AGCTGTGACAATTCTTCTCTCACCTATGATGCTCTCCCTTCACTTTTTTTCTTTCTT
TCTTTCTTTCTTTCTTTCTCTCTTATGCACACTCTCCTTCGCCTCTTCCTCCTCCCT
TCCGTGTCATTGCTGCTGCCGATTCCTTCTCCTCCTCTTCTTTGCCTTTGGGAAA



(B) MuH9-SSR12 (repeat $(CTTT)_6$)

Figure 3.21. Sequences of PCR product amplified by using microsatellite primers (A) MuH9-SSR1 ($(CT)_{11}$) and (B) MuH9-SSR12 ($(CTTT)_6$) on genomic DNA of obtained from *M. acuminata* 'Calcutta-4' .

CGATTTCATCCGCATCACTAGAACGAAGACAAATCAACCGGGAGATAATGTCA
TGAGTAGGTGCTGTTAATATATATCGCTCGAGTAGAAACAGAGAGTGTGTGAT
AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCGAGAGAGAGAGAGAGATGC
ATGCCAAGACAA

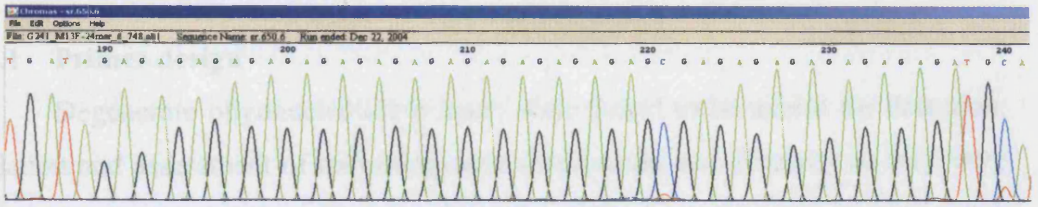


Figure 3.22. Sequence of PCR products amplified by using microsatellite primers MuG9-SSR12 ($(GA)_{14}GC(GA)_8$) on genomic DNA of obtained from *M. acuminata* 'Calcutta-4'.

3.4 Discussion

Microsatellites and retroelements are both abundant repetitive DNA sequence motifs that are dispersed over much of the nuclear genome of most plants. Both evolve rapidly, but largely by different mechanisms (uneven crossing-over or retro-transposition via an RNA intermediate). Hence, they are independent DNA markers which can be used to measure diversity and perhaps the diversity data can be used to infer relationships. The results above present a limited study carried out to assess the nature and value of these two marker systems and to learn about the BAC sequence data and BAC library which has become available in 2004/05 as a genomics resource.

3.4.1 Primer design

Degenerate oligonucleotide primers were found to be useful for detection, isolation and assessment of retrotransposon distribution and diversity as they were present in truncated and multi-copy elements. It has been reported in many plant species that retrotransposons belong to many families and/or subgroups (Flavell *et al.,* 1992; Hirochika *et al.*, 1992; Kubis *et al.,* 1998; Suoniemi *et al.,* 1998) with different sizes and hence PCR products might vary. This indicates that the nature of retrotransposons is related to integration, with big family clusters and/or subgroups that share a common sequence including over the RT domains.

Heterogeneity in *Metaviridae* retrotransposons resulted in difficulties for designing universal degenerate oligonucleotide primers which lead to variable PCR products or no amplification. Multiple bands observed for *Metaviridae* confirmed the low specificity of the degenerate primers, but also probably arose from truncation of sequences and insertion of the elements into pre-existing elements. As a comparison, degeneracy of the *Metaviridae* primers was higher than *Pseudoviridae* primers.

Direct sequencing of PCR products was not used because the products were quite polymorphic. Further, it was essential to identify primer sequences to know that products were from primer pairs rather than single primers (which could occur either because elements inserted into each other in opposite orientation or because the primer degeneracy and low annealing temperatures amplified non-retroelement sequences in the genome).

3.4.2 Diversity and organisation of LTR-Retroelements

Retrotransposons were a major fraction of the plant genome, and it was reported that spontaneous mutation occurred in *Drosophila* (Green, 1988) and maize (Varogana *et al.*, 1992). DNA fragment replication and transposition activities resulted increase of retrotransposons copy number in plant genomes and they were identified as repetitive sequences or insertions elements in cloned DNA or genes (Grandbastien, 1992). Sequence heterogeneity was variable amongst different related genomes (species) with an extreme case in *Solanum tuberosum* which has hundreds copies of these elements but not corresponding to any division of genomes (Flavell *et al.*, 1992). The sequence heterogeneity might be influenced by their copy numbers (Charlesworth, 1986).

Multiple bands in PCR product of *Metaviridae* primers showed this group has many targeted genes and might be truncated as compared to PCR product from *Pseudoviridae* primers. Thus, only dominant bands were selected for confirmation by sequencing. All clone amplified by g*ypsy*-like primers have either stop codons or frame-shifts in their RT sequences. Thus only partial *gypsy*-like elements were homologous to all the fragments cloned.

The occurrence might be related to rearrangement copies and insertion and/or translocation occurred in plant species or might be false amplification products by the degenerate primers. According to Gao *et al.*, (2004), *Metaviridae* has four times more elements and relatively high rates multiple bands to be amplified as compared to *copia*-like in rice, although whether this is true in other species remains to be shown.

Truncated sequences in *Metaviridae* family make them difficult to be evaluated and their higher degeneracy means they are less studied as compared to *Copia*-like family, as reported in conifers (Friesen *et al.*, 2001). There is no doubt that the grouping might have an ambiguity in certain samples as the bands are not clear and the experiment has to be repeated to increase the sample of *Metaviridae* sequences for more confirmation.

Although RT fragments in each clone for *Pseudoviridae* were single band, they differed in their sizes but were homologous to *Ty1-copia*-like retrotransposon sequences of W1 *Ty1-copia*-like retrotransposon, N18 *Ty1-copia* and M6 *Ty1-copia* isolated by Teo *et al.,* (2002). The activation and control mechanism for both ancient and recent retrotransposons is significantly affected by the stress environment of the plant, but the exact nature and timing of amplification bursts is uncertain, as each species reacts differently and the elements themselves consist of many families (clusters in the sequence tree) which apparently behave differently in the various species (genomes). Thus, retroelement phylogenetic study requires more individual accessions and species for further classification as each taxon has dispersed, and independent sequences, corresponding to different sub-clusters of the retrotransposons.

Tree construction, both from deduced amino acid and nucleotide RT sequences, was important to clarify the relationships among the *copia*-like retrotransposon clusters from the various *Musa* species. Distances in each clade correspond to divergence of the retrotransposon which has actively occurred throughout evolution. The position and distances showed occurrence of different activity in retrotransposons inter and/or intra species.

RT occurred in many *copia*-like families and the motif was not grouped together in the sequence analysis (Figure 3.8, Figure 3.9, Figure 3.11). Different *copia*-like families were reported in cereals (Gribbon *et al.,* 1999), *Solanaceae* (Costa *et al.,* 1999), gymnosperms (Stuart-Rogers and Flavell, 2001), *Pisum* (Pearce *et al.,* 2000), *Alstroemeria* (Kuipers *et al.*, 1998), *Silene latifolia* (Matsunaga *et al.*, 2002) and *Olea* (Stergiou *et al.,* 2002) where some of them share common fragments.

An evolutionary hypothesis would be that the elements are present in a current species through vertical transfer where the sequences are transferred from common ancestors or parents to their offspring. However, RT in retrotransposons demonstrated ambiguities and unspecific trends based on their multiple sequence alignments.

This might be explained because retrotransposons can be activated both under abiotic and/or biotic stress conditions such as cold activation in maize retrotransposons (Steward *et al.,* 2002), salinity and wounding (Kimura *et al.,* 2001) on oat climatic factors in wild barley retrotransposons (Kalendar *et al.,* 2000), tobacco tissue culture (Hirochika *et al.*, 1993) and virus infection in tobacco retrotransposons (Hirochika *et al.,* 1995). In the rice genome, active retrotransposons at transcriptional level was found under stress conditions (Hirochika *et al.,* 1997).

There are possibilities that one species can generate many members of retrotransposons under different conditions. This phenomenon can be explained through a mechanism called horizontal transfer where transposable elements can move around without bounds directly between plants (different lineages or even species). Horizontal transfer is proposed when high similarity between sequences from related or distant species is observed and the occurrences were not consistent with identity-by-descent (vertical transfer) when supported by phylogenetic trees (Capy *et al.*, 1994). This reason was supported by the horizontal transferred of P-elements in *Drosophila* corresponding to *copia*-like elements which originated from one species (Daniels *et al.,* 1990). A mechanism of transferring was described by Sugimoto *et al.* (1994) through virus infection that transferred transposable elements from maize to rice.

Both sequence trees revealed that RT fragments in *Musa* were not corresponding to the clusters of *copia*-like retroelements. High heterogeneity was found both at deduced amino acid and nucleotide sequences derived from single species. Although the phylogenetic trees were not coincident with individual species or genomic constitutions, it was clearly shown that four different groups of RT were observed related to genetic diversity in *Musa.*

In BAC clones, the single band size observed in all 10 BACs amplified with *Pseudoviridae* primers showed that *Pseudoviridae* retrotransposons are present in all BACs and hence there is at least one copy per 100kb (the approximate average insert size), and thus very high copy number. If the elements were purely randomly distributed, then a statistical analysis could predict the average distance between the elements, but there is as yet no support for the assumption of random distribution. Only three BAC clones were amplified by *Metaviridae* primers, showing dispersion of the *Metaviridae* retrotransposons in *Musa*, as represented by the families amplified by these particular primers, was much lower copy number than the *copia*-like elements. The RT sequences of *Pseudoviridae* retroelements matched with *M. acuminata* BAC clone MuG9 sequence as retrotransposon-like protein sequence (Aert *et al.,* 2004). This showed, as expected, that the *copia* sequence is present in multiple, highly diverged, copies in banana genomes.

A clear picture of localization and copy number of a retroelement family is shown in FISH (Figure 3.15) as the *Pseudoviridae* was found to be dispersed in all chromosomes of *Musa* spp., especially with some clusters at the centromeric regions. Thus, it is difficult to distinguish between A and B chromosomes and/or between samples studied. The same distribution was reported in clone *Radka*8 that is related to the *monkey* retrotransposon sequence (Valarik *et al.,* 2002). The 5S rDNA probe showed clearly distribution of ribosomal DNA in different *Musa* accessions. According to Dolezelova *et al.* (1998), *Musa* accessions are distinct in 5S sites and localised on two pairs of chromosomes with an exception in Mas where only five sites of 5S rDNA was observed; the results also agree with Bartos *et al.* (2005).

Both *Pseudoviridae* and *Metaviridae* retroelements might be used as molecular marker to study *Musa* diversity through PCR-based techniques for evaluating relationships and quantifying biodiversity. However, for a specific molecular marker amongst the species it remains unclear as the retroelements might be affected by environment such as abiotic and biotic stress.

The nature of retrotransposons in replicating and inserting at multiple sites into the genome supported with their highly copy numbers indicate that individual transposable elements might serve as markers for diversity study by analysis of insertional polymorphisms of single elements.

An excellent example of the use of the retroelement makers is given above with the group of cultivars selected from spontaneous variation in tissue culture (somaclonal variation) and their ancestral lines (Figure 3.2). Intan and Mutiara are Malaysian commercial triploid banana cultivars with constitution of AAA and AAB respectively. They were improved from Berangan and Rastali respectively based on high yield and good characteristics. The results using *Metaviridae* particularly, consistent with *Pseudoviridae* results, shows that the cultivars have much closer similarities to the ancestral accessions from which they were derived.

The specific taxonomic classification of *Musa* accessions and their genome arrangement, especially in hybrid species remains unclear, as there is occurrence of heterogeneity and diversity amongst both A and B genomes. The frequency and periodic occurrence of retroelement transposition also should be considered as these events might be a limitation in quantifying their insertion into, or modification of expression of, functional genes. Further, environmental condition affected the retrotransposons activities by technical aspects and biological reasons showing that the retrotransposons are not reliable to be used as a universal marker system, but are particularly useful in some cultivar analysis and mutation or somaclonal variation breeding programmes.

3.4.3 Fluorescent *in situ* hybridization.

High cytoplasm in *Musa* cells gave an obstacle even though pepsin was used to reduce the cytoplasm. It was observed that the *Pseudoviridae* probe from clone 2BAC-2A38 dispersed at different location in different variety. Six signals of 5S rDNA were found in Obino L'Ewai and five signals in Mas, whereas, four signals of 5S rDNA were located in both Giant Cavendish and Butuhan.

Musa chromosomes are small relative to other species. Thus only *Pseudoviridae* was used to study the physical distribution of the retroelements representing the LTR for Fluorescent in situ hybridization (FISH). The greatest obstacle to FISH was in slide preparations for the *in situ* in getting good and countable metaphase chromosome spreads. Further, the chromosomes were hindered by high cytoplasm in the cells.

Vigorous and healthy root tips good for chromosome spreads were obtained during summer and spring period (April to August). It was observed that *Musa* variety with B genome needs longer time for their incubation period for cell wall digestion. Further, different varieties required a different time for the incubation. Only good slides with high metaphase spreads were selected for FISH.

3.4.4 Genomic polymorphism based on SSR markers

All PCR products of the MuH9 primers were showed a single band on 2% agarose gel but their sizes and band thickness was not the same but specific to species and indicated polymorphism. Detailed analysis would require higher resolution methods. For SSRs, high levels of polymorphism occurred in hexaploid bread wheat (*Triticum estivum)* although with a degree of amplification of several artefact fragments from a single locus (Stephenson *et al.,* 1998). In cereals, most SSR primers are specific to one genome or species (Devos and Gale, 2000) although in palms, identical primers will amplify homologous loci across most of the family (Billotte *et al.,* 2001).

It is strongly believed that using 6% agarose gel or acrylamide gel, which is laborious and time consuming but gives high resolution of fragment size separation, would resolve some polymorphisms and separate the broad bands. Further, they can be improved by fluorescent labelled primers and sequencing machine with direct measurement of fragment size on each sample. Specificity of SSR primers was showed by MuG9-SSR1 and MuH9-SSR1 where both primers targeted on the same repeat of 'CT' but only primers MuH9-SSR1 was amplified as both of them have different flanking region.

Remarkably, primers for MuG9-SSR12 showed a much higher level of polymorphism than the MuH9 derived primers, with two or three distinct bands with sizes between 150 bp and 300 bp. Primers MuG9-SSR12 showed multi alleles for *Musa* hybrids. Their bands size and patterns might be possessing to proportion alleles of each hybrid and might be used for SSR markers.

Information about ploidy levels or genome constitution in some *Musa* samples was not entirely consistent with these SSR results: sample Lujugira 'Mbwazirume' (no 42), Lujugira 'Intokatoke' (No 43) Tiparot (No 9) *M. acuminata* subspecies *banksii* 'Paliama' (No 25) and Red Yade (No 45). Overall, the results agreed with Crouch *et al.* (1996) that restriction fragment length polymorphism (RFLP) and particularly random amplified polymorphism (RAPD) used for germplasm characterization in *Musa* are no longer methods of choice for routine application in *Musa* characterizations.

Wild accessions of *M. acuminata* were classified by SSR and RFLP molecular markers to consist of four major groups (sub-species) i.e. *malaccensis, banksii, zebrina* and *burmanicoides* (Lanaud *et al.,* 1992; Grapin *et al.,* 1998). As reported in CIRAD germplasm, *M. balbisiana* was grouped into three sub-species known as Type-1, Type-3 and Type-4.

Based on the five SSR primers designed, many polymorphic loci were detected especially in *M. acuminata* and the results supported them in having four sub-groups i.e. *malaccensis, banksii, zebrina* and *burmanicoides* (Carreel *et al.,* 1994; Grapin *et al.,* 1998). This suggests that high activities such as mutations and transitions occur in *M. acuminata* corresponding to wide genetic base and creating variability in their hybrids. In this work only two groups were observed based on five SSR primers tested. The SSR primers showed clearly that *M. balbisiana* sub-species type-4 'Klutuk wulung' was distinguished from the other *M. balbisiana* samples, even to Batu, which was grouped under the same sub-species. This suggests that *M. balbisiana* species has a narrow genetic base and more reliable markers are required to discriminate amongst them precisely.

SSR primers in this work cannot distinguish amongst *M. schizocarpa* samples and there was no specific region for amplification in *M. textilis*, indicating that there is no homologous primer binding sites flanking the microsatellite, or a deletion with respect to *M. acuminata* (i.e. null alleles).

Thus, another SSR primer which is more universal to all species, or else specific to these species, is required to distinguish among and between species. This shows that specificities of SSR to provide distinctive fingerprints in each samples and useful marker systems. During 2006, homologous regions have been sequenced in BACs originating from both A and B genomes, and careful examination of these sequences to find flanking sequences which are present in both these genomes might improve the chance of amplification in *M. textilis*, rather than giving null amplification. Thus, PCR products obtained from Butuhan (hybrid of *M. textilis* and *M. balbisiana*) were believed to come from the *M. balbisiana* ancestor. No PCR product was obtained from primer MuH9-SSR6 for *M. velutina* and two faint bands were observed on *M. ornata.*

Both wild type and hybrid *Musa* produced polymorphic patterns. These patterns can be further studied using acrylamide gel for better separation and sequencing for more specific mutation in the repeats. *Musa* hybrids are complicated and might originate from a wide range of wild diploids. Thus, individual wild type species must be clearly understood before any conclusion of origin for the hybrids can be made. Their diversity can be better studied using the whole genome level mapping in BACs insert libraries, *in situ* hybridization and informatics methods.

3.5 Conclusion

Retroelements are very diverse and highly abundant in *Musa,* but, within the *Pseudovirideae* and *Metavirideae* groups, do not show clear clusters by their sequences which relate to known phylogeny or genome composition. This is consistent with previous studies: retroelements are generally too unstable from technical and biological reasons and present at too high copy number for their internal sequences to be useful as inter-specific marker. However, both individual sequence variants and the presence of insertional polymorphisms for individual elements might be useful for detection of intra-specific variation and determining identity or relationships within small groups of species.

SSRs are useful as diversity markers and give indications about phylogeny and relationships. The results here show that development from BAC sequences is an efficient way to generate SSR-flanking primers and about 50% of the resultant PCR products are usable as genetic markers in a range of different accessions of *Musa*. Some also give multiple bands indicating number of alleles and because of their polymorphism were at least usable to indicate ploidy. Most of the primers were able to amplify from the A, B, S but not T genomes, a relatively wide range of SSR specificity.

The A genome generally was more polymorphic than the B genome accessions studied. BAC clones have high frequency of SSR and are a suitable source for designing SSR primers but they have diverse retroelements. It suggests that BAC clones derived from A and B genomes as reliable sources for sequences alignments to define primers which are homologous in flanking regions of both species; BACs from different genomic regions may give different variation in levels and nature of SSR polymorphism.

CHAPTER 4

4.0 Diversity and evolution of NBS-type disease resistance gene homologues in *Musa* sp.

4.1 Introduction

As an important fruit crop species, it is important to know about the diversity, origin and evolution of *Musa* (bananas and plantains) for *Musa* improvement and the development of new domesticated cultivars. A particular demand is related to improved resistance to disease. The main difficulties in improvement relate to its long life cycle, its large physical size (with consequences on time scales, large plantation areas and difficulties of experimental replication), and high sterility. These factors have limited the development of *Musa* genetics, and we have a very limited understanding of the diversity within unimproved germplasm, leading to a lack of understanding of possibilities for manipulation of the genes responsible for the control of important traits. Thus, biotechnological approaches, molecular techniques and bioinformatics are required to support and enhance breeding programmes.

Disease resistance is one of the most important properties of crop plants, whether they are grown under intensive, high-input conditions, or in smallholder-based systems. It has a major impact on the sustainability of production in both environmental and economic terms. Many plant disease resistance genes are members of large gene families, whose individual members may interact with a diverse spectrum of pathogens. However, the similarities in structure and the conserved domains they share have been shown to be the result of their participation in protein-protein interactions and signal transduction. Such gene families have been found in all plants examined to date, and have been extensively studied in *Arabidopsis*, tomato, tobacco, cereals including rice and wheat, soybean, peanut, chestnut rose (*Rosa roxburghii*), grape, cocoa, cassava, coffee and many other species.

Resistance genes (R-genes) are often arranged as tandem direct repeats, consistent with an origin via gene duplication over evolution (Ronald, 1998). Many of the genes implicated in resistance to viruses, bacteria, fungus, nematodes and even insects share conserved internal sequence domains, and these have been used to define a class of sequences known as resistance gene analogues (RGA) or resistance gene homologues (RGH) (Cannon *et al.,* 2002). For many of these sequences, functionality has yet to be determined or fully understood. However, the conservation of these sequences across many plant species is a particular advantage for identifying candidate genes for resistance in species such as *Musa*.

Homologous genes have common origins but do not necessarily have the same activities as analogous genes and vice versa. Furthermore, homologues are described by the degree of sequence match while analogs have a common activity throughout the evolution regardless of sequence. Higher selection pressure - either biotic and/or abiotic stress – may conserve motifs of important genes while nonessential genes diverge from each other. Thus resistance proteins are believed to be the key of plant survival and through mutation and/or changes in the protein sequences lead to novel plant resistance genes or the correct response to challenge by pathogens.

Sequence comparisons have been made between R-genes encoding proteins with Leucine Rich Repeats (LRR) and Nucleotide Binding Site (NBS) domains, which form an NBS-LRR complex (Pan *et al.,* 2000a). The NBS protein domain is essential for the catalytic activity of various proteins (Saraste *et al.,* 1990) and the domains are known as members of the largest and diverse families in plant species (Meyers *et al.,* 1999). The NBS domain is the most common of these conserved motifs, which also include Kinase-1 (P-loop), Kinase-2 and Kinase-3 (Hammond-Kosack *et al.,* 1997; Pan *et al.,* 2000b). Another resistance motif RNBS (resistance nucleotide binding sites) A, B, C and D has been located in the NBS domain (Trout, 1994; Meyers *et al.*, 1999).

R-gene homologs are abundant in plants. It is believed that the NBS domain participates in signalling transduction, while the LRR domain appears to be responsible for primary elicitor recognition. The LRR motif may act as a hydrophobic core, with the intervening residues performing a surface function related to ligand binding. Thus, the NBS-LRR complex works as a conformational medium and may release a cascade of defence responses which lead to cell death. The role of NBS-LRR genes in encoding plant resistance to disease and/or pests has been demonstrated by many researchers involved in gene discovery and isolation (Hulbert *et al.*, 2001; Michelmore, 1998; 2000).

The NBS-LRR families were subdivided into two broad groups based on the presence of other recognisable motifs, such as the Kinase-1 (P-loop), Kinase-2 and Kinase-3 at the N-terminus. The first group known as TIR (homology to *Drosophila* Toll *and mammalian Interleukin-1* receptor) and the second group codes for coiled-coil (CC) or sometimes from leucine zipper (LZ) known as TIR, which flanking near the N-terminus (Pan *et al.*, 2000a, 2000b; Meyers *et al.,* 1999). R-genes isolated to date include *L6, RPP5, M* and *RPP1* in the first group, and *Rps2, Rpm1, I2, Xa1, Mi* and *Dm3* in the second group.

Pathogen mutation is responsible for the breakdown of host resistance, so species survival is dependant on its ability to generate diversity at disease resistance loci (Hammond-Kosack and Jones 1997). Thus, it has been speculated that DNA rearrangement is responsible for the evolution of R-genes, but the mechanism of such a rearrangement remains unexplored. *M. acuminata* 'Calcutta-4' is currently being studied through the Global *Musa* Consortium as a potential source of resistance genes (INIBAP, 2001). PCR approaches via degenerate primers based on conserved NBS domains should allow for the identification of specific R-genes in *Musa* sp. Furthermore, these sequences could be utilised as probes and markers in identifying new disease resistance genes. Molecular markers have become the most rapid means to evaluate, identify and map genes. Once a gene for a particular trait has been defined, its isolation serves to locate homologues in other species.

Understanding diversity of the conserved region related to R-genes by cloning and sequencing eventually demonstrate the evolution, divergence, interaction and relationship amongst them. Further, it might lead to the resistance protein and further through gene expression might recognise the specificity of the pathogens which yet no evidence to support this hypothesis.

In the present work, I aimed to identify the nature and diversity of NBS domains (between the P-loop and the GLPL motif), which are important determinants of the functional properties of R genes. A secondary aim was to study the evolutionary and functional relationships across these sequences derived from the MIPS-Sputnik EST database and from genomic DNA clones extracted from A, B, S and T diploids, including both wild and cultivated accessions. This was achieved by the design and implementation of degenerate PCR primers.

4.2 Materials and Methods

4.2.1 Genomic DNA isolation

Total genomic *Musa* DNA was isolated and quantified as described in Chapter 2, excluding the samples obtained in dissolved DNA form from CIRAD. The 42 accessions used were Awak, Berangan, Butuhan, Calcutta-4', Giant Cavendish, Jari Buaya, Klutuk Wulung, Mas, Nangka, Obino L'Ewai, *M. ornata*, Radjah, M. textilis (2 varieties), *M. velutina*, Mutiara, Rastali, Batu, Bluguoe, Errans, Siamea, Banksii, Tani, Honduras, Saba, Yawa2, *Malaccensis*, IPTJ, Pahang, Madu, Borneo, Lilin, Gaba, *M. schizocarpa* (three varieties), Tiparot, Pelipita, Ambon, Mysore, Lujugira and Balbisiana (refer to Table 2.1 for accession details).

4.2.2 Primer design

A data base search for R-gene sequences was conducted with BLAST (http://www.ncbi.nlm.nih.gov/blast). These sequences were tested against the Sputnik, and NCBI and TIGR sequence databases. A significant number of homologous sequences were found in the MIPS-EST database.

The GenBank NBS sequences used for primer design and sequence alignment were *N* (U15605), M (U73916), *Prf* (U65391), *RPM1*(X87851), *RPS2* (U12860), *RPS5* (AF074916), *RPP1* (AF098962), *RPP5* (U97106), *RPP8* (AF089710), *I2C-1* (AF004878) and *XA-1* (AB002266). Protein sequences were aligned using Jalview Multiple Alignment Editor V 1.3 (EMBL-EBI Version) software in default settings. Degenerate oligonucleotide primers were designed based on conserved motifs of known R-genes containing an NBS domain. The expected size of the PCR product was approximately 600bp.

4.2.3 PCR amplification of genomic DNA

PCRs were performed in a TGradient Thermocycler (Biometra) in 25 μl volumes for 30 cycles as described in Chapter 2. The annealing temperature (Ta) for the combinations CNL298F/M1445R and CNL298F/NBSR1R were 50°C and 54°C respectively. PCR products were separated on 1.2% agarose gels. Products were excised from the gel and purified using Qiagen Gel extraction Kits following the manufacturer's protocol.

4.2.4 Cloning and sequencing of PCR products

Ligation was performed in 15μl reactions using the pGEM-T Easy Vector System I kit (Promega), and transformation was performed as described in Chapter 2.5. For each fragment, five recombinant plasmids were screened, confirmed and isolated as described in Chapter 2. Three clones were selected from each sample for sequencing.

4.2.5 Sequence analysis and alignment

For consistency, the analysis was restricted to the complete region between the P-loop and the GLPL motif. Wild type samples were first analyzed by multiple alignment to construct a sequence tree where *M. ornata* was used as an outgroup. As the tree distinguished four clusters, samples in each cluster were aligned separately and the consensus sequences from each cluster were then used to analyze cultivated *Musa* samples. Other NBS domains (*N* (U15605), M (U73916), *RPM1*(X87851), *RPS2* (U12860), *RPS5* (AF074916), *RPP1* (AF098962), *RPP5* (U97106), *RPP8* (AF089710), *I2C-1* (AF004878) and *XA-1* (AB002266)) were used in for a comparative phylogenetic analysis and a multiple alignments study.

4.3 Results

4.3.1 Primer design and PCR amplification

Currently, there is no public database of *Musa* R-gene homologues. Both Group I (TIR) and Group II (Coiled-Coil) sequences were used to design degenerate primers with modifications based on their conserved consensus domains. Multiple alignments showed that the sequence was variable at both ends. In between, three conserved regions – the Kinase-1 (P-loop), Kinase-2 and Kinase-3 - were identified. The variable sequences were in the LRR region, and this differed between species.

The first primer design attempt relied on a multiple alignment of a large collection of plant R-genes, putative R-genes, putative resistance proteins and resistance proteins from both Group I and Group II. This approach failed to identify consistent conserved motifs, making it difficult to design degenerate primers. Thereafter, only expressed R-genes from both TIR and non-TIR group were included (Figure 4.1). The degenerate NBS primers designed are shown in Table 4.1.

Two degenerate oligonucleotides (CNL298F and M1445R) were designed to amplify from the P-loop to the GLPL domain. Pilot amplifications from template of Calcutta-4 and Klutuk Wulung showed multiple PCR products. NBS1R is a modification to the reverse primer by reducing its degeneracy. The new combination of primers resulted in a single reproducible banded of PCR product at size approximately 500bp.

The primer combination CNL298F / M1445 produced multiple products in the size range 300-900 bp (Figure 4.2). Strongly amplified products of size approximately 500bp and 900 bp were found in the amplicon of all samples, whereas the 700bp product was only amplified from Calcutta-4, Klutuk Wulung, Nangka and Obino L'Ewai. The 700 bp and 900 bp product from Klutuk Wulung and Calcutta-4 were cloned and sequenced.

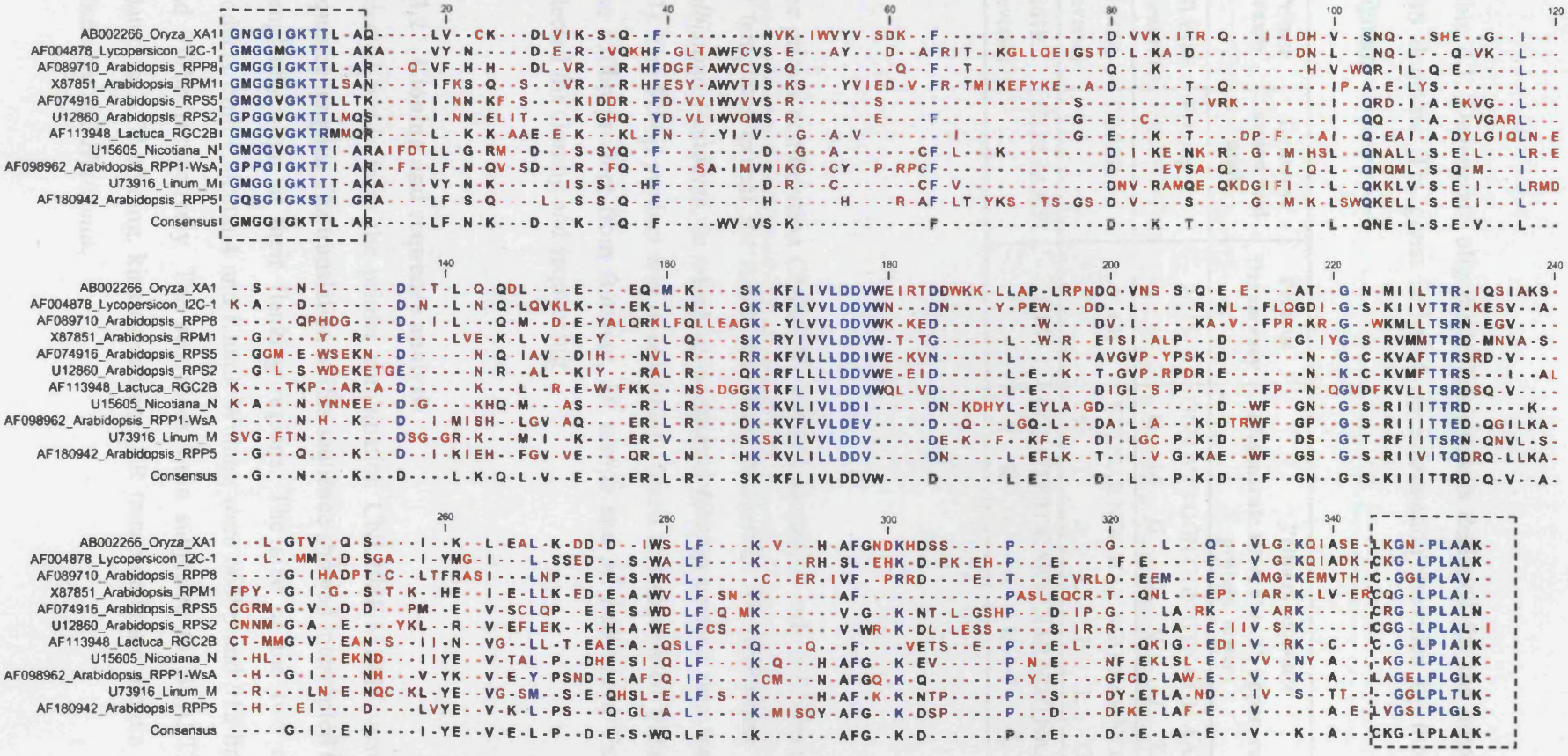


Figure 4.1. Multiple sequences alignment of NBS (P-loop to GLPL) domains of known plant resistance genes consists of TIR and non-TIR (coiled-coil), which known as Group I and Group II respectively (Pan *et al.*, 2000b). The sequences were obtained from NCBI GeneBank database (http://www.ncbi.com). Degenerate primers were design based on the consensus from the multiple sequence alignment as shown in boxes.

Table 4.1. Degenerate oligonucleotide primers designed for the amplification of NBS domains. Tm given is based on information provided by the synthesiser (Sigma).

Primer name	Conserve amino acid motif	Percent Degeneracy	Primers sequence (degenerate DNA 5' to 3', upper, related protein, lower)	Tm (°C)
CNL298F (Forward)	P-Loop	26.1%	GGN ATG GGN GGN GTN GGN AAR AC G M G G V G K T	62.7
M1445R (Reverse)	GLPLK	47.6%	YTT NAR NGC NAR NGG NAR NCC K L A L P L G	54.9
NBS1R (Reverse)	GLPLALKT	25%	CGT CTT TGC MGC NAR NGG NAA NCC T K L A L P L G	66.3

The primer combination CNL298F/ NBS1R amplified only a 520bp product from all templates except for Awak, Butuhan, *M. textilis, M. balbisiana* Tani and *M. balbisiana* Honduras, in which an additional 620 bp product was formed (Figure 4.3). The 520bp product was re-amplified, cloned and sequenced from all samples. The 620bp product from Butuhan, *M. textilis* and *M. balbisiana* Tani were also selected for cloning and sequencing.

4.3.2 Cloning and sequence analysis

Selected bands from the primer combination CNL298F/M1445 were of various sequence lengths. The translation of this sequence showed interrupted ORF and not completely align at their flanking regions. The size of the two amplification products from Calcutta-4 and Klutuk Wulung were 665bp and 912 bp, and 614bp and 917bp respectively. The sequences were subjected to BLASTX and were related to ATP-binding, kinases and an LRR transmembrane protein from *Oryza, Medicago* and *Heliconia*.

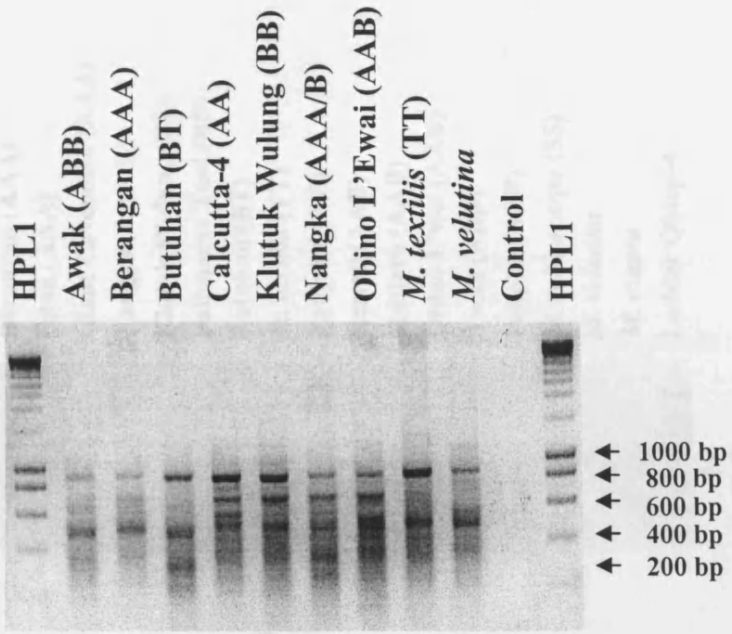


Figure 4.2. PCR products of genomic DNA obtained from NBS primers of CNL298F and M1445R spanning P-loop to GLPL motifs. HyperLadder 1 (HPL 1) is located at both sides of the gel. Fragments at 900 bp, 700 bp and 600 bp from Calcutta-4 and Klutuk Wulung were excised and cloned in pGEM-T Easy vector.

A total of 225 clones from 42 *Musa* samples were selected based on M13 amplification followed by *Eco*RI enzyme digestion for insert size confirmation. Of these, 135 clones were chosen for sequencing. BLASTP analysis showed that all clones share similarity to plant disease resistance genes and disease resistance proteins. Of these, 102 clones (75%) have continuous ORF sequences, corresponding to the NBS domain, and multiple alignments were performed from these sequences (Figure 4.4).

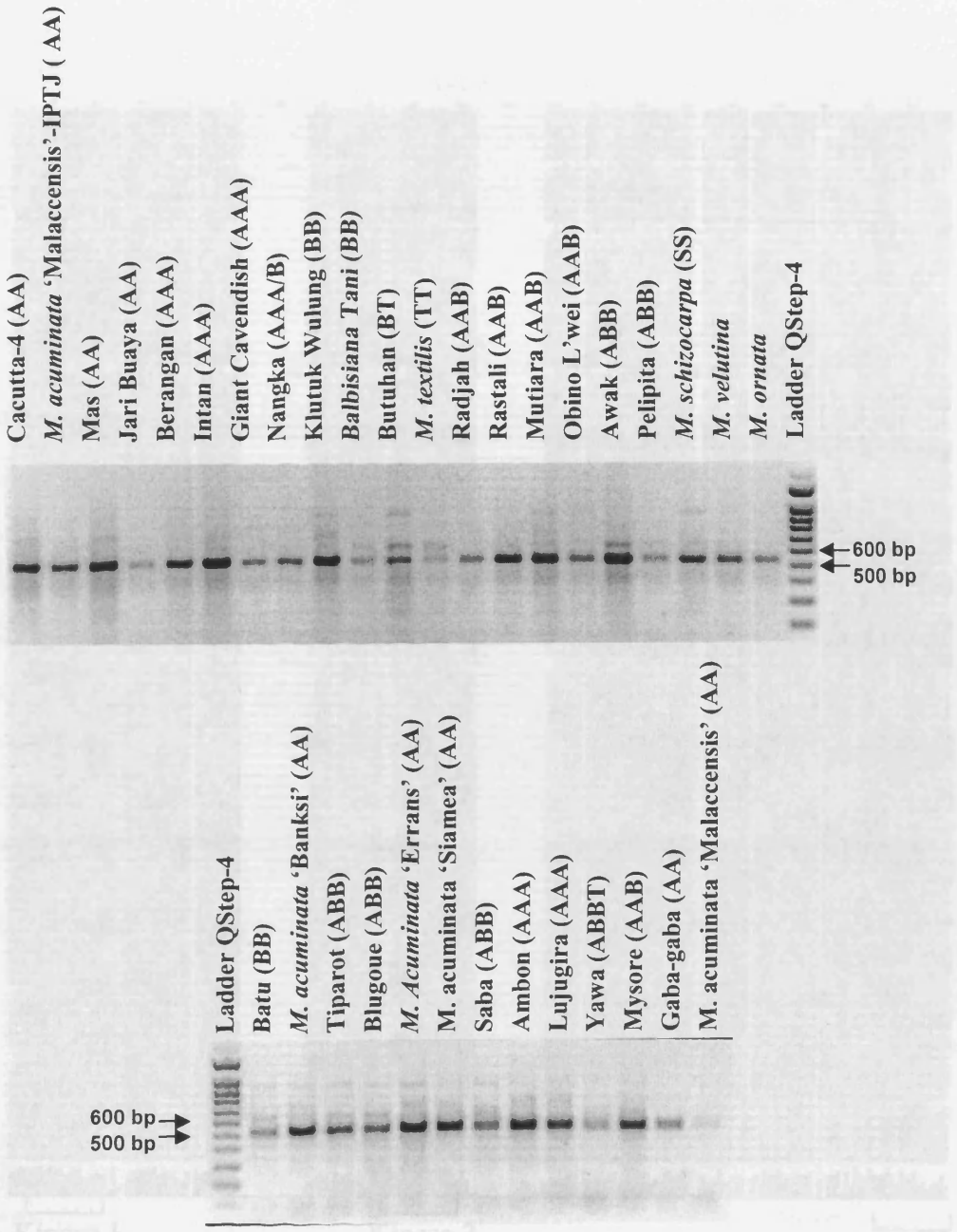


Figure 4.3. PCR products of genomic DNA obtained from NBS primers of CNL298F and NBSR1 spanning from P-loop to GLPL motifs at approximately 500 bp. QStep-4 was used as markers.

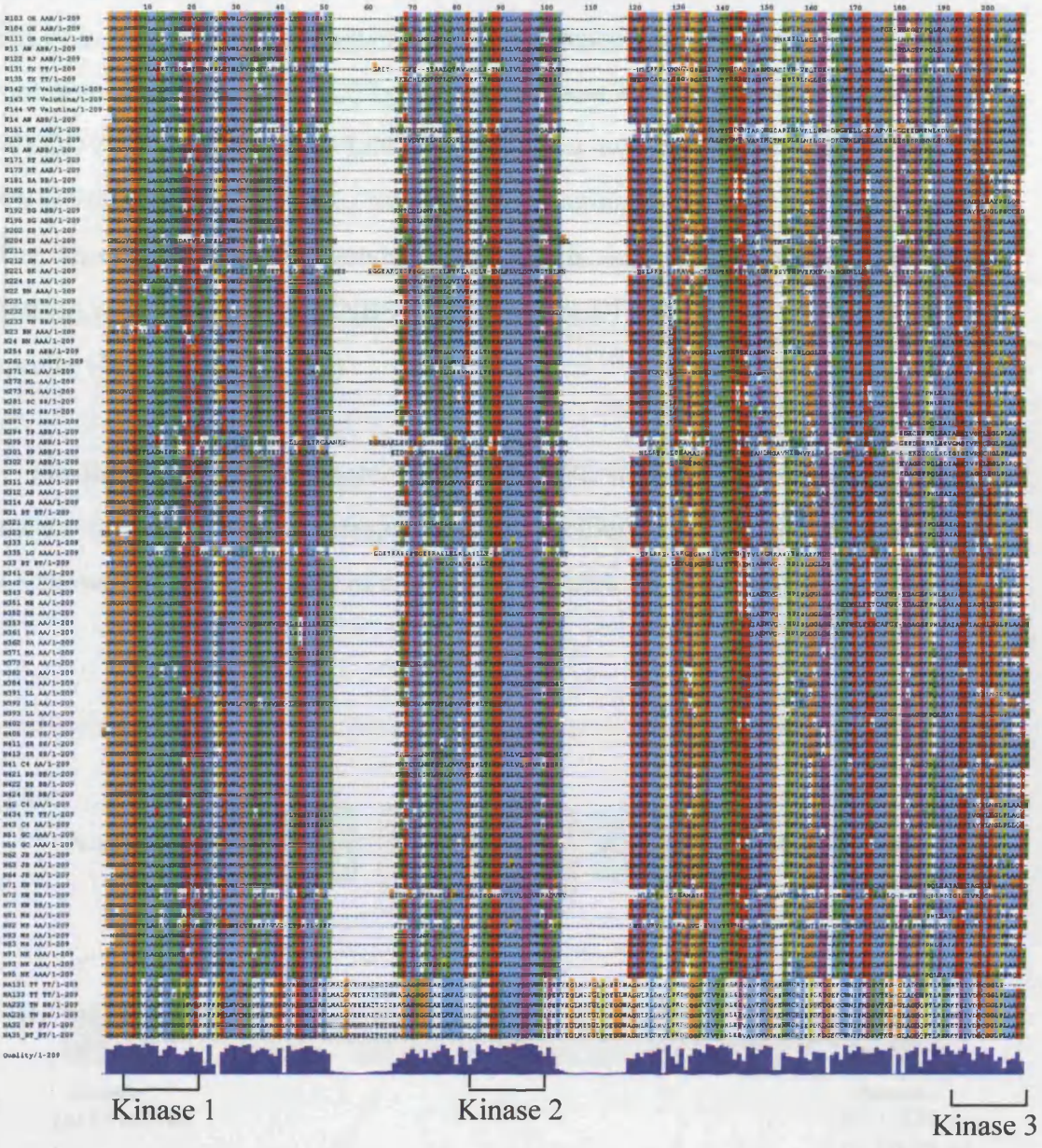


Figure 4.4. Multiple alignments corresponding to NBS domain performed uninterrupted ORF deduced amino acid from *Musa* clones. The NBS domain consists of Kinase-1 (P-loop), Kinase 2 and hydrophobic Kinase-3 (GLPL) domain in. Kinase-2 motifs is conserved (VLDDVW) in *Musa* NBS which, indicates the R-genes for *Musa* consists of Coiled-coil (non-TIR) in N-terminus as interaction platform for downstream signalling activations.

The conserved NBS domain in *Musa* consists of a P-loop (GMGGVGKT), Kinase-2 (VLDDVW), an intermediate sequence (KILTVTTR), and Kinase-3 (GLPL), similar to other known disease resistance genes. However, the Kinase-2 motifs in the 520 bp and 620 bp fragments show a replacement of Phenylalanine (VFDDVW) for Leucine (VLDDVW). Another two conserved NBS motifs were KVWVCVS and ASYWKK, but these were ambiguous in the 620 bp sequence. The *Musa* NBS domain contains novel motifs, known as RNBS-A (consensus: FDLxAWVCVSQxF) and RNBS-C (YEVxxLSEDEAWELFCKxAF) in other plant species (Mayers *et al.*, 1999).

Flexibility in the hydrophobic motifs of GLPL suggested that they may act as an anchor for LRR motifs, corresponding to specificity in R-gene activation. Figure 4.5 shows a predicted model of a complete R-genes for biotic resistance in *Musa*.

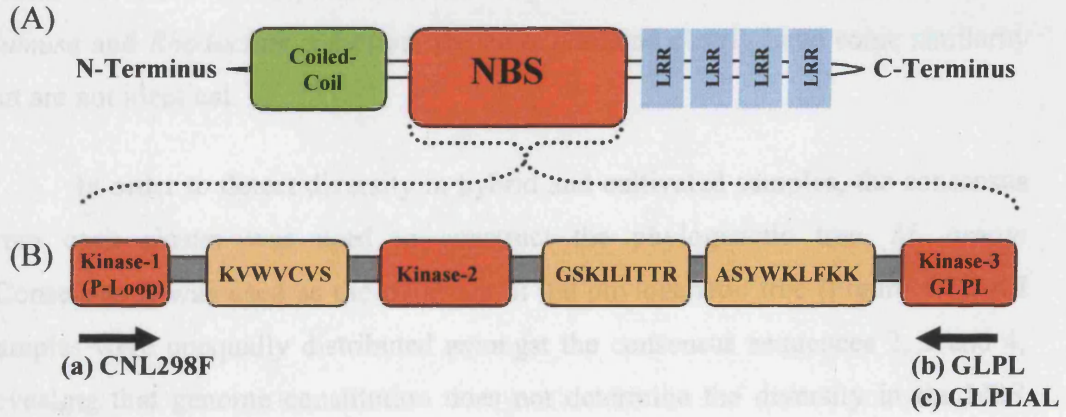


Figure 4.5. (A) Predicted model of complete R-Genes consists of coiled-coil (non-TIR) N-terminal domain and Leucine-Rich Repeat (LRR) C-Terminal domain (drawing is not in scale). B) Degenerate primers designed from conserved motifs of P-loop (Kinase-1) to hydrophobic GLPL (Kinase-3) for Nucleotide-binding site (NBS) to anchor the NBS region in *Musa sp*. Kinase-2 is a conserved motif of VLDDVW which, indicates the R-genes in *Musa* is having only Coiled-coil motifs in N-terminus as interaction platform for downstream signalling activations. Forward and reverse primers are (a) and (b) - (c) respectively.

In order to obtain an improved understanding of the diversity in *Mus*a, a phylogenetic tree was constructed separately for the wild type and cultivated samples. Figure 4.6 shows the UPGMA tree, where the bootstrap values allow the wild type accessions to be grouped into four clusters. The neighbour-joining algorithm (Saitou and Nei, 1987) (tree not shown) gave a similar positioning pattern for the clusters, but with lower bootstrap values, as this method subtracts the distance to all other nodes from the pair wise distance.

M. ornata was used as outgroup in this tree and grouped with Cluster I. Most of the members in Cluster-II were defined by the 620bp fragment. However both heterogeneous fragments from Butuhan (N33 and NA35) and *M. textilis* (NA131 and N131) were found in cluster II. This shows that the NBS motif may play a different role in R-genes. Cluster III consists of *M. acuminata* (A-genome), *M. balbisiana* (B-genome), *M. textilis* (T-genome) and *M. schizocarpa* (S-genome).

Samples in each cluster reflected that the NBS domains were highly similar and might share the functional R-genes. Cluster IV consists of *M. velutina* and Calcutta-4, separated on two branches. This shows that their NBS domains are similar but non-identical, and therefore may or may not share functionality. In the *Eumusa* and *Rhodoclamys* section, the NBS domains clearly have some similarity but are not identical.

In order to detect diversity in hybrid and cultivated samples, the consensus from each cluster was used to construct the phylogenetic tree. *M. ornata* (Consensus I) was used as the outgroup in the phylogenetic tree (Figure 4.7). All samples were unequally distributed amongst the consensus sequences 2, 3 and 4, revealing that genome constitution does not determine the diversity in the NBS domain. However, the phylogenetic model indicated three distinct groups (A, B and C) supported by high bootstrap values, suggesting that the NBS domains are diverse and might relate to specificity in the R-genes.

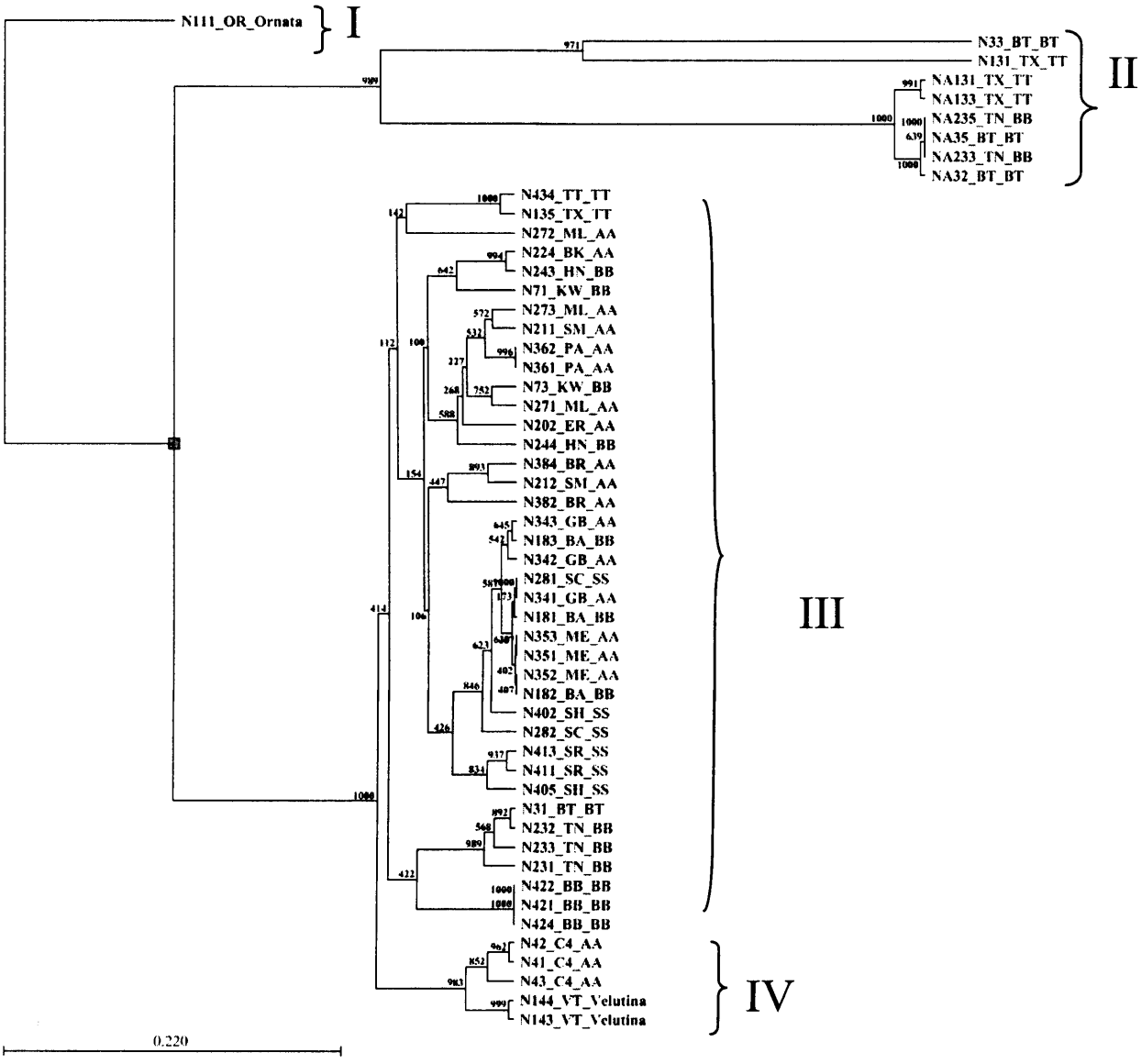


Figure 4.6. Comparison based on alignment of amino acid sequences from NBS domain of wild type *Musa sp.* Two wild type samples represent section *Rhodochlamys* whereas 20 wild type samples represent for section *Eumusa*. The sequence tree model suggests that *M. ornata* and *M. velutina* both from *Rhodochlamys* section are the most distantly related to *Eumusa* section. *M. ornata* has been used as outgroup taxon in this study. Branches corresponding to clusters of related NBS domain are labelled I-IV. The tree was constructed by using UPGMA method provided in *CLC Free Workbench 2.2.5* with bootstrapping of 1000 replicates (numbers near the branches). Branch lengths are proportional to the average amino acid substitutions per site as indicated by the scale.

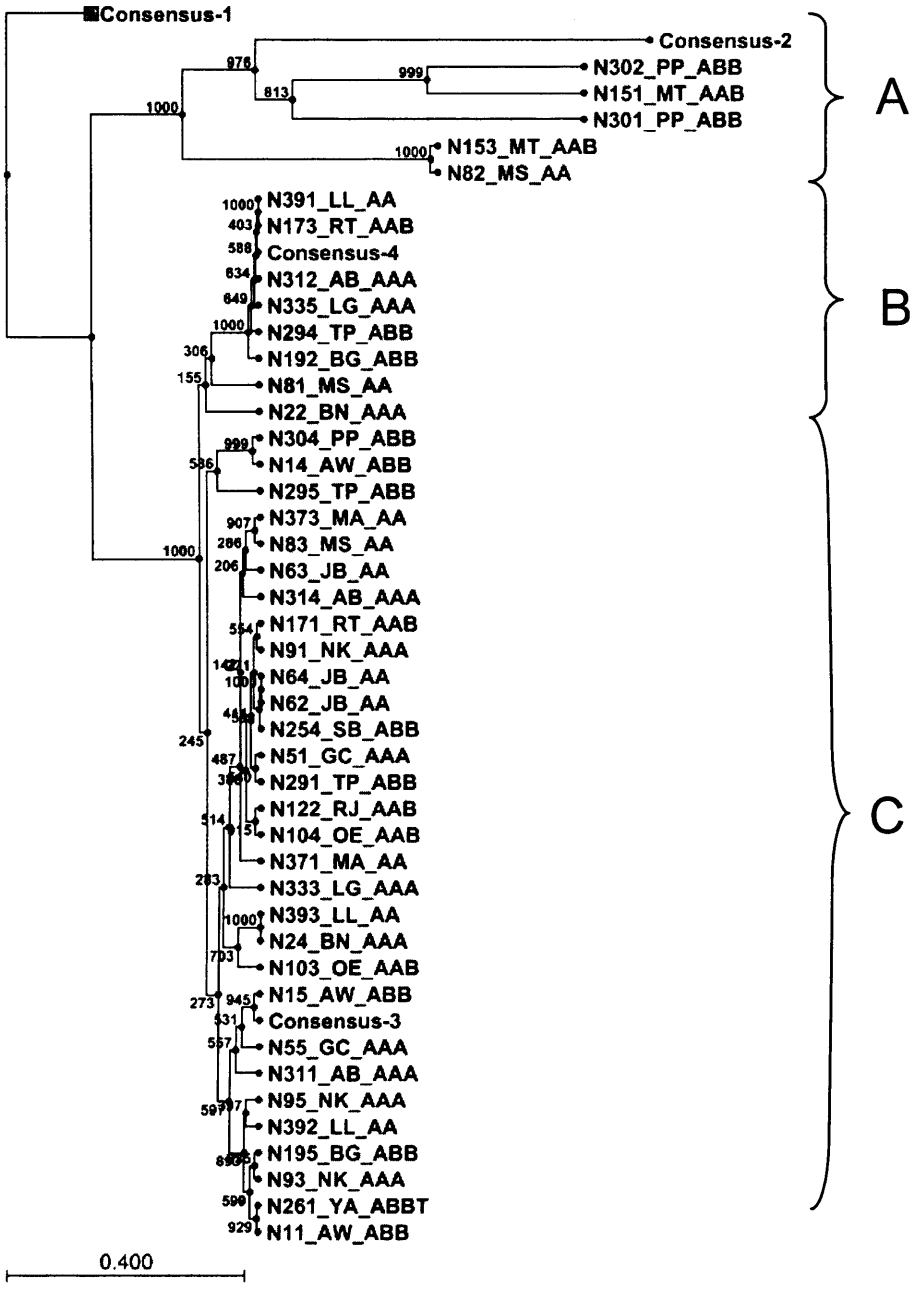


Figure 4.7. Comparison based on alignment of amino acid sequences from NBS domain of 20 hybrid *Musa* cultivars in *Eumusa* section. The sequence tree model suggests that *M. ornata* from *Rhodochlamys* section are the most distantly related to *Eumusa* section and used as outgroup taxon in this study. Clusters obtained from wild type *Musa* analysis were aligned separately and their consensuses were included in this study labelled as Consensus_1 to Consensus_4 to identify diversity of the hybrids and cultivated *Musa*. Diversity of NBS domains are not corresponding to *Musa* genomic constitution. Three distinct groups were identified labelled as A, B and C. The tree was constructed by using UPGMA method provided in *CLC Free Workbench 2.2.5* with bootstrapping of 1000 replicates (numbers near the branches). Branch lengths are proportional to the average amino acid substitutions per site as indicated by the scale.

4.3.3 R-genes, resistance and mutation-derived varieties

A comparative analysis was made based on multiple alignments between NBS domain consensus sequences of the *Musa* clusters (Figure 4.8), followed by a sequence tree construction. As shown in Figure 4.9, *Musa* consensus sequences were grouped together with *Arabidopsis RPP8* and *RPM1*, *Solanum* (formerly *Lycopersicon*) *I2C-1* and *Oryza XA1* genes. Further analysis needs to be carried to study the gene expression of the key R-genes. The phylogenetic tree indicated that the domains shared an identical motif at Kinase-1, Kinase 2 and Kinase-3. Furthermore they have similar residues in the RNBS-A, RNBS-B motifs, but are ambiguous in RNBS-C (Figure 4.8), making it hard to design informative primers at these regions.

A similarity analysis of the NBS domain was performed using CLUSTALW on the multiple sequences alignments of the *Musa* consensus and R-genes (Table 4.2). The percentage similarities between the *Musa* consensus and the other R-genes ranged from 27% to 87% and 8% to 47% respectively. Consensus-3 and consensus-4 are most similar to *Xa1* and *I2C* with 40 % to 42% similarity. A comparison between Mutiara and its derivative Rastali was made based on percentage similarity in the NBS domain to understand their relationship as both cultivars respond differently to *Fusarium* wilt. The percentage similarity for individual clones obtained from Mutiara and Rastali were 41% and 82% respectively.

The percentage similarity between the cultivars ranged from 38% to 47%, showing that the NBS of Mutiara has changed, and suggesting that these changes may be connected with the acquisition of tolerance to *Fusarium* wilt. The similarity comparison is shown in Table 4.4. Multiple alignments for the deduced amino acid sequences obtained from Mutiara, Rastali and their consensus sequences indicated five unique motifs where the changes had occurred (Figure 4.10).

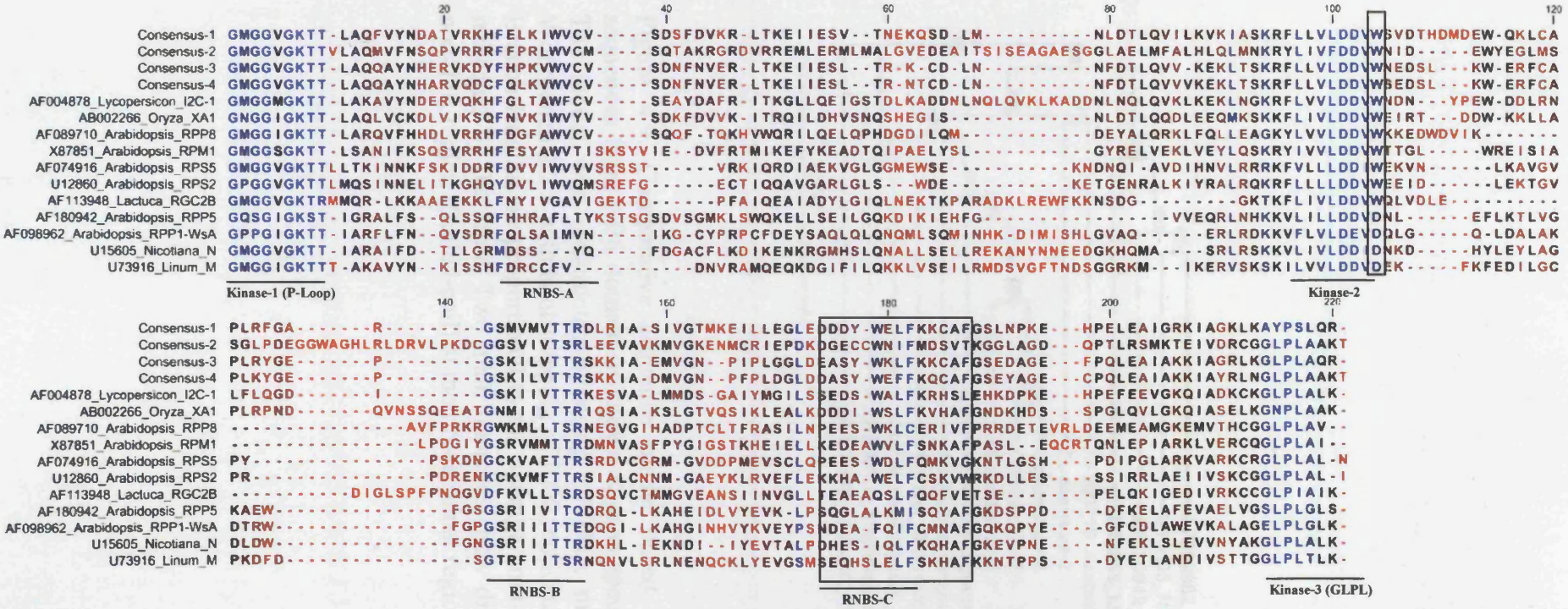


Figure 4.8. Comparative multiple alignments of deduced amino acid sequences correspond to NBS domain of *Musa* consensus and analogous domains of known R-genes performed by ***CLC Free Workbench 2.2.5*** . Gaps are indicated by '-'. The NBS domain conserved at Kinase-1 (P-loop), Kinase 2 and hydrophobic Kinase-3 (GLPL) domain in. Kinase-2 motifs is indicated that *Musa* consensus are in non-TIR Group and marked with boxes. The alignments shows RNBS-A and RNBS-B motifs are conserved but ambiguous in RNBS-C.

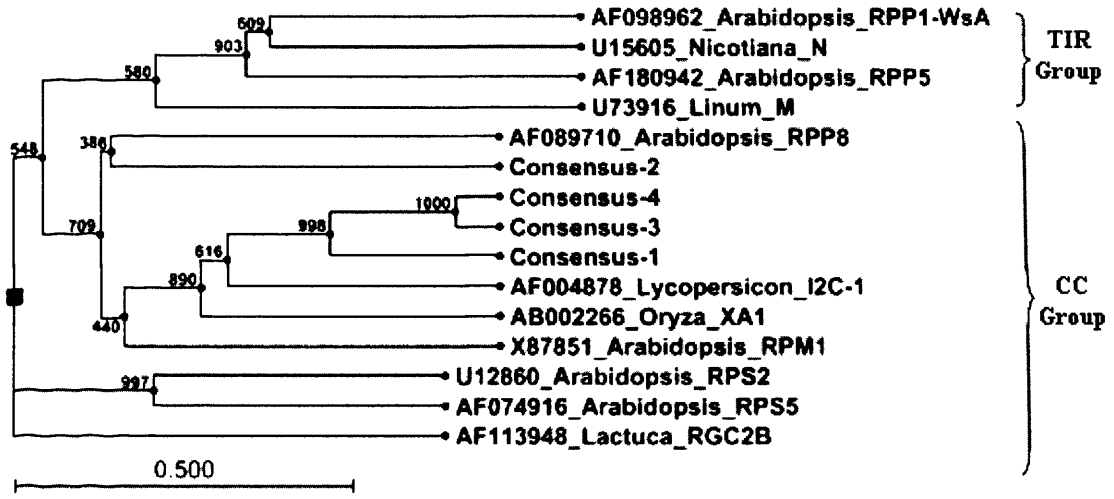


Figure 4.9. Comparative sequence model based on alignment of amino acid sequences from NBS domain of known R-gene species and *Musa* sp. consensuses. The model was constructed by using UPGMA method provided in CLC Free Workbench 2.2.5 with bootstrapping of 1000 replicates (numbers near the branches). Branch lengths are proportional to the average amino acid substitutions per site as indicated by the scale. Two main branches show differentiation between TIR and non-TIR region (Coiled-coil) at their N-terminal region.

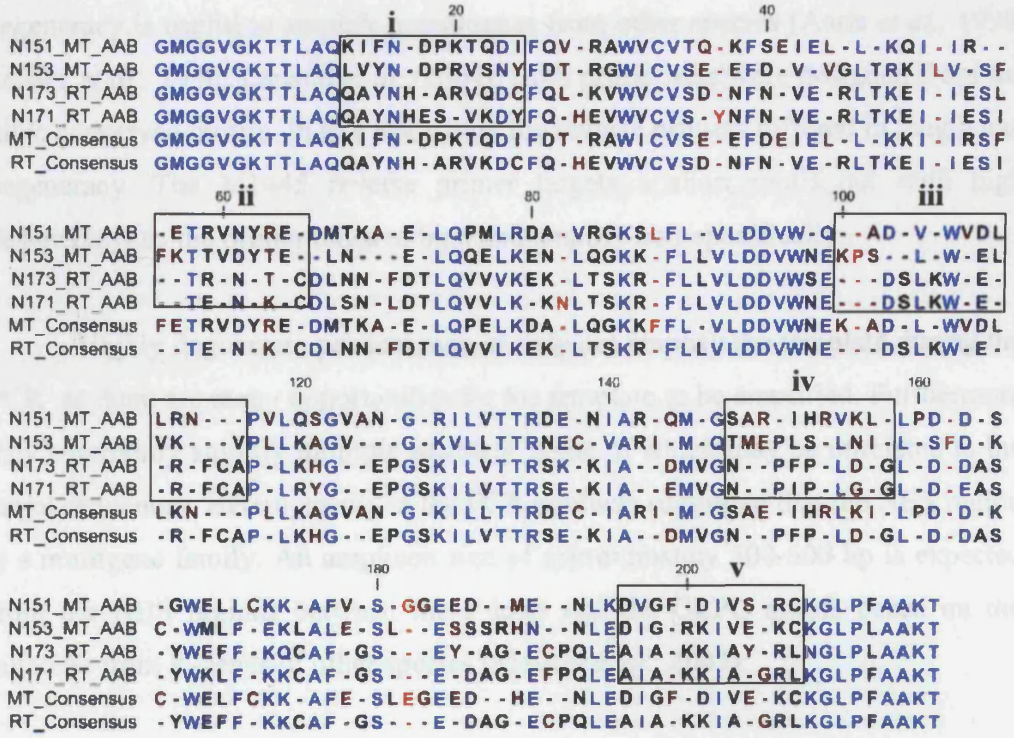
Table 4.2. Similarity percentage of deduced amino acid sequences of NBS domain from wild type *Musa* consensuses (shaded in green colours) and known R-genes. TIR and non-TIR group for other plant R-genes of NBS domain have been shaded in blue and yellow colours respectively. The similarity was performed using CLUSTAL W (1.83) Multiple Sequence Alignments.

	Consensus-1												
Consensus-2	27	**Consensus-2**											
Consensus-3	61	31	**Consensus-3**										
Consensus-4	58	30	87	**Consensus-4**									
I2C	43	32	42	41	***I2C***								
Xa1	47	29	42	40	35	***Xa1***							
RPP8	27	34	28	30	36	26	***RPP8***						
RPS5	28	25	19	31	29	29	23	**RPS5**					
RPM1	33	31	33	30	30	27	31	28	***RPM1***				
RPS2	25	24	31	29	30	29	30	41	28	***RPS2***			
RPP1	21	13	20	30	22	22	15	17	12	15	***RPP1***		
RPP5	14	22	19	14	22	16	18	8	18	19	43	***RPP5***	
N	24	24	28	30	30	26	19	25	26	22	37	31	***N***
M	19	19	25	25	30	23	25	24	25	24	27	30	37

Table 4.3. Similarity percentage of amino acid sequences from Kinase-1 (P-Loop) to GLPL motif between Rastali (RT) and Mutiara (MT) which are triploid AAB. Mutiara is known as tolerant to *Fusarium* wilt disease was derived from Pisang Rastali after been gradually challenged by *Fusarium* wilt disease at 'hot spot' field in Malaysia. The similarity was performed using CLUSTAL W (1.83) Multiple Sequence Alignments.

	Clone N151					
Clone N153	41	Clone N153				
Consensus_MT	78	69	Consensus MT			
Clone N171	40	47	45	Clone N171		
Clone N173	38	44	43	82	Clone N173	
Consensus_RT	38	46	46	90	91	Consensus RT

Figure 4.10. Multiple sequence alignments of deduced amino acid sequences obtained from Mutiara (MT) and Rastali (RT) from Kinase-1 (P-Loop) to GLPL motifs. Five sequence motifs marked with boxes indicated changes and unique deduced amino acid residues as compared between both clones.



4.4 Discussion

4.4.1 Primer designing and PCR-based amplification.

Interaction between resistance (R) and the corresponding avirulence (Avr) gene is the first step in the expression of R-gene encoded disease resistance. The R-protein recognises Avr-gene-dependent ligands and activates the defence mechanism signalling cascade(s). It is believed that the R-proteins are constitutively expressed in healthy and in unchallenged plants in readiness for the detection of pathogen attack (Hammond-Kosack *et al.*, 1997; Keen, 1990). Most of the R-gene encoded protein contains an NBS domain and an LRR region at the C-terminus.

In the isolation of plant resistance genes, it is important to understand and identify the signal cascade(s) believed to be related to the NBS domains. The present results have shown that the least conserved region of the NBS domain is at the Kinase-3 end. It is believed that the variable sequences are related to the LRR motifs which contribute to the specificity of the resistance genes resulted undesirable alignments.

A PCR-based technique based on less degenerate primers showed good potential to isolate and clone candidate resistance gene in *Musa* sp. Primers degeneracy is useful to amplify homologues from other species (Aarts *et al.*, 1998; Leister *et al.*, 1996; Lanaud *et al.*, 2004). Both primer sets were designed from the same conserved motifs. In this study only the reverse primers differed in length and degeneracy. The M1445 reverse primer targets a short motif but with high degeneracy, so the primer tends to bind and amplify non-specifically.

Highly degenerate primers may or may not amplify the template during the PCR as there are many opportunities for the template to be amplified. Furthermore they commonly amplify multiple products, some of which may be unrelated to the target sequences. Heterogeneity in the PCR products suggested that the NBS region is a multigene family. An amplicon size of approximately 500-600 bp is expected from the NBS regions between the P-loop and the GLPL motifs based on the outcome from R-genes of other species (Meyers *et al.*, 2003).

Primers of low degeneracy are thought to deliver reliable products, suitable for the isolation of highly divergent sequences such as the NBS domains (Noir *et al.,* 2001; Shen *et al.,*1998; Kanazin *et al.,* 1996, Leister *et al.*, 1996). In designing the degenerate primers, only sequences related to the target (R-genes) both from TIR and non-TIR have to be considered as large multiple alignments resulted in an inconsistent consensus and implied the design of highly degenerate primers. A 25% level of degeneracy resulted in the amplification 135 sequences, 75% of which contained uninterrupted ORF sequence. Four distinct NBS-like clusters were identified from 42 *Musa* sp. varieties, representing a wide range of diversity.

4.4.2 Characterization of NBS domain

The multiple sequence alignments demonstrated a level of homology between NBS regions in *Musa* species, and that these were distinguishable from, though similar to many of the resistance genes from other species. Amongst the wild type *Musa* species, four clusters were distinguished. R-gene clusters and their organisation have been reported in many other species (Lanaud *et al.,* 2004; Laili *et al.,* 2005; Yuksel *et al.*, 2005; Soriano *et al.,* 2005; Xu *et al.*, 2005) and their common motifs were also found in *Musa*.

It is not surprising in other species that R-genes can be inherited by crossing, but in *Musa* where parthenocarpy and sterility are the rule, these mechanisms are not available. Thus, it appears that mutation may have been induced through biotic stress, generating small and slow changes and creating novel genes which may have been selected in cultivation. Michlemore and Meyers (1998) have suggested that duplication, divergence and deletion have occurred during the evolution of other non-NBS R-genes such as *Xa21* (Song *et al,* 1995), *Cf-2* and *Cf-5*, (Hammond-Kosack *et al.,* 1997) and *Mlo* (Buschges *et al.*, 1997). The higher similarity percentage among the *Musa* consensus sequences supports the theory of natural mutation.

As the NBS domain is present in large gene families, it is possible that these genes encode functions other than disease resistance, especially for clades without any known resistance gene products. Sequence analysis based on construction of the trees suggest that the evolution of NBS domain in *Musa* have been originated from *M. textilis* (T-genome) and *M. balbisiana* (B-genome) (Figure 4.6 and Figure 4.7). Cluster II is represented in both genomes but they are distributed in cluster III and further emerge in a phylogenetic tree constructed for cultivated *Musa*.

Although the R genes share common motifs, their specificity is clear. Motif discovery is effective through sequence clustering as they correspond to functionally similar sequences, but analysis is limited by computational techniques and sequence analysis per sample. Thus, reliable computational techniques and a high number of sequences are important for efficiency and precision. According to Meyers *et al.* (1999), highly similar NBS domains are common in plant species and encoded R-gene products with unknown specificity are found, which require a challenge for their expression. Complexity amongst R-genes that cluster in the genome have been found in flax (Ellis *et al.,* 1995)), lettuce (Witsenboer *et al.,* 1995), barley (Jorgensen, 1992), bean (Geffroy *et al.,* (1999), tomato (Dickinson *et al.,* 1993), *Arabidopsis* (Noel *et al.,* 1999), rice (Song *et al.,* 1997), soybean (Kanazin *et al.,* 1996) and potato (Leister *et al.,* 1996).

The most conserved NBS motifs in *Musa* are the P-loop (GMGGVGKT), Kinase-2 (VLDDVW) and GLPL. These motifs are present in R-genes that confer resistance to a broad spectrum of pathogens (Baker *et al.,* 1997; Hammond-Kosack *et al.*, 1997). The P-loop (Kinase-1) has been well characterized and is involved in ATP and GTP binding protein (Bourne *et al.,* 1991; Trout, 1994), as it interacts directly with the phosphate of bound NTP (Saraste *et al.*, 1990). The Kinase-2 motif includes an aspartate residue important for the coordination of Mg^{2+} in phosphor-transfer reactions (Trout, 1994).

The Kinase-2 of *Musa* emerged in Group II associate to coiled-coil motifs at N-terminal sequence that they might have potential to contribute towards resistance genes in downstream signalling molecules. The most distinct changes of protein arrangement occurred between Kinase-1 and Kinase-2, and between the Kinase-2 and GLPL motifs which correspond to the specificity of the NBS downstream signalling activation of coiled-coil domains (Pan *et al.*, 2000b).

The coiled-coil structure was first described in animals as the main structural element in a class of fibrous protein which includes keratin, myosin and fibrinogen. (Pauline and Corey, 1953). It was then identified as a dimerization element in a class of transcription factors known as the Leucine zipper protein (Landschulz *et al.*, 1988). The non-TIR (Group II) genes have a coiled coil motif at their N terminus where the overall secondary structure is alpha helical, and hydrophobic residues are arranged on the side of the helices. The coiled coil structure is used to stabilize alpha helices in protein structure. Their stability is achieved by distinctive 'knob-into-hole' packing of the apolar side chains into a hydrophobic core as postulated by Crick in 1953 and later proved by Burkhard *et al.* (2001). The coiled-coil is known to be involved in dynamic motion and rearrangement in several families of transcriptional activators, dimerization and DNA-binding regions in response factors (Lupas, 1996), as the coiled-coil domain determines its oligomerization state, rigidity and ability to function as a molecular recognition system. Progresses have been made in understanding the factors that determine coiled-coil formation and its stability (Burkhard *at al.*, 2001).

The functional and relationship between Kinase-1/P-loop, Kinase-2, RNBS-A, RNBS-B, RNBS-C and Kinase-3 (GLPL) are not well understood. It has been suggested that consistent structural arrangement of the common motifs in NBS domains correspond to particular biochemical function (Meyers *et al.*, 1999). However, variation patterns of these motifs observed from the same clones and highly divergent between the conserved motifs may correspond to variable specificity, and be important for the activation of the NBS domain as a whole in the signalling process. Thus, more clones from the same cultivar or species are required for a better understanding of R-gene diversity and specificity.

Less than 50% similarity with known R-genes is observed among the NBS domains of different species. The consensus *Musa* NBS domain shows a level of homology ranging from 27% to 87%. As NBS domains are multigene proteins, they may or may not be directly involved determining R-gene functionality, but their presence is are crucial for R-protein function. The percentage similarity in the same or close related species was expected to be higher than that between different species.

The pathogens involved in the activation of R-genes include bacteria, fungi and virus, and these challenges have contributed and affected NBS domain similarity and diversity. Distinct signalling pathways have been identified by Aarts *et al.* (1998) for directing the diversification of NBS sequences.

Both UPGMA and Neighbour-joining algorithm distance based methods were performed to construct the molecular phylogenies, as each requires different assumptions, and the derived phylogenetic trees were supported by bootstrap confidence levels. The bootstrap is a computer-based technique for the statistical assessment of precision of an emerging 'true tree' (Page *et al.,* 1998; Efron *et.al.,* 1996; 1993). The UPGMA method was preferred, as it is able to estimate the occurrence of the sequence divergence at a constant rate of evolution at all branches of the tree.

This means that the branches are aligned to the extent that all the sequences including the root are estimated as part of the procedure. Further, the bootstrap values are higher than those from the NJ method and showed that there is a reasonably constant mutation rate in the *Musa* sp NBS domains. In addition, the tree pattern was supported by Neighbour-joining methods although with a low bootstrap value, as this method calculates and compares distance by assuming that there are different rates of evolution of the sequences. However, for larger numbers of sequence data the bootstrap values become impractical (Page *et al.,* 1998). Therefore, wild and cultivated *Musa* sp samples were separated in order to obtain a reliable 'true tree' of the NBS domain.

Distinct groups based on the consensus in clusters II to IV were observed suggests that the NBS domain has been manipulated and/or changed gradually based on environmental challenges. Mutiara, which is tolerance to *Fusarium* wilt disease, originated from Rastali, which is susceptible to the disease through gradual selection in a disease 'hot spot' (Ho, 1999). A low percentage of similarity among the clones obtained from Mutiara as compared to Rastali suggests that rapid changes have occurred in the evolution of the tolerant cultivar.

Mutiara groups to Cluster II together with Pelipita (PP) and Mas (MS), which may indicate a similar response to disease pressure. However, other samples also may have the same response to biotic stress as they grouped in the cluster consisting of resistant plants related to different pathogens. Pelipita (PP, usually a variety with 8 A genome and 25 B genome chromosomes; D'Hont *et al.,* 2000) is resistant to Moko disease (Stover and Richarson, 1968). Calcutta-4 and Klutuk Wulung, in clusters III and IV respectively, have resistance to Sigatoka.

As R-genes are multigene families, the diversity and evolution of R-genes depends on how the specificity of avirulence signalling may be related to the organisation (tandem arrays) of the genes. The high mutation rates typical of plant pathogens, affects the effectiveness of R-Genes as a defensive mechanism. Thus, the ability to survive depends on how well the plant can maintain and generate useful levels of diversity at disease resistance loci (Hammond-Kosack *et al.,* 1997). Clustering the NBS domain might be narrowed the specific and signal recognition on diverse pathogens.

A comparative analysis has shown that the *Musa* consensus NBS domain is distributed within known R-genes. This suggests that many types of R-gene are involved in defence and are specific to various pathogens. Low similarities in the NBS domains among the R-genes shown in table 4.2 supports the suggestion that a high specificity typifies the recognition and activation of signalling of the defence mechanism in different plant species. The similarities based on the percentages of NBS domain were not biased, as the primers designed to amplify the NBS domain targeted both the TIR and non TIR class of R-gene.

Among the *Musa* consensus sequences, a broad range (from 27% to 87%) indicated that these species have evolved and diversified over a long period or at a rapid rate, in order to sustain their genetic capacity to thrive under biotic stress. However, *Musa* NBS are of the non-TIR type, which to date are restricted to the monocotyledonous species. This suggests that the NBS domain in *Musa* corresponds to genes for resistance to biotic stresses of unknown specificity which are still unclear, and thus require future research.

The consensuses presented here are reliable not only for species classification but also may be suited to understanding the movement of the NBS domain from wild species to cultivars in the domestication process of *Musa* species as shown in Figure 4.6 and Figure 4.7.

4.5 Conclusion

The NBS motifs in *Musa* are conserved in four clusters and emerge within large gene families known as the coiled-coil class, which encode the R-gene proteins. However, this conservation does not correspond to genome constitution, indicating that R-genes are specific and exist independently. Diversity in the NBS domain can classify *Musa* varieties in a specific manner, corresponding to similar disease resistance. The tendency to identify functional genes through sequence homology and clustering of known properties may serve to enhance breeding and selection activities for resistance to critical diseases in *Musa*. Further research is required to confirm and identify the specificity of the NBS domains and the relationship to disease resistant phenotypes.

CHAPTER 5

5.0 Diversity and evolution of Leucine-Rich Repeat (LRR) motifs in *Musa* spp.

5.1 Introduction

Leucine-rich repeats (LRR) are protein domains that fold in an arc or horseshoe shape with β-strand and α-helix lining the inner and outer circumference respectively. The LRR motif comprises approximately 20-30 amino acids. They can be involved in protein-protein interactions which are part of many biological processes found from viruses to eukaryotes. LRR proteins exist in different independent subfamilies or classes with an irregularly spaced leucine motif with the most conserved pattern of LxxLxLx motif which forms the β-strand (L is a leucine residue and 'x' is any amino acid). The β-strand is stabilised through hydrophobic forces between the leucine consensuses. The LRR conserved segments are present in either 11-residue LxxLxLxxNxL or 12-residue LxxLxLxxCxxL motifs where 'L' might be replaced by valine, leucine or isoleucine, 'N' is asparagine, threonine, serine or cysteine, and 'C' is cysteine or serine, and 'x' residues are proposed to be solvent exposed (Jones and Jones, 1997; Kajava, 1998; Matsushima *et al.*, 2000). Other residues which form the α-helix are not conserved and differ amongst the proteins. Together, the LRR motif is tandemly repeated and is involved in a variety of functional proteins.

Based on their length and consensus, at least seven families of LRR protein have been identified, each sharing certain features such as cellular localization, origin and flanking regions. The LRR classes were known as typical type, RI-like (Ribonuclease Inhibitor), CC (Cysteine-Containing), PS (Plant Specific), SDS22-like, bacterial and TpLRR (*Treponema pallidum* LRR) (Malvar *et al.*, 1992; Ohyanagi *et al.*, 1997; Kajava, 1998; Enkhbayar *et al.*, 2003). New classes of LRR have been found and are being identified, but still there is limited knowledge of protein signalling, and the mechanisms which are associated to ligand binding and receptors shared in their motifs. In *Arabidopsis*, 400 genes encoding putative receptor-like kinases (PRKs) were found and classified based on their extracellular domains, with the LRR amongst the largest group (Shiu and Bleeckeer, 2002; Tichtinsky *et al.*, 2003).

Relationships between LRR domains and plant signalling mechanisms clearly showed signal specificity with minimal identity of the LRR domains creating an advantage to both pathogen and host. It is clear that the LRR genes can rapidly evolve with different pathogens or pathogen races invading the host, allowing the host to recognise the pathogen and build a defence mechanism. Thus, it is predicted that through evolution, the LRR in plant species might be altered, given high levels of both diversity and specificity.

Most of plant resistance genes consist of LRR motifs either at N-terminus (before NBS domain) or C-terminus (after NBS domain). The LRR has many unknown functions and inconsistent nucleotide repeats in the sequence. Hence, there are higher chances to get different activities in the expression of the resistance genes. Two distinct classes of intracellular LRR proteins known to correspond to plant pathogen resistance according to Jones (1997) are:

i) 23- or 24-amino acid consensus
LxxLxxLxxLxLxx(N/C/T)x(X)LxxIPxx
ii) 24- amino acid consensus
LxxLxxLxxLxLxxNxLxGxIPxx

In this chapter, a research experiment aimed to design oligonucleotide primers for PCR to isolate LRR regions from diploid A, B, S and T *Musa* genomes from wild and cultivated accessions. Many LRR families are involved in protein-protein interactions and most participate in signal transduction which appears to be responsible for elicitor recognition with challenges from pathogens. The intention was to investigate the diversity of LRR motifs in *Musa* which might lead us to their important roles as plant resistance genes. Further, the aim was to characterize their evolutionary and functional relationships, and compare this with diversity in homologues and available LRR motifs from *Musa* and reference species where the genes have been characterized with functional assays.

The PCR products were cloned, sequenced and clustered according to similarities of sequences for sequence tree construction and diversity analysis both at nucleic acid and protein levels. Diversity information might be useful to understand the evolution mode in diploid A, B, S, T genome and hybrids of *Musa* accessions.

This study is important to identify nature and diversity of the LRR motifs in *Musa* and to look at their evolutionary and functional relationships across the sequences in the MIPS-Sputnik *Musa* EST database available through the Global *Musa* Genomic Consortium and new genomic clones. Finally, Southern hybridization was used to characterize genomic diversity and abundance.

5.2 Material and methods

5.2.1 Primer design and PCR optimization

The first specific LRR primer sets followed De Barros *et al.* (2000) which produced a PCR product at approximately 2079 bp in soybean lines A45-10 (Accession No. AF215729). Gradient annealing temperature (T_a) of 47.4°C, 48.7°C, 50.6°C, 52.8°C, 54.9°C, 57.1°C, 59.2°C and 61.3°C were applied for the PCR optimization (Table 5.1). Klutuk Wulung was used as DNA template and the PCR reactions were performed as described in Chapter 2.

Table 5.1. Specific primers for LRR as described in De Barros *et al.*, (2000).

Primers Name	Sequence (5'—3')	T_m (°C)	T_a (°C)
MLRR-F	TTG ACT GGT GAT GTG CCT GT	64.4	55.0
MLRR-R	CAG AAT CCT ATG CAA GCT CC	61.3	

DNA and deduced amino acid sequence of accession AF215729 were compared against *Musa* Sputnik EST sequences, and aligned in ClustalW. Homology was found *Musa* sequence accession C_600099504T1 and six primers each for specific primers and degenerate primers were designed based on the DNA sequence and uninterrupted deduced amino acids respectively (Figure 5.1).

(A) Nucleotide sequence of *Musa* accession C_600099504T1

TATTTGAACGTAGATGAAGTACTAGCAGTTGTTGTAGATTTTCTCCCATCCATAG
CGTTATACTTCCGACAAGATTATTTTGAGCCACATCAAGAAACCATAAATTTGTACA
ATTTTTCAATGACAAAGGAACCTCACCCGAAAAACTGTTATTTTTCAGTTGCAATAGC
AGAAGCTCACTTAAAAAACCGATCGTGTGTGGAATTTCACCCGACAAGCGATTGTTG
CCCAAATTCAAGAAATAAAGATTTTGTAATGATCCTCCGAAACAATAAGGGATCTCA
CCAAGTAATTTGTTGTTCGAAAGGTCAAGGTATTCGAGATATGTCCATTGGCAGATT
GATGAAGATAAGCCACCATCAAGCATATTATCTGAGAGTGCCACTGAGTGAACATG
GGGCCATATCGGCAATTGCCCTGTAAAGGAATTATTGGAGAGGTATAGAGTATGAA
GTGTAGACGGTAGCATCGTAGGCAATGGACCTTCAAATCTATTGGAATCCAAATAT
AATATAACCAACTTGGTGAACTTTAAAGAAGATGGCAGCTTCCCTCCTATTTGGTTGT
TGGAAAGGCCTAAAAATATGATCGTAGAAGATGAAATATTCCAAAACCAAGCGGGC
ATCGTCCCTGCAATTTTACAGTCTGCCAAAAATAATTCTATTATCTGCGTTTGGAATT
GCAACCATTCTGGAAATCGAGGTCCCAACTGACATTTGGTTAAATATACAAATCTGA
GTTGGAAAGGAGGGACCCAACTCTGCCCGATTGATATGGTGATGGGGTTGCCAGACA
AGTCCAAGACTTGCAATCTGGTAAGATTCTCAAAATGAACCTCTGAAATAACACC
TCCAAGAGAATTTCC

(B) Deduced amino acid sequence of *Musa* accession C_600099504T1

GNSLGGVISEVHFENLTRLQVLDLSGNPITISIGQSWVPPFQLRFVYLTKCQLGPRFPEWLQFQ
TQIIELFLADCKIAGTMPAWFWNISSSTIIFLGLSNNQIGGKLPSSLKFTKLVILYLDSNRFEGPL
PTMLPSTLHTLYLSNNSFTGQLPIWPHVHSVALSDNMLDGGLSSSICQWTYLEYLDLSNNKL
LGEIPYCFGGSLQNLYFLNLGNNRLSGEIPHTIGFLSELLLLQLKNNSFSGEVPLSLKNCTNLW
FLDVAQNNLVGSITLWMGENLQQLLVLHLRSN

Figure 5.1. Sequence (A) Nucleotide and (B) Deduced amino acid sequences obtained from *Musa* accession C_600099504T1 corresponding to a resistance protein and consisting of LRR motifs. The accession is 40% and 32% similar to *Glycine max* accession AF215729 at nucleotide and deduced amino acid respectively. Three primer pairs were designed each from (A) as specific primers and from (B) as degenerate primers. The forward primers (5'-3') and the reverse primers (3'-5') are underlined and coloured as red and blue respectively.

Both specific and degenerate primer combinations were optimized by gradient temperature at six different Ta and Klutuk Wulung was used as DNA template for the PCR amplifications (Table 5.2 and Table 5.3). PCR reactions were performed as described in Chapter 2. Primer orientations of the LRR primers are shown in figure 5.2 and details for both specific and degenerate oligonucleotide primers are in table 5.2. All PCR products were separated on 1.2% agarose gels stained with ethidium bromide. Desired bands were excised and purified using QIAGEN Gel extraction Kits following the manufacture's protocol.

(A) Primers based on DNA sequence of *Musa* accession C_600099504T1.

(B) Degenerate primers based on protein sequence of *Musa* accession C_600099504T1.

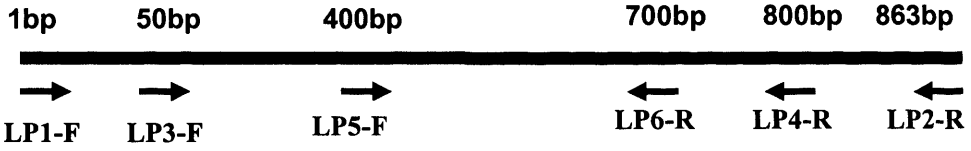


Figure 5.2. (A) Specific primers were designed based on DNA sequence of *Musa* accession C_600099504T1 from *Musa* Sputnik ESTs. (B) Degenerate primers were designed based on deduced amino acid of *Musa* accession C_600099504T1. The accession has 40% LRR motif similarity to clone accession AF215729 (*Glycine max;* De Barros *et al.*, 2000) from EMBO GenBank.

5.2.2 PCR amplification of genomic DNA

Selected *Musa* varieties of A, B, S, and T genome were amplified by using specific LRR primers as described by De Barros *et al.* (2000). Specific LRR primers designed from DNA *Musa* sequence accession C_600099504T1 was arranged in three combinations at different orientations labelled as no.1 (LINT1/LINT2), no. 2 (LINT3/LINT4) and no. 3 (LINT5/LINT6). The primer combinations no.1, no.2 and no.3 were optimised at 63°C, 63°C and 60°C respectively with expected different product sizes at approximately 800bp, 400 bp and 300 bp. The combination of primers with the longest fragment was chosen to study the LRR motifs diversity in *Musa.* After optimised all the primer combinations, genomic DNA from 40 *Musa* varieties was used to isolate LRR motifs by using primer combination of LINT1-F and LINT2-R. The PCR reaction mixtures, PCR cycles and related procedures were described in Chapter 2.

Table 5.2. Specific primers based on *Musa* EST sequence accession C_600099504T1.

No	Primer Name	Sequence (5'—3')	T_m (°C)	T_a (°C)
1	LINT1-F	ATT TGA ACG TAG ATG AAG TAC TAG CAG	61.6	63
2	LINT2-R	GAA ATT CTC TTG GAG GTG TTA TTT CAG	64.7	
3	LINT3-F	TTG TTG TAG ATT TTC TCC CAT CC	62.8	63
4	LINT4-R	TAA AAT TGC AGG GAC GAT GC	64.6	
5	LINT5-F	GTG CCA CTG AGT GAA CAT GG	64.4	60
6	LINT6-R	GGT TCA TTT TGA GAA TCT TAC CAG	61.5	

Table 5.3. Degenerate primers based on *Musa* EST sequence accession C_600099504T1.

No	Primer Name	Sequence (5'—3')	T_m (°C)	T_a (°C)
1	LP1-F	GTN CAY TTY GAR AAY YTN CAN CGN	58.9	No amplification
	LP2-R	YTC NCC CAT CCA NAR NTG NAT NSW NCC	65.3	
2	LP1-F	GTN CAY TTY GAR AAY YTN CAN CGN	58.9	No amplification
	LP6-R	YTC NCC NSW NAR NCG RTT RTT NCC	61.1	
3	LP1-F	GTN CAY TTY GAR AAY YTN CAN CGN	58.9	No amplification
	LP4-R	NGG NAR CAT NTG NGG NAR NGG NCC	62.8	
4	LP3-F	AAR ATN GCN GGN CAN ATG CCN GCN	65.7	51.2
	LP2-R	YTC NCC CAT CCA NAR NTG NAT NSW NCC	65.3	
5	LP3-F	AAR ATN GCN GGN CAN ATG CCN GCN	65.7	53.0
	LP6-R	YTC NCC NSW NAR NCG RTT RTT NCC	61.1	
6	LP3-F	AAR ATN GCN GGN CAN ATG CCN GCN	65.7	46.0
	LP4-R	NGG NAR CAT NTG NGG NAR NGG NCC	62.8	
7	LP5-F	TGG CCN CAY GTN CAY WSN GTN GCN	69.7	55.0
	LP2-R	YTC NC CCA TCC ANA RNT GNA TNS WNC C	65.3	
8	LP5-F	TGG CCN CAY GTN CAY WSN GTN GCN	69.7	55.0
	LP6-R	YTC NCC NSW NAR NCG RTT RTT NCC	61.1	

5.2.3 Cloning and sequence analysis

All fragments were cloned in pGEM-T Easy vector and labelled by an 'L' followed by number, plant code and constitution accordingly. The last digit of the number referred to clone number. Both DNA and deduced amino acids were analyzed in ***CLC Free Workbench 2.2.5*** software for multiple sequences analysis and sequence tree construction as described in Chapter 2. Protein repeats of deduced acid amino were analysed by using Rapid Automatic Detection and Alignment of Repeat (RADAR) available from www.ncbi.uk.com.

5.2.4 Southern hybridization

Restriction enzyme digestions, clone labelling and procedures for Southern hybridization were described in Chapter 2. Clone L44_C4_AA was used as probe.

5.3 Results

5.3.1 PCR primers and optimization

Sequence of accession AF215729 was found homology to *Musa* sequence accession C_600099504T1 in Sputnik with 40% identity at nucleotide and 32% identification at deduced amino acid alignment level.

Amplification of genomic DNA of Klutuk Wulung was optimised at 55°C producing a single band approximately 500bp by using specific primers MLRR-F and MLRR-R. Multiple bands were observed in most PCR products of selected *Musa* varieties which, representing the genomic constitution of A, B, S, and T genome (Figure 5.3).

Comparison amongst the same sample constitutions showed inconsistent patterns. No amplification was found in three samples i.e. ITC0539 (*M. textilis*), ITC1074 (Butuhan) and ITC1238 (Yawa2). The primers clearly showed that samples from *M. balbisiana* had three different patterns; ITC0545 (*M. balbisiana*) and ITC1156 (Batu), ITC1063 (PKW) and ITC1120 (Tani) and ITC0247 (Honduras).

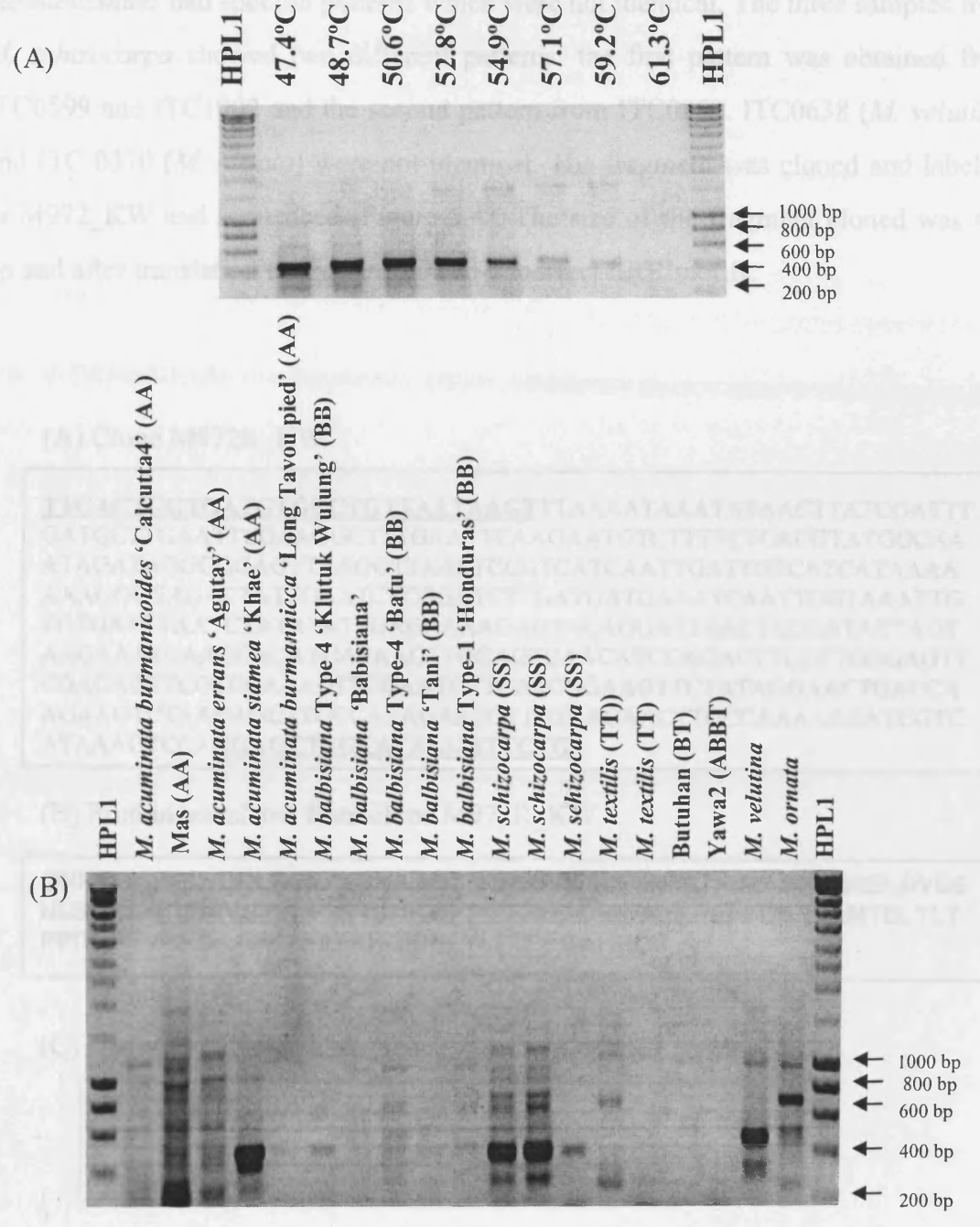


Figure 5.3. PCR product of specific LRR primers as described by De Barros *et al.* (2000). Gel photo (A) shows gradient PCR product amplified from genomic DNA of Klutuk Wulung at different temperatures. Gel photo (B) shows different genomic DNA of *Musa* spp. representing constitution of A, B, S, T and hybrids amplified by the LRR primers at T_a=55°C. HyperLadder 1(HPL1) was used as marker for both gel images.

Samples from *M. acuminata* subspecies *errans, siamea, burmannicca, burmanicoides* had specific patterns which were not identical. The three samples from *M. schizocarpa* showed two different patterns: the first pattern was obtained from ITC0599 and ITC1002 and the second pattern from ITC0852. ITC0638 (*M. velutina*) and ITC 0370 (*M. ornata*) were not identical. The fragment was cloned and labelled as M972_KW and sequenced (Figure 5.4). The size of the fragment cloned was 463 bp and after translation to protein showed imperfect LRR motifs.

(A) Clone M972R_KW

TTGACTGGTGATGTGCCTGTTAATAAGTTTAAAATAAATATAACTTATCGATTT
GATGCTTGAATTCGAAGGCTTTGAATTCAAGAATGTCTTTTCTCACGTATGGCAA
ATAGATAGGGGGAGTTAAGGTTAACTCCGTCATCAATTGATTGTCATCATAAAA
AAAGGGGAGACTATTGAATCTCGGATCTTGATGATGAAATCAATTGGTAAATTG
TGTGATCTAATCTATATATTGAGAAAAGAGTACAGGATTAACTACGATAGTAGT
AAGAAATGAAGCAGATGATAAGTTGGAGTCAACATCGAGACTTCGTTGGGAGTT
CGAGAGTTCGTCGAAAGTTCGAATGTTCATCGGAAGTTCTATAGGAACTGACCA
AGAAGTCTAAGAGCTTGCCAAAGAATCTTGTCAGAACTTGCCAAAAAGATCGTC
ATAAAGTCCA**GGAGCTTGCATAGGATTCTG**

(B) Protein translated from clone M972R_KW

QNPMQAPGLYDDLFGKF*QDSLASS*TSWSVPIELPMNIRTFDELSNSQRSLDVDS
NLSSASFLTTIVVNPVLFSQYID*ITQFTN*FHHQDPRFNSLPFFYDDNQLMTELTLT
PPIYLPYVRKDILEFKAFEFKHQIDKLYLF*TY*QAHHQS

(C) Deduced amino acid sequence in pattern of LRR repeats

PGLYDDlfgkfqDSLASSTSWSVPIELP
PFFYDD.....NQLMTELTLTPPIYLP

Figure 5.4. Sequences of LRR fragments amplified from Klutuk Wulung genomic DNA. (A) DNA sequence of 463bp obtained from clone M972_KW using the combination primers MLRR-F and MLLR-R. Both reverse and forward primers were underlined. (B) Protein translated from clone M972_KW shows LRR motifs. (C) Two repeats of LRR motifs observed at deduced amino acid sequence analyzed by RADAR.

In targeting more LRR motifs, degenerate LRR primers were tested on eight primer combinations labelled as no. 1 (LP1-F/LP2-R), no. 2 (LP1-F/LP6-R), no. 3 (LP1-F/LP4-R), no. 4 (LP3-F/LP-2-R), no. 5 (LP3-F/LP6-R), no. 6 (LP3-F/LP4-R), no. 7 (LP5-F/LP2-R) and no. 8 (LP5-F/LP6-R). No PCR amplification was found in combination no. 1, no. 2 and no. 3. Multiple bands were observed in primer combinations for no. 5, no. 6, no. 7 and no. 8. The annealing temperatures were optimized at 53.6°C, 46.4°C, 55.9°C and 55.9°C respectively. Combination no. 4 produced a single band of the PCR product at approximately 400 bp and optimised at 51.2°C. PCR optimization based on gradient annealing temperature is shows in figure 5.5. As the degenerate primer combinations resulted in multiple bands differing from the expected size, they were not reliable in isolating the LRR motifs as they might be pseudogenes or random genomic regions.

5.3.2 Isolation and cloning of LRR motifs

LRR motifs for *Musa* varieties were isolated by using the specific primer pair of LINT1-R and LINT2-R which produced a single PCR product approximately at 820 bp in most accessions. An additional PCR product was obtained from *M. acuminata* subsp. *errans* at approximately 1500bp (Figure 5.6).

The extra fragment of *M. acuminata* subsp. *errans* at 1500bp was labelled as 1L20. Two selected clones were sequenced and labelled as 1L201 and 1L202. The clones sequenced were readable only up to 700 bp and consisted of either forward or reverse primer flanking the insert. These results were expected as the sequencer has its limitation to sequence readable of less than 800 bp. It means that only outer regions of the fragment (1500 bp) were sequenced and a primer walking strategy was used: other specific primers were designed based on the sequence from clones 1L201 and 1L202 in order to amplify the whole 1500bp fragment. The primers were labelled as LINT7-F, LINT8-R and LINT9-R (Figure 5.7).

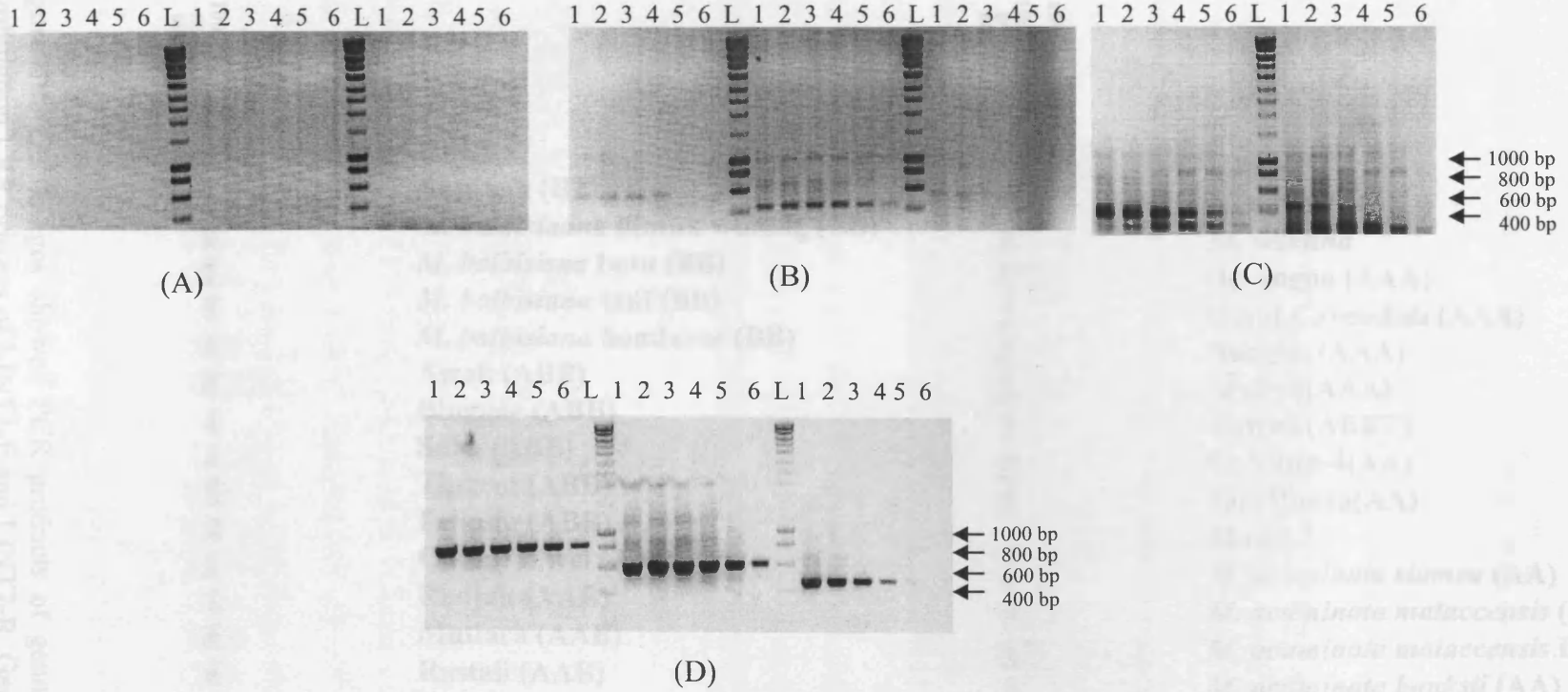


Figure 5.5. Gradient PCR at different temperature of eight pairs of degenerate primers and three pairs of specific primers using genomic DNA of Klutuk Wulung. Line 1 to 6 in gel photo (A), (B), and (C) show gradient temperature of 46.4°C, 48.8°C, 51.2°C, 53.6°C, 55.9°C and 58.1°C. Line 1 to 6 in gel photo (D) show gradient temperature of 59.0°C, 60.2°C, 61.4°C, 62.8°C, 63.8°C, 65.0°C. HyperLadder 1 (L) is used as marker.

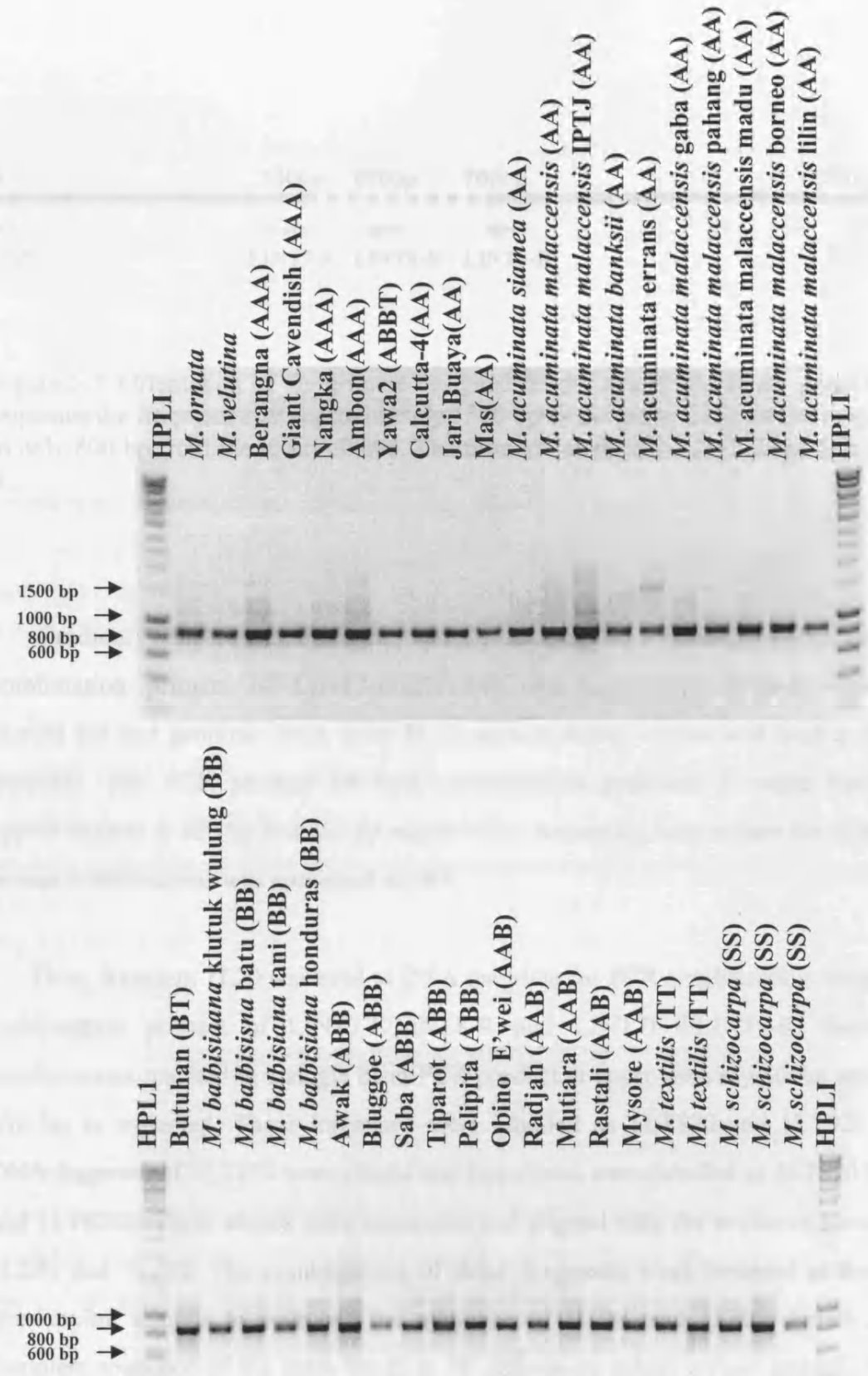


Figure 5.6. Gel images showed PCR products of genomic DNA amplified by combination LRR primers of LINT1-F and LINT2-R. Genomic constitutions were put in bracket. All bands at approximately 820 bp and 1500 bp (*M. acuminata* subsp *errans*) were excised, cloned in pGEM-T Easy vector and sequenced. HyperLadder I (HPL 1) was used as a marker at both sides of the gel.

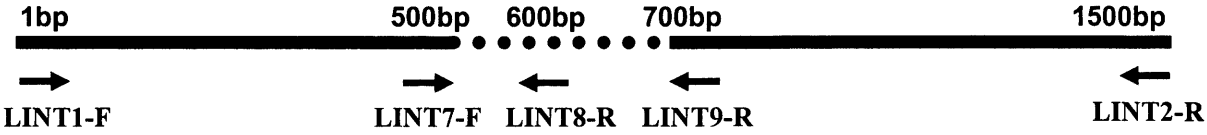


Figure 5.7. Orientation of the primers designed for *M. acuminata* subsp. *errans* to sequence the fragment size approximately 1500 bp as the sequencing can be read up to only 800 bp from the insert cloned. The primers labelled as LINT-7 and LINT8-R.

Gradient temperature PCR at 50.1°C, 52.7°C, 55.3°C and 57.9°C for combination primers of LINT7-F/LINT8-R and LINT7F-F/LINT9-R were carried out and genomic DNA from *M. acuminata* subsp. *errans* was used as a template. The PCR product for both combinations produced a single band approximately at 400 bp and 450 bp respectively. Annealing temperature for both primer combinations was optimised at 58°C.

Then, fragment 1L20 was used as DNA template for PCR amplification using combination primers of LINT7-F/LINT8-R and LINT7F-F/LINT9-R. Both combinations resulted in a single band PCR product at approximately 400bp and 450 bp as expected. These fragments were labelled as 1LT820 and 1LT920. DNA fragment of 1LT820 were cloned and two clones were labelled as 1LT8201 and 1LT8202. These clones were sequenced and aligned with the sequence from 1L201 and 1L202. The combinations of these fragments were trimmed at the overlapping regions to complete the sequence of 1L201 and 1L202. Finally, complete sequence of the extra bands in *M. acuminata* subsp. *errans* amplified by LINT-1F and LINT2-R was determined at 1494 bp. These sequences have the same size and were labelled as 1LT201_8201, 1LT201_8202, 1LT202_8201 and 1LT202_8202.

5.3.3 Sequence analysis

A total of 202 clones were obtained from 40 fragments of *Musa* varieties (including two clones which were sequenced separately at 1500 bp) amplified by LINT1-F and LINT2-R primers. The clones were strictly selected based on M13 amplification followed by *Eco*RI enzyme digestion to confirm the insert size prior to sequencing.

In total, 121 clones with approximately 820bp were sequenced. Seven imperfect clones were found without any flanking primer to the insert sequence and 114 clones were determined as perfect sequence with both reverse and forward primers flanking to the insert fragment. Another four perfect sequences were obtained from the clones combinations as described in 5.3.1 (1LT201_8201, 1LT201_8202, 1LT202_8201 and 1LT202_8202).

The single size of the PCR product for the perfect nucleotides in each clones varied between 838bp to 874bp. A pie chart showed 12 different sizes of LRR nucleotide in percentages based on total number of the clones sequenced (Figure 5.8), with 38% of clones being 864 bp. A specific individual nucleotide size was observed in Calcutta-4 (L44_C4_AA 838 bp), Awak (L11_AW_AAB 871 bp), Radjah (L121_RJ_AAB 873 bp) and Gaba (L341_GB_AA 878 bp). The imperfect clones were obtained from *M. balbisiana* 'Klutuk Wulung', *M. acuminata* subsp. *banksii*, *M. schizocarpa* (two clones), Tiparot, Mysore and Madu.

Fragments isolated from the same species showed numerous insertions and/or deletions in their sequence. It revealed that allelic differences in LRR motifs were complicated and their relationship amongst the *Musa* varieties was ambiguous. All the 118 sequences were aligned in ***CLC Free Workbench 2.2.5*** together with both forward and reverse primers flanking onto the sequence. As the view size of the multi alignment occupied big space, it was reduced to identify positioning of the sequence domains (Figure 5.9).

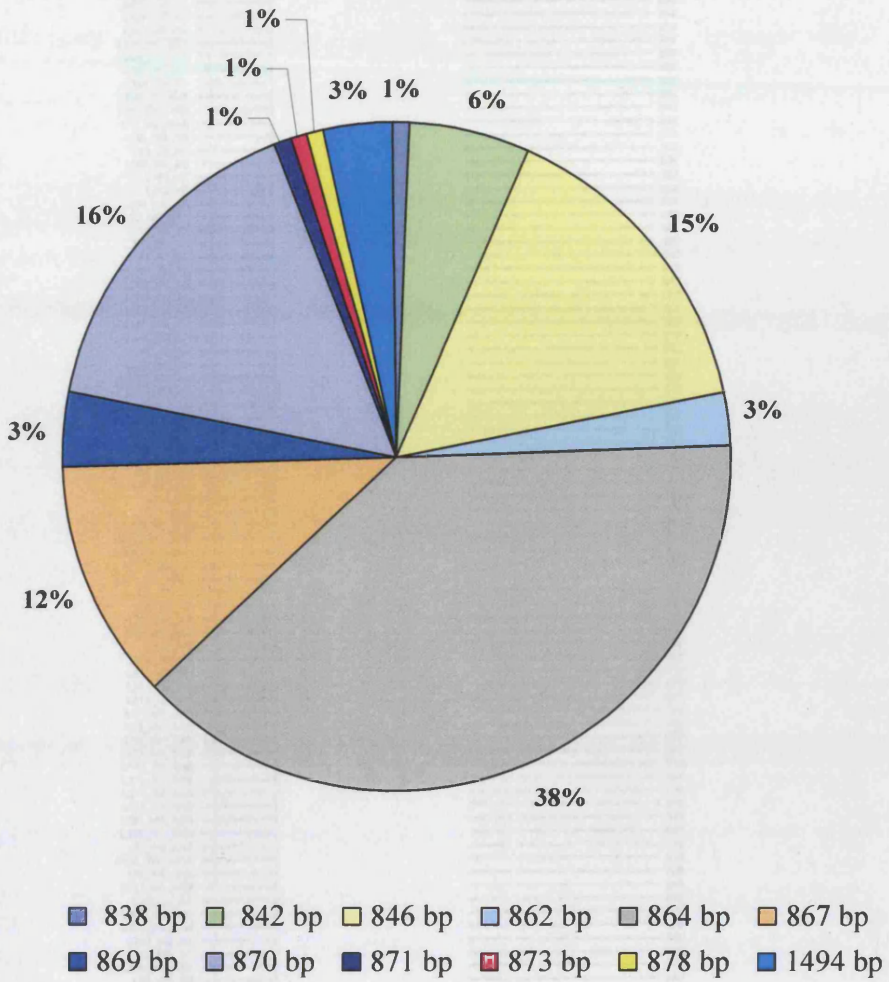


Figure 5.8. Percentage of 121 clones of LRR motifs based on nucleotide size amplified from DNA genomic of *Musa* varieties by using combination primer LINT1-F and LINT2-R. Sequence size of the clones were shown in the legend differentiate by colour.

The longest sequence of 1494 bp was derived from *M. acuminata* subsp. *errans* and a 631 bp fragment was not aligned to other shorter sequences. Excluding the non-aligned fragment, similarity amongst the clones ranged between 85-100%, putting an upper limit of 15% on the number of sites responding to selection. The lowest similarity was 85% between *M. acuminata* subsp. *errans* and *M. textilis* (ITC 1072).

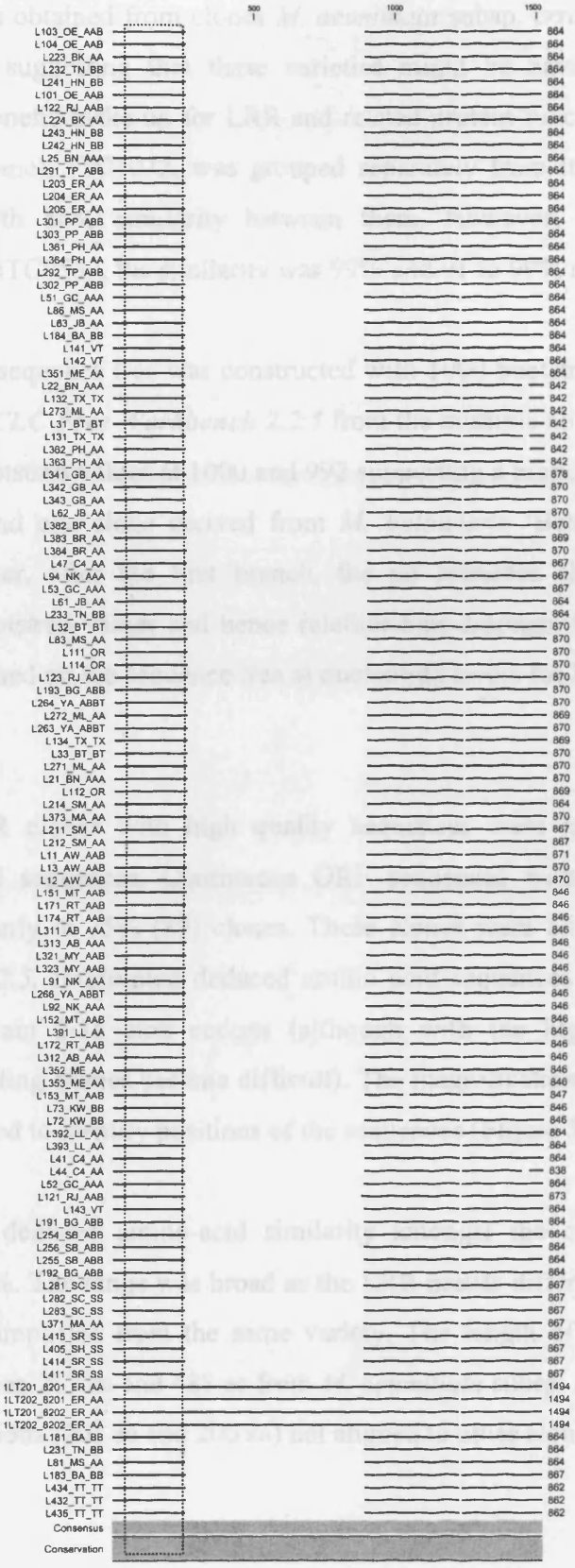


Figure 5.9. Nucleotide alignments of 118 clones consisting of LRR motifs. Sequences in box are conserved regions related to extracytoplasmic LRR which facilitates recognition of an extracelullar ligand.

All sequences obtained from clones *M. acuminata* subsp. *errans* and Gaba were 100% similar suggesting that these varieties might be having the most ancestral common genetic make up for LRR and related protein functions. One of *M. textilis* (TT genome), ITC1072, was grouped separately from its relative *M. textilis* ITC0539 with 86% similarity between them. However, amongst the ITC1072 clones and ITC0539, the similarity was 99% and 91 to 99% respectively.

An unrooted sequence tree was constructed with 1000 bootstraps using the UPGMA method in ***CLC Free Workbench 2.2.5*** from the multiple alignments. The tree showed high bootstrap values at 1000 and 992 supporting a branch with the *M. textilis* (ITC1072) and one clone derived from *M. balbisiana* 'Batu' (ITC1156) respectively. However, after the first branch, the no branches showed strong support from the bootstrap values and hence relationships amongst the clones are unclear. Diversity based on the sequence tree at nucleotide levels for LRR motifs is shown in figure 5.10.

All 118 LRR clones with high quality sequences were translated into deduced amino acid sequences. Continuous ORF sequences with LRR repeat motifs were found only in 75% (89) clones. These clones were aligned in ***CLC Free Workbench 2.2.5.*** Interrupted deduced amino acid sequences in 29 clones were found to contain 2-15 stop codons (although with the higher numbers determination of reading frames became difficult). The magnification of the multi-alignment was reduced to identify positions of the sequences (Figure 5.11).

Overall, the deduced amino-acid similarity amongst the clones ranged between 28 and 100%. The range was broad as the LRR motifs differed even from the same fragment amplified from the same variety. The length of the deduced peptide ranged between 267 aa and 488 aa from *M. acuminata* subsp. *errans* clones which had two insertions (201 aa and 200 aa) not aligned to other clones.

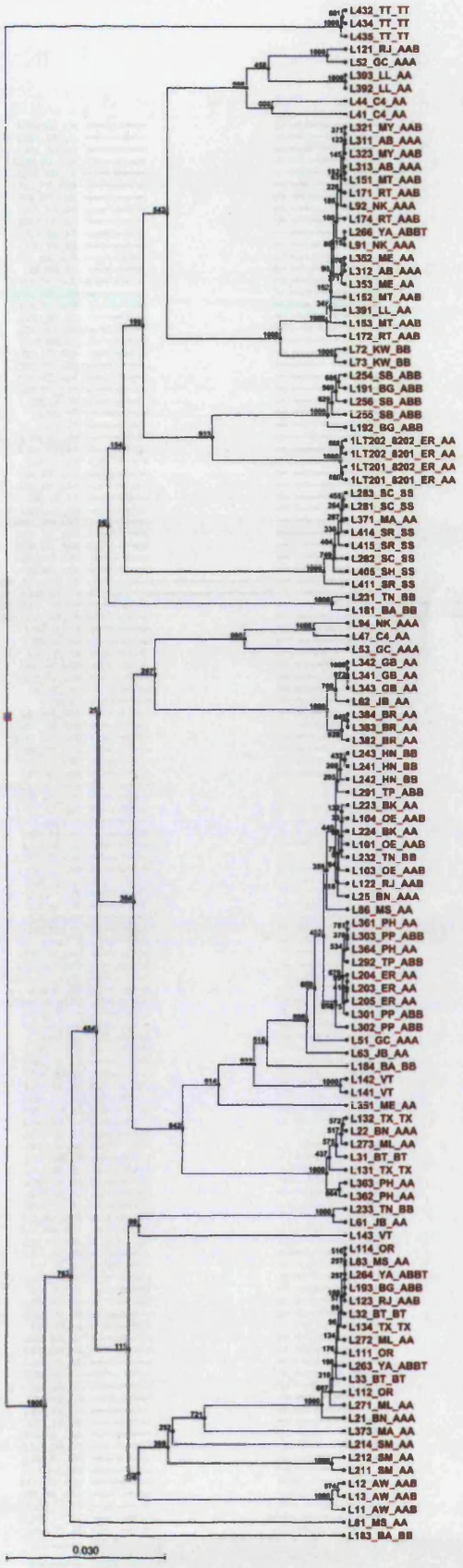


Figure 5.10. Sequence comparison model of unrooted tree based on alignment of nucleotide sequences of *Musa* LRR motifs. The tree was constructed by using UPGMA method provided in ***CLC Free Workbench 2.2.5*** with bootstrapping of 1000 replicates (numbers near the branches). Branch lengths are proportional to the average amino acid substitutions per site as indicated by the scale.

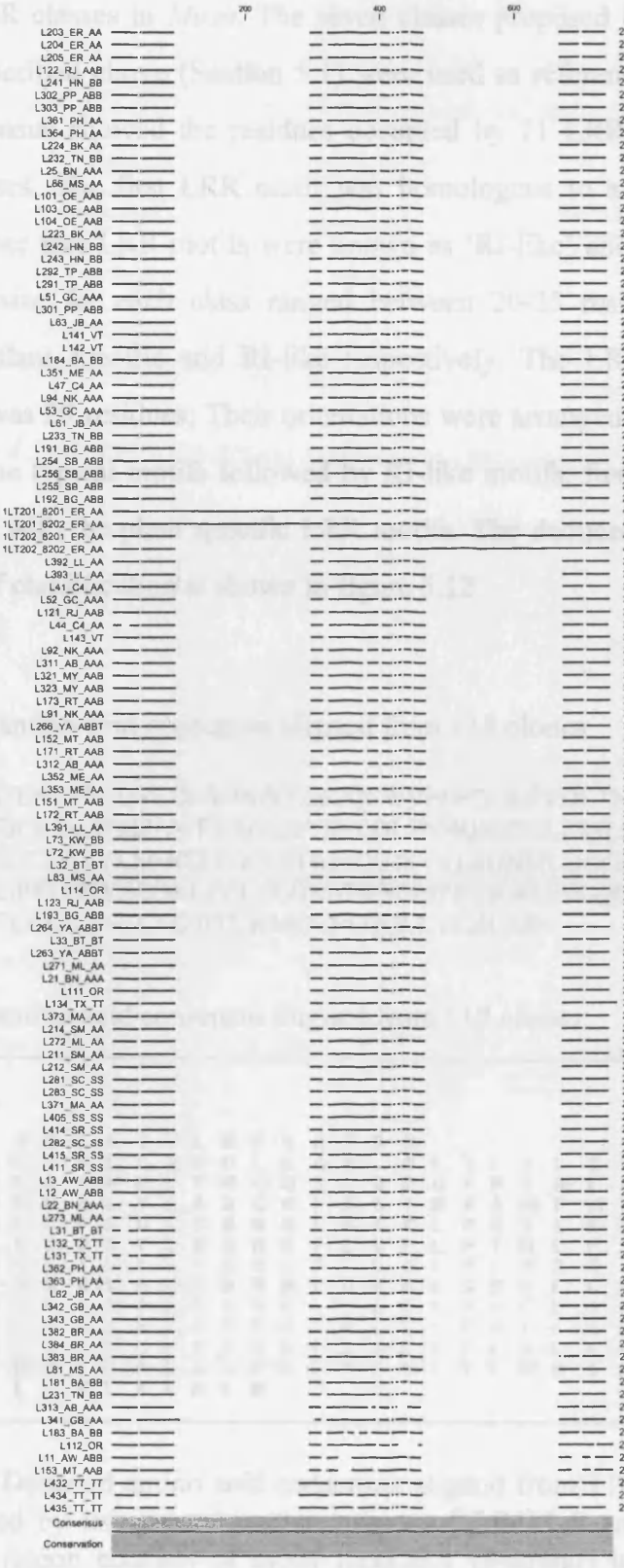


Figure 5.11. Deduced amino acid sequence alignments of 118 clones consisting of LRR motifs.

A multiple alignment consensus obtained from 118 clones was used to determine LRR classes in *Musa*. The seven classes proposed by Enkhbayar *et al.* (2003) described above (Section 5.1) were used as reference. The deduced peptide consensus showed the residues occupied by 11 LRR motifs in three different classes. The first LRR motif was homologous to a class known as typical. Another five LRR motifs were known as 'RI-like' and 'plant specific' classes. The size for each class ranged between 20-25 residues and 23-26 residues for plant specific and RI-like respectively. The LRR motif for the typical class was 29 residues. Their orientations were arranged accordingly and started with the typical motifs followed by RI-like motifs, fused in the middle before ending with the plant specific LRR motifs. The deduced amino acid for the LRR motif classification is shown in figure 5.12.

(A) Deduced amino acid consensus aligned from 118 clones

NSLGGVISEVHFENLTRLQVLDLSDNPITISIGQSWVPPFQLRYVDLTNCQLGPQFPEWLQ
FQTQIEELYLADCKIAGTMPAWFWNISSSTIIFLDLSNNQIGGKLPSSLKFTKLEILYLDSN
RFEGPLPTMLPSTLDTLYLSNNSFTGQLPIWPHVHSVALSDNMLDGGLSSSICQWTYLEY
LDLSNNKLLGEIPYCLGKSLQNLYFLNLDNNHFSGEIPHTIGFLSELQLLQLKNNSFSGEVP
LSLKNCTNLWFLDLAQNNLVGSITLWMGENLQQLLVLHLRSN

(B) Deduced amino acid consensus aligned from 118 clones

LRR Class	Repeat No	N	S	L	G	G	V	I	S	E	V	H	F	E	N													
Typical	1	L	T	R	L	Q	V	L	D	L	S	D	N	-	P	I	T	I	S	I	G	Q	S	W	V	P	P	Q
RI-like	2	L	R	Y	V	D	L	T	N	C	Q	L	-	G	P	Q	F	P	E	W	L	Q	F	Q	T	Q		
RI-like	3	I	E	E	L	Y	L	A	D	C	K	I	A	G	T	M	P	A	W	F	W	N	I	S	S	S	T	
RI-like	4	I	I	F	L	D	L	S	N	N	Q	I	G	G	K	L	P	S	S	L	K	F	T	K				
RI-like	5	L	E	R	L	Y	L	D	S	N	R	F	E	G	P	L	P	T	M	L	P	S	T					
PS	6	L	D	T	L	Y	L	S	N	N	S	F	T	G	Q	L	P	I	W	P	H							
RI-like	7	V	Q	S	V	A	L	S	D	N	M	L	D	G	G	L	S	S	S	I	C	Q	W	T	Y			
PS	8	L	E	Y	L	D	L	S	N	N	K	L	L	G	E	I	P	Y	C	L	G	K	S	L	Q	N		
PS	9	L	Y	F	L	N	L	D	N	N	H	F	H	G	E	I	P	H	T	I	G	F	L	S	E			
PS	10	L	Q	L	L	Q	L	K	N	N	S	F	S	G	E	V	P	L	S	L	K	N	C	T	N			
PS	11	L	W	F	L	D	L	A	Q	N	N	L	V	G	S	I	T	L	W	M	G	E	N	L	Q	Q		
		L	L	V	L	H	L	R	S	N																		

Figure 5.12. (A) Deduced amino acid consensus aligned from 118 clones of *Musa* varieties amplified by using combination primer of LINT1-F and LINT2-R. (B) LRR conserved region consists of motif LxxLxLx (β-strand) in the first seven residues followed by other residues (α-strand) which form a complete motif of LRR. The consensus sequence consists of 11 LRR motifs under three LRR classes known as Typical, RI-like (Ribonuclease Inhibitor) and Plant specific (PS). The top and the last sequences are related to primers and are not aligned to the LRR motifs.

Based on the percentage of the deduced peptide length, 17 different sizes of fragments were identified as presented by a pie chart (Figure 5.13), more than the 11 nucleotide classes. Size of 287 aa was the highest percentage (38%) consists of 39 clones. Single clones were obtained from Radjah (L121_RJ_AAB 291 aa), *M. textilis* (L134_TX_TT 301 aa), Borneo (L383_BR_TT 283 aa), Ambon (L313_AB_AAA 275 aa), Gaba (L341_GB_AA 290 aa) and Mutiara (L153_MT_AAB 267 aa).

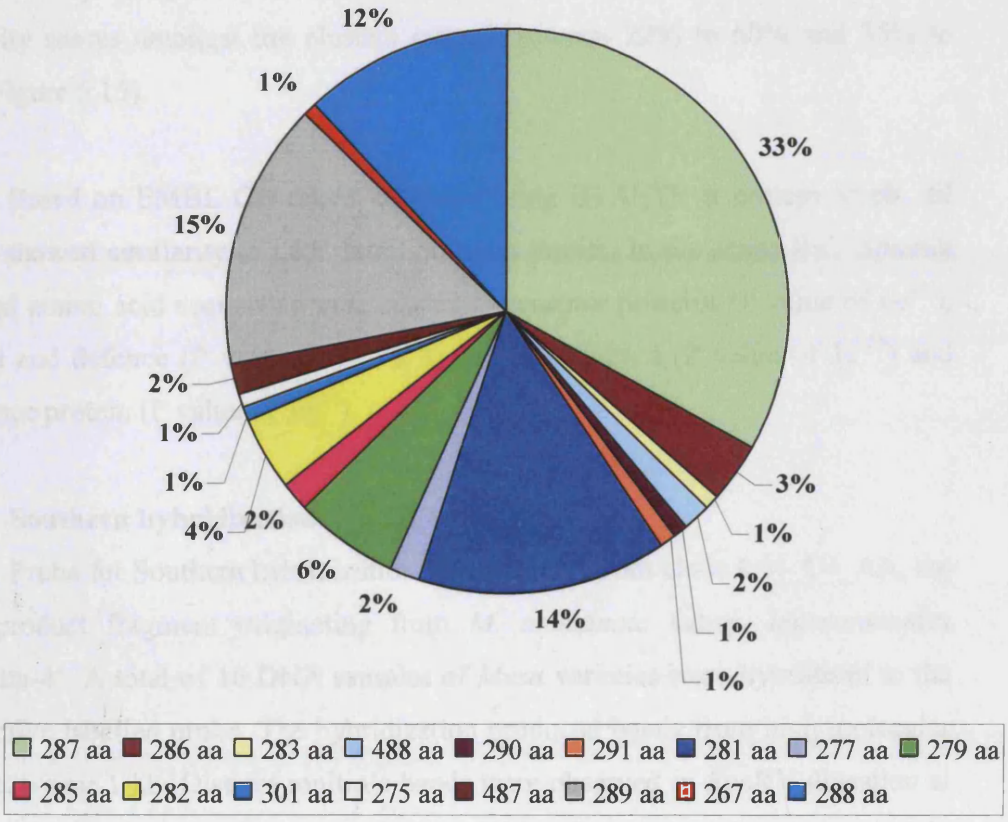


Figure 5.13. Pie chart showing percentage of size (deduced amino acid residues) consisting of LRR motifs obtained from DNA sequences of *Musa* varieties clones amplified by using primer combination LINT1-F and LINT2-R. The size of amino acid residues were shown in the legend differentiated by colour.

An unrooted sequence tree with 1000 bootstraps was constructed by using the UPGMA method provided in ***CLC Free Workbench 2.2.5*** from the multiple alignments. Four clusters supported with high bootstrap value were identified. The first cluster consisted of hybrid samples, Mutiara (L53_MT_AAB) and Awak (L11_AW_ABB). A single clone of *M. balbisiana* 'Batu' (L183_BA_BB) was determined as a second cluster. The third cluster consisted of 112 clones and the fourth cluster was only clones from *M. textilis* (L432_TT_TT, L434_TT_TT and L435_TT_TT). The clusters did not correspond to the *Musa* genomic constitution or the ploidy level. Diversity of LRR motifs at deduced amino acid levels was shown in figure 5.14.

Sequences in each cluster were aligned separately and a sequence tree was constructed by using UPGMA method from the cluster consensuses. Identity and similarity scores amongst the clusters ranged between 22% to 60% and 35% to 70% (Figure 5.15).

Based on EMBL GeneBank database using BLASTP at protein levels, all clones showed similarity to LRR family protein motifs. In the *Musa* EST Sputnik deduced amino acid sequences were related to receptor proteins (P value of $6e^{-75}$), disease and defence (P value of $5e^{-25}$), signal transduction (P value of $1e^{-23}$) and resistance protein (P value of $3e^{-24}$).

5.3.4 Southern hybridization

Probe for Southern hybridization was obtained from clone L44_C4_AA, the PCR product fragment originating from *M. acuminata* subsp. *burmanicoides* 'Calcutta-4'. A total of 16 DNA samples of *Musa* varieties were hybridised to the radioactive labelled probe. The hybridization produced bands from high molecular weight to near 1 kb. Distinct multiple bands were observed in *Eco*RV digestion at high molecular weight between 1kb to 10kb. The probe detected multiple polymorphic regions in *Musa* varieties. However identical patterns were observed in Mutiara (AAB) and Rastali, where the former originated from the later variety following mutagen treatment. A single band with a same length was observed in the *Eco*RI digestion. However, a smear of hybridization was observed in *Hin*dIII showing the probe was bound to many sites in the genome (Figure 5.16).

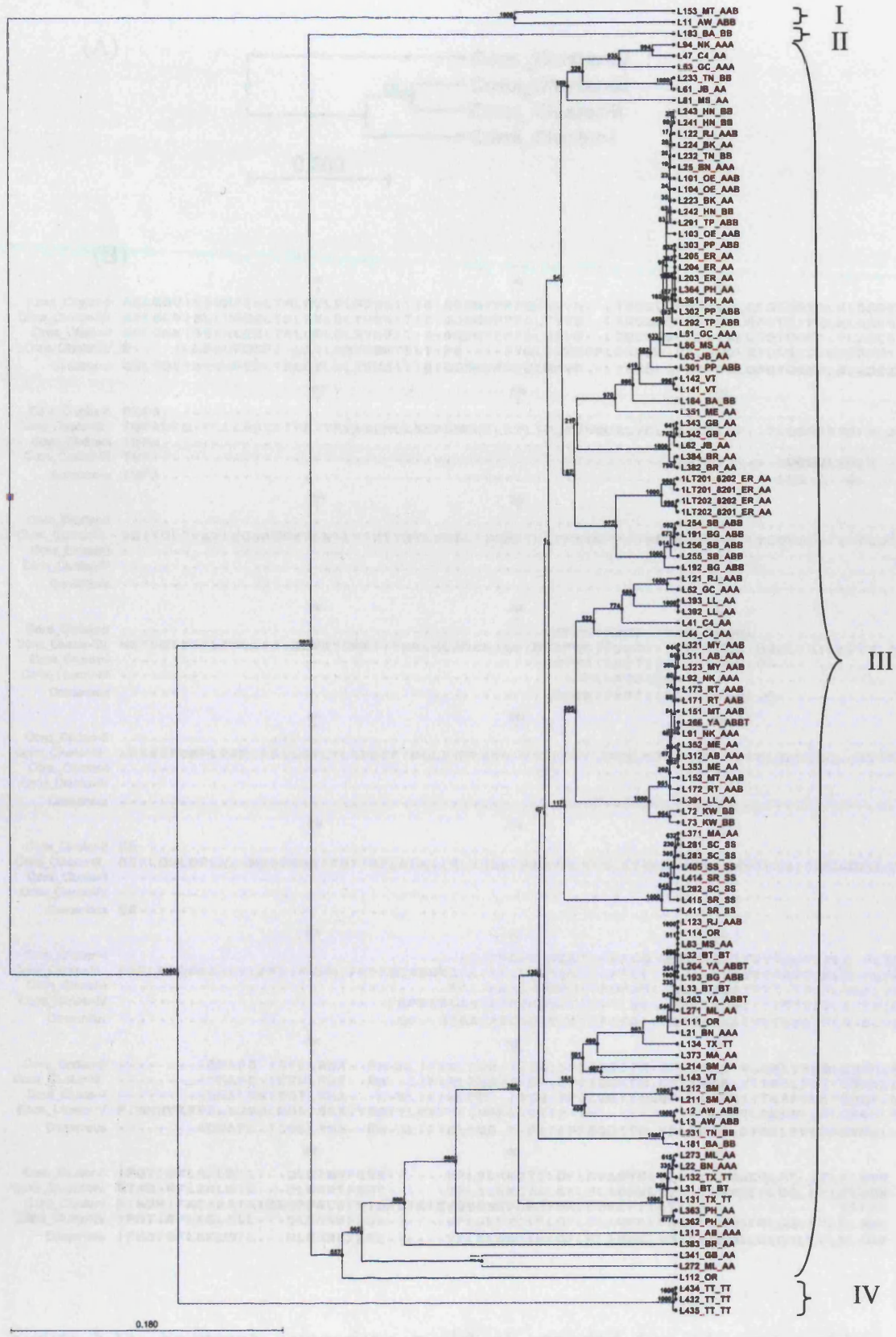
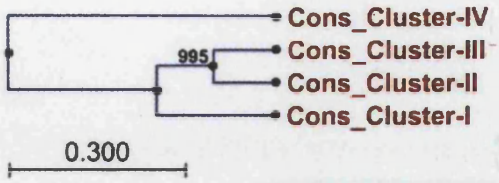


Fig 5.14. Sequence comparison model of unrooted tree (A) and multiple alignments (B) of deduced amino acid sequences obtained from LRR motifs of *Musa* consensuses. The tree was constructed by using UPGMA method with bootstrapping of 1000 replicates (numbers near the branches).

(A)



(B)

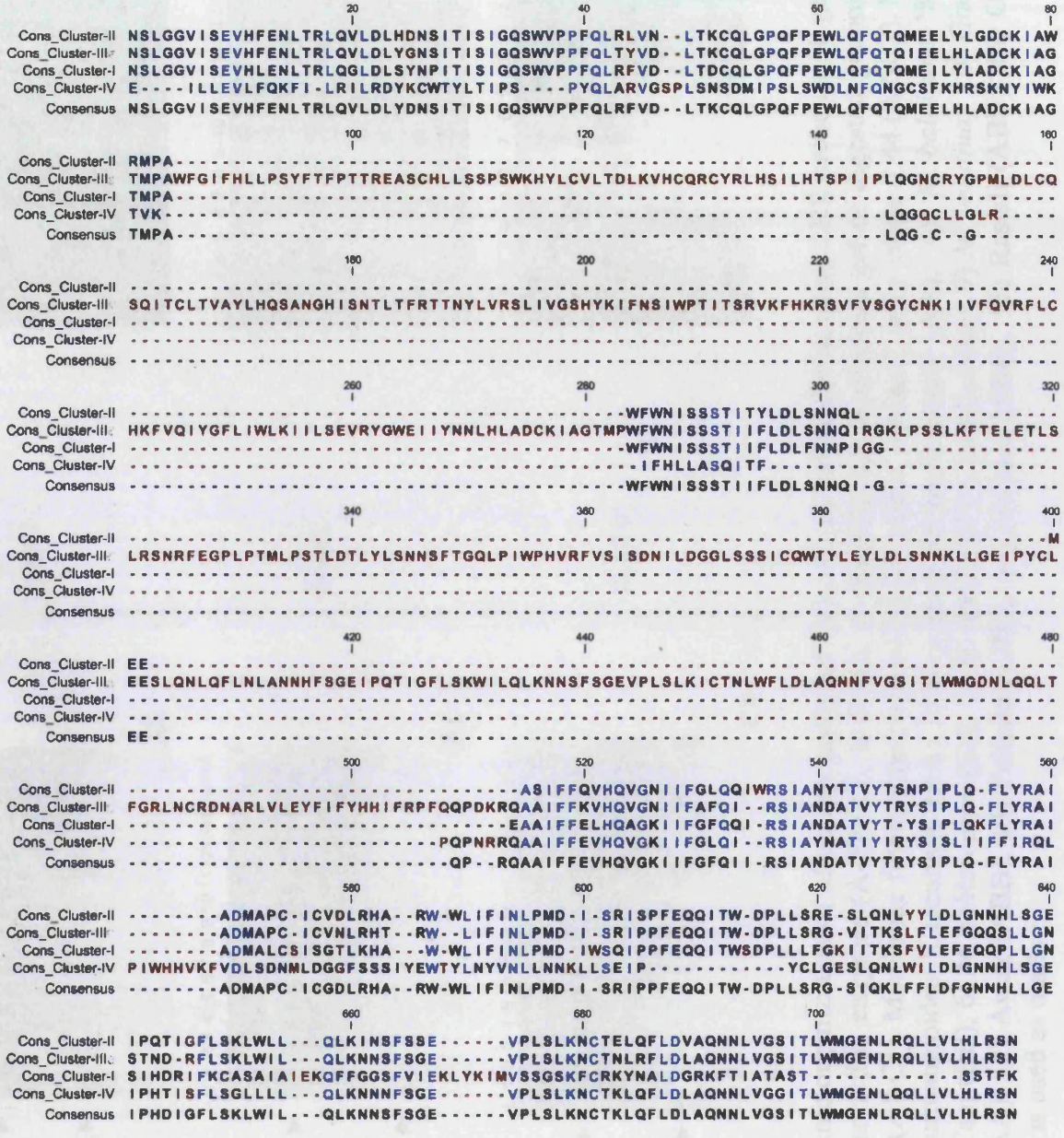


Figure 5.15. Sequence comparison model of unrooted tree (A) and multiple alignments (B) of deduced amino acid sequences obtained from LRR motifs of *Musa* consensuses. The tree was constructed by using UPGMA method with bootstrapping of 1000 replicates (numbers near the branches). Branch lengths are proportional to the average amino acid substitutions per site as indicated by the scale. Both tree and the alignment were provided by ***CLC Free Workbench 2.2.5.***

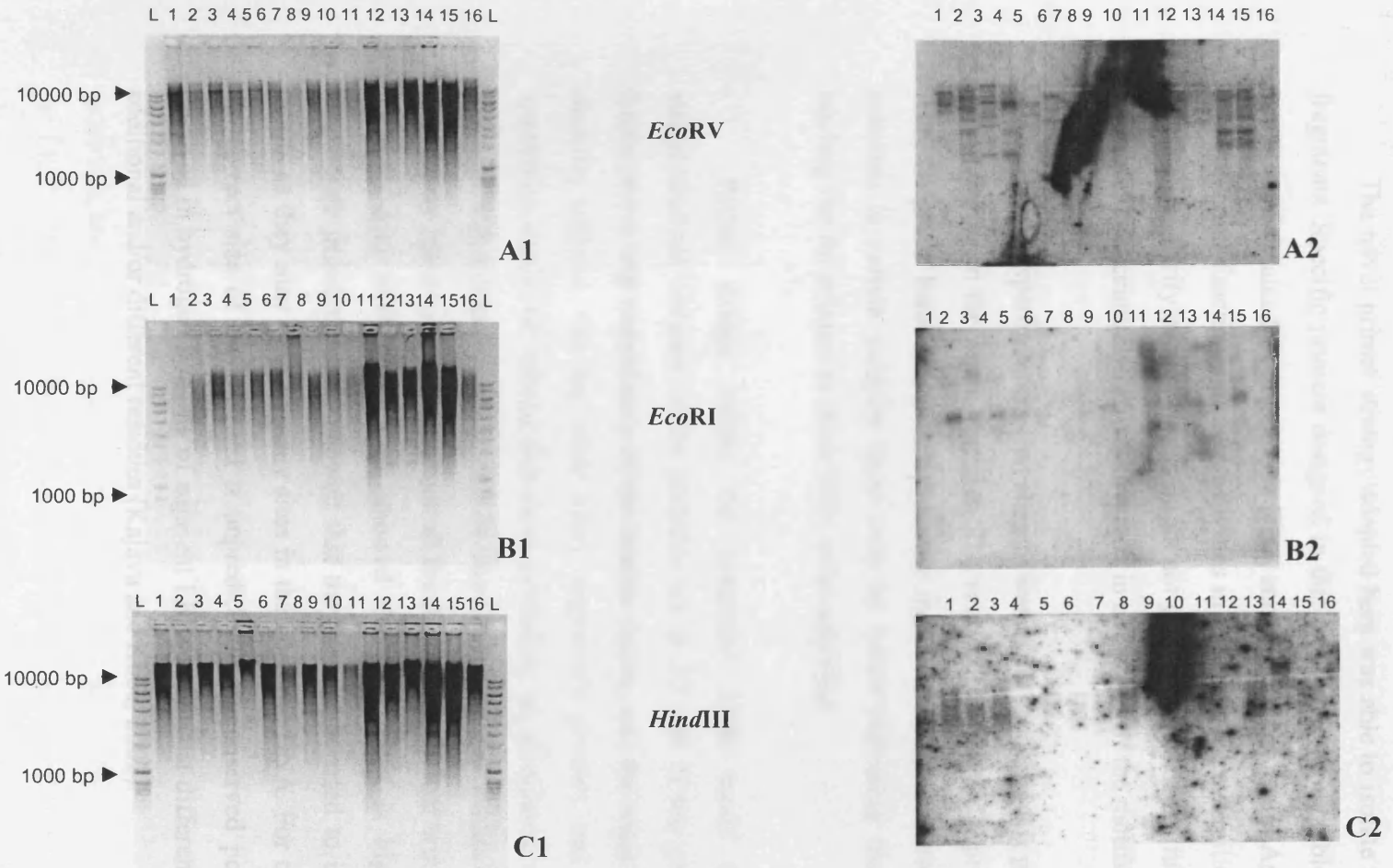


Figure 5.16. Genomic organization of *Musa* spp. analyzed by southern hybridization using LRR probe on 16 genomic DNA digested with 3 different restriction enzymes (A-*Eco*RV, B-*Eco*RI, C-*Hind*III). Electrophoresis gel of digested genomic DNA stained with ethidium bromide (A1-C1). Membrane filters hybridized with LRR probe obtained from clone L44 (A2-C2). Sample 1) Mas (AA), 2) *M. acuminata* burmanicoides 'Calcutta-4' (AA), 3) *M. acuminata* 'malaccensis' (AA), 4) . *M. balbisiana* 'Klutuk Wulung' (BB), 5) *M. balbisiana* Tani (BB), 6) *M. schizocarpa* (SS), 7) *M. textilis* (TT), 8) *M. ornata*, 9) *M. velutina*, 10) Giant Cavendish (AAA), 11) Berangan (AAA), 12) Awak (ABB), 13) Pelipita (ABB), 14) Mutiara (ABB), 15) Rastali (ABB), 16) Obino L'Ewai (AAB). HyperLadder 1(L) is used as marker.

5.4 Discussion

5.4.1 Primer limitations

The novel primer strategy adopted here was able to isolate a range of LRR fragments. Specific primers designed in this work were remarkable as only single bands were obtained from genomic DNA amplification in *Musa*. As the LRR motif is involved in functional protein activities in large sequence families, it is possible that the specificity of the primers might not give a picture of the full range of LRR diversity, but accurately reflects diversity in one group of the motifs.

In the experiments, a first attempt was using specific LRR primers designed from *Glycine* in the *Musa* varieties. The results from the PCR product showed a wide range of band sizes, and even the duplicate samples showed inconsistent patterns; in certain samples there were no bands indicating that there was no binding site for primers to allow DNA to be amplified.

Primer design within the conserved LRR motif itself was not straightforward because of the probable six to 20 loci in the genome including LRRs, the 6-fold redundancy of the leucine-codon, and the weak conservation of flanking regions. On the other hand, degenerate primers had a tendency to amplifying non-LRR related sequences (including an abundance of single-primer products with unknown relationship to true LRR loci). The results indicate that the primers are specific enough that not all the different LRR loci would be amplified although data mining for LRR showed that this motif is highly conserved. Degenerate primers clearly showed that they cannot be used to isolate the LRR motifs as they may bind to many sites in the genomic DNA. For the LRR motifs, the correct size of the α-strand is unpredictable as conserved polar side chains involved in hydrogen bonding of adjacent LRRs may twist differently ending with additional and/or different residues (Kajava and Kobe, 2002).

5.4.2 LRR motifs

As described in Chapter-1, LRR occurs in many different proteins, including tyrosine kinase receptors, virulence factors, extracellular matrix-biding glycoproteins, disease resistance and associated signal transduction proteins (www.pfam.wustl.edu).

Prediction model for LRR is that the common ancestor of the *Musa* species analysed here has multiple independent LRR genes (loci). Each one of these genes will have had several alleles in the ancestor, and will have changed (mutated) during evolution of the current *Musa* species. Furthermore, some genes with similar functions may be duplicated, and genes may have been lost or gained during evolution of the current species or become pseudogenes. The challenge of the analysis here is to distinguish these three alternatives: different LRR loci with different functions; evolutionary and allelic differences in LRR at one locus; and new/lost LRR loci.

PCR and Southern hybridization showed that the LRR motifs are present in a relatively high copy number but exact estimation is not possible as the sequences diverge considerably. The sequence tree showed that the copies are present in more than one structure, based on nucleotide sequences, as revealed in Radjah (AAB), Butuhan (BT), Mas (AA), Calcutta-4 (AA), BA (BB), *M. velutina, M. ornata*, Yawa2 (ABBT), Blugoe (ABB) and Berangan (AAA). These results showed the existence of LRR proteins in different subfamilies that can simultaneously occur independently and/or at irregularly spaced for patterns.

LRR motifs isolated from this work consist of 75% perfect ORFs indicating that they might correspond to structural elements associated with protein-protein interaction. LRR motif arrangements in *Musa* sequences were found to match to more than one of the LRR classes described by Kajava (1998). Thus, it is clear that the LRR motifs of one class are not necessarily similar in their repeat sequence as the motif in each class can be integrated due to similar repeat motifs. Further, each LRR motif might be involved with and/or support many functional proteins which require certain copy numbers or repeats.

Thus it is not possible that functional LRR motifs occupy more then one class as found in *Musa.* These situations suggest that the LRR motifs have emerged independently during evolution. It agrees with Kajava (1998) that the repeats from different classes retain a similar super helical form, but differed in three dimensional structures of individual repeats. However, the mechanisms of insertion and duplication were not clear and could be related to retroelements which have capability for sequence amplification. Thus, it is a challenge, and the finding of multiple families of *Musa* LRR motifs is of considerable interest.

Both typical and plant specific classes of LRR motifs are extracellular multi-domain membrane proteins which share common functions in signalling, and most of them mediate plant resistance to pathogen attack (Jones *et al.*, 1997; Jones and Jones, 1997). Surprisingly, the LRR motif of the RI-like class with known binding capability to ribonucleases (RNases), and which is commonly found in animals, also exists in *Musa.* It was revealed that the RNase not only catalyzes RNA degradation but also intervenes in angiogenesis, cytotoxicity and the host defence response (Leland and Raines, 2001). Activities of RNase can be found in the vacuole, extracellular space, endoplasmic reticulum and plastids, and influences gene expression levels during development and in response to biotic and abiotic stresses (Green, 1994; Booker, 2004).

It was found in *Zinnia elegans*, *Arabidopsis* and tomato leaves that an RNase gene was expressed, and multiple enzymes activities increased in response to pathogen and wounding (Ye and Droste, 1996; Lers *et al.,*1998; LeBrasseur *et al.,* 2002). Booker (2004) supported the suggestion that RNases might be involved in different processes, although some might be overlapping in their functions (Green, 1994, Ye and Droste, 1996). Multiple activities and their overlapping functions found in the RNAse, which also are related to the LRR, are suggested as corresponding to both abiotic and biotic specificity in defence protein functions.

Detailed analysis of the LRR classes and the function of the various isolated fragments will be an interesting part for future research, as they may affect the protein functions. In particular, the RNase activity and perhaps specificity would be very exciting to explore further to show if it has any relationship to disease resistance. It is suggested that the α-strand in the typical class of the LRR motifs evolves to match particular functions; however, structural information comparing both classes is currently not available.

In relation to disease resistance genes, the genes include either extracellular LRR motifs or cytoplasmatic LRR motifs (Jones and Jones, 1997). Multiple LRRs at the C-terminus were found in a cytoplasmic LRR comprised of R-genes in *N* from tobacco (Whitham *et al.,* 1994), *L6* from flax (Lawrence *et al.,* 1995) and from *Arabidopsis,* including *RPS2, RPM1* and *RPP5* (Bent *et al.*, 1995; Grant *et al.*, 1995; Parker *et al.,* 1997). Another class of disease resistance genes, *Cf,* consists of *Cf-2, Cf-4, Cf-5* and *Cf*-9, all genes encoding membrane-anchored proteins composed of extracellular LRR motifs. Comparison between the *Cf-2* and *Cf-5* gene revealed that the LRR motifs were extensively shared particularly within the C-terminus regions of their predicted proteins; the *Cf-5* disease resistance gene clearly showed allelic variation in the LRR copy numbers (Dixon *et al.*, 1998).

The LRR motif ORF found in *Musa* showed homology to ethylene-inducing xylanase (EIX) receptor 1 and 2 in *Solanum* (*Lycopersicon) esculentum* (accession AAR2877 and AAR28378) and a putative HcrVf3 protein in *Oryza sativa* (accession 483581) particularly in the LRR regions. Protein EIX was related to ethylene biosynthesis, electrolyte leakage, pathogenesis protein expression and hypersensitive response (HR) in specific plant species (Bailey *et al.,* 1990).

In tomato, the EIX domains were predicted to play roles in orientation and anchoring the protein to cell membrane (Mily and Avni, 2004). Thus, it is suggested that, although the LRR domain is widespread and abundant, particular LRR motifs are specific enough that they are clear candidates for indicating or screening purposes of resistance genes generally either in plants or BACs.

5.4.3 Diversity of LRR

Through evolution, both biotic and abiotic stress may have led to the accumulation of deletions, insertions and mutations in certain hot spot sites in the genome, especially in repeated motifs as LRR which are involved in signalling responses. Short repeats like LRR often correspond to distinct structural modules and become important as the repeats build up with a few changes, as in *Cf*-disease resistance families which are related to evolution of LRR between *Cladosporium fulvum* and tomato (*Lycopersicon* spp.).

Diversity and evolution of LRR corresponding to R-genes are clearly observed in tomato *Cf* genes including *Cf-2*, Cf-4, Hcr9-4E, Cf-5 and Cf-9 which were resistance to *C. fulvum* corresponding to *Avr* gene, Avr2, Avr4, Avr4E, Avr5 and Avr9 respectively (Joosten and De Wit 1999). Amongst them, *Cf-4* and *Cf-9* genes were the first studied and found specificity of Cf-4 and Cf-9 functional proteins consists of extracellular LRR differs as each has a total number of 25 and 27 LRRs respectively. The studied observed that the LRRs protein in Cf-4 were not fully utilised towards their specificity and reside in the N-terminal LRR-flanking domain and LRR at position order 13, 14 and 16, whereas LRRs protein of Cf-9 were distributed in all order numbers (Wulf *et al.,* 2001; De Wit *et al.*, 2002).

Multiple sequence alignments of LRR proteins were essential to identify functionally important sites and seeking protein functions and structure predictions by homology with well characterized sequences from reference species. However, individual sequences from each variety were generally too variable to be used alone in structure prediction, but by clustering the sequences, the LRR motifs could be differentiated. LRRs were also found in TIR and non-TIR domains further integrated with NBS-LRR complexes which might be involved in functionality and activities of the specificity R-genes. However the roles of the LRR in these domains are not being focussed directly and remain ambiguous. Polymorphism in NBS-LRR complexes were found between rice accessions reflected a strong contribution of LRR in diversify the R-genes and were used to measure their relatedness and linked with disease resistance although most studies have not focussed on between accession LRR diversity (Moumeni *et al.*, 2003).

A distance method through UPGMA for tree building construction was limited by its sensitivity and accuracy at deduced amino acid levels as translations had six alternative frames. Further, ambiguous relative placements of protein residues in the gaps may change the alignments. In general terms, all sequences present at the ends of deep branches which include multiple current species (accessions) are likely to represent allelic or evolutionary sequences from a single gene, or perhaps gene family. Within the branch, it is then possible to make sequence interpretations based on the model that all sequences on the branch evolved from a single common ancestral sequence. Sequence trees showed divergence of the LRR motifs and might be corresponding to sterility and vegetative propagation in *Musa*.

Comparative sequences analysis was relatively difficult as even the same fragments showed size polymorphism, further making the alignment complicated. On many occasions, clones derived from a single fragment in one variety were grouped into different clades in sequence tree although with higher similarity. This is a strong clue that all the LRR motifs isolated originated from the same ancestor.

LRR motifs isolated from genomic DNA of *Musa* amplified by using primer LINR1-F and LINT2-R managed to detect genomic polymorphism. It is suggested that *Musa* genome is polymorphic due to its broad genetic base, related to its domestication by human and/or natural selection through environmental stress with often without inbreeding in the parthenocarpic, sterile accessions.

5.5 Conclusion

Specific LRR primers designed to isolate LRR motifs for *Musa* accessions in this experiment were conserved amongst them, but the sequences do not correspond to *Musa* genomic constitutions. Degenerate primers were not found to be optimum for isolating the LRR motifs as they can bind to many template sites creating multiple bands including those from pseudogenes. The LRR motifs are conserved in *Musa* and exist in many copies, with independent sizes and/or specific structures. Based on deduced amino acid sequences, four clusters were identified from the LRR motifs.

Analysis of the LRR motif isolated from *Musa* revealed that eleven LRR motifs were present in three different classes, which correspond to both intra- and extra- cytoplasmic groups of the gene identified in other species. The motifs can be related to many functional signalling proteins and disease resistance proteins that have been found in reference species. However, disease resistance proteins are specific sequences related to specific genes which are triggered by specific elicitors in pathogens or pathogen strains. It may remain unexpressed and activate only following a biotic challenge (or indeed under a specific abiotic stress). Thus, further analysis is needed to characterize their expression and any relationship to specific stress responses.

CHAPTER 6

6.0 Identification of large-insert clones containing biotic and abiotic stress-related resistance genes in *Musa* BAC libraries

6.1 Introduction

Large-insert genomic DNA libraries are important for physical and genetic mapping in many organisms. Information derived from the analysis of cloned inserts would contribute knowledge of the structure and function of genes in their context including upstream and downstream regions with promoters and modifiers, which leads to understanding their homology, genetic linkages, conservation and evolution (including conservation of synteny with relatives) in plant species. For the past few years, bacterial artificial chromosome (BAC) vectors have been emerged as a choice for cloning the large-insert DNA from animal and plant species.

BACs are relatively easy to maintain, show very large inserts, high stability in host strains and a low degree of chaemerism as compared to alternatives such as yeast artificial chromosomes (YACs), P1, or phage (lambda) vectors. Many BAC libraries have been constructed in plant species i.e. coffee (Noir *et al.*, 2004), sugar beet (Fang *et al.*, 2004), *Triticum* (Shen *et al.*, 2005), Maize (Quint *et al.*, 2003) and *Cicer* (Rajesh *et al.*, 2004) to be useful not only for gene homology, conserved regions and functional genes analysis but also for determination of their marker systems; in particular, the laboratory of Hong-Bin Zhang in Texas has generated many hundreds of plant libraries used throughout the world (see listing at http://www.hbz7.tamu.edu). Varieties of BAC library from plants and animals species are listed in http://www.genome.clemson.edu.

Natural or induced selection has revealed involvement of various resistance genes related to biotic and abiotic stress amongst the plant species. It is believed that certain structures i.e. resistance protein motifs, have been maintained and conserved very widely among plants. In understanding and revealing the genes, molecular markers for tagging of interesting genes is reliable and rapid to identify and further isolate the complete genes for gene expression studies.

Biotic stress is related to biological disturbance and/or invaders from organisms involving avirulence (*avr*), virulence and Resistance genes (R-genes) recognising, signalling and activation of specific defence mechanisms. However, abiotic stress such as drought, salinity, chemical or metal toxicity and acceptant temperature is related to a series of morphological, physiological, biochemical pathways and molecular changes which might be associated and interconnected amongst them to destruct plant growth and productivities (Knight and Knight, 2001; Wang *et al.*, 2001; Zhu, 2001, 2002).

In Chapter 4, NBS domains which are related to disease resistance genes are shown to be conserved. In Chapter 5, specific LRR motifs were isolated which are involved in many protein interactions and showed occurrence of variation in their repeat motifs which might be related to NBS-LRR complexes for resistance specification. And retrotransposons as described in Chapter 3 are believed to be related to a mechanism in generating new variation in gene organisation and structures. These approaches are convincingly leading towards R-genes isolation for biotic stress. Abiotic stress response involves in many genes and pathway complexes, tagging protein related to drought, salinity and heat tolerance is still an obstacle. However, it might be the main approached towards understanding their associations through utilisation of BAC libraries.

Screening of BAC libraries for individual clones containing candidate genes of interest is possible through either a PCR- or hybridization-based approach (colony hybridization). Facilities and techniques for accessing the BAC libraries, including storage and DNA isolation in microtiter plates, robotic systems, arrayed pooled clones and high-density arrayed hybridization membrane filters, have been developed for efficient and rapid screening. To speed up the screening process, a pooling strategy screening was applied where the library was made into sets of clones from the BAC library, corresponding to plate, column and row pools (known as a multi dimensional or, here, three-dimensional pooling strategy) where each clone can be uniquely or with little ambiguity identified through PCR amplification in the sets of clones in which it occurs.

PCR-based screening is a straightforward procedure which requires specific primers designed from interesting genes or sequence motifs whereas colony hybridization-based screening requires high-density gridded macroarrays of entire clones in the library attached accordingly to membrane filter and hybridization of a probe for the interest genes. The probes used for the hybridization can be obtained from DNA fragments, PCR products or DNA oligonucleotides.

The first two probes are preferred as they consist of several hundred base pairs and ease of labelling by nick translation or random priming to generate high specificity activity probes (Hans *et al.,* 2000). However, very long probes have a disadvantage of having repeat elements which contributed to high detection of the BAC clones, resulting in false identification. In avoiding the repeats, oligonucleotide probes have been developed known as overgos (overlapping oligonucleotides) through annealing of two overlapping oligonucleotides (22-26 bases long). Overhanging primer bases are extended with Klenow enzyme and radio-labelled nucleotides.

As compared to DNA oligonucleotides probe and other hybridization methods, the overgos probe can avoid the repeats, giving better hybridization kinetics and lower rates of false positives, higher specificity labelling and lower background effects (Ross *et al.*, 1999; Hans *et al.*, 2000). Disadvantages of this method are requirement of more primers, re-array of secondary filters and at least two hybridization procedures must be completed in getting to the specific clones. After all this, the individual clone obtained still requires PCR procedure for confirmation.

Efficiency of PCR-based screening in combination with pooling strategies has been known to be effective in identifying BACs or YACs from libraries (Bruno *et al.,* 1995; Klein *et al.,* 2000). In identifying a single clone from a target sequence, condensing the BAC library into pools (i.e. plate, column and row pools representing overlapping groups of clones) and then screening can require a relatively large number of PCR. However, the procedures are straightforward, easy to handle, rapid and reliable for small numbers of target sequences.

In this Chapter, a research experiment is aimed to identify a reliable protocol for BAC library screening through combination approach of pooling strategy either with PCR-based or colony hybridization. Large-insert BAC libraries of *Musa* were screened to identify clones consisting of related sequences corresponding to both abiotic and biotic stress genes. High density filter arrays of individual BAC clones and PCR based techniques with multi-dimensional BAC DNA pools were used to identify a single BAC clone containing the candidate gene for both abiotic and biotic stress in *Musa,* where the related sequences were then confirmed by sequencing.

The information is useful for targeting resistance protein genes in breeding and crop improvement in *Musa*, and in revealing the full-length coding sequences and flanking promoters and regulatory regions, as well as the genomic context of the stress-related genes. BACs selected here have now be pipelined for complete sequencing in the Japanese NIAS programme within the Generation Challenge Programme, but analysis of the individual BAC clones is beyond the scope of this chapter.

6.2 Materials and Methods

6.2.1 *Musa* BAC libraries

Musa BAC libraries originated from *M. acuminata* 'Calcutta-4' (diploid 'A' genome) and *M. balbisiana* 'Klutuk Wulung' (diploid 'B' genome). They were obtained from the *Musa* Genome Resources Centre (MGRC), Institute of Experimental Botany (IEB), Czech Republic and described in Chapter 2. The BAC libraries were pooled accordingly as super-pools, plate-pools, column-pools (from 8 plates) and row-pools (also from 8 plates) and kindly supplied to us by Professor J. Dolezel, IEB. The BAC clones in each library were arranged and pooled accordingly in sets of eight plates to cover up 18 and 12 super-pools from MA4 (*M. acuminata* 'Calcutta-4' BAC library) and MBP (*M. balbisiana* 'Klutuk Wulung' BAC library) respectively (Figure 6.1). Membrane filters orientations and details were described in Chapter 2. For PCR application, the initial samples were diluted in 1:50 sterile dH_2O.

6.2.2 Colony hybridization

Six high-density membrane filters consisting of genomic DNA each from *M. acuminata* 'Calcutta-4' (four membranes: MA4-F1, MA4-F2, MA4-F3, C4BAM) and *M. balbisiana* 'Klutuk Wulung' (two membranes: MBP-F1, MBP-F2) were obtained from Institute of Experimental Botany (IEB), Czech Republic and described in Chapter 2. DNA from plasmids for abiotic stress was obtained from CIRAD, France as described in Chapter 2 were used as probe to hybridize the membrane filters (Table 6.1).

Table 6.1. Membrane filters arrangement against abiotic EST plasmid.

Probe Used (EST Clone)	Description	Membrane Filter
MA4LIMFES014A_G03	Drought Responsive Family Protein	MA4-F1 MBP-F2
MA4LIMFES011B_G02	Chloroplast Drought Induced Stress Protein	MA4-F3
MACVLIMFLS011D_A07	Salt Tolerance Protein (Member of the Constants Zinc Finger Family)	MA4-F2 MA4-F1
MA4LIMFES003C_G06	Low Temperature Induced Protein	C4BAM

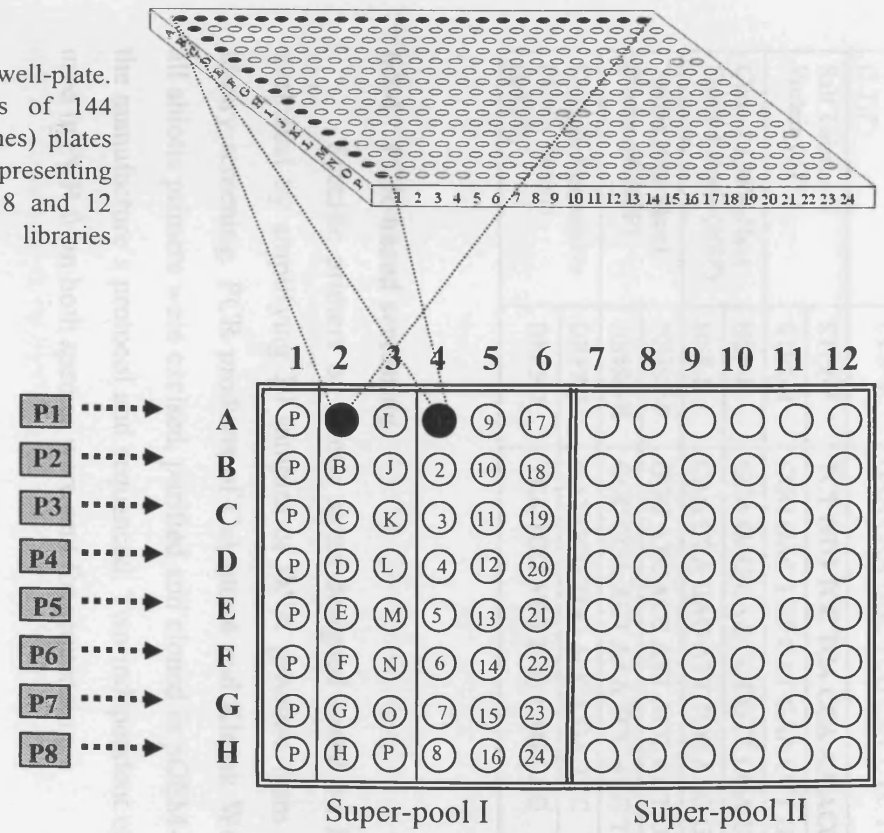
LRR motifs corresponding to biotic stress were used as probe obtained from clone L73 (insert amplified from genomic DNA of *M. balbisiana* 'Klutuk Wulung') and L41 (insert amplified from genomic DNA of *M. acuminata* 'Culcutta-4') described in Chapter 5. Colony hybridizations were carried out on two selected membrane filters each from *M. acuminata* and *M. balbisiana* library arranged as in Table 6.2. Probe labelling and membrane hybridization were described in Chapter 2.

Table 6.2. Membrane filters arrangement against LRR motifs.

Membrane Filter	Probe Used (LRR clone; see Chapter 5)
MA4-F1	L41
MA4-F3	L73, L41
MBP-F1	L73
MBP-F2	L41, L73

(A) Plate pool (384-well-plate)
Individual BAC clones in 384-well-plate. MA4 and MBP libraries consists of 144 (55296 clones) and 96 (36864 clones) plates respectively. Eight plate-pools representing one super-pool corresponding to 18 and 12 super-pools for MA4 and MBP libraries respectively .

(B) BAC-pooled plate
BAC-pooled plate consists of two BAC super-pools. 384 clones were pooled from 384-well-plate assigned as 'plate pools'. Pooled of eight 'plate pools' (P1 to P8) were assigned as BAC super-pools. The BAC super-pools consists of 3072 BAC clones.



(C) Two super-pools arrayed in 96-well-plate. Wells in black indicate the origin of the assemble pools from 384-well plate. Super-pools I arrayed in column 1-6 and Super-pools II arrayed in column 7-12. Column 2-3 and 8-9 consists of 192 BAC clones (24 columns x 8 plates) in each well assigned as 'column-pools'. Column 4-6 and 10-12 consists 128 BAC clones 16 rows and 8 plates) in each well assigned as 'row-pools'.

Figure 6.1. Schematic of pooling strategy for PCR-based screening of BAC libraries consisting of BAC super-pools, plate-pools, row pools and column pools. (A) Array of 384 individual BAC clones in 384-well plates. (B) Assembly of eight plate- pools to perform 18 BAC super-pools of MA4 and 12 BAC super-pools for MBP which were used as template first PCR-based screening. Each BAC plate-pool consists of all BACs from eight 384 well-plate corresponding to 3072 individual BAC clones. (C) Two sets of super-pools were organised in 96-well plates consist of 16 wells of 'row-pools' (A-P) and 24 wells of 'column-pools' (1-24). Each 'row pools' and 'column pools' consist of 3072 BAC clones. MA4 and MBP BAC libraries are 9x coverage from estimated genome size of 640 Mbp and 530 Mbp of *M. acuminata* 'Calcutta-4' and *M. balbisiana* 'Klutuk Wulung' respectively. **(Accessed from www.musagenomics.org redirecting to http://bioinfo.inibap.org/statusdb/display.php?page=baclibrary 18/10/06)**

6.2.3 Primer designed for abiotic stress

Specific primers for abiotic stress were designed as described in Chapter 2 which aimed to amplify as long amplicons as possible based on the EST sequences.

Table 6.3 Abiotic primers

Primer Description	Primer Label	Sequence (5' – 3')	T_m (°C)	Expected Product size (bp)
Low Temperature Induced Protein (LTIP)	LTIP-F	AAA TGC GGC ACT TTT CAT TC	63.7	502
	LTIP-R	GAA CAA GGC TCG CAT CTC TC	64.0	
Salt Tolerance Protein (STP)	STP2-F	TCT GTG TCC TCA CCA AA ACG	63.8	540
	STP2-R	CGG GAC TGT GAT GAA CCT G	64.3	
Class-1 LMW Heat Shock Protein (HSP)	HSP-F	GCA GAG CAA CCT GTT CGA C	63.7	457
	HSP-R	CAA CCA GAG ATT TCG ATG GAC	63.3	
70kDa Heat Shock Protein (70HSP)	70HSP-F	GGG ATC ACC ATT ACC AAT GAC	62.7	440
	70HSP-R	GGC CGA ACT AAA TCC ACC TC	64.5	
Drought Responsive Family Protein (DRFP)	DRFP-F	GAT GCG TGG AGC CGA TTC	66.9	224
	DRFP-R	TCA TCG ATG TGG CAG CAG	65.0	

6.2.4 PCR-based screening

Specific primers of abiotic stress designed based on EST sequences were evaluated by amplifying 20 samples of DNA genomic from *Musa* prior to BAC library screening. PCR products of Calcutta-4 and Klutuk Wulung obtained from all abiotic primers were excised, purified and cloned in pGEM-T Easy according to the manufacture's protocol and sequenced. Two independent clones in each primer used in PCR from both species were sent for sequencing.

Degenerate primers related to NBS domain (described in Chapter 4) and specific primers related to LRR (described in Chapter 5) were used for biotic stress screening in BAC libraries. Transposable element primers related to copia-like and gypsy-like retroelements (described in Chapter 3) were also appraised for the BAC library screening. All PCR reactions related to abiotic, biotic and retroelements primers and cycles were optimised as described in Chapter 2. Annealing temperatures for different primer combinations were as described accordingly in Chapter 3, 4 and 5.

Strategy in PCR-based screening firstly involved in amplification of BAC super-pools. Then, positively selected super-pools were determined and further, appropriate plate-, row- and column-pools were screened to identify the specific individual clones.

6.2.5 Sequence confirmation and homology

Individual BAC clones corresponding to those selected from the pools were requested from the *Musa* Genome Resources Centre, IEB. These were grown and PCR products obtained using the appropriate primers, including negative control BACs. PCR products of identified clones were confirmed through direct sequencing from PCR products. The PCR products were separated using 1% gel electrophoresis, excised and purified by using Qiagen Gel extraction Kits following the manufacture's protocol. Purified fragments were check by separated 5µl of the fragment on 1% agarose gel before sending to JIC for sequencing.

Homology of the sequences were searched and compared by BLASTX 2.2.15 (Altschul *et al.*, 1997) in EMBL for mining related protein databases. Significance of the match showed by E-value (significant at < 0.01) corresponding to a statistical analysis based on the similar likelihood match occurring by chance. Individual BAC which were confirmed and high homology to protein in public databases were proposed for full sequencing.

6.3 Results

6.3.1 Colony hybridization screening of BAC libraries

There were no clear results obtained from colony hybridization of membrane filters hybridized by the abiotic EST. X-ray films exposed were rather clean without spots or high background after they were exposed more then a week to the hybridized membrane filters. None of the MBP membrane filters produced double dots pattern hybridized with probe L41 and L73. MBP-F1 showed only background dots and MBP-F2 has an ambiguous orientation of double dots which do not clearly correspond to possible patterns on the robot-made grids.

Membrane MA4-F1 probed with L73 showed some dispersed background with unclear edges. Five clones were selected from this membrane, identified as MA4-34-C4, MA4-3A13, MA4-1O18, MA4-13-F2 and MA4-13E8 (Figure 6.2). Membrane filters MA4-F3 probed with L73 and L41 had no background, with ambiguous double dots signals. Two clones were selected from this membrane probe by L73 and identified as MA4-138-I17 and MA4-138-H17.

All seven clones were amplified with LRR primers for confirmation, resulting in PCR products at approximately 820 bp (Figure 6.3). Three clones from the seven clones selected were amplified with NBS primers. They were identified as MA4_13E8, MA4_138H17 and MA4_1O18 and amplified at approximately 520 bp (Figure 6.4.). The PCR proved that the selected colony were true hybridization dots and not background and/or debris from the colony hybridization.

PCR screening onto the seven BAC candidate-clones for abiotic stress related to drought, salinity and heat shock identified with *M. acuminata* 'Calcutta-4' as a positive control. The abiotic screening revealed six abiotic positive clones. The salinity clones identified by primer STP2 were MA4_13F2, MA4_1O18, MA4_138H17 and MA4_3A13) and by primer HSP were MA4_13E8 and MA4_138I11 (Figure 6.5).

6.3.2 PCR-based screening

6.3.2.1 Abiotic PCR and cloning of genomic DNA

The PCR products for each primer showed a single band for all *Musa* genomic DNA at the expected band size except *M. textilis* which amplified only with primers STP and LTIP (Figure 6.6). All PCR products were amplified from the abiotic primers and were significantly longer as compared to EST sequences (as expected) for designing the primers except HSP and 70HSP which were equal in their sizes. PCR products from LTIP, STP, HSP, 70HSP and DRFP were 615 bp, 615 bp, 456 bp, 440 bp and 312 bp respectively.

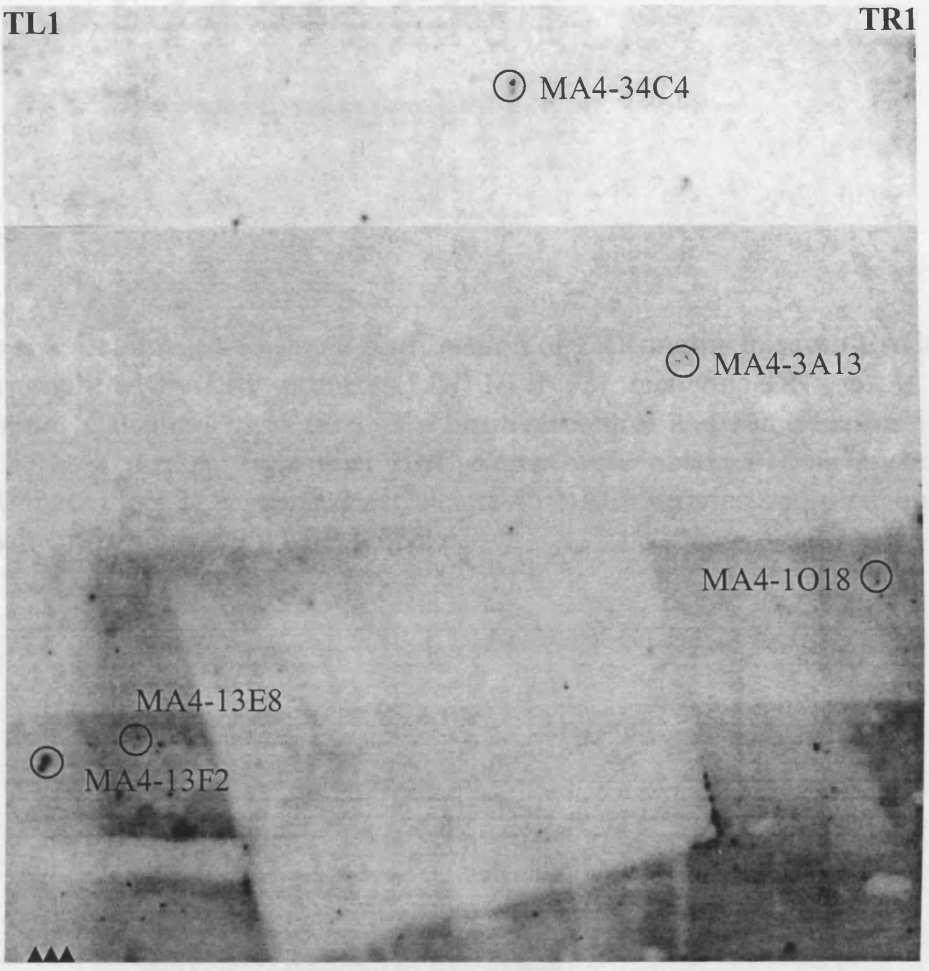


Figure 6.2. Colony hybridization of MA4 BAC Membrane Filter 1 by using LRR motifs probe obtained from Clone L73 (refer Chapter 5). The filter consists of six super-pool: MA4_1-8, MA4_9-16, MA4_17-24, MA4_25-32, MA4_33-40 and MA4_41-48. Each super-pools corresponding to eight plate (384-well-plate). The super-pools were arranged in two replication for confirmation based on pattern order as in Figure 2.1 (Chapter 2). The probe was prepared by *Eco* RI restriction enzyme digestion and extracted from 1% agarose gel using QIAGEN extraction kit according to manufacture protocol. The insert of clone L73 was originated from PCR amplification of *M. balbisiana* 'Klutuk Wulung' (BB) with LRR primers (LINT1-F/LINT2-R) and cloned in pGEM-T Easy. Five BAC clones were selected (in circle) based on 4x4 grid identified as MA4-34C4, MA4-13E8, MA4-13F2, MA4-1O18 and MA4-3A13.

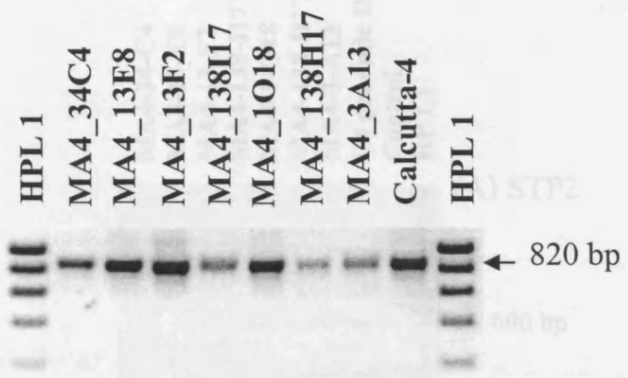


Figure 6.3. Gel image showing a confirmation of LRR motifs in seven BAC clones amplified by LRR primers (LINT1/LINT2). Genomic DNA of *M. acuminata* 'Calcutta-4' was used as a positive control and HyperLadder I (HPL 1) as a marker. Individual BAC clones were obtained from *Musa* GRMC based on *M. acuminata* 'Calcutta-4' BAC libraries screened by colony hybridization using LRR probe.

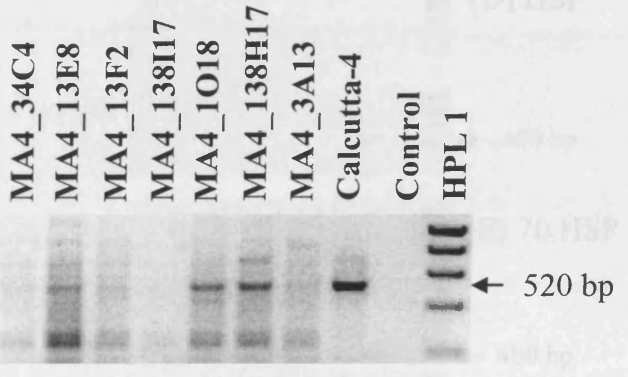


Figure 6.4. PCR products of 520 bp obtained from seven BAC clones amplified by using NBS primers (CNL298F/NBSR1) spanning from Kinase-1 (P-loop) to GLPL motifs. Genomic DNA of *M. acuminata* 'Calcutta-4' was used as positive control and HyperLadder 1 (HPL 1). As a marker. Three positive clones consist of NBS domain corresponding to protein resistance were identified; MA4_13E8, 1O108 and MA4_138H17.

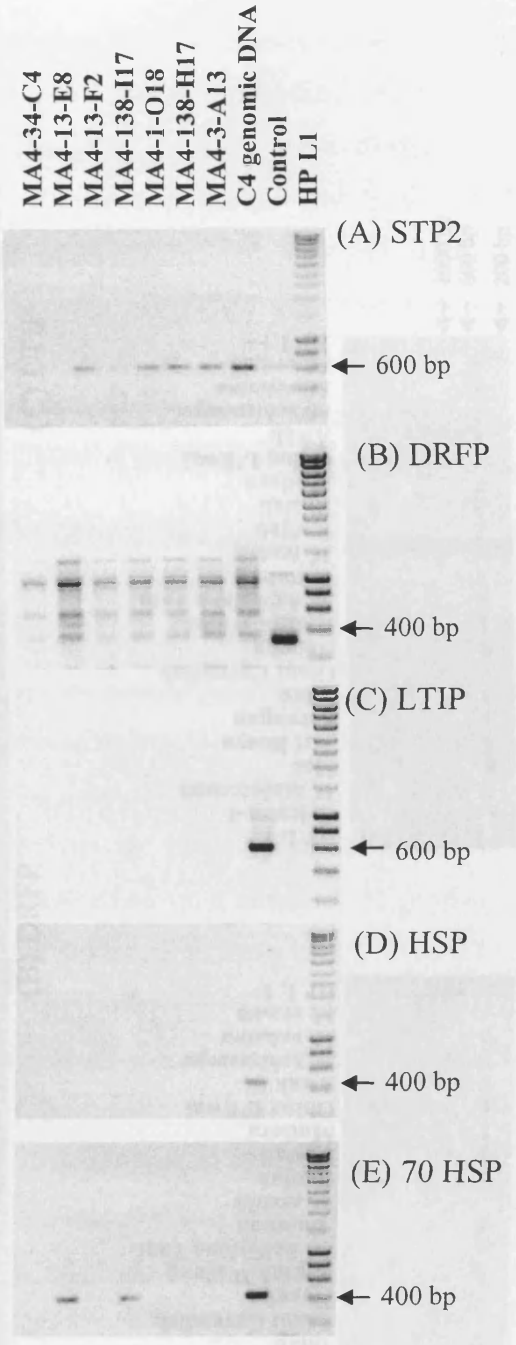


Figure 6.5. Gel image of seven BAC clones screened by PCR by using abiotic primers (A) Salt Tolerance Protein (STP), (B) Drought Responsive Family Protein (DRFP), (C) Low Temperature Induced Protein (LTIP), (D) Class-1 LMW Heat Shock Protein (HSP) and (E) 70 kDa Heat Shock Protein (70HSP). Genomic DNA of *M. acuminata* 'Calcutta-4' was used as a positive control and HyperLadder 1 (HPL 1) as a marker. Individual BAC clones were screened from MA4 BAC library of *M. acuminata* 'Calcutta-4' by colony hybridization using LRR motifs as a probe and were obtained from MGRC.

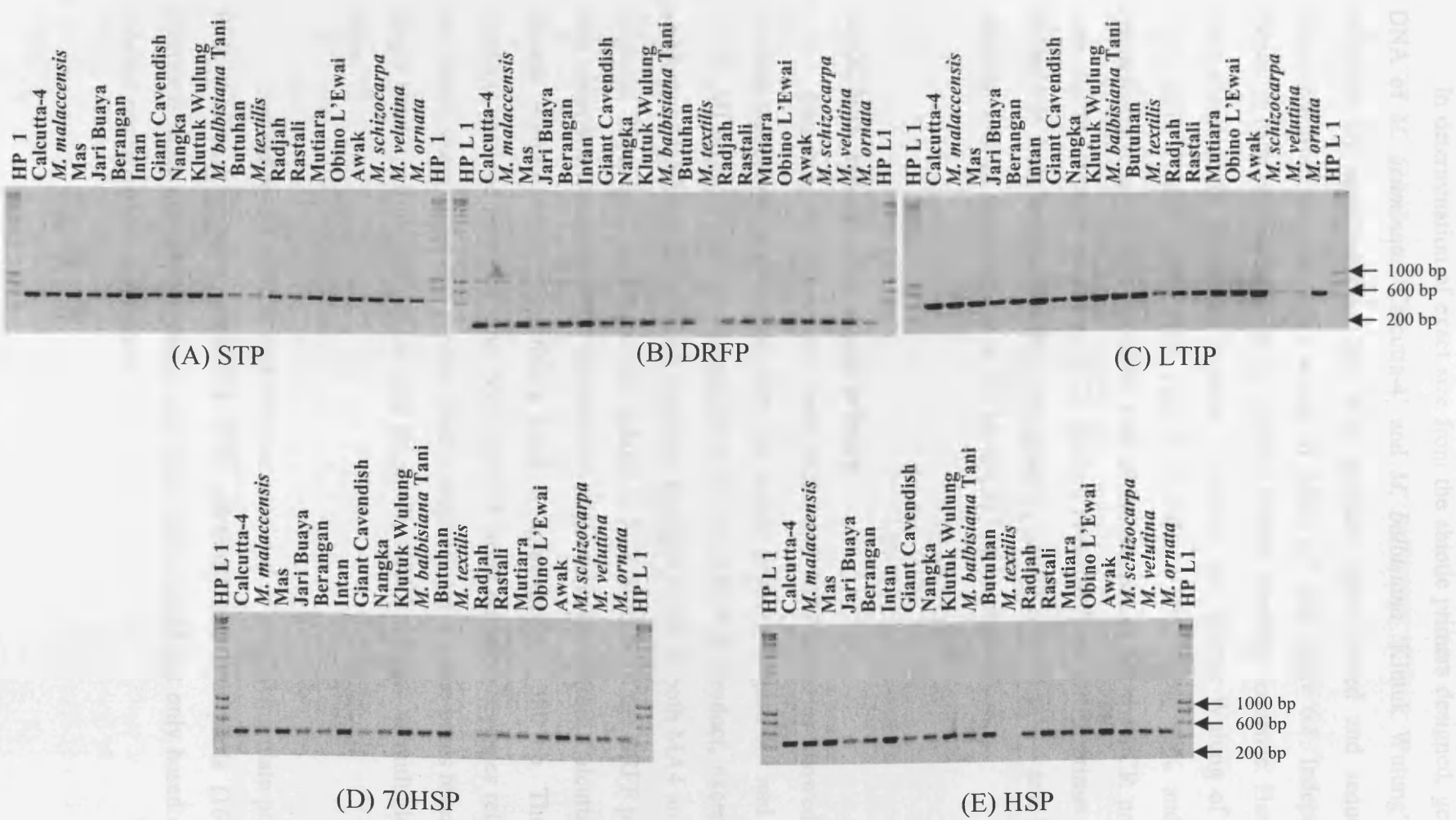


Figure 6.6. Gel image of PCR products amplified by using (A) Salt Tolerance Protein (STP), (B) Drought Responsive Family Protein (DRFP), (C) Low Temperature Induced Protein, (D) 70kDa Heat Shock Protein (70HSP) and (E) Heat Shock Protein Class 1 LMW (HSP) primers on genomic DNA of *Musa* varieties. HyperLadder 1 (HPL 1) are located at both sides of the gel.

In determination of exact size from the abiotic primers designed, genomic DNA of *M. acuminata* 'Calcutta-4' and *M. balbisiana* 'Klutuk Wulung' were amplified by each primers. The PCR products were cloned and sequenced. Sequences from the clones are shown in figure 6.7 and figure 6.8. Independent clones of each primers showed an identical (>98% identity) sequence. However, comparison of the sequences between Calcutta-4 and Klutuk Wulung of LTIP, STP, HSP, 70HSP and DRFP ranged at 99%, 100%, 94%, 99% and 80% respectively. Comparison based on size of the sequences for the PCR products were found identical except the PCR product obtained from DRFP primer which differed in their length. The PCR amplified by DRFP for Calcutta-4 and Klutuk Wulung resulted fragment size at 312 bp and 327 bp respectively.

6.3.2.2 BAC screening by abiotic primers

Screening of super-pools from MA4 and MBP libraries showed many positive pools which amplified from the abiotic primers (Figure 6.9 and Figure 6.10). All positive super-pools amplified as a single PCR product, except PCR products of DRFP primers which produced multiple bands in both MA4 and MBP libraries. Thus, associated band size, related to PCR product of the DRFP primers, was determined based on PCR product size of genomic DNA from Calcutta-4 and Klutuk Wulung which identified a band at approximately 380 bp. The PCR products obtained from the super-pools showed target domain sequences related to the abiotic stress differed in product band strength. Positive clone-pools revealed as single band in all abiotic primer used except for DRFP consists of multiple band sizes.

Second round of PCR amplification consisted of amplifying plate pools (8 PCR reactions), column pools (24 PCR reactions) and row pools (16 PCR reactions) using the abiotic primers and these were carried out only based on the selected positive MA4 super-pools.

(A) Clone C4_70HSP (size 440bp)

GGGATCACCATTACCAATGACAAGGGTAGGCTCAGCAAGGAGGACATTGAGAAGATGGTGCAGGAAGCG
GAGAAGTACAAGTCGGAGGACGAGGAGCACAAGAAGAAGGTGGAGGCGAAGAATGCTCTGGAGAACTAC
TCTTACAACATGAGAAACACCATCAAGGATGATAAGATTGCATCCAAGCTTGCGGCTGCCGATAAGAAGA
AGATCGAGGATGCTATTGATCAGGCAATCCAGTGGCTCGACGGCAACCAGTTGGCAGAAGCTGACGAGTT
CGAAGACAAGATGAAGGAGCTGGAGAGCATTTGCAACCCCATCATCGCCAAGATGTATCAGGGTGCCGGT
GCTGACATGGGTGGTGGAATGGATGATGACGCACCCTCTGCTGGTGCCAGTGGTGCAGGTCCCAAGATCG
AGGAGGTGGATTTAGTTCGGCC

(B) Clone C4_DRFP (size 312 bp)

GATGCGTGGAGCCGATTCTCCGCGGCCGCGTCGAAGCGCCACCAATCCGCGCTGCAGTCGCGATACGCCG
ACGCGTCTCGGGATCCCAGTCTTCGCGGCTTGAGAGGTGGGGGGTTTTTCGACAGGGGAAGGGGCTTGTGA
TCCCTGCCCTGATTCATGAGTGGGATTTTCTGATTGGTTGTTTCGTTCGGTAGATCTGTATCTTGGATTCGAG
GAGTTAGACGGAGGCGAGGATGACACCCGCGCGGAGTTTCCCTGCCCCTTCTGCTCTGAGGACTTCGATAT
TGTCGGGCTCTGCTGCCACATCGATGA

(C) Clone C4_HSP (size 456 bp)

GCAGAGCAACCTGTTCGACCCTTTCTCCCTCGACCTCTGGGATCCGATCGATGGCTTCCCCTTCGGCTCTTC
CTCCTCCCTTGCCCGCCCCTCCATCCTGTTCCCTAGCGAGGTATCTGCCTTCGTCGGCACCCACATCGACTG
GAAGGAGACACCGGAGGCGCACGTCTTCAAGGCCGACATCCCGGGGCTCAAGAAGGAGGAGGTGAAGGT
GGAGATAGAGGATGGCCGGTTCCTCCAGATCAGCGGAGAGCGGAAGAGCGAGGAGGAGAATACCGACAC
CTGGCACCGCGTGGAGCGGAGCCGCGGCAAGTTCCTGCGACGGTTCCGGTTACCGGAGACCGCCAGGGTG
GACCAGGTGAGGGCGGCCATGGAGGACGGCGTGCTCACCGTCACCGCTCCCAAGGAGGAGGAGATCAAG
GACTCCGACGTCCGGTCCATCGAAATCTCTGGTTG

(D) Clone C4_STP (size 615 bp)

TCTGTGTCCTCACCAAAACGTAAGTATAATATTTTAGGCAACGAAACATTAAGCTTTATCCAATATCAGGG
ACAGTGAAGTATTCCTCATCATCAGATATCTCAAGCCTAGGCTTCTTGAATGTTGTGCCATGTTTGTTAGCC
CTGTGGAAACCTGTGTTGCTTGCTTGTGATGTAGGGAGTTCAGGAACTTGAGCTGCTGCTTGGGCACCCTTT
GGCGTTTCATCATGGAAAAGACCAATGTCTTCAAACCATTCCAACTCTCCAAATCCAACAGGAGACTCCTG
GAACAGGTTGGATCAGAGTATGACCATCTTATGTGAACTACTTTAAATGGAAGTAAATCGACATATCATAC
CTTTTCACCAATTTCTTGGTCTGAAAGCTGAAGGAACTCATCAACAGCCCATGCAGAATGCATGAATGATG
GTGTGCTCTGTTGTGTTGCAGGAACTTTTGTTGCAATCAAGGCCGGTCCATGGCGTGGTGGTTCAGTGTTGT
CCTTTAGGTCCTTGTCGCACATCTTACTTACGGCCACACGGATTCCAGTAGCTAAATAACGCTGGTGGTTGC
CGGAAAGGGTACCAGCGACATGAACAGGTTCATCACAGTCCCG

(E) Clone C4_LTIP (size 615 bp)

AAATGCGGCACTTTTCATTCCAATGTCTTATCAGCAGGTAGAAGATTTCTCCCCTGAATGGTGGGAGTTGGT
GAAGACAACCCCATGGTTTCGTGAGCACTGGTTCCATCAGTATCAGGAGCAAGAAACATTTGGCAACGAT
GAGGAGGACATGGCAGACGCTTTGCCAGACTCTTTCGATCTTGGAATCACTGAATACTCTTTGCTGGAGGC
AGAGTCCGATGGGGCAGCTTACCATGAACCTGGGAAAGAGATGATTTCTGGTGCTACAGAGCGAGAGAAG
ATAGAGCATGCTATTCAACTGTATTGCAGTTTGATGGAAGGTTATGTATACACCAAAATTGCAACCATATG
CAATTTACAGTTGCTTTGTCAATTGAAGCCTTTGTGGTAGTGAATCTAATTCTTTTCCTCTCTGATTGGTTAT
ATGAGACAGAGCCCTATAGCGATGCAAAGGCAGCAATCAAGGATCTGAGCTTGAATTCTCCGAAGAATGG
TGTTAAGCCATTGTCTGTTCCAGCAAAGTATCGCGAGAAGCCTCTGCAATGCGTTAGTCCAAAGTACAGCC
CGCGGCGCATCATTCACCAGCCACGCTGAGAGATGCGAGCCTTGTTC

Figure 6.7. PCR products of five abiotic primers amplified from genomic DNA of *M. acuminata* 'Calcutta-4', cloned in pGEM-T-Easy vector. Clones are labelled as (A) Clone C4_70HSP (size 440bp), (B) Clone C4_DRFP (size 312 bp), (C) Clone C4_HSP (size 456 bp), (D) Clone C4_STP (size 615 bp) and (E) Clone C4_LTIP (size 615 bp).

(A) KW-70HSP (440 bp)

GGGATCACCATTACCAATGACAAGGGTAGGCTCAGCAAGGAGGACATCGAGAGGATGGTGCAGGAAGCGGA
GAAGTACAAGTCGGAGGACGAGGAGCACAAGAAGAAGGTGGAGGCGAAGAATGCTCTGGAGAACTACTCTT
ACAACATGAGAAACACCATCAAGGATGATAAGATTGCATCCAAGCTTGCAGCTGCCGATAAGAAGAAGATCG
AGGATGCCATTGATCAGGCAATCCAGTGGCTCGACGGCAATCAGTTGGCAGAAGCTGACGAGTTCGAAGACA
AGATGAAGGAGCTGGAGAGCATTTGCAACCCCATCATCGCCAAGATGTATCAGGGTGCCGGTGCTGACATGG
GTGGTGGAATGGATGATGACGCACCCTCTGCTGGTGCCAGTGGTGCAGGTCCCAAGATTGAGGAGGTGGATTT
AGTTCGGCC

(B) KW_DRFP (327 bp)

GATGCGTGGAGCCGATTCTCCGCGGCCGCGTCCAAACGCCATCAATCCGCGCCCCAGTCACGATACTCCGATG
CGTCTAGGGATCCCAATCTTCGCGGTTGGAAAGGTGGGTGGGTGGATTTCGATTTCGCCGAAGAGGGTTCGTG
TGTTCATTGCGCTGTTATTAAGTTCGTTCATTTTTTTGCCTGATCGGCCCTTTCGGTTGGCAGATTTGTATCTCG
GGTTCGAGGAGTTCGAAGGAGGAGAAGATGACCCCCGCGCGGAGTTCCCCTGCCCATTCTGCTCTGAGGATTT
CGACATCGTCGGACTCTGCTGCCACATCGATGA

(C) KW-HSP (459 bp)

GCAGAGCAACCTGTTCGACCCTTTCTCCCTCGACCTCTGGGATCCGATCGATGGCTTCCCCTTCGGCTCTTCCT
CCTCCCTTGCCCGCCCCTCCATCCTGTTCCCTAGCGAGGTCTCTGCCTTCGTCGGCACCCACATCGACTGGAAG
GAGACACCGGAGGCGCACGTCTTCGAGGCCGACATCCCGGGGCTCAAGAAGGAGGAGGTGAAGGTGGAGAT
AGAGGATGGCCGGTTCCTCCAGATCAGCGGAGAGCGGAAGAGCGAGGAGGAGGAGAATACCGACACCTGGC
ACCGCGTGGAGCGGAGCCGCGGCAAGTTCCTGCGTCGGTTCCGGTTGCCGGAGACCGCCAGGGTGGACCAGG
TGAGGGCGGCCATGGAGGACGGCGTGCTCACCGTCACCGTTCCCAAGGAGGAGGAGATCAAGAACTCCGACG
TCCGGTCCATCGAAATCTCTGGTTG

(D) KW_DRFP (327 bp)

GATGCGTGGAGCCGATTCTCCGCGGCCGCGTCCAAACGCCATCAATCCGCGCCCCAGTCACGATACTCCGATG
CGTCTAGGGATCCCAATCTTCGCGGTTGGAAAGGTGGGTGGGTGGATTTCGATTTCGCCGAAGAGGGTTCGTG
TGTTCATTGCGCTGTTATTAAGTTCGTTCATTTTTTTGCCTGATCGGCCCTTTCGGTTGGCAGATTTGTATCTCG
GGTTCGAGGAGTTCGAAGGAGGAGAAGATGACCCCCGCGCGGAGTTCCCCTGCCCATTCTGCTCTGAGGATTT
CGACATCGTCGGACTCTGCTGCCACATCGATGA

(E) KW-STP (615 bp)

TCTGTGTCCTCACCAAAACGTAAGTATAATATTTTAGGCAACGAAACATTAAGCTTTATCCAATATCAGGGAC
AGTGAAGTATTCCTCATCATCAGATATCTCAAGCCTAGGCTTCTTGAATGTTGTGCCATGTTTGTTAGCCCTGT
GGAAACCTGTGTTGCTTGCTTGTGATGTAGGGAGTTCAGGAACTTGAGCTGCTGCTTGGGCACCCTTTGGCGTT
TCATCATGGAAAAGACCAATGTCTTCAAACCATTCCAACTCTCCAAATCCAACAGGAGACTCCTGGAACAGGT
TGGATCAGAGTATGACCATCTTATGTGAACTACTTTAAATGGAAGTAAATCGACATATCATACCTTTTCACCA
ATTTCTTGGTCTGAAAGCTGAAGGAACTCATCAACAGCCCATGCAGAATGCATGAATGATGGTGTGCTCTGTT
GTGTTGCAGGAACTTTTGTTGCAATCAAGGCCGGTCCATGGCGTGGTGGTTCAGTGTTGTCCTTTAGGTCCTTG
TCGCACATCTTACTTACGGCCACACGGATTCCAGTAGCTAAATAACGCTGGTGGTTGCCGGAAAGGGTACCAG
CGACATGAACAGGTTCATCACAGTCCCG

(E) KW-LTIP1 (615 bp)

AAATGCGGCACTTTTCATTCCAATGTCTTATCAGCAGGTAGAAGATTTCTCCCCTGAATGGTGGGAGTTGGTG
AAGACAACCCCATGGTTTCGTGAGCACTGGTTCCATCAGTATCAGGAGCAAGAAACATTTGGCAACGATGAG
GAGGACATGGCAGACGCGTTGCCAGACACTTTCGATCTTGGAATCACTGAATTCTCTTTGCTGGAGGCAGAGT
CCGATGGGGCAGCTTACCATGTACCTGGGAAAGAGATGATTTCTGGTGCTACAGAGCGAGAGAAGATAGAGC
ATGCTATTCAACTGTATTGCAGTTTGATGGAAGGTTATGTATACACCAAAATTGCAACCATATGCAATTTACA
GTTGCTTTGTCAATTGAAGCCTCTGTGGTAGTGAATCTAATTCTTTTCCTCTCTGATTGGTTATATGAGACAGA
GCCCTATAGCGATGCAAAGGCAGCAATCAAGGATCTGAGCTTGAATTCTCCGAAGAATGGTGTTAAGCCATTG
TCTGTTCCAGCAAAGTATCGGGAGAAGCCTCTGCAATGCGTTAGTCCAAAGTACAGCCCCCGGCGCGTCATTC
ACCAGCCACGCTGAGAGATGCGAGCCTTGTTC

Figure 6.8. PCR products of five abiotic primers amplified from genomic DNA of *M. balbisiana* 'Klutuk Wulung', cloned in pGEM-T-Easy vector. Clones are labelled as (A) Clone KW_70HSP (size 440bp), (B) Clone KW_DRFP (size 327 bp), (C) Clone KW_HSP (size 456 bp), (D) Clone KW_STP (size 615 bp) and (E) Clone KW_LTIP (size 615 bp).

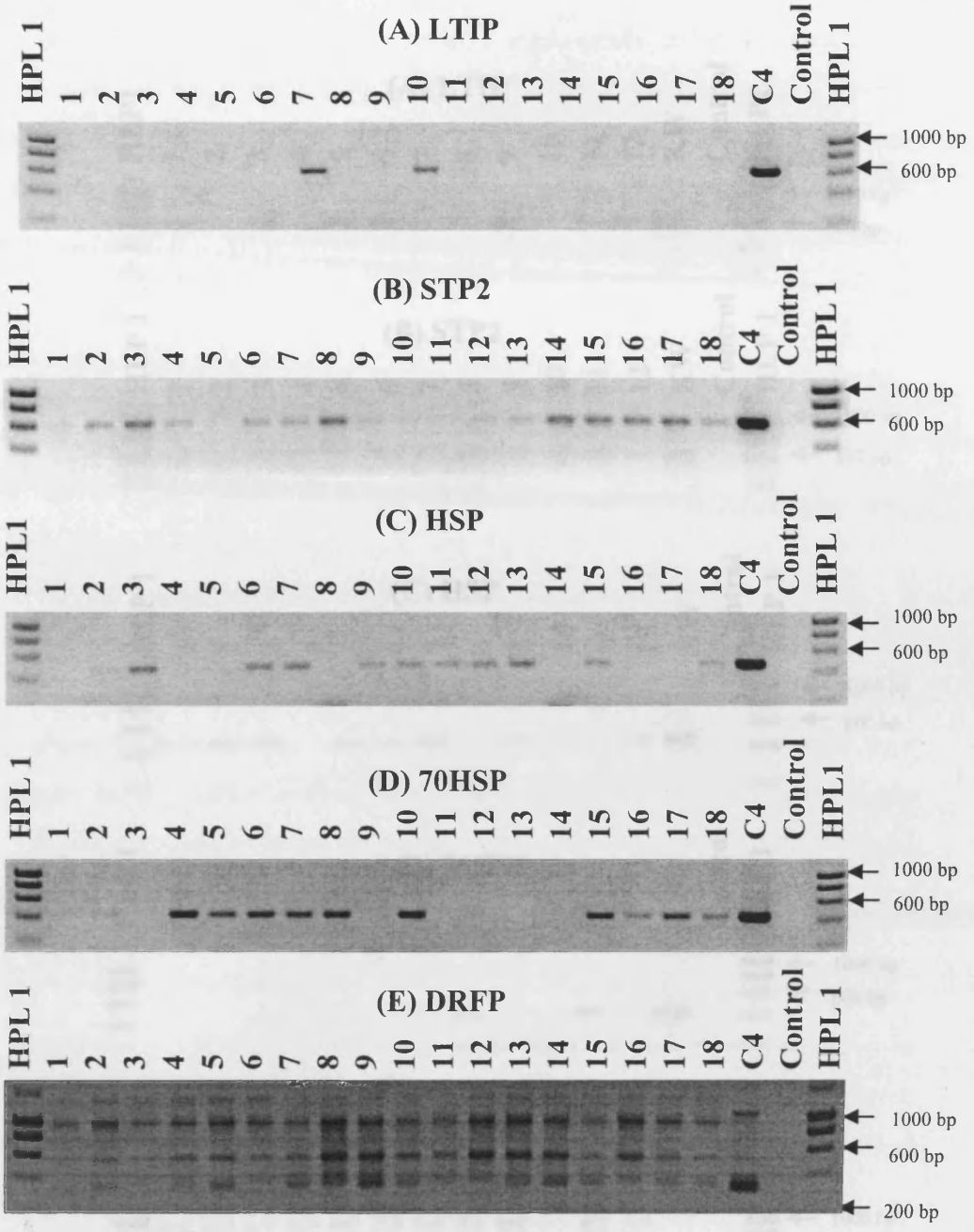


Figure 6.9. Screening of 18 MA4 BAC super-pools (*M. acuminata* 'Calcutta-4' BAC libraries) by PCR-based technique using abiotic stress primers. (A) Low temperature Induced Protein (LTIP primers), (B) Salt Tolerance Protein (STP2 primers), (C), Class-1 LMW Heat Shock Protein (HSP primers), (D) 70 kDa Heat Shock Protein (70HSP primers) and (E) Drought Responsive Family Protein (DRFP). Line 1-18 are BAC super-pool labeled as MA4_1-8, MA4_9-16, MA4_17-24, MA4_25-32, MA4_33-40, MA4_ MA4_41-48, MA4_49-56, MA4_57-64, MA4_65-72, MA4_73-80, MA4_81-88, MA4_89-96, MA4_97-104, MA4_105-112, MA4_113-120, MA4_121-128, MA4_129-136, MA4_137-144. Line 19 and 20 are positive control using genomic DNA *M. acuminata* 'Calcutta-4' (C4) and negative control without DNA respectively. HyperLadder 1 (HPL 1) are located at both side of the gel.

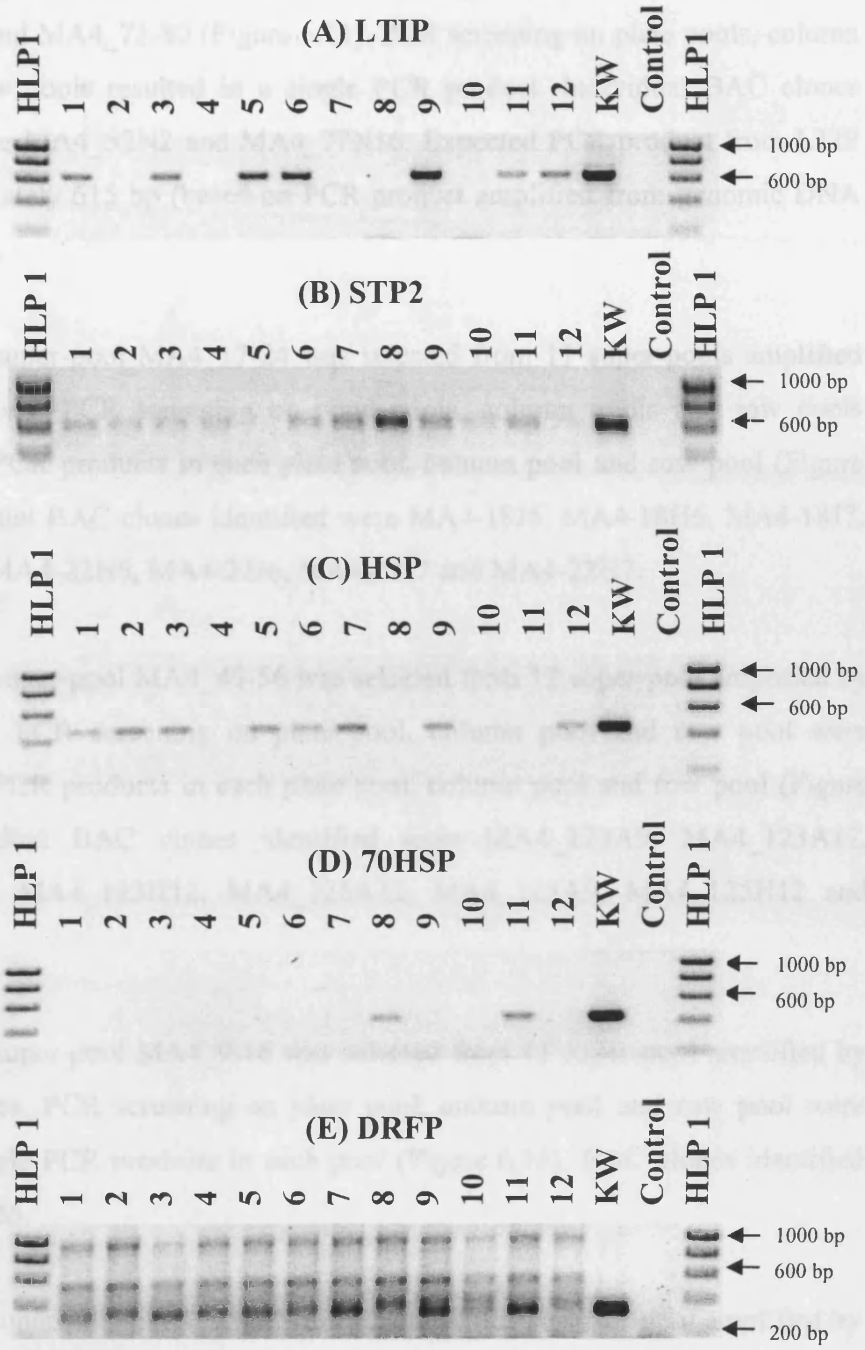


Figure 6.10. Screening of 12 MBP (*M. balbisiana* 'Klutuk Wulung' BAC libraries) BAC super-pools by PCR-based technique using abiotic stress primers. (A) Low temperature Induced Protein (LTIP primers), (B) Salt Tolerance Protein (STP2 primers), (C), Class-1 LMW Heat Shock Protein (HSP primers), (D) 70 kDa Heat Shock Protein (70HSP primers) and (E) Drought Responsive Family Protein (DRFP). Lanes 1-12 are BAC super-pool labelled as MBP_1-8, MBP4_9-16, MBP_17-24, MBP_25-32, MBP_33-40, MBP_ 41-48, MBP_49-56, MBP_57-64, MBP_65-72, MBP_73-80, MBP_81-88 and MBP_89-96. Lanes 13 and 14 are positive controls using genomic DNA *M. balbisiana* 'Klutuk Wulung' and negative control without DNA respectively. HyperLadder 1 (HPL 1) are located at both side of the gel.

LTIP primers amplified two positive super-pools in MA4 library i.e. MA4_49-56 and MA4_73-80 (Figure 6.11). PCR screening on plate pools, column pools and row pools resulted in a single PCR product. Individual BAC clones identified were MA4_52N2 and MA4_77N16. Expected PCR product from LTIP was approximately 615 bp (based on PCR product amplified from genomic DNA of Calcutta-4).

BAC super-pool MA4_17-24 was selected from 11 super-pools amplified by HSP primers. PCR screening on plate pools, column pools and row pools resulted two PCR products in each plate pool, column pool and row pool (Figure 6.12). Individual BAC clones identified were MA4-18J6, MA4-18H6, MA4-18J7, MA4-18H7, MA4-22H6, MA4-22J6, MA4-22J7 and MA4-22H7.

BAC super-pool MA4_49-56 was selected from 12 super-pool amplified by STP primers. PCR screening on plate pool, column pool and row pool were resulted two PCR products in each plate pool, column pool and row pool (Figure 6.13). Individual BAC clones identified were MA4_123A9, MA4_123A12, MA4_123H9, MA4_123H12, MA4_125A12, MA4_125A9, MA4_125H12 and MA4_125H9.

BAC super-pool MA4_9-16 was selected from 11 super-pool amplified by 70HSP primers. PCR screening on plate pool, column pool and row pool were resulted a single PCR products in each pool (Figure 6.14). BAC clones identified was MA4-12B6.

BAC super-pool MA4_25-32 was selected from 10 super-pool amplified by DRFP primers. PCR screening on plate pool, column pool and row pool were resulted a single PCR products in each pool (Figure 6.17). Individual BAC clones identified was MA4-30C11.

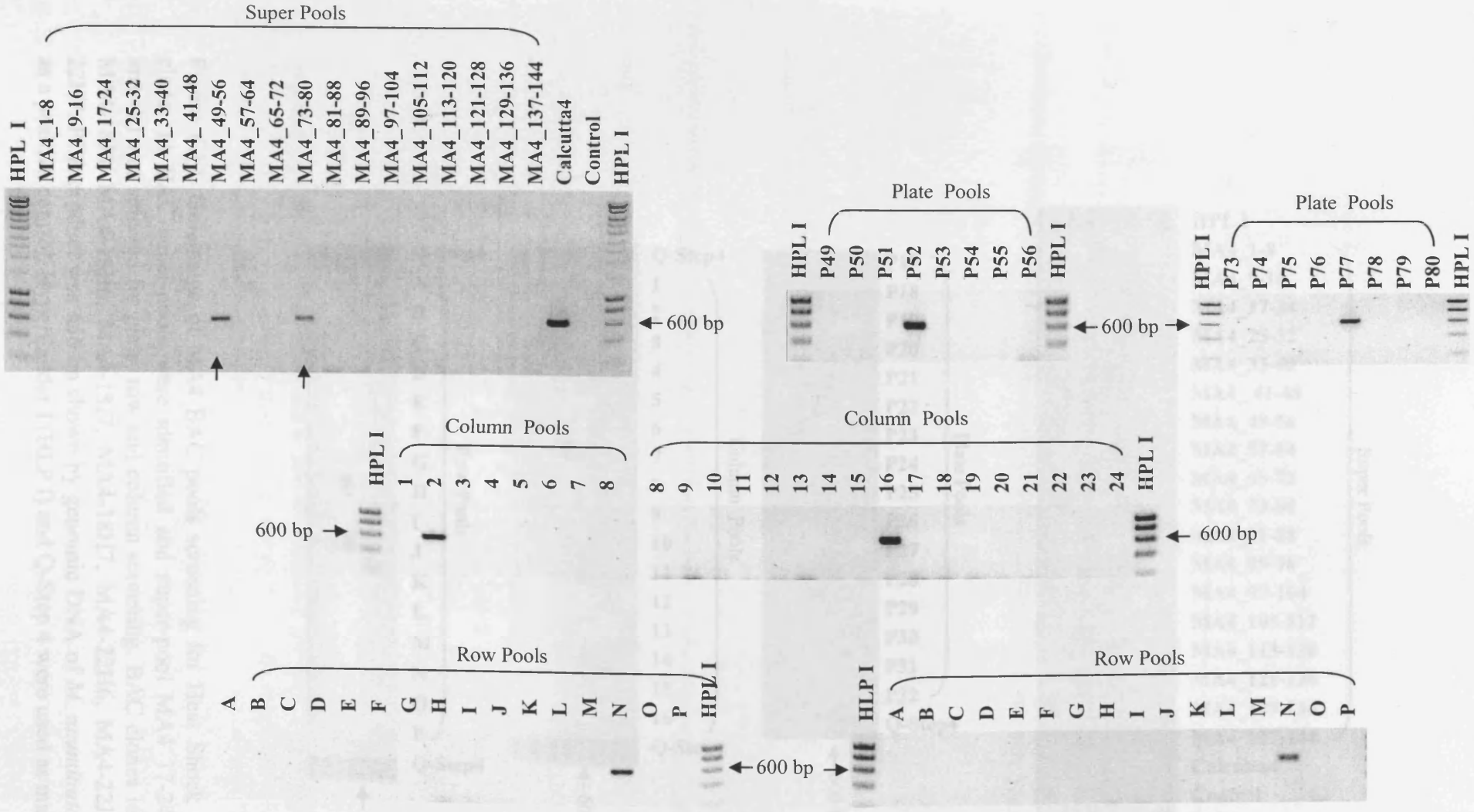


Figure 6.11. Gel image of MA4 BAC pools screening for Low Temperature Induced Protein BAC clone. 2 BAC super-pools were identified super-pool MA4_49-56 and super-pool MA4 73-80 (marked by arrows) were selected for plate, row and column screening. BAC clones identified was P52N2 and P77N16 respectively. PCR product was 615 bp shown by genomic DNA of *M. acuminata* 'Calcutta-4' as a positive control. Hyper Ladder I (HLP I) was used as markers.

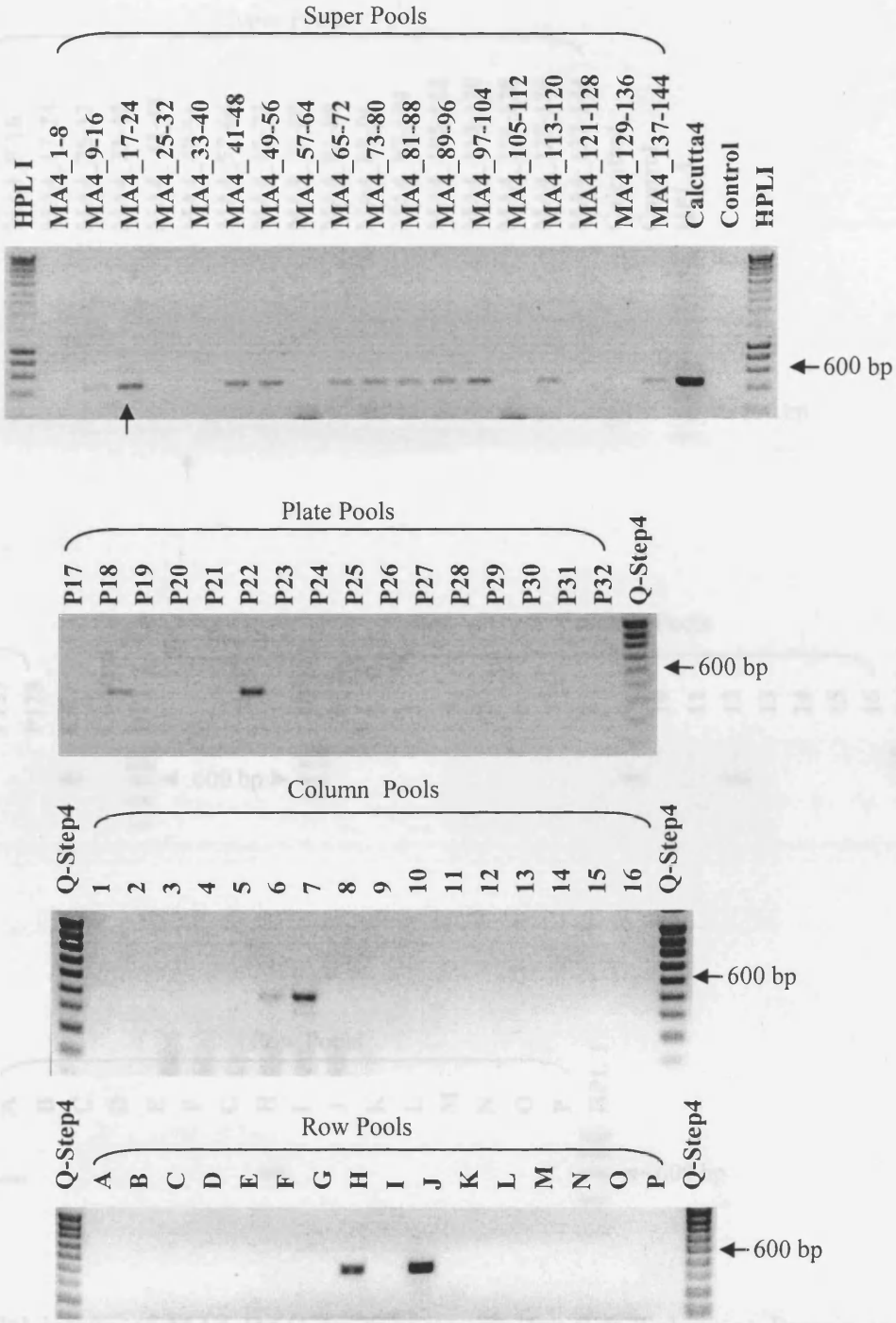


Figure 6.12. Gel image of MA4 BAC pools screening for Heat Shock Protein BAC clone. 11 BAC super-pools were identified and super-pool MA4_17-24 (marked by arrows) was selected for plate, row and column screening. BAC clones identified were MA4-18J6, MA4-18H6, MA4-18J7, MA4-18H7, MA4-22H6, MA4-22J6 and MA4-22H7. PCR product was 456 bp shown by genomic DNA of *M. acuminata* 'Calcutta-4' as a positive control. HyperLadder I (HLP I) and Q-Step 4 were used as markers.

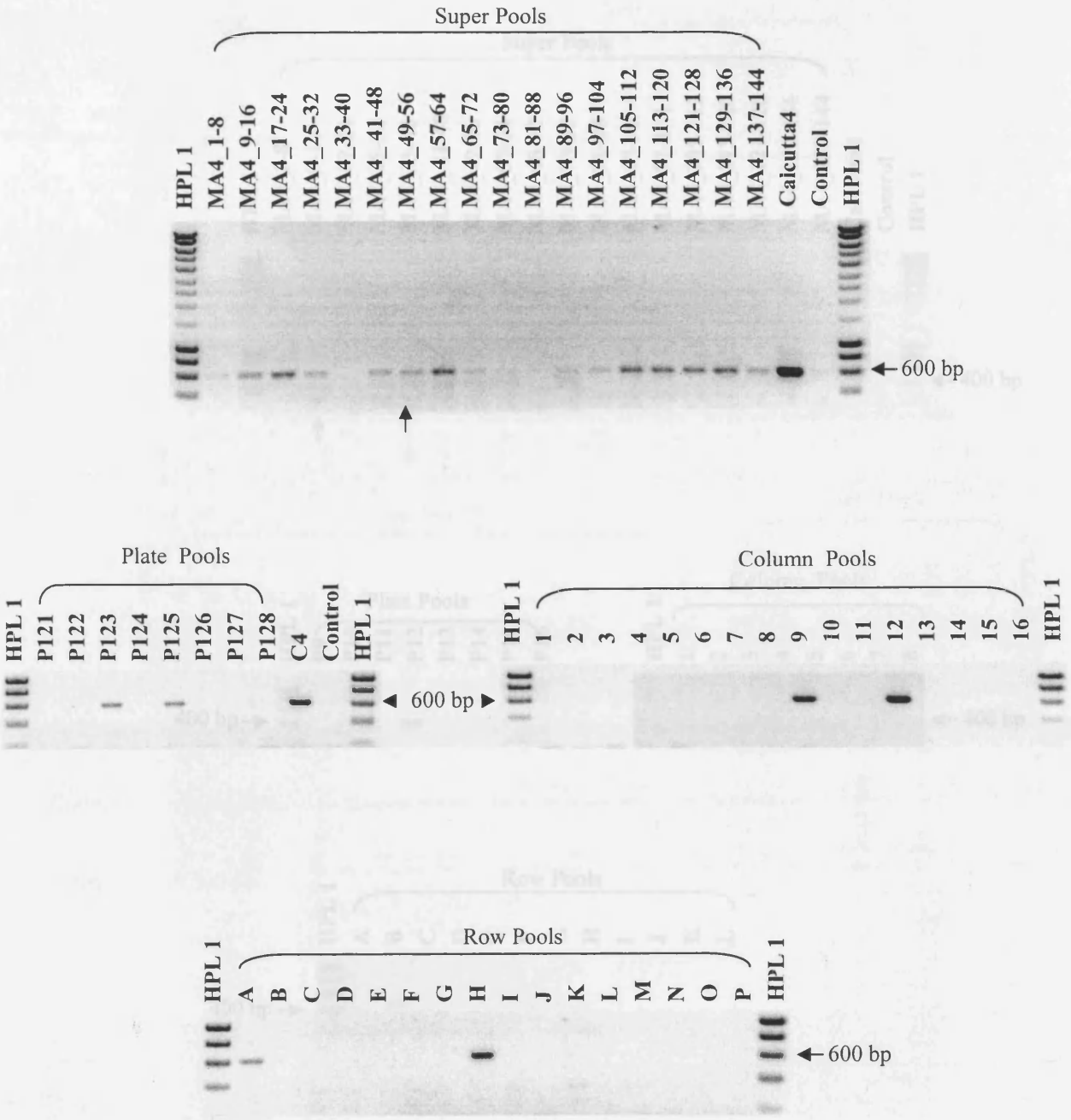


Figure 6.15 Gel image of MA4 BAC pools screening for Salt Tolerance Protein in BAC library. 11 BAC super-pools were identified and super-pool MA4_49-56 (marked by arrows) was selected for plate, row and column screening. BAC clones identified was MA4-123A9, MA4-123A12, MA4-123H9, MA4-123H12, MA4-125A12, MA4-125A9, MA4-125H12 and MA4-125H9. PCR product was 615 bp shown by genomic DNA of *M. acuminata* 'Calcutta-4' as a positive control. HyperLadder 1 (HPL 1) was used as markers.

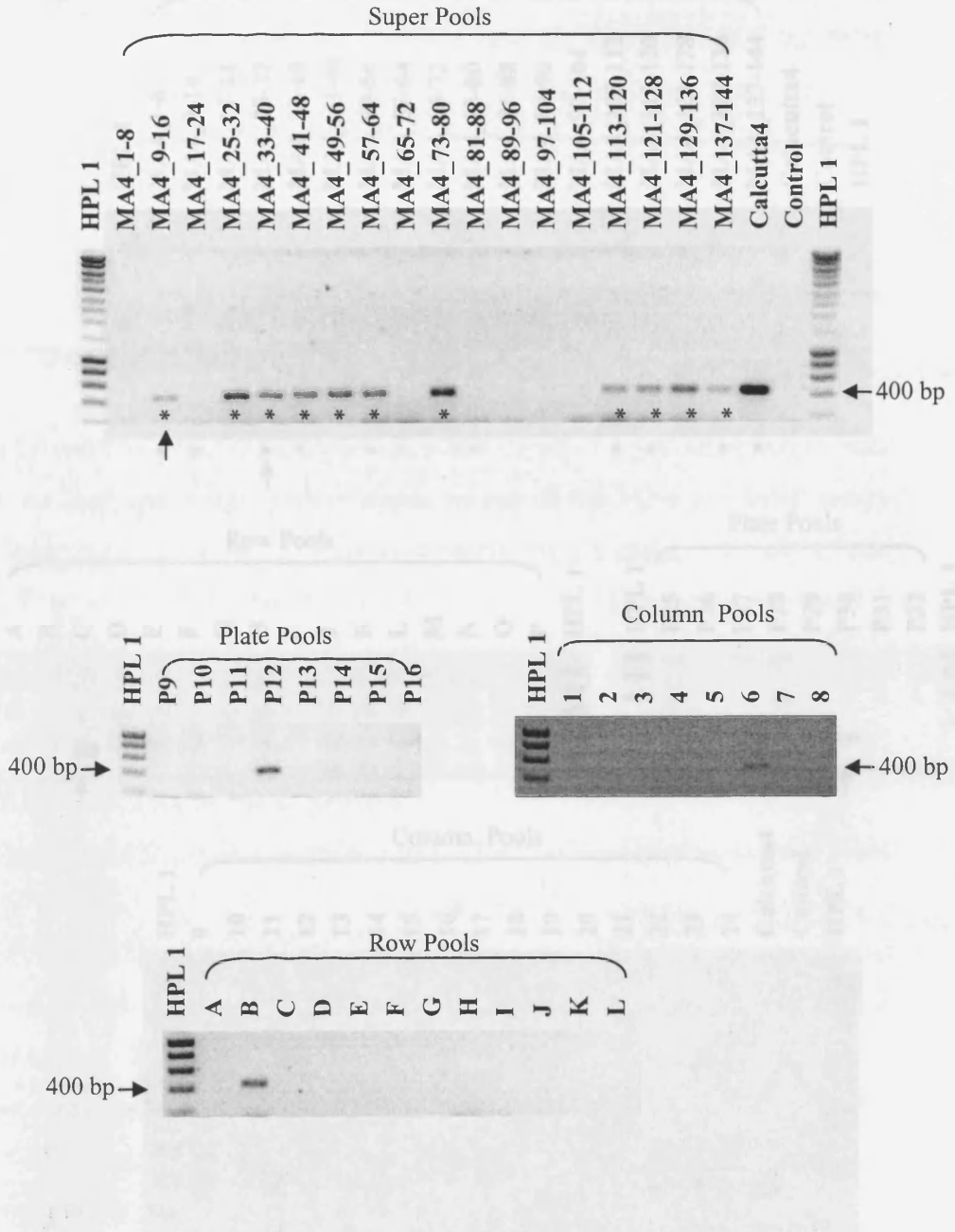


Figure 6.14. Gel image of MA4 BAC pools screening for 70kDa Heat Shock Protein BAC clone. 11 BAC super-pools were identified (marked by asterisks) and super-pool MA4_9-16 (marked by arrows) was selected for plate, row and column screening. BAC clones identified was MA4-12B6. PCR product was approximately 440 bp shown by genomic DNA of *M. acuminata* 'Calcutta-4' as a positive control. Hyper Ladder I (HLP I) and Q-Step 4 were used as markers.

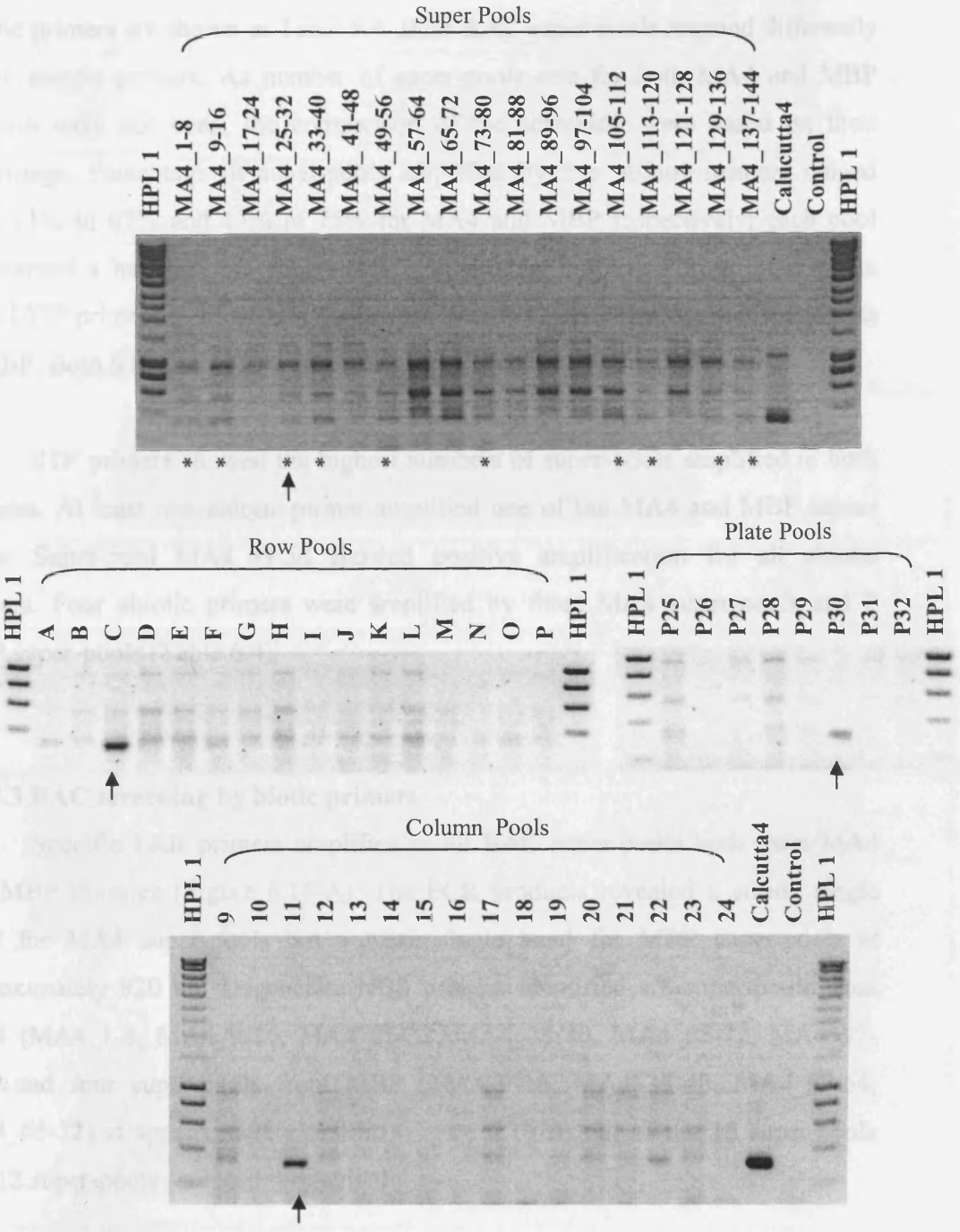


Figure 6.15. Gel image of BAC screening in identifying the Drought Responsive Family Protein (DRFP) clones by PCR-based technique using DRFP primers. BAC super pools identified 10 prospective clones marked by asterisks (MA4_1-8, MA4_9-16, MA4_25-32, MA4_33-40, MA4_49-56, MA4_73-80, MA4_105-112, MA4_113-120, MA4_129-136, MA4_137-144). Row pool and column pool amplified (marked by arrows) from Plate pool MA4_25-32 identified the individual BAC clone for DRFP as MA4-30-C11. Genomic DNA of *M. acuminata* 'Calcutta-4' and negative control are positive and negative control respectively. Hyper ladder 1 (HPL I) are located at both side of the gel.

Summary of PCR products of MA4 and MBP super-pools amplified by five abiotic primers are shown in Table 6.4. Both BAC super-pools respond differently to the abiotic primers. As number of super-pools sets for both MA4 and MBP libraries were not even, the comparison of the screening were based on their percentage. Percentage of super-pools amplified by five abiotic primers ranged from 11% to 67% and 17% to 75% for MA4 and MBP respectively; each pool represented a half-genome equivalent. PCR product amplified from super-pools from LTIP primers in MA4 was the lowest, whereas 70HSP primer was the lowest in MBP. Both STP and HSP were equally amplified in their super-pools.

STP primers showed the highest numbers of super-pools amplified in both libraries. At least one abiotic primer amplified one of the MA4 and MBP super-pools. Super-pool MA4_49-56 showed positive amplification for all abiotic primers. Four abiotic primers were amplified by three MA4 super-pools and 2 MBP super-pools (Table 6.4).

6.3.2.3 BAC screening by biotic primers

Specific LRR primers amplified in all BAC super-pools both from MA4 and MBP libraries (Figure 6.16-A). The PCR products revealed a strong single band for MA4 super-pools but a weak single band for MBP super-pools at approximately 820 bp. Degenerate NBS primers identified six super-pools from MA4 (MA4_1-8, MA4_9-16, MA4_25-32, MA4_33-40, MA4_65-72, MA4_97-104) and four super-pools from MBP (MA4_9-16, MA4_33-40, MA4_57-64, MA4_65-72) at approximately 520 bp (Figure 6.16-B) out of the 18 super-pools and 12 super-pools screened respectively.

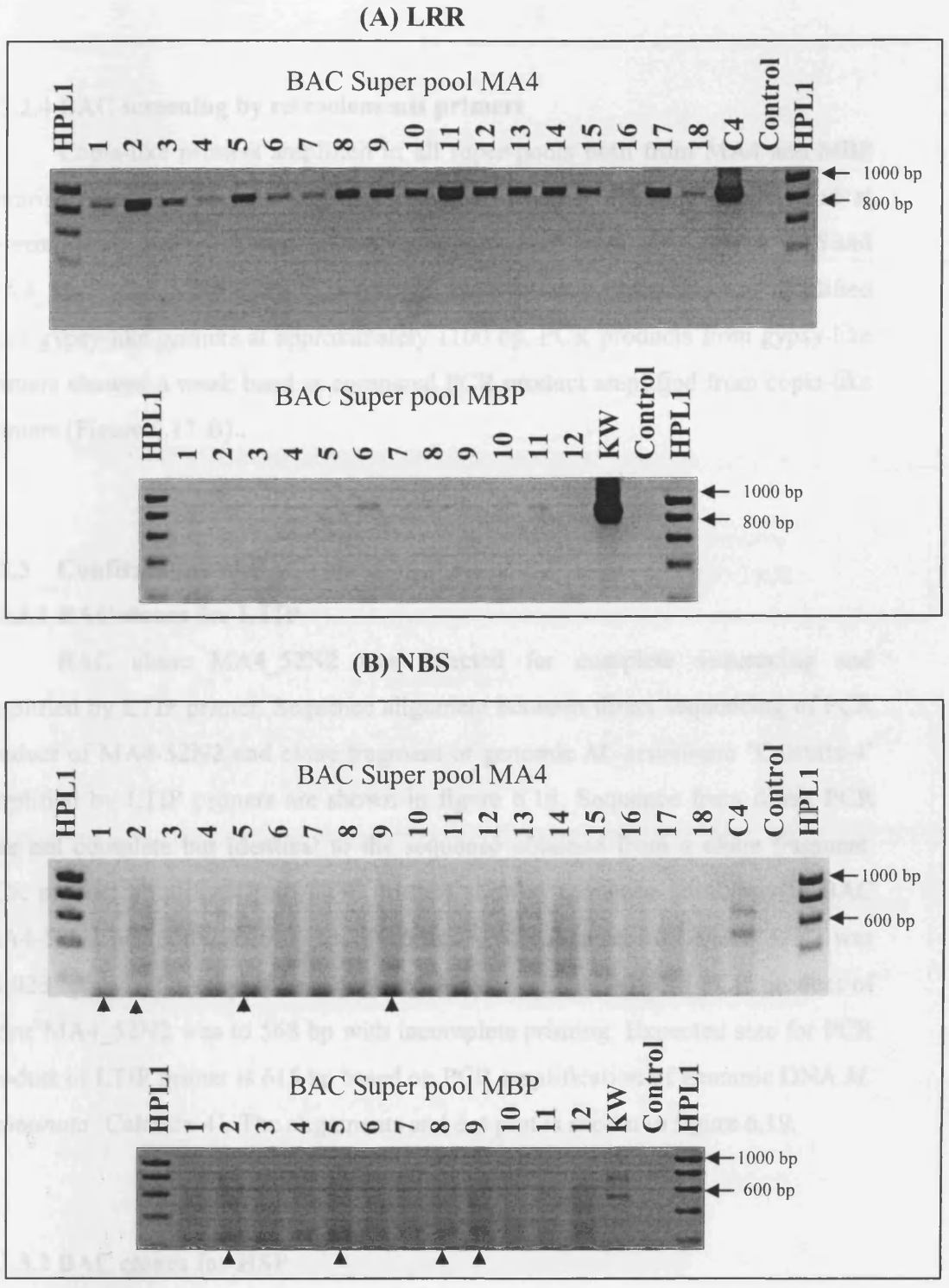


Figure 6.16. Screening of 18 MA4 BAC super-pools (*M. acuminata* 'Calcutta-4') and 12 MBP super-pools (*M. balbisiana* 'Klutuk Wulung') by PCR-based technique to identify (A) LRR motifs and (B) NBS domain. Line 1-18 (Gel image for MA4) and 1-12 (Gel image for MBP) are labelled corresponding to BAC super-pool 1-8, 9-16, 17-24, 25-32, 33-40, 41-48, 49-56, 57-64, 65-72, 73-80, 81-88, 89-96, 97-104, 105-112, 113-120, 121-128, 129-136, 137-144. MA4 and MBP BAC libraries consist of LRR motifs in all super-pools. BAC super-pool amplified by NBS primers were marked with arrows. The second and the last line before the ladder positive control using genomic DNA *M. acuminata* 'Calcutta-4' or *M. balbisiana* 'Klutuk Wulung' and negative control without DNA respectively. HyperLadder 1(HPL1) are located at both side of the gel.

6.3.2.4 BAC screening by retroelements primers

Copia-like primers amplified in all super-pools both from MA4 and MBP libraries (Figure 6.17-A). The PCR products revealed a strong single band at approximately 300 bp. However, two super-pools each from MA4 (MA4_9-16 and MA4_33-4) and MBP (MBP_73-80 and MBP_89-96) libraries were amplified from gypsy-like primers at approximately 1100 bp. PCR products from gypsy-like primers showed a weak band as compared PCR product amplified from copia-like primers (Figure 6.17-B)..

6.3.3 Confirmation of BAC clones

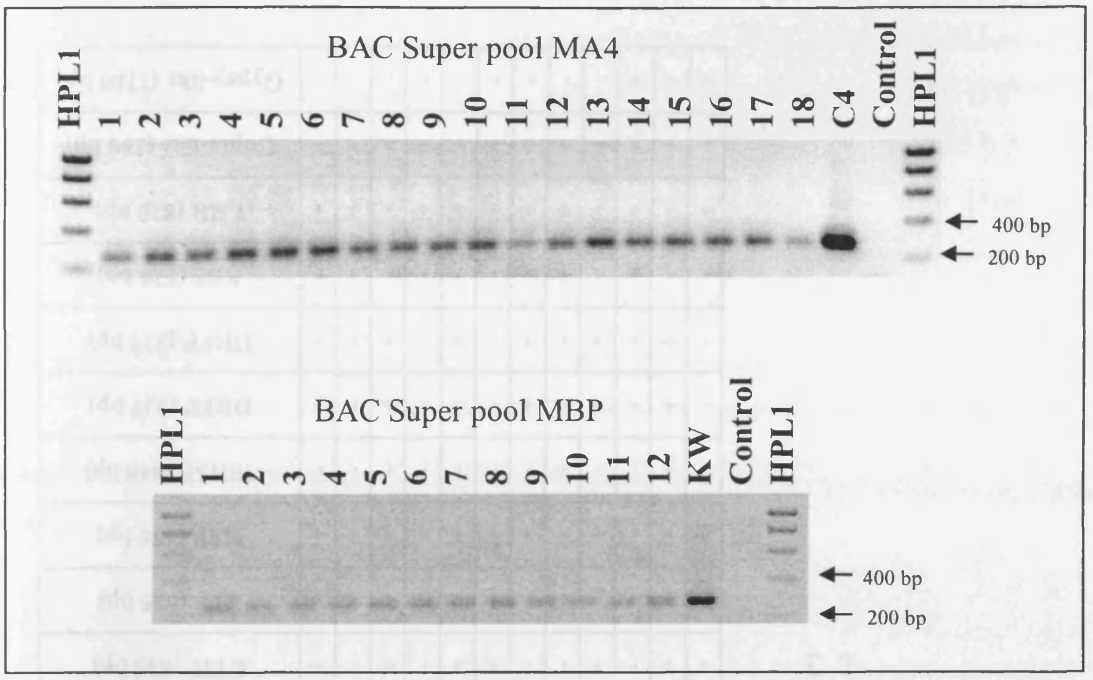
6.3.3.1 BAC clones for LTIP

BAC clone MA4_52N2 was selected for complete sequencing and amplified by LTIP primer. Sequence alignment between direct sequencing of PCR product of MA4-52N2 and clone fragment of genomic *M. acuminata* 'Calcutta-4' amplified by LTIP primers are shown in figure 6.18. Sequence from direct PCR was not complete but identical to the sequence obtained from a clone fragment. PCR product amplified from LTIP primers showed sequence homology in BAC MA4-52N2 position 27010 bp to 27466 bp. A complete size for MA4_52N2 was 73,023 bp (www.musagenomics.org) and direct sequencing from PCR product of clone MA4_52N2 was to 568 bp with incomplete priming. Expected size for PCR product of LTIP primer is 615 bp based on PCR amplification of genomic DNA *M. acuminata* 'Calcutta-4'. The alignments and dot plot is shown in figure 6.19.

6.3.3.2 BAC clones for HSP

BAC clone MA4_18J6 was selected for a complete sequence and amplified by LTIP primer. Sequence alignment between direct sequencing of PCR product of MA4-18J6 and clone fragment of genomic *M. acuminata* 'Calcutta-4' amplified by HSP primers are shown in figure 6.20.

(A) *Pseudoviridae*



(B) *Metaviridae*

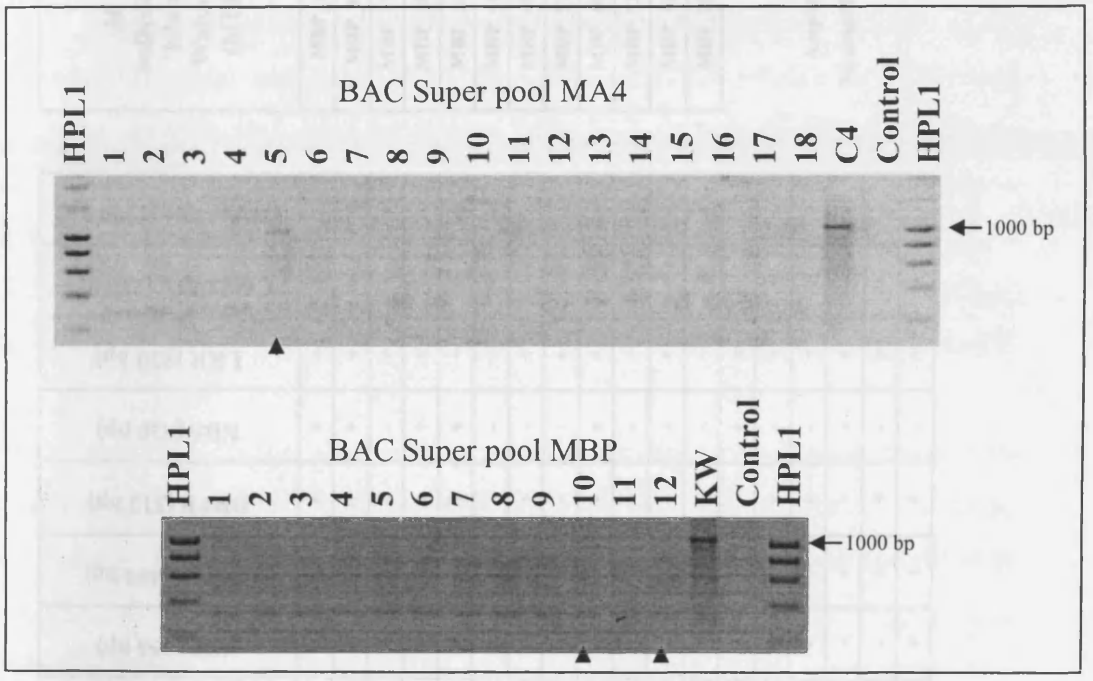


Figure 6.17. Screening of 18 MA4 BAC super-pools (*M. acuminata* 'Calcutta-4') and 12 MBP super-pools (*M. balbisiana* 'Klutuk Wulung') by PCR-based technique to identify (A) *Pseudoviridae* and (B) *Metaviridae* retroelements. Line 1-18 (Gel image for MA4) and 1-12 (Gel image for MBP) are labelled corresponding to BAC super-pool 1-8, 9-16, 17-24, 25-32, 33-40, 41-48, 49-56, 57-64, 65-72, 73-80, 81-88, 89-96, 97-104, 105-112, 113-120, 121-128, 129-136, 137-144. All BAC super-pool for MA4 consists of copia-like retroelements. Positive BAC super-pools MBP consists of gypsy-like retroelements are marked with arrows. The second and the last line before the ladder positive control using genomic DNA *M. acuminata* 'Calcutta-4' or *M. balbisiana* 'Klutuk Wulung' and negative control without DNA respectively. HyperLadder 1 (HPL1)are located at both side of the gel.

Table 6.4. *M. acuminata* 'Calcutta-4' (MA4) and *M. balbisiana* 'Klutuk Wulung' (MBP) BAC super-pools amplified by abiotic, biotic and retroelement primers.

Super-pool Sets	*M. acuminata* 'Calcutta4' (MA4)	LTIP (615 bp)	STP (615 bp)	HSP (456 bp)	70HSP (440 bp)	DRFP (312 bp)	NBS (520 bp)	LRR (820 bp)	Copia-like (280 bp)	Gypsy-like (1100 bp)
1	MA4_1-8	-	+	-	+	+	+	+	+	-
2	MA4_9-16	-	+	+	+	+	+	+	+	+
3	MA4_17-24	-	+	+	-	-	-	+	+	-
4	MA4_25-32	-	+	-	+	+	+	+	+	-
5	MA4_33-40	-	-	-	+	+	+	+	+	+
6	MA4_41-48	-	+	+	+	-	-	+	+	-
7	MA4_49-56	+	+	+	+	+	-	+	+	-
8	MA4_57-64	-	+	-	+	-	-	+	+	-
9	MA4_65-72	-	-	+	-	-	+	+	+	-
10	MA4_73-80	+	-	+	+	+	-	+	+	-
11	MA4_81-88	-	-	+	-	-	-	+	+	-
12	MA4_89-96	-	-	+	-	-	-	+	+	-
13	MA4_97-104	-	+	+	-	-	+	+	+	-
14	MA4_105-112	-	+	-	-	+	-	+	+	-
15	MA4_113-120	-	+	+	+	+	-	+	+	-
16	MA4_121-128	-	+	-	+	-	-	+	+	-
17	MA4_129-136	-	+	-	+	+	-	+	+	-
18	MA4_137-144	-	+	+	+	+	-	+	+	-

M. balbisiana 'Klutuk Wulung' (MBP)	LTIP (615 bp)	STP (615 bp)	HSP (456 bp)	70HSP (440 bp)	DRFP (327 bp)	DRFP (327 bp)	NBS (520 bp)	LRR (820 bp)	Copia-like (280 bp)	Gypsy-like (1100 bp)
MBP_1-8	+	+	+	-	+	+	-	+	+	-
MBP_9-16	-	+	+	-	-	-	+	+	+	-
MBP_17-24	+	+	+	-	+	+	-	+	+	-
MBP_25-32	-	+	-	-	-	-	-	+	+	-
MBP_33-40	+	-	+	-	+	+	+	+	+	-
MBP_41-48	+	+	+	-	-	-	-	+	+	-
MBP_49-56	-	+	+	-	+	+	-	+	+	-
MBP_57-64	-	+	-	+	-	-	+	+	+	-
MBP_65-72	+	+	+	-	+	+	+	+	+	-
MBP_73-80	-	+	-	-	-	-	-	+	+	+
MBP_81-88	+	+	-	+	+	+	-	+	+	-
MBP_89-96	+	-	+	-	-	-	-	+	+	+

Notes:

Amplification (+)

No amplification (-)

Sequence alignment between direct sequencing of PCR product of MA4-18J6 and clone fragment of genomic *M. acuminata* 'Calcutta-4' amplified by HSP primers are shown in figure 6.21. Sequence from direct PCR was not complete but identical to the sequence cloned fragment. PCR product amplified from HSP primers showed sequence homology in BAC MA4-18J6 position 48445 bp to 49060 bp. A complete size for MA4_18J6 was 76,327 bp (www.musagenomics.org) and direct sequencing from PCR product of clone MA4_18J6 was 315 bp with incomplete priming. Expected size for PCR product of HSP primer is 456 bp based on PCR amplification of genomic DNA *M. acuminata* 'Calcutta-4'.

6.4 Discussion

6.4.1 Primer design and sequence homologies

Degenerate (NBS and retroelements primers) and specific primers (LRR and abiotic primers) used in the PCR-based screening were reliable for PCR-based screening. In designing specific primers from EST sequences, Primer3 software was found to be reliable and convenient as it takes considerations in many factors such as oligo melting temperature, significant product length relative to particular interest regions, GC content, 3' stability, predicted secondary structure, primer-dimer formation between primers and options to give optimal PCR product lengths.

Smearing results obtained from NBS domain amplification from BAC clones might be related to unspecific target and multiple copies of the target sequence in BAC. Relatively screening of retroelements in BAC showed different band intensity from MA4 and MBP libraries. This might be related to copy number in both libraries and MA4 has more copy numbers than MBP. Comparison with positive control for both *Pseudoviridae* and *Metaviridae* significantly showed the difference, which support the contribution of the copy number in the retroelements. The finding can be related to genome size of *M. acuminata* 'Calcutta-4' a diploid 'A-genome' which is bigger than *M. balbisiana* 'Klutuk Wulung' a diploid 'B-genome'.

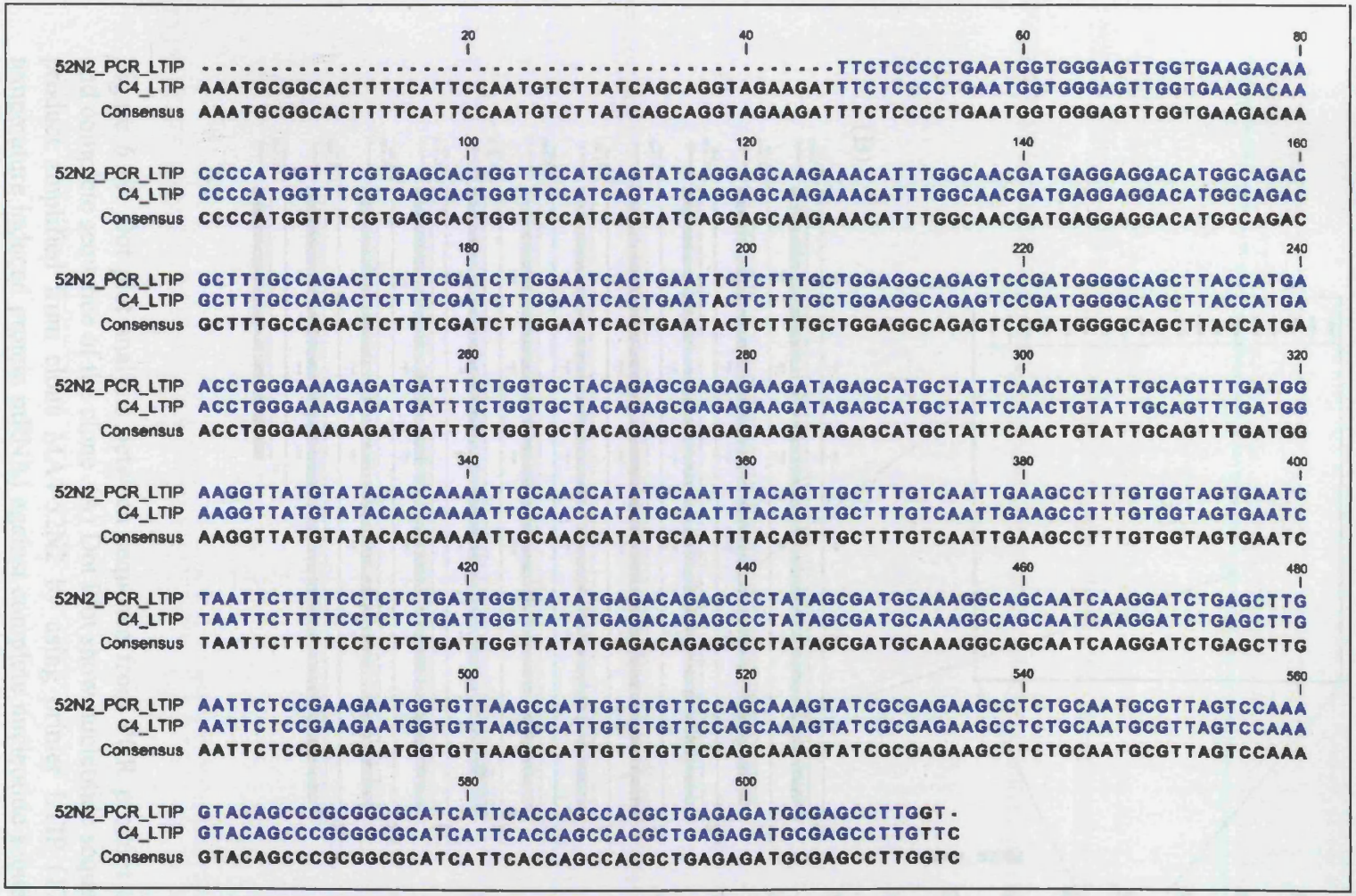


Figure 6.18. Sequence alignments between direct PCR product obtained from BAC MA4_52N2 and clone C4_LTIP amplified by LTIP primers (Low Temperature Induced Protein) LTIP-F (5'-3') : AAA TGC GGC ACT TTT CAT TC and LTIP-R (5' to 3'): GAA CAA GGC TCG CAT CTC TC. Direct PCR product produced 568 bp and expected size for PCR product of LTIP is 615 bp.

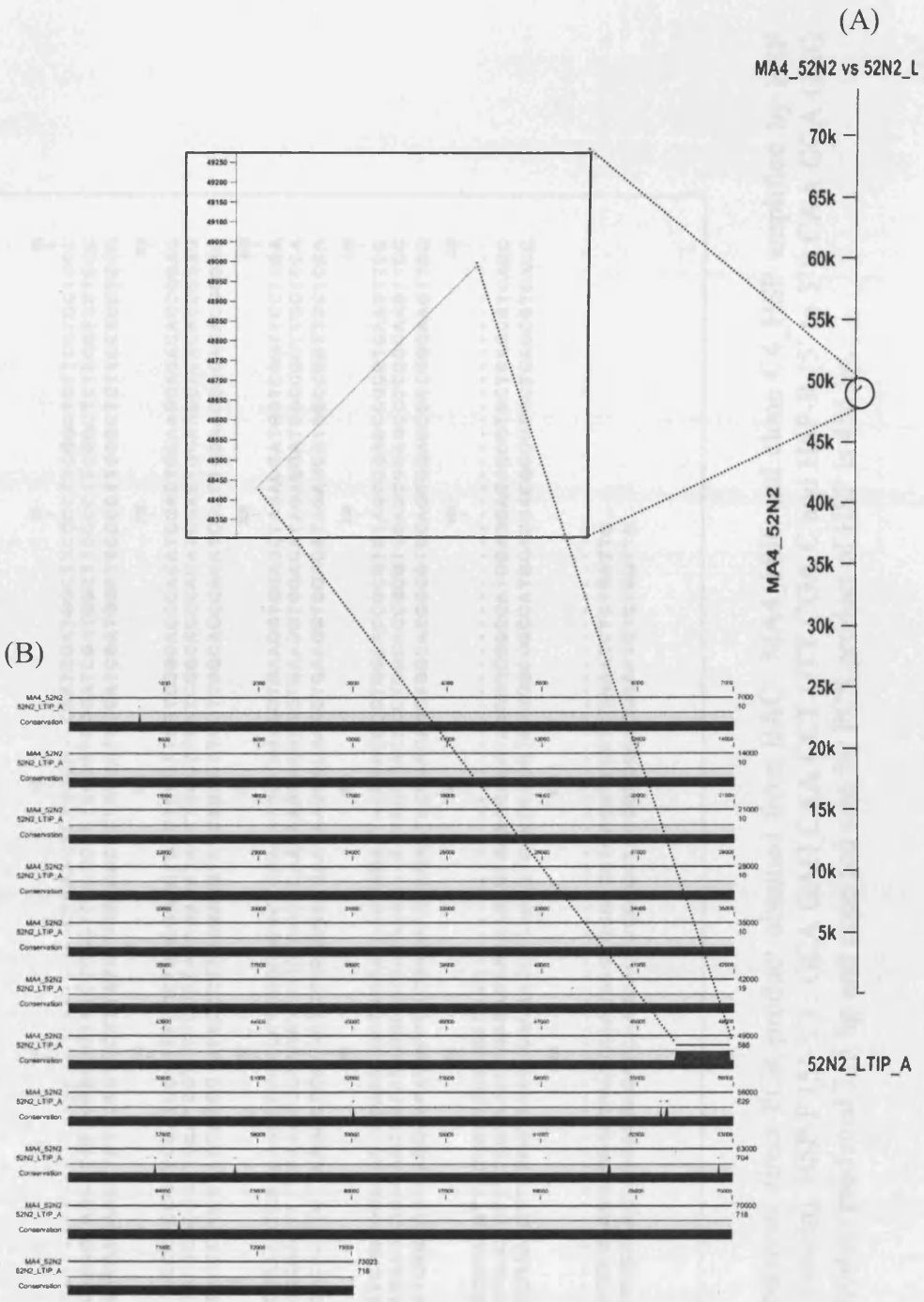


Figure 6.19. Dot plot analysis between sequence from PCR product of MA-52N2 and complete sequence of the clone (A) Dot Plot show nucleotide sequences of PCR product amplified from clone MA4-52N2 by using primer LTIP (17.7 kDa low temperature induced protein mRNA) against complete nucleotide sequence of clone MA4-52N2. (B) Alignments of the both sequences. Dot plot and alignments were provided in ***CLC Protein Workbench 3.0.***

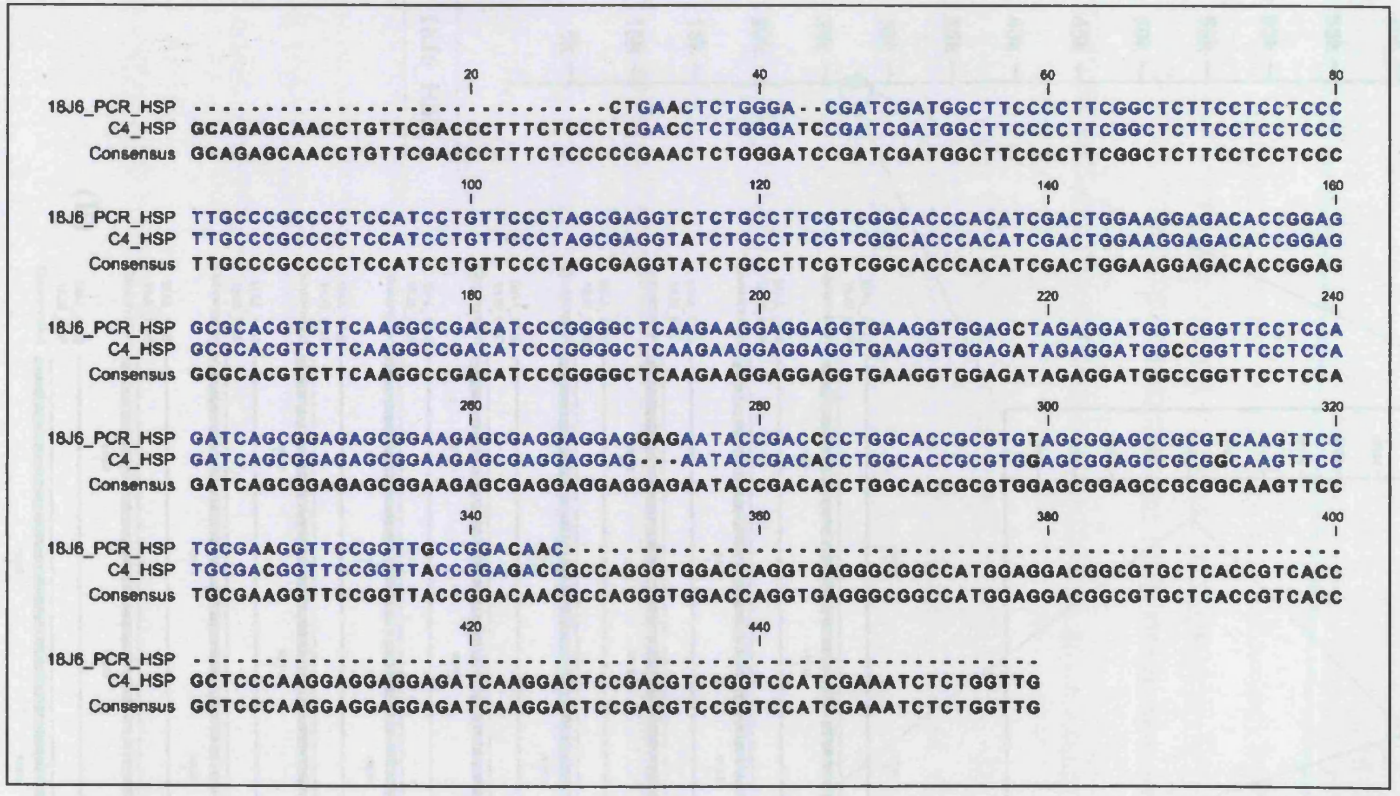


Figure 6.20. Sequence alignments between direct PCR product obtained from BAC MA4_18J6 and clone C4_HSP amplified by HSP primers (Class-1 LMW Heat Shock Protein), HSP-F (5'-3') : GCA GAG CAA CCT GTT CGA C and HSP-R (5' to 3'): CAA CCA GAG ATT TCG ATG GAC. Direct PCR product produced 315 bp and expected size for PCR product of HSP is 456 bp.

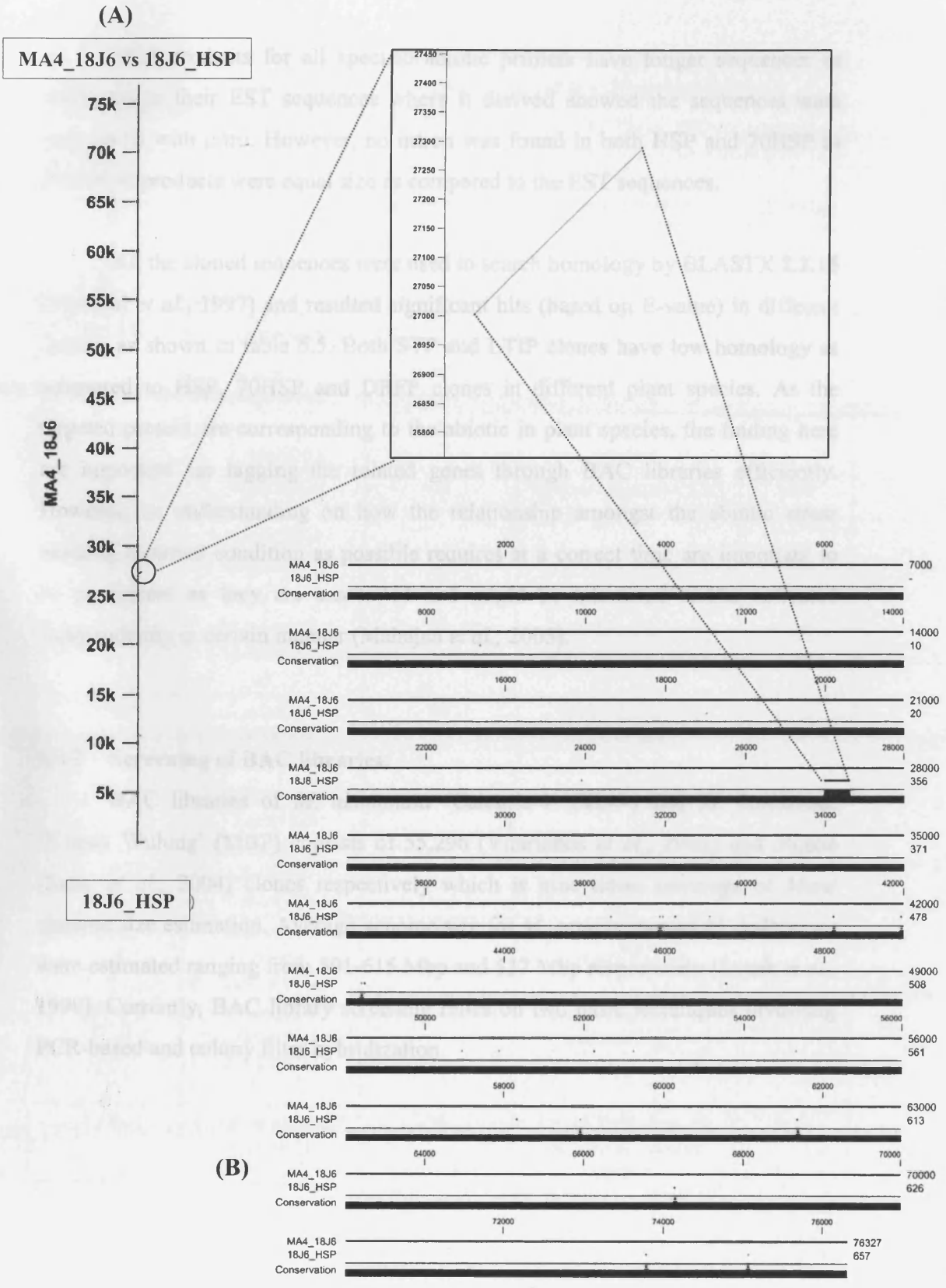


Figure 6.21 (A) Dot Plot shows nucleotide sequences of PCR product amplified from clone MA4-18J6 by using primer HSP (Heat Shock Protein) against complete nucleotide sequence of clone MA4-18J6. (B) Alignments of both sequences. Dot plot and alignments were provided in ***CLC Protein Workbench 3.0.***

PCR products for all specific abiotic primers have longer sequences as compared to their EST sequences where it derived showed the sequences ware interrupted with intro. However, no intron was found in both HSP and 70HSP as their PCR products were equal size as compared to the EST sequences.

All the cloned sequences were used to search homology by BLASTX **2.2.15** (Altschul *et al.*, 1997) and resulted significant hits (based on E-value) in different species as shown in table 6.5. Both STP and LTIP clones have low homology as compared to HSP, 70HSP and DRFP clones in different plant species. As the targeted protein are corresponding to the abiotic in plant species, the finding here are important for tagging the related genes through BAC libraries efficiently. However, in understanding on how the relationship amongst the abiotic stress working minimal condition as possible requires at a correct time are important to be considered as they are associated and might be interacted and/or activated independently in certain manner (Mahajan *et al.*, 2005).

6.4.2 Screening of BAC libraries.

BAC libraries of *M. acuminata* 'Calcutta-4' (MA4) and *M. balbisiana* 'Klutuk Wulung' (MBP) consists of 55,296 (Vilarinhos *et al.*, 2003) and 36,864 (Safar *et al.*, 2004) clones respectively which is nine times coverage of *Musa* genome size estimation. Average genome size for *M. acuminata* and *M. balbisiana* were estimated ranging from 591-615 Mbp and 537 Mbp respectively (Lysak *et al.*, 1999). Currently, BAC library screening relies on two basic techniques involving PCR-based and colony filter hybridization.

Table 6.5. Sequence homology in NCBI database from other plant species related to abiotic stress amplified from *M. acuminata* 'Calcutta-4' by using designed abiotic primers.

PCR product amplified from abiotic primer	Plant species (No. Accession)	Homology % identity	Description	E-value
STP	*Glycine max* ABB29467	48	salt-tolerance protein	2e32
	Arabidopsis NP_849598	44	SALT TOLERANCE (STO)	2e11
HSP	*Ananas comosus* AAM28293	69	class-1 LMW heat shock protein	2e57
	Pennisetum glaucum CAA63903	70	heat shock protein 17.9	4e56
	Oryza sativa BAC78583	67	heat shock protein 18	2e54
	Pisum sativum AAN74634	70	heat shock protein	6e53
70HSP	*Sandersonia aurantiaca* AAL85887	88	70 kDa heat shock protein	1e45
	Nicotiana tabacum AAP04522	87	heat shock protein 70	5e44
	Arabidopsis thaliana CAA05547	91	heat shock protein 70	6e44
DRFP	*Arabidopsis thaliana* AAL67123	75	Drought induced 19 protein (Di19)	2e10
	Gossypium hirsutum AAY43802	74	Drought induced 19 protein (Di19)	3e10
LTIP	*Oryza sativa* AAT72926	50	17.7 kDa low temperature induced protein	2e14

The first BIBAC library has demonstrated the colony filter hybridization in tagging black Sigatoka R-genes by amplifying NBS-LRR domain in *M. acuminata* Tuu Gia' using rice RGA9 primers and used as probe revealed 65 positive clones (Ortiz-Vazquez *et al.*, 2005). Hybridization pattern was found identical to another probe known as EST-TGBR where this probe originally obtained by differential display from wounded tissue of *M. acuminata* 'Tuu Gia' treated with *M. fijiensis* extracts. It is agreed that NBS-LRR is related to protein kinases but the signals obtained in the hybridization might be related to other protein kinases as the kinases and LRR motifs involve in many protein families. As described in Chapter 4, NBS domain in *Musa* sp. was found at 520 bp which does not associate to LRR domain. Thus, there is doubt about specificity of the probe used in the experiment as the expected size was 452 bp (original PCR product from rice) but the product from *M. acuminata* 'Tuu Gia' obtained was 400 bp. Further the EST-TGBR sequence was reported has little homology to resistance gene kinase from rice (Ortiz-Vazquez *et al.*, 2005).

Pooling strategy in BAC libraries showed reliability in handling the BAC screening at big scales. The pooling strategies can be employed by two-dimensional system (Gardiner *et al.*, 2004), three-dimensional system (Klein *et al.*, 2000) or four-dimensional system (Asakawa *et al.*, 1997) for screening the BAC libraries. According to Gardiner *et al.*, (2004) pooling strategies are effective when small numbers of target sequence need to be addressed to BACs but become complicated when it involved in multiple target sequences. In this experiment, both techniques were used and it was found that colony filter hybridization provided poor and ambiguous results in identification of positive BAC clones.

Three main factors played big role in colony hybridization screening. First, it requires high quality membrane filters with DNA concentration at microgram levels, second, a good probe and third control of hybridization stringency. BAC membrane filters and probe were assumed to be high quality coming from established sources. Hybridization conditions used followed from successful experiment in both Southern hybridization (Chapter 5) and in the lab for screening BAC filters (Alix *et al.*, 2005) previously.

Thus, by assumption of equivalent in hybridization properties, detection should provide same signals in all clones with specific target sequence. Colony screening will also give a quick estimate of copy number by counting the number of hybridization spots out of the thousands BACs screened. Similar information is not obtained with minimal PCR reaction strategies.

In the colony hybridization screening, failure results obtained from abiotic probes showed that the probe might be not appropriately labelled and might be related to low copy numbers and/or no target sequences in BAC clones. Even LRR motifs used as probe which can be considered has relatively high copy numbers resulted ambiguous results and were highly equivocal from poor hybridization. Clones identified based on the patterns ordered were not correlated well with the grid-spotting patterns. Ambiguous results in biotic probe with inconsistent double dot detection showed hybridization has occurred but the binding was not intense enough to be distinguished from no signal or background signal. These circumstances might be related to concentration of target sequence and/or failure in probe labelling. Strong incorporation of both radio-labelled nucleotides detected on the membrane filters after hybridization prior to X-ray film exposure was not corresponding to the results obtained. It was believed that double dots and background signals were mixed-up, making the orientation of the dot pattern ambiguous. Thus, the fourth factor might be considered leading to the wrong identification of colony hybridization is related to data interpretation (clarification and evaluation of the double-dot orientation).

Evidence of false clones selected in colony hybridization after confirmation by PCR showed that negative clones shared fragments with positive clones identified with the same markers (Cai *et al.,* 1998). Thus, confirmation of the BAC clones required PCR steps followed by sequencing to avoiding the false-positive clones. The factors altogether in the colony hybridization screening showed that the technique was not reliable under the conditions we used for the BAC screening at ultra large-scale of genomic mapping. Nevertheless, other partners in the Generation Challenge Programme have successfully identified BACs with biotic stress related genes (GCP report on web and musagenomics.org).

PCR-based screening offers several advantages over the colony hybridization especially screening big numbers of clones in a BAC library. High sensitivity and specificity in PCR technique showed a rapid, reliable, and efficient approach in targeting specific genes and/or motifs for genomic mapping. The pooling strategy showed a remarkable results as it can eliminate up to 90% of unrelated BAC clone with no targeted sequence. Numbers of expected clones for particular genes or motifs can be confirmed straight away at plate-pool screening by using mathematical conversion to ensure the pools and/or the BAC colonies are not contaminated.

Figure 6.22. Mathematical conversion of identifying positive BAC clones from PCR-based screening at 'plate-pool' level.

$\mathbf{n^3}$= no of clones in BAC library

Note:
n = no of positive plate-pool in super-pool sets
e.g.: $2^3 = 8$; $1^3=1$

The calculation also can be used as quality control and as a cross-check to address any false positive amplification during PCR.

Positive controls can be used to detect clone contamination or false positives and pseudo-genes in clones based on length size of PCR product. DRFP screening showed clearly the comparison between false clones and the real positive clones.

PCR-based technique requires good primers which amplify the target gene; if there are polymorphisms, particularly in the 3' end of the primers, no colonies will be amplified. In contrast, hybridization probes will identify all BACs with 85% (depending on hybridization stringency) homology to the probe sequence (e.g. the PCR primers don't amplify from *M. textilis*).

It was noted that amount of PCR product differed between reactions using different templates. PCR as carried out here is non-quantitative, and can only be fully interpreted by presence or absence of a band. Any qualitative differences (which may arise from variation in target copy number in the template or primer-target homology differences, as well as methodological differences such as template concentration or presence of different ions in the purified template) would need to be demonstrated by real-time PCR, which is beyond the scope of this study.

6.4.3 Relative efficiency of screens

BAC screening is a labour-intensive and expensive process because of the huge number of BACs involved in the screening processes. Preparation of BAC pools either for membrane filters or DNA pools requires huge number of pipetting steps – using many more operations than the number of the clones in the library. Each pool creation with the MA4 library, for example, needs some 150,000 pipetting steps (four times the number of BAC clones).

6.5 Conclusion

In the present experiments, scoring of the filters was strenuous and tricky because of low background hybridization making alignments problematical, and background spots giving possible real signals as they overlapped with weak double dot hybridization spots. Colony hybridizations by using probe from LRR motifs were highly equivocal from poor hybridization. And further clones identified based on the patterns were not correlated well with the grid-spotting patterns.

Pooling strategy in combination with PCR-based methods was much more convenient as compared to colony hybridization screening. The PCR-based screening was found more reliable, rapid, effective and effortless with precise and significant results for a large number of clones as BAC libraries. The PCR-based technique was succeeded in identification of two BAC clones related to LTIP from the whole MA4 BAC library.

CHAPTER 7

7.0 General Discussion

7.1 Techniques

Molecular techniques were applied to reach the goals ranging from designing oligonucleotide primers, PCR-based manipulation, cloning, Southern hybridization, *in situ* hybridization and finally bioinformatics to aid the conclusive high standard of data analysis. PCR techniques might be restricted by the homology of the PCR primers, thus Southern hybridization was used to verify and understand the genomic organization and the evolution on these sequences. All techniques complementing each other and have been widely used for characterization of gene families from plant species.

PCR-based techniques have become a powerful tool in gene isolation (Chapter 4 and Chapter 5), marker systems (Chapter 3), sequence identification and BAC library (Chapter 6) screening as demonstrated throughout this study. Both degenerate and specific primers have their own advantages in fulfilling the PCR work objectives. The degenerate primers are important elements in amplifying unknown sequences or genes based on protein alignments and/or protein homologies. By introducing more 'wobbles' in the degenerate primer sequences, they are more likely to bind target template but also unwanted targets, giving multiple bands (which may include pseudo-genes).

Sizes of the target sequence, and the primer design, also affect the PCR products. Extremely high GC content will generate non-specific fragments, and the same thing occurs with GC contents much below 50% and with short primers; however, the use of primer-design packages such as primer3 minimises such problems. The specificity of degenerate primers can be increased by increasing annealing temperatures to be close to the melting temperature of the primers in fully homologous primer-target hybridization situations.

In Chapter 3 and Chapter 4, degenerate primers have proven to be very useful tools in finding new domains which related to specific orthologous genes or gene families. There is no doubt that most of the genes come from families or groups which share structural similarity and common evolutionary history. Thus, this is a reliable source in designing the degenerate primers. Through alignments of the sequences from various related proteins, conserved regions where the sharing of motifs has occurred can be identified.

In this work, target sequences were amplified by PCR from genomic DNA. The sequences were variable reflected variability within and between the *Musa* species. These results suggest that the PCR primers used were able to amplify diverse members of the target gene families, although it is unknown whether the primers were too specific and did not amplify some important family members. Thus, the technique is important as a tool in understanding variation amongst *Musa* accessions. The information is reliable and could be used to construct sequence tress, give sequences that could be integrated in genetic maps in the future, and provides information for comparison with other plant species. PCR work presented here is non-quantitative and based on interpretation of the presence or absence of a band, and copy numbers or primer homology to amplified sequences cannot be rigorously interpreted based on the band's intensity. For assessing quantitative differences, real-time PCR is strongly proposed. Southern hybridization can be used to support the PCR results but still requires an explanation whether the sequences exist in discrete clusters or throughout the genome.

Gel electrophoresis is important in analysing the PCR products. The smallest fragments move the furthest in the gel, as they pass through the small pores in the gel more readily than do the larger fragments. Fragment migration is also affected by the composition and ionic strength of the TAE buffer. TE can also be used as a gel or a tank buffer, but has lower buffering capacity (Maniatis *et al.*, 1982). At low concentration and ionic strength, migration is limited, while excessive ionic strength induces higher electrical conductivity, which is associated with the generation of heat. This heat can be enough to damage and/or denature the double stranded DNA and even cause melting of the gel.

Exposure to UV is known to induce lesions in DNA such as pyrimidine cyclobutane dimers and/or 5-thyminyl-5,6-dihydrothumine (non cyclic dimers) and breaking of sugar phosphate backbone, cross-links between DNA strand which might contribute to DNA damage (Brunk *et al.*, 1977; Grundemann *et al.*, 1996). In reducing such circumstances, working under UV light was minimized during gel observation and gel excision for both cloning and probe preparation proposes. It was reported that the exposure to UV light was not damage a High Molecular Weight (HMW) DNA prepared for BAC cloning (Simkova *et al.*, 2003). Thus, it was assumed that any damage caused by visualizing DNA on UV trans-illuminator was minor before subsequent of DNA manipulation.

High-throughput DNA marker automation and sequencing technologies for high quality genome analysis, with a basic principle related to PCR-based techniques, have achieved a breakthrough in plant science and can be utilised in *Musa*. However the use of a wide range of DNA marker assays is necessary for precise characterization of the genes of interest, and for a comprehensive understanding of diversity evaluation within and between *Musa* cultivars and wild material. These methods are important in determining functional genes, comparative analysis amongst related and/or distantly related species, and gene mapping either for agronomic traits i.e. dwarf, high yielding, strong panicle, early flowering or value added residual characters i.e. aromatic, fibres, pulp texture and skin colour.

7.2 Diversity in *Musa* as source of genetic improvement

Diversity and distribution of both wild and cultivated *Musa* is scattered over the Asia-Pacific region which covers 70-85% of the gene pools and includes the centre of origin of *Musa* (INIBAP, 2006a). Over 1000 cultivars used by smallholders and farmers have been estimated to be present by taxonomists throughout these regions, but classification is hampered with many synonyms and widespread germplasm movement. Genetic sources for *Musa* improvement rely on genetic collections from cultivars, wild type species, domestication of new cultivars and intercrossed families amongst genomes A, B, S and T.

Most selections of edible *Musa* by farmers are mainly for cultivation originated from *M. acuminata* and *M. balbisiana* which contributed to 'A' and 'B' genome respectively. These classifications are mainly based on system described by Simmonds and Shepherd (1955) and revised by Silayoi and Chomchalow (1987) but are not considering cultivars with the S and T genome constitution. Thus, diversity from these constitutions might give a great impact in *Musa* diversity as their hybrids AS and BT have already been identified in Papua New Guinea.

The Taxonomic Advisory Group for *Musa*: Expert discussion forum on taxonomic and conservation issues (http://tag.inibap.org/) was set up during 2006 to address the issues regarding identification, relationships, synonymy and naming. In understanding *Musa* genomic resources, *Musa* Germplasm Information System (MGIS) provides information access to accessions and supported by *Musa* Genome Resources Centre in IEB which supplies BAC libraries, markers for genetic mapping and high density colony filters (INIBAP, 2006b).

Research work in retroelements showed high copy numbers in *Pseudoviridae* (copia-like) which exist in clusters whereas *Metaviridae* (Gypsy-like) clearly showed multiple patterns in diploid *Musa* species from *M. acuminata*, *M. balbisiana*, *M. textilis*, *M. schizocarpa*, *M. ornata* and *M. velutina* amplified by a PCR-based technique.

The information without diversity measurements clearly shows that the diversity occurs even amongst the same genome constitution and it is potentially good in providing sources for genetic improvement. New diversity can also be induced. Probable examples are Mas and Rastali, giving rise to Cinta Manis and Mutiara respectively through somaclonal variation and tissue culture materials. As the diversity of *Musa* is important for genetic improvement, conservation of these valuable materials has been made intensively through both *in vivo* and *in vitro* approaches.

Both approaches are important for taxonomy: morphological characteristics and trait evaluations which are related to field planting are essential, and isolated areas are hard to collect, highly vulnerable to nature disaster especially diseases and climates, and disease spread is uncertain and can destroy material in the field.

In order to sustain the conservation, regular planting is required in a certain cycle, and re-allocation of the planting area must be carried out to refresh the soil and vigour. The in vitro approach is more convenient and reliable for maintaining germplasm as it is related to storage in the form of tissue culture materials but cannot give much agronomic information about accessions.

High capacity of samples storage can be catered for by tissue culture and kept in the same place ready for distribution of material with known disease status. Maintenance by sub-culturing and further monitoring is required to ensure genetic integrity and freedom from diseases and without somaclonal variation. Currently, slow growth techniques are used extensively and cryopreservation is being established to reduce laborious work in culturing and provide a low maintenance feature in handling the materials with the possibility of introducing somaclonal variation. Nevertheless, observational data from the ITC suggests induction of somaclonal variation is not a major problem during slow-growth maintenance of *Musa* germplasm (INIBAP, 2006b).

Diversity of *Musa* is the source of new varieties with better agronomic performance including high yield, resistance to both abiotic and biotic stress, and improved morphological and nutritional characters which leads to sustained production of this crop of benefit to smallholders, commercial farmers and consumers. Besides variation present in wild collections, diversity generated through somaclonal variations and mutation breeding also contributed and can be used as reliable sources towards diversity and improvement of *Musa*.

7.3 Molecular markers as technology for genetic improvement

The Zingiberales including *Musa* are a diverse order of eight families including species from ginger to ornamental plants. Attempts for domestication and classification of *Musa* species have been made long ago based on morphological characteristics (Simmonds and Shepherd 1955) and biochemical markers (Horry and Jay 1988; Lebot *et al.*, 1993) but as in many groups with extensive hybrids and vegetative propagation, the morphological approaches do not give enough information to form a comprehensive understanding of *Musa.* As there are four genome constitutions in *Musa* currently known i.e. A, B T and S, the characterization become more complicated and further are hampered with many synonyms for the same cultivar.

In breeding strategies, both ancestral diploids and hybrids are important sources of diversity and can be used to determine genetic and evolutionary drift. A comparative analysis of wild types and cultivars has lead to some models of how 2n (=3x) and x gametes can be formed in triploid cultivars.

The diploid wild type *M. acuminata burmanicoides* 'Calcutta-4' (2n=2x=22) which has been used as model in breeding works is highly heterozygous (Crouch *et al.*, 1998) which is not ideal to be used as a genotype tester or for some purposes as a source of parental plants. Thus, mechanisms in transferring the genes and/or motifs of the desirable traits including abiotic and biotic stress are still unclear, and breeding and sexual gene-transfer strategies are still under development in the *Musa* genus.

Mobile elements as discussed in chapter 3 which can amplify, move and jump around throughout the genomes are also a source of genetic variation and continuous mutation pressure on the genome. However, the time frame and condition of the movements are also not clear as the expression and/or the existence of the elements are also related to silencing and stress activated mechanisms.

Molecular markers are very important for genetic characterization in identify, allocating, probing and mapping genes both animal and plant species. Impacts of marker technology can be shown in the ability of accessing biodiversity, genetic relationships, understanding structural organisation and interactions amongst the species. Each marker method has their pros and cons in terms of capabilities. For example in AFLP which has ability in identifying large number of polymorphic bands but the patterns are restricted and must be treated as dominant markers. Besides, to perform this approach, high quality and large amounts of DNA are required as poor DNA may lead to incomplete digestion resulting in false polymorphisms.

The third chapter showed microsatellites are reliable and promising as rapid tools in identifying polymorphism amongst *Musa* but proper primer design is relatively cost demanding. Non-coding sequences as SSRs, i.e. mono- , tri-, tetra- and/or pentanucleotide, corresponding to particular locus are ubiquitous repeats in all eukaryotic genome as through evolution they might undergo mutated, slipped and/or impaired changes leading to either expansion or contraction. Use can be challenging as less information is known about their sequences and virtually no complementary data sets available.

The reduced costs of sequencing and better understanding of sequence variation, combined with automated SSR identification and primer design approaches are rapidly making high throughput SSR analysis easier. The results obtained need to be furnished with more primers and integrated to other techniques such as microarray, Southern hybridisation and/or *in situ* hybridization for more conclusive information to prove copy number and localisation.

It is important to use a range of techniques that are independent and can be integrated to achieve reliable and conclusive data in generating information, patterns and links amongst the species and genotypes of interest. However, in approaches to selection, cost and technical demands (skills and knowledge) should be fair enough and appropriate to apply in routine work in breeding programmes for screening and evaluation of samples.

In breeding, the molecular markers are extensively being used to understand agronomic traits, genetic drift and gene segregation to facilitate movement and/or accumulation of desirable genes amongst breeding lines leading to commercially fine selections and propagation of desirable plants.

PCR-based techniques are the most powerful tool and highly sensitive to facilitate identification at molecular level; these techniques have been chosen as a core in molecular marker assessments for germplasm characterisation, cultivar identification and genetic evolution including mutation and repeats motifs.

Large insert recombinant DNA libraries from the *Musa* genomics programme through BACs is already having an impact on understanding the diversity which leads to discovery of genes or DNA motifs of interest. Through EST (mRNA/cDNA sequences) discoveries of common and specific express genes such as abiotic and biotic resistance can be mined efficiently.

ESTs provide a comprehensive prospective of specific gene similarities which can be linked through bioinformatics tools. Thus by knowing their diversity, more collection and/or new cultivated *Musa* can be identified and compared which leads to more accessible intercrossed varieties. Construction of BAC libraries is a new era in understanding part of the genome size in plant genome organisation. The libraries are easy to handle, involving more markers that are expected to be derived, which will allow physical mapping studies on many species and understand gene and genome organisation in the hybrid *Musa*s.

Genome mapping consists of three approaches either by genetic mapping, cytogenetic mapping or physical mapping. Physical mapping is important in localization and isolation of genes to study genome organisation and evolution. DNA consists of functional elements which are related to coding and non-coding sequences. Retroelements which are coding or degenerate coding sequences related to *Pseudoviridae* and *Metaviridae* are relatively straightforward, mostly known and comparable to other species.

Although they are abundant and not coding for essential genes, they are important for maintaining chromosome structure, gene alteration and manipulation throughout evolution (Schmidt and Heslop-Harrison, 1998). Complementary data sets are available in ESTs although not yet from fully public databases. Challenges remain in measuring allelic diversity in different accessions which has to be properly documented.

7.4 *Musa* genomics in relation on stress resistance

Genes involved in stress resistance are varied and specific, or sometimes broadly responsive, to the various stress conditions which respond either directly and/or through several biochemical pathways. Regulatory sequences are a small part of the nucleotide component in genome organisation but have a great impact in function for plant survival. As they are inherited, the sequences are more or less conserved both at intra- and extra- species levels; however, the principles are the same which requires a recognition event for the stress, signalling and execution of the stress defence. Efficiency and effectiveness of the mechanisms remain unclear as different varieties or species differ, and many factors might be involved and integrated. For example signalling in biotic stress involves NBS which are related to intracellular activities and require another extracellular substituted domain to be anchored and activated during pathogen invasion. Changes in genome size are sometimes corresponding to mobile elements which are abundant but have similarity at their RT region and have a great impact on plant functional genes either through expansion, mutation and/or trans-activation. Expression of various abiotic and biotic stresses in tobacco which are related to retrotransposons *Tnt1* was also found in tomato and *Arabidopsis* (Mhiri *et al.*, 1997; 1999).

In *Musa*, genomic approaches are very important as there are no inbred lines of homozygosity available and highly heterozygosity amongst diploids. Double haploid in *Musa* is excessively being developed in *M. acuminata* cv. Pahang for characterization and understand its diversity and evolution (Assani *et al.*, 2003; http://www.generationcp.org/arm/ARM06/day_3/Roux_part_1.pdf).

However, the markers obtained have their limitations: various origins or sources, genetic drift and stability related to unclear genetic changes in *Musa*, and restricted understanding of the genes. Nevertheless, the opportunity is there to provide the breeders precision, with high efficiency in identifying phenotype of selected traits.

As a monocotyledon plant (APG2, 2003), *Musa* could represent a useful comparison to model plants like *Oryza* (monocot) and *Arabidopsis* (dicot). *Oryza,* in the *Poales*, is a sister group to the *Zingiberales* (both lying within the commelinids) while *Arabidopsis* is a core eudicot in the eurosids II. High throughput propagation and screening from both tissue culture and field evaluation with a breeding system for both sexual and vegetative in combination with parthenocarpic fruit production are ideally perfect to put *Musa* as one of model plant species for breeders.

Tissue culture practises provide the earliest stage in storage, screening and selection of the novel plants prior planting. As molecular assisted markers are increasingly developed and can rapidly be used to detect interesting genes, selection becoming a routine procedure in achieving breeding strategies.

Molecular markers will speed up the screening process of culturing and/or planting generation procedures. But issues might be raised in accelerating resistance genes related to biotic and abiotic stress from wild species of *Musa* and/or related species to be integrated in *Musa* for superior cultivars as the markers are limited to the numbers of samples can be analysed and numbers of lines can be improved in certain time frame period.

7.6 General conclusion

The diversity of abundant repetitive DNA in the *Musa* genome is related to transposable elements and SSRs, but there is little clear correlation with genome constitution. The transposable elements exist independently and can be amplified by degenerate primers. BAC sequences are useful in designing specific primers for SSR, but identification of specific BACs has to be supported with other interesting genes which might reflect relationships between coding (from genes of interest) and non-coding regions (SSR).

Conserved regions in the genes of interest, related to biotic stress at amino acids residues particularly NBS, have enabled a breakthrough in isolating R-gene domains which will lead to full length sequences of the resistance-related genes. Clusters generated by amplification of genomic DNA from *Musa* accessions provide information regarding specification of the R-genes signalling for defence mechanisms. The information gained can be expanded for gene expression. 'Wobbling' motifs of LRR exist in big families so it is difficult to design primers sets based on their conserved repeats: inconsistent motifs between the repeats are not conserved, although might be related to specificity and responses to stress environments.

EST sequences are reliable for tagging genes of interest and are useful in understanding diversity of the genes in *Musa.* Retroelements, SSR, NBS and LRR are not corresponding to genome constitutions of *Musa* species. Again, BACs are an important resource for gene discovery, and pooling strategies provide a remarkable application in finding specific BACs for plant gene studies.

7.7 Future work

The research presented here is related to diversity in coding and non coding sequences from transposable elements, repeats motifs and protein sequences which contribute to abiotic and biotic stress in *Musa*. Characterization of the retroelements i.e. *Copia-like* and *Gypsy-like* should go further for full length and details in understanding the mechanism and impact to diversity, evolution and domestication of the *Musa*, as will be found as more BAC sequences become available (www.musagenomics.org). Currently, large scale genome sequencing projects under The Global *Musa* Genomic Consortium are in progressing which brings together expertise from 20 countries. This global collaboration enables information and technologies to be shared and achieved more goals under a concise timeframe. Further, members of the consortium also actively participated in Generation Challenge Programme (GCP) which contributes and supports development of markers for *Musa* characterization as well as comparative study between *Musa* and *Oryza* (INIBAP, 2006b).

Microsatellites which are related to non coding sequences showed great value as a specific marker systems in *Musa* and requires more primers to get comprehensive and conclusive picture on how the repeats organised; again, more sequence information and comparative data between A, B and other genomes will be valuable for development of more microsatellite markers. Although microsatellites are abundant, evidence showed that non-coding regions might be significant in control of transcribed regions and involved in functional genes (Kashi *et al.*, 1997; Li *et al.*, 2002, 2004) whether through chromatin structure or nuclear organization.

Tolerance to environmental factors such water, temperature and salinity have involved common proteins throughout evolution. As discussed here, challenges from bacteria, fungi and viruses show responses involving proteins with a specific domain for signalling related to NBS-GLPL domain, and this is conserved and clustered amongst *Musa* genotypes. The signalling is supported by imperfect motifs which related to LRR for gene specification and triggering a defence mechanism.

The orientation, size, number, type and arrays of the LRR are not yet defined but their contribution in both abiotic and biotic stress are acknowledged, and again sequence information from larger genomic regions (e.g. from BACs or else genomic sequencing) will be now the most efficient approach to understanding their organization. Questions about LRR which remain to be answered, in *Musa* as well as other species, are related to the mechanism of manipulating the motif sequences specific to the stress, the LRR motif clustering, copy numbers and their involvement in triggering the signalling mechanism.

Primers designed for related abiotic and biotic stress responsive sequences are useful in partial gene isolation and move research towards complete gene and promoter sequences and gene expression studies. Sequences obtained are important in identification of valuable alleles which can be used as genetic markers for traits. Similar approaches also can be applied for post harvest study which is related to fruit ripening, sweetness, texture and shelf life. Thus, it is also crucial to understand and ascertain complementary applications between marker selections and conventional breeding to develop a dynamic conclusive strategy in *Musa* breeding programmes.

Designed primers, sequencing data and clones presented in this research for both abiotic and biotic stress are valuable tools in understanding genome organisation, evolution and biodiversity in *Musa* especially for gene discoveries. Availability of large inserts DNA through BAC libraries for screening, well documented EST in public databases, cDNA libraries, micro/macro arrays and DNA-chips would significantly achieve towards novel genes and domestication of new cultivars. New cultivars are expected to be rapidly screened, selected and introduced to public with ideal traits through breeding programs. Thus, future investigations are widely open based on the information gathered in this research work.

REFERENCES

References

Aarts M.G., B. te Lintel Hekkert, E.B. Holub, J.L. Beynon, Stiekema W.J. , Pareira A . 1998. Identification of R-gene homologous DNA fragments genetically linked to disease resistance loci in *Arabidopsis thaliana. Mol. Plant Microb Interact.* 11 (4):251-258.

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

Agarwal P.K. 1987. Cytogenetical investigations in Musaceae II. Meiotic studies in eight male sterile triploid banana varieties of India. *Cytologia.* 52:451-454.

Alix K., Ryder C.D., Moore J., King G.J., Heslop-Harriosn J.S. 2005. The genomic organisation of retrotransposons in *Brassica oleracea. Plant Mol. Biol.* 59:839-851.

Allard R. W. 1960. *Principles of Plant Breeding.* New York: John Wiley & Sons Inc.

Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W., Lipman D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.

Anderson P.A., Lawrence G.J., Morrish B.C., Ayliffe M.A., Finnegan E.J., Ellis J.G. 1997. Inactivation of the flax rust resistance gene *M* associated with loss of a repeated unit within the leucine-rich repeat coding region. *Plant Cell.* 9:614-651.

APG2. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical J. of the Linnean Society.* 141:399-436.

Arabidopsis-Genome-Initiative. 2000. The (2000) analysis of the genome sequence of the flowering plant *Arabidopsis thaliana. Nature.* 408:796-813.

Arnaud E., Horry J.P. 1997. Musalogue: a catalogue of *Musa* germplasm. Papua New Guinea collecting missions, 1988-1989. *International Network for the Improvement of Banana and Plantain, Montpellier, France.*

Asakawa S., Abe I., Kudoh Y., Kishi N., Wang Y., Kubota R.,Kudoh J.,Kawasaki K., Minoshima S., Shimizu N. 1997. Human BAC library: Construction and rapid screening. *Gene.* 191:69-79.

Assani A., Bakry F., Kerbellec F., Haïcour R., Wenzel G., Foroughi-Wehr B. 2003. Production of haploids from anther culture of banana, *Musa balbisiana* (BB). *Plant Cell Rep.* 21:511-516.

Audic S., Claverie J. 1997. The significance of digital gene expression profiles. *Genome Research* 7:986-995.

Bai J., Lourdes A. Pennill, Jianchang Ning, Se Weon Lee, Jegadeesan Ramalingam, Craig A. Webb, Bingyu Zhao, Qing Sun, James C. Nelson, Jan E. Leach, Scot H. Hulbert. 2002. Diversity in nucleotide binding site-leucine-rich repeat genes in cereals. *Genome Research.* 12 (12):1871-1884.

Bailey B. A., Dean F.D.J., Anderson A. D. 1990. An ethylene biosynthesis-inducing endoxylanase elicits electrolyte leakage and necrosis in *Nicotiana tabacum* cv Xanthi leaves. *Plant Physiol.* 94:1849-1854.

Baker B., Zambryski P., Staskawicz B., Dinesh-Kumar S.P. 1977. Signaling in plant-microb interactions. *Science.* 276:726-733.

Bakry F., Carreel F., CaruanaM. L., Cote F. X., Jenny C.,Tezenas du Montcel H. 2001. Banana. In:*Tropical Plant Breeding*, edited by H. S. Charrier M.J.A., Nicolas D. Enfield. NH, USA.: Science Pub.

Baltimore D. 1970. RNA-dependant DNA polymerase in virions of RNA tumour-viruses. *Nature.* 226:1209-1211.

Bartos 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): Taxanomi implication. *Cytogenetic Genome Research.* 109:50-57.

Bent A.F., Kunkel B.N., Dahlbeck D., Brown K.L., Schmidt R., Giraudat J., Leung J., Staskawicz B.J. 1994. *RPS2* of *Arabidopsis thaliana*: A leucine-rich repeat class of plant disease resistance genes. *Science.* 265:1856-1860.

Bergelson J., Kreitman M., Stahl E.A., Tian D. 2001. Evolutinary dynamics of plant R-genes. *Science.* 292:2281-2285.

Bevan M. 1998. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana. Nature.* 391:485-488.

Biffen R. H. 1905. Mandel's laws of inheritance and wheat breeding. *J. Agric. Sci.* 1:4-48.

Billotte N., Risterucci A.M., Barcelos E., Noyer J.L., Amblard P., Baurens F.C. 2001. Development, characterisation and across-taxa untility of oil palm (*Elaeis guineensis* Jacq.) microsatellite markers. *Genome.* 44:413-425.

Bingham P. M., Zachar Z. 1989. Retrotransposons and the FBtransposon from *Drosophila malanogaster*. In Mobile DNA. *American Society for Microbiology*:485-502.

Birmeta G., Nybom H., Bekele E. 2004. Distinction between wild and cultivated enset (*Ensete ventricosum*) gene pools in Ethiopia using RAPD markers. *Hereditas.* 140:139-148.

Bohnert H.J., Nelson, D.E., Jensen, R.G. 1995. Adaptations to environmental stresses. *Plant Cell.* 7:1099–1111.

Bonaldo M., Lennon G., Soares M. 1996. Normalization and substraction: Two approaches to facilitate gene discovery. *Genome Res.* 6:791-806.

Booker F. L. 2004. Influence of ozone on ribonuclease activity in wheat (*Triticum aestivum*) leaves. *Physiologia Plantarum.* 120:249-255.

Botella M. A., Coleman M. J., Hughes D. E., Nishimura M. T., J. D. G., Sommerville S. C. 1997. Map positions of 47 *Arabidopsis* sequence with sequence similarity to disease resistance genes. *Plant J.* 12:1197-1211.

Botstein D., White R., Skolnick M., Davis R.W. 1980. Construction of genetic linkage map using Restriction Fragment Length Polymorphism. *American J. of Human Genetics.* 32:314-331.

Bourne H. R., Sanders D. A., McCormick F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature.* 349:117-127.

Brandes A., Heslop-Harrison, J. S., Kamm, A., Kubis, S., Doudrick, R. L. and Schmidt T. 1997. Comparative analysis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. *Plant Mol. Biol.* 33:11-21.

Brunk C. F., Simpson L. 1977. Comparison of various ultraviolet sources for fluorescent detection of ethidium bromide-DNA complexes in poly acrylamide gels. *Anal. Biochem.* 82:455-462.

Bruno W.J., Knil E., Balding D.J., Bruce D.C., Doggett N.A., Sawhill W.W., Stallings R.L., Whittaker C.C., Torney D.C. 1995. Efficient pooling designs for library screening. *Genomics.* 26:21-30.

Buchanan S.G., Gay N.J. 1996. Structure and functional diversity in the leucine-rich repeat family of protein. *Prog. Biophys. Mol. Biol.* 65:1-44.

Buddenhagen I. W. 1987. Disease susceptibility and genetics in relation to breeding of bananas and plantains. In *Bananas and Plantain Breeding Strategies*, edited by Persley G. J. De Langhe E. A. ACIAR. Canberra, Australia.

Bureau T. E., Wessler S. R. 1992. *Tourist*: a large family of inverted-repeat element frequently associated with maize genes. *Plant Cell.* 4:1283-1294.

Burkhard P., Stetefel J., Strelkov S.V. 2001. Coiled coils: a highly versatile protein folding motif. *TRENDs in cell Biology.* 11 (2):82-88.

Buschges R., Hollricher K., Panstruga R. 1977. The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell.* 88:695-705.

Cai W., Reneker J., Chow C., Vaishnav M., Bradley A. 1998. An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. *Genomics.* 54:387-397.

Cannon S.B., Zhu H., Baumgarten A.M., Spangler R., May G.,Cook D.R., Young N.D. 2002. Diversity, distribution and ancient taxonomic relationship within the TR and non-TIR NBS-LRR resistance gene subfamilies. *J. Mol. Evol.* 54:548-562.

Capy P., Anxolabehere D., Langin T. 1994. The strage phylogenies of tranposable elements: are horizontal transfer the only explanation? *Trends Genet.* 10:7-12.

Carreel F. 1994. Evaluation de la diversite genetique chez les bananiers diploides (*Musa* sp.). *Genet. Sel. Evol.* 26 (Suppl 1):125-136.

Chang C.L., Marra G., Chauhan P., Ha H. T., Chang D. K., Ricciardiello L., Randolph A., Carethers J M., Boland C. R. 2002. Oxidative stress inactivates the human DNA mismatch repair system. *Am. J. Physiol. Cell Physiol.* 283:C148-C154.

Charlesworth B. 1986. Genetic divergence between transposable elements. *Genet. Res. Camb.* 48:111-118.

Clare J., Farabaugh P. 1985. Nucleotide sequence of yeast Ty element: evidence from an unusual mechanism of gene expression. *Proceedings of the National Academy of Sciences of USA.* 82:2829-2833.

Costa A. P. P., Scortecci K. C., Hashimoto R.Y., Araujo P.G., Granbastien M. A., Van Sluys M.A. 1999. Retrolyc1-1, a member of the Tnt1 retrotranposon super- family in the *Lycopersicon peruvianum* genome. *Genetica.* 107:65-72.

Crick F., C.H. 1953. The packing of alpha helices : simple coiled coil. *Acta Crytallographica.* 6:689-697.

Crouch H.K., Crouch J.H., Jarret R.L., Cregan P.B., Ortiz R. 1998. Segregation of microsatellite loci in haploid and diploid gemetes of *Musa*. *Crop Science.* 38:211-217.

Crouch J.H., Ortiz R. 1996. Availability of IITA's improved *Musa* Germplasm. *MusAfrica.* 9:1-2.

Dangl J., L. Jones D.G. 2001. Plant pathogens and integrated defense responses to infection. *Nature.* 411:826-833.

Daniels S.B., Peterson K.R., Straubaugh L.D., Kidwell M.G., Chovnic A. 1990. Evidence for horizontal transmission of the P transposable elements between *Drosophila* species. *Genetics.* 124:339-355.

De Barros E. G., Tingey S., Rafalski J.A. 2000. Sequence charaterization of hypervariable regions in the soybean genome: Leucine-rich repeats and simple sequence repeats. *Genetics and Molecular Biology.* 23 (2):411-415.

De Wit P. J. G. M., Brandwagt B. F., Van de Burg H. A., Cai X., Van der Hoorn R. A. L., De Jong C. F., Van 't Klooster, De Kock M.J. D., Kruijt M., Lindhout W. H., Luderer R., Takken F.L.W., Westerink N., Vervoort J.J.M., Joosten M. H. A.J. 2002. The molecular basis of co-evolution between *Cladosporium fulvum* and tomato. *Antonie van Leeuwenhoek.* 81:409-412.

Devos K.M., Gale M. D. 2000. Genome relationships: The grass model in current research. *Plant Cell.* 12:637-346.

D'hont A., Paget-Goy A., Escoute J., Carreel F. 2000. The interspecific genome structure of cultivated bananas, *Musa* spp. revealed by genomic DNA in situ hybridisation. *Theor. Appl. Genet.* 100:177-183.

Diaz-Perez S.V., Crouch V.W., Orbach M.J. 1996. Construction and characterization of *Magnaporthe grisea* bacterial artificial chromosome library. *Fungal Genet. Biol.* 20:280-288.

Dickinson M.J., Jones D.A. Jones J. D. 1993. Close linkage between the *Cf-2/Cf-5* and *Mi* resistance loci in tomato. *Mol. Plant-microbe Interact.* 6:341-347.

Dievart A., Clark S. E. 2004. LRR-containing receptors regulating plant development and defense. *Development.* 131:251-261.

Dixon M.S., Jones D.A., Keddie J.S. Thomas C.M., Harrison K., Jones J.D.G. 1996. The tomato *Cf-2* disease resistance locus comprise two functional genes encoding leucine-rich repeat protein. *Cell.* 84:451-459.

Dixon M.S., Hatzixanthis K., Jones D.A., Harrison K., Jones J.D.G. 1998. The tomato *Cf-5* disease resistance gene and six homologs show pronounced allelic variation in leucine-rich repeat copy number. *The Plant Cell.* 10:1915-1925.

Dodds K. S., Simmonds N. W. 1948. Sterility and parthenocarpy in diploid hybrids of *Musa. Heredity.* 2:101-117.

Dolezel J., Dolezelova M., Novak F. J. 1994. Nuclear DNA amount in diploid bananas (*M. acuminata* and *M. balbisiana*). *Biologia Plantarum.* 36:351-357.

Dolezel J. 1998. Flow cytometry, its application and potential for plant breeding. In *Current Topics in Plant Cytogenetics Related to Plant Improvement.*, edited by T. Lelley: Universitätsverlag, Vienna, Pp. 80-90.

Dolezelova M., Valarik M., Swennen R., Horry J.P., Dolezel J. 1998. Physical mapping of the 18S-25S and 5S ribosomal RNA genes in diploid bananas. *Biologia Plantarum.* 41 (4):497-505.

Efron B., R. Tibshirani. 1993. *An introduction to the bootstrap*: Chapman & Hall, London.

Efron B., Halloran E., Holmes S. 1996. Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA.* 93:7085-7090.

Eickbush T.H. 1994. Origin and evolutionary relationships of retroelements. In *The Evolutionary Biology of Viruses.*, edited by S. S. M. (Ed.): Raven Press, New York.

Ellis J., Lawrence G., Finnegan E., Anderson P. 1995. Contrasting complexity of two rust resistance loci in flax. *Proc. Natl. Acad. Sci. USA.* 92:4185-4188.

Enkhbayar P., Kmiya M., Osaki M., Matsumoti T., Matsushima N. 2003. Structural Principles of Leucine-Rich Repeat (LRR). *Protein: Structure, Function and Bioinformatics* 54:394-403.

Ewing B., Green P. 2000. Analysis of expressed sequence tags indicates 35,000 human genes. *Nat. Genet.* 25:232-234.

Fang X., Gu S., Xu Z., Chen F. 2004. Construction of a binary BAC library for an apomictic monosomic addition line of *Beta corollifora* in sugar beet and identification of the clones derived from the alien chromosome. *Theor. Appl. Genet.* 108:1420-1425.

FAOSTAT. 2004. http://apps.fao.org/default.jsp.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 39:783–791.

Feng Q., Zhang, Y. J., Hao, P., Wang, S. Y., Fu, G., Huang, Y. C., Li, Y., Zhu, J. J.,Liu, Y. L., Hu, X. 2002. Sequence and analysis of rice chromosome 4. *Nature.* 420:316-320.

Feschotte C., Wessler S.R. 2001. Treasures in the attic. Rolling circle transposons discovered in eukaryotic genomes. *PNAS.* 98 (16):8923-8924.

Flavell A.J., Dunbar E., Anderson R. Pearce S.R., Hartley R., Kumar A. 1992. *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. *Nucleic Acids Res.* 20:3639-3644.

Flor H.H. 1956. The complementary genetic system in flax and flax rust. *Adv. Genet.* 8:29-54.

Frengen E., Weichenhan D., Zhao B.H., Osoegawa K., van Geel M., de Jong,. 1999. A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics.* 58:250-253.

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.

Fu H., Dooner, H.K. 2000. A gene-enriched BAC library for cloning large allele-specific fragments from maize: isolation of a 240-kb contig of the bronze region. *Genome Res.* 10:866-873.

Gao L., McCarthy E. M., Ganko E. W., McDonald J.F. 2004. Evolutionary history of *Oryza sativa* LTR retrotransposons : a preliminary survey of the rice genome sequences. *BMC Genomics.* 5 (18):1-18.

Garcia-Mas J., Oliver M, .G., Gomez-Paniagua H., De Vicente M.C. 2001. Comparing AFLP, RAPD, RFLP markers for measuring genetic diversity in melon. *Theor. Appl. Genet.* 101:860-864.

Gardiner J., Schroeder S., Polacco M. L., Villeda-Sanchez H., Fang Z., Morgante M., Landewe T., Fengler K., Useche F., Hanafey M., Tngey S., Chou H., Wing R., Soderlund C., Coe Jr. E. H. 2004. Anchoring 9,371 maize expressed sequence tagged unigenes to the Bacterial Artificial Chromosome contig map by two-dimensional overgo hybridization. *American Society of Plant Biologist.* 134:1317-1326.

Gawel N. J., Jarret L. 1991. A modified CTAB DNA extraction procedure for *Musa* and *Ipomea. Plant Mol. Biol. Rep.* 9:262-266.

Gawel N. J., Jarret L., Whittemore A.P. 1992. Restriction fragment length polymorphism (RFLP)-based phylogenetic analysis of *Musa. Theor. Appl. Genet.* 84:286-290.

Geering A. D. W., Olszewski N. E., Glyn H., Lockhart B.E.K., Hull R., Thomas J. E. 2005. Banana contains a diversity array of endogenous badnaviruses. *J. of General Virology.* 86:511-520.

Geffroy V., Sicard D., de Oliveira J.C., Sevignac M., Cohen S., Gepts P., Neema C., Langin T., Dron M. 1999. Identification of an ancestral resistance gene cluster

involved in the co-evolution process between *Phaseolus vulgaris* and its fungal pathogen *Colletotrichum lindemuthianum. Mol. Plant-Microbe Interact.* 12:774-784.

Gerlach W. L., Dyer T. A. 1980. Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. *Nucleic Acids Res.* 8 (21):4851-4865.

Grandbastien M. A. 1992. Retroelements in higher plants. *Trends Genet.* 8:103-108.

———. 1998. Activation of plant retrotransposons under stress. *Trends in Plant Science.* 3:181-187.

Grant M.R., Godiard L., Straube E., Ashfield T., Lewald J., Sattler A., Innes R.W. Dangl J.L. 1995. Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science.* 269:843-846.

Grapin A., Noyer J.L., Carreel F., Dambier D., Baurens F.C., Lanaud C. Lagoda P.J.L. 1998. Diploid *Musa acuminata* genetic diversity assayed with sequence tagged microsatellite sites. *Electrophoresis.* 19:1374-1380.

Green M. M. 1988. Mobile DNA elements and spontaneous gene mutation. In *Eukaryotic Tranposable Elements as Mutagenic Agents.*, edited by M. D. J. F. Lambert M.E., Weinstein I.B. (eds).

Green P.J. 1994. The ribonuclease of higher plants. *Annals Rev. Plant Physiol. Plant Mol. Biol.* 45:421-445.

Gribbon B.M., Pearce S.R., Kalendar R., Schulman A.H., Paulin L., Jack P., Kumar A., Flavell A.J. 1999. Phylogeny and tranpositional activity of *Ty-1-copia* group retrotransposons in cereal genomes. *Mol. Gen. Genet.* 261:883-891.

Grundemann D., Schomig. 1996. Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. *BioTechniques.* 21:898-903.

Gupta P. K., Roy J. K., Prasad M. 2001. Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *Current Science.* 80 (4).

Gupta P.K., Varshney R.K., Sharma P.C., Rameshi B. 1999. Molecular markers and their applications in wheat breeding. *Plant Breed.* 118:369-390.

Gupta S., Zink D., Kom B., Vingron M., Haas S. 2004. Strengths and weaknesses of EST-based prediction of tissue-specific alternative splicing. *BMC Genomics.* 5 (72):1-8.

Hong C.P., Lee S.J., Park .Y., Plaha P., Park Y.S., Lee Y.K., Choi J.E., Kim K.Y., Lee J.H., Lee J., Choi S. R., Lim Y.P. 2004. Construction of a BAC library of Korean ginseng and initial analysis of BAC-end sequences. *Mol. Gen. Genomics.* 271 (709-716).

Horry J. P., Arnaud E. 1997. Musalogue: a catalogue of *Musa* germplasm. Papua New Guinea collecting missions, 1988-1989. *International Network for the Improvement of Banana and Plantain, Montpellier, France.*

Horry J.P., Jay M. 1988. Distribution of anthocyanin in wild and cultivated banana varieties. *Phytochem.* 27:2667-2672.

Howell E.C., Newbury . J., Swennen R.L., Withers L. A., Ford-Lloyd B. V. 1994. The use of RAPD for identifying and classifying *Musa* germplasm. *Genome.* 37:328-332.

Hulbert S. H., Webb C.A., Smith A. M., Sun Q. 2001. Resistance gene complexes: evolution and utilization. *Annual Rev. Phytopathol.* 39:285-312.

INIBAP. 2001. Strategy for the global *Musa* Consortium. Report on a meeting held in Arlington, USA.

———. 2006a. Global Conservation Strategy for *Musa* (Banana and Plantain). A consultative document prepared by INIBAP with the collaboration of numerous partners in the *Musa* research-and-development community: International Network for the Improvement of Banana and Plantain, Montpellier, France.

———. 2006b. INIBAP Annual Report 2005. Celebrating 20 years of networking banana and plantain: International Network for the Improvement of Banana and Plantain, Montpellier, France.

Inouye M., Inouye S. 1991. Retroelements in bacteria. *Trends Biochem. Sci.* 16:18-21.

IPGRI-INIBAP/CIRAD. 1996. *Descriptors for Banana (Musa spp.).* International Plant Genetic Resources Institute, Rome, Italy/International Network for the Improvement of Banana and Plantain, Montpellier, France/Centre de cooperation internationale en recherche agronomique pour le developpement, Montpellier, France.

Jackson R. W., Athanassopoulos E., Tsiamis G., Mansfield J.W. Sesma A., Arnold D. L., Gibbon M. J., Murillo J., Taylor J.D. Vivian A. 1999. Identification of a parthenogenecity island, which contains genes for virulence and avirulence on a large native plasmid in the bean pathogen *Pseudomonas syringae* pathovar *phaseolicola. Proc Natl. Acad. Science. USA.* 96:10875-10880.

Jamil M. A. 1998. Banana Cultivation in Malaysia - Country Status. *Proceedings 1st. National Banana Seminar, Genting Highland, Malaysia. pp 12.*

Jenny C., Carreel F., Bakry F. 1997. Revision on banana taxonomy: "Klue Tiparot" (*Musa* sp.) reclassified as a triploid. *Fruits.* 52 (2):89-91.

Jones D.A., Thomas C.M., Hammond-Kosack K.E. Balint Kurti .J., Jones J.D.G. 1994. Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by tranposon tagging. *Science.* 266:789-793.

Jones D.A., Jones J.D.G. 1997. The role of leucine-rich repeats in plant defenses. *Advances in Botanical Research.* 24:91-167.

Jones J. D. G. 1997. Plant disease resistance- a kinase with keen eyes. *Nature.* 385:397-398.

Jones J. D. G., Dangl J. L. 2006. The plant immune system. *Nature.* 444 (7117):323-329.

Joosten M. H. A. J., De Wit P. J. G. M. 1999. The Tomato *Cladosporium fulvum* interaction: a versatile experimental system to study plant-pathogen interactions. *Annu. Rev. Phytopathol.* 37:335-367.

Jorgensen J., H. 1992. Multigene families of powdery mildew resistance gene in locus *Mla* on barley chromosome 5. *Plant Breed.* 108:53-59.

Jurka J., Kapitonov V.V., Smit A. F.A. 2005. Repetitive elements: Detection. In *Encyclopedia of Life Sciences*: John Wiley & Sons, Ltd.

Kaemmer D., Peraza-Echeverrý a L, Canche´ B, Arroyo A, Ortiz E, Zhang H. B, James A. 2002. Characterization and utilization of the first banana BAC library of *Musa acuminata* spp. burmannicoides type 'Calcutta IV'. In*: Program and abstracts of the 3rd. International Symposium on molecular and cellular biology of bananas.* Leuven: INIBAP.

Kajava A.V. 1998. Structural diversity of leucine-rich repeat protein. *J. Mol. Biol.* 277:519-527.

Kajava A.V., Kobe B. 2002. Assessment of the ability to model proteins with leucine-rich repeats in light of the latest structural information. *Protein Science.* 11:1082-1090.

Kalendar R., Tanskanen, J., Immonen, S., Nevo, E. and Schulman, A. H. 2000. Genome evolution of wild barley (*Hordeum spontaneum)* by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence., Proceedings of the National Academy of Sciences., USA.

Kamarudin S, Siti Hawa J. 1996. Development of information resources on banana and plantain in Malaysia - The MARDI perspective. *Proceedings of a Consultation/Workshop on Reginal Information System for Banana and Plantain -- Asia and the Pacific Los Banos, Laguna, Philippines, 1-3 April 1996. Pp 23-27.*

Kamate K., Brown S., Durand P., Bureau J.M., De Nay D., Trinh T. H. 2001. Nuclear DNA content and base composition in 28 taxa of *Musa*. *Genome.* 44:622-627.

Kanazin V., Marek L.F Shoemaker R.C. 1996. Resistance genes analogs are conserved and clustered in soybean. *Proc. Natl. Acad. Science. USA.* 93 (21):11746-11750.

Kantety R. V., La Rota M. V., Matthews D. E., Sorrells M. E. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol. Biol.* 48:501-510.

Kapitonov V.V., Jurka J. 2001. Rolling-circle transposons in eukaryotes. *Proc. Natl. Acad. Science. USA.* 98:8714-8719.

Karamura D.A. 1998. Numerical Taxonomic Studies of the East African Highland Bananas (Musa AAA-East Africa) in Uganda. *The University of Reading.*

Kashi Y., King D., Soller M. 1997. Simple sequence repeats as a source of quantitative genetic variation. *Trends Genet.* 13:74-78.

Kaufmann H., Mattiesch L., Lorz H., T. Debener. 2003. Construction of BAC library of *Rosa rugosa* Thunb. and assembly of a contig spanning *Rdr1*, a gene that confers resistance to blackspot. *Mol. Gen. Genomics.* 268:666-674.

Keen N.T. 1990. Gene-for-gene complementary in plant-pathogen interactions. *Annual Rev. Genet.* 24:447-163.

Kidwell M.G. 2002. Transposable elements and evolution of genome size in eukaryotes. *Genetica.* 115:49-63.

Kim U.J., Birren B.W., Slepak T., Mancino V., Boysen C., Kang H.L., Simon M.I., Shizuya H. 1996. Construction and characterization of human bacteria arficial chromosome library. *Genomics.* 34:213-218.

Kimura Y., Tosa Y., Shimasa S., Sogo R., Kusaba M., Sunaga T., Betsuyaku S. Eto Y., Nakayashiki H., Mayama S. 2001. *OARE-1,* a *Ty-1 copia* retrotransposons in Oat activated by abiotic and biotic stresses. *Plant Cell Physiol.* 42 (12):1345-1354.

Klein P. E., Klein R.R., Cartinhour S.W., Ulanch P. E., Dong J., Obert J. A., Morishige D. T., Schlueter S.D., Childess K.L., Ale M., Mullet J. E. 2000. A high throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. *Genome Res.* 10:789-807.

Knight H., Knight M. R. 2001. Abiotic stress signaling pathways specificity and cross-talk. *Trends Plant Sci.* 6:262-267.

Kobe B., Deisenhofer J. 1993. Cristal structure of porcine ribonuclease inhibitor, a protein with leucine-rich repeats. *Nature.* 366:751-756.

Kobe B., Kajava A. V. 2000. When protein folding is simplified to protein coiling: the continuum of solenoid protein structures. *Trends Biochem. Science.* 25:509-515.

———. 2001. The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* 11:725-732.

Kosman E., Leonard J. 2005. Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid and polyploid species. *Molecular Ecology.* 14:415-424.

Krauss U., Soberanis W. Jarra J. 2001. Evaluation of FHIA hybrids in comparison with local *Musa* clones in a Black Sigatoka-free area of Eastern Peru. *INFOMUSA.* 10 (1):21-26.

Kubis S.E., Heslop -Harrison J.S., Desel C., Schmidt T. 1998. The genome organisation of non-LTR retrotransposons (LINES) from three Beta species and five other angiosperms. *Plant Mol. Biol.* 36:821-831.

Kubis S.E., Castilho A.M.M.F., Vershinin A. V., Heslop -Harrison J.S. 2003. Retroelements, transposons and methylation stautus in the genome of oil palm (*Elaeis guineensis*) and the relationship to somaclonal variation. *Plant Mol. Biol.* 52:69-79.

Kuipers A. G. J., Heslop-Harrison J. S. P., Jacobsen E. 1998. Characterization and physical localization of Ty-1-*copia*-like retrotransposons in four *Alstroemeria* species. *Genome.* 41:357-367.

Kumar, A. and J. L. Bennetzen. 1999. Plant retrotransposons. *Annals Rev. Genet.* 33:479-532.

Laili D.A., Decroocq V., Blenda A.V., Schurdi-Levraud V., Garay L., Le Gall O., Damsteegt V., Reighard G.L., Abbott A.G. 2005. Identification and mapping of resistance gene analogs (RGAs) in *Prunus:* a resistance map for *Prunus. Theor. Appl. Genet.* 111 (8):1504-1513.

Lanaud C., Du-Montcel H.T., Jolivot M.P., Glszmann J. C., De Leon D.G. 1992. Variation of ribosomal gene spacer length among wild and cultivated banana. *Heredity.* 68:147-156.

Lanaud C., Ristericci A.M., Pieretti I., N'Goran J.A. K., Fargeas D. 2004. Characterization and genetic mapping of resistance and defence gene analogs in cocoa (*Theobroma cacao L.*). *Molecular Breeding.* 13:211-227.

Landschhulz W., H. M., Johnson P.F and McKnight S.L. 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science.* 240 (4860):1759-1764.

Lawrence G.J., Finnegan E.J., Ayliffe M.A., Ellis J.G. 1995. The *L6* genes for the flax rust resistance is related to the *Arabidopsis* bacterial resistance genes *RPS2* and the tobacco viral resistance gene *N. Plant Cell.* 7:1195-1206.

Le Q.H., Wright S., Yu Z., Bureau T. 2000. Transposon diversity in *Arabidopsis thaliana. PNAS.* 97:7376-7381.

Lebot V., Aradhya K.M., Manchardt R., Meilleur B. 1993. Genetic relationships among cultivated bananas and plantains from Asia and the Pacific. *Euphytica.* 67:163-175.

LeBrasseur N.D., MacIntosh G.C., Parez-Amador M.A. Saitoh M., Green P.J. 2002. Local and systemic wound-induction of RNase and nuclease activities in *Arabidopsis.* RNS1 as marker for a JA-independent systemic signaling pathways. *Plant J.* 29:393-403.

Lee Y., Tsai J., Sunkara S., Karamycheva S., Perta G., Sultana R., Antonescu V., Chan A., Cheung F., Quakenbush J. 2005. The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res.* 33:D71-D74.

Leeuwen H.V., Monfort A., Zhang H.B., Puigdomenech P. 2003. Identification and characteristics of a melon genomic region containing a resistance gene cluster from a constructed BAC library. Microcolinearity between *Cucumis melo* and *Arabidopsis thaliana. Plant Molecular Biology.* 51:703-718.

Leister D., Ballvora A., Salamini F.,Gebhardt C. 1996. A PCR-based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. *Nat. Genet.* 14:421-429.

Leland A.P., Raines R.T. 2001. Cancer chemotheraphy-ribonuclease to the rescue. *Chemistry and Biology.* 8:405-413.

Lers A., Khalchitski A., Lomaniec E., Burd S., Green P.J. 1998. Senescence-induced RNases in tomato. *Plant Mol. Biol.* 36:439-449.

Lewin, B. 2000. Genes. *Oxford University Press.*

Li Y. C., Korol A. B., Fahima T., Beiles A., Nevo E. 2002. Microsatellites: genomic distribution, putative functions and mutational mechanism: a review. *Mol. Ecol.* 11:2453-2465.

Li Y. C., Korol A. B., Fahima T., Nevo E. 2004. Microsatellites within genes: Structure, function and evolution. *Mol. Biol. Evol.* 21 (6):991-1007.

Lijavetzky D., Muzzi, G., Wicker T., Keller B., Wing, R., Dubcovsky J. 1999. Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. *Genome.* 42:1176-1182.

Linnaeus, Karl. 1753. The origin of modern botanical nomenclature. *Species Plantarum.*

Loh Y.T., Martin G.B. 1995. The disease resistance gene *Pto* and the fenthion-sensitivity gene *Fen* encode closely related functional protein kinases. *Prot. Natl. Acad. Sci. USA.* 92:4181-4184.

Lupas A. 1996. Coiled-coils: new structure and new function. *Trends Biochem Science.* 21:375-382.

Lysak M. A., Dolezelova M., Horry J. P., Swennen R., J. Dolezel. 1999. Flow cytometric analysis of nuclear DNA content in *Musa*. *Theor. Appl. Genet.* 98:1344-1355.

Mahajan S., Tuteja N. 2005. Cold, salinity and drought stress: An overview. *Archives of Biochemistry and Biophysics.* 444:139-158.

Malvar T., Biron R. W., Kaback D.B., Denis C.L. 1992. The CCR4 protein from *Saccharomyces cerevisiae* contains a leucine-rich repeat region which is required for its control of ADH2 gene expression. *Genetics.* 132:951-962.

Maniatis T., Fritsch E. F., Sambrook J. 1982. *Molecular cloning . A laboratory manual*: Cold Spring Harbor Laboratory Press New York.

Marlov R.L., Parkhurst S.M., Corces V.G. 1986. The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral protein. *Molecular and Cellular Biology.* 6:1129-1134.

Martin G. B., Bogdanove A. J., Sessa G. 2003. Understanding the functions of plant disease resistance proteins. *Annual Rev. Plant Biol.* 54:23-61.

Martin G.B., Brommonshenkel S.H., Chungwongse J., Frary A., Ganal M.W. 1993. Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science.* 262:1432-1436.

Mason S. H., Warzecha H., Mor T., Arntzen C.J. 2002. Edible plant vaccines: applications for prophylactic and therapeutic molecular medicine. *TRENDs in cell Biology.* 8 (7):324-329.

Matsunaga S., Yagisawa F., Yamamoto M., Uchida W., Nakao S., Kawano S. 2002. LTR retrotranposons in the dioecious plant *Silene latifolia*. *Genome.* 45:745-751.

Matsushima N., Ohyanagi T., Tanaka T., Kretsinger R.H. 2000. Supermotifs and evolution of tandem leucine-rich repeats within the small proteoglycans-biglycan, decorin, lumican, fibromodulin, PRELP, keratocan, osteoadherin, epiphycan and osteoglycin. *Protein.* 38:210-225.

McDowel J. M., Woffenden B. J. 2003. Plant disease resistance genes: recent insights and potential applications. *TRENDS Biotechnol.* 21:178-182.

Meyers B. C. , Allan W. Dickerman, Richard W. Michelmore, Subramoniam Sivaramakrishnan, Bruno W. Sobral, Nevin D. Young. 1999. Plant disease resistance genes encode members of ancient and diverse protein family within the nucleotide-binding superfamily. *The Plant J.* 20 (3):317-332.

Meyers B. C., Alexander Kozik, Alyssa Griego, Hanhui Kuang, Richard W. Michelmore. 2003. Genome-wide analysis of NBS-LRR-Encoding genes in *Arabidopsis*. *The Plant Cell.* 15:809-834.

Mhiri C., Morel J.B., Vernhettes S., Casacuberta J.M., Lucas H., Grandbastien M.A. 1997. The promoter of the tobacco *Tnt1* retrotransposon is induced by wounding and by abiotic stress. *Plant Mol. Biol.* 33:257-266.

Mhiri C., de Wit P.J.G.M., Grandbastien M.A. 1999. Activation of the promoter of the *Tnt1* retrotransposon in tomato after inoculation with the fungal pathogen *Clodosporium fulvum*. *Mol. Plant-microbe Interact.* 12:592-603.

Michelmore R.W., Meyers B. C. 1998. Clusters of resistance genes in plants evolve by divergent selection and birth-and-death process. *Genome Res.* 8:1113-1130.

Michelmore R.W. 2000. Genomic approaches to plant disease resistance. *Curr. Opin. in Plant Biology* 3:125-131.

Michener C., Sokai R. (1957). 1957. A quantitative approach to a problem in classification. *Evolution.* 11:130-162.

Mily R., Avni A. 2004. The receptor for the fungal elicitor ethylene-inducing xylanase is a member of a resistance-like gene family in tomato. *American Soc. of Plant Biologist.* 16:1604-1615.

Mindrinos M., Katagiri F., Yu G. L., Ausubel F. M. 1994. The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell.* 78:1089-1099.

Moran J.V., Morrish J.V. 2005. Chromosomes: Noncoding DNA (Including satellite DNA). In *Encyclopedia of Life Science*: John Wile & Sons, LTD.

Moullet O., Zhang H.B., Lagudah E.S. 1999. Construction and characterisation of a large DNA insert library from the D genome of wheat. *Theor. Appl. Genet.* 99:305-313.

Moumeni A., Samadi Y. B., Wu J., Leung H. 2003. Genetic diversity and relatedness of selected Iranian rice cultivars and disease resistance donors assayed by simple sequence repeats and candidate defense gene markers. *Euphytica.* 131:275–284.

Mount S.M., Rubin G.M. 1985. Complete nucleotide sequence of the *Drosophila* transposable element *copia*: Homology between *copia* and retroviral protein. *Molecular and Cellular Biology.* 5:1630-1638.

Mueller L. A., Solow H. T., Taylor N., Skwarecki B., Buels R., Binns J., Lin C., Wright H. M., Ahrens R., Wang Y., V.Herbst E., KeyderR. E., Menda N., Zamir D. S. 2005. The SOL genomic DNA network. A comparative resource for S*olanaceae* biology and beyond. *Plant Physiol.* 138:1310-1317.

Noel L., Moores T.L., van Der Biezen E.A., Parniske M., Daniels M.J., Parker J.E., Jones J.D. 1999. Pronounced intra specific haplotype divergence at RPP5 complex disease resistance locus of Arabidopsis. *Plant Cell.* 11:2099-2112.

Noir S., Combes M.C., Anthony F., Lashermes P. 2001. Origin, diversity and evolution of NBS-type disease-resistance gene homologous in coffee trees (*Coffea L.*). *Mol. Gen. Genet. Genomics.* 265:654-662.

Noir S., · Patheyron S., Combes M. C., Lashermes P., Chalhoub B. 2004. Construction and characterisation of a BAC library for genome analysis of the allotetraploid coffee species (*Coffea arabica* L.). *Theor. Appl. Genet.* 109:225–230.

Nowakunda N., P.R. Rubaihayo, M.A. Ameny, W. Tushemereirwe. 2000. Consumer acceptability of introduced bananas in Uganda. *INFOMUSA.* 9 (2):22-25.

Nybom H. 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology.* 13:1143-1155.

Ohyanagi T., Matsuhima N. 1997. Classification of tendem leucine-rich repeat within a great variety of proteins. *11*:A949.

Ortiz R., Vuylsteke D. 1992. Inheritance of black sigatoka resistance and fruit parthenocarpy in triploid AAB plantain. In *Agronomy abstracts.*: American society of Agronomy. Madison, WI. USA.

Ortiz-Vazquez E., Kaemmer D., Zhang H. B. 2005. Construction and characterization of a plant transformation-competent BIBAC library of the black Sigatoka-resistant banana *Musa acuminata* cv. Tuu Gia (AA). *Theor. Appl. Genet.* 110:706-713.

Oselebe H. O., Tenkouano A., Pillay M. 2006. Ploidy variation of *Musa* hybrids from crosses. *African J. of Biotechnology.* 5 (11):1048-1053.

Osuji J.O., Harrison G., Crouch J., Heslop-Harrison J. S. 1997. Identification of the genomic constitution of *Musa* L. lines (bananas, plantain and hybrids) using molecular cytogenetics. *Annals of Botany.* 80:787-793.

Page R. D. M., Holmes E. C. 1998. *Molecular evolution, a phylogenetic approach*: Blackwell Science Ltd.

Pan Q., Jonathan W., Robert Fluhr. 2000a. Divergent evolution of plant NBS-LRR resistance gene homologues in dicot and cereal genomes. *J. Mol. Evol.* 50:203-213.

Pan Q., Liu Y. S., Hadrin O.B., Sela M., Goren L.C., Zamir D., Fluhr R. 2000b. Comparative genetics of nucleotide binding site-leucine rich repeat resistance gene homologues in the genome of two dicotyledons: Tomato and *Arabidopsis. Genetics.* 155:309-322.

Pandolfini T., RotinoG. L., Camerini S., Defez R., Spena A. 2002. Optimization of transgene action at the post-transcriptional level: high quality parthenocarpic fruits in industrial tomatoes. *BMC Biotechnology.* 2 (1).

Parker J.E., Coleman M.J., Szabo V., Frost L.N., Schmidt R., van de Biezen E. A., Moores T., Dean C., Daniels M.J., Jones J. D. G. 1997. The *Arabidopsis* downy mildew resistance gene *RPP5* shares similarity to the toll and interleukin-1 receptors with N and L6. *Plant Cell.* 9:879-894.

Pauling L., and Corey R.B. 1953. Compound helical configurations of polypeptide chains: structure of proteins of the alpha-keratin type. *Nature.* 171 (4341):59-61.

Pearce S.R., Knox M., Ellis T.H. N., Flavell A.J., Kumar A. 2000. Pea *Ty1-copia* group retrotransposons: transpositional activity and use as markers to study diversity in *Pisum. Mol. Gen. Genet.* 263:898-907.

Peterson A. H., Bowers J.E., Burow M. D., Draye X., Elsik C.G., Jiang C.X., Katsar C. S., Lan T.H., Lin Y.R., Ming R. Wright R. J. 2000. Comparative genomics of plant chromosome. *The Plant Cell.* 12:1523-1539.

Piepho H.P., Koch G. 2000. Codominant analysis of banding data from a dominant marker system by normal mixtures. *Genetics.* 155:1459-1468.

Pillay M., Nwankanma D.C., Tenkouano A. 2000. Identification of RAPD markers linked to A and B genome sequences in *Musa* L. *Genome.* 43:763-767.

Pink D. A. C. 2002. Strategies using genes for non-durable resistance. *Euphytica.* 124:227-236.

Quint M., Duble C. M., Melchinger A. E. 2003. Identification of genetically linked RGAs by BAC screening in maize and implications for gene cloning, mapping and MAS. *Theor Appl. Genet.* 106:1171-1177.

Rajesh P.N., Coyne C., Meksem K., DerSharma K., Gupta V., Muehlbauer F.J. 2004. Construction of a *Hind*III bacterial artificial chromosome library and its use in identification of clones associated with disease resistance in chickpea. *Theor. Appl. Genet.* 108:663-669.

Raman V.S., Sree Rangaswamy S. R., Alikhan W.M., Manimekalai G. 1970. A comparative cytomorphological analysis of diploid bananas. *Madras Agric. J.* 57:604-611.

Ronald P., C. 1998. Resistance gene evolution. *Curr. Opin. in Plant Biology.* 1:294-298.

Ross M.T., LaBrie S., McPherson J., Stanto Jr. V. P. 1999. Screening large insert library by hybridization. In *Curr. Protocol In Human Genetics.*, edited by B. A. (Ed.). New York: Wiley.

Rowe P. R. 1984. Breeding Bananas and Plantain . *Plant Breeding Reviews.* 2:135-155.

Rowe P. R., Rosales F.E., ed. 1993. *Banana and Plantains.* Edited by Janick J. and More J. N., *Advances in Fruit Breeding.*: 2nd ed.Timber Press.

Rowe P. R. 1998. Latest developments in the FHIA banana and plantain breeding programme: Bred hybrids are now being grown commercially. *In Memorias. Seminario Internacional sobre Producción de Plátano. CORPOICA, Armenia, Colombia*:pg. 33-43.

Rozen S., and Helen J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology.*, edited by M. S. Krawetz S: Humana Press, Totowa, NJ.

Safar J., Noa-Carrazana J. C., Vrana J., Bartos J, Alkhimova O., Sabau X. Simkova H., Lheureux F., Caruana M. L., Dolezel J., Piffanelli P. 2004. Creation of a BAC resource to study the structure and evolution of the banana (*Musa balbisiana)* genome. *Genome.* 47:1182-1191.

Saito N., Nei M. 1987. The neighbour-joining method: a New method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4 (4):406-425.

Salmeron J.M., Oldroyd G.E., Rommens C.M., Scofield S.R., Kim H.S., Lavelle D.T., Dahlbeck D., Staskawicz B.J. 1996. Tomato *Prf* is a member of the leucine-rich repeat class of plant disease resistance genes and lies embedded within the *Pto* kinase gene cluster. *Cell.* 86:123-133.

Sambrook J., Fritsch E.F., Maniatis T. 1989. *Molecular cloning: A laboratory manual.* 2nd. ed: Cold Spring Harbor, New York. Cold Spring Harbor Laboratory Press.

San Miguel P., Gaut B.S., Tikhonov A., Nakajima Y., Bennnetzen J.L. 1998. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* 20:43-45.

Saraste M., Sibbald P.R., Wittinghofer A. 1990. The P-loop: A common motif in ATP- and GTP- binding proteins. *Trends Biochem. Science.* 15:430-434.

Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y., Wu J.Z., Niimura Y., Cheng Z. K., Nagamura Y. 2002. The genome sequence and structure of rice chromosome 1. *Nature.* 420:312-316.

Schmidt T., Heslop-Harrison J.S. 1996. The physical and genomic organisation of microsatellites in sugar beet. *Proc. Natl. Acad. Science. USA.* 93:8761-8765.

———. 1998. Genome, genes and junk: the large-scale organisation of plant genomes. *Trends Plant Sci.* 3:195-199.

Schmitt A., Specht T., Beckmann G., Dahl E., Pilarsky C., Hiznmann B., Rosenthal A. 1999. Exhaustive mining EST libraries for genes differentially expressed in normal and tumor tissues. *Nucleic Acids Res.* 27:4251-4260.

Schwarzacher T., Heslop-Harrison J. S. 2000. *Practical in situ hybridization:* Bios Scientific Publishers Ltd., Oxford UK.

Shen B., Wang D. M., McIntyre C. L., Liu C. J. 2005. A 'Chinese Spring' wheat (*Triticum aestivum* L.) bacterial artificial chromosome library and its use in the isolation of SSR markers for targeted genome regions. *Theor. Appl. Genet.* 111:1489-4194.

Shen K.A., Meyers B.C., Islam-Faridi M. N., Chin D.B., Stelly D.M., Michelmore R.W. 1998. Resistance genes candidates identified by PCR with degenerate oligonucleotide primers map to clusters of resistance genes in lettuce. *Mol. Plant Microb Interact.* 11:815-823.

Shepherd K., Ferrira F.R. 1984. The PNG biological foundation banana collection at Laloki, Port Moresby, Papua New Guinea. *South-east Asia Newsletter* 8:28-34.

Shepherd K. 1990. Genetic improvement of bananas in Brazil , aspects related to resistance to the genus *Mycosphaerella.* In *Sigatoka Leaf Diseases of Bananas*, edited by Fullerton R. A. and Stover R H. Montpellier, France: pp. 237-242, INIBAP.

Shiu S.H., Bleecker A.B. 2001a. Plant receptor-like kinase gene family: diversity, function and signaling. *Sci. STKE.* 113:RE22.

———. 2001b. Receptor-like kinases from *Arabidopsis thaliana* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl. Acad. Sci. USA.* 98:10763-10768.

———. 2002. Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiol.* 132:530-543.

Shizuya H., Birren B., Kim U.J., Mancino V., Slepak T., Tachiiri Y., Simon M. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci.* 89:8794-8797.

Shizuya H., Kouros-Mehr H. 2001. The development and applications of the bacterial artificial chromosome cloning system. *Keio J. Med.* 50 (1):26-30.

Silayoi B., Chomchalow N. 1987. Cytotaxonomi and morphological studies of Thai banana cultivars. In *Banana and Plantain Breeding Strategies*, edited by Persley G. J. DeLanghe E. A.: ACIAR proceedings. 21:157-160.

Simkova H., Cihalikova J., Virana J., Lysak M.A., Dolezel J. 2003. Preparation of HMW DNA fro plant nuclei and chromosomes isolated from root tips. *Biologia Plantarum.* 46 (3):369-373.

Simmonds N. W. 1953. Segregation in some diploid bananas. *J. Genet.* 51:458-469.

Simmonds N. W., Shepherd K. 1955. The taxonomy and origins of the cultivated bananas. *J. Linn. Soc. Bot.* 55:302-312.

Simmonds N. W. 1962. *Evolution of bananas*: Longman, London UK.

Siti Hawa J. 1998. Commercial Exploitation of the Banana Diversity in Malaysia. *Proceeding 1st. National Banana Seminar, 23-25 November 1998, Genting Highland, Malaysia. Pp. 13.*

Sneath P., Sokal R. 1973. *Numerical Taxonomy.*: Freeman. San Francisco.

Song W.Y., Wang L.L., Kim H.S. 1995. A receptor kinase-like encoded by rice disease-resistance gene, *Xa21*. *Science.* 270:1804-1806.

Song W.Y., Pi L.Y., Wang G.L., Gardner J., Holsten T., Ronald P.C. 1997. Evolution of the rice Xa21 disease resistance gene family. *Plant Cell.* 9:1279-1287.

Soriano J.M., Vilanova S., Romero C., Llacer G., Badenes M.L. 2005. Characterization and mapping of NBS-LRR resistance gene analogs in apricot (*Prunus armeniaca L.*). *Theor. Appl. Genet.* 110:980-989.

Stephenson P., Bryan G., Kirby J., Collins A., Devos K., Bussco C., Gale M. 1998. Fifty new microsatellite loci fir the wheat genetic map. *Theor. Appl. Genet.* 97:946-949.

Stergiou G., Katsiotis A., Hagidimitriou M., Loukas M. 2002. Genomic and chromosomal organisation of *Ty1-copia*-like sequences in *Olea europaea* and evolutionary relationships of *Olea* retroelements. *Theor. Appl. Genet.* 104:926-933.

Steward N., Ito M., Yamaguchi Y., Koizumi N., Sano H. 2002. Periodic DNA methylation in maize nucleosomes and demethylation by environmental stress. *J. Biol. Chem.* 277:37741-37746.

Stover R.H., Richardson D. L. 1968. 'Pelipita'. An ABB Blugoe-type plantain resistant to bacterial and *Fusarium* wilt. *Plant Dis. Reporter.* 52:901-903.

Strong S.J., Ohta Y., Litman G. W., Amemiya C.T. 1997. Marked improvement of PAC and BAC cloning is achieved using electroelution of pulse-field gel-seperated partial digests of genomic DNA. *Nucleic Acids Res.* 25 (19):3959-3961.

Stuart-Rogers C., Flavell A.J. 2001. The evolution of *Ty1-copia* group retrotranposons in gymnosperms. *Mol. Biol. Evol.* 18:155-163.

Sugimoto K., Otsuki Y., Saji S., Hirochika H. 1994. Transposition of the maize Ds element from a viral vector to the rice genome. *Plant J.* 5:863-871.

Suoniemi A., Tanskanen J., Pentikainen O., Johnson M. S., Schulman A. H. 1998. The core domain of retrotransposon integrase in *Hordeum*: predicted structure and evolution. *Mol. Biol. Evol.* 15:1135-1144.

Susko E., Roger A. J. 2004. Estimating and comparing the rates of gene discovery and expressed sequence tag (EST) frequencies in EST surveys. *Bioinformatics.* 20 (14):2279-2287.

Tanksley S. D. 2004. The genetic, development and molecular bases of fruit size and shape variation in tomato. *The Plant Cell.* 16:S181-S189.

Tao Q., Zhao H., Qiu L., Hong G. 1994. Construction of a full bacterial artificial chromosome (BAC) library of *Oryza sativa* genome. *Cell Res.* 4:127-133.

Tautz D. 1989. Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucleic Acids Res.* 17:6463-6471.

Temin H.M., Mizutani S. 1970. RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature.* 226:1211-1213.

Teo C. H., Tan S. H., Othman Y. R., Schwarzacher T. 2002. The cloning of Ty1-*copia*-like retrotransposons from 10 varieties of banana (*Musa* sp.). *J. Biochem. Mol. Biophys.* 6:193-201.

Thiel T., Michalek W., Varshney R.K., Graner A. 2003. Exploiting gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106:411-422.

Thomas C.M., Jones D.A., Parniske M., Harrison K., Balint-Kurti P.J., Hatzixanthis K., Jones J.D.G. 1997. Charaterization of the tomato *Cf-4* gene for resistance to *Cladosporium fulvum* identifies sequences that determine recognitional specificity in *Cf-4* and *Cf-9*. *Plant Cell.* 9:2209-2224.

Tichtinsky G., Vanoostuyse V., Cock J. M., Gaude T. 2003. Making inroads into plant receptor kinase signalling pathways. *Trends Plant Sci.* 8:231-237.

Trout T. W. 1994. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. *Eur. J. Bioscience.* 222:9-19.

Ude G., Pillay M., Nwakanma D., Tenkouano A. 2002. Analysis of genetic diversity and sectional relationships in *Musa* using AFLP markers. *Theor. Appl. Genet.* 104:1239-1245.

Ude G., Pillay M., Ogundiwin E, Tenkouano A. 2003. Genetic diversity in an African plantain core collection using AFLP and RAPD markers. *Theor. Appl. Genet.* 107:248-255.

Uptmoor R., Wenzel W., Friedt W., Donaldson G., Ayisi K., Ordon F. 2003. Comparative analysis on the genomic relatedness of *Sorghum bicalor* accessions from Southern Africa by RAPDs, AFLPs and SSRs. *Theor. Appl. Genet.* 106:1316-1325.

Valarik 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 Research.* 10:89-100.

Valmayor R.V., Jamaluddin S.H., Silayoi B.,Kusumo S., Danh L.D., Pascua O.C., Espino R.R.C. 2000. Banana Cultivars Names and Synonyms in Southeast Asia. *INIBAP.*

Varagona M.J., Purugganan M., Wessler S.R. 1992. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell.* 4:811-820.

Varmus H., Brown P. 1989. Retroviruses. In: Mobile DNA. *American Society for Microbiology*:53-108.

Varoquaux F., Blanvillain R., Delseny M., Gallois P. 2000. Less is better: new approaches for seedless fruit production. *Trends Biotech.* 18:233-242.

Vershinin A. V., Druka A., Alkhimova A. G., Kleinhofs A., Heslop-Harrison, J. S. 2002. LINEs and gypsy-like retrotransposons in *Hordeum* species. *Plant Mol. Biol. Evol.* 49:1-14.

Vilanova S., Romero C., Abbott A.G., Llacer G., Badenes M. L. 2003. An apricot (*Prunus armeniaca* L.) F2 progeny linkage map based on SSR and AFLP markers, mapping plum pox virus resistance and self-incompatibility traits. *Theor. Appl. Genet.* 107:239-247.

Vilarinhos A. D., Piffanelli P., Lagoda P., Thibivilliers S., Sabau X., Carreel F., D'hont A. 2003. Construction and characterization of a bacterial artificial chromosome library of banana (*Musa acuminata* Coll). *Theor. Appl. Genet.* 106:1102-1106.

Virk P.S., Zhu J., Newbury H.J., Bryan G.J., Jackson M.T., Ford-Lloyd B.V. 2000. Effectiveness of different classes of molecular marker for classifying and revealing variation in rice (*Oryza sativa)* germplasm. *Euphytica.* 112:275-284.

Vos P., Hogers R., Bleeker M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research.* 23:4407-4414.

Vuylsteke D., Swennen R. , Ortiz R. 1993. Development and performance of Black Sogatoka resistant tetraploid hybrids of plantain (*Musa spp.* AAB group). *Euphytica.* 65:33-42.

Walker J. C., Zhang R. 1990. Relationship of a putative receptor protein kinase from maize to the S-locus glycoproteins of *Brassica. Nature.* 345:743-746.

Wang G.L., Holsten T.E., Song W. Y., Wang H.P., Ronald P. C. 1995. Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. *Plant J.* 7:525-533.

Wang Q., Zhang K. Qu X., Jia J., Shi J.,Jin D., Wang B. 2001. Construction and characterization of a rice bacterial artificial chromosome library of peach. *Theor. Appl. Genet.* 103:1174-1179.

Whisson S. C., Van der Lee T., Bryan G.J., Waugh R., Govers F., Birch P.R.J. 2001. Physical mapping across an avirulence locus of *Phytophthora infestans* using a highly representative, large-insert bacterial artificial chromosome library. *Mol. Genet. Genomics.* 266:289-295.

Whitham S., Dinesh-Kumar S.P., Choi D., Hehl R., Corr C., Baker B. 1994. The product of the tobacco mosaic virus resistance gene *N*: Similarity to toll and the interleukin-1 receptor. *Cell.* 78:1011-1015.

Wild J., Hradecna, Z., and Szybalski, W. 2002. Conditionally amplifiable BACs: switching from single-copy to high-copy vectors and genomic clones. *Genome Res.* 12:1434-1444.

Williams J. G. K., Kubelik A.R., Livak K. J., Rafalski J. A., Tingsey S. V. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18:6531-6535.

Witsenboer H., Kesseli R.V., Fortin M. G., Stanghellini M., Michelmore R.W. 1995. Sources and genetic structure of a cluster of genes for resistance to three pathogens in lettuce. *Theor. Appl. Genet.* 91:178-188.

Wong C., Kiew R., Argent G., Set O., Lee S. K., Gan Y. Y. 2002. Assessment of the validity of the sections in *Musa* (Musaceae) using AFLP. *Annals Botany.* 90:231-238.

Woo S.S., Jiang J., Gill B.S., Paterson AS.H., Wing R.A. 1994. Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor. Nucleic Acids Res.* 22:4922-4931.

Wulf B. B. H., Thomas C. W., Smoker M., Grant M., Jones J.D.G. 2001. Domain swapping and gene shuffling identify sequences required for induction of AVR-dependent hypersensitive response by tomato *Cf-4* and *Cf-9* proteins. *Plant Cell.* 13:255-272.

Xiong Y., Eickbush T.H. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9:3353-3362.

Xu Q., Wen Xiaopeng, Deng X. 2005. Isolation of TIR and non-TIR NBS-LRR resistance gene analogues and identification of molecular markers linked to a powdery mildew resistance locus in chestnut rose (*Rosa roxnurghii* Tratt). *Theor. Appl. Genet.* 115:819-830.

Yang Y., Shah, J., Klessig, D.F. 1997. Signal perception and transduction in plant defense responses. *Genes Dev.* 11:1621-1639.

Ye Z. H., Droste D.L. 1996. Isolation and characterization of cDNAs encoding xylogenesis-associated and wounding-induced ribonucleases in *Zinnia elegans. Plant Mol. Biol.* 30:697-709.

Yoo E.Y., Kim S., Kim J.Y., Kim B. D. 2001. Construction and characterization of Bacterial Artificial Chromosome library from Chili pepper. *Molecules and Cells.* 12:117-120.

Young N. D. 2000. The genetic architecture of resistance. *Curr. Opin. in Plant Biology.* 3:285-290.

Young N.D. 1999. A cautiously optimistic vision for marker assisted-breeding. *Molecular Breeding.* 5:505-510.

Yuksel B., Estill J.C., Schulze S.R. 2005. Organisation and evolution of resistance gene analogs in peanut. *Mol. Gen. Genomic.* 274:248-263.

Zhang, H. B., Martin, G.B., Tanksley, S.D., Wing, R.A. 1994. Map-based cloning in crop plants: tomato as a model system II. Isolation and characterization of a set of overlapping yeast artificial chromosomes encompassing the jointless locus. *Mol. Gen. Genet.* 244:613-621.

Zhu J. K. 2001. Cell signaling under salt, water and cold stresses. *Curr. Opin. in Plant Biology* 4:401-406.

———. 2002. Salt and drought stress signals transduction in plants. *Annu. Rev. Plant Biol.* 53:247-277.

Zietkiewicz E., Rafalski A., Labuda D. 1994. Genomic fingerprinting by simple sequence repeat (SSR)-anchored PCR amplification. *Genomics.* 20:176-183.

APPENDIX

Appendix

Original list of *Musa* genomic DNA samples from CIRAD

Numéro attribué (extraction)	Reference	Species	Sub-species	Name	Country of origin	ITC code	Record number in MGIS
1	Eumusa	AAB	Nadan	Lady Finger ou AA Sucrier	India ?	ITC.0582	NEU0297
2	Eumusa	AAB	Pome / Prata	Foconah	Cameroon	ITC.0649	NEU0298
3	Eumusa	AAB	Pome / Prata	Prata Ana	Brazil	ITC.0962	NEU0310
4	Eumusa	*balbisiana*	type 4	P. Klutuk Wulung, IDN 056	Indonesia	ITC.1063	NEU0054
5	Eumusa	*balbisiana*	type 4	P. Batu, IDN 080	Indonesia	ITC.1156	NEU0055
Bank (6)	Eumusa	acuminata	Banksii	Banksii 623	Papua new guinea		
7	Eumusa	*acuminata*	*microcarpa*	Borneo	Malaysia, S/E Borneo	ITC.0253	NEU0028
8	Eumusa	*acuminata*	*burmannicoides*	Calcutta 4	India, Calcutta	ITC.0249	NEU0017
9	Eumusa	ABB	Ind ABB	K. Tiparot=?Te(e)parod, THA 020	Thailand	ITC.0652	NEU0383
10	Eumusa	AAB	Plantain	Orishele	Nigeria	ITC.1325	NEU0256
11	Eumusa	ABB	Pelipita	Pelipita	Philippines	ITC472	NEU0360
12	Eumusa	ABB	Bluggoe	Dole		ITC.0767	NEU0334
13	Eumusa	AAA	Cavendish	Grande Naine	Guadeloupe		NEU0172
14	Eumusa	AAA	Orotava	Pisang Kayu, (IDN098)	Indonesia	ITC0420	NEU0208
15	Eumusa	*acuminata*	*errans*	Agutay	Philippines	ITC.1028	NEU0033
16	Eumusa	*acuminata*	*siamea*	Khae (Phrae), THA 015	Thailand	ITC.0660	NEU0025
17	Eumusa	AAB	Figue Pomme / Silk	Figue Pomme Géante	Guadeloupe	ITC.0769	NEU0285
18	Eumusa	ABB	Saba	Saba	Philippines	ITC.1138	NEU0361
19	Eumusa	AAA	Ambon	Pisang bakar , IDN106	Indonesia,	ITC.1064	NEU0229
20	Eumusa	ABB	Monthan	Monthan	India ?	ITC0046	NEU0350
Tani (21)	Eumusa	balbisiana		Tani			
22	Eumusa	*acuminata*	*burmannica*	Long Tavoy pied		ITC.0283	NEU0016
23	Eumusa	AB cv		Safet Velchi	India	ITC.0245	NEU0152
24	Eumusa	AAA	Cavendish	Petite Naine		ITC.0654	NEU0174
25	Eumusa	*acuminata*	*banksii*	Paliama, PNG067	Papua New Guinea, East Sepik	ITC.0766	NEU0079
26	Eumusa	AAA	Cavendish	Poyo	Nigeria		NEU0165
27	Eumusa	AAB	Popoulou/Maia Maoli	Popoulou	Cameroon	ITC.0335	NEU0277
28	Eumusa	ABB	Peyan	Simili Radjah	From india through Zaire		NEU0357
29	Eumusa	AAA	Gros Michel	Gros Michel	Guadeloupe		NEU056
30	Eumusa	AS		Wompa, PNG063	Papua New Guinea	ITC.1152	NEU0020
31	Eumusa	AB cv		Kunnan	India, Kerala	ITC.1034	NEU0155
32	Eumusa	AAcv (18)	type P.jari buaya	P. Jari Buaya/BS312	Malaysia, Kelatan, Thai border	ITC.0312	NEU0117
33	Eumusa	AAcv (2)	type P.mas	P. mas / Figue Sucrée	Malaysia	ITC.0653	NEU0108
34	Eumusa	AAB	Pisang rajah	P. Raja Bulu, IDN 093	Indonesia	ITC.0843	NEU0276
35	Eumusa	AAA	Rio	Leite		ITC.0277	NEU0226
36	Eumusa	ABB	Ney mannan	Ice Cream	?	ITC020	NEU0353
37	Eumusa	*acuminata*	*zebrina*	Zebrina	Indonesia	ITC.1177	NEU0029
38	Eumusa	AAcv	Cooking AA	Tomolo, (PNG023)	Papua New Guinea, East New Britain	ITC.1187	NEU0082
39	Eumusa	*balbisiana*	type 1	Honduras	seeds from Honduras	ITC.0247	NEU0049
40	Eumusa	*balbisiana*	type3	Lal Velchi	India		NEU0051
41	Eumusa	ABB	Pisang awak	Namwa Khom, THA011	Thailand	ITC0659	NEU0347
42	Eumusa	AAA	Lujugira/Mutika	Mbwazirume	Burundi	ITC.0084	NEU0222
43	Eumusa	AAA	Lujugira/Mutika	Intokatoke	Burundi	ITC.0082	sent
44	Eumusa	AAA	Ibota	Yangambi KM5	Cameroon	ITC.1123	NEU0212
45	Eumusa	AAB	Plantain	Red Yade		ITC.1140	NEU0244
46	Eumusa	AAB	Nendra padaththi	P. Rajah	Brazil	ITC.0243	NEU0282
47	Eumusa	ABBB		Yawa 2, PNG 072	Papua New Guinea, East New Britain	ITC1238	NEU0384
48	Eumusa	AAB	Mysore	P. Ceylan	Thailand	ITC1441	NEU0284