THE APPLICATION OF IRAP MARKERS IN THE BREEDING OF PAPAYA

BATHUSHA SHAREEF BIN KADER SHAIK ALI

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2013

THE APPLICATION OF IRAP MARKERS IN THE BREEDING OF PAPAYA

BATHUSHA SHAREEF BIN KADER SHAIK ALI

DISSERTATION SUBMITTED IN FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2013

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **BATHUSHA SHAREEF BIN KADER SHAIK ALI**

I/C/Passport No: **810426075777**

Regisration/Matric No.: **SGR050021**

Name of Degree: **MASTER OF SCIENCE**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

"THE APPLICATION OF IRAP MARKERS IN THE BREEDING OF PAPAYA"

Field of Study: **PLANT BIOTECHNOLOGY**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work,
(2) This Work is original,
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work,
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work,
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained,
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

(Candidate Signature) Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name: **PROFESSOR DR ROFINA YASMIN OTHMAN**

Designation:

Witness's Signature Date:

Name: **ASSOC. PROF. DR KAMALUDIN A. RASHID**

Designation:

ABSTRAK

Di Malaysia, suatu program pembiakbakaan betik telah dimulakan oleh Institut Penyelidikan dan Pembangunan Pertanian Malaysia (*MARDI*) untuk menghasilkan betik-betik rintang virus. Walaubagaimanapun, pemilihan betik-betik rintang daripada cubaan ladang memakan masa dan tenaga. Oleh yang demikian, untuk memudahkan pemilihan betik-betik rintang daripada program pembiakbakaan, potensi kaedah pengecapjarian DNA telah dianalisa dalam kajian ini untuk diaplikasikan pada masa hadapan. Bahan iaitu daun betik (*Carica papaya* L.) daripada variasi Tainung 5, Eksotika 6 dan kacukan antara kedua-duanya telah digunakan dalam penilaian penanda molekul.

Daripada analisis penanda molekul Polimorfisma Diamplifikasi Antara-Retrotransposon (*IRAP*), lima kombinasi daripada sejumlah 45 kombinasi primer yang mungkin, menunjukkan polimorfisma corak jalur DNA yang signifikan. Seterusnya, corak jalur DNA telah digunakan untuk menentukan hubungan filogenetik antara sampel-sampel betik tersebut. Daripada lima kombinasi tersebut, hanya satu kombinasi primer menunjukkan corak jalur DNA yang menyamai pengagihan fenotip daripada keputusan cubaan ladang MARDI. Penanda *IRAP* boleh diuji selanjutnya untuk diaplikasikan dalam Pemilihan Dibantu Secara Molekul (*MAS*) dan Pembiakbakaan Dibantu Secara Molekul (*MAB*) untuk memudahkan pemilihan betik rintang virus bintik cincin betik.

Untuk kajian pada masa akan datang, lebih banyak primer-primer Polimorfisma Diamplifikasi Antara Retrotransposon boleh direka berdasarkan jujukan DNA Hujung Panjang Berulang (*Long Terminal Repeats*) dari genom betik. Penanda *IRAP* ini juga boleh digunakan dengan jujukan primer-primer lain seperti dari mikrosatelit (*REMAP*). Ini bertujuan untuk mendapatkan corak jalur DNA yang lebih bermaklumat dan dengan itu penanda IRAP dapat dieksploitasi sepenuhnya untuk memudahcara program

pembiakbakaan betik. Oleh yang demikian, diharapkan ia dapat memberi manfaat pada industri penanaman betik.

ABSTRACT

In Malaysia, a papaya breeding program was initiated by the Malaysian Agricultural Research and Development Institute (MARDI) to produce resistant papayas. Nevertheless, selection of resistant papayas from field trials is laborious and time-consuming. In order to facilitate selection of resistant papayas in breeding programs, potential of DNA fingerprinting method was analysed in this study for future applications. Materials from leaves of *Carica papaya* L. from varieties Tainung 5, Eksotika 6 and their hybrids were used in the assessment of molecular markers.

From an analysis of Inter-Retrotransposons Amplified Polymorphism Reaction (IRAP) molecular markers, five combinations out of 45 possible primers combinations showed significant polymorphism in DNA banding patterns. Furthermore, the banding patterns were used to assess the phylogenetic relationship between the papayas sample tested.

Out of the five combinations, only one combination of primers showed DNA banding patterns which is similar to the phenotypic segregation of MARDI's field trial results. IRAP markers could be further tested to be applied in Molecular Assisted Selection (MAS) and Molecular Assisted Breeding (MAB) to facilitate selection of PRSV resistant papayas.

For future studies, more IRAP primers can be designed based on Long Terminal Repeat (LTR) sequences of the papaya genome. IRAP markers could also be used in combination with other primer sequences such as from microsatellites. (Retrotransposon-microsatellite Amplified Polymorphism). This is to obtain more informative banding patterns and thus IRAP markers could be fully exploited to facilitate papaya breeding programmes. Hopefully, this would benefit the papaya planting industry.

ACKNOWLEDGEMENTS

All praise to Allah for enabling and guiding me to complete this humble piece of manuscript. My salaam and salute to Prophet Muhammad (peace be upon him) for being a mercy for mankind.

I would like to thank my mother, father and brother for their continuing support and motivation.

To my supervisors, Prof. Dr. Rofina Yasmin Othman and Assoc. Prof. Dr. Kamaludin A. Rashid who have been very helpful and supportive, thank you so very much.

I would also like to extend my thanks to Dr. Chan and Dr. Johari Sarip from Malaysian Agriculture Research and Development Institute (MARDI) for providing me with papaya samples for this project.

Not forgetting all professors, lecturers and staffs at the Unit of Genetics and Molecular Biology, Institute of Biological Sciences and Faculty of Science, thank you for your kind support and understanding.

Last but not least, my lab mates and friends; Dr. Suraya Faziella, Dr. Syarifah, Mohtaram, Walftor, Zalina, Florence and all BGM lab members who had helped me along the way, thank you very much and wish you all the best.

TABLE OF CONTENTS

TITLE			PAGE
ABSTRACT			v
ACKNOWLEDGEMENTS			vi
LIST OF FIGURES			viii
LIST OF TABLES			x
LIST OF SYMBOLS AND ABBREVIATIONS			xi
LIST OF APPENDICES			xiv
CHAPTER:			
1	INTRODUCTION		1
	1.1	Introduction	1
	1.2	Objectives	6
2	LITERATURE REVIEW		7
	2.1	Papaya Origin, Genetics and Taxonomic Classification	7
	2.2	Morphology and Cultivars	9
	2.3	Importance	12
	2.4	Pests and Diseases	13
	2.5	Strategies for Breeding Resistant Papayas	15
	2.6	DNA Fingerprinting and the Use of Molecular Markers	16
	2.7	Retrotransposons	21
	2.8	Inter-Retrotransposons Amplified Polymorphism Marker	26
	2.9	Phylogenetic tree	30
3	METHODOLOGY		32
	3.1	Plant Material	32
	3.2	Nucleic Acid Extraction	32
	3.3	DNA Purity, Quantitative and Qualitative	34
	3.4	Inter-Retrotransposons Amplified Polymorphism-Polymerase Chain Reaction (IRAP-PCR)	35
	3.5	Phylogenetic analysis	39
4	RESULTS		40
	4.1 Papaya Tissue		40
	4.2 DNA Purity, Quantitative and Qualitative		40
	4.3 IRAP-PCR		43
	4.4 Band Analyses		50
	4.5 Phylogenetic Tree Construction		51
5	DISCUSSION		57
	5.1 Papaya DNA Extraction		57
	5.2 IRAP-PCR		58
	5.3 Phylogenetic Analyses		61
	5.4 Phylogenetic Tree Inferences		61
	5.5 Results Analyses and Future Studies		62
6	CONCLUSION		70
	REFERENCES		91

LIST OF FIGURES

Figure	Title	Page
2.2(a)	Papaya tree	10
2.2(b)	Male papaya flower (center)	11
2.2(c)	Papaya leaves	11
2.5	Papaya fruit infected with Papaya Ringspot Virus (PRSV)	14
2.8	Structural features of the Ty1-*copia* retrotransposons showing the Long Terminal Repeats in direct orientation at each ends	24
2.9	IRAP schemes	28
4.2	Bands showing presence and quality of DNA extracted from papaya leaves	41
4.3(a)	IRAP-PCR using primers combination of LTR 6150 and 5'LTR1	45
4.3(b)	IRAP-PCR using primer combination of LTR 6150 and Nikita	46
4.3(c)	IRAP-PCR using primer LTR 6149	47
4.3(d)	IRAP-PCR using primers combination of LTR 6150 and RT (reverse)	48
4.3(e)	IRAP-PCR using RT (forward and reverse) primers combination	49
4.5(a)	Phylogenetic tree of *Carica papaya* L. based on 6150 and 5'LTR1 primers.	52
4.5(b)	Phylogenetic tree of *Carica papaya* L. based on 6150 and Nikita primers.	53
4.5(c)	Phylogenetic tree of *Carica papaya* L. based on 6149 primer	54
4.5(d)	Phylogenetic tree of *Carica papaya* L. based on 6150 and RT (reverse) primers	55

4.5(e)	Phylogenetic tree of *Carica papaya* L. based on RT (forward and reverse) primers	56
6.0	IRAP-PCR with 5'LTR1 & 3LTR primers	80
6.1	IRAP-PCR with 5'LTR2 & Sukkula LTR primers	80
6.2	IRAP-PCR with 5'LTR2 primer	81
6.3	IRAP-PCR with 5'LTR2 & Sukkula LTR primers	81
6.4	IRAP-PCR with 5'LTR1 primer	82
6.5	IRAP-PCR with LTR 6149 and LTR 6150 primers	82
6.6	IRAP-PCR with RT (forward and reverse) primers	83
6.7	IRAP-PCR with 5'LTR2 & Sukkula LTR primers	83
6.8	IRAP-PCR with 5'LTR1 primer	84
6.9	IRAP-PCR with LTR 6149 and Nikita LTR primers	84
7.0	IRAP-PCR with 5'LTR2 and 3'LTR primers	85
7.1	IRAP-PCR with LTR 6150 and RT (forward) primers	85
7.2	IRAP-PCR with 6149 and Nikita LTR primers	86
7.3	IRAP-PCR with 6150 primers	86
7.4	IRAP-PCR with LTR 6150 and 3'LTR primers	87
7.5	IRAP-PCR with Nikita and 5'LTR2 primers	87
7.6	IRAP-PCR with 3'LTR and Sukkula LTR primers	88
7.7	IRAP-PCR with 6150 and 5'LTR1 primers	88
7.8	IRAP-PCR with 5'LTR1 and Nikita LTR primers	89
7.9	IRAP-PCR with LTR 6149 and 5'LTR1 primers	89
8.0	IRAP-PCR with 3' LTR and Nikita LTR primers	90
8.1	IRAP-PCR with LTR 6149 and Sukkula LTR primers	90

LIST OF TABLES

Table	Title	Page
3.1	Description and explanatory notes of papaya samples	33
3.4 (a)	IRAP-PCR primers and their optimized Tm values	37
3.4 (b)	IRAP-PCR primers and their corresponding location to the sequence in Genbank database	38
4.2	DNA purity (O.D readings) and concentration of papaya samples	42
4.3	The degree of polymorphism of IRAP markers in *Carica papaya*	44
4.4	Band scoring for primers combination of LTR 6150 and 5'LTR1	50
4.5	Inter-Retrotransposon Amplified Polymorphism-Polymerase Chain Reaction (IRAP-PCR) primers combinations and their annealing temperatures (°C) for papaya	51

LIST OF SYMBOLS AND ABBREVIATIONS

Symbols	Abbreviations
°C	degree Celsius
%	percentage
bp	base pairs
cDNA	complementary DNA
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dsDNA	double stranded DNA
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
EDTA	ethylenediamine tetraacetic acid
β-gal	β-galactosidase
g	gram
µg	microgram
mg	milligram
ng	nanogram
kg	kilogram
µl	microlitre
ml	millilitre
min	minute
M	molar

mM	milliMolar
mmol	milllimoles
pM	picomolar
m	meter
cm	centimeter
nm	nanometer
U	unit
UV	ultra violet
V	volt
GMO	genetically modified organism
HCL	hydrochloric acid
IPTG	isopropyl-β-D-thiogalactopyranoside
Kb	kilobase pairs (duplex nucleic acid)
	kilobases (single-stranded nucleic acid)
lbs	pound
LINES	long interspersed elements
LTR	long terminal repeat
MARDI	Malaysian Agricultural Research and Development Institute
MAPPS	Malaysia Plant Protection Society
$MgCl_2$	magnesium chloride
mRNA	messenger RNA
nt	nucleotide
PCR	polymerase chain reaction
PRSV	papaya ringspot virus
RNase A	ribonuclease A
RNase H	ribonuclease H

rpm	revolution per minute
RT	reverse transcriptase
SINES	short interspersed elements
ssDNA	single-stranded DNA
TE	transposable elements
US	United States
X-gal	5-bromo-4-chloro-3-indolyl-β-D- galactopyranoside

LIST OF APPENDICES

APPENDIX	PAGE
APPENDIX A	77
APPENDIX B	79
APPENDIX C	80

CHAPTER 1

INTRODUCTION

1.1 Introduction

Papaya is an herbaceous fruit tree grown in most parts of the world while its wild relatives thrive in Southern America. As a member of the Brassicales, papaya shares a common ancestor with *Arabidopsis* from about 72 million years ago (Wikström *et al.,* 2001). Papaya is ranked first on nutritional scores among 38 common fruits, based on the percentage of the United States Recommended Daily Allowance for Vitamin A, Vitamin C, potassium, folate, niacin, thiamine, riboflavin, iron and calcium, plus fibre. Consumption of its fruit is recommended for preventing Vitamin A deficiency, a cause of childhood blindness in tropical and subtropical developing countries. The fruits, stems, leaves and roots of papaya are used in a wide range of medical applications, including production of papain, a valuable proteolytic enzyme (Ming, R. *et al*., 2008).

According to the Food and Agriculture Organization of the United Nations (FAO), Malaysia was ranked the world's nineteenth (19^{th}) largest papaya producer in 2010 (http://faostat.fao.org). Additionally, statistics from Malaysia's Department of Agriculture showed that Malaysia produced 49,760 metric ton of papaya in 2010 worth RM 68, 419, 982.

The extensive adaptation of this plant and wide acceptance of the fruit offers considerable promise for papaya as a commercial crop for local and export purposes. Like banana, pineapple and mango, papaya is one of the important cash crops in the tropics and subtropics. However, the production of this economically important fruit crop is limited by the destructive disease caused by papaya ringspot virus (PRSV) and the fragile and perishable nature of the fruit limit large-scale exportation, with the result that papaya lags behind banana and pineapple in world markets (Yeh *et al*., 2007).

Between 1993 and 1997, Hawaiian papaya production declined by forty percent due to invasion of the papaya ringspot virus (PRSV): one of the islands' major industries was thus threatened with extinction. By inserting a gene for just part of the virus's coat protein into the papaya's genome, researchers from Washington's University was able to create papaya resistant to attacks by the virus (Watson and Berry, 2004).

In May 1998, PRSV-CP transgenic papaya Rainbow and SunUp were developed in Hawaii were deregulated and granted approval for commercialisation, representing the first successful application of a transgenic fruit tree in the world (Yeh *et al*., 2006). Although the transgenic varieties are not resistant to most other PRSV strains from different geographic areas, the breakdown of the resistance in Hawaii has not been recorded suggesting that this approach had good potential.

In Malaysia, research at the Malaysian Agriculture Research Development Institute (MARDI) included a breeding program for developing new papaya varieties with improved eating qualities and yield for both export and local demands. Among them is Eksotika papaya variety. The Eksotika developed by MARDI in 1987 is a pure line like many popular varieties in the world today. Before the introduction of Eksotika, papayas were grown in backyards or as a cash crop during the early establishment of rubber or oil palm plantations (www.grain.org).

Eksotika is a MARDI backcross involving the backcrossing of local Subang with Sunrise Solo. Since the introduction, the papaya variety has been grown extensively in Johor and Perak for the domestic and international market. With Eksotika, permanent papaya farms were established to cater to the new export markets, ranging in size from 1-2 ha monoculture farms to large-scale plantations of 500 ha (www.grain.org).

After seven years of breeding, Eksotika II is a new Malaysian hybrid released by MARDI on 15th October 1991. Unfortunately, PRSV disease was detected in Malaysia in the southern state of Johor in 1991. The disease had affected papaya industry and most varieties commonly grown in the country including Eksotika are very susceptible to the PRSV infection.

To overcome this problem, MARDI had initiated breeding program for PRSV resistance in 1991. In this program, Tainung 5 (T5) variety which is a tolerant parent towards PRSV was crossed with variety Eksotika 6 (E6), which is a local but susceptible parent. The F1 hybrids, showed very good tolerance. Field test was carried out on 12 elite lines and the results of the trials indicated that three lines (Line 41, Line 90 and Line 248) exhibited high tolerance. Fruits of Line 90 rarely showed any PRSV symptoms while Lines 41 and 248 were very tolerant, showed only mild symptoms and continued yielding fruits long after susceptible trees were killed (Chan, 2002).

With the success of papaya breeding program, faster ways of evaluating hybrids were explored other than field testing. Among the faster ways of evaluating hybrids is with the use of biotechnology.

Genetic engineering and biotechnology hold great potential for plant breeding as it promises to expedite the time taken to produce and select crop varieties with desirable characters. With the use of molecular techniques it would now be possible to speed up the transfer of desirable genes from related wild species. Techniques which are particularly promising in assisting selection for desirable character involve the use of molecular markers such as retrotransposons based markers and random-amplified polymorphic DNAs (RAPDs) using F2 and back-cross populations, near-isogenic lines, doubled haploids and recombinant inbred lines (Mohan *et al*., 1997).

Once molecular markers closely linked to desirable traits are identified, marker-assisted selection can be performed in early segregating populations and at early stages

of plant development. Significantly marker-assisted selection (MAS) or identification can be used to pyramid the major genes including resistance genes, with the ultimate goal of producing varieties with more desirable characters. Thus, with MAS it may now be possible for the breeder to conduct many rounds of selection in a year. Molecular marker technology is now integrated into existing plant breeding programmes all over the world in order to allow researchers to access, transfer and combine genes at a rate and with a precision not previously possible (Mohan *et al*., 1997).

Different types of molecular markers target different parts of the plants' genome, such as restriction sites and repetitive elements.

Except for high copy number genes, repetitive elements have often been considered junk DNA with no function. However, recent studies suggest that they may play an important role as drivers of genome evolution in several regards, such as response to environmental cues, determination of continuous phenotypic characters and gene regulation (Nagarajan *et al*., 2008).

These repetitive elements otherwise known as retrotransposons can be used as markers because integration of a daughter copy creates new joints between genomic DNA and the conserved Long Terminal Repeats (LTRs). To detect polymorphisms for retrotransposon insertion, the marker systems that have been recently developed generally rely on Polymerase-Chain Reaction (PCR) amplification between an LTR and some component of the flanking genomic DNA (Kalendar and Schulman, 2006). This method is known as Inter Retrotransposons Amplified Polymorphisms – Polymerase Chain Reaction or IRAP-PCR.

There are two types of retrotransposons in a genome; LTR and non-LTR retrotransposons. The LTR retrotransposons, or Type I transposable elements, replicate by a process of reverse transcription resembling that of the lentivirus [such as Human Immunodeficiency Virus (HIV)]. The retrotransposons themselves encode the proteins

needed for their replication and integration back into the genome. Their “copy and paste” life cycle means that they are excised in order to insert a copy elsewhere in the genome. Hence, genomes diversify by the insertion of new copies, but old copies persist. Their abundance in the genome is generally highly correlated with genome size. Large plant genomes contain hundreds of thousands of these elements, together forming the vast majority of the total DNA (Kalendar and Schulman 2006).

Protocols for the IRAP techniques includes PCR amplification either with a single primer or with two primers and agarose gel electrophoresis of the product using optimal electrophoresis buffers and conditions. The protocol can be completed in one to two days (Kalendar and Schulman, 2006).

The genotype of the individual plant whether for resistance or susceptibility, could then be directly ascertained by the presence or absence of the marker band on the gel. Only the materials in the advanced generations would be required to be tested in disease and insect nurseries. The breeder would require little amount of DNA from each of the individual plants to be tested without destroying the plant (Mohan *et al.* 1997).

1.2 Objectives

The overall objective of this study was to asses a DNA based fingerprinting approach, Inter-Retrotransposons Amplified Polymorphism (IRAP) to potentially identify molecular markers which could be used to support papaya breeding programmes in particular papaya bred for resistance to Papaya Ringspot Virus's infection. The specific objectives of this study were:

i) To screen combinations of IRAP primers which produce polymorphic banding patterns.

ii) To score polymorphic bands from IRAP-PCR.

iii) To construct phylogenetic trees and to study pattern of inheritance in papaya samples.

iv) To identify IRAP markers which could be applied for selection of papaya resistant papayas towards papaya ringspot virus (PRSV).

The analysis was carried out using a test set of papaya plants obtained from MARDI's papaya ringspot virus (PRSV) resistance breeding programme.

CHAPTER 2

LITERATURE REVIEW

2.1 Papaya Origin, Genetics and Taxonomic Classification

Carica papaya Linnaeus or its few common names like paw-paw or papaw in some countries, is an angiosperm and a tropical dicot plant. This giant herb has no real wood, reproduces through seeds and originated from tropical South American countries. The Spaniards initially introduced papayas in the Philippines during the 16th century. Later, they were brought to Malacca and India by the Portuguese (Chan *et al*., 1994).

Papaya has a relatively small genome of 372 megabases (Mb), diploid inheritance with nine pairs of chromosomes, a well-established transformation system, a short generation time (9–15 months), continuous flowering throughout the year and a primitive sex-chromosome system (Ming *et al*., 2008). The small genome size of papaya and the fact that it can produce fruits in as nine months make it a potential model organism for fruit-producing tree crops (Ming *et al*., 2001). And added incentive to analyse this particular plant genome is provided by the identification of a primitive sex chromosome in papaya (Liu *et al*., 2004), which has commercial implications, as hermaphrodites have preferred agronomic characteristics (Chun *et al*., 2006).

In April 2008, papaya's genome had fully been sequenced and published. Papaya taxonomic classification is as follows (http://plants.usda.gov/java/nameSearch):

Kingdom: *Plantae*

Subkingdom: *Tracheobionta*

Superdivision*:* *Spermatophyta*

Division*:* *Magnoliophyta*

Class: *Magnoliopsida*

Subclass: *Dilleniidae*

Order: *Violales*

Family: *Caricaceae*

Genus: *Carica* L.

Species: *Carica papaya* L.

Within the *Caricaceae* family, there are five genera. They are the neotropical genera *Carica*, *Vasconcellea*, *Jarilla*, and *Jacaratia* and the equatorial African genus *Cylicomorpha*. The *Caricaceae* are herbaceous, shrubby or arborescent plants with a well-developed system of articulated laticifers. *Papaya* is the only species in the genus *Carica* (Kyndt *et al*., 2005).

Vasconcellea is a wild papaya. *Vasconcellea* is considered the most important genus of the *Caricaceae* as a consequence of its size and genetic diversity (Van Droogenbroeck *et al*., 2004), however papaya is the only *Caricaceae* species of worldwide economic importance (Manshardt and Drew, 1998).

2.2 Morphology and Cultivars

Papaya trees grow from three to eight meters high and maximum up to ten meters. They can live up to 15 years or more but their economical lifespan is two years only. They normally have no branches, so the leaves and fruits grow directly from the trunk. Branches do form only in rare cases when the trunk is damaged. The trunk can have a diameter of up to 20 cm (Tietze, 2001).

The leaves are located on the top of the trunk, placed in a spiral fashion, have mostly five to nine main fingers with a diameter of between 30 and 70 cm. The stems of the leaves can grow from 30 cm to one metre long. Leaves are being produced the whole year round and when they are four to six months old, they fall from the tree (Tietze, 2001).

Underneath the leaves, fruits are produced one by one or in groups. Female flowers will give rise to oval or round shaped fruits. On the other hand, hermaphrodite flowers will give rise to a longer oval shaped fruits. The fruits' flesh is soft and is yellow, orange or red in colour. Inside, 500-1000 seeds can be found from a complete pollinated fruit. But certain cultivars which are parthenocarpy might produce seedless normal fruits.

There are two types of papaya, the first one is a dioecous and another is gynodioceous. Dioecous type tree contain male and female on separate trees whilst gynodioecous type contain male in one tree and female with hermaphrodite in another. Others at certain seasons produce short-stalked male flowers, at other times perfect flowers. This change of sex may occur temporarily during high temperatures in midsummer. Male or bisexual plants may change completely to female plants after being beheaded. Certain varieties have a propensity for producing certain types of flowers (Tietze, 2001).

Insects' pollination (entomophily) sometimes compliments wind effective pollination (anemophily). Insects like moths would unintentionally pollinate papaya flowers although the flowers don't contain nectars. Unfortunately, midge and other pests also visit papaya flowers in order to obtain the stigmatic liquid which is rich with amino acid. Dioecious plants use wind to pollinate pollen grains from male trees to female trees. Compared to gynodioecious population which (cleistogamy) tend to release pollen grain before antesis. This would guarantee own pollination which doesn't affect plants (Tietze, 2001).

Papaya grows best in the tropics with the average temperature ranging from 24°C to 25°C and tolerates a soil pH of 4.3 up to 8.0 but the plant can grow with the readings over or below this value. Nevertheless, the ideal soil pH is 6 (Tietze, 2001).

Figure 2.2(a): Papaya tree.

Figure 2.2(b): Male papaya flower (center).

Figure 2.2(c): Papaya leaves.

Among papaya cultivars are Sunrise Solo (gynodiecous), Waimanalo (gynodiecous), Subang 6, Tainung, Sitiawan (gynodiecous) and Baru Arang (gynodiecous).

In Malaysia, the cultivars being planted for commercial purposes are Eksotika and Eksotika II which are both female and hermaphrodite. Eksotika is a crossed between Hawaiian Sunrise Solo (excellent eating qualities but with poor yield and small fruit) and locally adapted Subang 6 (large fruited).

Due to its shortcomings, Eksotika was subsequently crossed with its sister line (Line 19) which was resistant to freckles and had better keeping qualities. The resultant

hybrid was named Eksotika II and released in 1991. Eksotika II was more robust, higher yielding and had much improved fruit cosmetics and keeping quality. Unfortunately, Eksotika and Eksotika II varieties are susceptible to papaya ringspot virus.

Other cultivars besides these are Hong Kong, Solo, Kamiya, Mexican Red, Sunrise (Sunrise Solo), Sunset (Sunset Solo) and Vista Solo. One obvious difference between different cultivars is the size of the fruits (Tietze, 2001).

2.3 Importance

Papayas are grown commercially in most part of the world for their fruits. The green fruit of papaya is the source of the enzyme papain (protease), which is used in meat tenderizers. Besides that, papayas are used for preventive and restorative medicine. In Malaysia, commercial plantation of papaya has reached up to 49,760 metric ton with sales up to RM 68, 419, 982 in the year 2010 (Department of Agriculture Malaysia).

These plantations are mostly located in the southern part of peninsular Malaysia, which is in the state of Johor. The commercial production of papaya in Malaysia is headed by MARDI. Eksotika II is the commercially produced variety of papaya exported by Malaysia especially to Hong Kong and Taiwan. The plantation of papaya is also an important source of income for a lot of farmers in Malaysia.

Papaya ripe fruits are the richest in nutrients. They contain vitamin A, C and the Vitamin B complex, amino acids, calcium, iron, niacin, fibre, fat, carbohydrate, beta carotene, ascorbic acid, enzymes and so on. The protein in papaya is highly digestible. Dehydrated candied fruit, fruit rolls, chilli sauce, fruit cocktail are also produced from papayas (Tietze, 2001).

2.4 Pests and Diseases

Most commercial crops in the world face problems of pests and diseases including papaya. There are numerous bacterial, fungal, nematodes, parasitic, phytoplasmal, virus and viroid diseases related to papayas. Among all of these problems, the most feared in the papaya industry is the papaya ringspot virus type P (PRSV-P). This virus is considered the most significant pathogen of papaya, and losses as high as 70% of expected yield have been reported in affected areas (Manshardt and Drew, 1998).

In the year 1975, PRSV was first being reported in Southern Taiwan. From there, the virus spread over the West Coast and most of the commercial papaya orchards were destroyed (W.H. Lim and Doon Y., 1989). Now the virus is prevalent in most parts of the world including Malaysia.

Papaya ringspot virus is non-enveloped; rod shaped and is between 760-800 nanometres long and 12 nanometres in diameter. It is transmitted between plants by sap inoculation and non-persistently by a few aphid species like *Myzus persicae* and *Aphis gossipii*. Papaya Ringspot Virus classification is as follows:

Group: IV(+) sense RNA Viruses

Family: Potyviridae

Genus: Potyvirus

Species: Papaya Ringspot Virus

Although the symptoms induced on papaya are variable depending on the stage of infection, plant vigour, temperature, virus strain and plant size. Below are general symptoms of this virus infection:

i) Leaves mottling, malformation, chlorosis and yellowing.

ii) Ring spots and streaking on fruit, stems and petioles.

iii) Stunted plants.

iv) Prominent vein clearing on seedlings.

v) Fruits deformed, smaller in size, flavour and aroma impaired.

Plants that are found to be infected by this virus are slashed immediately and sprayed with *Paraquat* at concentration of 60 mL/4L of water transportation of infected plantlets and fruits from one place are also not recommended (Lim *et al*., 1994). To date, the only other resistance strategy has been transgenic, and although this approach has been successful it does not meet all the requirements of the different papaya growing regions (Dillon *et al*., 2006).

Figure 2.5: A papaya fruit infected with Papaya Ringspot Virus (PRSV). (http://the.honoluluadvertiser.com/dailypix/2004/Oct/17)

2.5 Strategies for breeding resistant papayas.

Genetic modification using coat-protein mediated resistance in the development of the first transgenic papaya with commercial potential, 'Rainbow', was reported by the University of Hawaii (Gonsalves, 1998). In conventional breeding for PRSV resistance, there is evidence that resistant genes exist in some cultivars such as 'Cariflora', 'Tainung 5', and Sinta. 'Tainung 5' is a gynodioecious cultivar bred from a cross between Florida FL77-5 and Costa Rica Red in Taiwan (Lin *et al*., 1989).

The Malaysian Agricultural Research and Development Institute (MARDI) initiated a breeding program for developing papayas with resistance to PRSV in 1991. The programme was divided into four phases i.e. 1. Hybridization of the susceptible 'Eksotika' with tolerant cultivars *viz*. 'Cariflora' and 'Tainung 5', followed by field evaluation of the F_1 hybrids. 2. Using the Single Seed Descent Method to develop F_5 inbred lines from the F_1 population by repeated self-pollination at each generation. 3. Screening and selection of F_5 seedlings for PRSV disease resistance by manual inoculation. 4. Evaluation of field tolerance and performance of selected lines (Chan *et al*., 2002).

Without the use of latest technological applications, the screening and selection of resistant plants have to be done manually. All plants being tested have to be grown individually to determine whether or not they are resistant to the virus. Furthermore, disease development in all the plants grown has to be observed and recorded until the plants mature. With the advent in the field of plant biotechnology, these observation and selection processes can be shorten with the use of molecular markers in DNA fingerprinting.

2.6 DNA Fingerprinting and the Use of Molecular Markers

How is DNA fingerprinting defined? The term was originally introduced by Jeffreys *et al*. to describe a method for the simultaneous detection of many highly variable DNA loci by hybridization of specific multilocus "probes" to electrophoretically separated restriction fragments. In recent years, several modifications of the basic technique have appeared, and related strategies have been developed. Most importantly, DNA polymorphisms become detectable by the polymerase chain reaction (PCR). Some of the new methods are still called DNA fingerprinting, but "DNA profiling," "DNA typing," or more specific terms have also been introduced (Weising *et al*., 1995).

According to this definition, DNA fingerprints are mainly obtained by either of two strategies which are "classical" hybridization-based fingerprinting and PCR-based fingerprinting. PCR-based fingerprinting involves the in vitro amplification of particular DNA sequences with the help of specifically or arbitrarily chosen oligonucleotides ("primers") and a thermostable DNA polymerase; the electrophoretic separation of amplified fragments, and the detection of polymorphic banding patterns by such method as staining (Weising *et al*., 1995).

The term marker is used in the sense of genetic marker. It is always synonymous with marker locus. A marker locus is a polymorphic locus that indicates the genotype of the individual that carries it: it is for this purpose that markers are used in population genetics. A marker locus also indicates the genotype of one or several loci linked to the marker; the applications here range from positional cloning to marker-assisted selection (de Vienne *et al*., 2003).

According to established terminology, the most commonly used genetic markers are morphological markers, molecular markers (at DNA level), and biochemical markers (isozymes, proteins). An "ideal" genetic marker is polymorphic: the geneticist's "raw material" is variability, multiallelic, codominant (a heterozygous hybrid simultaneously presents the traits of the homozygous parents; in a progeny, the heterozygous can be distinguished from each of the homozygotes), non-epistatic (where its genotype can be inferred from its phenotype, whatever the genotype at other loci may be), or, "neutral" (where the allelic substitutions at the marker locus do not have phenotypic or selective effects. Almost all molecular polymorphisms are neutral), and insensitive to the environment (de Vienne *et al*., 2003).

In contrast to morphological and biochemical markers, markers at the DNA level (molecular markers) are nearly infinite in number and are independent of the stage or organ analysed, since DNA is the same in all the tissues. Moreover, they have the advantage that they can be more directly used for further applications in molecular biology (de Vienne *et al*., 2003).

The development of so-called "molecular markers," which are based on polymorphisms found in proteins or DNA, has greatly facilitated research in a variety of disciplines such as taxonomy, phylogeny, ecology, genetics and plant breeding. For quite a long period of time, allozymes have been the molecular markers of choice. In recent years, attention has increasingly focused on the DNA molecule as a source of informative polymorphisms. Because each individual's DNA sequence is unique, this sequence information can be exploited for any study of genetic diversity and relatedness between organisms (Weising *et al*., 1995).

In the field of plant genetics and biotechnology, molecular markers can be specifically applied for molecular assisted breeding (MAB), molecular assisted selection (MAS), studies of genetic diversity and/or relatedness, population genetics studies, identification of genotypes, quantitative trait loci (QTL) analysis and construction of genetic linkage maps.

The utility of molecular markers for assisted selection of superior genotypes has been demonstrated in a large number of crop species (Jeffries *et al*., 2003, Lecouls *et al*., 2004, Narayanan *et al*., 2004, Yi *et al*., 2004).

Genetic maps made using single- or low-copy DNA sequences and molecular markers are now available for many species. The markers and maps are proving valuable for marker-assisted selection in plant breeding programmes, for targeted gene cloning, and for assisting fundamentals investigations of plant genome structure (Schwarzacher, 1994).

Numerous types of molecular markers are available. In order to select the right type, considerations should be given to use molecular markers that have good criteria such as highly polymorphic, high reproducibility, co-dominant inheritance (which allows us to discriminate homo- and heterozygotic states in diploid organism), frequent occurrence in the genome, non-epistatic, neutral behaviour, insensitive to the environment, even distribution throughout the genome, easy access, easy and fast assay (Weising *et al*., 1995)

Examples of several types of molecular markers are Inter Retrotransposons Amplified Polymorphism (IRAP), Random Amplified Polymorphic DNA (RAPD), Sequence Tagged Microsatellite Sites (STMS), Sequence-Characterized Amplified Region (SCAR), Single Strand Conformation Profile (or Polymorphism) (SSCP) and Amplified Fragment Length Polymorphism (AFLP).

Each types of molecular markers mentioned previously differs from each other in terms of cost of conducting experiments, speed in obtaining results, reproducibility level, equipments needed, amount of sequence information available from sample and distance of marker to the gene of interest (Pearce *et al*., 1999).

Marker systems based on transposable elements, in contrast to other methods, detect large changes in the genome. By comparison, RFLP (Restriction Fragment Length Polymorphism) and to some extent, AFLP (Amplified Fragment Length Polymorphism) detect single nucleotide changes that are bi-directional (have a fairly high reversion frequency), whereas SSR (Simple Sequence Repeat) or microsatellite polymorphism track the gain or loss of generally less than 20 nucleotides. Microsatellites alleles differ by the number of SSRs they contain and, like single nucleotide changes, also suffer from homoplasy because the number of SSRs can increase or decrease reversibly, making it impossible to distinguish ancestral and derived states (Kalendar and Schulman, 2006).

Retrotransposons have several advantages as molecular markers. Their abundance and dispersion can yield many marker bands, the pattern possessing a high degree of polymorphism due to transpositional activity (Kalendar and Schulman, 2006 and Kalendar *et al*., 2011)

The terminal sequences of long-terminal repeat (LTR) retrotransposons are a source of powerful molecular markers for linkage mapping and biodiversity studies. The major factor limiting the widespread application of LTR retrotransposon-based molecular markers is the availability of new retrotransposon terminal sequences (Pearce *et al*., 1999).

With multiple applications, molecular markers are becoming an important tool in biology. Molecular markers are applicable in plant and animal breeding, phylogeny, taxonomy and ecology thus allowing researchers all over the world to access, transfer

and combine genes at a rate and with a precision not previously possible. They also allow data comparison from various labs across the world and save time, money and cost in determining morphological characteristics of organisms. Below is previous work as examples of association of molecular markers with the desired traits in different crops:

Papaya:

The codominant CAPS marker Psilk4 develop in the study permits reliable detection of the PRSV-P resistant allele in hybrids of *V.cundinamarcensis* and *V. Parviflora* (Dillon *et al*., 2006).

Banana:

Analysis of somaclonal variation of tissue cultured banana plants derived from various explants using Random Amplified Polymorphic DNA (RAPD) markers (Bathusha and Othman, 2004).

Potato:

Race-specific resistance to *Phytophthora infestans R1* allele. Flanking RFLP loci-GP21 and GP179 and 2 AFLP loci not separated from *R1* by recombination (MeKsem K *et al.*, 1995).

Rice:

Resistance to blast caused by Pyricularia oryzae. *Pi-2(t)* gene located 2.8 cM from RG64 on chromosome 6 *Pi-4(t)* gene located 15.3 cM from RG 869 on chromosome 12. *Pi-10t* gene tagged with RAPD markers RRF6 and RRH18 on chromosome 5 (Yu *et al*., 1991).

Wheat:

Triticum tauschii cereal cyst nematode resistance Ccn-D1 gene. Complete cosegregation with RAPD fragment csE20-2 (Eastwood *et al*., 1994).

Tomato:

Phenotypic and molecular characterization of selected tomato recombinant inbred lines derived from the cross *Solanum lycopersicum* x *S. pimpinellifolium* (Pratta *et al*., 2011).

Sunflower:

DNA markers linked to the R2 rust resistance gene in sunflower (*Helianthus anuus* L.) facilitate anticipatory breeding for this disease variant (Lawson *et al*., 2011)

Primers specific for three pea LTRs have also been used to reveal polymorphisms associated with the corresponding retrotransposons within the *Pisum* genus (Pearce *et al*., 1999). However, one of the weaknesses of molecular markers is cross-over between the marker and the gene when the linked marker used for selection is at a distance away from the gene of interest. This produces a high percentage of false-positives/negatives in the screening process (Mohan *et al*., 1997).

2.7 Retrotransposons

Plant genomes have acquired a variety of repeat elements that account for up to 97% of nuclear DNA. For practical purposes, repetitive sequences can be divided into three main classes (1) transposable elements (TEs), which are the best-defined category and constitute the most abundant component of many genomes, ranging from 40%-80%, (2) tandem repeats, where individual copies are arranged adjacent to each other

forming tandem arrays of the monomeric unit, comprising hundreds or thousands of repeats. Finally, (3) high copy number genes, such as ribosomal or histone genes, are also an important part of the repeatome. TEs can be further divided into RNA-mediated class I retrotransposons and DNA-mediated class II transposons (Nagarajan *et al*., 2008 and Kalendar 2011).

RNA-mediated class I retrotransposons, or Long Terminal Repeat (LTR) retrotransposons are mobile genetic elements that transpose through reverse transcription of an RNA intermediate. The most studied group of LTR retrotransposons is the Ty1-*copia* group, named after the best characterised elements in *Saccharomyces cerevisiae* (Ty1) and *Drosophila melanogaster* (copia). Those that do not have LTR (non-LTR) are LINEs and SINEs group of retrotransposons (Kumar and Bennetzen, 1999).

It is now well known that retrotransposons of the Ty1-*copia* group are ubiquitous in plants (Flavell 1992, Flavell, Pearce & Kumar, 1994, Kumar 1996) and constitute a significant fraction of all plant genomes. 40% or more of the total genome may consist of retrotransposon sequences, based on estimates using the reverse transcriptase domain of Ty1-*copia* as a probe in broad (or field) bean (*Vicia faba*; Pearce et al., 1996a; 1996b).

A great variety of different Ty1-*copia* retrotransposons exist in most higher plants (Flavell *et al*., 1992, Matsuoka and Tsunewaki 1996, Pearce *et al*. 1996, SanMiguel *et al*. 1996, Vanderwiel *et al*., 1993, Voytas *et al*., 1992). Some of these, which have been transposing in the recent past, are extremely polymorphic within species and may be used as markers for linkage analysis and intra-specific diversity studies (Ellis *et al*., 1998, Waugh *et al*., 1997).

Other retrotransposons which were active several million years ago (SanMiguel *et al*., 1998), should be more useful in elucidating phylogenetic relationships between species (Pearce *et al*., 1999).

The general pattern of the genomic distribution of Ty1-*copia* group retrotransposable elements revealed by *in situ* hybridisation shows the elements are distributed throughout most of the length of plant chromosomes, with a few regions of higher and lower relative concentration in many species (Pearce *et al*., 1996a, 1997; Brandes *et al*., 1997).

These repetitive sequences account for around 56% of the papaya genome with the TEs being the most abundant at 52%, tandem repeats at 1.3% and high copy number genes at 3%. Most common types of TEs are represented in the papaya genome with retrotransposons being the dominant class, accounting for 40% of the genome. The most prevalent retrotransposons are Ty3-*gypsy* (27.8%) and Ty1-*copia* (5.5%) (Nagarajan *et al*., 2008).

On a different note, re-analysis of the repetitive DNA content of the papaya male-specific region of the Y chromosome (MSY) from papaya genome project's BACs, to include the new papaya-specific repeat families identified herein, increased the new papaya-specific repeat families identified herein, increased the average repeat sequence to 85.6%, with 54.1% Gypsy and 1.9% Copia retro-elements. This compares with an earlier estimate of 17.9% using the *Arabidopsis* repeat database alone (Ming *et al*., 2008).

Plant Ty1-*copia* group retrotransposons potentially provide an excellent system for developing a multiplex DNA-based marker system. This is largely due to the following properties; they are ubiquitous in the plant kingdom, they are present in high copy number, they possess a high degree of sequence heterogeneity, they are widely

dispersed within all chromosomes and they show insertional polymorphism both within and between species (Kumar *et al*., 1997).

Additionally, the genetic, structural and dispersion properties of the Ty1-*copia* group of retrotransposons make them an ideal tool for studying genetic biodiversity and for the generation of genetic markers in plant breeding. Extreme heterogeneity in the sequence of the Ty1-*copia* retrotransposons from potato, faba beans, barley, rye and onion was revealed following sequence analysis of reverse transcriptase fragments (Kumar *et al*., 1997).

The Ty1-*copia* group reverse transcriptase gene domains are concentrated in the centromeric regions, collocalizing with the 180bp satellite sequence pAL1 (Heslop-Harrison *et al*., 1997).

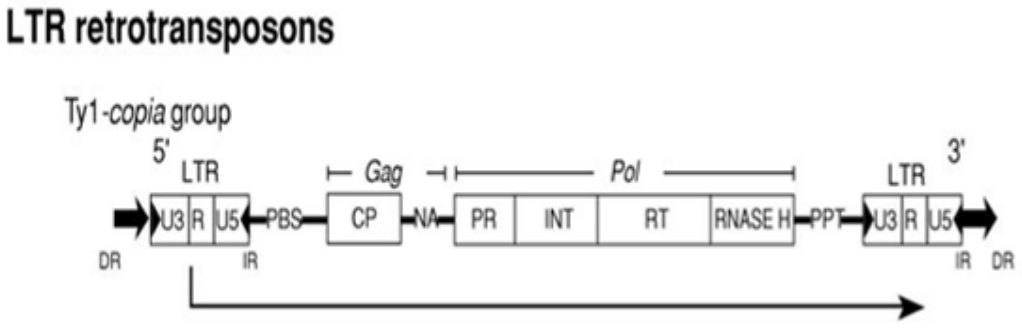

Figure 2.7: Structural features of the Ty1-*copia* retrotransposons showing the Long Terminal Repeats (LTR) in direct orientation at each ends (Kumar and Bennetzen, 1999).

Figure 2.7 above shows the Long Terminal Repeats (LTR) structural features of the Ty1-*copia* retrotransposons. Within these are U3, R and U5 regions, which contain signals for initiation and termination of transcription. Also showing the unique gene order within two open reading frames encoding for capsid-like protein (CP), protease (Pr), integrase (Int), reverse transcriptase (RT), and RNase-H, PBS-primer binding site, PPT-polypurine tract, NA-nucleotide acid binding (Kumar *et al*., 1997).

The long terminal repeats flank the genes encoding a core protein (gag protein) and a poly protein (pol). The poly protein consists of four domains, protease, integrase, reverse transcriptase (RT) and ribonuclease H (RNase H). Both groups of LTR retrotransposons are structurally similar to retroviruses with the major difference being that they lack a gene encoding the envelope (*env*) protein (Bennetzen 2000).

However, the two groups of LTR retrotransposons are distinguished by amino acid difference in their catalytic enzymes and the order of reverse transcriptase and integrase domains in their pol genes (Rogers and Paul, 2000).

The LTR sequences of Ty1-*copia* retrotransposons are present at each end of the retrotransposon and are identical. The LTRs contain no conserved motifs which would allow their direct amplification by PCR. The nearest conserved sequence motifs to the 3' LTR is within the adjacent RNaseH gene (Pearce *et al*., 1999).

Retrotransposition proceeds by transcription from DNA to RNA, reverse transcription from RNA to DNA, and finally, the synthesis of double-stranded DNA prior to integration in the host genome. Cellular RNA polymerases catalyze the first step, whereas reverse transcriptases do the rest. Reverse transcriptases as well as RNA polymerases lack exonucleolytic proofreading activity and therefore, a high error rate must be expected in this type of replication (Preston, 1996).

There are five major control points of the transposition cycle of LTR retrotransposon: transcription from DNA to RNA, translation and control of the gag/pol ratio, RNA packaging and virus-like particle assembly, reversely transcription to cDNA, and integration into host genomes (Grandbastien 1998). Restriction at any of these steps may hinder transposition (Tahara *et al*., 2004).

On the other hand, the sequences of copies transposed in the distant past are likely to have been differentiated by non-directional and independent mutations. The degree of differentiation within a family of copies therefore should be proportional to the age of the individual transposition events (Tahara *et al*., 2004).

2.8 Inter-Retrotransposon Amplified Polymorphism Marker

Retrotransposons can be used for markers because their integration creates new joints between genomic DNA and their conserved ends. The features of integration, activity, persistence, dispersion, conserved structure and sequence motifs and high copy number together suggest that retrotransposons are well suited genomic features on which to build molecular marker systems. Marker systems based on transposable elements, in contrast to other methods, detect large changes in the genome (Kalendar and Schulman, 2006).

Ruslan Kalendar and Alan H. Schulman from University of Helsinki developed a new method, known as IRAP (Inter-Retrotransposon Amplified Polymorphism), which requires neither restriction enzyme digestion nor ligation to generate marker bands (Kalendar *et al*., 2011).

IRAP is a dominant, multiplex marker system that examines variation in retrotransposon insertion sites. IRAP fragments between two retrotransposons are generated by PCR, using outward-facing primer annealing to LTR target sequences. Fragments are separated by high-resolution agarose gel-electrophoresis (Kalendar *et al*., 1999, 2011).

The IRAP method requires comparatively little sequence information before implementing them in a new plant species. The primary requirement is the sequence of an LTR end, harvested either from a database or produced by cloning and sequencing the genomic DNA that flanks conserved segments of retrotransposons. From their

initial description, IRAP have been applied in species ranging from barley, wheat and their relatives, to oat, apple, banana, citrus, grapevine, pea and sea grass (Kalendar and Schulman, 2006, Kalendar *et al*., 2011).

The IRAP method detects retrotransposon insertional polymorphisms by amplifying the portion of DNA between two retroelements. It uses one or two primers pointing outwards from an LTR, and therefore amplifies the tract of DNA between two nearby retrotransposons. IRAP can be carried out with a single primer matching either the 5' or 3' end of the LTR but oriented away from the LTR itself, or with two primers. The two primers may be from the same retrotransposon element family or may be from different families. The PCR products, and therefore the fingerprint patterns, result from amplification of hundreds to thousands of target sites in the genome. The complexity of the pattern obtained will be influenced by the retrotransposons copy number, which mirrors genome size, as well as by their insertion pattern and by the size of the retrotransposon family chosen for analysis. Furthermore, thousands of products can neither be simultaneously amplified to detectable levels nor resolved on a gel system. Hence, the pattern obtained represents the result of competition between the targets and products in the reaction. As a result, the products obtained with two primers do not represent the simple sum of the products obtained with the primers individually. The IRAP bands are generated from two nearby LTRs using outward-facing primers annealing to LTR target sequences (Kalendar and Schulman, 2006).

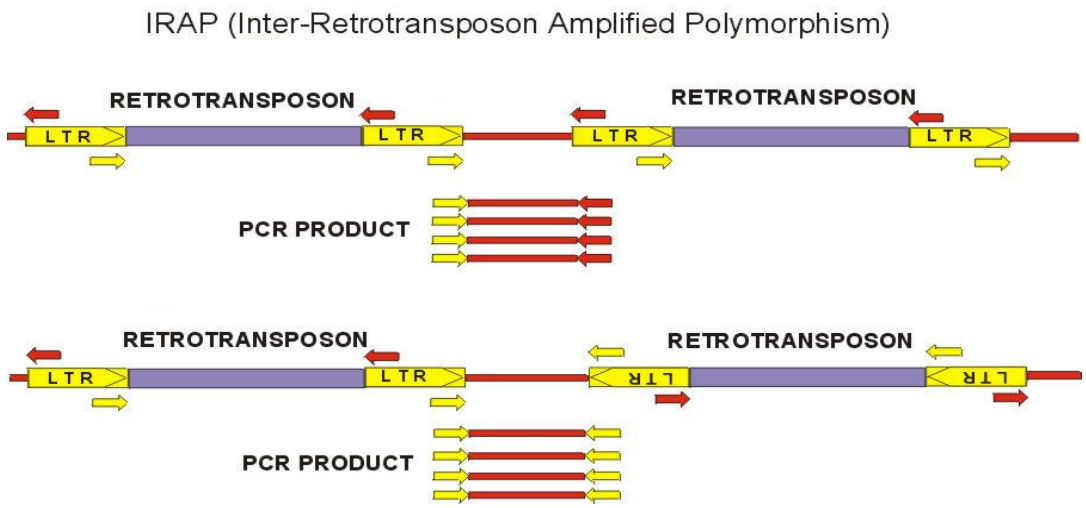

Figure 2.8: IRAP schemes.

(http://www.biocenter.helsinki.fi/bi/bare-1_html/images/irap.jpg)

IRAP is a PCR (Polymerase Chain Reaction) based marker technique that uses the proximity of LTR (Long Terminal Repeat) regions of adjacent retrotransposons to generate markers by use of outward facing primers that anneal to LTR target sequences. The LTR regions tend to be highly conserved because they contain sequences, essential for expression. New insertions lead to polymorphism in IRAP banding patterns (Lightbourn and Veilleux, 2003). It uses one or two primers pointing outwards from an LTR, and therefore amplifies the tract of DNA between two nearby retrotransposons (Kalendar and Schulman, 2006). It could also be used for unravelling the evolutionary history of retrotransposons insertions in plants (Price, 2002).

If retrotransposons were fully dispersed within the genome, IRAP would either produce products too large to give good resolution on gels or target amplification sites too far apart to produce products with the available thermostable polymerases. However, IRAP has succeeded for all genomes tried to date. This is because retrotransposons generally tend to cluster together in "repeat seas" surrounding "genome islands", and may even nest within each other. IRAP with BARE-1 primers

displays a range of products from 100bp upwards of 10kb (Kalendar and Schulman, 2006).

Steps that are involved in carrying out IRAP-PCR are sample preparation, designing IRAP primers, running polymerase chain reaction (PCR), casting agarose gel, loading PCR samples and running gel electrophoresis, DNA visualization and analyses.

The IRAP fingerprinting patterns can be used in a variety of applications, including measurement of genetic diversity and population structure, determination of essential derivation, marker assisted selection and recombinational mapping. In addition, the method can be used to fingerprint large genomic clones (e.g. BAC) for the purpose of assembly. The method can be extended, as well, to other prevalent repetitive genomic elements such as MITEs (Kalendar and Schulman, 2006).

Once the positions and matches of fingerprint bands have been scored, the data are ready for interpretation. The three main application areas for PCR- and hybridization-based DNA fingerprinting are (1) the identification of genotypes, (2) the assessment of genetic diversity and/or relatedness, and (3) segregation and linkage analysis for genetic mapping. In some of these areas, qualitative evaluation of the fragment pattern by eye may be sufficient to give a quick answer to the investigated problem. In general, however, the data have to be analyzed quantitatively with the help of various statistical methods (Weising *et al.*, 1995).

Once results of molecular markers have been obtained, analysis of fragment patterns is carried out by using statistical procedures. First, matches of fingerprint bands have to be scored. Then, the data will be ready for constructing phylogenetic trees. To date, there are many statistical software products that are available to construct phylogenetic trees.

2.9 Phylogenetic tree

A tree is a mathematical structure which is used to model the actual evolutionary history of a group of sequences or organisms. This actual pattern of historical relationships is the phylogeny or evolutionary tree which we try and estimate. A tree consists of nodes connected by branches (also called edges). Terminal nodes (also called leaves, OTU [Operational Taxonomic Units], or terminal taxa) represent sequences or organisms for which we have data, they may be either extant or extinct. Internal nodes represent hypothetical ancestors; the ancestor of all the sequences that comprise the tree is the root of the tree (Page and Holmes 1998).

The genetic distance can serve as a basis for the classification of the materials of interest (e.g., lines, populations). Many algorithms are presently used but the most commonly used are those of the average linkage (commonly called UPGMA), where the distance between two groups is defined as the mean distance between elements of the groups, and the method of Ward (1963), which minimizes the variation within the groups formed. These analyses can be complemented by other descriptive approaches of the variability, such as principal component analysis or factorial analysis of correspondences (de Vienne 2003).

The polymorphism parsimony method was first used by Felsenstein, and the results published (without a clear specification of the method) by Inger (1967). The method was independently published by Farris (1978) and by Felsenstein (1979). The method assumes that we can explain the pattern of states by no more than one origination (0-->1) of state 1, followed by retention of polymorphism along as many segments of the tree as are necessary, followed by loss of state 0 or of state 1 where necessary. The program tries to minimize the total number of polymorphic characters, where each polymorphism is counted once for each segment of the tree in which it is

retained. The assumptions of the polymorphism parsimony method are in effect (Felsenstein, J., 2004):

1. The ancestral state (state 0) is known in each character.
2. The characters are evolving independently of each other.
3. Different lineages are evolving independently.
4. Forward change (0-->1) is highly improbable over the length of time involved in the evolution of the group.
5. Retention of polymorphism is also improbable, but far more probable that forward change, so that we can more easily envisage much polymorphism than even one additional forward change.
6. Once state 1 is reached, reoccurrence of state 0 is very improbable, much less probable than multiple retentions of polymorphism.
7. The lengths of segments in the true tree are not so unequal that we can more easily envisage retention events occurring in both of two long segments than one retention in a short segment.

Among the software products to construct phylogenetic trees are PAUP*, JOINMAP 2.0, PHYLIP and MEGA. These software differ in terms of price whereby some are freely available through distribution by the author such as PHYLIP and some are available by purchase such as PAUP*. They also differ in terms of complexity and speed of use where some software products need the instructions to be keyed in to run programs while some others simply use the point and click method.

Upon successful construction of phylogenetic trees, the next step is to infer the trees, which means giving meaning to them.

CHAPTER 3

METHODOLOGY

3.1 Plant Material

Leaves of *Carica papaya* L. from varieties Tainung 5 (T5), Eksotika 6 (E6) and 14 of their hybrids (Table 3.1) were collected from plants at the Malaysian Agriculture Research and Development Institute (MARDI) in Serdang. The leaves of these plants were used as a source of genomic DNA (gDNA) for PCR amplification and analysis of Inter-Retrotransposon Amplified Polymorphism bands.

3.2 Nucleic Acid Extraction

Prior to plant DNA extraction, the following solutions were prepared: Extraction buffer (consisting of 5% hexadecyltrimethylammonium bromide (CTAB), 1.4 M NaCl, 0.4% 2-mercaptoethanol, 20 mM EDTA, 100 mM Tris-HCL, pH 8.0), TE buffer (10mM Tris-HCL, pH8, 1mM EDTA), washing solution (76% of ethanol, 10mM ammonium acetate, chloroform:isoamylalcohol (24:1) solution, 70% and 96% ethanol). Extraction protocol involved a modified CTAB procedure of Doyle and Doyle (1990).

Approximately 3.0 gram of young papaya leaf (*Carica papaya* L.) sample were ground with a pre-chilled mortar and pestle in the presence of liquid nitrogen. The frozen powder was transferred into a polypropylene tube containing 15ml of preheated extraction buffer (pH 8). The extraction buffer and frozen powder were mixed evenly and incubated at 60°C for 60 minutes with occasional gentle inverted mixing every ten minutes. After that, the mixture was cooled at room temperature and equal volume of chloroform: isoamyl alcohol (24:1) was added. The mixture was then centrifuged at 5000 rpm for ten minutes at 25°C.

Table 3.1: Description and explanatory notes of papaya samples

Sample Name	Description	Notes
T5	Tainung 5 Parental Line	Samples from one plant. Gynodioecious cultivar bred from a cross between Florida FL77-5 and Costa Rica Red in Taiwan. Most resistant towards PRSV.
E6	Eksotika 6 Parental Line	Samples from one plant (most susceptible towards PRSV from Malaysia). Onset of the disease and fruit symptoms.
L13A	Hybrid of T5 and E6	Samples from one plant each. F_5 inbred line derived from single seed descent method with seedling screening for PRSV resistance. From F_1 to F_4 selection was made without prior pressure on yield and quality. Among the most resistant lines. Field test showed disease symptoms low and build-up gradually. Fruits size resemblance to E6.
L13B		
L13C		
L13D		
L33A	Hybrid of T5 and E6	Samples from one plant each. F_5 inbred line selected after three rounds of manual inoculation with the same batch of seeds as Line 13. Susceptible towards PRSV. Showed gradual plant and fruit symptoms in field trial results.
L33B		
L33C		
L41A	Hybrid of T5 and E6	Samples from one plant each. F_5 inbred line selected after three rounds of manual inoculation with the same batch of seeds as Line 13 and Line 33. Among the most resistant towards PRSV.
L41B		
L41C		
L41D		
L90A	Hybrid of T5 and E6	Samples from one plant each. The only line selected for field trial after three rounds of manual inoculation from second batch of seeds. Immune and most impressive in term of fruit symptoms (hardly any symptoms like ringspot or necrosis).
L90B		
L90C		

DNA was precipitated by adding 0.6 volume of ice-cold isopropanol and incubated at -20°C for 30 minutes. Precipitated DNA was centrifuged at 5000 rpm for 10 minutes at 4°C. The supernatant was decanted carefully and the pellet washed with 20ml of washing solution. The pellet was then agitated gently and centrifuged at 5000 rpm for 10 minutes at 4°C. After that, 2 volume of 96% ethanol was added, mixed by inversion, stored at -20°C for 1 hour and later centrifuged at 5000 rpm for ten minutes at 4°C. Next, the pellet being washed finally with 70% ice-cooled ethanol and centrifuged at 5000 rpm for 10 minutes at 4°C . The pellet was then dried in a speed vacuum for 20 minutes before finally being dissolved in 1ml of deionised water. After the DNA dissolved in the water, the tube was kept at -20°C as stock for further analysis.

3.3 DNA Purity, Quantitative and Qualitative

The purity of the nucleic acid samples was determined by calculating the ratio between the optical density readings at 260 nm and 280 nm. High purity will give a ratio of 1.8 to 2.0 of optical density (OD). For estimating DNA concentration, spectrophotometric readings of UV absorbance at 260 nm using Eppendorf's Biophotometer machine was taken. An optical density of 1.0 at 260 nm is equivalent to 50mg/mL of double-stranded DNA. Prior to Biophotometer readings, all DNA samples were diluted 100-1000 times using deionised water.

3.4 Inter-Retrotransposon Amplified Polymorphism - Polymerase Chain Reaction (IRAP-PCR)

The IRAP method was carried out as previously described by using long terminal repeat (LTR) primers derived from the barley genome (*Hordeum vulgare*) (Kalendar *et al.,* 1999, 2000 ; Boyko *et al.,* 2002). Primer sequences and sources of retrotransposons are shown in Table 3.4(a). The Sukkula and Nikita LTR primers were designed facing outward from the *Sukkula* LTR and *Nikita* LTR, respectively. The LTR 6149, LTR 6150, 5' LTR1 and 5' LTR2 were designed facing outward from the 5' LTR region of the *BARE*-1a retrotransposons sequence whereas the 3' LTR was designed facing outward from the 3' ends of the *BARE*-1a LTR.

The IRAP-PCR primers were designed from the long terminal repeats (LTR) of the *BARE*-1 retrotransposon and the internal region of reverse transcriptase (RT) of papaya Ty1-*copia*-like retrotransposons. The IRAP-PCR primers and their corresponding locations to the sequence in Genbank database were summarized in Table 3.4(a).

The reverse primer LTR 6150 is complementary to bases 418-439 of the 5' LTR of the *BARE*-1a sequence (accession Z17327, 309-2137), and the forward primer LTR 6149 matched bases 1993-2012 of the 5' LTR of the *BARE*-1a sequence (accession Z17327). The location of the primers corresponded to the conserved stretches of the *BARE*-1 LTR, particularly at the 3' ends of the primers, based on initial alignments (Suoniemi *et al.*, 1997).

The 5' LTR1 primer matched to bases 36369-36394 of the 5' LTR of the *BARE*-1 sequence (accession AF254799, 34586-36407) whereas the 5' LTR2 primer is complementary to bases 310-331 of the 5'LTR of the *BARE*-1a sequence (accession Z17327). The location of the primers corresponds to conserved stretches of the 5' ends of the *BARE*-1 LTR.

The 3’ LTR primer is complementary to bases 14971-14994 of the 3’ LTR of the *BARE*-1 retrotransposon (accession AF254799, 14967-16796). The Sukkula LTR primer matched bases 4935-4959 of the *Sukkula* LTR-1 (accession AF254799, 6245-11204) and the Nikita LTR primer matched bases 1623-1645 of the *Nikita* LTR-1 (accession AF254799, 38443-40171).

For IRAP-PCR, the total genomic DNA of papaya samples was diluted with sterile dH_2O to 100 ng/µl concentration. The IRAP-PCR was performed in a 20 µl reaction mixture containing 100ng DNA, 1X PCR buffer (Promega, USA), 2mM $MgCl_2$, 16 pmol of each primer, 200 µM dNTP mix, 1 U *Taq* 50polymerase (Promega, USA) in 0.2 ml tubes. Amplification was performed using Eppendorf Mastercycler 5330.

The PCR reaction program consisted of: 1 cycle at 95°C, 2 min; 1 cycle at 95°C, 60s; annealing temperature (Tm) [Table 3.4(b)], 60s; ramp +0.5°Cs^{-1} to 72°C; 30 cycles of 72°C, 2min + 3s; 1 cycle at 72°C, 10 min; 20°C.

After the IRAP-PCR reaction, the IRAP-PCR products were analyzed by electrophoresis on 1.6% (w/v) agarose and detected by ethidium bromide staining. The agarose gel was analysed using Quantity One® version 4 for Windows® (Bio-Rad, USA).

Table 3.4(a): IRAP-PCR primers and their corresponding location to the sequence in Genbank database.

Primers	**Location / Accession Number**
RT (reverse)	Internal region of reverse transcriptase of the Ty1-*copia*-like retrotransposon
RT (forward)	Internal region of reverse transcriptase of Ty1-*copia*-like retrotransposon
LTR 6149	Bases 1993-2012 of the 5' LTR of the *BARE*-1a sequence / Z17327
LTR 6150	Bases 418-439 of the 5' LTR of the *BARE*-1a sequence / Z17327
5'LTR2	Bases 310-321 of the 5' LTR of the *BARE*-1a sequence / Z17327
5'LTR1	Bases 36369-36394 of the 5' LTR of the *BARE*-1 sequence / AF254799
3' LTR	Bases 14971-14994 of the 3' LTR of the *BARE*-1 sequence / AF254799
Sukkula LTR	Bases 4935-4959 of the *Sukkula* LTR-1 sequence / AF254799
Nikita LTR	Bases 1623-1645 of the *Nikita* LTR-1 sequence / AF254799

Table 3.4(b): IRAP-PCR primers and their optimized Tm values.

Primer	**Primer Sequence (5’ to 3’)**	**Priming Site**	**Tm (°C)**
5’ LTR2	GCCTCTAGGGCATAATTCCAAC	5’ LTR	50
3’ LTR	TTGGTTTCCCATGCGACGTTCCCC	3’ LTR	56
Sukkula LTR	GATAGGGTCGCATCTTGGGCGTGAC	Sukkula LTR	58
Nikita LTR	AAGAAGTGCCTATGGACAAATCC	Nikita LTR	48
LTR 6149	CCACTACATCACCCGCGTATATT	5’ LTR	50
LTR 6150	TTGTCTATGTATCCACACATGTA	5’ LTR	45
5’ LTR1	AACTATATTTATTATTGCCTCTAGGG	5’ LTR	47
RT (forward)	ACNGCNTTYYTNCAYGG	RT	40
RT (reverse)	ARCATRTCRTCNACRTA	RT	40

3.5 Phylogenetic analysis

Phylogenetic trees were constructed from tables consisting of scores for presence or absence of bands of particular mobility in accessions. Only bands that are clear and distinguishable were taken into account for the scoring. A value of 1 indicates the presence of a band of a particular mobility compared to a DNA ladder and a value of 0 indicates the absence of that particular band.

In this study, all phylogenetic trees were constructed using the Dollop program from PHYLIP (PHYLogeny Inference Package) software (version 3.69). PHYLIP software is distributed freely by the author. All the trees were bootstrapped with 500 replicates and were not rooted.

Unrooted trees lack a root, and hence do not specify evolutionary relationships in quite the same way, and they do not allow us to talk of ancestors and descendants. Furthermore, sequences that may be adjacent on an unrooted tree may need not be evolutionarily closely related (Page and Holmes, 1998).

Upon completion of band scoring table for all primers combinations, the data were written into an input file according to a format specified by PHYLIP. Then, SeqBoot program was run using the prepared input file. Here, the data (input files) were bootstrapped with 500 replicates.

After that, the output files (results) were run using Dollop program (becomes the new input files) which specify polymorphism parsimony as the method. Next, all the trees produced by Dollop were analyzed and the program used for this purpose was Consense. Consensus trees are trees that represent the commonality (if any) among a set of trees (Page and Holmes, 1998).

CHAPTER 4

RESULTS

4.1 Papaya Tissue

Fresh leaves were collected from a total of 16 plants (two parental lines of T5 and E6 and 14 hybrids labelled L13A, L13B, L13C, L13D, L33A, L33B, L33C, L41A, L41B, L41C, L41D, L90A, L90B and L90C. These plants were used for the IRAP analyses throughout the study.

4.2 DNA Purity, Quantitative and Qualitative Analysis

Papaya DNA extraction was successful with all 16 samples collected showing the presence of generally intact DNA bands suitable for PCR applications (Figure 4.2). The DNA concentration of the samples obtained ranged from low concentration to highly concentrated DNA as observed from the DNA qualitative gel analysis. Optical density (OD) readings showed sufficient level of purity for use in IRAP-PCR. For use in IRAP-PCR, further purification such as RNase treatment was not necessary. All 16 papaya samples have OD readings that ranged between 1.7 to 2.1 while DNA concentration ranged between 0.1 µg/µL to 2.5 µg/µL (see Table 4.2).

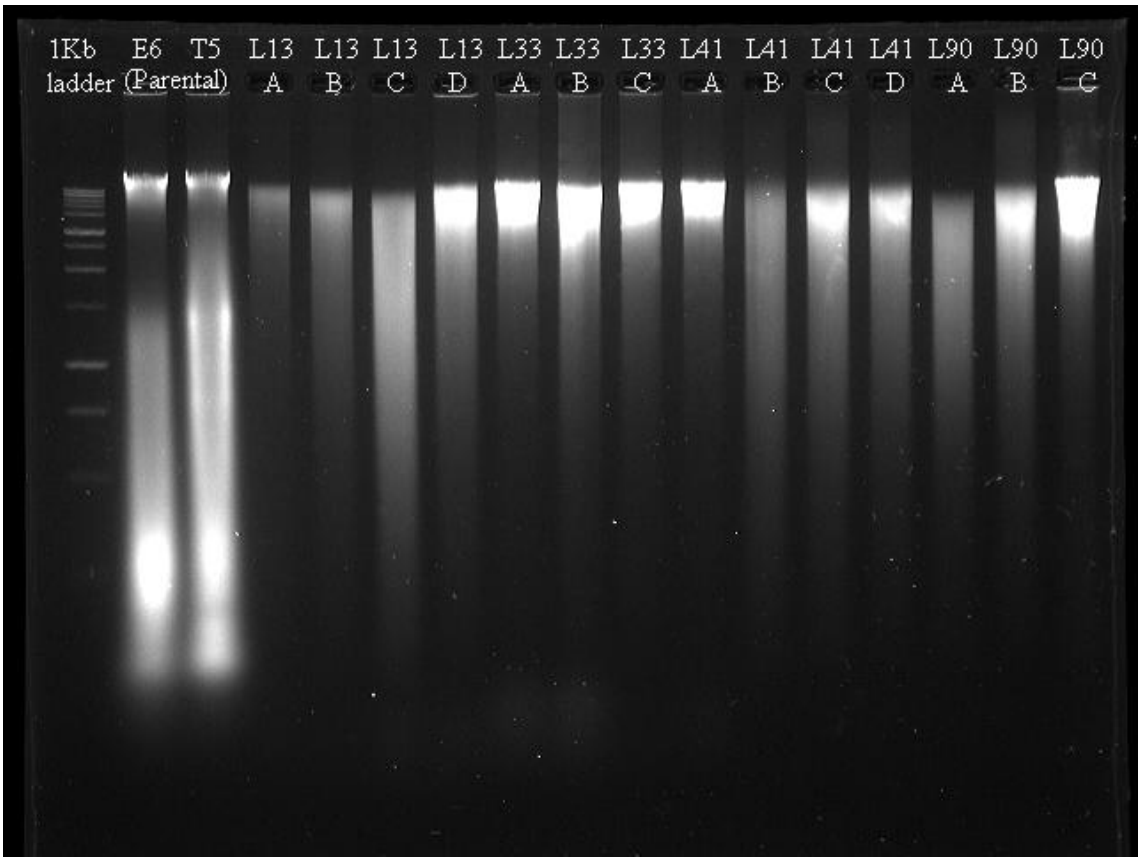

Figure 4.2: Bands showing presence and quality of DNA extracted from papaya leaves. 1Kb ladder was used in lane 1. Lane 2-3 represents parental papayas (E6 and T5). Lane 4-17 represents 14 hybrid papayas from four lines (L13-L90).

Table 4.2: DNA purity (O.D readings) and concentration of papaya samples.

Sample	**Purity (O.D reading)**	**DNA Concentration (µg / µL)**
T5	1.73	1.728
E6	2.03	1.913
L13A	1.88	0.516
L13B	1.91	0.337
L13C	2.00	0.496
L13D	1.80	1.534
L33A	1.90	1.795
L33B	1.91	2.178
L33C	1.87	1.208
L41A	1.79	1.771
L41B	1.70	0.155
L41C	2.08	0.519
L41D	1.92	0.911
L90A	1.95	0.612
L90B	1.89	0.352
L90C	2.10	1.199

4.3 IRAP-PCR

IRAP-PCR was carried out using primer combinations as shown in Table 4.3. Each analysis was repeated in triplicate. The annealing temperatures were calculated based on the sequences of the primers [see Table 3.4(b)]. From 28 possible IRAP and RT primers combinations, a total of 45 possible primers were tested. Initial analysis (Table 4.3) showed that only five primers combinations successfully produced significant polymorphic banding pattern results. These combinations were: LTR 6150 with 5'LTR, LTR 6150 with Nikita LTR, LTR 6149 with itself, LTR 6150 with RT Reverse and finally, RT forward with RT reverse.

The results showed that the five primers combinations produced significant polymorphic banding patterns that could potentially discriminate between the parental samples [Figures 4.3(a) - 4.3(e)]. Representative gel analysis results for less discriminatory primer pairs are shown in Appendix B.

The results suggest that these five primers combinations could be developed to determine parental-progeny relations and possibly to classify progenies based on their resistance towards PRSV (resistant progenies to resistant parent and vice versa). Based on the IRAP-PCR results, the PCR bands were then scored and the data used to construct phylogenetic trees.

Table 4.3: Inter-Retrotransposon Amplified Polymorphism-Polymerase Chain Reaction (IRAP-PCR) primer combinations and their annealing temperatures (°C) for papaya.

Primer	LTR 6149	LTR 6150	5' LTR1	5' LTR2	3'LTR	SUKKULA LTR	NIKITA LTR	RT REVERSE	RT FORWARD
LTR 6149	40.5	-	-	-	-	-	-	-	-
LTR 6150		-	45.5	-	-	-	45.5	40.0	-
5' LTR1			-	-	-	-	-	-	-
5' LTR2				-	-	-	-	-	-
3' LTR					-	-	-	-	-
SUKKULA LTR						-	-	-	-
NIKITA LTR							-	-	-
RT REVERSE								-	40.0
RT FORWARD									-

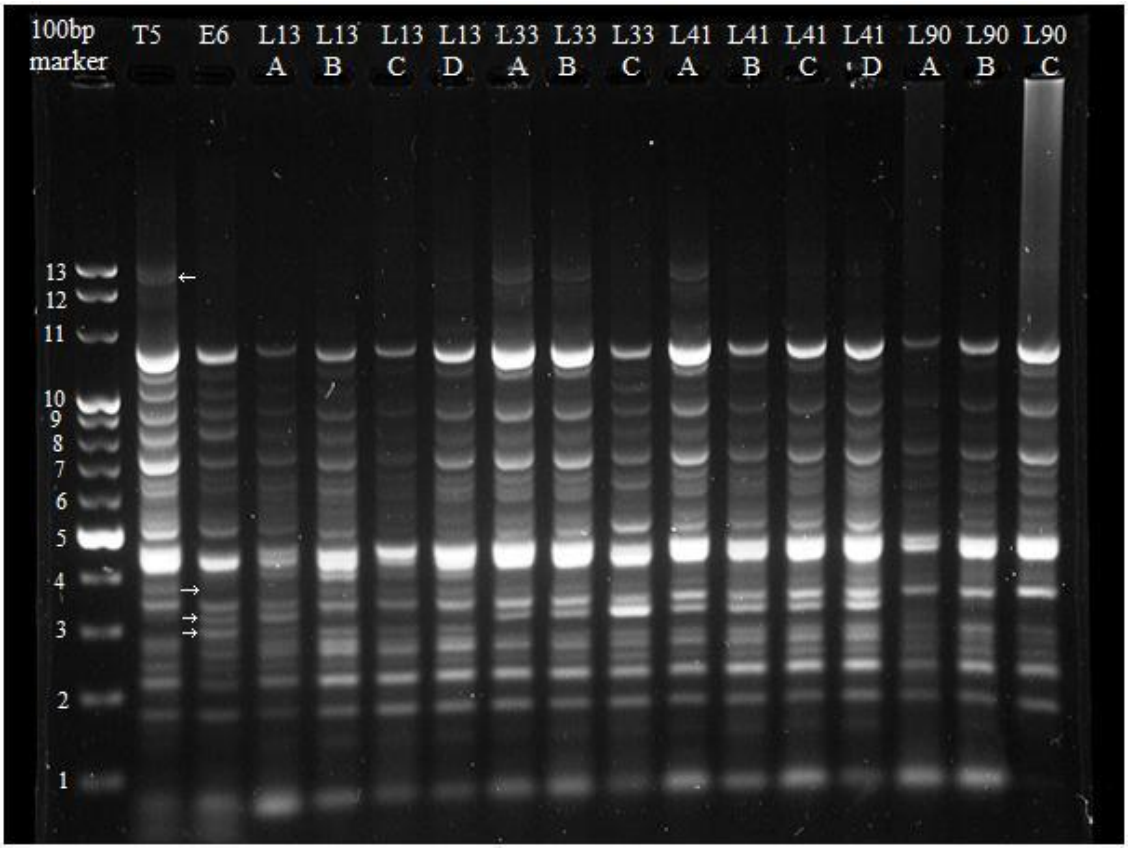

Figure 4.3(a): IRAP-PCR using primer combination of LTR 6150 and 5'LTR1. The molecular weights of the 100bp DNA ladder are: 1 = 100bp, 5 = 500bp, 7 = 700bp, 10 = 1000bp and 13 = 1500bp. Arrows indicating bands that are present in either one of the parents only.

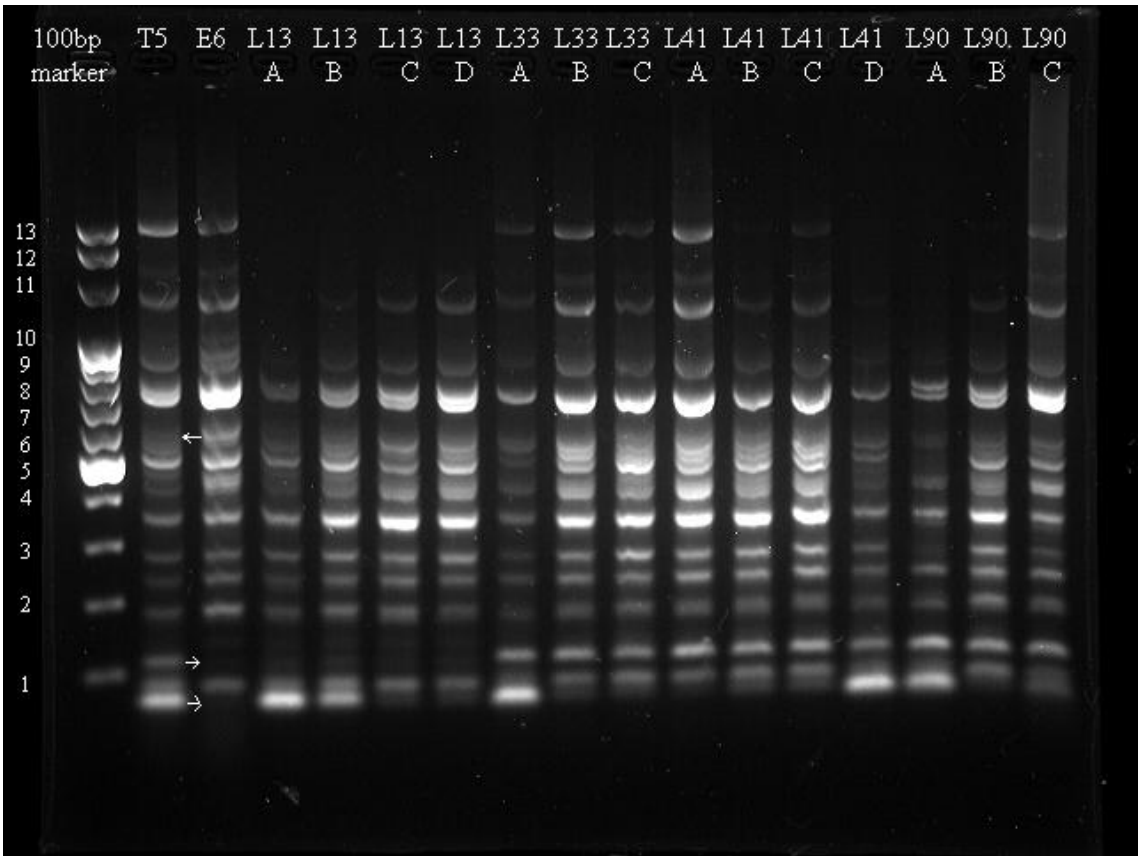

Figure 4.3(b): IRAP-PCR using primer combination of LTR 6150 and Nikita LTR. The molecular weights of the 100 bp DNA ladder are: 1 = 100bp, 5 = 500bp, 7 = 700bp, 10 = 1000bp and 13 = 1500bp. Arrows pointing to bands that are present in one of the parents only.

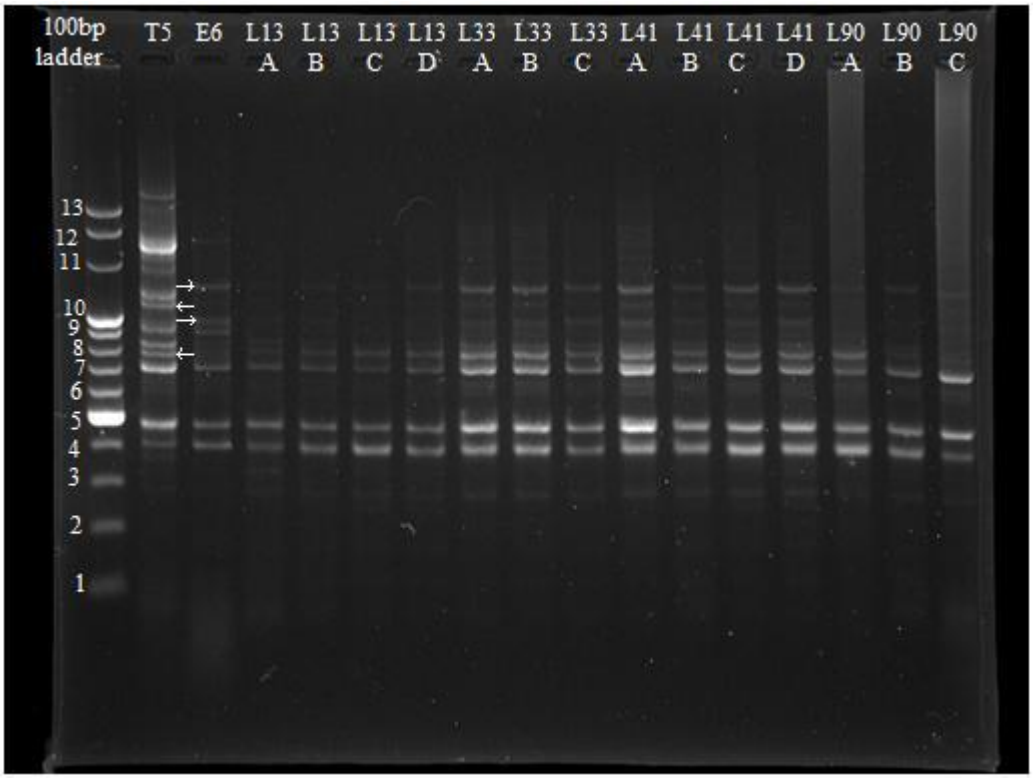

Figure 4.3(c): IRAP-PCR using primer LTR 6149 (combination of both LTR 6149). The molecular weights of the 100 bp DNA ladder are: 1 = 100bp, 5 = 500bp, 7 = 700bp, 10 = 1000bp and 13 = 1500bp. Arrows showing bands that are present in one of the parents only.

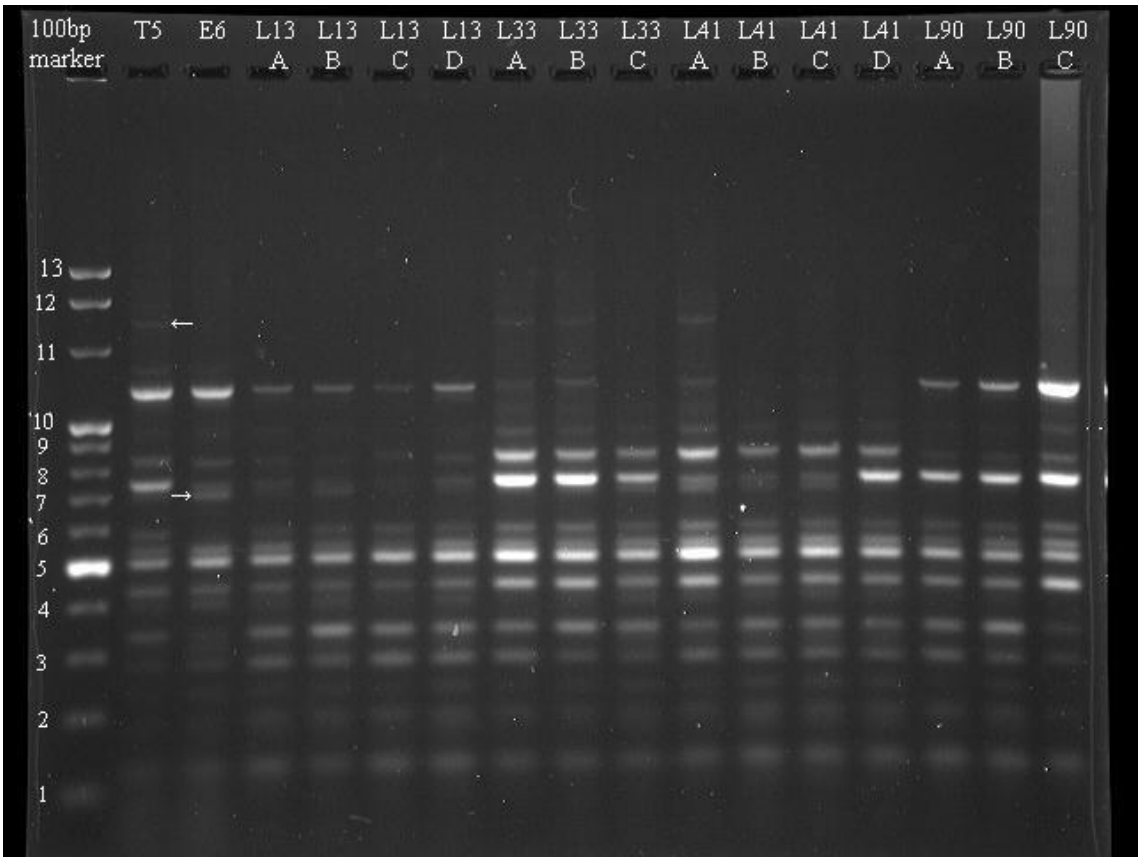

Figure 4.3(d): IRAP-PCR using primer combination of LTR 6150 and RT (reverse). The molecular weights of the 100 bp DNA ladder are: 1 = 100bp, 5 = 500bp, 7 = 700bp, 10 = 1000bp and 13 = 1500bp. Arrows indicating bands that are present in one of the parents only.

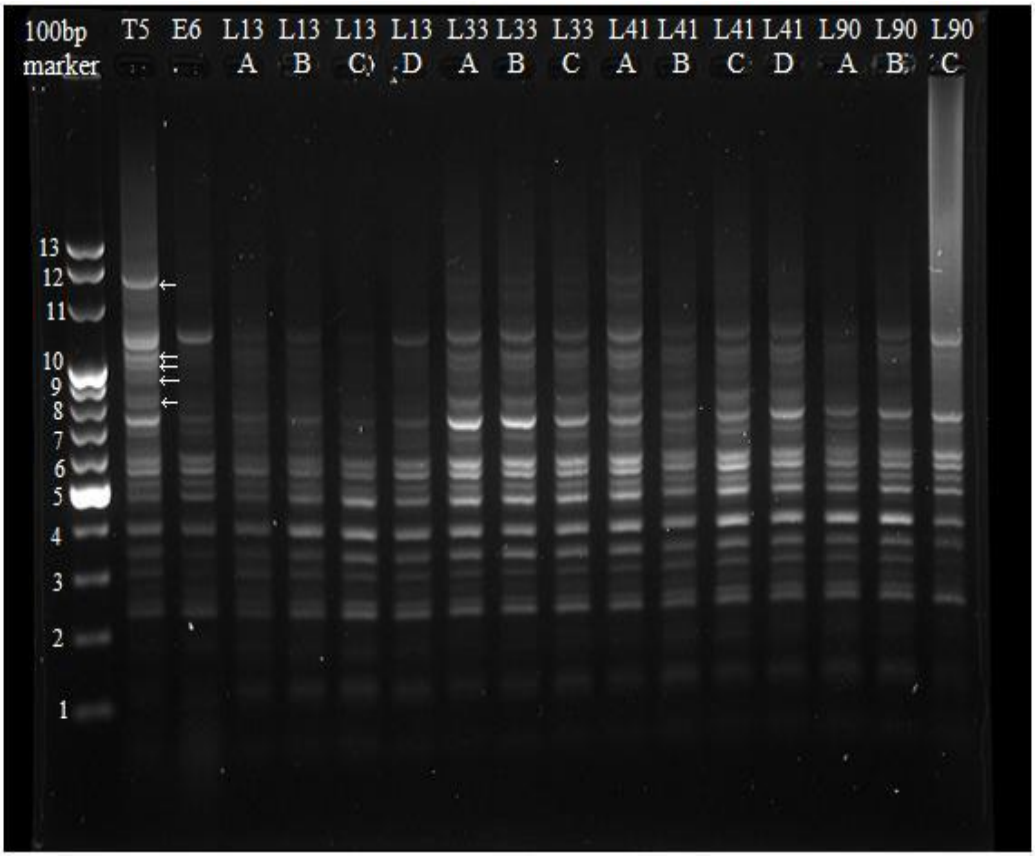

Figure 4.3(e): IRAP-PCR using RT (forward and reverse) primer combination. The molecular weights of the 100bp DNA ladder are: 1 = 100bp, 5 = 500bp, 7 = 700bp, 10 = 1000bp and 13 = 1500bp. Arrows pointing to bands that are present in one of the parents only.

4.4 Bands Analyses

The analyses of the bands showed that the degree of polymorphism of IRAP products in *Carica papaya* is high in this study, as it was observed to be between 39% to 75% (Table 4.4). In total, about half of the IRAP products were significantly polymorphic bands with overall 54% degree of polymorphism. The percentage of polymorphism was calculated based on the total number of polymorphic bands over the total number of bands scored.

The combination of IRAP primers with the lowest degree of polymorphism were for primers RT Forward & RT Reverse with 39% polymorphic bands generated while the primer combination of LTR 6149 and LTR 6149 generated the highest degree of polymorphism which was 75% of polymorphic bands.

The smallest band was approximately 70 base pairs and the largest band observed was approximately 1500 base pairs long. Both parental lines T5 and E6 samples produced polymorphic bands which show inter-specific banding patterns.

There were in total 81 polymorphic bands among all the samples, with several shorter fragments amplified from most or all lines, suggesting conservation of the internal organization of parts of retroelements.

Table 4.4: The degree of polymorphism of IRAP markers in *Carica papaya* L. Results are based on scoring of 16 samples.

Primers combination	**Total number of bands**	**Number of polymorphic bands**	**Percentage of polymorphism**
LTR 6150 & 5'LTR1	20	11	55%
LTR 6150 & Nikita LTR	16	8	50%
LTR 6149 & LTR 6149	12	9	75%
LTR 6150 & RT (reverse)	15	9	60%
RT (forward & reverse)	18	7	39%
TOTAL	81	44	54%

4.5 Phylogenetic Tree Construction

Table 4.5 shows a sample of the band scoring for primers combination of LTR 6150 and 5'LTR1. The final stage was to draw the constructed tree using Drawtree program. The results of the analyses are shown in Figures 4.5(a) – 4.5(e).

Table 4.5: Band scoring for primers combination of LTR 6150 and 5'LTR1.

Band size / **Samples**	**300 bp**	**320 bp**	**370 bp**	**1500 bp**
T5	0	0	1	1
E6	1	1	0	0
L13A	1	1	0	0
L13B	1	0	1	0
L13C	1	0	1	0
L13D	1	0	1	1
L33A	0	1	1	1
L33B	0	1	1	1
L33C	0	1	1	0
L41A	0	1	1	1
L41B	0	1	0	0
L41C	0	1	1	1
L41D	0	1	1	1
L90A	1	0	0	0
L90B	0	0	0	0
L90C	0	0	1	0

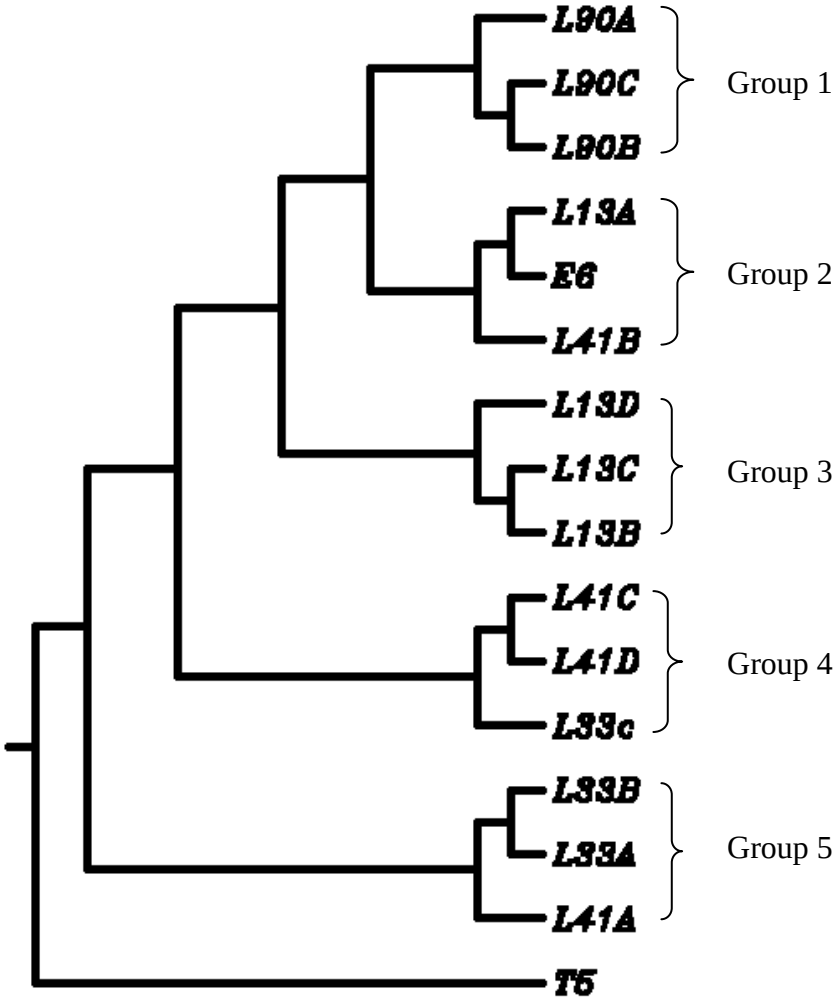

Figure 4.5(a): Phylogenetic tree of *Carica papaya* L. based on LTR 6150 and 5'LTR1 primers.

Phylogenetic tree in Figure 4.5(a) has five groups. Sample T5 which is the outgroup, is closely related to sample L41A followed by samples L33A and L33B. In Group 2, another parental sample E6 is closely related to sample L13A followed by sample L41B. Each group has three papaya samples.

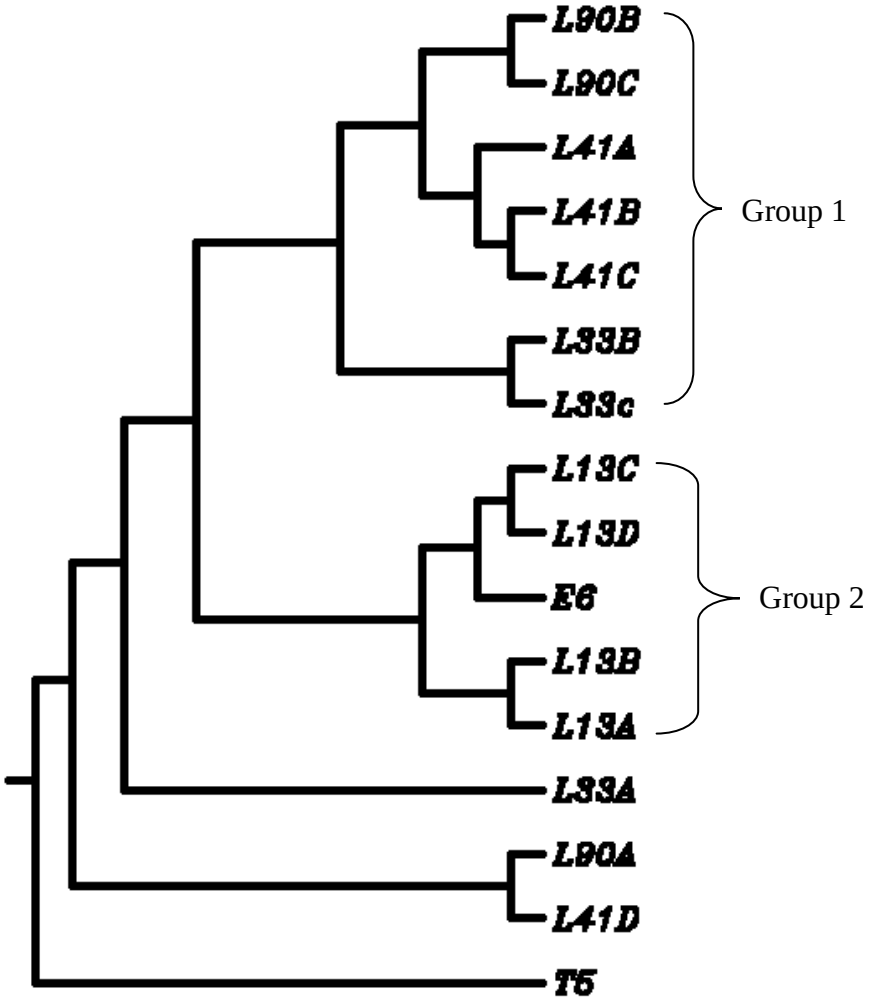

Figure 4.5(b): Phylogenetic tree of *Carica papaya* L. based on LTR 6150 and Nikita LTR primers.

The phylogenetic tree from Figure 4.5(b) is divided into two groups which are further divided into three and two subgroups each. Parental sample T5 which is the outgroup is closely related to samples L41D and L90A. Another parental sample E6 is closely related to samples L13C and L13D from Group 2. Group 1 showed samples from Lines 90 and Lines 41 are closely related.

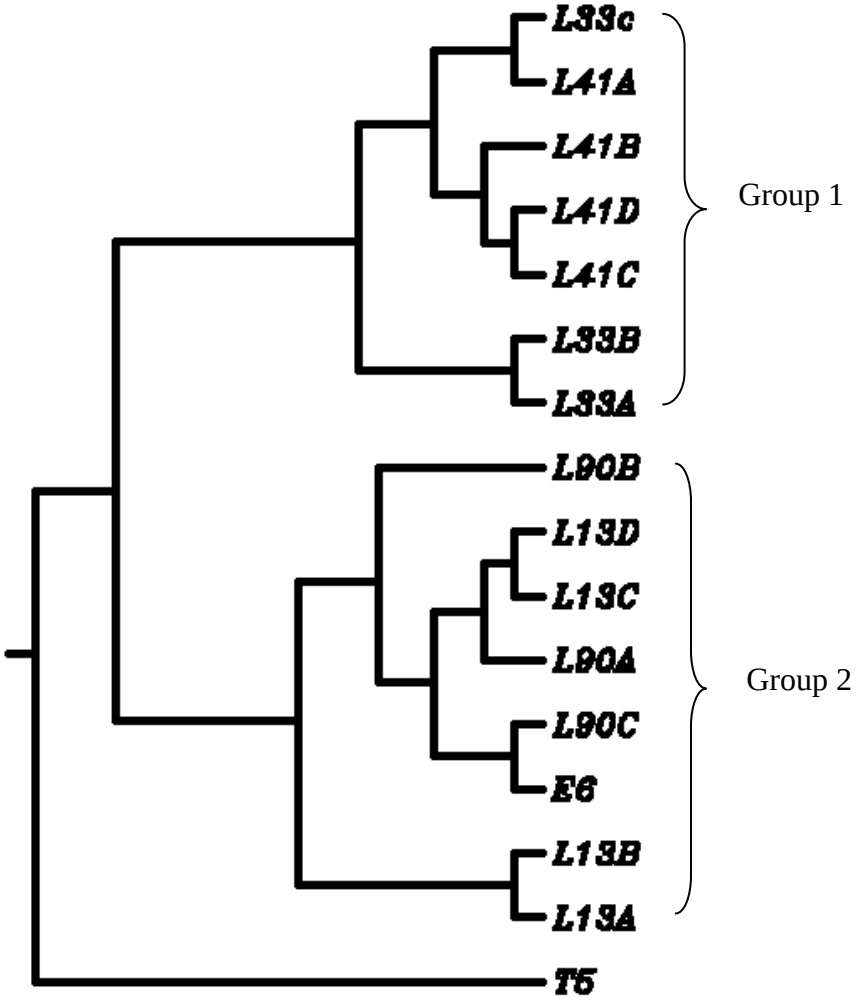

Figure 4.5(c): Phylogenetic tree of *Carica papaya* L. based on LTR 6149 primers.

The phylogenetic tree from Figure 4.5(c) is divided into two major groups. Each major groups is further divided into three subgroups making up a total of six subgroups overall. From the topology, parental sample T5 (outgroup) is closely related to samples L13A and L13B. Another parental sample E6 is closely related to samples L90C.

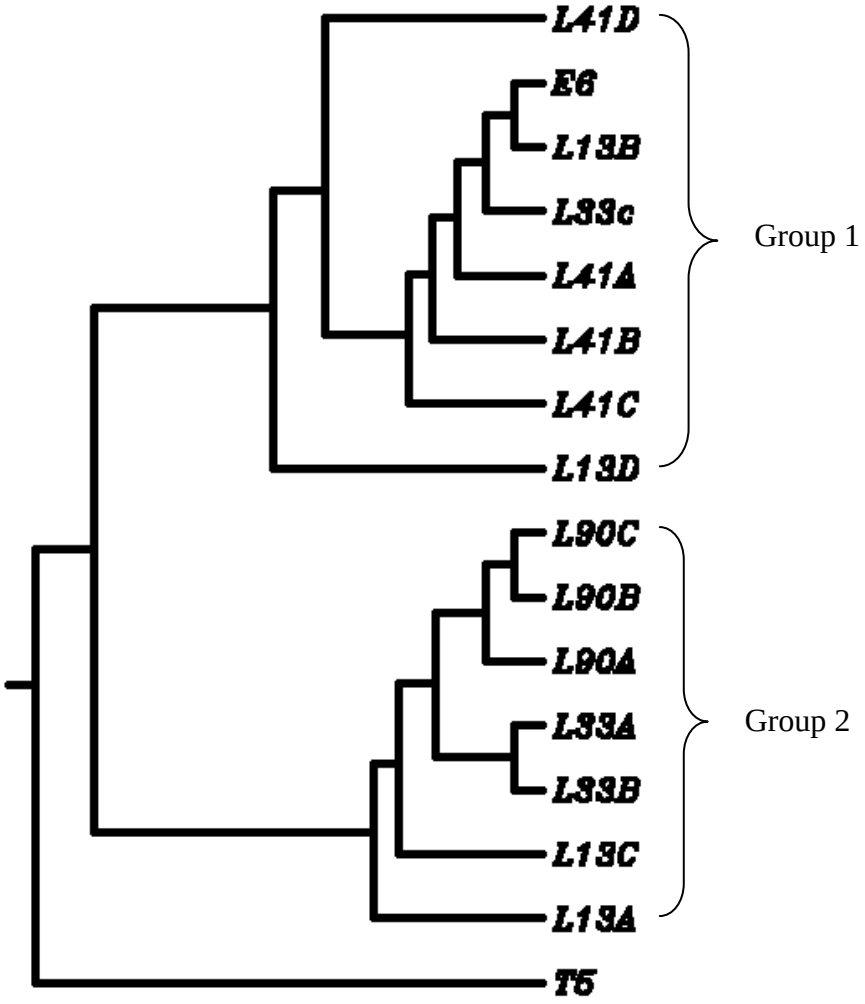

Figure 4.5(d): Phylogenetic tree of *Carica papaya* L. based on LTR 6150 and RT (reverse) primers.

Phylogenetic tree from Figure 4.5(d) splits into two major groups and three subgroups. Based on this, sample T5 is closely related to samples L13A and L13D. Another parental sample E6 is closely related to sample L13B.

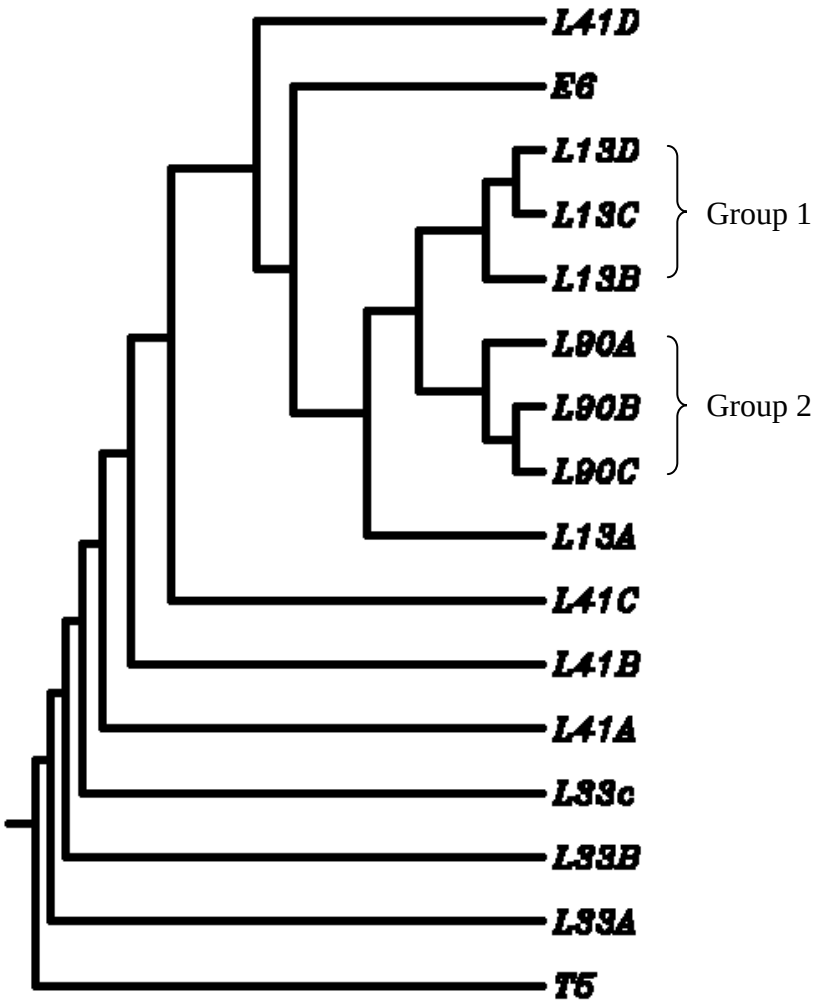

Figure 4.5(e): Phylogenetics tree of *Carica papaya* L. based on RT (forward and reverse) primers.

The phylogenetics tree in Figure 4.5(e) has two groups. Parental sample T5 is the outgroup and closely related to samples L33A. Sample E6 is closely related to sample L41D. The two groups comprise of samples from Line L90 and Line 13 each exclusively.

CHAPTER 5

DISCUSSION

5.1 Papaya DNA Extraction

There are several methods designed to extract plants DNA but the protocol of Doyle and Doyle (1990) involving CTAB, the DNA-binding detergent which is included in the isolation buffer, was found to be the best method for large scale extraction of papaya DNA. Large amounts of papaya DNA was obtained using this modification of the Saghai-Maroof *et al*. (1984) method. This protocol omitted cesium chloride (CsCl) / EtBr density gradient centrifugation step that is tedious and costly (Croy *et al*., 1993).

A modified CTAB procedure based on the protocols of Saghai-Maroof et al., Rogers and Bendich, and Doyle and Doyle is the method of choice for obtaining good-quality total DNA from many plant species and also from fungi. CTAB is a cationic detergent which solubilises membranes and forms a complex with DNA. After cell disruption and incubation with hot CTAB isolation buffer, proteins are extracted by chloroform-isoamyl alcohol, and the CTAB-DNA complex is precipitated with isopropanol. The DNA pellet resulting after centrifugation is washed, dried and redissolved. Depending on the species, additional purification steps may or may not be necessary in order to remove RNA, polysaccharides, polyphenols, and other contaminating substances (Weising *et al*. 1995).

No additional purification steps were taken following extraction using this method as the papaya DNA quality was found to be sufficient for IRAP-PCR. During extraction work, working efficiently and quickly was important to avoid any avoidable DNA degradation or contamination.

However, a variety of essential compounds which are protecting DNA from degradation are contained in the majority of extraction buffers. For example, EDTA (ethylenediamine tetraacetate) is generally included in DNA isolation buffers and storage solutions, since this compound chelates bivalent cations and thereby inhibits metal- dependant DNases. Reducing agents such as β-mercaptoethanol are also usually included to inhibit oxidization processes, which either directly or indirectly cause damage to DNA. Still another strategy to reduce the amount of undesired contaminants (and actually the original strategy of CTAB-based DNA isolation procedures) makes use of the fact that DNA-CTAB complexes are soluble in high salt only (Weising *et al*., 1995).

5.2 IRAP-PCR

The DNA quality is very important, as it is for most PCR-based methods. DNA purification with a spin-column containing a silica-gel membrane (such as Qiagen, www.qiagen.com/) is not a guarantee of high DNA quality for all plant samples or tissues. One sign of DNA contamination is that, after some period of time (a month or more) in storage, only short bands can be amplified. Careful casting of gels is also critical to success. Small, undissolved agarose inclusions in the gels will result in bands with spiked smears. Finally, a high-quality gel scanner with good sensitivity and resolution is also very important. Other still-video systems, which may be suitable for checking the success of restriction digests, cloning reactions or simple PCR reactions, are not suitable for analysis of complex banding patterns (Kalendar and Schulman 2006).

During samples preparation and mixing PCR ingredients, good pipetting technique was observed to avoid any contamination and to obtain exact required volume of PCR reactions.

Careful loading of PCR products in agarose gels was practised to avoid PCR products being over-spilled outside the wells or causing leakage at the bottom of the wells.

In order to ease DNA visualization for bands scoring and size determination purposes, agarose gel electrophoresis was carried out at a lower voltage and for a longer period of time. Furthermore, the electrophoresis was run until the end of the gel as indicated by the loading dye. This is so that IRAP-PCR bands will be well separated between each other.

Not all primers (derived from either retrotransposons or microsatellites) will work in the PCR. The genome may contain too few retrotransposon or microsatellite target sites, or they may be too dispersed for the generation of PCR products. Alternatively, sequence divergence in old retrotransposons or polymorphisms between heterologous primers and native elements may lead to poor amplification. Some primers generate smears under all PCR conditions. Many sources can contribute to this problem, ranging from primer structure to variability in the target site and competition from other target sites. Generally, it is more efficient to design another primer than try to identify the source of the problem. Furthermore, primers which produce a single, very strong band are not suitable for fingerprinting (Kalendar and Schulman 2006).

The primers for different retrotransposons can be combined in many ways to increase the number of polymorphic bands to be scored. Furthermore, the length and conservation of primers to the LTRs facilitate cloning of interesting marker bands and the development of new retrotransposons for markers (Kalendar and Schulman 2006).

However, one of the major drawbacks is when the linked marker used for selection is at a distance away from the gene of interest, leading to cross-overs between the marker and the gene. This produces a high percentage of false-positives/negatives in the screening process. Also, a marker developed for a gene in one cross may not be

useful in other crosses eventhough the same gene may be segregating in the second cross, unless the marker is from the gene itself (Mohan *et al*., 1997).

Among all 45 possible IRAP and RT primers combinations, only five primers combination successfully produced significant polymorphic banding pattern results.

Furthermore, the five primers combinations produced significant polymorphic banding patterns that discriminate between parental samples. This is important to determine parental-progeny relations and to classify progenies based on their resistance towards PRSV (resistant progenies to resistant parent and vice versa). Based on the IRAP-PCR results, the PCR bands are then scored and the data used to construct phylogenetic trees. Phylogenetic trees are constructed as they are able to depict parental-progenies relations of the samples and thus the pattern of inheritance.

These results obtained confirm the transferable nature of the retrotransposon-based marker system between genus (Teo *et al*., 2005) in this case between barley derived-retrotransposons and their application to the papaya genome.

All six universal primers amplified multiple fragments of defined sizes from papaya genomic DNA under the PCR conditions used, with different level of polymorphisms (Table 4.3). The number of bands and level of polymorphism shown through percentage of polymorphic bands may also reflect the copy number and distribution of LTR retrotransposons in papaya.

5.3 Phylogenetic Analyses

Dollop program carries out the Dollo and polymorphism parsimony methods. The method chosen in this study is polymorphism parsimony. These program and method were chosen because they generate fully-resolved phylogenetic trees, produce results quickly and are easy to use.

Other available method such as Distance is not suitable in this study because when applied, the distance between T5 and E6 would become infinity. This is as such because based on the resistance towards PRSV, samples T5 and E6 are at the opposite end of the tree as T5 is resistant and E6 is not. The interest of this study is to estimate the closeness of association of the progeny samples with their parents which then might reflect their resistance phenotype. The use of distance method is suitable when comparing the relationship of samples between different cultivars or species. The results could then be the basis of selection of progeny in future breeding experiments using these two lines.

In contrast to distance methods, discrete methods operate directly on the sequences, or on functions derived from the sequences, rather than on pairwise distances. Hence, they endeavour to avoid the loss of information that occurs when sequences are converted into distances (Page and Holmes, 1998).

5.4 Phylogenetic Tree Inferences

Among all five phylogenetic trees constructed, the phylogenetic tree constructed based on the bands scoring of primers combination LTR 6150 and Nikita [refer Figure 4.3(b)] was found to correlate with known groupings (papaya resistant progenies with resistant parent, and vice versa). From the phylogenetic tree in Figure 4.5(b), progenies from Lines 90 were found to be more closely related to PRSV resistant parent Tainung

(T5) while progenies from Lines 13 and Lines 33 were closely related to the non-resistant parent, Eksotika (E6).

This finding corroborates field trial findings by Chan *et al*. (2002). L90 with near symptom-free fruits was the most tolerant insofar as fruit symptoms were concerned. At 12 months, when fruit symptom expression was the most severe, L90 fruits had hardly any symptoms like ringspots or necrosis. L13 was found to have TSS and fruit weight closest to Eksotika and had disease tolerance only just below that of the three most tolerant lines. L41 and L90 would make useful parents for development of F1 hybrids in future breeding programs (Chan *et al.*, 2002).

The other phylogenetic trees did not show any classification between resistant and susceptible papayas. Therefore, the trees lineages do not follow exactly the findings of field trial results. Probably these primers are located far from the resistant gene embedded in the genome of Tainung 5 and hence did not follow its pattern of inheritance.

5.5 Results Analyses and Future Studies

To ease bands scoring and comparison, medium or large scale DNA gel electrophoresis systems should be used so that it can run more samples and the use of thicker combs to improve band resolution.

Whenever possible, samples to be compared to each other should be run in adjacent lanes. This is particularly important if high-precision comparisons are required (e.g. in paternity analyses). Large sample numbers, however, will make it necessary to determine band matching between lanes which are widely separated and even between lanes derived from different gels. In these cases, appropriate standards have to be included at least every four to five lanes on each gel. In most cases, molecular weight markers are used for this purpose. However, one of the investigated samples may also

serve as a standard, especially if it contains a set of invariable bands present presents in all individuals (Weising *et al*., 1995).

In plants, retrotransposons have been extremely successful as evident to their abundance (Kumar and Bennetzen, 1999). Their ubiquity in the plant kingdom suggests that they are of very ancient origin (Bennetzen, 2000). In addition, their abundance has played a major role in plant genome structure and evolution (Bennetzen, 2002). In this regard, the possibility that retroviruses might exist in plants had been discussed (Kumar, 1998; Kumar and Bennetzen 1999), but it is only very recently that plant genomes have been shown to contain retroviral-like sequences. In other words, the detection of *env*-like gene. It is noteworthy that the presence of an *env*-like gene that encodes a transmembrane protein is generally considered to be a predictor of a retroelement's infectious nature (Peterson-Burch *et al*., 2000)

Therefore, it is possible that retrotransposons type 1 originated from retroviruses. As mentioned earlier, the DNA sequence of LTR retrotransposons is the same as retroviruses except for the *env* gene. But recently, *env*-like gene has also been detected.

In retrospect, probably many years ago the incident that first gave rise to these retrotransposons was during a retrovirus infection, the plant's defence mechanism successfully split the *env* gene from the retrovirus sequence. Consequently, retrotransposons type 1 has become the remnants of the retroviruses and the *env*-like gene (defective or functional) being located at a different location in the plant's genome.

Based on that, copy number and distribution of retrotransposons type 1 could possibly be an indication of the time this event took place in plants and might also reflect the number of times these plants had undergone stress, diseases or attacks in the past.

The sequences of copies transposed in the distant past are likely to have been differentiated by non-directional and independent mutations. The degree of differentiation within a family of copies therefore should be proportional to the age of the individual transposition events (Tahara *et al*., 2004).

As mentioned earlier, retrotransposons are active during stress or wound. It could also be possible that during this time, the cell's tight regulation of retrotransposons activity is down, thus enabling retrotransposons to transpose into new sites and making new copies of itself. Cells' tight regulation is understandably to avoid retrotransposons causing gene inactivation by transposing themselves into important genes such as genes which are vital for the cell's survival.

Since it was also hypothesized that retrotransposons are activated by tissue culture, it could be suggested that retrotransposon activity is actually behind a phenomena known as somaclonal variation.

Somaclonal variation is used to describe the occurrence of genetic variants derived from in vitro procedures; it is also called tissue or culture-induced variation (Soniya *et al*., 2001). Such variation arises in tissue culture as manifestation of epigenetic influence or a change in the genome of differentiating vegetative cells induced by tissue culture and is expected to generate stable plants carrying interesting heritable traits (Soniya *et al*., 2001). Four critical variables for somaclonal variation: genotype explants origin, cultivation period and the cultural condition in which the culture is made (Sheidai *et al*., 2008).

It has also been suggested that the activation of elements may occur in the widespread and successful opportunistic asexually reproducing (apomictic) dandelion (Taraxacum officinale; Richards, 1989). It is also possible that some of the effects described as 'somaclonal variation' in species as diverse as oil palm and tomato may be

due to activation and insertion of retroelements into genes (Heslop-Harrison *et al*., 1997).

If this proved to be true, therefore, it could be speculated that IRAP markers would also be an ideal marker to study somaclonal variation among different explants compared to other markers as the variation occurred is probably retrotransposons induced.

More importantly, tissue culture and the phenomena of somaclonal variation (whereby suspected that retrotransposon activity is causing them) could be exploited to activate retrotransposons and increase the possibility of insertions (retrotransposition events) into strategic new sites (near resistance genes). Perhaps even better, they might also introduce new resistance genes in the plants.

Among the reasons why retrotransposon was chosen as a molecular marker in this study is the possibility of retrotransposon to transpose itself near resistance genes. IRAP-PCR products might actually contain the resistance genes being much searched for. Based on Figure 2.9 about IRAP scheme, it is clear that IRAP-PCR products are sequences amplified between retrotransposons. Therefore, if the retrotransposons are located near the resistance gene, (based on banding pattern which is similar to resistant plant banding pattern) there is a high probability that cloning and sequencing those very bands might also reveal the resistance gene itself.

It can be deduced that the orientation and sequence of primers **Nikita** and **LTR 6150** primers are strategically located near to PRSV resistant gene in the papaya genome which are being amplified through IRAP-PCR. This in turn caused Nikita and LTR 6150 DNA banding pattern to classify papaya genotypes according to their resistance towards PRSV. Therefore, there is a similarity between the results obtained in IRAP molecular marker studies based on phylogenetic tree groupings with field trial analysis of papaya parents and hybrids in MARDI breeding program.

To increase the chances of IRAP markers being near to the gene of interest (PRSV resistance gene), more papaya-derived IRAP markers should be designed and tested. This is because the nearer the IRAP markers to the resistant gene, the more accurate the IRAP banding patterns which follows papayas' resistance trait towards PRSV banding patterns.

Another aspect of the results is eventhough barley-derived, the IRAP markers used in this study have proven to be applicable, useful and informative to papaya for selection of tolerant papayas in a Malaysian breeding program. This in turn could expedite the time taken to evaluate candidates for future studies (crosses) and to facilitate farmers in selection such as molecular assisted breeding and molecular assisted selection. The same protocols and primers could be used to develop a similar screen for future breeding lines.

From this study, it is found that universal TY1-*copia* Reverse Transcriptase (RT) primers (reverse and forward primers) could also be used together with other IRAP primers. Due to its location and orientation, the cloning and sequencing of the bands produced from PCR using RT with IRAP primers would enable new papaya-specific LTR sequence to be obtained. Once papaya LTR sequences are obtained, new papaya-derived IRAP primers could then be designed and tested.

The use of RT primers matching highly conserved domains (Flavell *et al*., 1992, Voytas *et al*., 1992, Teo *et al*., 2002) allows the detection of insertions into virtually any *copia*-like retrotransposon. A combination of LTR and RT primers can be used to track nested insertions events whereby a retrotransposons is integrated into the LTR or internal region of another retrotranspsons. Such nested patterns of insertions are common at least in barley and maize (Shirasu *et al*., 2000; San Miguel *et al.*1996). Methods have been developed for rapidly isolating native retrotransposons based on conserved domains within retrotransposons (Pearce *et al*., 1999).

Furthermore, with the availability of papaya whole genome sequence database and papaya repeat database such as CPR-DB, retrotransposons native LTR ends can be analysed and compared. Using these databases, IRAP primers can be designed through the native LTR ends.

For example, CPR-DB is a database of papaya genome repeats which was created to shed light on papaya genome organization and specifically on the role of repetitive elements. CPR-DB is divided into three main categories: transposable elements (TEs), tandem repeats and high copy number genes (Nagarajan *et al*., 2008).

Apart from that, another strategy in designing papaya-derived IRAP markers is by cloning and sequencing full-length Ty1-*copia* / Ty3-*gypsy* retrotransposons sequences. Nonetheless, this method would depend on the activation of papaya's retrotransposons. Among the strategies to induce retrotransposon's activity is through tissue culture.

As mentioned earlier, retrotransposons are generally thought to be activated by tissue culture. The *Tos10, Tos17* and *Tos 19* elements of rice (Hirochika *et al*., 1996) are so activated; *Tos 17* is in fact only active in tissue culture, its copy number of *Tos17* increasing with prolonged culture. The copy number of retrotransposon *Tto1*, normally an inactive element, can increase 10-fold during tissue culture (Hirochicka, 1993).

Tissue culture-induced mutations have been reported in many plant species and seem to be ubiquitous (Hirochika *et al*., 1996). The first active retrotransposon to be identified in plants, Tnt-1 of tobacco, was isolated by virtue of its mutagenicity in tissue culture Grandbastien *et al*., 1989), a property of retrotransposons reported also for rice (Hirochika *et al*., 1996).

As such, tissue culture could be included in the experimental design of obtaining full-length retrotransposons since it increases retrotransposons activity and thus provides a means of isolating newly inserted, active members for further analysis.

Together with that, IRAP could also be used to track the integration of new retrotransposons into the genome of papayas during tissue culture. For future studies, more papaya samples from other genotypes could also be tested in to further compare marker results with field trial results.

Since Tainung 5 is resistant towards infection by PRSV, it is obvious that Tainung 5 possesses natural PRSV resistant gene as compared to the transgenic papaya which depends on PRSV coat protein gene to develop resistance. Thus, it is suggested that this resistant gene be identified and characterized. Besides utilizing the resistant gene for producing resistant papayas, these genes could also be used as marker gene to detect papaya resistance together with other markers. And now with the availability of papaya genome database, other resistant genes that might be present in other papaya wild varieties could be mined, aligned and compared.

Indeed this will also be a preparatory measure if in some unfortunate circumstances that papaya faces different strains of PRSV in the future or the current PRSV strain mutates and cause the current resistant papayas susceptible. This is in view of the difference in pathogenic strains of PRSV circulating across the regions.

As one of genetics major discoveries, genetic transformation of plants with PRSV code protein has successfully enabled papayas 'cross protection.' And with that, there are numerous explanations of the resistant mechanism which produced this amazing result. For the sake of general discussion, I would humbly speculate that the transgenic resistance plant (tobacco, papaya, etc) which contain the virus' coat protein gene would cause the expression of the coat protein gene both from the plant's DNA and from the viruses.

This in turn would cause coat protein competition between the protein expressed from the plant and from the virus DNA to wrap up or package the virus' genetic material (post-translational template competition or in other words, protein to

protein competition to bind).Thus, the failure to wrap up the virus' genetic material due to this competition would consequently result in the failure of the whole infection altogether and deemed the plant to be resistant altogether.

Other applications remain for the most part in the domain of research. This is particularly the case of selection of quantitative traits for which the main principles have been defined, but for which few experimental results are so far available. Nevertheless, markers have already indisputably enriched the range of methods available to the breeder in managing and exploiting genetic variability (Young, 2000; Dekkers and Hospital, 2002; de Vienne 2003).

Nonetheless, the selection of resistant papayas based on molecular markers is without consideration of yield and quality. Tolerance or resistance to disease per se in fruit crops would be rather meaningless without consideration of yield and quality. This is more so in the present case where the genotypes were derived from single seed descent without prior selection pressure on yield and quality (Chan et. al 2004).

Most lines seemed to bear quite well under PRSV infected conditions. Their yield ranged from 29.5 to 41.7 kg / tree, which is quite comparable to 35-60 kg / tree previously reported for the tolerant PRSV tolerant 'Sinta' hybrid (Villegas *et al*., 1996) (Chan *et al*., 2004). Yield and quality might be affected even though the plant is able to resist virus infection.

CHAPTER 6

CONCLUSION

In this study, 45 combinations of IRAP primers were screened using a test set of six papaya genotypes from a domestic papaya breeding programme. Five combinations which produced polymorphic banding pattern were scored, phylogenetic trees constructed and population segregation/pattern of inheritance were studied. IRAP markers which are able to discriminate for plants that were tolerant towards papaya ringspot virus (PRSV) infection were identified through comparison with field trial results.

Here, significant polymorphic bands are defined as bands which can discriminate parental samples. This is to enable clustering progenies according to their closest parent.

The repetitive, dispersed nature of many long terminal repeat (LTR)-retrotransposons families has been successfully exploited for classifying progenies in *Carica papaya* L. hybrids. The elements' insertional polymorphism was studied with seven published primers facing outward from the LTRs and reverse transcriptase (RT) domain of the retrotransposon and generated specific amplification patterns showing the universal applicability of this marker type.

Plant retrotransposons have thus great potential as useful tools for detecting genetic diversity (Kumar *et al*., 1997). In the final analysis, the success will depend on identifying marker(s) as close to the gene as possible for its utility across all publications (Mohan *et al*., 1997). Most commontypes of TEs are represented in the papaya genome with retrotransposons being the dominant class, accounting for 40% of the genome. The most prevalent retrotransposons are Ty3-gypsy (27.8%) and Ty1-copia (5.5%) (Nagarajan *et al*., 2008).

The PRSV resistance trait is heritable and detectable using IRAP markers by identifying patterns of inheritance in papaya. IRAP is an ideal marker for papaya especially in detecting resistant lines. This is probably due to its location being near to the resistance genes.

For dominant retrotransposons markers, the absence of an amplicon may be the consequence of mutation at the locus carrying the insertion. The mutation could affect the binding site for the retrotransposon primer or, in SSAP, led to the gain or loss of a restriction site. In practise, this problem does not arise in the application of retrotransposon markers to segregating populations generated by deliberate crosses, because the allesles are determined by the parents (Kalendar *et al*., 2011).

All primers that produced significant polymorphic banding patterns are good and useful but the best primers combination is 6150 and Nikita in term of analysis of its banding pattern which showed groupings in accordance with known relationship.

The association of bands with particular genome types can be explained as the result of the integration of new retrotransposon copies after divergence of the ancestral genomes. The power of IRAP to identify the genome type was confirmed by comparing this study results with genome classification of local papayas observed from field trial.

Phylogenetic trees analysis of Malaysian and Taiwanese papaya cultivars and their hybrids (six genotypes) using IRAP markers revealed information about resistance towards papaya ringspot virus's infection based on their groupings. In applied work, this method which is based on PCR-based assay, is amenable to the large-scale throughput demands of screening breeding populations and is applicable to any crop.

With regard to the PRSV resistant gene, we can deduce that Tainung 5 is a heterozygous dominant. This is as a result of the cross with susceptible Eksotika 6 which also produced susceptible progenies beside resistant progenies (based on field

trial results). Therefore, Eksotika 6 can also be deduced as being homozygous recessive parent based on the same charateristic.

Retrotransposons markers are transferable from one crop to another and in particular, barley-derived sequences being used for identification and classification of papaya cultivars in papaya breeding program. The products obtained with two primers do not represent the simple sum of the products obtained with the primers individually.

Retrotransposons-based marker such as IRAP has a high chance of being located near to virus resistant gene. This is because amplification of retroelements in some regions might occur not through reverse transcription and re-insertion but through replicative mechanisms (also involved with other non-transposable repetitive sequences) such as unequal crossing over, replication slippage or perhaps even transposition through DNA intermediates and preferential insertion into linked sites (Heslop-Harrison *et al*., 1997).

Indeed, the promoters of retrotransposon expression are wound-and stress-inducible (Mhiri *et al*., 1997). The genomic divergence following retrotransposon activity may lead to speciation because meiotic pairing would be disturbed in hybrids if the insertions occurred in regions of the genome critical for chromosome alignment (Heslop-Harrison *et al*., 1997). With respect to target site preferences for retrotransposon insertion, there is evidence that at least some sub-families show preferential insertion in some genomic regions.

The copy number of these retrotransposons rapidly increases due to their replication mode of transposition, which in turn may cause host genome expansion and evolution (Dixit *et al*., 2005).

Recent developments in DNA marker technology together with the concept of marker-assisted selection provide new solutions for selecting and maintaining desirable genotype. Once molecular markers closely linked to desirable traits are identified,

marker assisted selection can be performed in early segregating populations and at early stages of plant development (Mohan *et al*., 1997).

Conventional methods for the selection of resistant papaya lines are time-consuming. Procedures involving the use of molecular markers and the selection of resistant plants can reduce this time considerably.

Of the various techniques available, IRAP detects high levels of polymorphism and have the advantages over other techniques such as no DNA digestions, ligations or hybridization are needed to generate the marker data, thus increasing the reliability and robustness of the assay. The LTRs of retrotransposons contain sequences that are essential for expression (promoter and processing signals) and integration. From the conserved LTR regions, primers for PCR amplification are designed to amplify the DNA between the closely dispersed members of a retrotransposons family. The use of RT primers matching highly conserved domains allows the detection of insertions into virtually any copia-like retrotransposon (Teo *et al*., 2005).

IRAP primers LTR 6150 and Nikita are the closest linked to desirable trait identified which is resistance towards PRSV. It will also be ideal if future markers designed are unique to one of the parents involved so that the banding patterns produced would definitely discriminate between parental samples and therefore ease bands scoring analysis.

For future studies, it is recommended that primers LTR 6150 and Nikita be tested for use in Molecular Assisted Breeding (MAB) and Molecular Assisted Selection (MAS). The availability of data-mining programs such as LTR_STRUC which is a novel search and identification program for LTR retrotransposons would definitely facilitate the design of such powerful markers.

Besides that, the conventional method of searching for LTR retrotransposons in papaya genome databases also will allow researchers to compare, align and analyse

different LTR retrotransposons sequences. This will contribute towards the designing of more IRAP markers.

In general, if one wishes to study closely related varieties or breeding lines, one should develop a retrotransposon marker system based upon the most polymorphic TE available. This process begins with amplification and sequencing of variable regions close to the outer termini of the TE, development of primers specific for the retrotransposon families found and testing these for their efficacy as markers *(*Pearce *et al.,* 1999, Jing *et. al,* 2005, Kalendar *et al.,* 2010).

It maybe necessary to clone and sequence hundreds of clones to obtain a few good primers. In practise, up to one person–year of time is needed to develop and apply a fully functioning novel retrotransposon-based marker system in a new species. However, this is a one-time investment that can be appied thereafter to the corresponding species and its close relative (Kalendar *et al*., 2011).

The most polymorphic retrotransposons are likely to include those that are currently active. To identify these, one could amplify and sequence unconserved regions between conserved domains (for example, within the integrase or reverse transcriptase domains) in RNA, and then use a primer from the unconserved region and an adapter primer in a genome-walking approach to isolate the corresponding LTR (Kalendar *et al*., 2010)

Theoretically, bands produced from PCR using IRAP primers and RT primers would enable full-length sequence of LTR retrotransposons to be obtained without using technique which requires the activation of LTR retrotransposons (transcription). Based on the regions amplified, RT primers are positioned in the middle of one LTR retrotransposons while IRAP primers are positioned at the end of another LTR retrotransposons (5’ or 3’ LTR ends), therefore any bands produced would definitely

include the LTR sequence of the first retrotranspososon (which also contain RT sequences).

Focus should also be given towards using LTR sequences from Ty3-*gypsy* retrotransposons as Ty3-*gypsy* is found to be more abundance in papaya genome compared to Ty1-*copia* retrotransposons. Therefore, the complexity and polymorphism of IRAP markers based on Ty3-*gypsy* retrotransposons are suspected to be higher. Furthermore, the probability of Ty3-*gypsy* retrotransposons being near to resistance genes in the papaya genome would definitely be higher compared to Ty1-*copia*.

Close analysis of larger members of genome-specific bands enabled the relationships of the individual genomes to be ascertained. This will help breeders to clearly identify and screen the genotypes with better products value and manage the genotype resources of papaya that some misleading by the synonymous names. The practical implication is that retrotransposon-based primers may be applied directly across divergent genera in orphan crops.

The effectiveness of IRAP markers in detecting resistant plants would enable breeders to save time, money and effort in selecting resistant plants and speed up the process of obtaining papaya gene pools of resistance traits in the long run. This is because conventional methods for the selection of papaya resistant lines are time-consuming.

In conclusion, this study has demonstrated that the use IRAP markers is an effective procedure for developing papaya resistant lines in that genotyping and selection can be carried out in early generations on those individuals bearing the desired resistant trait. Primers LTR 6150 and Nikita are suggested to be further tested with other papaya lines (genotypes) and applied in Molecular Assisted Breeding (MAB) or Molecular Assisted Selection (MAS) for papaya breeding programs.

The results of many studies have indicated that MAS may be more efficient, accurate and simpler strategy for breeding selection than selection based only on phenotype (Kwon *et al*., 2001). The former is mainly effective for the selection of early generations and traits that could overcome the difficulty of phenotypic detection (Oliveira *et al*., 2010)

Hopefully, these IRAP markers would prove to be useful in improving the papaya industry in Malaysia, *Insya*Allah.

APPENDIX A

Instruments and Apparatuses

All the instruments and apparatuses required for the study are listed below:

Centrifuges:

1) Centrifuge 5417R and 5415D (Eppendorf, Germany)
2) E-centrifuge (Wealter, Taiwan)
3) Mini Spin Plus Centrifuge (Eppendorf, Germany)

Electrophoresis apparatuses:

1) Power PAC MP-250 (Major Science, Taiwan)
2) Power PAC 300 (Bio-Rad, USA)
3) Shorter Mini Horizontal Gel Electrophoresis System (Major Science, Taiwan)

Other equipments and apparatuses:

1) AlphaImager 2200 Gel Documentation System (Alpha Innotech, USA)
2) Water Purification System Arium 611 (Sartorius, UK)
3) Autoclave machine (TOMY, Japan)
4) Balances or weighers
5) Beakers (100 mL, 500 mL, 1000 mL)
6) BioPhotometer (Eppendorf, Germany)
7) Centrifuge tubes, round bottom (Nalgene, USA)
8) Desiccators
9) Drying oven
10) Bottles (100 mL, 250 mL, 500 mL, 1000 mL) (Schott Duran, UK)
11) -20°C freezer

12) -80°C freezer

13) Gloves

14) Ice maker

15) Liquid nitrogen and tank

16) Magnetic stirrer

17) Mastercycler 5330 (Eppendorf, Germany)

18) Microcentrifuge tubes (0.5 mL and 1.5 mL)

19) Microwave oven

20) 0.2 mL PCR tubes

21) pH metre

22) Pipettes (Eppendorf, Germany)

23) Pipette tips (blue, yellow and white)

24) Speed vac

25) 15 mL and 50 mL tubes (Falcon, Becton Dickinson, USA)

26) Universal bottles (30 mL)

27) Vortex mixer

28) Wash bottles

29) Water distiller

30) Mortar and pestle

31) PCR racks

32) PCR reaction kit and 6x loading dye

APPENDIX B

Formulation for Solutions

70% ethanol (100 mL)

- 70 mL ethanol
- 30 mL dH_2O

5X Tris-Borate (TBE) Buffer (1L)

- 54 g Tris base
- 27.5 Boric acid
- 20 mL 0.5 M EDTA (pH 8.0)
- Add dH_20 to final volume

APPENDIX C

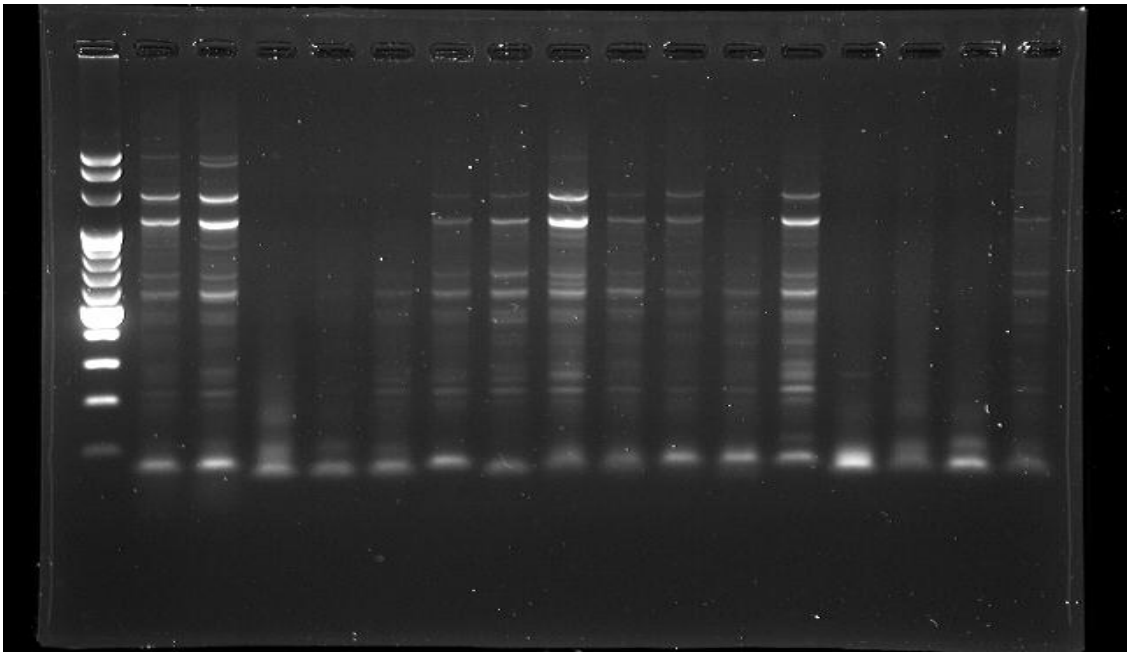

Figure 6.0: IRAP-PCR with 5’LTR1 and 3’LTR primers.

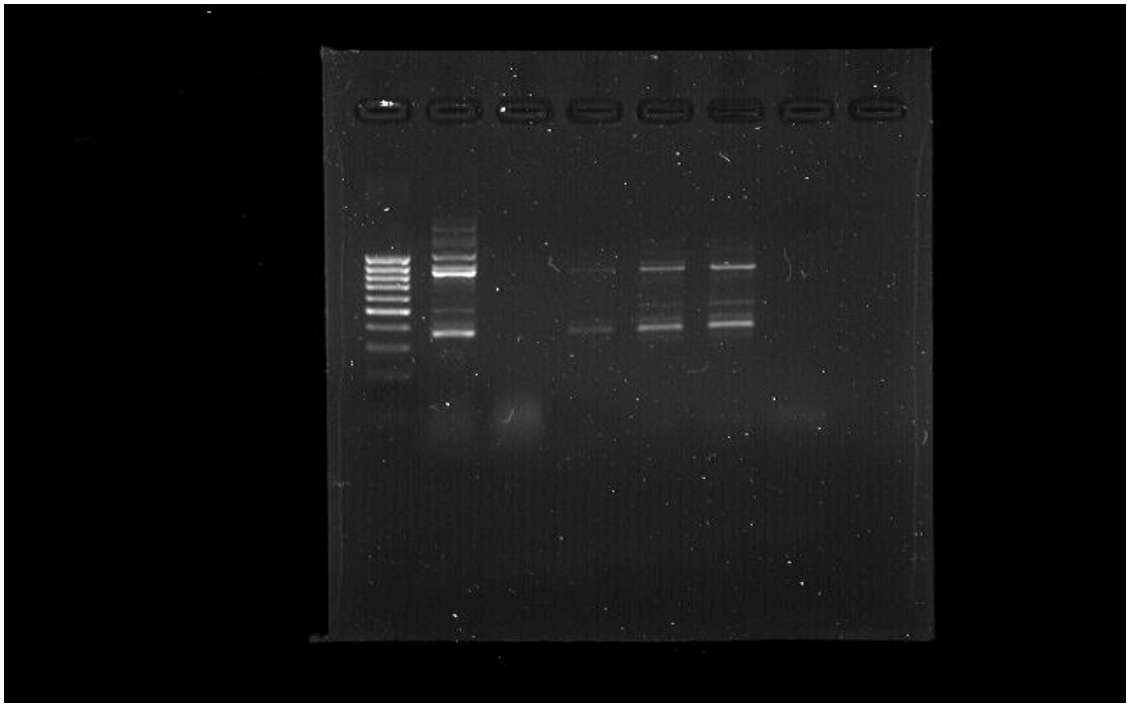

Figure 6.1: IRAP-PCR with 5’LTR2 and Sukkula LTR primers.

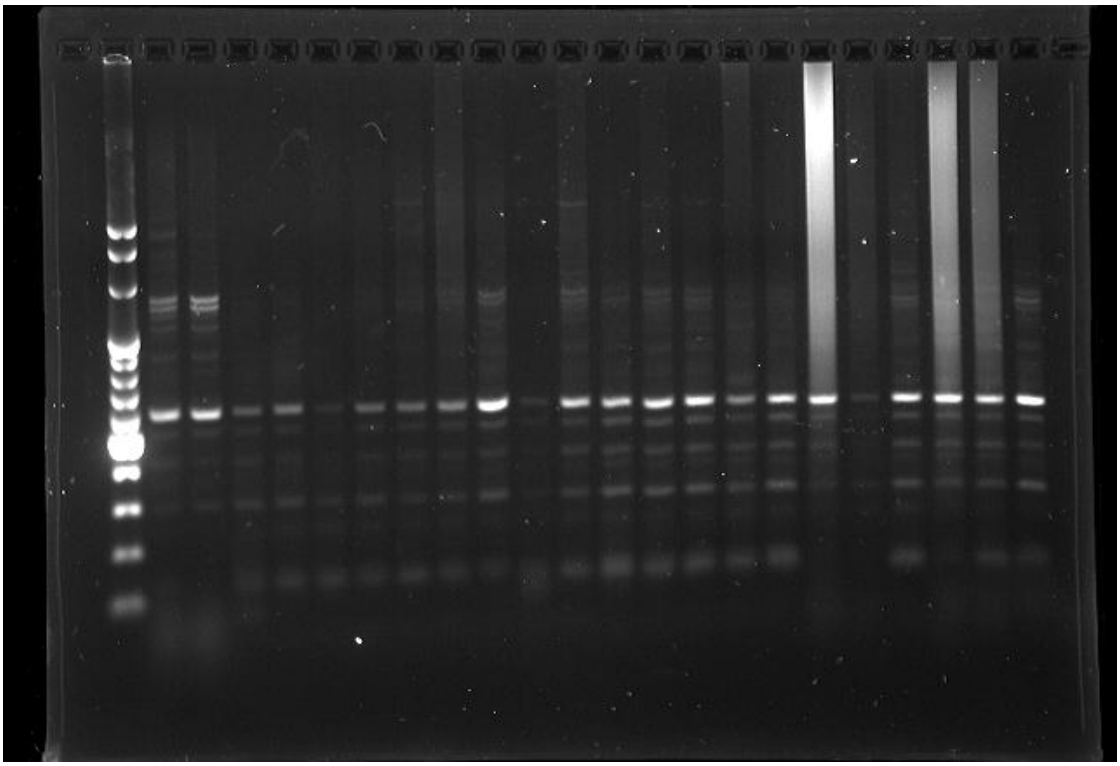

Figure 6.2: IRAP-PCR with 5’LTR2 primer.

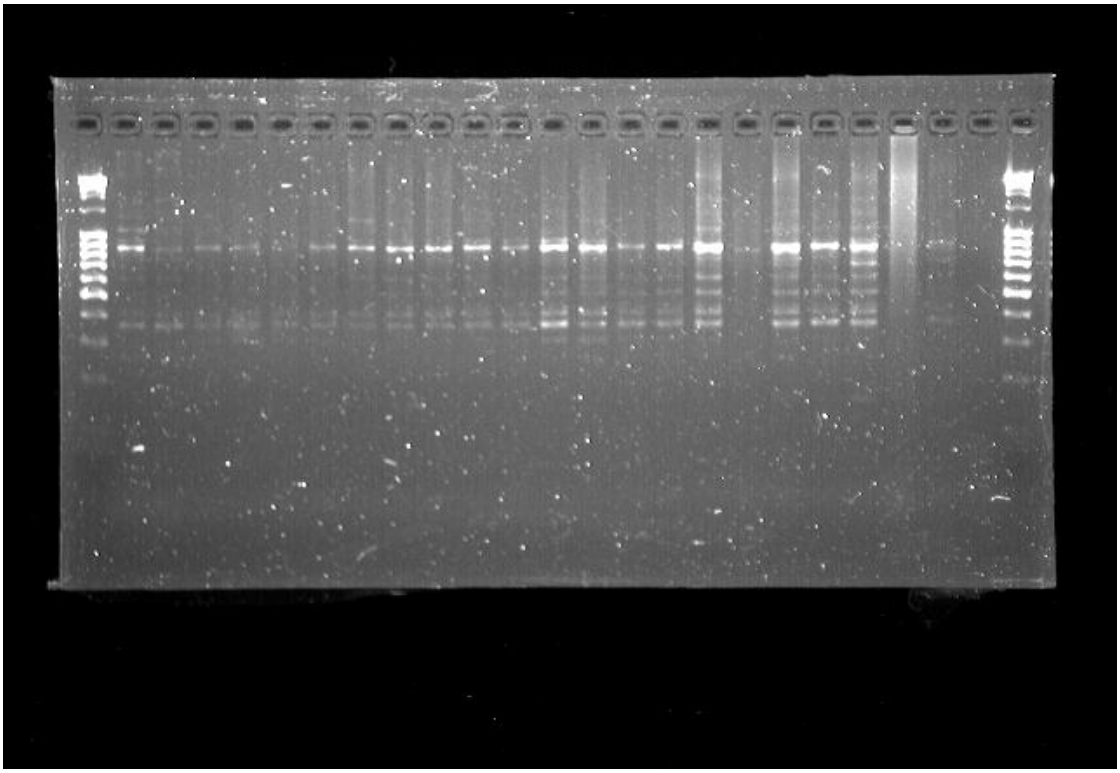

Figure 6.3: IRAP-PCR with 5’LTR2 and Sukkula LTR primers.

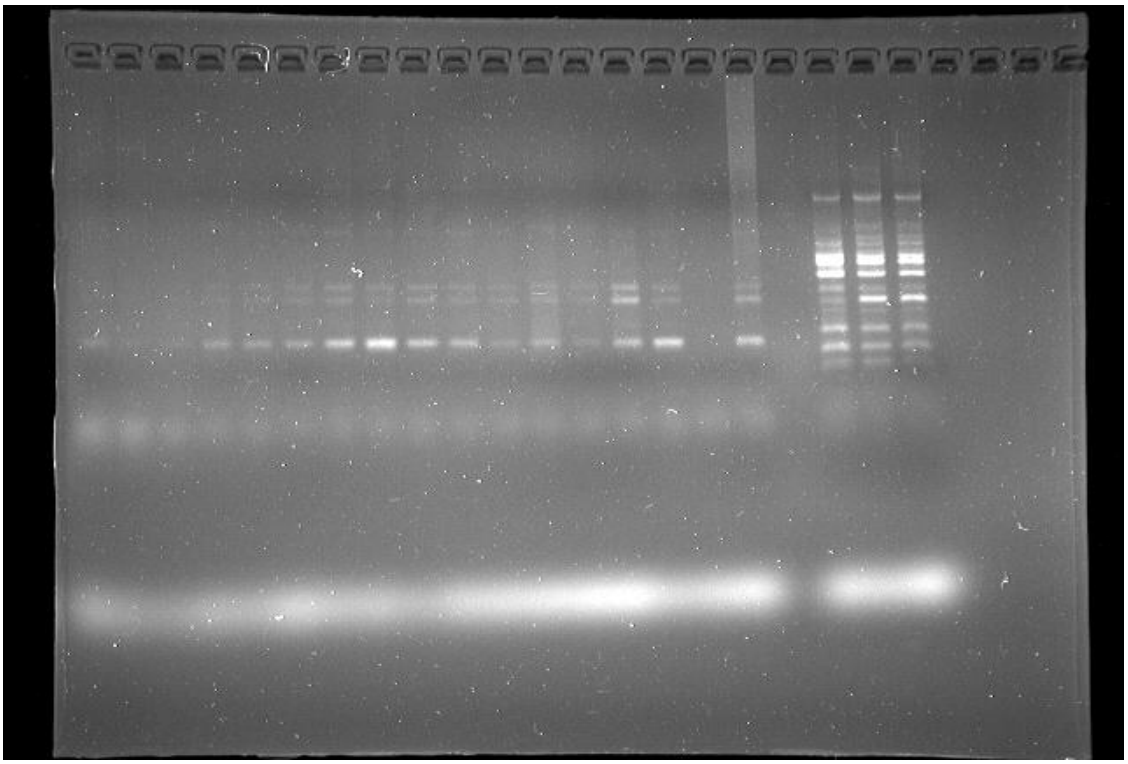

Figure 6.4: IRAP-PCR with 5’LTR primer.

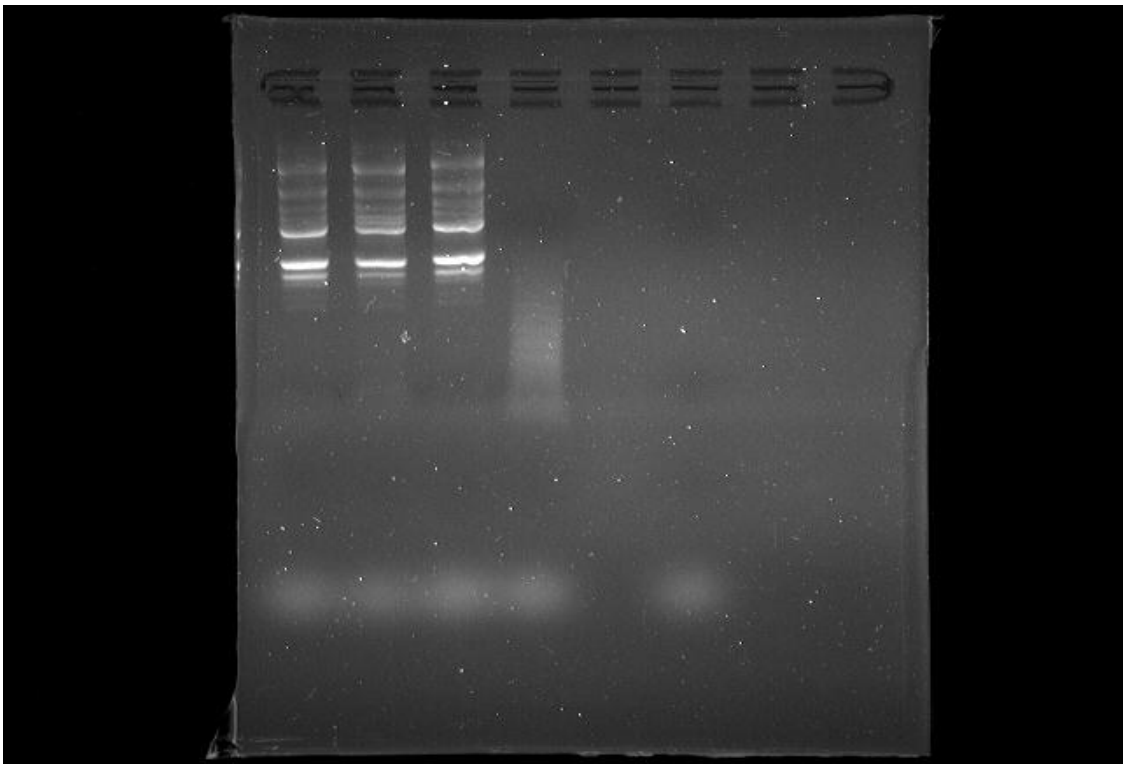

Figure 6.5: IRAP-PCR with LTR 6149 and LTR 6150 primers.

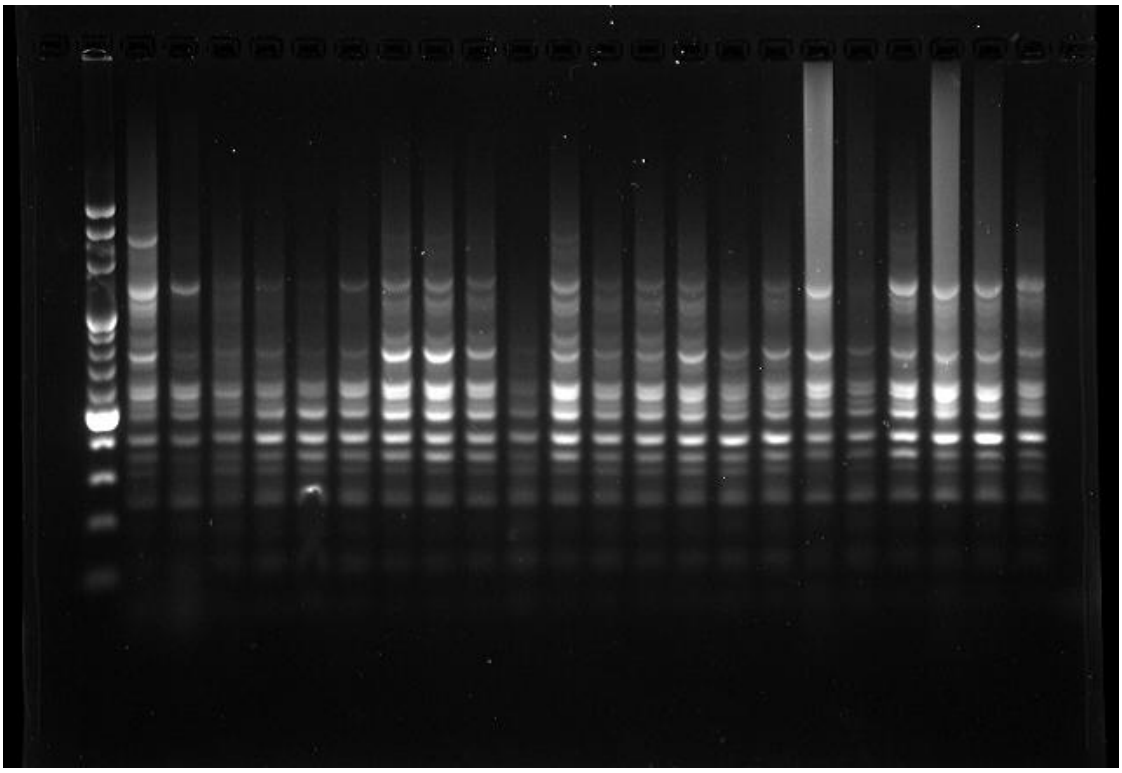

Figure 6.6: IRAP-PCR with RT (forward and reverse) primers.

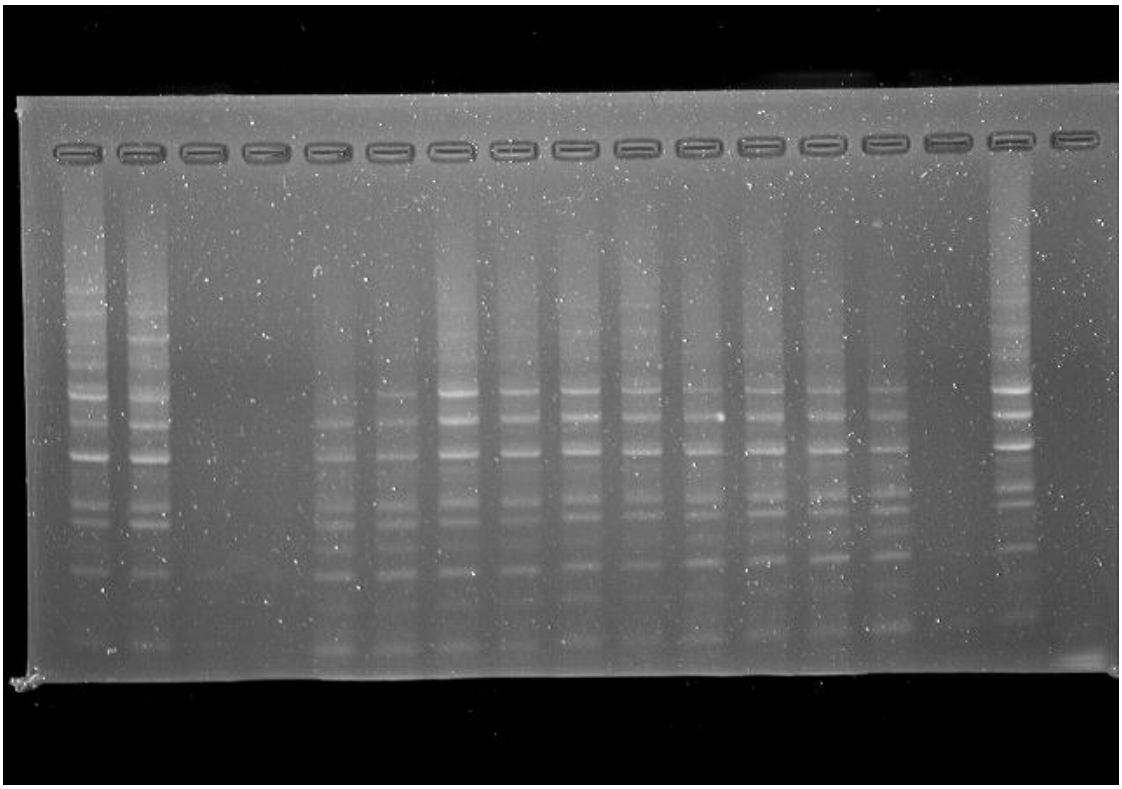

Figure 6.7: IRAP-PCR with 5’LTR2 and Sukkula LTR primers.

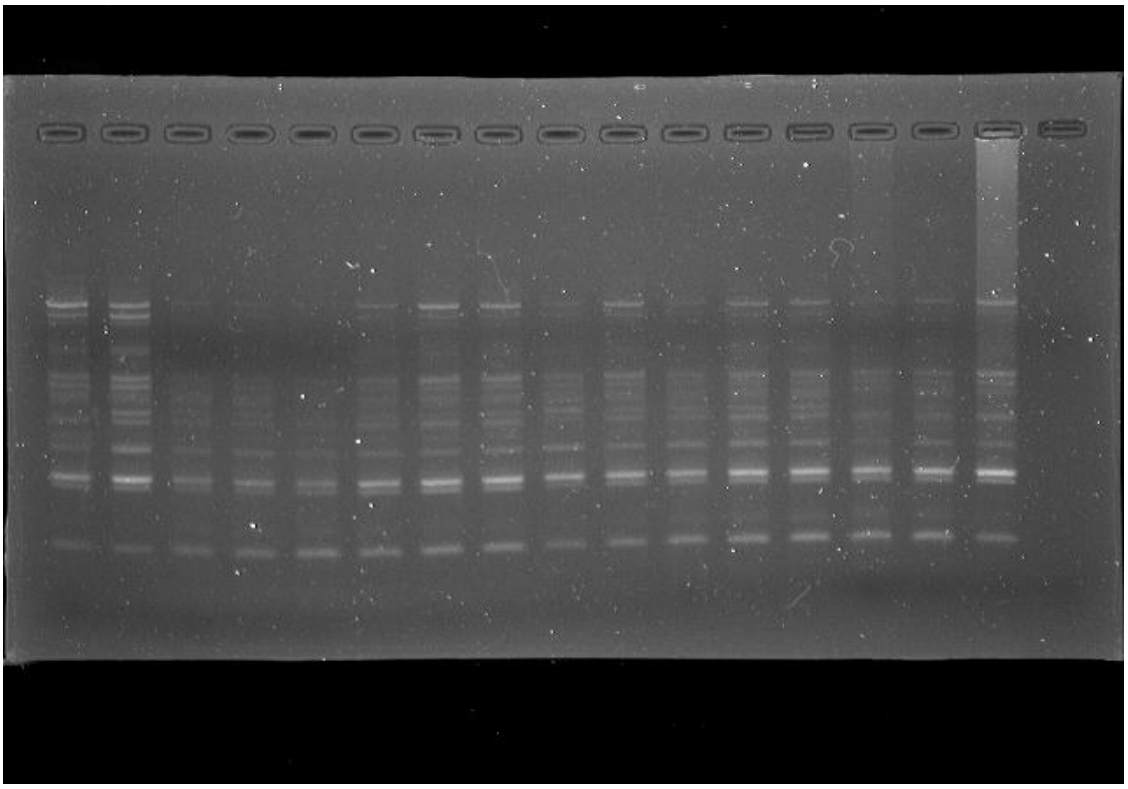

Figure 6.8: IRAP-PCR with 5’LTR1 primers.

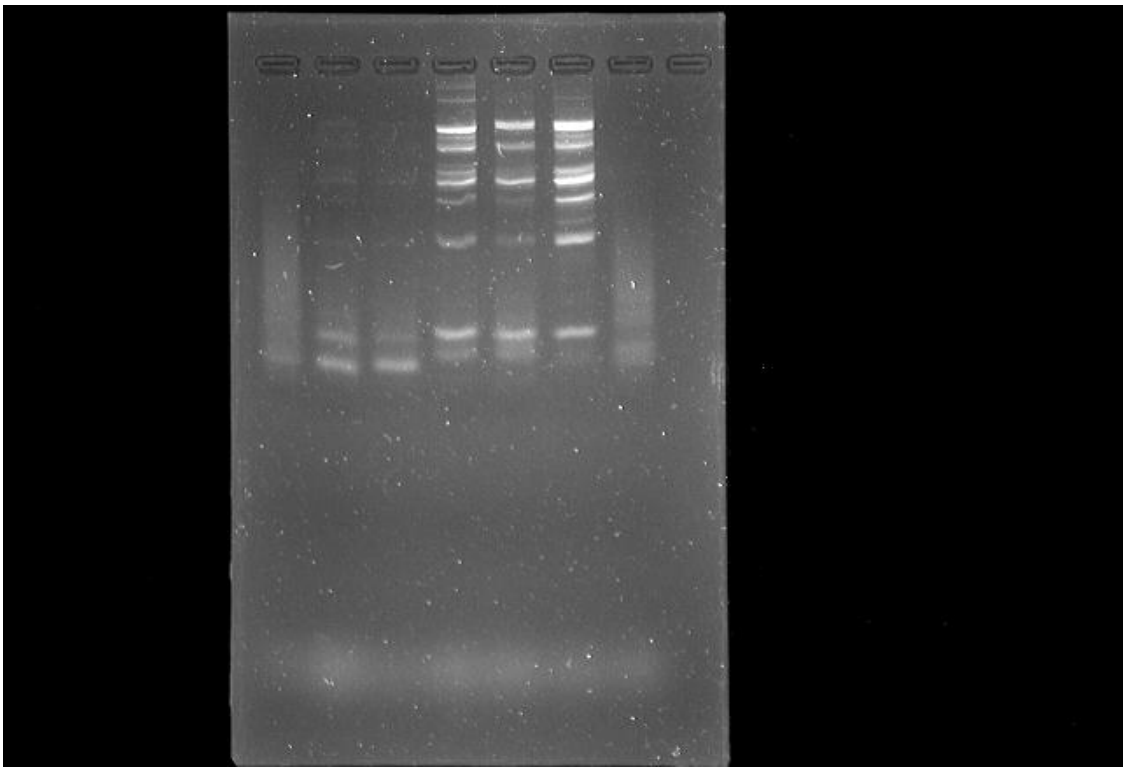

Figure 6.9: IRAP-PCR with LTR 6149 and Nikita LTR primers.

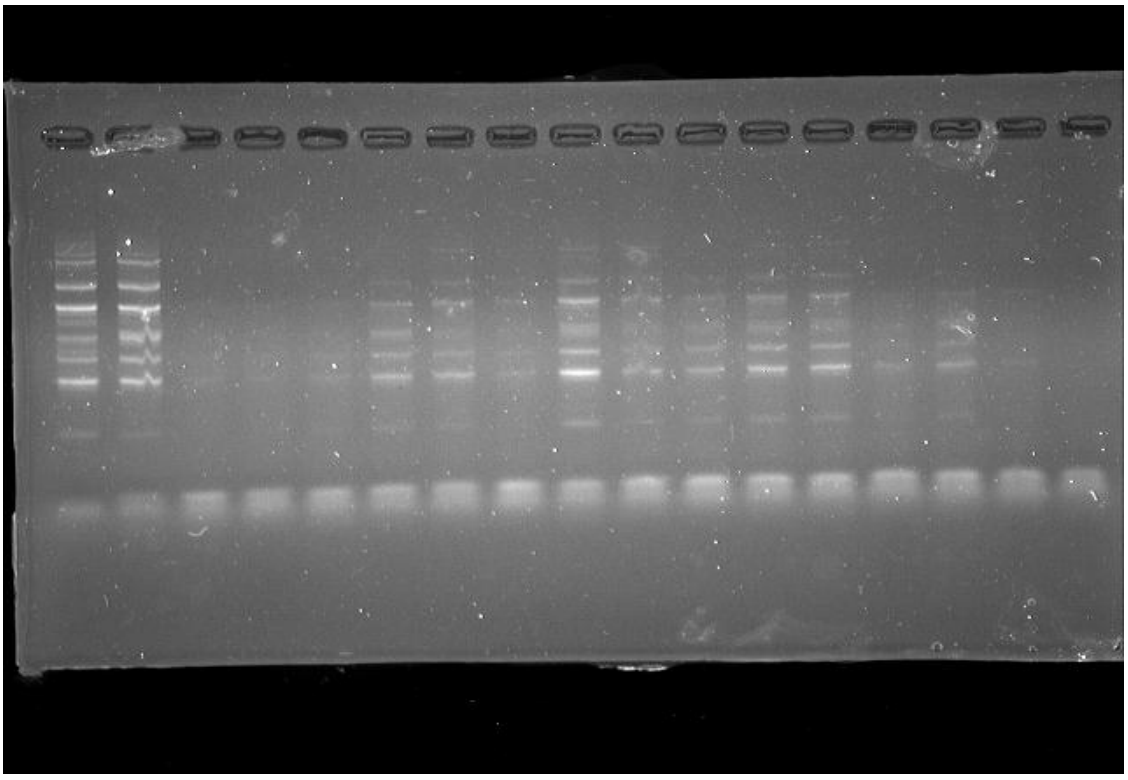

Figure 7.0: IRAP-PCR with 5’LTR2 and 3’LTR primers.

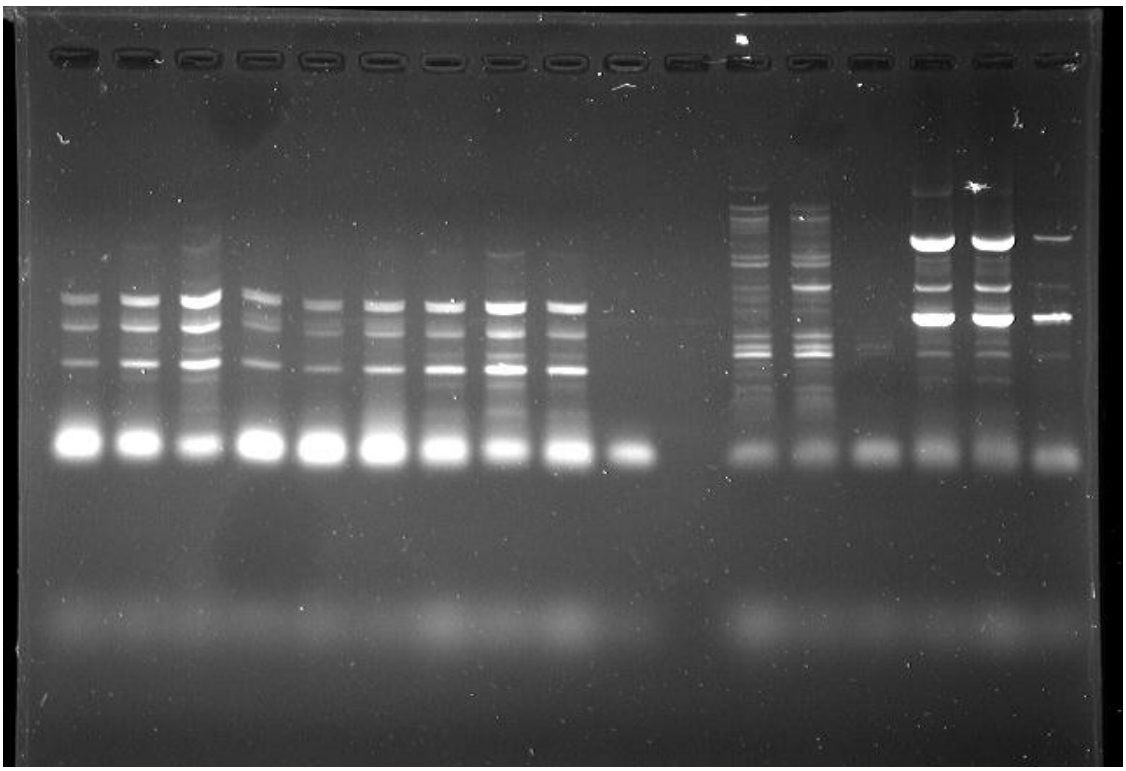

Figure 7.1: IRAP-PCR with LTR 6150 and RT (forward) primers.

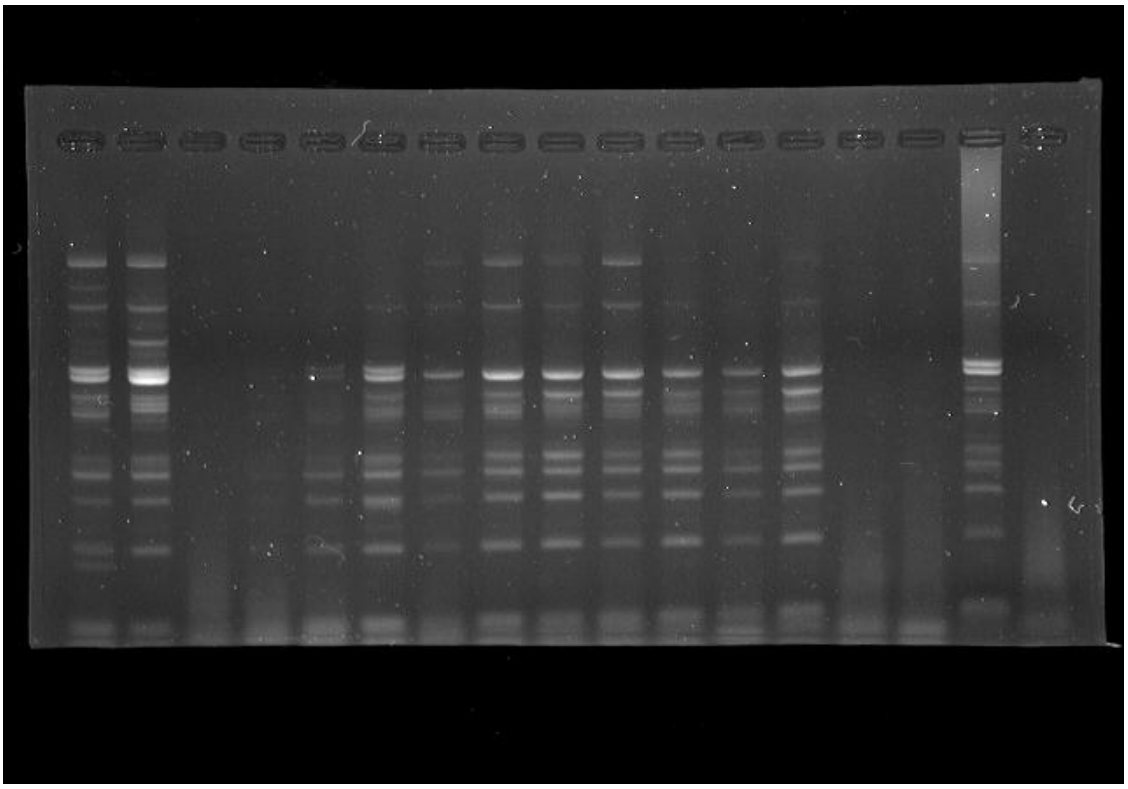

Figure 7.2: IRAP-PCR with 6149 and Nikita LTR primers.

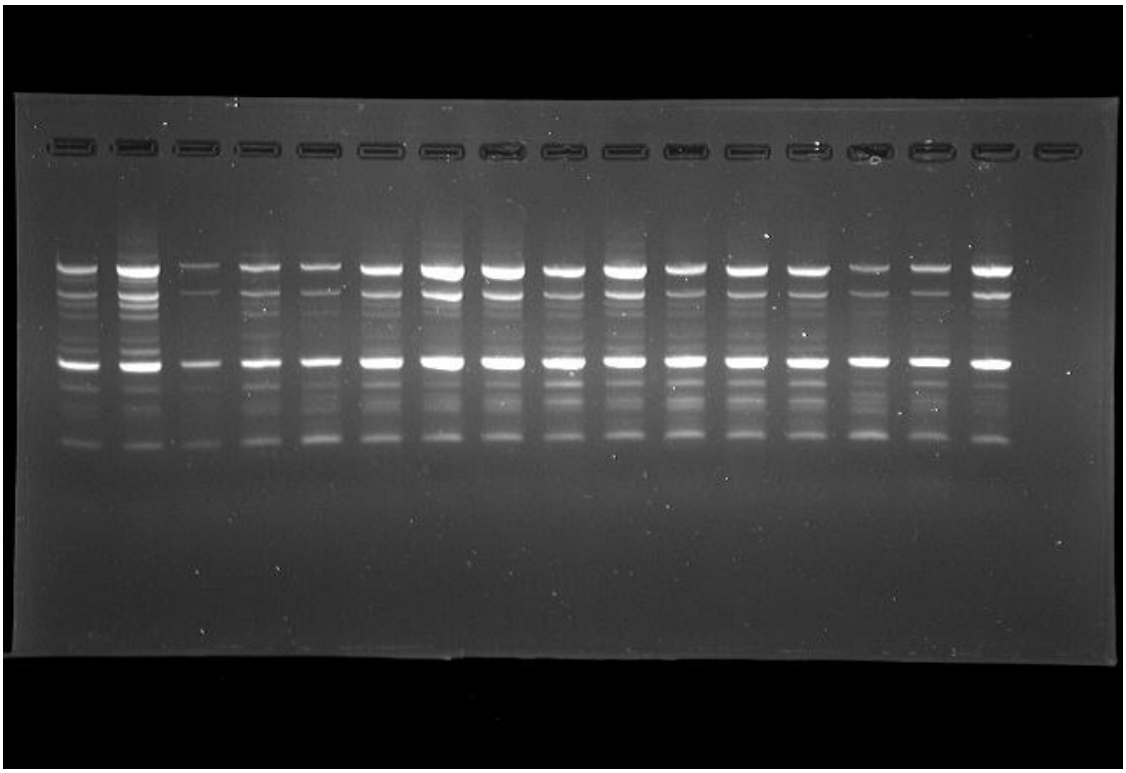

Figure 7.3: IRAP-PCR with LTR 6150 primers.

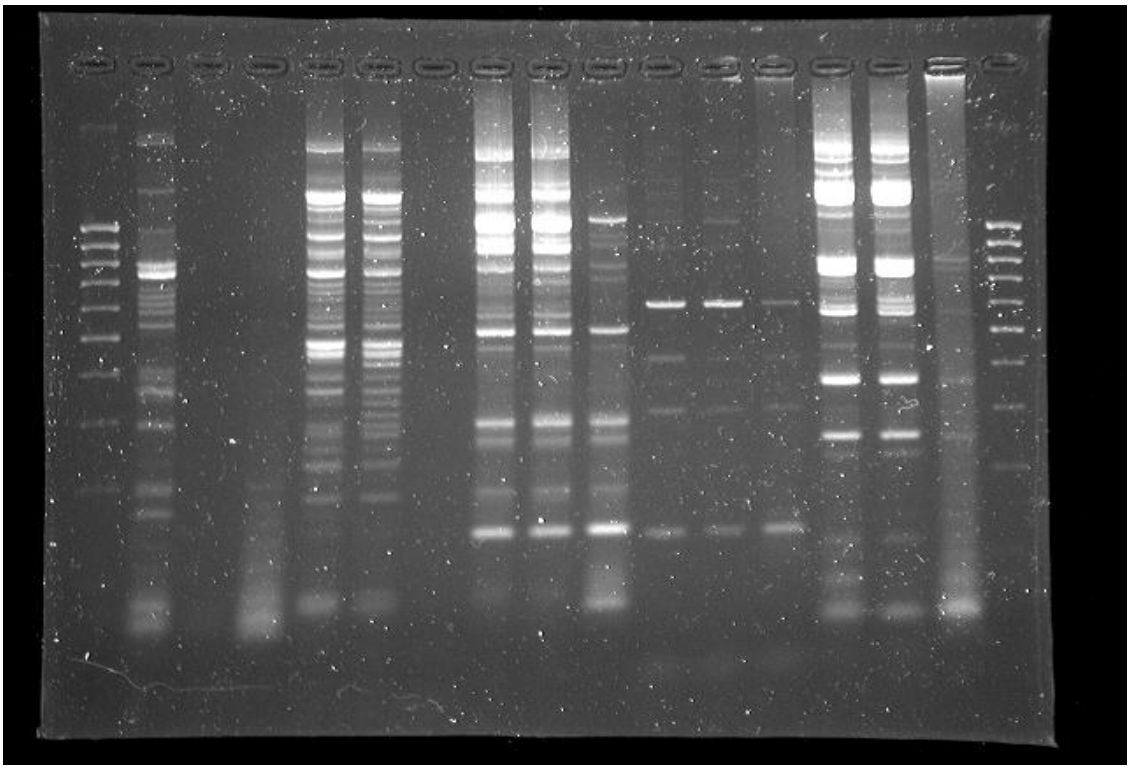

Figure 7.4: IRAP-PCR with LTR 6150 and 3’LTR primers.

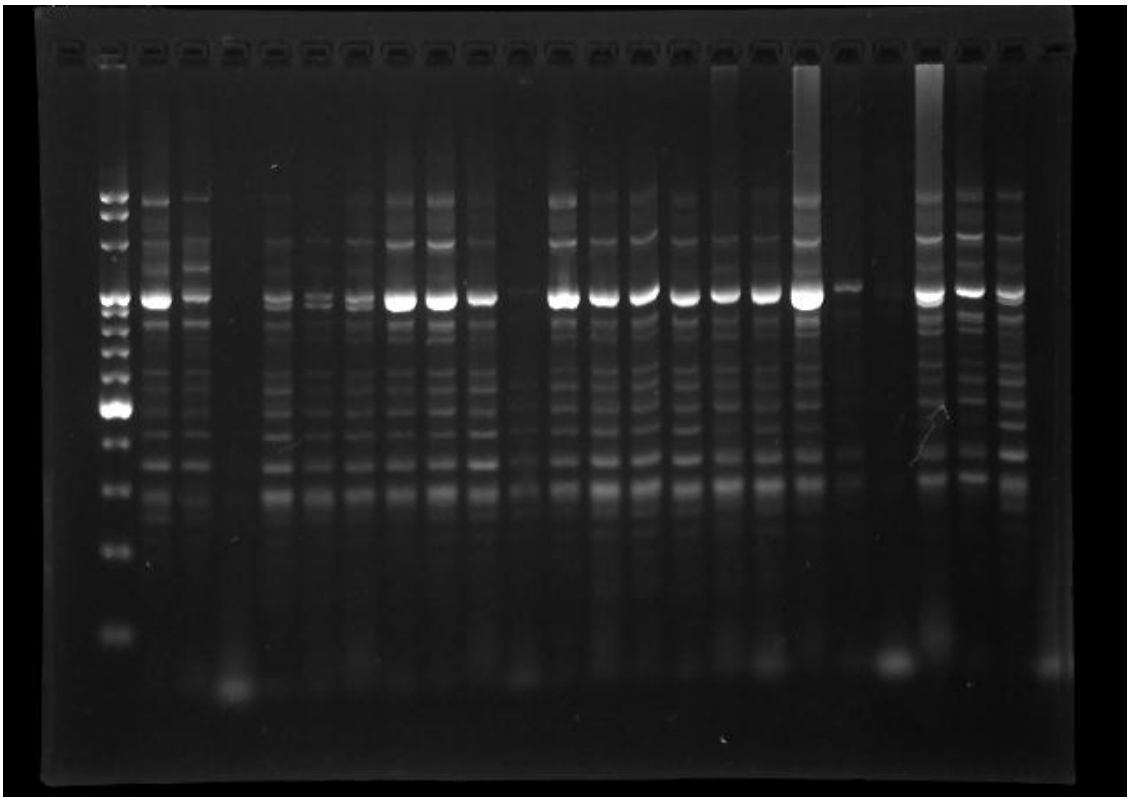

Figure 7.5: IRAP-PCR with Nikita LTR and 5’LTR2 primers.

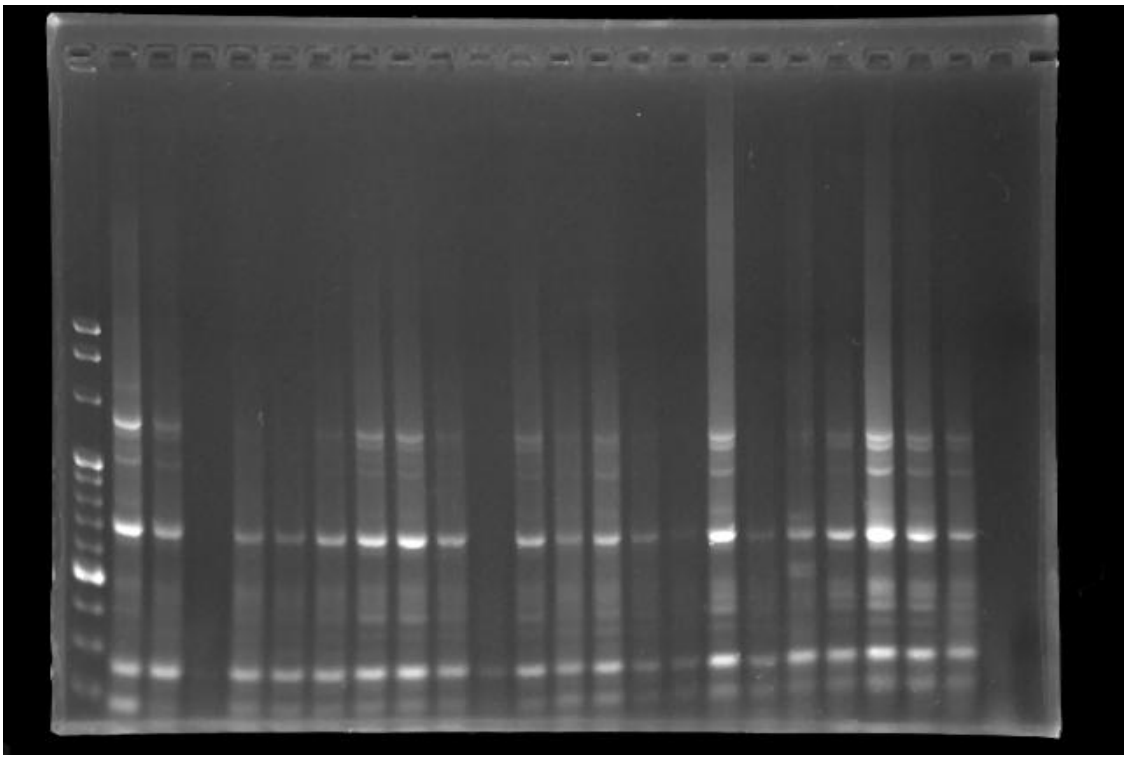

Figure 7.6: IRAP-PCR with 3’LTR and Sukkula LTR primers.

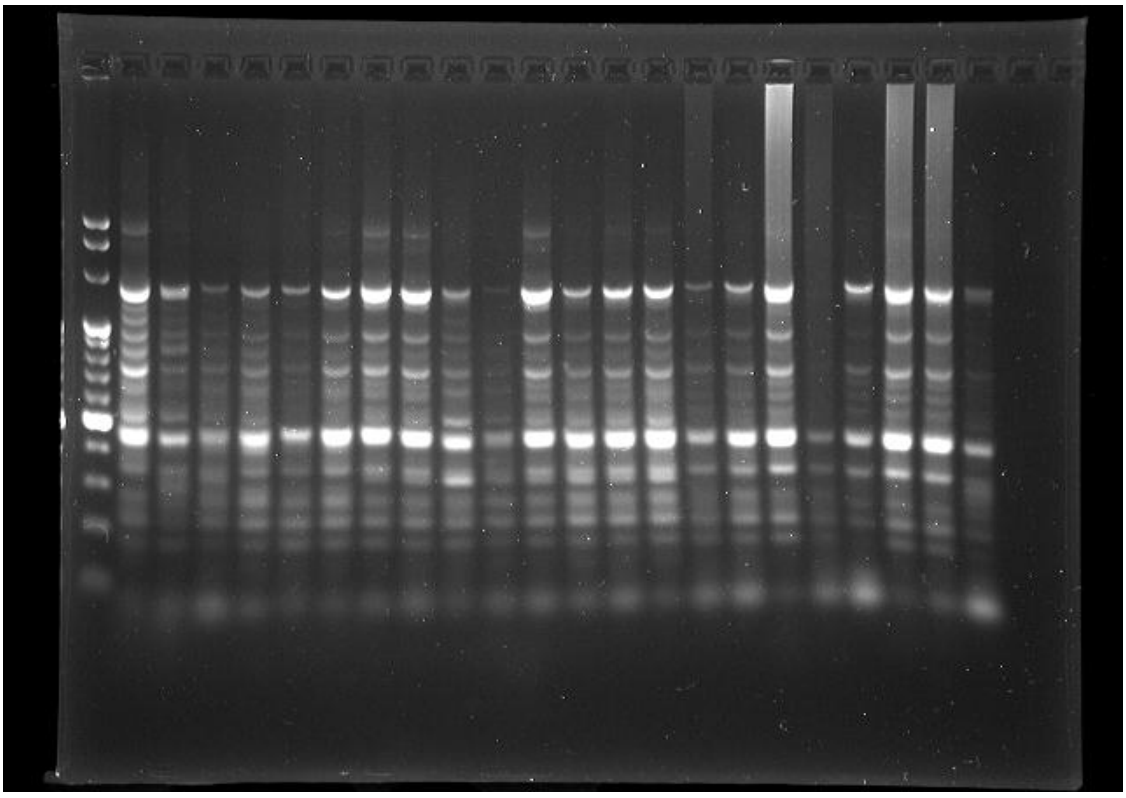

Figure 7.7: IRAP-PCR with LTR 6150 and 5’LTR1 primers.

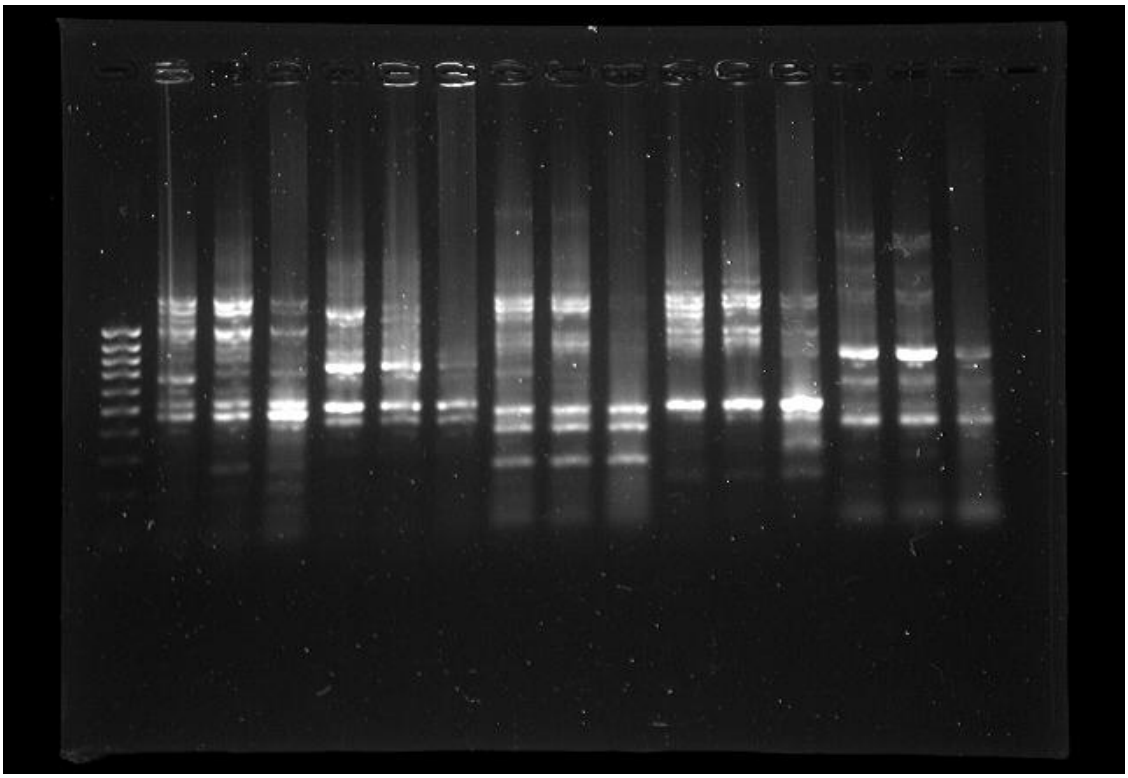

Figure 7.8: IRAP-PCR with 5’LTR1 and Nikita LTR primers.

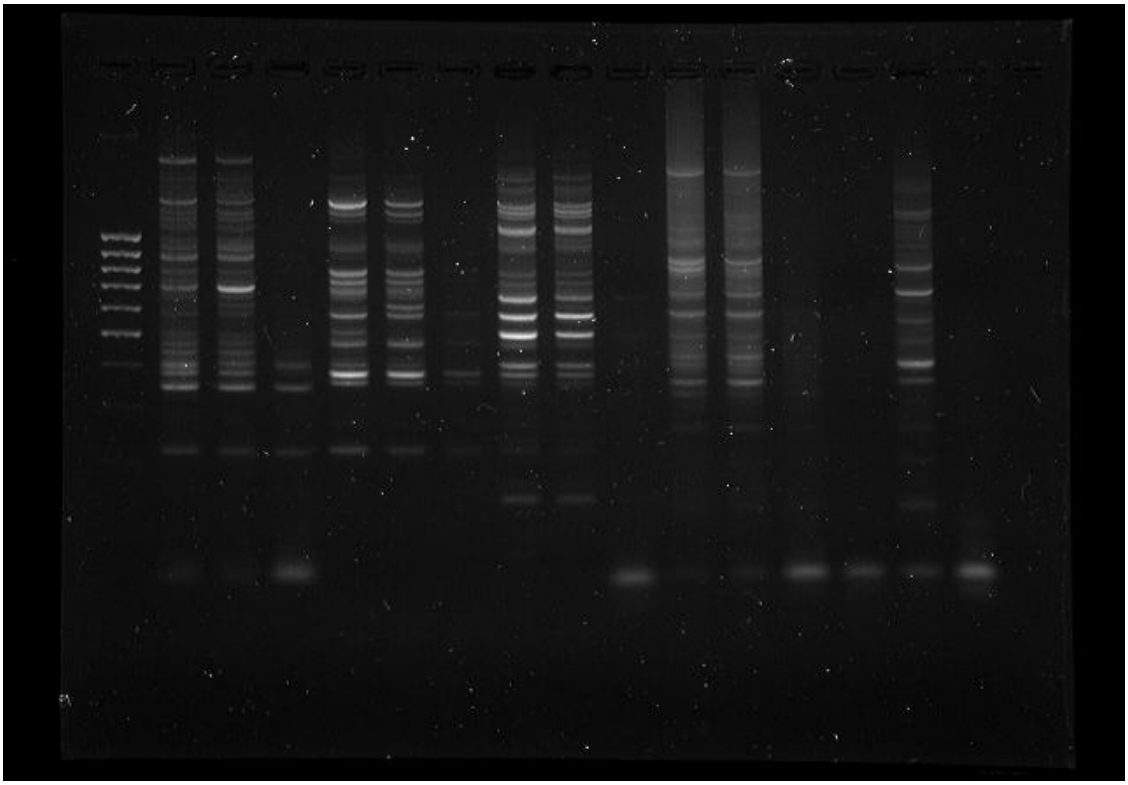

Figure 7.9: IRAP-PCR with LTR 6149 and 5’LTR1 primers.

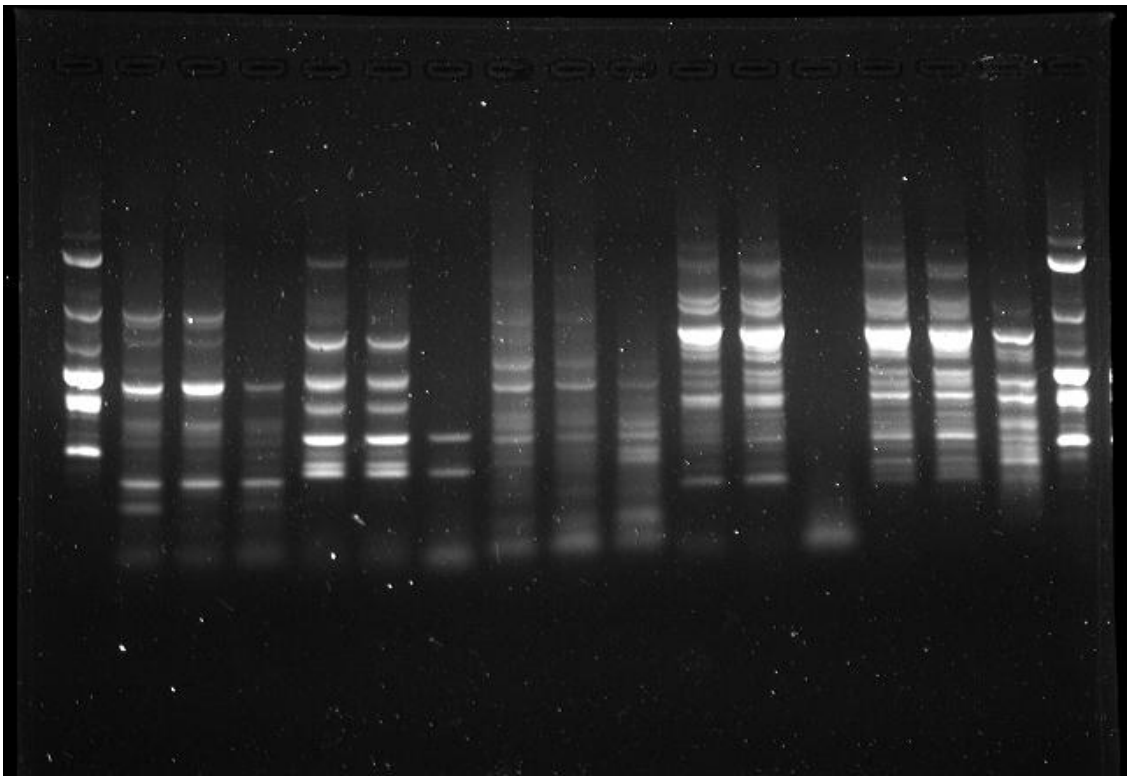

Figure 8.0: IRAP-PCR with 3’ LTR and Nikita LTR primers.

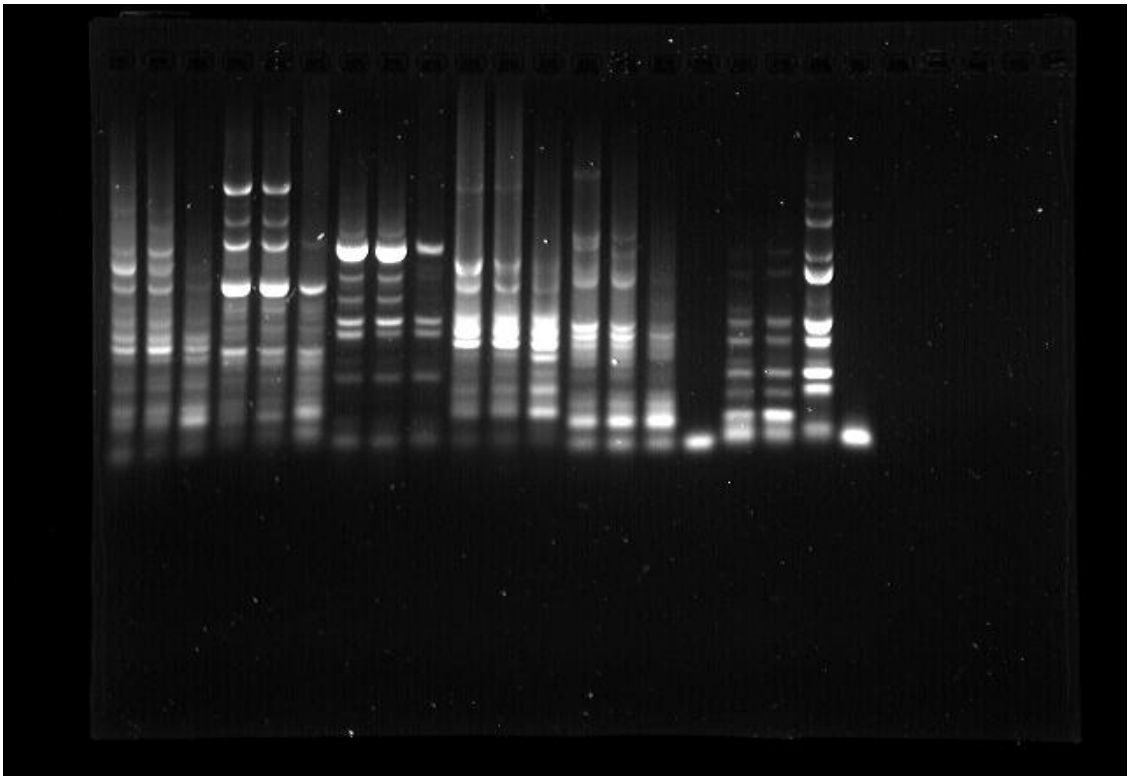

Figure 8.1: IRAP-PCR with LTR 6149 and Sukkula LTR primers.

REFERENCES

Bathusha S. and Othman R.Y., (2004). Analysis of somaclonal variation of tissue cultured banana plants derived from various explants using Random Amplified Polymorphic DNA (RAPD) markers. International Banana Congress 6-9 July 2004.

Bennetzen, J.L., (2000). Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42:251-269.

Bennetzen, J.L., (2002). Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica. 115:29-36.

Boyko E., Kalendar, R., Korzun, V., Gill B. & Schulman, A.H. Combined mapping of *Aegilops tauschii* by retrotransposon, microsatellite, and gene markers (2002). Plant Mol. Biol. 48, 767-790.

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

Chan, Y.K., P. Raveendranathan, M.L. Raziah amd S.T Choo (1994). Penanaman Betik. Kuala Lumpur, Malaysian Agriculture Research and Development Institute (MARDI).

Chan, Y.K. (2002). Breeding papaya for resistance to ringspot virus disease in Malaysia. Symposium 20 (S20): Citrus and other subtropical and tropical fruit crops: issues, advances and opportunities (Pg. 98). Metro Toronto Convention Centre.

Chun W.J.L, Yu Q., Hou S., Skelton R.L., Jones M.R., Lewis K.L.T., Murray J., Eustice M., Guan P., Agbayani R., Moore P.H., Ming R. and Presting G.G. (2006). Analysis of papaya BAC end sequences reveals first insights into the organization of a fruit genome. Mol Gen Genomics 276:1-12.

Croy E.J., Ikemura T., Shirsat A. and, Croy, R.R.D. (1993). Plant nucleic acids. Croy, R. R.D. (Ed.), Plant Molecular Biology. Oxford: BIOS Scientific.

de Vienne, D. (2003). Molecular markers in plant genetics and biotechnology. INRA, Paris, Science Publishers.

Dekkers, J.C., and Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. Nat. Rev. Genet., 3, 22-32.

Dillon S., Ramage C., Ashmore S. and Drew R.A. (2006). Development of a codominant CAPS marker linked to PRSV-P resistance in highland papaya. Theor Appl Genet 113:1159-1169.

Dixit A., Ma K.H., Yu J.W., Cho E.G., and Park Y.J., (2005). Reverse transcriptase domain sequences from Mungbean (Vigna radiate) LTR retrotransposons: Sequence characterization and phylogenetic analysis. Plant Cell Rep., 00:1-12.

Doyle, J.J. and Doyle J.L, (1990). A rapid total DNA preparation procedure for fresh plant tissue. Focus 12:13-15.

Eastwood R.F., Lagudah E.S., Appels R. (1994). A directed research for DNA sequences tightly linked to cereal cyst nematode resistance genes in *Triticum tauschii*. Genome 37:311-319.

Ellis, T.H.N., Poyser, S.J., Knox, M.R., Vershinin, A.V. and Ambrose, M.J. (1998). Ty1-copia class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. Mol. Gen. Genet. 260, 9±19.

Felsenstein, J. (2004). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

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

Flavell A.J., S.R.Pearce and A. Kumar (1994). Plant transposable elements and the genome. Curr.Op.Genet.Devel. 4:838-844.

Gonsalves D. (1998). Control of Papaya Ringspot Virus in Papaya: A Case Study. Annu. Rev. Phytopathol., 36, 415-477.

Grandbastien M.A., (1998). Activation of plant retrotransposons as ubiquitous components of plant genomes. Jpn J Genet 68:35-46.

Heslop-Harrison J.S., Brandes A., Taketa S., Schmidt T., Vershinin A.V., Alkhimova E.G., Kamm A., Doudrick R.L., Schwarzacher T., Katsiotis A., Kubis, S. Kumar, A. Pearce, SR.= Flavell, A.J, Harrison, G.E. (1997). The chromosomal distribution of Ty1-copia group retrotransposable elements in higher plants and their implications for genome evolution. Genetica 100:197-204.

Hirochika F. and Hirochika R. (1993). Ty1-copia group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet 68:35-46.
Hirochika, H., K. Sugimoto, Y. Otsuki, H. Tsugawa & M. Kanda, (1996). Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci USA 93: 7783-7788.

Jefferies S.P., King B.J., Barr A.R., Warner P., Logue S.J., Langridge P. (2003). Marker-assisted backcross introgression of the Yd2 gene conferring resistance to barley yellow dwarf virus in barley. Plant Breed 122:52–56.

Jing R., Knox M.R., Lee J.M., Vershinin A.V., Ambrose M., Ellis T.H. *et al*. (2005). Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in *Pisum* species. Genetics 171: 741-752.

Kalendar, R., Grob, T., Regina, M., Suoniemi, A. And Schulman, A.H. (1999). IRAP and REMAP: two new retrotransposons-based DNA fingerprinting techniques. Theor Appl Genet 98: 704-11.

Kalendar, R., J. Tanskanen, S. Immonen, E. Novo and A.H Schulman (2000). Genome evolution of wild barley (Hordeum spontaneum) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA 97: 6603-6607.

Kalendar, R., Flavell, A.J., Ellis, T.H.N., Sjakste, T., Moisy, C. And Schulman, A.H. (2011). Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106: 520-530.

Kalendar, R. and Schulman, A.H. (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. Nature Protocols 1(5): 2478-2484.

Kalendar R., Antonius K., Smykal P. and Schulman A.H. (2010). A universal method for DNA fingerprinting and retrotransposon isolation. Theor Appl Genet (doi:10.1007/ s00122-010-1398-2).

Kumar, A. (1996). The adventures of the Ty1-*copia* group of retrotransposons in plants. Trends Genet. 12: 41-43.

Kumar , A., Pearce, S.R., McLean, K., Harrison, G., Heslop-Harrison, J.S., Waugh, R., Flavell, A.J.(1997). The Ty1-copia group of retrotransposons in plants: genomic organisation, evolution, and use as molecular markers. Genetica 100: 205-217.

Kumar, A. (1998). The evolution of plant retroviruses: moving to green pastures. Trends Plant. Sci. 3:371-374.

Kumar, A. and Bennetzen, J.L. (1999). Plant retrotransposons. Annu Rev Genet 33: 479-532.

Kwon Y.S., Eun M.Y. and Sohn J.K (2001). Marker-assisted selection for identification of plant regeneration ability of seed-derived calli in rice (*Oryza sativa* L.).
Mol Cells 1: 103-106

Kyndt T., Droogenbroeck B.V., Romeijn-Peters E., Romero-Motochi J.P., Schedelman X., Goetghebeur P., Van Dammne P.,Gheysen D.(2005). Molecular phylogeny and evolution of Caricaceae based on rDNA internal transcribed spacers and chloroplast sequence data. Molecular Phylogenetics and Evolution Elsevier 37: 442-449.

Lawson W. R., Jan C. C, Shatte T., Smith L., Kong G.A. and Kochman, J.K. (2011). DNA markers linked to the R2 rust resistance gene in sunflower (*Helianthus anuus* L.) facilitate anticipatory breeding for this disease variant. Mol Breeding 28:569–576.

Lecouls A.C., Bergougnoux V., Rubio-Cabetas M.J., Bosselut N.,Voisin R., Poessel J.L., Faurobert M., Bonnet A., Salesses G., Dirlewanger E., Esmenjaud D. (2004). Marker-assisted selection for the wide-spectrum resistance to root-knot nematodes conferred by the Ma gene from Myrobalan plum (*Prunus cerasifera*) in interspecic Prunus material. Mol Breed 13:113–124.

Lightbourn, G. and Veilleux R. (2003). Retrotransposons based markers to characterize somatic hybrids and assess variation induced by protoplast fusion of monoploid potato. Acta Horticulturae (ISHS) 619:35-43.

Lim, W.H. and Doon Y. (1989). Major diseases of papaya and their management. Presented at Seminar Betik Eksotika MARDI-MAPPS, Johor Baru, 10pp.

Lim, W.H. and Doon, Y. (1990). Major diseases of papaya and their management. MARDI-MAPPS Seminar On Eksotika Papaya: Problems and Prospects (7-8 March 1989: Johor Bahru) Bangi. Malaysian Agriculture Research and Development Institute (MARDI) 35-42.

Lim, J.K. and Simmons J.M. (1994). Gross chromosome rearrangements mediated by transposable elements in *Drosophila melanogaster*. Bioessays 16:269-73.

Lin, C.C., Su, H.J. and Wang, D.N. (1989). The control of papaya ringspot virus in Taiwan R.O.C. Tech. Bull. ASPAC 114:1-13.

Liu, Z.W., Fu, T.D., Tu, J.X. Chen, B.Y. (2004). Inheritance of seed colour and identification of RAPD and AFLP markers linked to the seed colour gene in rapeseed (Brassica napus L.). Theor. Appl. Genet., 106, 107-111.

Manshardt, R. and Drew R.A. (1998). Biotechnology of papaya. In: Drew RA (ed) Proceedings of international symposium on tropical and subtropical fruits, part 2. Acta Horticulture, Netherlands, pp 285-291.

Matsuoka, Y. and Tsunewaki, K. (1996) Wheat retrotransposon families identifed by reverse transcriptase domain analysis. Mol. Biol. Evol. 13, 1384±1392.

MeKsem K, Leister D, Peleman J, Zabeau M, Salamini F, Gebhardt C (1995). A high-resolution map of the vicinity of the R1 locus on chromosome V of potato based on RFLP and AFLP markers. Mol Gen Genet 249:74-81.

Mhiri C., J-B. More, S. Vernhettes, J.M. Casacuberta, H. Lucas & M-A. Grandbastien, (1997).The promoter of the tobacco Tnt1 retrotransposons is induced by wounding and by abiotic stress. Plant Mol. Biol. 33: 1-10.

Ming R., Moore P.H., Zee F., Abbey C.A., Ma H. and Paterson A.H. (2001). Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome. Theor Appl Genet 102:892–899.

Ming, R. *et al*.(2008). The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya Linnaeus*). Nature 452: 991- 996.

Mohan M., Nair S., Bhagwat A., Krishna T.G., Yano M., Bhatia C.R., Sasaki T. (1997). Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3: 87-103.

Narayanan N.N., Baisakh N., Oliva N.P., VeraCruz C.M., Gnanamanickam S.S., Datta K. and Datta S.K. (2004). Molecular breeding: marker-assisted selection combined with biolistic transformation for blast and bacterial blight resistance in Indica rice (cv. CO39). Mol Breed 14:61–71.

Nagarajan N., Navajas-Perez R., Pop M., Alam M., Ming R., Paterson A.H., Salzberg, S.L. (2008). Genome-Wide Analysis of Repetitive Elements in Papaya. Tropical Plant Biol. 1:191-201.

Oliveira E.J., Silva A.S., Carvalho F.M., Santos L.F., Costa J.L., Amorim V.B.O. and Dantas J.L.L. (2010). Polymorphic microsatellite marker set for Carica papaya L. and its use in molecular-assisted selection. Euphytica 173: 279-287

Page, R. D. P. and E. C. Holmes (1998). Molecular Evolution: A Phylogenetic Approach. Blackwell, Oxford.

Pearce S.R., G. Harrison., Li, D., J.S. Heslop-Harrison., A. Kumar and A.J. Flavell (1996a). The Ty1-copia group retrotransposons in Vicia species: copy number, sequence heterogeneity and chromosomal localisation. Mol. Gen.Genet, 250, 305±315.

Pearce S.R., G. Harrison, J.S. Heslop-Harrison, A.J. Flavell, & A. Kumar (1996b). The Ty1-copia group retrotransposons of Allium cepa are distributed throughout the chromosomes but are enriched in the telomeric heterochromatin. Chromosome Res. 357-364.

Pearce S.R., G. Harrison, J.S. Heslop-Harrison, A.J. Flavell & A. Kumar, (1997). Characterisation and genomic organisation of Ty1-*copia* group retrotransposons in rye (*Secale cereale*). Genome, in press.

Pearce S.R., Stuart-Rogers C., Knox M.R., Kumar A., Ellis T.H.N. and Flavell A.J. (1999). Rapid isolation of plant Ty1-copia group retrotransposon LTR sequences for molecular marker studies. The Plant Journal 19 (6), 711-717.

Peterson-Burch B.D., Wright D.A., Laten H.M. and Voytas D.F. (2000). Retroviruses in plants. Trends Genet. 16: 151-152.

Pratta G.R., Rodriguez G.R., Zorzoli R., Valle E.M. and Picardi, L.A. (2011). Phenotypic and molecular characterization of selected tomato recombinant inbred lines derived from the cross *Solanum lycopersicum* x *S. pimpinellifolium.* Journal of Genetics, 90 (2): 229-237

Preston, B.D., (1996). Error-prone retrotransposition: Rime of the ancient mutators. Proc Natl Acad Sci USA 93:7427-7431

Price Z., Dumortier F., MacDonald D.W. and Mayes, S. (2002). Characterization of copia-like retrotransposons in oil palm (Elaeis guineensis Jacq.) Theor Appl Genet 104:860-867.

Rogers S.A., Pauls K.P., (2000). Ty1-copia like retrotransposons of tomato (Lycopersicon esculentum Mill.). Genome 43:887-894

SanMiguel P. and Bennetzen J.L. (1998). Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann Bot 81:37-44

SanMiguel P., Tikhonov, A., Jin Y.K., Motchoulskaia N., Zakharov D., Berhan A.M., Springer P.S., Edwards K.J., Lee M., Avramova Z. and Bennetzen J.L. (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science 274, 765±768.

Schwarzacher T., A.R. Leitch and J.S Heslop Harrison, 1994, DNA:DNA in situ hybridisation – methods for light microscopy, pp. 127-155 in Plant Cell Biology: A Practical Approach, edited by N. Harris & K.J. Oparka. Oxford University Press, Oxford.

Sheidai M., Yahyazadeh, Farahanei F. and Noormohammadi Z. (2008). Genetic and morphological variations induced by tissue culture in tetraploid cotton (Gossypium hirsutum L.) Acta Biol Szeged 52(1):33-38.

Shirasu, K., A.H. Schulman, T. Lahaye and P. Schulzelefert (2000). A contiguous 66Kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res. 10: 908-915.

Soniya E.V., Banerjee N.S. and Das M.R. (2001). Genetic analysis of somaclonal variation among callus-derived plants of tomato. Cur Sci 80:1213-1215.

Tahara M., Aoki T., Suzuka S., Yamashita H., Tanaka M., Matsunaga S., and Kokumai, S., (2004). Isolation of a transcriptionally active element of high copy number retrotransposons in sweet potato genome. Biochemistry.

Teo C.H., Tan S.H., Othman R.Y. and Schwarzacher T. (2002). The Cloning of Ty1-copia-like Retrotransposons from 10 Varieties of Banana (Musa Sp.). J Biochem Mol Biol Biophys 6:193-201.

Teo C.H., Tan S.H., Ho C.L., Faridah Q.Z., Othman R.Y., Heslop-Harrison J.S., Kalendar R. and Schulman A.H. (2005). Genome constitution and classification using retrotransposon-based markers in the orphan crop banana. Journal of Plant Biology 48: 96-105.

Tietze, H.W. (2001). Papaya as medicine: A safe and cheap form of food therapy. Subang Jaya, Selangor. Pelanduk Publications.

Vanderwiel P.L., Voytas D. and Wendel J.F. (1993). Copia-like retrotransposable element evolution in diploid and polyploidy cotton (Gossypium L.) J. Mol. Evol. 36, 429±447.

Van Droogenbroeck B., Kyndt T., Maertens I., Romeijn-Peeters E., Scheldeman X., Romero-Motochi J., Van Damme P., Goetghebeur P., Gheysen G. (2004). Phylogenetic analysis of the highland papayas (Vasconcellea) and allied genera (Caricaceae) using PCR-RFLP. Theor Appl Genet 108:1473-1486.

Voytas D.F., Cummings M.P., Konieczny A., Ausubel F.M. and Rodermel S.R. (1992). Copia-like retrotransposons are ubiquitous among plants. Proc. Natl Acad. Sci. USA, 89, 7124±7128.

Watson, J.D. and Berry, A. (2004). DNA The secret of life.United Kingdom: Arrow Books.

Waugh R., McLean K., Flavell A.J., Pearce S.R., Kumar A., Thomas B.T. and Powell, W. (1997). Genetic distribution of BARE-1 retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol*. Gen. Genet. 253, 687±694.

Weising K., Nybom H., Wolff K. and Meyer W. (1995). DNA fingerprinting in plants and fungi.United States of America: CRC Press, Inc.

Wikström N., Savolainen V., V. and Chase, M.W. (2001). Evolution of the angiosperms: calibrating the family tree. Proc. R. Soc. Lond. B. 268, 2211-2220.

Yeh S.D., Bau H.J., Kung Y.J. and Yu T.A. (2006). Assessment of Emerging Virus Threats for Application of Transgenic Papaya Resistant to Papaya Ringspot Virus. International Symposium Ecological and Environmental Biosafety of Transgenic Plants 1-19.

Yeh S.D., Bau H.J., Kung Y.J. and Yu T.A. (2007). Papaya. Biotechnology in Agriculture and Forestry. Vol. 60 Transgenic Crops V: 73-96.

Yi G., Lee S.K., Hong Y.K., Cho Y.C., Nam M.H., Kim S.C., Han S.S., Wang G.L., Hahn T.R., Ronald P.C. and Jeon J.S. (2004). Use of Pi5(t) markers in marker-assisted selection to screen for cultivars with resistance to *Magnaporthe grisea*. Theor Appl Genet 109:978–998.

Young N.D. (2000). A cautiously optimistic vision for marker-assisted breeding. Mol. Breeding 505-510.

Yu Z.H., Mackill D.J., Bonman J.M. and Tanksley S.D. (1991). Tagging genes for blast resistance in rice via linkage to RFLP markers. Theor Appl Genet 81: 471-476.

Yu K. and Pauls K.P. (1993). Identification of a RAPD marker associated with somatic embryogenesis in alfalfa. Plant Mol Biol 22: 269-277.

Websites cited:

http://faostat.fao.org

http://plants.usda.gov/ java/nameSearch

http://the.honoluluadvertiser.com/dailypix/2004/Oct/17

http://www.biocenter.helsinki.fi/bi/bare-1_html/images/irap.jpg

http://www.grain.org

http://www.mardi.com.my