UNIVERSITI PUTRA MALAYSIA

CLASSIFICATION OF THE MALAYSIAN BANANA VARIETIES USING RETROTRANSPOSON MARKERS

TEO CHEE HOW

FSMB 2002 8

CLASSIFICATION OF THE MALAYSIAN BANANA VARIETIES USING RETROTRANSPOSON MARKERS

By

TEO CHEE HOW

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

April 2002

Dedicated to my parents

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Master of Science

CLASSIFICATION OF THE MALAYSIAN BANANA VARIETIES USING RETROTRANSPOSON MARKERS

By

TEO CHEE HOW

April 2002

Chairman: Tan Siang Hee, Ph.D.

Faculty: Food Science and Biotechnology

Retrotransposons are the most common class of eukaryotic transposable elements that make up over 50% of nuclear DNA found in many plants with large and complex genomes. The long terminal repeat (LTR) retrotransposons have a high degree of autonomy and encode at least five distinct protein components, which are required for their movement in the genome. We have exploited the repetitive, dispersed nature of many LTR-containing retrotransposon families for the visualization of genomic polymorphism (genome constitution) and the visualization of varieties polymorphism (varieties classification). The IRAP-PCR (*I*nter-*R*etrotransposon *A*mplified *P*olymorphism-Polymerase Chain Reaction) markers are generated by the proximity of two LTRs or reverse transcriptase (RT) genes using outward-facing primers annealing to LTR target sequence. This method could distinguish between banana genome constitution (A and B genome) and varieties. This method can track new insertion of retrotransposon in banana tissue culture materials. The evolution of Malaysian bananas can also be tracked using this phylogenetic analysis. Since the method is PCR-based, it may allow the detection and isolation of the active, newly inserted retrotransposon copies easily, by analyzing

the subsequent generations within an otherwise genetically homogeneous material. This method is applicable to other plants with dispersed families of LTR-retrotransposons where either the LTR or nearby internal sequences are known.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

KLASIFIKASI VARIETI PISANG DI MALAYSIA DENGAN MENGGUNAKAN PETUNJUK RETROTRANSPOSON

Oleh

TEO CHEE HOW

April 2002

Pengerusi: Tan Siang Hee, Ph.D.

Fakulti: Sains Makanan dan Bioteknologi

Retrotransposon adalah kelas eukariotik 'transposable' elemen yang paling biasa ditemui dan ia merangkumi lebih daripada 50% nuklear DNA di dalam tumbuhan yang mempunyai genom yang besar and kompleks. 'Long terminal repeat' (LTR) retrotransposon mempunyai darjah otonomi yang tinggi dan mengkodkan sekurang-kurangnya lima bahagian protein yang berbeza yang diperlukan untuk pergerakannya di dalam genom. Kami telah menggunakan ciri-ciri penyebaran dan berulangan semulajadi kebanyakan famili retrotransposon yang mempunyai LTR untuk visualisasi kepelbagaian genomik (genom konstitusi) dan visualisasi kepelbagaian variety (klasifikasi varieti). Petunjuk IRAP-PCR (*I*nter-*R*etrotransposon *A*mplified *P*olymorphism-Polymerase Chain Reaction) terhasil dengan perdekatan dua LTRs atau gen reverse transcriptase (RT) dengan menggunakan pencetus yang menghala keluar dan melekat kepada tapak jujukan LTR. Kaedah ini boleh membezakan genom konstitusi pisang (A dan B genom) dan varietinya. Kaedah ini boleh mengesan jejak tapak kemasukan baru retrotransposon di dalam bahan kultur tisu. Evolusi pisang di Malaysia boleh juga dikesan dengan menggunakan analisis filogenetik ini. Oleh kerana kaedah ini adalah berasaskan PCR, ia membolehkan

pengesanan dan pengasingan salinan retrotransposon baru yang aktif dengan mudah dengan menganalisis generasi seterusnya di antara bahan genetik yang biasanya homogen. Kaedah ini boleh digunakan untuk sebarang tumbuhan yang mempunyai famili LTR-retrotransposon yang tersebar di mana LTR atau jujukan dalamannya diketahui.

ACKNOWLEDGEMENTS

'This project paper is dedicated to my parents and my family members'

First of all, I would like to give my heartiest appreciation to my supervisory committee, Dr. Tan Siang Hee, Associate Professor Dr. Rofina Yasmin Othman, Dr. Ho Chai Ling and Dr. Faridah Qamaruz Zaman for their unrelenting guidance and supervision throughout my project. Most of the suggestions were constructive and practical.

I would like to thank Dr. Ruslan Kalendar and Dr. Alan Schulman from the Institute of Biotechnology, University of Helsinki, Finland and Prof. J.S. Heslop-Harrison from John Innes Centre, U.K., for providing me the primer of *BARE*-1 LTR and Ty1-*copia*-like retrotransposon, respectively.

Deep gratitude is also acknowledged to Choong Chieh Wean and Foong Chuen Yi, who had always helped me and supported me.

I would like to thank the other Master students in the Plant Molecular Biology Group of the Faculty of Food Science and Biotechnology, such as Teoh Wan Chin, Nancy Liew and See Pao Then. Thanks to Mr. Ong for creating a very good and clean environment for the experiment, and sorting out the bills.

The banana leaf samples were kindly provided by Dr. Siti Hawa Jamaluddin of MARDI.

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Science.

AINI IDERIS, Ph.D.
Professor/Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AFLP	Amplified Fragment Length Polymorphism
AP-PCR	Arbitrarily Primed Polymerase Chain Reaction
bp	Base pair
CTAB	Cetyltrimethylammonium bromide
EDTA	Ethylene diamine tetra-acetate
E-H	Flower Emergence to Harvest Duration
FAO	Food and Agriculture Organization
GOT	Glutamate oxalacetate transaminase
IBRGR	International Board of Plant Genetic Resources
IITA	International Institute of Tropical Agriculture
INIBAP	International Network for Improvement of Banana and Plantain
IRAP	*I*nter-*R*etrotransposon *A*mplified *P*olymorphism
IRAP-PCR	*I*nter-*R*etrotransposon *A*mplified *P*olymorphism-Polymerase Chain Reaction
kb	Kilo base
LER	Leaf Emergence Rate
LINE	Long Interspersed Nuclear Elements
LTR	Long Terminal Repeat
M	Molar
MARDI	Malaysian Agricultural Research Institute
MAS	Marker-Assisted Selection
MDH	Malate dehydrogenase
mg	Milligram
mm	Millimeter

mM	Millimolar
NaCl	Sodium chloride
nt	Nucleotide
PCR	Polymerase Chain Reaction
PVP-40	Polyvinylpyrrolidone with molecular weight 40000
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RER	Primary Root Extension Rate
RFLP	Restriction Fragment Length Polymorphism
RIRE	Rice retrotransposon
RT	Reverse Transcriptase
SINE	Short Interspersed Nuclear Elements
Spm	*Suppressor-mutator*
S-SAP	Sequence-Specific Amplified Polymorphism
SSR	Simple Sequence Repeat
SSRLP	Simple Sequence Repeat Length Polymorphism
TE	Tris-EDTA
t ha^{-1}	Tonnes per hectare
%	Percentage
°C	Degree Celsius
Σ	Sum
x	Times
g	Gram

CHAPTER I

INTRODUCTION

Banana is one of the sixteen fruit types that have been identified by the Malaysian Ministry of Agriculture as having commercial potential either as fresh or processed fruit. Banana is of great socio-economic importance in the moister areas of tropical agriculture and is an important fruit crop in Malaysia. They are soil-conservative, productive, and almost non-seasonal and they yield diverse foods from sweet fruits to staple starches as well as numerous useful secondary products, from fibres to wrappings. Banana is the staple food for millions of people in the 120 countries where they are popularly grown. In addition, banana is an important cash crop for small farmers. In some countries where commercial plantations exist, considerable revenues from exporting bananas are generated.

The main banana germplasm collection of Malaysia is located at the Malaysian Agricultural Research and Development Institute (MARDI). It was initiated in 1973 and now has about 200 accessions. However, more than 50% are duplicates or triplicates due to synonymy. Description and documentation have been carried out for about 80% of the accessions using the recommended IBPGR Descriptors for Banana (1986). The germplasm collection is under the care of Dr. Siti Hawa Jamaluddin (MARDI). The popular dessert cultivars for Malaysia are Pisang Mas, Pisang Berangan, Pisang Rastali, and Pisang Embun while the popular cooking types are Pisang Raja, Pisang Nangka, Pisang Tandok, Pisang Awak, and Pisang Abu

Nipah. However, only Pisang Mas is grown for the export market while other cultivars are for domestic consumption.

The genus *Musa* can be divided into four sections, which include both seeded and non-seeded (parthenocarpic) types: *Callimusa*, *Australimusa*, *Eumusa* and *Rhodochlamys*. In addition, there are three species for which the relevant section has not been determined: *Musa ingens*, *M. lasiocarpa* and *M. boman*. Major cultivated bananas arose from the *Eumusa* section and this is the biggest section in the genus and the most geographically widespread, with species being found throughout South East Asia from India to the Pacific Islands. Bananas, which is under *Eumusa* section consist of two groups of plants: (1) the cultivars, which are clones maintained exclusively through vegetative propagation; and (2) the wild plants, which are diploid and are derived from two main species (*M. acuminata* and *M. balbisiana*).

Systematic scoring of characters diagnostic of the two parental species and chromosome counting jointly are sufficient to diagnose the main cultivated groups. They are designated by the genome constitution: AA, AAA, AB, AAB and ABBB. The other groups (AAAB and AABB) have not yet been fully classified but should present no great difficulty. Moreover, there are cultivars derived from hybridisations with *M. schizocarpa* (S genome) and there are also cultivars derived from hybridisations between *M. balbisiana* and *M. textiles* (Butuhan). Furthermore, several landraces containing the two genomes *acuminata* and species from the Australimusa section (T genome) and two landraces containing A, B and T genomes have been found in Papua New Guinea.

Retrotransposons are DNA sequences that can move or *transpose* themselves from one site to another site within the genomic DNA by a mechanism involving reverse transcription. Retrotransposons can be divided into two main classes based on their structural differences: those flanked by long terminal repeats (LTR) and non-LTR retrotransposon. The LTR retrotransposon can be further divided into two groups: Ty1-*copia* group (endonuclease domain is found upstream of the reverse transcriptase domain) and Ty3-gypsy group (the organization of coding domain is found to be similar to retroviruses), named after the well-characterized elements in *Saccharomyces cerevisiae* and *Drosophila melanogaster*. Retrotransposons contain long, defined and conserved sequences that can be used for cloning of specific markers and flanking sequences. Retrotransposons that are actively transposed will produce new insertion in the genome, which would lead to polymorphisms. These new insertions may then be detected and used to temporally order insertion events in a lineage, thereby helping to establish phylogenies. These structures and replication strategies of retrotransposons give them several advantages to serve as molecular markers.

Lack of a suitable molecular marker system for bananas has made it difficult to characterize the genome constitution, varieties and biodiversity of bananas. By carrying out this project, the retrotransposons can be utilized as molecular markers for identification and characterization of bananas in Malaysia. Thus the main objective of this study is to classify the genome constitution and varieties of local bananas using retrotransposon-based marker system.

CHAPTER II

LITERATURE REVIEW

The Distribution and Importance of Banana

Origin and Distribution

The genus *Musa* can be divided into four sections, these include both seeded and non-seeded (parthenocarpic) types. *Callimusa* and *Australimusa* contain species with a chromosome number of 2n = 20 whereas *Eumusa* and *Rhodochlamys* contain species with a basic chromosome number of 11 (2n = 22). Moreover, there are three species for which the relevant section has yet to be determined: *M. ingens* (2n = 14), *M. lasiocarpa* (2n = ?) and *M. boman* (2n = ?) (Horry *et al.*, 1997).

The *Eumusa* is the biggest section in the genus and progenitor of the major cultivated bananas. This section is the most geographically widespread, with species being found throughout South East Asia from India to the Pacific Islands. *Eumusa* section contains about 11 species (Horry *et al.*, 1997). Most cultivars are derived from two species, *M. acuminata* (A genome) and *M. balbisiana* (B genome). However, Shepherd and Ferreira (1982) have been successfully identified cultivars derived from hybridisations with *M. schizocarpa* (S genome), and this was confirmed by Carreel *et al.* (1994). Furthermore, a Philippine clone (Butuhan) is considered to be the result of an ancient hybridisation between *M. balbisiana* and *M. textiles* (T genome). Several landraces containing the two genomes *acuminata* and species from

the *Australimusa* section (T genome) and two landraces containing all three A, B and T genomes, have been found in Papua New Guinea (Carreel, 1994).

Musa acuminata (AA) and *M. balbisiana* (BB) are both diploids with $2n$ = 22. *Musa acuminata* is the most widespread of the *Eumusa* species being found throughout the range of the section. Chromosome structural changes that have occurred spontaneously or as a result of recombination events have resulted in the development of natural reproductive barriers within the species, causing subspecies divergence and genetic diversity in the species.*Musa balbisiana* which is considered to be more drought and disease resistant than *M. acuminata* and such characteristics are often found in cultivars containing a 'B' genome. *Musa ornata*, which is an ornamental banana, is under the *Rhodochlamys* section with a basic chromosome number of 11 (2n = 22). The genome constitution of *M. ornata* is yet to be determined.

The first step in the evolution of the edible bananas was the development of parthenocarpy and seed-sterility in *M. acuminata,* which are under human selection. This gave rise to the edible diploid cultivars (AA) in South East Asia. From the AA cultivars, by chromosome restitution at meiosis, there arose the AAA (*acuminata*) triploids, one of the three most important groups. Another step that was thought to have occurred in South East Asia, was the crossing of AA cultivars with *M. balbisiana* (BB). *Musa balbisiana* is a hardier and more drought-tolerant plant than *M. acuminata*, so the hybrid groups do not only extended the range of plant characters and quality features but also help to extend the geographical range of the bananas out of the wetter tropics into the seasonally drier zone.

In South East Asia, the bananas have been first recorded probably several thousand years ago. About 2000 years ago, they spread in the hands of travellers, eastwards to the remoter Pacific Islands and westwards to Africa (probably via Madagascar). The first European visitors to West Africa found them there and several clones were taken to the New World very soon after discovery. There, the crop spread very rapidly. The present distribution is roughly 30° north and south and bananas are grown wherever there is frost-freedom and enough rain. The history of the crop in Africa still present many problems and it is not clear why the varieties grown in eastern and western Africa could be so different if they had a common ancestor (Simmonds, 1962; 1976).

Importance as a Food Crop

Plantains and bananas are staple foods for rural and urban consumers in the humid tropics and an important source of rural income, particularly in some locations where smallholders produce them in compound or home gardens. Hence, fruit harvested from bananas and plantains are important components of food security in the tropical world and provide income to the farming community through local and international trade. The banana production in Malaysia is around 545 thousand tonnes annually (FAO, 2000), of which bananas cultivated for the export trade account for only 7.2%.

Bananas are chiefly eaten raw as a dessert fruit, because in the ripe state they are sweet and easily digested. Plantains are generally much starchy and can be eaten either ripe or unripe. They are usually boiled, fried or roasted. Gowen (1988) has

highlighted the apparent ambiguity of using the word 'plantain'. To many, plantain implies a cooking banana but in Spanish the word can also be used to mean dessert forms. There appear to be no accepted botanical distinction between the starchy types that have to be cooked, and the sweeter types that can be eaten raw. As fruits ripen, a conversion of starch to sugar occurs and this process is slower in cooking varieties, which contain *M. balbisiana* characteristics. All varieties derived from *M. acuminata* only are more sugary and are mainly eaten as a dessert fruit when ripe. However, such types are often popular as a starchy food when cooked in the unripe state. There is thus considerable overlapping between bananas and plantains, with respect to the way they are consumed.

In terms of world trade of *Musa*, plantains and cooking bananas are insignificant, but for domestic use as staple food they are vitally important in some countries. This is especially so in the equatorial belt of Africa stretching from East to West where they are the major staple food. About 70 million people in West and Central Africa are estimated to derive more than one quarter of their food energy requirement from plantains. The unripe fruits are peeled, wrapped in banana leaves and steamed. They are then pounded into porridge and eaten, the starchy dish being called 'matoke'. Unripe or ripe fruits may sometimes be baked, roasted of fried. In Uganda and Tanzania a nutritious beer is also brewed from plantains and large quantities of this are consumed in the region. Plantains and cooking bananas also form part of the daily diet of people in the Caribbean and Latin America.

Apart from their major uses as dessert fruits eaten raw, or as starchy fruits cooked before eating, the proportion of banana and plantain production put to other