UNIVERSITI PUTRA MALAYSIA

DNA FINGERPRINTING OF THEOBROMA CACAO

NORLIA BT BASHERUDIN

FSMB 1998 21

DNA FINGERPRINTING OF *THEOBROMA CACAO*

BY

NORLIA BT BASHERUDIN

Thesis Submitted in Fulfillment of the Requirement for the Degree of Master Science in Faculty of Food Science and Biotechnology,
University Putra Malaysia.

April 1998

ACKNOWLEDGEMENTS

In the name of ALLAH, the Merciful and the most Beneficent. All praise are due to ALLAH for giving me the opportunity, patience and guidance in completing this study.

I would like to express my sincere appreciation to the following people for helping me during my studies:

Dr. Suhaimi Napis, Dr. Harikrishna Kulaveerasingam and Dr. Raha Abdul Rahim for their guidance, suggestion, support and constructive criticism throughout the course of my study.

Meilina Ong Abdullah and husband, who have kindly and willingly helped me in completed this thesis, without them probably this thesis would never be completed.

All members of the biotechnology laboratory of UPM, including Liza, Jamilah, Chin Ching, Shari, Meta and Fadilah for the sincere help and good company throughout the long days spent in the lab.

Dr. Syed Kamaruddin Syed Wazir, Dr. Sapiyah Subali and all the workers at MARDI Hilir Perak for their help and co-operation

Dato' Dr. Abdul Razak Mohd. Ali, Director-general of FRIM, and Dr. Baskaran Krishnapilai for allowing me to get away from a busy work for a month in order to finish out this thesis.

All members of the genetic unit of FRIM, especially Dr. Wickneswari, Norwati Mohammad, Norwati Adnan, Chai Ting and Salwana for their support and encouragement.

Finally, I would like to express my highest gratitude to my husband, Azhar Saad for his continuous support, sacrifice and patience during the course of the study. To my son, Shahmi for giving me his smiles and to my parents who has always had complete faith in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF ABBREVIATIONS

AFLP	amplified fragment length polymorphism
$AgNO_3$	silver nitrate
AP-PCR	arbitrarily primed polymerase chain reaction
APS	ammonium persulphate
bp	basepairs
bis-acrylamide	N,N'- methylenebisacrylamide
BSA	bovine serum albumin
cDNA	complementary DNA
CsCl	cesium chloride
CTAB	cethylmethylammoniumbromide
DAF	DNA amplification fingerprinting
DNA	deoxyribonucleic acids
dATP	2' - deoxy-adenosine-5'- triphosphate
dCTP	2' - deoxy-cytidine-5' - triphosphate
dGTP	2' - deoxy-guanosine-5' - triphosphate
dTTP	2" - deoxy-thymidine-5' - triphosphate
DMSO	dimethyl sulfoxide
EDTA	ethylene diaminetetraacetic acid
EtBr	ethidium bromide
g	gram
HCl	hidrochloric acid
MARDI	Malaysian Agriculture Research Development Institute
Mbp	mega base pair
$MgCl_2$	magnesium chloride
min	minute (s)

mM	milimolar
mt	metric ton
MW	molecular weight
ηg	nanogram
NaCl	sodium chloride
Na_2CO_3	sodium carbonate
NaOH	sodium hidrocide
OD	optical density
PCR	polymerase chain reaction
PVP	polyvinylpyrrolidone
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
rDNA	ribosomal DNA
RE	restriction enzyme
rpm	revolution per minute
SCAR	sequence characterised amplified regions
SDS	sodium dodecyl sulphate / sodium lauryl sulphate
SSC	standard saline citrate
Taq	*Thermus aquatus*
TEA	triethylamine
TEMED	N,N,N',N'-tetramethylethlenediamine
μg	microgram
μl	microlitter
UV	ultraviolet

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

DNA FINGERPRINTING OF *THEOBROMA CACAO*

By

NORLIA BT BASHERUDIN

April 1998

Chairman: Dr. Suhaimi Napis

Faculty: Food Science and Biotechnology

Traditionally, the characterisation of *Theobroma cacao* is based on morphological characteristics and geographical distribution. Three major groups have been identified, namely, Criolla, Forestero and Trinitario (Wood, 1985). Crosses between these three groups have constitute the major breeding strategies used during the last few years. Currently, large numbers of cocoa clones derived from these crosses have been introduced in cocoa plantations.

Two different techniques, Random Amplified Polymorphic DNA (RAPD) with random primers and Restriction Fragment Length Polymorphism (RFLP) using universal probes, has been investigated as a mean to differentiate 12 local cocoa clones with 1 imported clone as a comparison. 40 primers have been used to amplify genomic DNA from 13 cocoa clones. Agarose and acrylamide gel electrophoresis used to separate the RAPD products were further stained with ethidium bromide and sliver nitrate, respectively. NTSYS-PC computer programme was used to analyse the RAPD data. Meanwhile, two

types of probe were used in RFLP study, *viz* M13 phage DNA and a fragment from a RAPD product.

The RAPD technique was found to be able to differentiate between local cocoa clones. Combining the result with UPGMA analysis, 13 tested clones were determined to fall into 3 main clusters. A band map has been formed as a future source of reference to determine the best combination of markers to differentiate two or more clones. M13 DNA, as a probe in the RFLP study, failed to give a consistence results in differentiating cocoa clones. An RAPD product, as a probe however was able to hybridise to few digested genomic DNA fragments. This indicates that the product is truly amplified fragment and was confirmed to be mid-repetitive. However the probe did not produce any polymorphic bands between the clones.

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

ANALISA 'DNA FINGERPRINTING' KE ATAS *THEOBROMA CACAO*

Oleh

NORLIA BASHERUDIN

April 1998

Pengerusi : Dr. Suhaimi Napis

Fakulti : Sains Makanan dan Bioteknologi

Pengenalpastian *Theobroma cacao*, pada tradisinya adalah berasaskan kepada bentuk morfologi dan taburan geografi. Tiga kumpulan utama telah dikenal pasti iaitu Criolla, Forestero dan Trinitario (Wood, 1985). Kacukan daripada tiga kumpulan ini berperanan besar dalam menyusun strategi pembiakan koko sejak sekian lama. Pada masa kini banyak klon-klon koko hasil dari kacukan tiga kumpulan tersebut telah diperkenalkan untuk penanaman koko.

Dua teknik yang berbeza, 'Random Amplified Polymorphic DNA' (RAPD) dengan primer rambang dan 'Restriction Fragment Length Polymorphism' (RFLP) menggunakan prob yang umum telah diselidiki untuk membezakan 12 klon koko tempatan dan satu klon impot sebagai perbandingan. 40 jenis primer telah digunakan untuk mengamplifikasi genom DNA daripada 13 klon koko. Elektroporesis gel agarose dan acrilamid telah digunakan untuk memisahkan serpihan hasil dari RAPD dan seterusnya di'stain' dengan ethidium bromide untuk agarose dan silver nitrate untuk acrilamid. Program komputer

NTSYS-PC telah digunakan untuk menganalisa data hasil dari RAPD. Sementara itu dua prob telah digunakan di dalam kajian RFLP iaitu DNA darlpada faj M13 dan serpihan dari hasil RAPD.

Teknik RAPD mampu untuk membezakan klon-klon koko tempatan. Keputusannya yang digabungankan dengan analisa UPGMA telah mengumpulkan 13 klon yang diuji kepada 3 kumpulan utama. Pemetaan 'band' telah juga dihasilkan untuk rujukan dl masa depan dalam mencari gabungan penanda yang sesuai untuk membezakan di antara dua atau lebih klon-klon koko. DNA M13 sebagai prob untuk kajian RFLP gagal menghasilkan keputusan yang sama pada setiap pecubaan, dalam membezakan klon koko. Walaubagaimanpun, serpihan hasil dari RAPD sebagai prob mampu untuk hibridisasi kepada beberapa serpihan genomic DNA koko yang telah dipotong. Ini menandakan hasil RAPD tersebut adalah serpihan teramlifikasi yang tulen dan ianya adalah 'mid-repetative'. Namun begitu prob ini tidak menunjukkan sebarang 'polymorphic band' di antara klon-klon tersebut.

CHAPTER I

INTRODUCTION

Theobroma cocoa is a tropical tree crop of great economic importance. Cocoa is a dicotyledonous plant classified taxonomically to be in the Sterculiaceae family. Cocoa originated near the head water of Amazon river (Samah, 1993) and was introduced to South East Asia in Seventeenth Century by the Spanish.

Cocoa has been traditionally propagated by seeds, however this technique usually produces considerable variation in a cocoa plantation (Ibrahim *et al.*, 1984). Vegetative propagation by leaf bud cutting, multiple bud cutting, marcoting, budding, grafting and layering has also been carried out. Consequently, a great number of cocoa clones have been introduced locally by MARDI, namely KKM (Koko Malaysia) and MHP (MARDI Hilir Perak) series. Harrisons Malaysia Plantation Bhd. properties has also introduced a number of clones, namely the PBC (Prang Besar clone) series.

The most important economic part of the cocoa tree is the cocoa beans, which are the source of the multi-billion dollars chocolate and cocoa butter industries. Cocoa is the third most important crop in Malaysia after oil palm and rubber (Thong *et al.*,1992) , while in world cocoa production, Malaysia was placed fifth after Cote d'Ivory, Brazil, Indonesia and Ghana (Malaysia Cocoa Board, 1995).

Breeding of high yielding plants contributed to the boost of cocoa bean production. Until now, all breeding and improvement programmes had relied on crosses between clones of various groups to obtain fine tasting chocolate from vigorous and productive trees (Laurent *et al*., 1993). It is important to understand the structure of the genetic variability of the species by refining the current classification. Beside breeding, plant breeders are also interested in producing improved varieties such as disease and pest resistant plants. Traditionally, plant breeders have produced improved varieties by selection based on useful agronomic traits and morphological descriptors (Wood,1985).

Although morphological characters have been used since a long time ago to differentiate species and clones of cocoa, however this technique is affected by environmental effects which make measurement difficult due to continuos variation. Further more many of the morphological descriptions can only be assessed at maturity. Isozymes, as a protein markers, have been used to differentiate cocoa clones (Atkitson *et al*., 1986), but this technique is limited by the low level of protein polymorphism detected in cocoa so far.

DNA markers, that are not subjected to environmental influences, provide an opportunity to examine the genetic relationship between cocoa clones. RFLP is the most reliable and powerful technique in a genetic variability study. However the RFLP technique is labour intensive and its use of radio-isotopes makes it difficult to extend its utility.

The availability of new and simpler genetic markers, has allowed some progress to be made in cocoa clone differentiation. The techniques, such as RAPD, which has been used widely in DNA fingerprinting (Wilde *et al.*, 1992 ; Ngoran 1994), diversity study (Laurent *et at*., 1994) and to establish genetic linkage maps (Carlson *et al*., 1991). Although RAPDs are

CHAPTER II

LITERATURE REVIEW

Cocoa (*Theobroma cacao*) - Background

Cocoa was introduced to South East Asia by the Spanish, who brought it from Latin America to the Philippines in 1670. In the Eighteenth Century, this crop was brought over to Indonesia and Sabah (Samah, 1993). In Malaysia, even though cocoa trees have survived in some other places, the first cocoa tree that produced pods was found in Melaka in 1778 (Koening, 1984). The planting of cocoa was initiated on small plantation areas in the Serdang Agriculture Station, Serdang, Selangor and also at the Agriculture Research Centre in Silam, Sabah (Samah, 1993).

Cocoa is a dicotyledonous plant classified taxonomically in the super order *Dillesiedae*, order *Malvaves*, family *Sterculiaceae,* genus *Theobroma* and species *cacao.* There are 20 species in the genus but the common cocoa tree, *Theobroma cacao,* is the only one that is widely cultivated. The other known species in the genus are *Theobroma bicolar* and *Theobroma grandiflorum* (Wood, 1985). *Theobroma cacao* is a diploid plant (Glicenstein *et al.*,1989) with a haploid number of 10 chromosomes (Martinson,1975). Its nuclear genome size is 200 megabase pairs (Mbp) and this makes it the second smallest plant genome known (Fritz, per. Comm.). *Theobroma cacao* is 'cauliflarous', which means that the flowers and fruits are produced on the older leafless parts of the stem and branches (Urquhart, 1955).

The natural habitat of the genus *Theobroma* is in the lower storey of the evergreen rain forest. All the species of the genus are found wild in the rain forest of the Western Hemisphere from 18°N to 15°N. In this habitat, rainfall is heavy, the temperature is relatively uniform through out the year, humidity is highly constantly and the shade is dense. Under these condition, *Theobroma cacao* flowers sparsely and bears only few pods (Wood, 1985).

Based on morphological characteristics and geographical distribution 3 major types of cocoa can be distinguished. They are Criolla, Forastero (upper and lower Amazon) and Trinitario. Forastero clones have originated from the Amazonian region of South America and possess green pods with predominantly purple seed. Most cultivated clones are Forestero which represent over 80% of the world cocoa production. Forastero has been further sub-divided into the Upper Amazon and Lower Amazon based on their geographic locations and also due to some distinctive characteristics. Criolla was the first domesticated cocoa and it originated from Central America and northern regions of South America. The beans are large, white or rosy and this gives a highly desirable chocolate although they have poor agronomic characteristics. Trinitario is a hybrid between Criolla and Forastero (Ngoran, 1994). Crosses between these three groups have a certain heterosis and the resulting hybrids has constituted the major breeding materials used during the later years.

Seed propagation is the easiest and cheapest form of producing planting material. This method results in highly variable progeny. Seeds can be produced either by cross pollination or hand-pollination. Oii and Chew in 1985 reported that cumulative dry bean yield from seed-garden mixed hybrid were 8% lower than seed-garden identified hybrid, while production from seed-garden identified hybrid were 10% lower than that of hand-pollinated seeds. Therefore, hand-pollinated seeds were the best choice among all the seeds for planting. However, Ibrahim *et al.,* in 1984 reported that there was a great difference observed in product and fruit index within progeny of plants derived from hand-pollinated seeds. This will then lead to a non-uniform product and quality of

cocoa. Furthermore, hand-pollination is technically demanding and requires expertise in order to achieve good results.

Vegetative propagation has an important role in reproducing true to type cocoa trees. Propagation by leaf bud cutting, multiple bud cutting, marcotting, budding, grafting and layering has been carried out in cocoa (Wood, 1985). From 1983 to 1989, MARDI has introduced 15 cocoa clones which have been proven to be good in production. These clones were known as KKM (Koko MARDI) series. Cocoa clones KKM 1, 2, 5, 6 and 7 were introduced in 1983, KKM 3, 4, 15, 17, 19, 25, 26 and 27 in 1986, and KKM 25 and 28 in 1989 (Saamin *et al.,* 1990). Cocoa clones beside being able to produce uniform bean size, are also a good source of planting materials (Engles, 1983).

Economic Importance

Cocoa seeds are a source of the multi-billion dollars chocolate and cocoa butter industries, which represent an important cash crop for more than 40 developing countries located within 20 degrees of the equator around the world (Urquhart, 1955). The Cocoa bean contains about 15% starch, 15% protein and 50% fat (Cuatrecasas, 1964).This makes it a valuable source of vegetable fat. The cocoa bean can be processed into chocolate liquor which is the main ingredient of chocolate. From the chocolate liquor, cocoa powder can be produced and subsequently used to prepare beverages or to make flavouring ingredients. Cocoa beans can also be processed into cocoa butter which is used in the manufacturing of chocolate confectioneries, pharmaceutical ointments and toiletries (Cuatrecasas, 1964).

World cocoa bean products have increased about 2.2% in 1993/94, from 2.376 million metric ton in 1992/93 to 2.429 million metric ton. This is due to African cocoa bean product which has increased by about 22% from 697 000 metric ton in 1992 to

850 000 metric ton in 1993/94. On the other hand, cocoa production in Malaysia has decreased from 200 000 metric ton to 180 000 metric ton, 10% lower than in 1993. Consequently Malaysia is the fifth largest world cocoa producer after Cote d'Ivoire, Brazil, Indonesia and Ghana (Malaysia Cocoa Board, 1995)

Cocoa have been reported to be the third most important crop in Malaysia after oil palm and rubber (Thong *et al.*, 1992) and the Malaysia Cocoa Board (1995) reported that export of cocoa products alone in 1994 has contributed for about RM722.74 million. 45% of the exports were in the form of cocoa butter, 36% in the form of cocoa bean, 9% in the form of chocolate , 6% in the form of cocoa powder while another 4% were in the form of cocoa paste. Cocoa products were exported to more than 10 countries such as Singapore, United Kingdom, United States of America, Germany, Philippines and others.

Genetic Diversity in Cocoa

Concern over the genetic erosion of crop plants, has led to an increase in plant exploration and in the establishment of numerous gene banks. For cocoa, in order to provide a secure conservation base for the future, a sufficiently wide range of diversity was needed and two international 'base collections' sited in Trinidad and Costa Rica have been designated (Wilde *et al.*, 1991). The objective of such a living collection is to include minimum redundancy and maximise the genetic diversity of a crop species and its wild relatives (Brown and Clegg, 1983).

Enriquez and Soria in 1908 reported that many clones have a common source of origin and thus have similar characteristics which might lead to confusion in identification, especially when they were grown and transported to other places. Even though some clones were morphological distinct, in many cases they have been mis-identified when transported from one place to another. Due to identification errors which were common in almost every experimental station where cocoa clones were

maintained, it was recommended at several cocoa conferences that a catalogue be prepared to facilitate identification of cocoa clones. Those definitions emphasised the need for a systematic evolution and characterisation of the genetic resources presented in a living collection.

Basically three methods have been used to analyse variability in cocoa: morphological traits, protein markers and DNA markers.

Morphological Characteristic

In cocoa populations, variability and relatedness have been based on morphological descriptors to select desirable cocoa plants for breeding programmes. The morphological descriptors include the measurement of leaf, flower, fruit and seed characteristics according to the International Board for Plant Genetic Resources (IBPGR) (Sirju-Charran *et al.*,1991). A number of researchers studied morphological characteristics of cocoa in order to systematically describe the crop (Cuatrecasas, 1964; Chessman, 1944 and Soria, 1970).

At one time the shape of the pod was used as a means for identification of cocoa bean quality. Various shapes were used to name bean *viz*. Angolata, Cundeamor, Amelonado and Calaballo (Cuatrecasas, 1964). However, the relationship of pod shape and quality has not been substantiated. With the exception of Amelonado, these names have largely gone out of use.

Certain morphological characters of pod and bean were used as basic classification of cocoa. Chessman in 1944 proposed three main types, as groups of population in *Theobroma cacao,* namely Criolla, Forestero and Trinitario. The grouping were based on texture and colour of pod rust characters as well as the average of bean number per pod and colour of cotyledon. Later on, Soria in 1970 has divided Criollo into 4 types,

based on geographic distribution. They were Mexican Criollo, Pentagoua or Lagorto Criollo, Nicaraguan Criollo and Colombian Criollo.

Morphological traits have a disadvantage of being influenced by environmental factors and this may not accurately estimate or represent genetic relationships. The effectiveness of using these traits to estimate genetic diversity has been questioned by several workers (Gottlieb, 1977; Brown, 1979). The long generation time of most perennial crops as well as the fact that many of the morphological descriptors can only be assessed at maturity demonstrates the need for the development of new techniques of germplasm assessment.

Protein Markers

The most popular used protein markers are isozymes. Isozymes analysis has become particularly prominent in systematic and evolutionary biology as well as in agronomy (Tanksley and Orton,1983). Isozymes or multiple molecular forms of enzymes, are enzymes that share a common substrate but differ in electrophoretic mobility (Market and Muller, 1959). They are revealed when tissue extracts are subjected to electrophoresis in various types of gels and subsequently submersed in a solution containing enzyme specific stains.

Genetic analysis may indicate that some of the variant electromorphs are encoded by alternate alleles at a single locus, in which case the allelic products are termed allozymes (Prakash *et al.*, 1969). Data retrieved from electrophoretic gels consist of the number and relative mobilities of various enzyme products, which with appropriate genetic analysis, become transformed into single or multilocus genotypes for each individual analysed.