THE ISOLATION AND IDENTIFICATION OF MOLECULAR MARKERS LINKED TO SUGAR PRODUCTION IN PINEAPPLE (*Ananas comosus* var. *comosus*)

THIEN YONG NAM

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITI MALAYSIA SABAH
2006

THE ISOLATION AND IDENTIFICATION OF MOLECULAR MARKERS LINKED TO SUGAR PRODUCTION IN PINEAPPLE (*Ananas comosus* Var. *comosus*)

THIEN YONG NAM

THIS THESIS IS SUBMITTED AS PART OF THE REQUIREMENT TO OBTAIN THE MASTER OF SCIENCE DEGREE WITH HONOURS

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITI MALAYSIA SABAH
2006

BORANG PENGESAHAN STATUS TESIS

JUDUL: **THE ISOLATION AND IDENTIFICATION OF MOLECULAR MARKERS LINKED TO SUGAR PRODUCTION IN PINEAPPLE (*Ananas comosus* var. *comosus*)**

IJAZAH: SARJANA SAINS (BIOTEKNOLOGI TUMBUHAN)

SESI PENGAJIAN: 2004-2006

Saya, THIEN YONG NAM mengaku membenarkan tesis Sarjana ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:

1. Tesis adalah hakmilik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian saya.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. TIDAK TERHAD

Disahkan oleh

(Penulis: THIEN YONG NAM)

ANITA BINTI ARSAD
PUSTAKAWAN KANAN
UNIVERSITI MALAYSIA SABAH

(TANDATANGAN PUSTAKAWAN)

(Penyelia: Dr Vijay Kumar)

Tarikh: 7 Mac 2006

Tarikh: 2/4/07

DECLARATION

The materials in this thesis are original except for quotations, excerpts, summaries and references, which have been duly acknowledged.

THIEN YONG NAM
PS04-013-001
01 March 2007

ACKNOWLEDGEMENT

First and foremost, I would like to express my gratitude to Dr. Vijay Kumar, Assoc. Prof Datin Dr. Mariam Abd Latip and Ms. Teoh Peik Lin (my supervisors) for all their guidance and advice that they gave to me throughout the project. It was them who gave me the opportunity, chance and motivation to discover things and I had not known previously.

Not forgetting, I would also like to thank all the lecturers at the Biotechnology Research Institute (BRI) who have taught me throughout the past two years, especially Dr. Michael Wong and Ms. Chang Pheh Ping who gave me advice on the writing up of this thesis. My gratitude also goes to Prof. Datin Dr. Ann Anton for giving me motivation and training in many fields.

Next, I would like to thank the lab assistants Mdm. Vidarita Maikin and Mr. Richard Dailis for their technical supports in the lab especially when dealing with equipment and chemicals. My appreciation also goes to the administration staff at BRI in particular, Mdm. Shanty Ripah, Ms. Sahati Ambotang, Mr. Zulkifi Muhammad, Mr. Johan Alexander and Mr. Rudy Boliku for their help and advice in the handling of seminars, documentations and other administration tasks that I was responsible for.

Last but not the least, I would like to also thank my fellow colleague especially Ms. Chan Kher Xing and Mr. Kenneth Rodriguezs for helping me to proofread my thesis. I also wish to thank the other postgraduate students, Mr. Awang Muhammad Sagaf Abu

Bakar, Mr. Adrian Ng D'Victor, Mr. Melvin John Kinsuat, Ms. Chelven Lim, Ms. Grace Joy Chin and Mr. Gordon John Thomas for their constant guidance, support and understanding during the times when I was performing my laboratory work together with them in the same laboratories. Lastly, I would like to thank all those who had given me support and motivation during the times when I stumbled across problems, especially during the whole two years project, directly or indirectly.

ABSTRACT

THE ISOLATION AND IDENTIFICATION OF MOLECULAR MARKER LINKED TO SUGAR PRODUCTION IN PINEAPPLE (*Ananas comosus* var. *comosus*)

Most pineapple (*Ananas comosus* var. *comosus*, Family: *Bromelieceae*) varieties produced fruits with low sugar content quality which affects the agronomical and breeding potential of pineapple. Thus, a better knowledge of the genetic and molecular basis underlying fruit quality would benefit pineapple breeding programmes. Isolation and identification are the essential steps for the development of molecular markers to be used in marker assisted selection for plant breeding programmes. Hence, two analyses, the Bulked Line Analysis (BLA) and the Differential Gene Expression Analysis were employed in this study. Genomic DNA (leaf) and total RNA (fruit tissue) were isolated and grouped into two groups "Acid" (<11% Brix) and "Acidless" (>17% Brix) prior to polymerase chain reaction (PCR) and reverse transcription-polymerase chain reaction (RT-PCR). In the DNA study, the Random Amplified Polymorphic DNA (RAPD) assay and the Direct Amplified Length Polymorphisms (DALP) assay were used in conjunction with Bulked Line Analysis to identify sugar-content linked markers in pineapple. The experiments showed no polymorphism was observed between the "Acid" and "Acidless" groups of pineapples. This showed that both the RAPD and the DALP methods were unable to differentiate the desired characteristic within clonally propagated individuals. However with the limited number of primers used, it was difficult to conclude that DNA based markers are not suitable in this particular investigation. In the RNA study, newly synthesized double stranded cDNAs were subjected to PCR using three different sets of primers from Amplified Differential Gene Expression (ADGE), Differential Gene Expression Based Annealing Control Primer analysis (DEG-ACP) and specific markers linked to sugar content from peach and tomato. Differential gene expression analysis indicated five markers linked to the sugar content in pineapple. Three of the markers (M1, M3 and M4) were from ADGE, marker HK1 was from specific primer linked to sugar content and one marker (TYN4) was from DEG. A comparison of the differential expression between the cDNA from the "Acid" and "Acidless" groups of pineapple revealed differences in markers M1, M3, M4, and the markers HK1 and TYN4. These markers can be used in marker assisted selection (MAS) to aid in identifying Quantitative Trait Loci since the majority of quantitative traits in crop plants are controlled by polygenes. Selection of quantitative trait loci for crop improvement such as sweetness in pineapple, will result in significant reduction in the time taken for reducing the productivity of low sugar content pineapples and increasing the productivity of high sugar content pineapples.

ABSTRAK

*Kebanyakan variati Nanas (*Ananas comosus var. comosus*, Keluarga:* Bromelieceae*) menghasilkan buah yang mengandungi kandungan gula yang rendah. Kualiti kandungan gula yang rendah ini boleh menjejaskan agronomi dan potensi pembiakbakaan nanas. Justeru, permahaman yang mendalam terhadap genetik dan molekular mengenai kualiti buah boleh meningkatkan mutu pembiakbakaan nanas. Kaedah pengasingan dan kaedah pengenalpastian merupakan kaedah utama untuk mencari pananda molekular. Penanda molekular ini boleh digunakan sebagai "marker assisted selection" untuk program pembiakbakaan nanas. Dengan itu, dua analisis telah digunakan iaitu, "Bulked Line Analysis (BLA) dan Differential Gene Expression Analysis" dalam kajian ini. Genomik DNA (Daun) dan RNA (Tisu buah) diekstrak dan dikumpulkan ke dalam dua kumpulan iaitu "Acid" (<11%) dan "Acidless" (>17%) sebelum proses tindakbalas berantai polimerase (PCR) dan trankripsi berbalik tindakbalas berantai polimerase (RT-PCR) dilakukan. Dalam kajian DNA, "Random Amplified Polymorphic DNA (RAPD) dan "Direct Amplified Length Polymorphisms (DALP) diamplikasi bersama BLA untuk mengenalpasti penanda berkait dengan penghasilan gula dalam nanas. Keputusan eksperimen tidak menunjukkan perbezaan di antara kumpulan "Acid" and "Acidless". Dengan itu, ia menunjukkan kedua-dua RAPD dan DALP tidak dapat mebezakan perwatakan yang dikehendaki di antara klon. Penggunaan primer yang terhad menunjukkan penggunaan RAPD dan DALP tidak sesuai untuk kajian ini. Dalam Kajian RNA, jujukan cDNA diamplikasi dengan tiga set primer iaitu dari "Amplified Differential Gene Expression (ADGE), Differential Gene Expression Based Annealing Control Primer analysis (DEG-ACP)" dan primer berkait dengan kandungan gula dari "peach" dan tomato. Lima penanda molekular berkait dengan penghasilan gula telah dihasilkan daripada perbezaan antara pengekspresan gen kumpulan "Acid" dan "Acidless". Tiga penanda (M1, M3 and M4) dari ADGE, penanda HK1 dari primer berkait dengan kandungan gula dari "peach" dan tomato. Manakala satu penanda (TYN4) dari (ACP-PCR). Dengan pembangunan penanda ini boleh digunakan sebagai "marker assisted selection (MAS)" untuk membantu mengenalpasti "Quantitative Trait Loci" pada nanas. Ini kerana kebanyakan cirri-ciri kuantitatif pada tumbuhan pertanian adalah dikawal oleh poligen. Dengan mengenalpasti QTL ini boleh digunakan untuk pembaikbakaan tanaman seperti kandungan gula dalam nanas, dengan mengurangkan masa, mengurangkan penghasilan buah kandungan gula rendah dan meningkatkan penghasilan buah nanas dengan kandungan gula tinggi.*

LIST OF SYMBOLS/ABBREVIATIONS

ADGE	Amplified Differential Gene Expression
AFLP	Amplified Fragment Length Polymerase
ACP	Annealing Control Primer
bp	Base pair
BLA	Bulked Line Analysis
BSA	Bulked Segregant Analysis
CO_2	Carbon dioxide
cm	Centimeter
cDNA	Complimentary deoxyribonucleic acid
°C	Degrees Clesius
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleoside triphosphate
DEPC	Diethylpyrocarbonate water
DEG	Differential Expressed Gene
DALP	Direct Amplification of Length Polymorphism
dH_2O	Distilled water
EtBr	Ethidium bromide
EST	Expressed Sequence Tag
Kb	Kilobase
LMW-PEG	Low molecular weight polyethyglycol
mRNA	Messenger ribonucleic acid

µl	Micro liter
ng	Nanogram
%	Percentage
pmol	Pico mol
PCR	Polymerase chain reaction
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
rpm	Revolutions per minute
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
SDS	Sodium dodecyl sulphate
Tris-HCl	Tris hydrochloride
UV	Ultraviolet
V	Volts
w/v	Weight over volume

µl	Micro liter
ng	Nanogram
%	Percentage
pmol	Pico mol
PCR	Polymerase chain reaction
QTL	Quantitative Trait Loci
RAPD	Random Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
rpm	Revolutions per minute
RNA	Ribonucleic Acid
SNP	Single Nucleotide Polymorphism
SDS	Sodium dodecyl sulphate
Tris-HCl	Tris hydrochloride
UV	Ultraviolet
V	Volts
w/v	Weight over volume

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Ananas comosus (Family: *Bromelieceae*), commonly known as pineapple, is an important tropical fruit in Malaysia. The sweet taste of pineapples is one of the important reasons why it is in high demand in the food industry for making jams, canned food products, biscuits and fruit juices. According to the Federal Agricultural Marketing Authority (FAMA, 2006), the average world production of pineapples from 2004-2005 was 4,643,891.5 metric tones which accounts for RM6.89 million of the average value of world export. In terms of the balance of the trade, Malaysia earned about RM 9.25 million, which contributed a lot to the nation's economy (FAMA, 2006). In the state of Sabah, farmers earn about RM 80,000 per hectare from pineapples plantations annually (FAMA, 2006). This indicates that the pineapple is an important contributor to the Malaysian economy. Currently, there are three pineapples cultivars which are renowned for its sweetness in Malaysia, namely Smooth Cayenne (Sarawak), Mauritius and Josapine (a hybrid developed by MARDI).

Pineapples originated from South America but are now widely grown throughout the tropics and the subtropics such as in Hawaii, the Philippines, the Caribbean area, Malaysia, Taiwan, Thailand, Australia, Mexico, Kenya and South Africa (Bartolome *et al.*, 1994). During 1700, the Dutch used the cross breeding technique to improve the quality of the pineapples (Bartolome *et al.,* 1994). However, most pineapple varieties produced fruits with low sugar content quality. Thus, a better knowledge of the genetic and molecular basis underlying fruit quality would benefit pineapple breeding programmes. Development of molecular markers is important for

marker-assisted selection in plant breeding programmes. It is based upon the principle that a gene(s) conferring a trait of interest is linked to an easily identifiable molecular marker. It may be more efficient to select in a breeding program for the marker than for the trait itself. Application of markers to introgression programs can result in a reduction in the number of breeding cycles needed by improving the selection efficiency, particularly at the early stage. Identification for the molecular marker such as Ristriction fragment length polymorphism (RFLP) associated with Rf gene in rice using Bulked Line Analysis (BLA) (Tan *et al.,* 1998), is one of the example how marker can helped in time saving especially in preparation of genetic stock.

Bulked Line Analysis (BLA) is the best methodology to quickly identify the linkages between target trait (sugar content) and genetic markers (Tan *et al.*, 1998). In this study, bulked DNA from low sugar content pineapples are compared to the bulked DNA of high sugar content pineapples by evaluating the differences in the pattern of DNA fragments. This can be done by amplification for DNA with appropriate molecular marker prior of comparison between the groups. Then the fragment predicted to be the gene controlling the production of sweetness trait can be isolated and identified through DNA sequencing. The development of DNA based genetic markers has had a revolutionary impact on plant genetics. With DNA markers such as mitochondrial DNA, Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), Microsatellite, Direct Amplified of Length Polymorphism (DALP), Single Nucleotide Polymorphism (SNP) and Expressed Sequence Tag (EST) markers, it is theoretically possible to observe and exploit the genetic variation in the entire genome (Liu *et al.,* 2004). The application of DNA markers in this study will allow rapid progress in the investigation of the genetic variability and inbreeding, parentage assignments, species and strain identification and the construction of high