CHARACTERIZATION OF A MADS-BOX GENE FROM *MUSA SP.* (PISANG BERANGAN)

JAGDISH KAUR CHAHIL A/P SANTOKH SINGH

SCHOOL OF SCIENCE AND TECHNOLOGY
UNIVERSITI MALAYSIA SABAH
2008

CHARACTERIZATION OF A MADS-BOX GENE FROM *MUSA SP.* (PISANG BERANGAN)

JAGDISH KAUR CHAHIL A/P SANTOKH SINGH

THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE

SCHOOL OF SCIENCE AND TECHNOLOGY
UNIVERSITI MALAYSIA SABAH
2008

UNIVERSITI MALAYSIA SABAH

DECLARATION OF THESIS STATUS FORM [@]

TITLE : **CHARACTERIZATION OF A MADS-BOX GENE FROM *MUSA SP.* (PISANG BERANGAN)**

DEGREE : **MASTER OF SCIENCE**

STUDIES SESSION : **2001- 2008**

I, **JAGDISH KAUR CHAHIL A/P SANTOKH SINGH** declared for this thesis to be kept in Library UMS, Universiti Malaysia Sabah to serve the following purposes:

1. This thesis is a proprietary rights of Universiti Malaysia Sabah
2. Library UMS is permitted to make copies for studies purposes.
3. Library UMS is allowed to make copies as an exchange material among other higher institutions.
4. NOT LIMITED

Declared by,

(Author: JAGDISH KAUR CHAHIL A/P SANTOKH SINGH)

ANITA BINTI ARSAD
PUSTAKAWAN KANAN
UNIVERSITI MALAYSIA SABAH

(Librarian Signature)

(Supervisor: *ASSOCIATE PROFESSOR DR. LEE PING CHIN*)

Date: 9 July 2008

NOTE: [@] Thesis meant is Doctor of Philosophy thesis, Masters by Research thesis or dissertation for Masters by Coursework and Research or First Degree Project Report.

DECLARATION

I hereby declare that the materials in this thesis are my own except for quotations, excerpts, equations, summaries and references, which have been duly acknowledged.

09 JULY 2008

JAGDISH KAUR CHAHIL A/P SANTOKH SINGH
PS2001-001-321

CERTIFICATION

TITLE : **CHARACTERIZATION OF A MADS-BOX GENE FROM *MUSA SP.* (PISANG BERANGAN)**

DEGREE : **MASTER OF SCIENCE**

VIVA DATE : **14TH APRIL 2008**

DECLARED BY

1. **SUPERVISOR**
   (*ASSOCIATE PROFESSOR DR. LEE PING CHIN*)

2. **CO-SUPERVISOR**
   (*DR. DOUGLAS B. FURTEK*)

ACKNOWLEDGEMENT

At one point, completing this dissertation was a loosing battle with countless adversities but I am grateful for the grace of Waheguru that made this dissertation possible.

I am very grateful to the wonderful individuals who moonlighted as my angels during these trying times. I would like to acknowledge my beloved parents, Mr. Santokh S. and Mdm. Abnash K., my loving uncles Mr. Naranjan S. and Dr. Amarjeet S., my understanding grandma Mdm. Bhagwan K. and my brothers, Hardeep, Sandeep and Satwant. A verbal thank you on this paper will never suffice for the financial, spiritual and emotional support given to me when determination was thinning. I will forever be indebted to all of you. Thank you for being my pillars of strength and keeping me in your prayers during my best and worst days, for it's with your prayers and love this dissertation has been possible.

I would like to express my gratitude to my supervisors Dr. Lee Ping Chin and Dr. Douglas Bruce Furtek for the academic and technical advice. Thank you both for the excellent supervision, guidance, patience and understanding. I would also like to thank Universiti Malaysia Sabah (UMS) for the fundamental research grant (B-09-01-01-PR/U013) granted to Dr. Jualang Azlan to fund this project. Thank you to the other Biotechnology program lecturers especially to Dr. Zaleha Aziz as well as to the Biotechnology staff for your advice and support.

A token of appreciation is extended to my fellow comrades, Ms. Fiana Yong, Ms. Helda Souki and Ms. Roseline Baun Ajang for enduring years of joy and frustration together with me. I am also very grateful for having Ms. Helena Biun, Ms. Tan Su Hui, Ms. Roslina Md. Shah, Ms. Norasekin Tamchek, Mr. Awang Mohd Sagaf, Mr. Adrian Ng and Mr. Kenneth Rodriguez as colleagues. I appreciate all the constant academically inclined advice and support that I received from all of you through-out this period.

My gratitude is further extended to En. Matnuddin Dais from the Agricultural Department of Tuaran and Cik Susie who supplied the banana inflorescences used in this study.

I am also taking this opportunity to thank all others whom I have failed to mention their names, who have been involved directly or indirectly in this project.

A million thanks for making this a success.

ABSTRACT

CHARACTERIZATION OF A MADS-BOX GENE FROM MUSA SP. (PISANG BERANGAN)

Homeotic genes are known to play important roles in the development of fruits and flowers. Most of these genes belong to a large family of regulatory genes that have a characteristic DNA binding domain known as the MADS-box. As a pioneer effort in the study of homeotic genes from banana, a full length MADS-box cDNA from an inflorescence of *Musa acuminata* var. Berangan designated as *MADS3* has been successfully isolated and characterized. The approach of this research was by isolating Poly A^+ mRNA from the inflorescence of Pisang Berangan using oligo-dT magnetic beads. First strand cDNA was synthesized using a dT_{18} anchored primer directly onto those beads. A degenerate primer designed by aligning the sequences of the 180bp MADS domain from known MADS-box genes was used as the forward primer to perform a 3′ RACE. This generated a complete coding sequence. To obtain the 5′ untranslated region on the N terminal side of the MADS-box domain a 5′ RACE was carried out. NCBI BLAST analyses were done on these sequences to confirm authenticity to known MADS-box genes. The PHYLIP package was used to further analyze these sequences. The putative 244 amino acid sequence deduced from this research supporting the MADS-box region and the K domain of the corresponding gene suggests that it is a member of the Type II family. *MADS3* belongs to the AGL clade and is a member of the *SEPALLATA* subfamily of MADS-box transcription factors.

ABSTRAK

CHARACTERIZATION OF A MADS-BOX GENE FROM MUSA SP. (PISANG BERANGAN)

Gen homeotik memainkan peranan penting dalam pembesaran serta pembentukan buah dan bunga. Kebanyakkan gen ini tergolong dalam sebuah gugusan besar gen regulatori yang mempunyai ciri pengenalan tersendiri iaitu sebuah domain yang berupaya membendung DNA yang dikenali sebagai MADS-box. Sebagai perintis dalam kajian gen homeotik daripada pisang, kami telah berjaya mengasingkan dan membuat pencirian ke atas satu jujukan lengkap cDNA daripada jantung buah pisang Musa acuminata var. Berangan. cDNA ini dilabelkan sebagai MADS3. Pendekatan yang diambil untuk menyempurnakan disertasi ini adalah dengan mengasingkan Poly A^+ mRNA daripada jantung buah Pisang Berangan menggunakan bebuli magnet oligo-dT. Jujukan cDNA dihasilkan secara terus ke atas bebuli magnet oligo-dT tersebut menggunakan primer dT_{18} yang mempunyai rangkaian bes-bes lain sebagai sauh. Primer 'forward' pula merupakan primer yang direka berdasarkan susunan jujukan-jujukan MADS region daripada gen MADS-box yang diketahui. Langkah ini telah menghasilkan sebuah jujukan pengkod yang lengkap (complete coding sequence). *Bagi memperolehi 5'* untranslated region *di sebelah terminal-N domain MADS-box, 5' RACE telah dibuat. Transkrip lengkap cDNA diperoleh dengan menindih jujukan hasil* 5' RACE *dan* 3' RACE. *Analisis menggunakan NCBI BLAST telah dibuat ke atas jujukan lengkap cDNA ini bagi mengesahkan kesahihannya terhadap jujukan-jujukan gen MADS-box yang diketahui. Pakej PHYLIP turut digunakan untuk analisis lanjutan. Jujukan jangkaan 244 amino asid ini mempunyai MADS-box region dan K domain adalah berasal dari gen MADS Jenis II. MADS3 digolongkan dalam kumpulan AGL dan merupakan sebahagian anggota subfamili SEPALLATA.*

LIST OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

A	Adenine
AG	AGAMOUS
AGL	AGAMOUS-LIKE
amp	ampicillin
AMV	Avian Myelobastosis Virus
AP1	APETALA1
AP2	APETALA2
AP3	APETALA3
ATP	Adenosine Triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
$CaCl_2$	Calcium chloride
CAL	CAULIFLOWER
cDNA	complementary DNA
CDS	Complete coding sequence
CGIAR	Consultative Group on International Agricultural Research
CIRAD	*Centre de coopération internationale en recherché agronomique pour le développement*
CIP	Calf Intestinal Alkaline Phosphatase
dH_2O	Distilled water
ddH_2O	Double distilled water
dNTP	deoxynucleotide Tri-Phosphate
DEPC	Diethylpyrocarbonate
DEF	DEFICIENS
DMF	Dimethylformamide
DNA	Deoxyribonucleic Acid
DNase	Deoxyribonuclease
EDTA	Ethylenediamine tetracetic acid
e.g.	*exemplie gratia* (Latin) - example
et al.	*et alia* (Latin) - and others
FM	Fitch-Margoliash
FOR	Forward (primer)
FUL	FRUITFUL
GSP	Gene specific primer
GLO	GLOBOSA
G	Guanine
IBPGR	International Board for Plant Genetic Resources
INIBAP	International Network for the Improvement of Banana and Plantain
IPGRI	International Plant Genetic Resources Institute
KAc	Potassium acetate
Kg.	Kampung
kg	Kilogram
KOH	Potassium hydroxide
LB media	Luria Bertanni media
LFY	LEAFY
M	Molar

max.	Maximum
Mb	Mega basepair
$MgCl_2$	Magnesium chloride
mRNA	messenger RNA
min.	Minutes
mL	Millilitre
mm	Millimetre
µL	Microlitre
µg	Microgram
MMLV	Moloney Murine Leukimia Virus
MYA	Million years ago
NaCl	Natrium chloride
NaOH	Natrium hydroxide
NCBI	National Center for Biotechnology Information
ng	Nanogram
$(NH_4)_2SO_4$	Ammonium sulphate
nm	Nanometer
OD	Optical density
Oligo(dT)	Oligodeoxythymidylic acid
PHYLIP	Phylogeny inference package.
P1	PISTILLATA
PLE	PLENA
PNAS	Proceedings of the National Academic of Sciences (USA)
PCR	Polymerase Chain Reaction
pg	Pico gram
poly A	Polyadenylated
pos.	Position
psi	Pressure per square inches
RACE	Rapid Amplification of cDNA Ends
REV	Reverse (primer)
RNA	Ribonucleic Acid
RNase	Ribonuclease
rpm	Revolution per minute
rRNA	ribosomal RNA
RTase	Reverse transcriptase
RT-PCR	Reverse Transcription PCR
s	seconds
SDS	Sodium dodecyl sulphate
SEP	SEPALATTA
SHP	SHATTERPROOF
SQUA	SQUAMOSA
T	Thymine
TAP	Tobacco Acid Pyrophosphatase
U	Uracil
UTR	Untranslated Region
w/v	weight/volume
X-gal	5-bromo-4-chloro-3-indolyl- β-D-galactopyranoside

SYMBOLS

~	Approximately
µ	Micro
°C	Degrees Celsius
Ω	Ohm
-	Minus
%	Percent
OD_{600}	Optical density at 600nm
®	Registered trademark
™	Trademark

CHAPTER ONE

INTRODUCTION

1.1 GENERAL INTRODUCTION

In this modern era, we are in a race between growing population and food production. The challenge of feeding a world population increasing by up to 160 people every minute sounds daunting. It is forecasted that, by 2050, the world's population will increase from the current level of 6 billion to more than 8 billion people (Hoisington *et al.*, 1999). Therefore, the International Board for Plant Genetic Resources/International Plant Genetic Resources Institute (IBPGR/IPGRI) defined over 50 priority crops that should be looked into urgently to maintain the continuous food supply to mankind. The boom of Green Biotechnology further enhanced research in agriculture as certain crops such as the starchy bananas and plantains are important in food production (Panis and Swennen, 1995).

Genome projects aimed at identifying genes and sequencing the complete genomes of as many crops as possible for instance of maize, rice, wheat, oat and barley have been initiated as reported by the Genome Project database of the National Center for Biotechnology Information (NCBI). The collection and interpretation of all the new data is important as it provides a powerful platform for gaining insights into the genome of other species to set groundwork to surmount susceptibility to diseases and climacteric changes. Joining this race, in July 2001, the Global *Musa* Genomics Consortium announced its plans to sequence the banana genome (INIBAP, 2001a). The status of this ongoing project is, however undisclosed.

This sequencing project is deemed a good start as Clarke (2001) in his article entitled 'Banana genome unpeeled', painted a depressing perspective of bananas going extinct in years to come. Most commercially available cultivars are seedless triploids thus sterile and lack variety (Siti Hawa, 1999) hence one would assume Clarke's prediction to be true. Although scientists have suggested using female flower primordia to induce *in vitro* flowering and modification of female fertility (Novak,

1992), it is the diseases such as Black Sigatoka and Fusarium wilt (Berrocal-Lobo and Molina, 2007) that are constraints to banana production. However, neither the diseases nor the sterility would bring this crop to extinction (Ploetz, 2003).

Bananas and plantains are major food crops in developing countries as well as important export crop to industrial nations. The Food and Agriculture Organization of the United Nations (FAO) Statistical estimated that total world exports of bananas accounted for 15.5 million tonnes in 2003, making this crop an important staple commodity for many developing countries.

In Malaysia, banana is grown both for the local and export markets. Being non-seasonal, banana cultivation very much support the current National Agricultural Policy (NAP3 1998-2010) which stresses on maximizing land use, increasing private sector involvement along with an increase in farmers' income and export earnings. Two main banana cultivars in our country are 'Berangan' (AAA) (Berangan will be used without inverted commas henceforth) and 'Mas' (AA). The most popular dessert variety with an annual per capita consumption of 2.7kg is 'Mas' (Thai *et al.*, 2005). Berangan is the third most popular cultivar at 0.5kg per person per year but with cultivation covering almost 4000 hectares (Zabedah, 2001) is Malaysia's most exported dessert banana (Thai *et al.*, 2005).

Banana has a unique inflorescence structure. Each tree possesses an inflorescence with either reddish or violet bracts. These bracts are made up of numerous little flowers. Male flowers are located in the superior part of the inflorescence, while the female ones are located below (Simmonds, 1966). Research on homeotic genes has increased the understanding of floral development. The potential of previous research has resurrected wide interest among the scientific society with researchers venturing into economically important plants. Homeotic genes are master regulators of other genes that control the development of segment-specific structures (Zik and Irish, 2003). By studying homeotic mutants in *Arabidopsis thaliana* (Coen and Meyerowitz, 1991), in which the organs of one floral whorl adopt the identity of organs belonging to a different whorl, it has been possible to identify several of the genes responsible for conferring organ identity. Most of the floral organ identity genes encode members of a conserved family of transcription factors, the MADS-box factors, which are extensively found within the plant kingdom

(Messenguy and Dubois, 2003). Therefore, banana being an economically important crop we propose to study a homeotic gene from the banana inflorescence to add to the existing gene-pool.

Transcription factors are proteins that control the first step of gene expression, the transcription of DNA into RNA sequences (Liu *et al.*, 1999a). These proteins modify expression of target genes by binding to regulatory DNA motifs (Martin, 1996). These factors act as key switches in plant development, therefore either inducing or repressing expression of structural genes within pathways (Travers, 1993). Generally, the developmental fate of an organ is controlled by homeotic genes (Benfey and Weigel, 2001). Plant MADS box genes represent a highly conserved multigene family of transcription factors (Shore and Sharrocks, 1995) that play important roles in different phases of plant growth. Unravelling the pathways that regulate how flowers, fruits and seeds develop has significant implications for agriculture.

All plant MADS-box genes studied to date, from more than 20 species, have specific patterns of spatial and temporal expression that are thought to control plant development. The first function associated with MADS-box genes involved defining floral organ or meristem identity (Coen and Meyerowitz, 1991). Subsequently, MADS-box genes that are expressed mainly or wholly outside the flower have been isolated and these genes probably play a wider role in development. Reproductive transition, inflorescence architecture, meristem patterning, and floral organ identity have been studied as distinct research areas in plant science.

Taking the *Arabidopsis* pathway as a model, some of the homeotic genes involved in flowering have been elucidated. *LEAFY* (*LFY*), *APETALA1* (*AP1*) and *CAULIFLOWER* (*CAL*) regulate the identity of floral meristems and subsequently regulate the expression of floral organ identity genes including *APETALA2 (AP2), APETALA3 (AP3), PISTILATTA* (*PI*) and *AGAMOUS* (*AG*). These genes are involved in specifying floral morphogenesis based on the landmark ABC model postulated by Coen and Meyerowitz (1991).

Recently, three related genes of the E class designated as *SEPELLATA 1/ 2/ 3* (*SEP1/ 2/ 3*) were also shown to be required for the activity of the B and C class genes in the control of floral organ formation (Pelaz *et al.*, 2000). Except for *LFY* and

AP2 (Takatsuji, 1998; Goto *et al.*, 2001), all of these genes are members of the MADS-box gene family. The ABC model was later refined and renamed as the ABCDE model with the introduction of the *SEP* class genes (Messenguy and Dubois, 2003).

Contemporary work on floral patterning started with a series of mutants appearing in the inappropriate whorls despite developing as normal floral organs. Hans Stubbe and Maarten Koornneef collected such mutants from *Antirrhinum majus* (garden snapdragon) and *Arabidopsis thaliana* (a mustard relative) respectively (Lohmann and Weigel, 2002). The importance shown by the extensive research on these mutants initiated interest in understanding MADS-box genes which act as regulators of meristem and organ identity. The research was expended to the isolation of the corresponding genes from other species such as in rice (Favaro *et al.*, 2002), maize (Becker *et al.*, 2002), sweetgum (Liu *et al.*, 1999b) and apple (Yao *et al.*, 2001).

1.2 OBJECTIVE

A large number of homeotic genes studied throughout a decade revealed the existence of a conserved region coined MADS after its founding members: MCM1, AGAMOUS, DEFICIENS and SRF (Shore and Sharrocks, 1995). These genes were named MADS-box genes. Although most studies so far had been done on dicotyledonous plants, a smaller percent of this study had been dedicated to MADS-box genes in a few monocotyledons.

The aims of this study are listed as follows:

i) To extract total RNA from inflorescence of Pisang Berangan

ii) To construct and identify cDNA clones of a MADS-box gene expressed in the inflorescence of Pisang Berangan

iii) To characterize the MADS-box gene by phylogenetic study

1.3 RATIONALE OF THIS RESEARCH

Banana has been an extensively researched crop. However, most published work is on banana diseases and pests (Gold *et al.*, 2006; Jaufeerally-Fakim *et al.*, 2006) and fruit biochemistry (Imsabai *et al.*, 2006; Ünal, 2007). Considerable amount of research has also been carried out on tissue culture (Strosse *et al.*, 2006) and the dietary aspects of the fruit (Zhang *et al.*, 2005).

The *Musa* Genome Project has been initiated in 2001 but the status of this project is undisclosed to the public. The database was last checked on 10th April 2008 and we found information stating that only 1% of each Genome A and Genome B have been sequenced. Molecular characterization done in different flowering plants and crops proposed that the conserved characteristics have lasted over a million years suggest great functional importance. Consequently, these MADS-box genes, being conserved throughout vast taxa may unfold some surprising insights into the evolution of banana.

In contrast to the well-studied dicotyledonous plants like *Arabidopsis* and *Antirrhinum*, agronomically important monocotyledon species are a little less emphasized. MADS-box gene studies on cereals such as maize (Schmidt *et al.*, 1993) and rice (Greco *et al.*, 1997; Pelucchi *et al.*, 2002) have been reported but research on monocotyledonous fruits is still at its infancy nevertheless gaining great interest. Although molecular studies have been reported on *Musa* but to date only one abstract was publicly available (Friedman *et al.*, 2007) on MADS-box genes.

Therefore this study was proposed and initiated in 2001. Taking all that into consideration, the main interest of this research is to set a foundation for the study of MADS-box genes in *Musa sp.* with the aim of identifying at least one gene involved in inflorescence development.

CHAPTER TWO

LITERATURE REVIEW

2.1 BANANA

2.1.1 Banana: An Introduction

Bananas are the third world countries fourth most important food crop after rice, wheat and maize in terms of gross value of production. It is an important source of income to millions of poor (Zhang, 2005). This crop is grown in more than 120 countries throughout the tropics and sub-tropics. Annual world production is around 95 million tonnes, of which a third is produced in each of Africa, Asia-Pacific and Latin America (Figure 2.1). Around 87% of all the bananas grown worldwide are produced by small-scale farmers for home consumption and for sale in local or regional markets. Bananas play an important role in poverty alleviation and have been considered as a useful tool to deliver edible vaccines as they are used as staple for 70 million people in Africa (INIBAP, 2002).

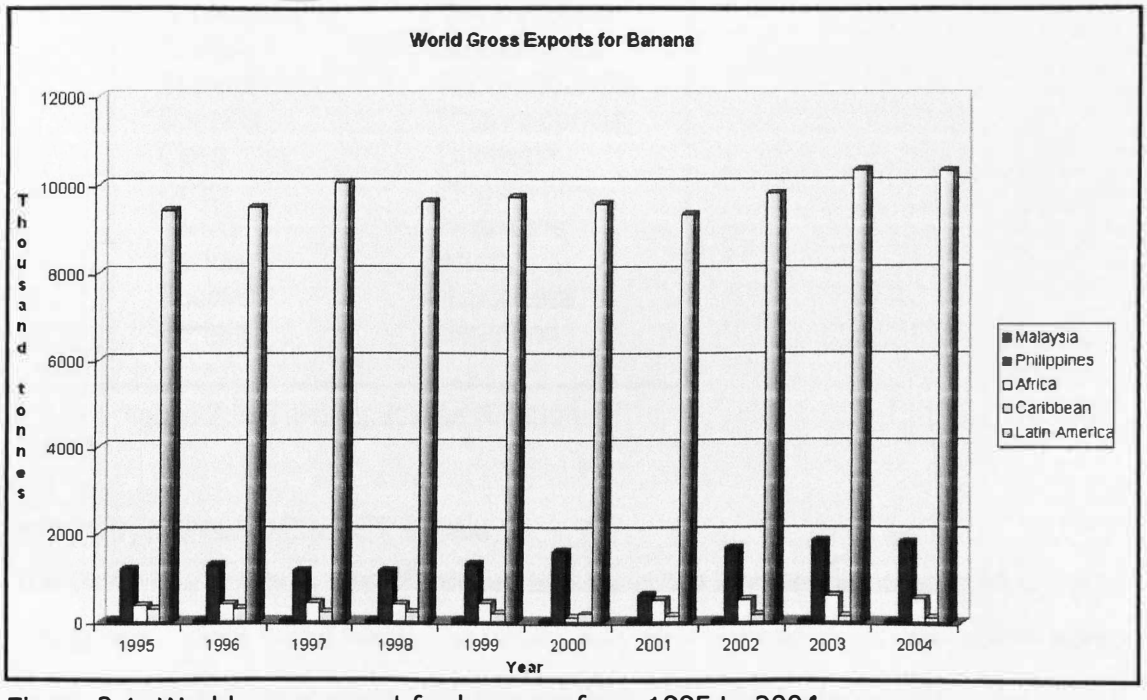

Figure 2.1: World gross export for bananas from 1995 to 2004.
Source: Food and Agriculture Organization (FAO) (2006)

Nutritionally, a single banana finger (approximately 100g) contains 15% of vitamin C, 11% of phosphorus and 16% of the dietary fibre needed each day for good health and contains no fat, sodium or cholesterol (Kahlon and Smith, 2007). Banana also provides 20% of the Recommended Daily Allowance of vitamin B6, which is significant in the synthesis of antibodies in the immune system and an antidote to high stress and anxiety. The banana fruit contains 75% water, 22% carbohydrate, 1% protein, and is rich in vitamin A (Zabedah, 2001). Banana is also the only fruit rich in tryptophan, an amino acid that helps body to create serotonin, a mood enhancer and contains potassium, magnesium, calcium and iron (Wall, 2006).

For developing nations where banana is a source of an annual income, banana puree is made into baby food and ice cream, as well as baked desserts. Flour is derived from dried fruits, and used for pastries or mixed with other flours. Dried fruit of both banana and plantain are commonly made into chips by frying slices in oil and salting. The fermented juices are made into beer and wine, commonly in Africa (INIBAP, 2001d). The young leaves and terminal inflorescence buds are edible.

2.1.2 Taxonomy and Botanical Background

Kingdom	Viridiplantae
Subkingdom	Tracheophyta
Phylum	Streptophyta
Superdivision	Spermatophyta
Division	Magnoliophyta
Class	Liliopsida
Order	Zingiberales
Family	Musaceae
Genus	*Musa*
Species	*Acuminata*
Variant	Berangan

Figure 2.2: Taxonomic classification

a) Systematic Classifications

The Musaceae family is one of the six families in the Zingiberales order and contains two genera, *Musa* and *Ensete*. The genus *Ensete* contributes fibre and edible starch in parts of southern Ethiopia (INIBAP, 2001b). The genus *Musa* (Figure 2.2) has a relatively small haploid genome of 500 to 600 Mb (25% larger than rice). It is one of