SEQUENCING AND CHARACTERIZATION OF MADS-BOX GENE IN OIL PALM (*Elaeis guineensis*)

TAN SU HUI

THESIS SUBMITTED IN FULFILMENT FOR THE DEGREE OF MASTER OF SCIENCE

SCHOOL OF SCIENCE AND TECHNOLOGY
UNIVERSITI MALAYSIA SABAH
2007

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS @

JUDUL : Sequencing and characterization of MADS-box gene in oil palm (*Elaeis guineensis*)

IJAZAH : Sarjana Sains (Bioteknologi Tumbuhan)

SESI PENGAJIAN: 2000-2006

Saya, TAN SU HUI 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 sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. TIDAK TERHAD.

Disahkan oleh

(Penulis: TAN SU HUI)

(TANDATANGAN PUSTAKAWAN)

Alamat Tetap:
17-2-6, LEBUH MAHSURI,
BAYAN BARU,
11950 BAYAN LEPAS.
PULAU PINANG.

(Pengerusi: PROF. DATUK DR. KAMARUZAMAN AMPON)

Tarikh: 9 April 2007

Tarikh: 19/4/07

Ahli-Ahli Jawatankuasa

(Dr Lee Ping Chin)
Tarikh: 25/4/07

(Dr Zaleha Abdul Aziz)
Tarikh: 25/4/07

(Dr Douglas Furtek)
Tarikh: 23 April 2007

CATATAN: @ Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana secara penyelidikan, atau disertasi bagi pengajian secara kerja kursus dan penyelidikan, atau Laporan Projek Sarjana Muda (LPSM).

DECLARATION

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

TAN SU HUI
PS2000-001-220
28 JULY 2006

ACKNOWLEGEMENTS

In my point of view, doing research is like running an off-road track. One can know in which direction wants to go; however, one cannot foresee what will be coming along the way. The varying nature of challenges one meets stimulate ones mind and the passage of each obstacle gives a feeling of satisfaction and adds to the excitement of what is to come.

Now, close to the finishing line of this run, I would like to address my sincere gratitude and appreciation to the persons who have accompanied me along the research and those who have been there by the side to support me.

Special thanks to my supervising committees, Prof. Datuk Dr. Kamaruzaman Ampon as president and the committee members are Dr. Lee Ping Chin and Dr. Zaleha Abd. Aziz (University Malaysia Sabah), and Dr. Douglas Furtek (Malaysian Cocoa Board) for their guidance, supervision, dedicated efforts and patience throughout this research. Thanks for having confidence in me, encouraging me in my steps to become an independent researcher.

I wish to express my thanks to Ministry of Science, Technology and Environment for sponsoring my master degree program research through the scholarship award (National Science Development and Research Council Grant). Acknowledgment also extended to UMS Fundamental Fund for funding the research project. Appreciation is also dedicated to Dr Jualang Azlan Gansau for being the project leader for the UMS Fundamental Fund.

Special thanks to my former supervisor, Prof. Dr. Paranjothy A/K Karthigesus for his earlier guidance and knowledge.

I am grateful to Mr Boris Soon (Agronomist) from Oil Palm Plantation of Borneo Samudera Pvt. Ltd. (Membakut, Sabah) for giving me permission to take oil palm samples at the plantation. I am also grateful to FIRST BASE Laboratories Sdn. Bhd. for performing the sequencing and primer synthesis.

Sincere gratitude is also extended to all laboratory assistants. They are Miss Rokiah Ibrahim and Miss Radizah Darun.

My greatest appreciation to my friends and colleagues: Roslina Mohd. Shah, Helena Biun, Norasekin Bt Tamchek, Jugdish Kaur, Fiana Yong, Roseline Baun Ajang, Cheah Hwen-Yee, Chong Tong Seng, Foo Sek Hin, Hew Chaw Sen, Puah Seok Hwa, Ong Si Mon and Ho Wei Loon for their support, cooperation and friendship.

Last, I would like to direct my deepest thanks to my beloved parents for the loves, invaluable support and encouragement. To my beloved boyfriend, Lum Mok Sam for his loves, support and understanding. To my dearest God sister, Chua Lee Chin for cheering me up all the time.

ABSTRAK

PENJUJUKAN AND PENCIRIAN GEN KOTAK MADS DALAM KELAPA SAWIT (Elaeis guineensis)

Pada tumbuh-tumbuhan, keluarga gen-gen kotak MADS memainkan peranan yang penting dalam perkembangan, termasuk, perkembangan vegetatif, bunga dan buah. Penyelidikan ini memfokuskan pada gen-gen kotak MADS dalam kelapa sawit yang merupakan tananam pertanian yang penting. Untuk tanaman ini, penghasilan buah dan minyak adalah bergantung pada pembungaan. Kajian ini telah memencilkan satu gen kotak MADS yang dinamakan EMADS1 daripada bunga betina kelapa sawit (Elaeis guineensis), melalui kaedah 3' dan 5' RACE. Panjang cDNA EMADS1 adalah 996 bp, yang mengandungi 5' UTR (75 bp), 3' UTR (192 bp) dan ekor poli(A)+. cDNA EMADS1 mengandungi satu kawasan pengekodan yang mengekodkan 242 asid-asid amino. Jujukan asid amino EMADS1 menunjukkan sifat homolog yang tinggi dengan faktor-faktor transkripsi ahli-ahli keluarga gen-gen kotak MADS dan mengandungi satu domain MADS sebanyak 60 asid-asid amino (pada kedudukan 76 – 255), satu domain K sebanyak 67 asid-asid amino (pada kedudukan 346 – 546), kawasan I sebanyak 30 asid-asid amino (pada kedudukan 256 – 345) dan penghujung-C sebanyak 86 asid-asid amino (pada kedudukan 547 – 804) termasuk kodon henti. Jujukan asid amino EMADS1 mengandungi domain MADS jenis seperti MEF2 dan domain K. Ini mencadangkan EMADS1 merupakan gen kotak MADS jenis 'Type II'. Jujukan EMADS1 menunjukkan kesamaan yang tinggi dengan AOM1 (80%) daripada asparagus, OM1 (80%) daripada orkid, LtAGL9 (78%) daripada Liriodendron tulipifera, PaAGL9.2 (83%) daripada avokado, VvMADS4 (79%) daripada pokok anggur, FBP2 (76%) daripada petunia, TM5 (82%) daripada tomato, SEP1 (63%), SEP2 (61%) dan SEP3 (70%) daripada Arabidopsis, dan EgAGL2-2 (100%), EgMADS8 (98%), EgAGL2-1 (97%), EgAGL2-3 (92%), EgAGL2-5 (65%) dan EgAGL2-4 (61%) daripada kelapa sawit. Antara gen-gen kotak MADS SEPALLATA yang dikenalpasti daripada Arabidopsis, jujukan protein EMADS1 menunjukkan kesamaan tertinggi dengan SEP3 (AGL9). Analisis pokok filogenetik menunjukkan EMADS1 tergolong dalam kumpulan AGL2. Dalam kumpulan AGL2, EMADS1 adalah dibahagikan semula kepada subkumpulan AGL9 dan subcabang dengan gen-gen OM1 dan AOM1. Nilai peratus bootstrap pada EMADs1 dan OM1 adalah 75 (neighbor-joining), 48.5 (parsimoni maksimum) dan 36 (kebarangkalian maksimum). Jujukan protein EMADS1 adalah paling berhomolog dengan gen-gen kumpulan AGL2. Analisis pengukapan RT-PCR dengan jumlah RNA ekstrak daripada kudup-kudup bunga jantan dan betina yang kurang matang, serta embrio dalam bernih menunjukkan gen EMADS1 dapat dikesan di kudup-kudup bunga muda jantan dan betina yang kurang matang. Corak ekspresi EMADS1 pada kudup-kudup bunga jantan dan betina yang kurang matang adalah serupa dengan EgAGL2-1, EgAGL2-2, EgAGL2-3, EgAGL2-4, EgAGL2-5 dan EgMADS8 daripada kelapa sawit. Transkrip EMADS1 juga terdapat dalam embrio semasa perkembangan bernih. Keputusan kajian ini mencadangkan bahawa gen EMADS1 bukan saja memainkan peranan penting dalam kawalan perkembangan organ bunga kelapa sawit, tetapi juga penting dalam perkembangan embrio dalam bernih.

Kata kunci: *Elaeis guineensis*; Kelapa sawit; Bunga; Embrio; Kotak MADS; *SEPALLATA*

ABSTRACT

In plants, MADS-box genes family plays a key role in developments, including vegetative, floral and fruit development. This study focused on the MADS-box genes of oil palm, a crop of agricultural importance. In this plant, flowering is directly related to the yield of fruits. In this study, *Elaeis guineensis* MADS-box 1 (*EMADS1*) was isolated from the female flower of oil palm (*Elaeis guineensis*), using the 3' and 5' RACE (Rapid Amplification of cDNA Ends) strategies. The *EMADS1* cDNA is 996 bp long, including a 5' UTR (75 bp), 3' UTR (192 bp) and a poly(A)+ tail. The *EMADS1* cDNA contains a 729 bp coding region encoding 242 amino acids. The deduced amino acid sequence of *EMADS1* indicated high homology with members of the MADS box family of transcription factors and contained a MADS domain of 60 amino acids (at position 76 – 255) and a K domain of 67 amino acids (at position 346 – 546), as well as I region of 30 amino acids (at position 256 – 345) and C-terminal of 86 amino acids (at position 547 – 804) included the stop codon. *EMADS1* amino acids sequence contained the *MEF2*-like MADS domain and the K domain and thus suggested that *EMADS1* is a Type II MADS-box gene. The *EMADS1* protein sequence showed a high sequence similarities to *AOM1* (80%) of asparagus, *OM1* (80%) of orchid, *LtAGL9* (78%) of *Liriodendron tulipifera*, *PaAGL9.2* (83%) of avocado, *VvMADS4* (79%) of grapevine, *FBP2* (76%) of petunia, *TM5* (82%) of tomato, *SEP1* (63%), *SEP2* (61%) and *SEP3* (70%) of *Arabidopsis*, and *EgAGL2-2* (100%), *EgMADS8* (98%), EgAGL2-*1* (97%), *EgAGL2-3* (92%), *EgAGL2-5* (65%), and *EgAGL2-4* (61%) of oil palm. Among the *SEPALLATA* MADS-box genes identified in *Arabidopsis*, *EMADS1* protein sequence showed the highest similarity to *SEP3* (*AGL9*). Phylogenetic tree analysis showed that *EMADS1* fall into the *AGL2* group. In *AGL2* group, *EMADS1* was subdivided to the *AGL9* subgroup and at the same sub-branch with *OM1* and *AOM1* genes. Percentage of bootstrap values at *EMADS1* and *OM1* were 75 (neighbor-joining), 48.5 (maximum parsimony) and 36 (maximum likelihood). *EMADS1* protein sequence was most likely homologous to *AGL2* group genes. RT-PCR analysis using total RNA isolated from immature male and female flower buds, and seed embryo showed that the *EMADS1* was ubiquitously expressed in the immature male and female flower buds. The expression pattern of *EMADS1* in immature male and female flower buds was similar to *EgAGL2-1*, *EgAGL2-2*, *EgAGL2-3*, *EgAGL2-4*, *EgAGL2-5*, and *EgMADS8* of oil palm. The *EMADS1* transcript also accumulated in embryos of the developing seeds. These results suggested that *EMADS1* may not only function in the important events controlling floral organ development in oil palm, but also in embryo development in seeds.

Keywords: *Elaeis guineensis*; Oil palm; Flower; Embryo; MADS-box; *SEPALLATA*

LIST OF CONTENTS

			Page
RESEARCH TITLE			i
DECLARATION			ii
ACKNOWLEDGEMENTS			iii
ABSTRAK			iv
ABSTRACT			v
LIST OF CONTENTS			vi
LIST OF TABLES			xi
LIST OF FIGURES			xii
ABBREVIATIONS			xiv
CHAPTER 1	**INTRODUCTION**		
1.1		Introduction	1
1.2		Aims of the project	4
1.3		Expected outputs of this study	4
CHAPTER 2	**LITERATURE REVIEW**		
2.1		African oil palm – *Elaeis guineensis* Jacq.	5
2.2		Scientific Classification	5
2.3		Origin and history	7
2.4		Botanical description of the oil palm	8
	2.4.1	Plant	8
	2.4.2	Inflorescences	8
	2.4.3	Pollination	9
	2.4.4	Fruit	9
	2.4.5	Mantled	11
	2.4.6	Folklore, medicinal properties, food and non-food usage	12
2.5		Flower development	14
2.6		The making of a flower: Genetic and molecular interactions	15

2.7		Classical ABC model of floral organ identity	16
2.8		Extended ABC model	19
	2.8.1	The Class-D floral identity genes and function	19
	2.8.2	The Class-E genes and function	20
2.9		The quartet model of flower organ identity	23
2.10		The MADS-box gene family	25
2.11		Evolution of the plant MADS-box gene family	27
2.12		Overview: strategies of cloning and analysis of expressed genes	29
2.13		Polymerase chain reaction (PCR)	30
	2.13.1	PCR oligonucleotides primer	32
2.14		Reverse transcriptase mediated PCR (RT-PCR)	33
2.15		Rapid amplification of cDNA ends (RACE)	33
2.16		PCR cloning vector system – TA cloning	34
2.17		DNA Sequencing	35
	2.17.1	The Sanger's dideoxy chain termination sequencing method	35
	2.17.2	Automated DNA sequencing	38
2.18		Analysis of Gene Expression	38
	2.18.1	Reverse transcription PCR analysis	39
CHAPTER 3		**MATERIALS**	
3.1		Chemicals, enzymes, oligonucleotides and commercial kits	40
3.2		Plant materials	42
3.3		Bacterial strain and cloning vector	44
	3.3.1	Bacterial strain	44
	3.3.2	Cloning vector	44
3.4		Strategies to produce an ribonuclease-free environment	46
	3.4.1	Gloves and contact	46
	3.4.2	Glassware, plasticware and disposable items	47
	3.4.3	Reagents	47
3.5		Buffers	47

	3.5.1	pH buffers	48
	3.5.2	Extraction/ lysis buffers and solutions	49
	3.5.3	Electrophoresis buffers	49
	3.5.4	Transformation buffer	50
3.6		Organic reagents	51
	3.6.1	Determining the pH of phenol	52
CHAPTER 4		**METHODOLOGIES**	
4.1		Overview of strategy	53
4.2		Design of MADS domain-specific degenerate primer for 3' RACE	55
4.3		Basic rules for designing primers	57
4.4		Agarose gel electrophoresis	58
4.5		The 3' RACE of MADS-box DNA fragments	59
4.6		Cloning of PCR products and transformation	62
4.7		Preparation of competent cells	62
4.8		Screening of positive colonies and isolation of plasmid	64
	4.8.1	Isolation of plasmid DNA	64
	4.8.2	Plasmid digestion with single restriction enzyme	65
4.9		The 3'-end DNA sequence determination	66
4.10		BLAST search and statistical parameters	66
4.11		Design of MADS-box gene-specific primer for 5' RLM-RACE	68
4.12		The 5' RLM-RACE of MADS-box gene from oil palm	68
4.13		Total RNA Isolation	71
4.14		The 5'-end DNA sequence determination	71
4.15		Sequence assembly	73
4.16		Coding region identification	73
4.17		Databanks	74
4.18		Sequence alignment	74
4.19		Phylogenetic analysis	74
4.20		Reverse Transcription PCR analysis	78

CHAPTER 5 RESULTS AND DISCUSSION

5.1 Designed degenerate oligonucleotide PCR primers and test 79

5.2 Isolation of an oil palm cDNA clone encoding MADS-box protein 81

5.2.1 The 3' RACE of MADS-box cDNA clone 82
5.2.2 The 5' RACE of MADS-box cDNA clone 84

5.3 Screening of positive colonies 86

5.3.1 Transformation efficiency and blue/ white selection of transformants 86
5.3.2 Screening of 3' RACE DNA fragment 87
5.3.3 Screening of 5' RACE DNA fragment 89

5.4 Sequence analysis of Oil palm cDNA clones 90

5.4.1 The 3'-end DNA sequence of cDNA clones 90
5.4.2 The 5'-end DNA sequence of cDNA clones 91
5.5 Sequence assemble of cDNA clones 91

5.6 Primary structure of oil palm MADS-box gene 92

5.7 Conserved domain 95

5.8 Protein sequence comparison 96

5.9 Phylogenetic analysis 99

5.10 Comparison with other *AGL2* group 105

5.11 Comparison to oil palm MADS box genes 106

5.12 RT-PCR analysis of *EMADS1* in different organs 109

CHAPTER 6 CONCLUSION

6.1 General conclusion 110

6.2 Recommendation for future works 112

REFERENCES 113

APPENDIX A 124

APPENDIX B 127

APPENDIX C 128

APPENDIX D 132

APPENDIX E 133

APPENDIX F 134

APPENDIX G 138

APPENDIX H 141

APPENDIX I 143

APPENDIX J 145

APPENDIX K 146

LIST OF TABLES

		Page
Table 2.1	Phenotypic effects of mutations in each of the three functions in *Arabidopsis*	18
Table 2.2	Ectopic expression of ABC and *SEP* genes	21
Table 3.1	List of commercial kits	41
Table 4.1	Maize and rice MADS-box genes used to design degenerate PCR primers	55
Table 4.2	Degenerate genetic code	56
Table 4.3	Range of separation in gels containing different amounts of agarose	59
Table 4.4	PCR reagents and their final concentration per reaction	61
Table 4.5	Single ligation reaction	62
Table 4.6	Plasmid DNA digestion per reaction	65
Table 4.7	Several BLAST homology search programs developed at NCBI	66
Table 5.1	The number of blue, white and light blue colonies obtained from the blue/white selection of 3' and 5' RACE transformants.	87
Table 5.2	Percentage of identity of MADS and K domains among *EMADS1* and other plant MADS-box proteins.	99
Table 5.3	The *AGL2* group genes	105
Table 5.4	Summary of thirteen oil palm MADS-box genes identified and characterized at CIRAD/IRD Palm Developmental Biology Laboratory, France by Adam and colleagues (2006)	107
Table 5.5	Overview and annotation of isolated oil palm MADS box genes	108

LIST OF FIGURES

		Page
Figure 2.1	Composition of *Elaeis guineensis* (African oil palm)	6
Figure 2.2	Fruit forms of the oil palm	10
Figure 2.3	Abnormal male inflorescence	11
Figure 2.4	Cross-sections of oil palm fruits originating from somaclones	12
Figure 2.5	Basic structure of a typical flower	15
Figure 2.6	The four whorl regions of a floral primodium	16
Figure 2.7	The classical ABC model for flower organ identity in *Arabidopsis*	17
Figure 2.8	The ABCD model	19
Figure 2.9	The ABCDE model of floral organ identity in *Arabidopsis*	22
Figure 2.10	Flower structure and the putative quaternary complexes bound to a target DNA sequence.	24
Figure 2.11	Schematic representation of the protein domains of plant type I (SRF-like) and type II (MEF2-like) MADS-domain proteins	26
Figure 2.12	Polymerase chain reaction (PCR)	31
Figure 2.13	TA cloning	34
Figure 2.14	Diagram of a dideoxynucleotide	36
Figure 2.15	Chain termination sequencing of DNA: Sanger method	37
Figure 3.1	Oil palm immature male inflorescence	42
Figure 3.2	Oil palm immature female inflorescence	43
Figure 3.3	Young oil palm fruit (a) and seed embryo (b)	43
Figure 3.4	The promoter and multiple cloning sequence of the pGEM®-TEasy Vectors	45
Figure 3.5	pGEM®-T Easy Vector circle map and sequence reference points	45
Figure 4.1	Flow chart of the overall strategy	54
Figure 4.2	The schematic diagram of two degenerate PCR primers location at the highly conserved portions of the MADS domain amino acid sequences MG/VRGRV/IEL/IK and EVALII/VFS.	55
Figure 4.3	Chemical structure of agarose	58
Figure 4.4	Overview of the 3' RACE procedure	60
Figure 4.5	The schematic diagram of SH8 reverse primer	68

Figure 4.6	Overview of the 5' RLM-RACE procedure	69
Figure 5.1	PCR amplification products	81
Figure 5.2	The 3' RACE on oil palm MADS-box gene	83
Figure 5.3	Total RNA of the flower of immature female inflorescences of oil palm	84
Figure 5.4	The 5' RACE on oil palm MADS-box gene	85
Figure 5.5	The 3' RACE DNA fragment screening	88
Figure 5.6	The 5' RACE DNA fragment screening	89
Figure 5.7	GenBank flatfile	93
Figure 5.8	Nucleotide and deduced amino acid sequences of *EMADS1* cDNA	94
Figure 5.9	NCBI conserved domain search result for *EMADS1*	95
Figure 5.10	Alignment of deduced amino acid sequence of *EMADS1* with other plant MADS-box proteins	98
Figure 5.11	Neighbor-joining tree illustrating relationship among *EMADS1* and other representative MADS-box genes	101
Figure 5.12	Phylogentic tree derived from parsimony analysis.	103
Figure 5.13	Phylogenetic tree derived from maximum likelihood analysis.	104
Figure 5.14	RT-PCR analysis of *EMADS1* genes in various organs in oil palm	109

ABBREVIATIONS

-	minus
%	percent
:	ratio
≈	is approximately equal to
'	prime
®	registered as trademark
°C	degree Celsius
α	alpha
β	beta
µg	micro gram
µl	micro liter
µM	micromolar
bp	base pair
cfu	colony forming units
cm	centimeter
cm^2	square centimeter
ddATP	dideoxyadenosine triphosphate
ddCTP	dideoxycytosine triphosphate
ddGTP	dideoxyguanine triphosphate
ddTTP	dideoxythiamine triphosphate
DEPC	diethyl pyrocarbonate
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
e.g.	exempligratia
EDTA	ethylenediaminetetraacetic acid
et al.	et alia
ft	foot/ feet
g	gram
Inc.	Incorporated
IPTG	isopropyl-1-thio-β-D-galactoside
kb	kilobase
kg	kilogram
m	meter

M	molarity
mg	milligram
ml	milliliter
mM	millimolar
mm	millimeter
MΩ	mega ohm
ng	nanogram
nm	nanometer
OD_{600}	optical density at 600 nm
oligo(dT)	oligodeoxythymidylic acid
pH	a number from 0 to 14 that describes how acid or alkaline a substance is
PIPES	piperazine-*N*,*N*'-bis(2-ethanesulfonic acid)
poly(A)+	polyadenylated mRNA
psi	pound per square inch
Pvt. Ltd.	Private Limited
RLM-RACE	RNA ligase-mediated RACE
RM	Ringgit Malaysia
RNA	ribonucleic acid
RNase A	ribonuclease A
SDS	Sodium Dodecyl Sulfate
Taq	*Thermus aquaticus* DNA polymerase
TE	tris/ EDTA
™	trademark
Tris	tris(hydroxymethy)aminomethane
Tris-Cl	tris hydrochloride
UV	ultraviolet
V	volts
w/v	weight per volume
x	times
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

CHAPTER 1

INTRODUCTION

1.1 Introduction

Tissue culture propagated oil palm produces a higher incidence of flowering abnormalities called mantling. This inhibited the use of this potentially powerful method of multiplication. The abnormality may occur in either male or female inflorescences or in both (Corley *et al.*, 1986). In the flower of the male inflorescence, the whorl of stamen primordium tissues develops aberrantly into carpel-like structures resulting in absence of pollen. In the flower of the female inflorescence, the central whorl that contains the ovary does not develop; instead develops a ring of stamen primordium tissues to form carpelloid structures surrounding the fruit. This phenotype leads to fruits subtended by a ring of fruitlets. Mantled fruits are known to be failed to sustain development of the bunch to maturity and ripeness, and thus directly affecting oil production (Rao & Donoughs, 1990).

Numerous genes that control the development of various organs in plants and animals have been isolated over a decade (Alvarez-Buylla, *et al.* 2000b; De Bodt, *et al.*, 2003). These genes have been termed homeotic genes, as they are involved in the spatial arrangement of cells and tissues within the organism. A homeotic gene is a gene that plays a role in determining a tissue's identity during development. A homeotic mutation causes the cells to misinterpret their position in the blueprint and become normal organs in inappropriate positions, thus alter the overall body plan. In other words, homeotic mutations cause the conversion of one organ into another and this

distinctive phenotype implicates the genes in the control of organ function. Therefore, the control of flower development by homeotic genes is very important.

Homeotic genes have been conserved throughout evolution, reflecting their fundamental importance in developmental regulation (Jordan & Anthony, 1993). The practical relevance of the homeotic genes comes from the opportunities that allow changes in the number and size of organs. A simple example is broccoli (hardy variety of cauliflower with greenish flower heads) where floral development continues almost indefinitely to create an enormous inflorescence.

Many of these homeotic genes act as transcription factors, regulating developmental cascades unique to each organ. Most of these homeotic genes belong to a large family of regulatory genes and possess a characteristic of DNA binding domain known as the MADS-box. The MADS-box is a highly conserved domain of 56 to 60 amino acids, named after four of the originally cloned members: MCM1, AGAMOUS, DEFICIENS and SFR (Shore & Sharrocks, 1995).

In plants, MADS-box genes have been shown to play fundamental roles in flower development by controlling floral meristem and floral organ identity (Coen & Meyerowitz, 1991; Davies & Schwarz-Sommer, 1994). Additionally, MADS-box genes also regulate fruits, leaves, roots and embryos development. The role of floral organ identity genes in model systems, such as *Arabidopsis thaliana*, *Antirrhinum majus* and *Petunia hybrida Vilm.,* are well studied and characterized. The floral organ identity genes can be subdivided into four different classes, termed A-, B-, C- and D-function genes, whose members provide four different homeotic functions, with A specifying the formation of sepals, A+B, petals; B+C, stamens; C, carpels; and D, ovules (Weigel & Meyerowitz,1994; Angenent & Columbo, 1996). In 2000, the

ABCD model has been amended to the ABCDE model, in which an E-function gene is required for the identity of the three inner whorls and possibly also for ovule identity (Pelaz *et al.*, 2000; Theissen & Saedler, 2001).

According to this amended model, which is called the quartet model, the combinatorial tetramers of four classes of floral MADS-domain proteins regulate the development of the four floral components (Honma & Goto 2001; Theissen 2001). Sepals are specified by class-A genes; petals by class A, B, and E genes; stamens by class B, C, and E genes, and carpels by class C and E genes. Class A, C, and E genes are also involved in floral meristem development.

The palaeontological record shows flowering plants have evolved and it appears that the essential mechanisms underlying flower development are very similar among different species. Gene sequences closely related to *AGAMOUS*, for example, have been cloned from tobacco and tomato (Meyerowitz, 1994). This homology enables us to use the available cloned genes from *Arabidopsis* and *Antirrhinum* to study floral development in oil palm. However, it is important to recognize that the same genes may not have the same function in different species.

In this study, it would be useful to focus attention on crops of agricultural importance where flowering is directly related to yield – oil palm. Oil palm is chosen in this project because it is the most important agricultural crop in Malaysia. Malaysia is the largest exporter and producer of palm oil in the world. Oil palm samples are available throughout the year. Therefore, the opportunities to study the role of MADS-box genes in controlling diverse developmental processes ranging from shoot and root development to flower and fruit development are vast.

1.2 Aims of the project

- To isolate and clone a MADS-box cDNA expressed in the flowers of oil palm.
- To obtain a full-length cDNA of MADS-box gene in oil palm.
- To investigate the evolutionary relationship of *EMADS* with other plants MADS-box gene.
- To compare the transcript accumulation of MADS-box gene in other organs in oil palm.

1.3 Expected outputs of this study

a) Isolation of an oil palm MADS-box gene that is potentially involved in floral development and cloning into a vector.

b) Determination of the nucleotide and deduced amino acids sequences of the oil palm MADS-box cDNA.

c) Identification of the MADS-box gene sequence relatedness using sequence alignment and phylogenetic analysis.

d) Comparison of the transcript accumulation of this MADS-box gene in other organs.

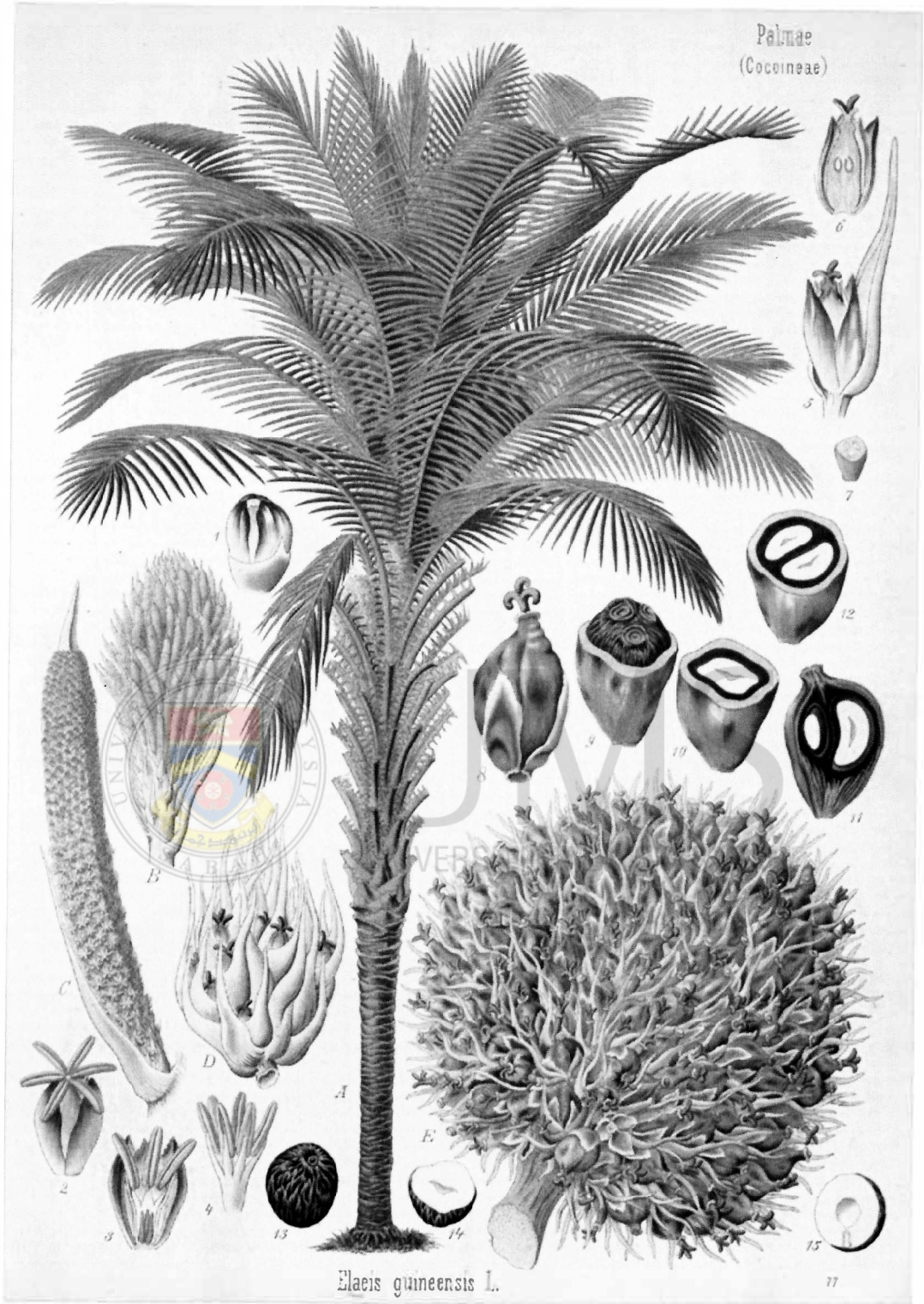

Figure 2.1: Composition of *Elaeis guineensis* (African oil palm). A oil palm tree. B male inflorescence. C male spikelet. D female inflorescence. E fruit bunch. 1 bud; 2 male flower; 3 that in the along cut; 4 staminaltubus; 5 female flower; 6 that in the along cut; 7 ovary in cross section; 8 fruit; 9 that after the skin and mesocarp are removed; 10 fruit in cross section; 11 that in the along cut; 12 a sweisamiger fruit; 13 seed; 14 that in the in cross section; 15 that in the along cut. (Köhler, 1887).

2.3 Origin and history

The African oil palm, *Elaeis guineensis* Jacq., is a native to the tropical rain forest region of West Africa. It was introduced to Malaysia through the Botanical Gardens, Singapore in 1870 as an ornamental plant. The first commercial planting in Malaysia began in 1917 at Tennamaran Estate in Kuala Selangor. Since 1961 most of the planting materials were the Tenera (Deli Dura x Pisifera) which has become the sole commercial planting material in Malaysia (Tan, 1987).

Today, oil palm is the leading agricultural crop in Malaysia, 3.87 million hectares of which over two million are in Peninsular Malaysia and the rest is in the East Malaysian states of Sabah and Sarawak (Appendix A). Production is divided between large estates managed by publicly listed companies, smaller independent estates, independent smallholders and government smallholder settler schemes.

Malaysia is not only the world's largest producer and exporter of palm oil, but is also the biggest exporters of oils and fats in the world (Appendix A). The palm oil industry continues to be an important foreign exchange earner for the country. According to Malaysian External Trade Statistics (June, 2006) palm oil industry is the third largest export revenue earner, accounting for 5.0% or RM14.1 billion of total exports. Oil palm is one of the economically most important oil-bearing crops, being by far the highest oil yielder per unit land area in the world. In 2005, palm oil production almost equaled that of soybean, which had been the number one oil crop for many years (Appendix A).

2.4 Botanical description of the oil palm

2.4.1 Plant

Oil palm is monocotyledons with a stout and straight trunk of about 30 to 60 cm in diameter, which allows oil palm to grow to 15 – 30 m in height (Bricks, 2000). In plantations, the palms are cut down before they reach such a height, as harvesting becomes more and more difficult with increase in height. A mature palm may carry a crown of 25 – 40 leaves, producing 24 – 26 leaves per year (Ng, 1980). Leaves (fronds) are 7 – 8 m in length, with 150 – 250 leaflets per leaf and about 1 – 1.5 m long. Leaflets are 3 – 5 cm wide at mid point and 80 – 120 cm in length (Ng *et al.*, 2003). In commercial oil palms leaves are pruned off when bunches are harvested.

2.4.2 Inflorescences

The oil palm is monoecious (male and female flowers occur separately on the same plant), producing male and female inflorescences in leaf axils (Ng, 1980). The male and female flowers mature in a sequence such that the oil palm is obligatorily cross-pollinated. The inflorescence of both sexes is a compound spike or spadix carried on a stout peduncle 30 – 45 cm in length and enclosed in a woody spathe or bract that splits 2 or 3 weeks prior to anthesis. The spikelets are arranged in a spiral on a central axis (Ng *et al.*, 2003).

The female flower has an off-white appearance and consists of a perianth (outer part of flower) of six segments in two whorls, a tricapellate (rarely 4 or 5 carpels) ovary (each carpel with a single ovule) and a trifid stigma (Ng, 1980; Tan, 1987; Ng *et al.*, 2003). The receptive faces of the three stigma lobes remain pressed on each other when young, but open out when mature (Ng *et al.*, 2003). The stigma

persists in the mature fruit as a hard, three-pointed structure. There may be 100 - 300 spikelets and over two thousand individual flowers (Ng *et al.*, 2003).

The male inflorescence is borne on a long peduncle and consists of long finger-like cylindrical spikelets, each comprising 700 - 1,200 male flowers (Ng, 1980; Ng *et al.*, 2003). The male flower has a fuzzy, bronze or gray appearance and has six perianth parts, and a tubular androecium with six stamens (Ng, 1980; Ng *et al.*, 2003). Flowers begin to open from the base of the spikelet.

2.4.3 Pollination

Oil palms were originally thought to be wind pollinated, but more evidence suggests that they are primarily insect pollinated. *Thrips hawaiiensis* once played an important role in the oil palm pollination in Malaysia (Tan, 1987). However, this native species was not an efficient pollinator. In 1981, *Elaeidobius kamerunicus* (weevils) was released into the oil palm plantations in Malaysia (Tan, 1987; Ng *et al.*, 2003). The insects are attracted to the male inflorescences where they forage for pollen and lay eggs, and move to female flowers by accident since the female flowers produce the same scent as male flowers.

2.4.4 Fruit

The oil palm fruits are drupes. Fruits range in size from 2 to 7 cm and vary in shape from nearly spherical to ovoid or elongated (Ng, 1980; Ng *et al.*, 2003). It consists of a thin exocarp (skin), an oily mesocarp, a hard stony endocarp (shell), a large endosperm (kernel) and one to three small embryos. The shell, kernel and embryo together form the opaque white seed. The mesocarp, from which palm oil is derived, is fibrous and oily, and palm kernel oil is derived from seeds.