



**UNIVERSITI PUTRA MALAYSIA**

**A CATALOGUE OF OIL PALM EXPRESSED SEQUENCE TAGS  
(ESTS) FROM A FLORAL CDNA LIBRARY AND THE ANALYSIS OF  
TWO MADS BOX GENES**

**KWAN YEN YEN.**

**FSMB 2004 13**



**A CATALOGUE OF OIL PALM EXPRESSED SEQUENCE TAGS (ESTs)  
FROM A FLORAL cDNA LIBRARY AND THE ANALYSIS OF TWO MADS  
BOX GENES**

**By**

**KWAN YEN YEN**

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

**July 2004**



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

**A CATALOGUE OF OIL PALM EXPRESSED SEQUENCE TAGS (ESTs)  
FROM A FLORAL cDNA LIBRARY AND THE ANALYSIS OF TWO MADS  
BOX GENES**

By

**KWAN YEN YEN**

**July 2004**

**Chairman : Ho Chai Ling, Ph.D.**

**Faculty : Food Science and Biotechnology**

Flowering is a fundamental process in plant development. The transition of flowering, the first step in flower development, is triggered by a number of environmental and endogenous signals. A variety of genes combined with external and internal cues are involved in a series of biochemical and physiological changes leading to floral induction. The commercial importance of oil palm is dependent on its fruit production. Hence, flowering plays a key role in the oil palm production as the flower is an important introductory step to fruit formation. This study was conducted to isolate sufficiently large numbers of expressed sequence tags (ESTs) from floral tissue in order to obtain gene expression information on a genome wide scale. In addition, EST also provides a valuable resource of candidate genes that can be selected for further study. A random EST approach may result in the isolation of clones representing highly repeated transcripts in the floral tissue whereas low abundance transcripts may



be absent. Hence, a cold plaque screening technique was employed in this study to reduce the prevalence of highly abundant transcripts as well as to facilitate isolation of low abundance transcripts. In plants, many low abundance genes encode transcription factors or proteins that are involved in signal transduction. A total of 1300 clones were isolated through cold plaque screening. Ninety-six percent of the ESTs showed homology to known sequences in the database. The remaining 4% showed no matches with anything in the database, indicating that these genes are plant-specific. Sequence analysis revealed that a large proportion of the ESTs (about 26%) encoded unknown proteins with no known functional definition in the database. The remaining ESTs were further classified into 12 groups according to their encoded putative functions. MADS-box genes are known to be important for the development of flowers and fruit. Two MADS-box cDNAs, 8-3 and 9-32, were cloned from oil palm flowers. Both 8-3 and 9-32 showed high sequence similarities to *AP1*-like MADS-box protein of rice and *SEP3* from *Arabidopsis* respectively. Clone 8-3 was expressed predominantly in the bract, spikelet primordia and the inner two whorls of the floral primordium. The second gene, 9-32, was expressed in flower primordia and the inner three whorls of the floral primordium. The gene expression patterns suggest that these genes have a role in regulating oil palm flower development. Oil palm is a monoecious plant that produces unisexual flowers by suppression of the development of either male or female organs in a particular whorl. Clone 9-32, which is expressed in the early stage of flowering has a different expression profile in male and female flowers.



The onset of 9-32 expression in the male flower is much later than in the female flower, indicating that it may be involved in sex differentiation in oil palm flowers.



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

**KATALOG TAG JUJUKAN TEREKSPRES KELAPA SAWIT DARI  
PERPUSTAKAAN cDNA FLORAL DAN PENGANALISAAN DUA GEN  
BERKOTAK MADS**

Oleh

**KWAN YEN YEN**

**Julai 2004**

**Pengerusi : Ho Chai Ling, Ph.D.**  
**Fakulti : Sains Makanan dan Bioteknologi**

Pembungaan adalah suatu proses yang fundamental dalam perkembangan tumbuhan. Peralihan kepada pembungaan merupakan langkah pertama di dalam perkembangan bunga, dicetuskan oleh sebilangan isyarat dari persekitaran dan endogen. Pelbagai gen yang diperkombinasikan dengan isyarat luaran dan dalaman terlibat di dalam suatu siri perubahan biokimia dan fisiologi yang menuju kepada induksi floral. Kepentingan komersial kelapa sawit adalah bergantung kepada penghasilan buah. Oleh itu, pembungaan memainkan peranan yang penting di dalam pengeluaran kelapa sawit kerana bunga adalah langkah pertama bagi penghasilan buah. Kajian ini dijalankan untuk mengasingkan sejumlah tag jujukan terekspres yang secukupnya dari tisu floral untuk memperolehi maklumat pengekspresan gen dalam skala-luas genomik. Tambahan pula, tag jujukan terekspres juga menyediakan sumber yang berharga bagi memperbolehkan calon gen dipilih untuk kajian seterusnya.

Kaedah tag jujukan terekspres secara rawak membolehkan pemencilan klon yang memperlihatkan kehadiran pengulangan klon-klon yang tinggi di dalam tisu floral manakala kelebaran transkripsi yang rendah mungkin tidak hadir. Oleh itu, kaedah penyaringan plak sejuk digunakan dalam kajian ini untuk mengurangkan kelaziman kelebaran transkripsi tinggi di samping memudahkan pemencilan kelebaran transkripsi rendah. Dalam tumbuhan, kebanyakan gen berkelebaran rendah mengkodkan factor-faktor transkripsi atau protein yang terlibat dalam isyarat transduksi. Sejumlah 1300 klon telah dipencilkan melalui penyaringan plak sejuk. Sembilan puluh enam peratus daripada tag jujukan terekspres menunjukkan homologi kepada jujukan yang tersimpan di dalam pangkalan data. Baki 4% menunjukkan tiada padanan di dalam pangkalan data, menunjukkan gen-gen tersebut adalah spesifik kepada tumbuhan. Penganalisan jujukan memperlihatkan sebahagian besar tag jujukan terekspres (hampir 26%) mengkodkan protein yang tidak mempunyai definisi fungsi yang diketahui di dalam pangkalan data. Lebihan tag jujukan terekspres seterusnya diklasifikasikan kepada 12 kumpulan selaras dengan pengekodan fungsi yang diramalkan. Pengetahuan mengenai kepentingan gen berkotak MADS dalam perkembangan bunga dan buah telah diperakui. Dua cDNA berkotak MADS, 8-3 dan 9-32, telah diklonkan. Kedua-dua 8-3 dan 9-32 masing-masing menunjukkan persamaan jujukan yang tinggi kepada protein berkotak MADS *AP1* dari padi dan *SEP3* dari *Arabidopsis*. Klon 8-3 diekspres di 'bract', primordia 'spikelet' dan lengkaran dua luaran primordium floral. Gen yang kedua, klon 9-32, diekspres di primordia bunga dan lengkaran tiga

dalam primordium floral. Corak pengekspresan gen mencadangkan bahawa kedua-dua gen tersebut memainkan peranan dalam pengawalan perkembangan bunga kelapa sawit. Kelapa sawit merupakan tumbuhan monoecius yang menghasilkan bunga uniseksual melalui penindasan perkembangan salah satu organ pembiakan samada jantan atau betina di dalam lengkaran yang khusus. Klon 9-32 yang diekspres pada peringkat awal pembungaan, mempunyai profil pengekspresan yang berbeza di dalam bunga jantan dan betina. Titik permulaan pengekspresan 9-32 di dalam bunga jantan adalah lebih lambat berbanding dengan bunga betina, menunjukkan 9-32 mungkin terlibat dalam perbezaan jantina di dalam bunga kelapa sawit.





## ACKNOWLEDGEMENTS

I would like to thank Associate Professor Dr. Harikrishna for his guidance, patience, advice and understanding throughout this project. Besides that, I would also like to express my appreciation to Dr. Sharifah Shahrul Rabiah Syed Alwee and Dr. Ho Chai Ling, my committee members for their guidance and advice.

I am thankful to all the members of Genetic Lab (UPM) and Tissue Culture Lab (MPOB). Special thanks are extended to Chooi Mei Choi, Ooi Siew Eng, Zaidah Rahmat, Lee Yang Ping and Choong Chieh Wen. I would not have completed my project without their help and guidance.

Lastly, my deepest gratitude to my family for their love and support.



## TABLE OF CONTENTS

	<b>Page</b>
ABSTRACT	II
ABSTRAK	V
ACKNOWLEDGEMENTS	VIII
APPROVAL	IX
DECLARATION	XI
LIST OF TABLES	XV
LIST OF FIGURES	XVI
LIST OF ABBREVIATIONS	XVIII
<b>CHAPTER</b>	
<b>1 INTRODUCTION</b>	<b>1</b>
<b>2 LITERATURE REVIEW</b>	<b>5</b>
2.1 Oil Palm	5
2.1.1 Oil Palm Flowers	6
2.1.2 Female Inflorescence and Flower	6
2.1.3 Male Inflorescence and Flower	7
2.2 Flowering	10
2.2.1 Induction of Flowering	11
2.2.2 Flower Development	14
2.2.3 Oil Palm Flower Development	16
2.2.4 MADS-Box Gene	18
2.2.5 Meristem Identity Gene	21
2.2.6 Organ Identity Gene	24
2.3 Floral Abnormality	29
2.4 Expressed Sequence Tag	30
<b>3 MATERIALS AND METHODS</b>	<b>33</b>
3.1 Plant Materials	33
3.2 RNA Extraction	33
3.3 Construction of the Ovule cDNA Library	36
3.3.1 Purification of mRNA	36
3.3.2 Synthesis of cDNAs from mRNAs	36
3.3.3 Ligation of cDNAs to vector	37
3.3.4 Generation of a cDNA library	38
3.4 Screening the Library	39
3.4.1 Probe Preparation	39
3.4.2 Plaque Lift	41
3.4.3 Hybridisation and Chemiluminescent Detection	42



3.4.4	Isolation of Putative Clones	43
3.4.5	Single Clone <i>In Vivo</i> Excision	44
3.4.6	Plasmid minipreparation	45
3.4.7	Polymerase Chain Reaction	46
3.4.8	Sequence Analysis	47
3.5	Reverse Transcription–Polymerase Chain Reaction (RT-PCR)	49
3.6	Southern and Northern Analysis	50
3.6.1	Southern Blotting	51
3.6.2	Northern Blotting	53
3.6.3	Radiolabelling of DNA Probe	53
3.6.4	Southern and Northern Hybridisation	53
3.7	Semi-Quantitative RT-PCR Analysis	54
3.7.1	First-Strand cDNA Synthesis	54
3.7.2	Quantification of cDNA Synthesised	55
3.7.3	PCR Analysis with Gene Specific Primers	56
3.8	<i>In situ</i> RNA hybridisation	57
3.8.1	Probe preparation	58
3.8.2	Fixation of Materials on Slides	59
3.8.3	Prehybridisation and Hybridisation	60
3.8.4	Post-Hybridisation Steps	61
3.8.5	Immunological Detection	61
4	RESULTS	63
4.1	Construction of cDNA Library	63
4.1.1	RNA Extraction	63
4.1.2	Quality of the Constructed cDNA Library	64
4.2	Library Screening	66
4.2.1	Computational Analysis of ESTs	70
4.2.2	Classification of ESTs	72
4.3	Isolation of MADS-Box Genes	74
4.3.1	Sequence Analysis	75
4.3.2	Comparison Analysis	79
4.3.3	Genomic Southern Analysis	88
4.3.4	Expression Studies of MADS-Box Clones	89
4.3.4.1	Northern Blot Hybridization	89
4.3.4.2	Semi-quantitative RT-PCR	92
4.3.4.3	<i>In situ</i> hybridisation	97
4.3.4.3.1	Expression Pattern of 8-3	98
4.3.4.3.2	Expression Pattern of 9-32	100
5	DISCUSSION	102
5.1	Cold Plaque Screening	102
5.2	Expressed Sequence Tag	104
5.3	Classification of the Putative Function of Oil Palm Floral ESTs	106



5.4	Gene Expression in Female Flower	109
5.5	The Possible Function of MADS-Box Like Genes 8-3 and 9-32	120
6	CONCLUSIONS	125
	BIBLIOGRAPHY	127
	APPENDICES	135
	BIODATA OF THE AUTHOR	



## LIST OF TABLES

<b>Table</b>		<b>Page</b>
1	Summary of the cDNA clones selected for screening for screening for full-length MADS-box genes	71
2	Summary of randomly selected cDNA clones generated by M16-2 probe	73
3	Summary of randomly selected cDNA clones generated by S49 probe	74



## LIST OF FIGURES

Figure		Page
1	Female flower of the oil palm	8
2	Male flower of the oil palm	9
3	A simplified depiction of the genetic hierarchy that controls flower development in the eudicot model plant <i>Arabidopsis</i> .	13
4	The ABC model of floral organ identity	25
5	Schematic representation organ identity model for wild type and mutant flowers	26
6	The ABC and <i>SEP</i> genes specify floral organ identity	28
7	Electrophoresis of total RNAs on denaturing formaldehyde agarose gel	60
8	Synthesized double-stranded cDNA using ZAP-cDNA synthesis kit which were electrophoresed on 1.2% agarose with marker 1 kb DNA ladder	61
9	Autoradiograph of primary screening	64
10	Autoradiograph of secondary screening	64
11	Probe synthesis of cDNA from total RNA of female flower using SMART-cDNA synthesis kit	65
12	Amplification of cDNA inserts of <i>opff</i> clones with T <sub>3</sub> and T <sub>7</sub> primers	66
13	The score distribution of clones based on BLASTX algorithm	68
14	Overview of the distribution of ESTs in plants	68
15	Summary of ESTs in biochemical categories	69
16	The nucleotide and deduced amino acids sequence of clone 8-3	77



17	The nucleotide and deduced amino acids sequence of clone 9-32	78
18	Amino acid sequence alignment of 8-3 with other related <i>AP1</i> genes from other species	80
19	Amino acid sequence alignment of 9-32 with other related <i>AGL9</i>	81
20	Phylogenetic tree of oil palm MADS-box protein sequences with <i>Arabidopsis</i> MADS-box protein sequences	83
21	Genomic Southern blot analyses of (A) 8-3 and (B) 9-32	84
22	Analysis of expression of (A) 9-32 gene in various sizes of abnormal female flower	86
23	Analysis of expression of (A) 9-32 gene in various sizes of normal male flower	87
24	The integrity of first-strand cDNAs produced for semi-quantitative analysis	90
25	Determination of PCR optimum cycle for semi-quantitative RT analysis	91
26	Expression of 8-3 and 9-32 genes in various tissues as described in Figure 24	91
27	Confirmation of specificity of semi-quantitative analysis	92
28	<i>In situ</i> hybridisation analysis of 8-3 expression in normal female flowers of oil palm	95
29	<i>In situ</i> hybridisation analysis of 9-32 expression in oil palm flowers	97



## LIST OF ABBREVIATIONS

$\alpha$	alpha
$\beta$	beta
$\lambda$	lambda
$\mu\text{g}$	microgramme
$\mu\text{l}$	microliter
$^{\circ}\text{C}$	degree centigrade
%	percentage
Amp	ampicillin
bp	base pair
BLAST	basic local alignment search tool
BSA	bovine Serum Albumin
cm	centimeter
DNA	deoxyribonucleic acid
Dnase I	deoxyribonuclease 1
cDNA	complementary DNA
dNTPs	deoxynucleotides
dATP	2'-deoxy-adenosine-5'-triphosphate
dCTP	2'-deoxy-cytidine-5'-triphosphate
dGTP	2'-deoxy-guanosine-5'-triphosphate
dTTP	thymidine-5'-tryphosphate
dH <sub>2</sub> O	distilled water





DEPC	diethyl pyrocarbonate
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol bis-( $\beta$ -aminoethyle ether)
EtBr	ethidium bromide
g	gramme
GTE	glucose-Tris-EDTA
hr	hour
Jacq.	Jacquin
kb	kilobase-pair
LB	luria-bertani
LiCl	lithium chloride
M	molar
MADS	MCM1-AGAMOUS-DEFICIENS-SRF
mg	milligram
min	minute(s)
mm	millimeter
mM	millimolar
MOPS	3-(N-morpholino) propanesulfonic acid
MPOB	Malaysian Palm Oil Board
mRNA	messenger RNA
NaCl	sodium chloride
NaOH	sodium hydroxide



NCBI	National Center for Biotechnology Information
ng	nanogramme
OD	optical density
PCI	phenol : chloroform : isoamyl
PCR	polymerase chain reaction
<i>pfu</i>	plaque forming units
Poly (A) <sup>+</sup>	polyadenylated (mRNA)
PVP	polyvinylpyrrolidone
PVPP	polypolyvinylpyrrolidone
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolution per minute
SDS	sodium dodecyl sulphate
SSC	sodium chloride-sodium citrate buffer
TAE	tris acetate EDTA
TE	tris-HCL-EDTA
UV	Ultraviolet
v/v	volume per volume
w/v	weight per volume



## CHAPTER 1

### INTRODUCTION

Flower development is a key process for all angiosperms. Although the flowering responses are very diverse in different species, it can be classified into two general phases: vegetative and reproductive. The transition to flowering, the first step in flower development, is triggered by a number of environmental and endogenous signals that gives rise to the transformation of the vegetative shoot apical meristem into an inflorescence meristem that in turn initiates floral meristems laterally in a spiral pattern. These flower meristems produce organ primordia in a whorled arrangement, which subsequently differentiate into numerous distinct cell types. After the maturation of the flower, fertilization of the ovules leads to fruit development and the production of seeds, closing the reproductive cycle.

Throughout their life cycle, plants and plant cells continually respond to signals that they use to alter their physiology, morphology and development. The specification of plant development involves a network of interacting genes, which are composed of mainly transcription factors. Transcription factor genes comprise a substantial fraction of all eukaryotic genomes and the majority can be grouped into a handful of different, often large gene families according to the type of DNA binding domain that they encode (Riechmann and Ratcliffe, 2000).



Transcription factors can be viewed as molecular switches that link signal transduction pathways to gene expression.

The establishment of meristem identity, floral organ identity and, to an extent, cell identity have been shown to require the combinatorial activity of members of the MADS-box gene family, acting in a series of networks involving other transcription factors. MADS-box genes are numerous in plants and they control many aspects of plant development, most notably flower development. Besides that, MADS-box genes also play important roles in controlling flowering time, leaf production and, ovule, fruit and root development (Levy and Dean, 1998; Mandel and Yanofsky, 1998).

Flowers and fruits are an integral part of seed production. The economic value of oil palm is much dependent on its production of fruits, which gives rise to oil. Furthermore, in the years to come, the demand for palm oil is expected to increase. In order to fulfil the increasing demand for palm oil, an improvement in yield is required. Hence, flowering and fruiting habits of the oil palm are important. Nevertheless, the occurrence of floral abnormalities such as mantling and andromorphic male inflorescences had reduced the ability to increase yield of oil palm through use of clonal plantings. The inadequate number of fertile fruits to sustain bunch development results in total bunch failure and loss of yield (Rao and Donough, 1990). The androgynous inflorescence is similar to an *Arabidopsis* mutant with a loss of B function.



The expressed sequence tag (EST) approach is an efficient way of obtaining vast amounts of gene resources and has revolutionized the way of obtaining functional genes from many organisms. It has provided an invaluable resource for the analysis of gene expression, and is also seen as an inexpensive and rapid means to identify large numbers of expressed genes from various developmental stages.

The random EST approach relies essentially on stochastic sampling of transcripts which is strongly biased by expression level. Highly expressed genes are often over represented in EST sets, whereas genes expressed at low levels may be absent. Low abundance genes have been shown to encode transcription factors, signal transduction components and membrane receptors, that play key roles in establishing structures, patterns and regulating developmental processes. Hence, cold plaque screening was employed to reduce the prevalence of highly abundance genes and enhance the chances to isolate low abundance genes.

The objectives of my study are I) to isolate and characterize low abundance clones from oil palm floral cDNA library, II) to isolate and characterize MADS box genes known to be involved in flower initiation and III) to study the role played by these MADS box genes in flower development.



Plant MADS-box genes control diverse aspects of plant development. Hence, the identification and isolation of MADS box genes have open up avenues for the study of vital processes involved in flower formation in oil palm. The understanding of the role played by these MADS box genes in normal and abnormal development of oil palm flowers would enable the manipulation of floral architecture as well as the directed improvement of floral and fruit development in oil palm to be realized.

