UNIVERSITI PUTRA MALAYSIA

ISOLATION OF EXPRESSED SEQUENCE TAGS (ESTS) FROM AN OIL PALM (*ELAEIS GUINEENSIS* JACQ.) MALE AND FEMALE FLORAL cDNA LIBRARY

CHOI MEI CHOOI

FSMB 2003 11

ISOLATION OF EXPRESSED SEQUENCE TAGS (ESTS) FROM AN OIL PALM (*ELAEIS GUINEENSIS* JACQ.) MALE AND FEMALE FLORAL cDNA LIBRARY

By

CHOI MEI CHOOI

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

February 2003

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

ISOLATION OF EXPRESSED SEQUENCE TAGS (ESTS) FROM AN OIL PALM (*ELAEIS GUINEENSIS* JACQ.) MALE AND FEMALE FLORAL cDNA LIBRARY

By

CHOI MEI CHOOI

February 2003

Chairman : Assoc. Prof. Dr. K. Harikrishna

Faculty : Food Science and Biotechnology

The bottleneck of the current knowledge in understanding the oil palm flower development in term of abnormalities and sex ratio is the low number of known flower genes available. The shortage of these genes is a serious constraint in studying the mechanisms regulating gene expression in flowers. Therefore, it is vital to obtain as much information as possible on the complexity and type of genes expressed in the oil palm floral organ. To dissect the molecular mechanisms underlying these development processes, a vast amount of genetic resources are required as markers that would enable a complete picture of the complex floral development processes in oil palm to be determined. In this study, expressed sequence tags (ESTs) were used as a genetic resource to facilitate the identification of new gene markers on a large-scale basis and also to provide information on gene expression patterns.

To date, about 1,600 EST clones were isolated and sequenced using cold plaque screening method. Among the 1,600 ESTs generated from cDNA libraries of oil palm

flowers, 266 independent clones with insert size of more than 500bp were subjected to reverse Northern analysis and this has resulted in classification into 5 major subpopulations (*opff* and *opmf*) based on the level and specificity of expression. Subpopulation A consisted of clones that were highly expressed in all the tissues (41%) whereas subpopulation B contained clones that were female predominant and expressed at medium levels (23.3%). Subpopulation C and D consisted of clones that were expressed at low levels predominantly in the flower (17.7%) and young leaf (6%), respectively. The last group, subpopulation E contained all the clones that showed no hybridization signals to all the tissues (12%).

Further characterization of selected *opff* and *opmf* clones by sequence analysis has revealed 3 major classes of ESTs. Class A consisted of sequences with similarity to known proteins in the database (56%) while class B showed sequences similar to proteins with unknown function (30%) and class C showed no sequence similarity to proteins in the database (14%).

Two clones were selected among the cDNA clones identified that have sequence homology to known sequences in database to be further characterized. They were putatively known as Squamosa Promoter Binding Protein (OPSBP) and GA-stimulated transcript 1 (OPGAST), respectively. OPSBP was found to be constitutively expressed throughout flower developement and was localized to bracts, rachis and carpels. On the other hand, OPGAST was found ubiquitously expressed in all flower tissues but the transcript levels were higher in the shoot apex. This study suggested that OPSBP appears to be up-regulated during early flower development.

Further characterization of these clones by Southern genomic hybridization showed that OPSBP and OPGAST exist as a single copy gene and multi-copy genes, respectively, in the oil palm genome.

In summary, the results indicate that the EST approach was efficient in identifying and isolating oil palm floral genes that range from those that are expressed at certain developmental stages to those that have undetectable expression throughout flower development. This achievement also demonstrated that genes isolated by cold plaque screening are not restricted to genes that are tissue specific or developmental stage specific but the technique also enables the isolation of genes that are generally expressed at low levels.

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

TAG PENGEKSPRESAN JUJUKAN (EST) DARIPADA KOLEKSI cDNA BUNGA JANTAN DAN BETINA KELAPA SAWIT (*ELAEIS GUINEENSIS* JACQ.)

Oleh

CHOI MEI CHOOI

Februari 2003

Pengerusi : Prof Madya. Dr. K. Harikrishna

Fakulti : Sains Makanan dan Bioteknologi

Masalah semasa di dalam pemahaman perkembangan bunga kepala sawit dari segi kenormalan dan nisbah seks ialah bilangan gen bunga yang diketahui dan tersedia ada adalah sangat rendah. Kekurangan gen ini adalah penghalang utama untuk memahami mekanisma pengawalan ekspresi gen dalam pembungaan. Oleh kerana itu, maklumat mengenai jenis gen yang dizahirkan dalam organ bunga kelapa sawit perlu diketahui sebanyak yang mungkin. Untuk memahami mekanisma molekul dalam proses perkembangan bunga kelapa sawit selanjutnya, sumber bekalan genetik yang luas perlu dijadikan penanda untuk memberi satu gambaran lengkap tentang proses perkembangan bunga kelapa sawit yang komplek. Oleh itu, di dalam kajian ini tag pengekspresan jujukan (EST) digunakan sebagai suatu sumber bekalan genetik untuk memudahkan pengenalan penanda gen baru secara besaran dan juga membekalkan maklumat berkaitan dengan corak ekspresi gen.

Pada masa kini, lebih kurang 1,600 klon EST telah dipencil dan dijujuk dengan menggunakan cara penyalingan "cold plak". Di antara 1,600 klon EST yang terhasil

daripada koleksi cDNA bunga kelapa sawit, 266 klon individu dengan saiz sisipan cDNA melebihi 500bp telah dijalankan analisi Northern "berbalik" dan ini menghasilkan 5 kelas subpopulasi utama berdasarkan tahap dan kespesifikan ekspresi dalam tisu. Subpopulasi A terdiri daripada klon yang menunjukkan ekspresi tinggi dalam semua tisu (41%), manakala subpopulasi B mengandungi klon yang menunjuk ekspresi sederhana pada bunga betina (23.3%). Subpopulasi C dan D masing-masing terdiri daripada klon yang menunjukkan ekspresi rendah dalam bunga betina dan daun muda. Kelas terakhir iaitu subpopulasi E mengandungi semua klon yang tidak menunjukkan sebarang isyarat penghibridan kepada semua tisu (12%).

Klon *opff* dan *opmf* yang terpilih untuk pencirian selanjutnya telah diketegori kepada 3 kumpulan yang besar berdasarkan analisi penjujukan. Kategori A terdiri daripada jujukan yang sama dengan protin yang terketahui dalam pengkalan data (56%) sementara kategori B menunjukkan jujukan sama dengan protin yang tidak tahu fungsinya (30%) dan kategori C terdiri daripada jujukan yang tidak menunjukkan sebarang persamaan dengan protin dalam pengkalan data (14%).

Dua klon telah dipilih daripada cDNA klon yang mempunyai jujukan sama dengan pengkalan data yang sedia ada. Mereka masing-masing dikenali sebagai SQUA Promoter Binding Protein (OPSBP) dan Giberelin Stimulated Transkript 1 (OPGAST). Gen OPSBP menunjukkan ekspresi berterusan di sepanjang perkembangan bunga dan ia didapati di 'bract', 'rachis' dan 'carpel'. Manakala, gen OPGAST menunjukkan ekspresi yang sama di semua tisu bunga dengan kadar yang lebih tinggi di puncak pucuk. Kajian ini mencadangkan OPSBP mungkin dikawal

pada peringkat awal perkembangan bunga. Pencirian selanjutnya klon-klon ini dengan cara pnghibridan genomik 'Southern' menunjukkan OPSBP dan OPGAST masing-masing muncul sebagai gen individu dan gen pelbagai salinan dalam genomik kelapa sawit.

Sebagai rumusan, keputusan di atas menunjukkan pendekatan EST cekap dalam pengenalan dan pemencilan gen-gen bunga kelapa sawit termasuk gen-gen yang penzahirannya pada peringkat perkembangan bunga tertentu dan gen-gen yang penzahirannya tidak dapat dikesan. Pencapaian ini juga menunjukkan gen yang terpencil dengan penyalingan 'cold plak' tidak saja terhad kepada gen tisu spesifik atau peringkat perkembangan spesifik, teknik ini juga membolehkan pemencilan gen yang menunjukkan ekspresi dalam kadar yang rendah.

ACKNOWLEDGEMENTS

A **"BIG"** Thank You I would like to give it to my supervisor, Assoc. Prof. Dr. K. Harikrishna for all his guidance, advice and patient throughout my study. I know I give you a lot of problems during this period of time, and sometimes you lost you patient on me, but I do appreciate because without your patient guidance I would not achieve what I have right now. Thanks for all the pushes to keep me going.

My greatest gratitude goes to my co-supervisor, Dr.Sharifah and Dr. Ho Chai Ling for their constant advice and guidance. Thanks for all your support and encouragement while I was conducting the project. A special thank to the Director of Malaysian Palm Oil Board for letting me doing radioactive work there.

My heartfelt thanks to all the members of the Genetic Lab, UPM including Dr. Tan, Wan Chin, Siaw San, Chieh Wean, Pick Kuen, Weng Wah, Sew, Yang Ping, Yen Yen, Mr.Ong, Seddon, Nancy, Musliana and Chuen Yi for their invaluable friendship that spiced up my life.

My sincere appreciation also dedicated to those people always support, encourage and help me throughout my study. All the staff members of plant development lab; Dr. Meilina Ong, Kak Zah, Siew Eng, Zaidah, Mas, Shamsul and Roslan for helping me around when I conducting my radioactive work in MPOB.

My friends, Yoanne, Siew Lian, Yean Hui, Chin Ming, Pik Lian, Ah Guek and my housemates, Ah Piah, Ee Fong, Ah Peng, Waei Gean for their encouragement, support and love.

The most importnant, I would like to express my sincere thanks to my parents, sister and brothers for their love and support throughout my life. My deepest gratitude to my sister, Poh Choo for her endless love and support.

TABLE OF CONTENT

		Page
ABSTACT		ii
ABSTRAK		v
ACKNOLEDGEMENTS		viii
APPROVAL SHEETS		x
DECLARATION FORM		xii
LIST OF TABLES		xv
LIST OF FIGURES		xvi
ABBREVIATIONS		xviii
CHAPTER		
1	**INTRODUCTION**	1
2	**LITERATURE REVIEW**	
	2.1 The Morphology of Oil Palm Flower	4
	2.2 Abnormalities Associated with Oil Palm Floral	6
	2.3 Possible Causes of Floral Abnormalities	7
	2.4 The Molecular Biology of Flower Development	9
	2.4.1 Flowering Time Genes	12
	2.4.2 Floral Meristem Identity Genes	12
	2.4.3 Organ Identity Genes	14
	2.4.4 MADS-box Genes	15
	2.5 Expressed Sequence Tags	18
	2.6 DNA Microarray	22
	2.7 Bioinformatics	23
	2.8 Gibberelin Signal Transduction	24
3	**MATERIALS AND METHODS**	
	3.0 Plant Materials	29
	3.1 Probe Preparation	29
	3.1.1 First Strand cDNA Synthesis and RT-PCR	31
	3.2 Cold Plaque Screening	32
	3.2.1 Preparation of Bacterial Culture for Infection	32
	3.2.2 Plaque lift	33
	3.2.3 Random Labeling of Double Stranded Probe	34
	3.2.4 Incorporation Assay	35
	3.2.5 Prehybridization and Hybridization of Plaques Membrane	35
	3.2.6 Autoradiography	36
	3.2.7 Random Selection of Plaques	36
	3.2.8 Single clone *In vivo* Excision	37
	3.2.9 96-well Plasmid Miniprep	39
	3.3 Characterization of Isolated cDNA Clones	40

3.3.1 Polymerase Chain Reaction (PCR) 41
3.3.2 Southern Blotting and Reverse Northern Analysis 42
3.4 Probe Removal 44
3.5 Automated DNA Sequencing 44
3.5.1 Cycle Sequencing 45
3.5.2 Preparation of Acrylamide Denaturing Gel 46
3.5.3 Sample Loading and Sequencing 46
3.5.4 Sequencing Analysis 47
3.6 RNA Extraction 48
3.6.1 Method by Schultz *et al.* (1994) 48
3.6.2 Method by Jill Winter *et al.* (1986) 50
3.6.3 Preparation of Denaturing Formaldehyde Agarose Gel 52
3.6.4 Northern Blot Analysis 53

4 RESULTS AND DISCUSSIONS
4.1 Isolation of EST's from an Oil Palm Floral cDNA Library 54
4.1.1 Cold Plaque Screening 54
4.1.2 Single *In Vivo* Excision Processes 59
4.1.3 PCR Amplification of Isolated cDNA Clones 59
4.1.4 Reverse Northern Analysis 60
4.1.5 Sequences Analysis of Putative Clones 64
4.2 Characterization of the Selected cDNA Clones 74
4.2.1 OPSBP Clone 78
4.2.2 OPGAST Clone 98

5 CONCLUSION 106

BIBLIOGRAPHY 109
APPENDICES
Appendix A 124
Appendix B 156
Appendix C 157
Appendix D 158
BIODATA OF THE AUTHOR 160

LIST OF TABLES

Table		Page
1	The 5 subpopulation of *opff* clones identified by reverse Northern analysis.....	62
2	Summary of stages of oil palm flower development.....	87

LIST OF FIGURES

Figures		Page
1	A schematic summary of the stable development phases of the transition from indeterminate, vegetative development to determinate, floral development	11
2	The workflow of cold plaque screening method	30
3	The PCR amplified cDNA probe from total RNA of oil palm 4-6cm female flower and young leaves used for reverse Northern analysis	55
4	The autoradiograph of primary screening with cDNA probe originated from the total of oil palm female flower	57
5	An autoradiograph of secondary screening of the oil palm female flwoer cDNA library with cDNA probe from total RNA of oil palm female flower	57
6	Reverse Northern analysis of *opff* clones	61
7	A summary of oil palm floral ESTs separated into their biochemical Categories	68
8	Redundancy of the oil palm floral EST sequences	70
9	Nucleotide and deduce amino acid sequence of clone OPSBP	75
10	Nucleotide and deduced amino acid sequence of OPGAST clone	77
11	The AP1 promoter contains a region with sequence similarity to the binding sites of SBP1 and SBP2 in the promoter region of the *A. Majus* SQUA gene	81
12	Alignment of deduced amino sequences of OPSBP with the homologues of Squamosa Promoter Binding Protein genes	84
13	Gene expression analysis of OPSBP during different development stages of normal flower	89
14	Gene expression analysis of OPSBP during different development stages of abnormal flower	90

15 Gene expression analysis of OPSBP during different development stages and floral organs in normal and abnormal flower..... 92

16 A comparison of expression patterns between OPSBP and OPAP1..... 93

17 Genomic DNA gel blot analysis of OPSBP clone..... 95

18 Alignment of deduced amino sequence of OPGAST with GA-regulated genes..... 100

19 Gene expression analysis of OPGAST during different development stages in normal flower..... 103

20 Gene expression analysis of OPGAST during different development stages in abnormal flower 104

21 Genomic Southern of OPGAST from oil pam..... 105

LIST OF ABBREVIATIONS

Symbol	Description
%	precentage
α	alpha
β	beta
λ	lambda
μg	microgram
μl	microliter
μm	micrometer
2-BE	ethyleneglycolmonobutylether
2-BE	ethylene glycol monobutyl ether
AG	AGAMOUS
AGL	AGAMOUS-LIKE
Amp	Ampicillin
BAP	benzylaminopurine
BLAST	Basic Local Alignment Research Tool
bp	base-pair
BSA	Bovine Serum Albumin
cDNA	Copy Deoxyribonucleic Acid
cm	centimeter
D X P	Dura X Pisifera
dATP	2'-Deoxy-adenosine-5'-triphosphate
dCTP	2'-Deoxy-cytidine-5'-triphosphate

DEPC	Diethyl Pyrocarbonate
dGTP	2'-Deoxy-guanosine-5'-triphosphate
dH_2O	distilled water
dNTPs	dioxynucleoside trisphosphate
DTT	Dithiothreithol
dTTP	Thymidine-5'-triphosphate
EDTA	Ethylenediaminetetraacetic Acid
EGTA	Ethylene Glycol Bis-(β-aminoethyle Ether)
EtBr	Ethidium Bromide
FLO	FLORICAULA
g	gram
GLO	GLOBOSA
GTE	Glucose-Tris-EDTA
HCl	Hydrochloric Acid
hr	hour
Jacq.	Jacquin
kb	kilobase-pair
KCl	Potassium Chloride
LB	Luria-Bertani
LFY	LEAFY
LiCl	Lithium Chloride
M	Molar
MADS	MCM1-AGAMOUS-DEFICIENS-SRF

mg	milligram
$MgCl_2$	magnesium chloride
$MgSO_4$	magnesium sulphate
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	nangram
°C	degree Centigrade
OD	Optical Density
OPAP1	Oil Palm APETALA 1
OPSBP	Oil Palm SQUAMOSA PROMOTER BINDING PROTEIN
ORF	Open reading frame
PCI	phenol : chloroform : isoamyl
PCR	Polymerase Chain Reaction
Pfu	plaque forming units
Poly $(A)^+$	polyadenylated (mRNA)
PVP	Polyvinylpyrrolidone

PVPP	Polypolyvinylpyrrolidone
RNA	Ribonucleic Acid
RNase	Ribonuclease
rpm	Revolution Per Minute
RT	reverse transcriptase
SDS	Sodium Dodecyl Sulphate
SSC	Sodium Chloride-Sodium Citrate buffer
TAE	Tris Acetate EDTA
TAE	Tri-Acetate-EDTA Buffer
TE	Tris-HCL-EDTA
TEMED	*N,N,N',N'*-tetramethyl-ethylenediamine
TPNS	triisopropylnaphtalene sulfonic acid
tRNA	Transfer Ribonucleic Acid
UTR	untranslated region
UV	Ultraviolet
V/v	volume per volume
W/v	weight per volume

CHAPTER I

INTRODUCTION

Malaysia is the world's leading producer of palm oil. On the world scale of production, Malaysia has recorded the highest share of globally traded palm oil (Yap, 1999). As demand for palm oil and oil palm seeds increase year to year, it is believed that the production of palm oil will reach up to 12.1 million tonnes by the year 2020 in Malaysia (Basiron *et al.*, 1999).

Since palm oil is of great economic value, a lot of emphasis has been placed on research in terms of quality improvement. Flowering is one of the areas of research interest. By controlling various aspects of oil palm flowering, it will be possible to improve the production of palm oil and oil palm seeds as flowering is an important introductory step to fruit formation.

Flower development is determined by an interaction between both genetic and environmental factors, but primarily is under genetic control (Gasser *et al.*, 1989). The use of polyembryogenic cultures for generating uniform selected plantlets has been applied to reduce the variation that could come with seed planting of hybrid planting material. However, polyembryogenic cultures are known to result in the occurrence of abnormal flowering (Corley *et al.*, 1986). These floral abnormalities can result in great crop losses whereby the abnormal fruits fail to ripen and often rot before oil synthesis occurs. However, little is known about the molecular biology of these abnormalities.

Various techniques and strategies have been used to understand oil palm flower development such as biochemistry, cytology and molecular biology (Shahrul, 1998). However, the number of known oil palm flower specific genes are currently very low. This shortage of genes is a serious constraint to studying the mechanisms regulating gene expression in flowers. Therefore, expressed sequence tags (ESTs) were used as a genetic resource in this project to facilitate the identification of new gene markers on a large-scale basis and also to provide information on gene expression patterns.

As the complete genome sequence of oil palm is unavailable, high-throughput EST generation combined with bioinformatics was used to rapidly discover and identify new gene sequences and their expression profiles. This information will provide essential resources for many biological studies, such as metabolic, development and signal tranduction. As a result, ESTs have emerged as a rapid way to establish an inventory of expressed genes by determining their sequences via single pass partial sequencing of randomly or selectively chosen cDNA clones (Adam *et al.*, 1991).

To better understand floral development in terms of abnormalities and sex ratio, partial cDNA sequences have been used to obtain as much information as possible on the complexity and type of genes expressed in oil palm floral organs. Thus, the objectives of this study were to isolate low abundance genes from cDNA libraries of oil palm flowers using a cold plaques screening method and to further identify and characterize these clones in term of tissue specificity. Partial cDNA

sequencing was used to obtain 1,600 ESTs from the cDNA libraries of oil palm flowers. These clones included 1,100 from a female flower library and 500 from a male flower library. Two cDNA clones were selected for further analysis and their expression patterns during oil palm floral development were characterized. These clones were selected based on the possibility that they are involved in the regulation of flower initiation and development. They are Squamosa Promoter Binding Protein (SBP: Klein *et al.*, 1996) and GA-stimulated transcript 1 (GAST1: Shi *et al.*, 1992). Northern and Southern analysis were performed to study their temporal and spatial expression pattern during flower development as well as the gene family copy number in the oil palm genome.

This study examined the potential and efficiency of an EST approach at identifying new genes isolated with the cold plaque screening method as a means to isolate low abundance cDNA clones from an oil palm flower cDNA library. It is hoped that this preliminary study will provide some valuable resources for the annotation of an oil palm database of sequences that may be used to facilitate the study of oil palm flower development in the future.