



**UNIVERSITI PUTRA MALAYSIA**

**GENERATION AND CHARACTERIZATION OF EXPRESSED  
SEQUENCE TAGS OF OIL PALM (*ELAEIS GUINEENSIS*) ROOT  
SYSTEM**

**NG WAI HAR.**

**FBSB 2005 25**



**GENERATION AND CHARACTERIZATION OF EXPRESSED SEQUENCE  
TAGS OF OIL PALM (*ELAEIS GUINEENSIS*) ROOT SYSTEMS**

By

NG WAI HAR

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

April 2005



## DEDICATION

I would like to dedicate this thesis to my beloved parents and family who have the confidence on me to overcome all the obstacles on my journey to success.



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement for the Degree of Master of Science

**GENERATION AND CHARACTERIZATION OF EXPRESSED SEQUENCE TAGS OF OIL PALM (*ELAEIS GUINEENSIS*) ROOT SYSTEMS**

**By**

**NG WAI HAR**

**April 2005**

**Chairman : Associate Professor Suhaimi Napis, PhD**

**Faculty : Biotechnology and Biomolecular Sciences**

Root system is an essential organ in plant for water and nutrient absorption, food storage and to support the plant within the soil. Construction of a cDNA library and analysis of the cDNA clones through ESTs approach have provided several advantages in acquiring data and gathering information on many aspects of plant biology at the molecular level.

A cDNA library was constructed using mRNA of oil palm root tissues. A total of 4,200 cDNA clones were randomly selected from the library, and about 3,365 ESTs with insert size ranging between 500 bp and 3,000 bp were generated. Similarity searches and analysis of the ESTs against GenBank databases were carried out. Characterization and classification of the ESTs were done based on their putative functions and biological roles. Results have shown that about 16.26% of the total ESTs did not show significant homology to the databases, whereas 21.84% were encoding hypothetical proteins in the oil palm root cDNA library. The results show that most of the functions of the isolated ESTs are still unknown. There is a need to carry out

further analysis and study in order to determine and identify the putative functions of these ESTs in the root growth and development.

In addition, an open reading frame of NAC gene was cloned from the oil palm root. This 1125 bp length NAC gene was found to show high similarity to the OsNAC protein (64%), GRAB1 (58%) and OsNAC 3 protein (58%). Alignment of this gene with NAC gene from other species has found that a conserved NAC domain with 156 amino acids length was located at the N-terminal of the oil palm amino acid from position of 86 to 240. It also carries a putative nuclear signal peptide, which corresponds to the NAC gene to function as a transcriptional factor in the nucleus. The existence of the highly basic regions of NAC domain suggests their involvement in DNA binding, as well as their interaction with the DNA.

Further analysis of the oil palm root ESTs using DNA microarrays technology was carried out to examine and study the differential gene expression patterns between the roots and stem tissues of oil palm. About 2,261 ESTs were spotted onto the solid substrate slides, and hybridized to the fluorescently labeled cDNA probes. Result indicates that about 69.65% of the ESTs isolated were specifically expressed in oil palm roots, while only 4.50% were expressed in roots, young leaf as well as in stem tissues of oil palm. This suggests that different tissues or organs in an organism have shared their gene expressions among each other for their growth and development.

Lastly, the results obtained have shown that the analysis of cDNA clones using ESTs approach is an effective and rapid way in isolating and identifying the genes expressed in an organism. It is also expected that the data collected from this study would be useful and informative for better understanding of molecular mechanism and development pathways in oil palm.



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

**PENGHASILAN DAN PERCIRIAN TEG BERTURUTAN EKSPRESI DARI  
AKAR POKOK KELAPA SAWIT (*ELAEIS GUINEENSIS*)**

Oleh

**NG WAI HAR**

**April 2005**

**Pengerusi : Professor Madya Suhaimi Napis, PhD**

**Fakulti : Bioteknologi dan Sains Biomolekul**

Tisu akar adalah organ yang penting dalam tumbuhan untuk penyerapan air dan nutrien, penyimpanan makanan serta sebagai penyokong bagi tumbuhan tersebut dalam tanah. Pembinaan perpustakaan cDNA dan analisa klon cDNA melalui kaedah tag berturutan ekspresi (ESTs) telah memudahkan dan memberi banyak kelebihan dalam pengumpulan maklumat dan data secara besar-besaran dari sudut biologi molekul sesuatu tumbuhan.

Satu perpustakaan cDNA telah dibina dengan menggunakan mRNA akar pokok kelapa sawit. Sejumlah 4,200 klon cDNA telah dipilih secara rawak daripada perpustakaan tersebut, dan sebanyak 3,365 ESTs bersaiz antara 500 bp hingga 3,000 bp telah dihasilkan. Pencarian persamaan dan analisis ESTs ke atas tapak data GenBank telah dijalankan. Percirian dan pengelasan ESTs telah dilakukan dengan mengikuti fungsi jangkaan dan tugas biologi bagi ESTs tersebut. Keputusan yang didapati menunjukkan bahawa sebanyak 16.26% daripada jumlah ESTs tidak mempunyai persamaan kepada tapak data am, manakala sebanyak 21.84% ESTs

berfungsi untuk menghasilkan protein di mana pemasukan kekuncinya tidak berasal di dalam perpustakaan cDNA akar pokok kelapa sawit. Ini menunjukkan bahwa kebanyakan ESTs yang ditemui dari akar pokok kelapa sawit mempunyai fungsi-fungsi yang belum dikenalpasti. Analisis dan kajian yang lebih lanjut diperlukan untuk mengenalpasti fungsi-fungsi jangkaan, serta penglibatan ESTs tersebut dalam perkembangan akar tumbuhan.

Tambahan pula, rangkaian bacaan terbuka bagi gen NAC telah diklonkan dari akar pokok kelapa sawit. Gen NAC yang bersaiz 1125 bp ini didapati mempunyai persamaan yang tinggi kepada protin OsNAC (64%), gen GRAB1 (58%) dan protin OsNAC 3 (58%). Penyusunan gen ini dengan gen NAC dari spesis-spesis lain telah menunjukkan bahawa suatu domain NAC tersimpan yang berpanjang asid amino 156 terletak di terminal-N antara asid amino 86 dan 240 dalam gen pokok kelapa sawit. Domain ini juga didapati mempunyai suatu isyarat peptid nuklear jangkaan yang berpadan kepada NAC gene untuk berfungsi sebagai faktor transkript dalam nukleus. Dengan ini, kemunculan NAC domain yang mempunyai kawasan tapak tinggi ini telah mencadangkan penglibatannya dalam pengikatan DNA serta interaksinya dengan DNA.

Analisis lebih lanjut ke atas ESTs akar pokok kelapa sawit dengan menggunakan teknologi susunan-mikro DNA telah dijalankan untuk memeriksa dan mempelajari corak ekspresi gen pembezaan antara akar dan tisu dahan pokok kelapa sawit. Sebanyak 2,261 klon ESTs telah disusun dan dicetakkan pada suatu slaid substrak pejal, dan dihibrid kepada prob



cDNA yang dilabelkan dengan fluorezen. Keputusan yang didapati menunjukkan bahawa sebanyak 69.65% ESTs yang ditemui adalah gen-gen unggul yang mengekspres di dalam akar pokok kelapa sawit, manakala hanya sebanyak 4.50% ESTs yang mengekspres di dalam akar, daun muda dan tisu dahan pokok kelapa sawit. Ini mencadangkan bahawa tisu-tisu atau organ-organ dalam suatu organisma berkongsi pengeksprei gen-gen mereka antara satu sama lain untuk pertumbuhan dan kemajuan tumbuhan tersebut.

Secara keseluruhannya, keputusan yang didapati telah menunjukkan bahawa analisa klon cDNA melalui kaedah ESTs merupakan suatu kaedah yang efektif dalam pengasingan dan pengenalpastian gen-gen yang diekspres di dalam sesuatu organisma. Data-data ESTs yang dihasilkan dalam pengajian ini juga dijangkakan adalah bermanfaat dan informatif untuk mengenali mekanisma molekul dan haluan pertumbuhan yang terlibat dalam pokok kelapa sawit dengan lebih lanjut dan mendalam.

## ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude and appreciation to Assoc. Prof. Dr. Harikrishna, Dr. Ho Chai Ling and Dr. Tan Siang Hee for their valuable guidance, advise, support and encouragement throughout the study.

I am also greatly indebted to all the members in the Genetic Lab at UPM, especially Lee Weng Wah, Sew Yun Shin, Nelson Chen Kin Foong, Au Sian Long, Mr. Ong, Choong Chieh Wen, Nancy Liew, Chan Pick Kuen, Choy Mei Chooi, Lee Yang Ping, Musliayana, Seddon Teoh, Bird Teo Chee How, Tan Boon Keat, Ng Chyan Leong, Sim Khean Hong, Nyu Ka Jin and others for their assistance, support and construction discussion and companionship throughout this project.

Not forgetting staff members in National Cancer Centre of Singapore, who extended their guidance and support during this thesis write-up, especially to those in Laboratory of Applied Human Genetics. They are Prof. Kon Oi Lian, Dr. Alvin Eng, Doris Ma, Adrian Khoo, Louise Lee Sze Sing, Leong Siew Hong, Cheryl Lee I-Ping, Jaichandran s/o Sivalingam, Mark Tan Eng Hui, Lai Li Yun, Bernice Wong Hui Min, Ang Kok Heong.

I also owe a special debt to my dearest friends, Gary Chew Chian Giap, Pang Chun Meng, Francis Tan Chong Kheam, Ivy Lau for their encouragement.

Last but not least, my highest appreciation to my family, my beloved and his family for all the moral support that I needed in completing this thesis.



## TABLE OF CONTENT

	Page
<b>DEDICATION</b>	<b>2</b>
<b>ABSTRACT</b>	<b>3</b>
<b>ABSTRAK</b>	<b>6</b>
<b>ACKNOWLEDGEMENT</b>	<b>9</b>
<b>APPROVAL</b>	<b>11</b>
<b>DECLARATION</b>	<b>13</b>
<b>LIST OF FIGURES</b>	<b>18</b>
<b>LIST OF TABLES</b>	<b>20</b>
<b>LIST OF ABBREVIATIONS</b>	<b>21</b>
<b>CHAPTER</b>	
<b>1 INTRODUCTION</b>	<b>25</b>
1.1 Introduction	25
1.2 Objectives	29
<b>2 LITERATURE REVIEW</b>	<b>30</b>
2.1 The Oil Palm, <i>Elaeis guineensis</i> , in Malaysia	30
2.1.1 Botany	30
2.1.2 History	33
2.1.3 Development and Economic Aspects	34
2.1.4 Market	35
2.2 The Root Systems of Plants	36
2.2.1 Anatomy and Morphology	37
2.2.2 Cellular Structure	42
2.2.3 Molecular Biology	43
2.2.4 Environmental Effects	44
2.3 Expressed Sequences Tags (ESTs)	44
2.3.1 EST Projects	45
2.3.2 EST Approach	46
2.3.3 Model Plant Species	49
2.3.4 Bioinformatics	50
2.3.4.1 Similarities Analysis of ESTs Clones	51
2.3.4.2 Databases and Tools	51
2.4 DNA Microarrays	53
2.4.1 Applications	56

<b>3</b>	<b>METHODOLOGY</b>	<b>58</b>
3.1	Plant Materials	58
3.2	Total RNA Extraction from Oil Palm Roots	58
3.2.1	Total RNA Extraction using CTAB/polyvinylpyrrolidone Method	58
3.2.2	Total RNA Extraction using Cesium Chloride (CsCl) Gradient	60
3.3	Total RNA Extraction from Oil Palm Young Leaf and Stem Tissue	62
3.4	Formaldehyde-Agarose Gel Electrophoresis	63
3.4.1	Formaldehyde RNA Gel Preparation	64
3.4.2	RNA Sample Preparation	64
3.5	Poly-A <sup>+</sup> mRNA Isolation	65
3.6	Construction of cDNA Library from Oil Palm Roots	66
3.6.1	Synthesizing First-Strand cDNA	67
3.6.2	Synthesizing Second-Strand cDNA	67
3.6.3	Blunting the cDNA Termini	68
3.6.4	Ligating the <i>Eco</i> RI Adapters	68
3.6.5	Phosphorylating the <i>Eco</i> RI Ends	69
3.6.6	Digesting with <i>Xho</i> I	69
3.6.7	Size Fractionation	69
3.6.8	Purification of cDNA Frgaments from Agarose Gel	70
3.6.9	Quantitating the cDNA with Ethidium Bromide Plate Assay	71
3.7	Ligation of cDNA into Uni-ZAP XR Vector	72
3.8	Preparation of the Host Bacteria	72
3.9	Packaging of Phages	73
3.10	Plating and Titering of Unamplified cDNA Library	73
3.11	Amplification of the cDNA Library	75
3.12	Plating of Phages	76
3.13	Single-Clone Excision of Bacteriophage in Large-Scale	76
3.14	Growth of Bacteria in Suspension Culture using 96-well Culture Plate	78
3.15	Preparation of Glycerol Stock of Bacterial Culture	78
3.16	Isolation of Plasmid DNA from Bacterial Culture	78
3.16.1	Mini-Preparation of Plasmid DNA	79
3.16.2	Purification of Plasmid DNA	79
3.17	Analysis of Cloned cDNA Fragments with PCR	80
3.18	Gel Electrophoresis	81
3.19	Automated DNA Sequencing	81

3.19.1	DNA Sequencing using ABI PRISM™ 377 DNA Sequencer	82
3.19.2	DNA Sequencing Using MegaBASE™ 1000	83
3.20	Sequences Data Analysis	84
3.21	Cloning of NAC Gene from Oil Palm	84
3.21.1	First-strand cDNA Synthesis	84
3.21.2	5'-RACE PCR Reaction	85
3.21.3	Confirmation and Amplification	86
3.21.4	Cloning of PCR Products and Transformation	86
3.21.5	Determination of Oil Palm NAC Gene Sequences	88
3.22	Analysis of ESTs Clones using DNA Microarray Technology	88
3.22.1	Preparation of Target DNA	88
3.22.2	Purification of PCR-Amplified DNA	89
3.22.3	Array Printing	90
3.22.4	Preparation of Probes for Hybridization	91
3.22.5	Labeling of Probe for Hybridization	91
3.22.6	Glass Slide Pre-Hybridization	92
3.22.7	Hybridization of Array Slide	93
3.22.8	Slide Washing	93
3.22.9	Slide Scanning and Analysis	94
<b>4</b>	<b>RESULTS AND DISCUSSIONS</b>	<b>95</b>
4.1	Extraction of Total RNA from Oil Palm Root Tissues	95
4.2	Extraction of Total RNA from Oil Palm Young Leaf and Stem Tissues	99
4.3	Isolation of mRNA	102
4.4	Construction of Oil Palm Root cDNA Library	103
4.4.1	cDNA Synthesis	103
4.4.2	Size Fractionation of cDNA	106
4.4.3	Ligation of cDNA into Vector	109
4.4.4	Plating and Titering of cDNA Library	109
4.4.5	Evaluation of the cDNA Library	110
4.5	Amplification of Packaged cDNA Library	112
4.6	Plating and Isolation of Single Recombinant Plaques	112
4.7	PCR Amplification of the cDNA Inserts	113
4.8	Single-Pass Partial Sequencing of cDNA Clones	117
4.9	Analysis of Expressed Sequence Tags	119
4.10	Characterization of Oil Palm Root ESTs	126
4.11	Gene Expression Profiles of Oil Palm Root ESTs	133

4.12	Metabolism of Oil Palm Roots	135
4.12.1	Catabolism	135
4.12.2	Anabolism	136
4.12.2.1	Synthesis of Monosaccharide and Polysaccharide	136
4.12.2.2	Synthesis of Fatty Acids	137
4.12.2.3	Assimilation of Nitrogen and Sulphur	138
4.12.2.4	Synthesis of Protein	139
4.12.3	Secondary Plant Products	140
4.12.3.1	Plant Hormones	140
4.12.4	Absorption, Secretion and Translocation	142
4.12.5	Growth and Differentiation	143
4.12.6	Cytoskeleton	144
4.12.7	Signal Transduction	145
4.12.8	DNA and RNA Metabolism	146
4.12.9	Stress and Defense	147
4.12.10	Specific Genes	149
4.13	Genes of Interest in Oil Palm Root cDNA Library	150
4.13.1	Chitinases	150
4.13.2	Phenylalanine Ammonia Lyase (PAL)	151
4.13.3	Nodulin Homeobox (NDX1) Protein	152
4.13.4	Omega-3 Fatty Acid Desaturase	153
4.13.5	<i>GIGANTEA (GI)</i>	154
4.14	Cloning of NAC Gene from Oil Palm	156
4.15	Uses of ESTs	171
4.16	DNA Microarray Analysis of Oil Palm Root ESTs	172
<b>5</b>	<b>CONCLUSIONS</b>	<b>185</b>
	<b>BIBLIOGRAPHIES</b>	<b>190</b>
	<b>APPENDICES</b>	<b>212</b>
	<b>BIODATA OF AUTHOR</b>	<b>239</b>



## LIST OF FIGURES

Figure		Page
1	Oil palm, one of the important crops in Malaysia.	26
2	Fruit forms of the oil palm.	32
3	The root system of oil palm.	38
4	A simplified representation of roots structural components.	40
5	Cross-section of the oil palm roots.	41
6	The flow chart of the construction of cDNA synthesis.	48
7	The strategy in analyzing gene expression of a tissue using DNA microarrays technology.	55
8	Extraction of total RNA from 3-month-old oil palm roots using CTAB/polyvinylpyrrolidone (PVP) method.	96
9	Extraction of total RNA from oil palm roots using CsCl gradient method.	98
10	Extraction of total RNA from oil palm young leaf and stem tissues using modified phenol/lithium chloride precipitation method.	101
11	Sequence of the 50-base oligonucleotide linker primer.	105
12	Size fractionation of cDNA by electrophoresis on 1.5% (w/v) agarose gel.	107
13	Determination of recovered cDNA concentration using EtBr plate assay.	108
14	Evaluation of the cDNA library by digesting the 24 recombinant clones with <i>Xho</i> I and <i>Eco</i> RI.	111
15	Electrophoresis of the amplified cDNA inserts on the 1.2% (w/v) agarose gel for size determination.	115
16	The size distribution of randomly selected cDNA clones from oil palm roots cDNA library.	116
17	The chromatogram profile and the plain text file of one of the cDNA clones.	118
18	Blast result of the EST clone.	121



19	Percentage of the ESTs which showed significant homology to the protein databases against the ESTs with no homology.	122
20	Functional Characterization of Oil Palm Root ESTs.	131
21	Electrophoresis of the PCR products using nested primer OPVR93-R1 on 1.5% (w/v) agarose gel.	157
22	An open reading frame nucleotide sequence and the deduced amino acid sequence of clone OPVR_93.	159
23	Sequence alignment of 15 NAC proteins and the NAC domain.	166
24	A neighbour-joining phylogeny of the conserved NAC domain of NAC proteins.	169
25	Southern blot analysis of the oil palm genomic DNA hybridized with the clone OPVR_93.	170
26	Electrophoresis of the purified EST clones on 1% (w/v) agarose gel for quality determination before spotting onto array slides.	175
27	Distance between each spot on the array slide.	176
28	The pseudocolours that appeared on the array slide after microarray hybridization.	177
29	Histogram to show the intensity distribution of Cye-3 and Cye-5.	179
30	Distribution of the ESTs hybridized to the Cye-3 or Cye-5 in DNA microarray analysis.	181

## LIST OF TABLES

Table		Page
1	The EST clones which showed similarities to the genes originated from oil palm family in the protein databases.	125
2	The classification and distribution of 3,365 ESTs and 1,379 distinct ESTs assigned to their putative functions.	127
3	ESTs that were categorized as non-plant proteins in oil palm root cDNA library.	132
4	Abundant ESTs found in the oil palm root cDNA library.	134
5	ESTs that have different gene expression areas in oil palm <i>Elaeis guineensis</i> .	182



## LIST OF ABBREVIATIONS

1:1	Ratio 1 to 1
1:50	Ratio 1 to 50
1:100	Ratio 1 to 100
1x	1 time
10x	10 times
24:1	Ratio 24 to 1
25:24:1	Ratio 25 to 24 to 1
~	About
α-	Alpha-
β-	Beta-
°C	Degree Celcius
μg	Microgram
μl	Microlitre
μg/μl	Microgram per microlitre
μg/ml	Microgram per millilitre
μM	Micromolar
%	Percentage
®	Registered trademark
x g	Relative centrifugal force (rcf)
A <sub>260</sub>	Absorbance at wavelength of 260 nm
A <sub>320</sub>	Absorbance at wavelength of 320 nm
A <sub>280</sub>	Absorbance at wavelength of 280 nm
A	Adenosine
ATP	Adenosine triphosphate
bp	Base pair
C	Cytidine
cDNA	Complementary DNA
cGMP	Cyclic guanosine 3', 5'-monophosphate
CO <sub>2</sub>	Carbon dioxide
CTAB	Hexadecyltrimethylammonium bromide
dATP	2'-deoxy-adenosine-5'-triphosphate

dCTP	2'-deoxy-cytidine-5'-triphosphate
ddATP	Dideoxy-adenosine-5'-triphosphate
ddCTP	Dideoxy-cytidine-5'-triphosphate
ddGTP	Dideoxy-guanosine-5'-triphosphate
ddTTP	Dideoxy-thymidine-5'-triphosphate
dGTP	2'-deoxy-guanosine-5'-triphosphate
dNTP	Deoxynucleotide
dTTP	2'-deoxy-thymidine-5'-triphosphate
dUTP	2'-deoxy-uridine-5'-triphosphate
DEPC	Diethyl pyrocarbonate
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acids
DTT	Dithiothreitol
<i>e.g</i>	For example
EDTA	Ethylenediamine tetraacetic acid
ESTs	Expressed sequence tags
<i>et al.</i>	And others
EtBr	Ethidium bromide
g	Grams
G	Guanosine
GMP	Guanosine 5'-monophosphate
GTP	Guanosine triphosphate
g/ml	Gram per milliliter
<i>i.e.</i>	In other words
IPTG	Isopropyl-beta-D-thiogalactopyranoside
K <sup>+</sup>	Potassium ion
kb	Kilo base pair
KCl	Potassium chloride
kDa	Kilo Dalton
L	Litre
LB broth	Luria-Bertani broth
LiCl	Lithium chloride
log <sub>10</sub>	Base-10 logarithm of a number

M	Molar
Mb	Mega base pair
mg	Milligram
MgCl <sub>2</sub>	Magnesium chloride
MgSO <sub>4</sub>	Magnesium sulphate
mg/ml	Milligram per millilitre
ml	Millilitre
mm	Millimeter
mM	Millimolar
MOPS	3-[N-morpholino]propane-sulfonic acid
mRNA	Messenger RNA
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NaOAc	Sodium acetate
ng	Nanogram
ng/μl	Nanogram per microlitre
O <sub>2</sub>	Oxygen
OD <sub>260/280</sub>	Optical density at wavelength of 260 nm over 280 nm
OD <sub>600</sub>	Optical density at wavelength of 600 nm
PCR	Polymerase chain reaction
pfu	Plaque forming units
pfu/ml	Plaque forming units per millilitre
pmoles	Picomoles
PVP	Polyvinylpyrrolidone
®	Registered trademark
RM	Ringgit Malaysia
RNA	Ribonucleic acids
RNase	Ribonuclease
RNase A	Ribonuclease A
rpm	Revolution per minute
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulfate
SSC	Standard Saline Citrate
ssDNA	Single strand DNA
T	Thymidine

TAE buffer	Tris-acetate-EDTA buffer
TE buffer	Tris-EDTA buffer
™	Trade mark
tRNA	Transfer RNA
U	Uridine
U	Unit
USD	US dollar
UV	Ultraviolet
U/μl	Unit per microlitre
V	Volts
v/v	Volume per volume
w/v	Weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside

UPM

© COPYRIGHT