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 fulfilment 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 lenght 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 berasas di dalam perpustakaan cDNA akar pokok kelapa sawit. Ini menunjukan bahawa 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, rangkaan 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 disusunkan dan dicetakkan pada suatu slaid substrak pejal, dan dihibrid kepada prob

cDNA yang dilabelkan dengan fluoresen. 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 pengekspresi 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 effektif 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

LIST OF FIGURES

LIST OF TABLES

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_{260}	Absorbance at wavelength of 260 nm
A_{320}	Absorbance at wavelength of 320 nm
A_{280}	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_2	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
e.g	For example
EDTA	Ethylenediamine tetraacetic acid
ESTs	Expressed sequence tags
et al.	And others
EtBr	Ethidium bromide
g	Grams
G	Guanosine
GMP	Guanosine 5'-monophosphate
GTP	Guanosine triphosphate
g/ml	Gram per milliliter
i.e.	In other words
IPTG	Isopropyl-beta-D-thiogalactopyranoside
K^+	Potassium ion
kb	Kilo base pair
KCl	Potassium chloride
kDa	Kilo Dalton
L	Litre
LB broth	Luria-Bertani broth
LiCl	Lithium chloride
$\log_{10}$	Base-10 logarithm of a number

M	Molar
Mb	Mega base pair
mg	Milligram
$MgCl_2$	Magnesium chloride
$MgSO_4$	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_2	Oxygen
$OD_{260/280}$	Optical density at wavelength of 260 nm over 280 nm
OD_{600}	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