UNIVERSITI PUTRA MALAYSIA

MOLECULAR CHARACTERIZATION AND FUNCTIONAL ANALYSIS OF SELECTED EXPRESSED SEQUENCE TAGS FROM OIL PALM CELL SUSPENSION CULTURE

LE VINH THUC

FBSB 2009 37

MOLECULAR CHARACTERIZATION AND FUNCTIONAL ANALYSIS OF SELECTED EXPRESSED SEQUENCE TAGS FROM OIL PALM CELL SUSPENSION CULTURE

By

LE VINH THUC

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in fulfilment of the Requirements for the Degree of Doctor of Philosophy

November 2009

This Thesis is Special Dedicated to

My Parents, Mr. Le Van Khoanh and Mrs. Tran Thi Kim Dang

And also to my beloved wife, Ngo Thi Thanh Truc

Their Sacrifice and Infinite Love Led Me to Present Achievements

Abstract of thesis presented to the senate of Universiti Putra Malaysia in fulfilment of the requirements for the degree of Doctor of Philosophy

MOLECULAR CHARACTERIZATION AND FUNCTIONAL ANALYSIS OF SELECTED EXPRESSED SEQUENCE TAGS FROM OIL PALM CELL SUSPENSION CULTURE

By

LE VINH THUC

November 2009

Chairman: **Parameswari A/P Namasivayam, PhD**

Faculty: **Biotechnology and Biomolecular Sciences**

A large quantity of Expressed sequence tags (ESTs) are available from various cDNA libraries of oil palm. The information from oil palm EST databases has been utilized to identify several interesting transcripts that are upregulated in cell suspension culture for molecular characterization and functional analysis. The first part of this study was to carry out the molecular characterization of selected ESTs of oil palm cell suspension culture which were Eg583 (Accession number: EU795363), Eg707 (Accession number: FJ196136) and EgHAD (Accession number: FJ196137).

The Eg583 sequence is highly similar to an unknown protein from rice. This predicted protein might be a transcription factor due to the presence of SIN3 domain and motifs of casein kinase II phosphorylation. The expression of this gene was not detectable in all tested tissues. This gene might be a member of a multigene family in the oil palm genome. The Eg707 sequence is highly similar to an unknown protein from *Arabidopsis* and might be a putative nuclear protein. Its amino acid sequence contains a Ald-Xan-dh-C2 domain that may be involved in ABA biosynthesis. *Eg707*

might be present as a single copy gene in the oil palm genome and its transcripts were highly expressed in tissue cultured materials compared to vegetative tissues. *Eg707* might have a role during oil palm somatic embryogenesis or at very early stage of embryo development. The EgHAD sequence is similar to a putative haloacid dehalogenase (HAD) superfamily hydrolase from monocots and phosphate hydrolase from dicots. However, the phylogenetic relationship of *EgHAD* is closer to monocots than dicots. *EgHAD* might be a member of a multigene family gene in the oil palm genome. It was highly expressed in leaves and meristem but lower expression was found in roots, female flowers, non-embryogenic and embryogenic calli in comparison to the oil palm cell suspension culture.

Functional analysis was carried out in rice by over-expressing *Eg707* and *EgHAD*, driven by a constitutive double Cauliflower Mosaic Virus 35S promoter. The constructs were made using the gateway technology with clonase (Invitrogen, USA). Transgenic plants over-expressing Eg707 protein had small sized, rolled and erect leaves, less tillers, empty seeds and higher total chlorophyll content. The phonotypes of these and the presence of Xan-dh-C2 domain in Eg707 protein, strongly suggest its involvement in ABA biosynthesis, particularly during somatic embryogenesis. Functional analysis of *Eg707* through RNAi-mediated gene silencing was unsuccessful since the T_1 seeds failed to germinate. Over-expression of *EgHAD* gene in rice produced more lateral roots and tillers than the wild type plants. However, it also reduced plant size, produced empty seeds and many tiny seeds which were not found in wild type plants. The suppression of the *EgHAD* orthologues in rice did not show any changes in the phenotype. *EgHAD* might be a metabolic protein involved in phosphate starvation mechanism and its expression might be necessary for seed germination.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENCIRIAN MOLEKULAR DAN ANALISIS BERFUNGSI TAG JUJUKAN TEREKSPRES TERPILIH DARIPADA KULTUR AMPAIAN SEL KELAPA SAWIT

Oleh

LE VINH THUC

November 2009

Pengerusi: **Parameswari A/P Namasivayam, PhD**

Fakulti: **Bioteknologi and Sains Biomolekular**

Sebilangan besar tag jujukan terekspres (ESTs) boleh didapati daripada pelbagai perpustakaan cDNA bagi kelapa sawit. Maklumat daripada pangkalan data EST kelapa sawit telah digunakan untuk mengenalpasti beberapa transkrip menarik yang telah ditingkat-kawalaturkan di dalam kultur ampaian sel untuk pencirian molekul dan analisis fungsian. Bahagian pertama kajian ini adalah untuk menjalankan pencirian molekul pada EST yang terpilih daripada ampaian kultur kelapa sawit iaitu Eg583 (Nombor kemasukan: EU795363), Eg707 (Nombor kemasukan: FJ196136) dan EgHAD (Nombor kemasukan: FJ196137).

Jujukan untuk klon Eg583 adalah amat seiras dengan protein yang belum dikenalpasti daripada padi. Protein yang diramalkan ini berkemungkinan merupakan faktor transkripsi disebabkan kewujudan domain SIN3 dan motif untuk pemfosforilan kasein kinase II. Ekspresi ungkapan gen ini tidak dapat dikesan dalam semua tisu yang diuji. Gen ini berkemungkinan merupakan ahli kepada keluarga gen berbilang di dalam genom kelapa sawit. Jujukan *Eg707* adalah sangat seiras dengan

protein yang belum dikenalpasti daripada *Arabidopsis* dan ia berkemungkinan merupakan putatif protein pada nukleus. Jujukan asid amino ini mengandungi domain Ald-Xan-dh-C2 yang mungkin terlibat dalam biosintesis ABA. *Eg707* berkemungkinan hadir sebagai salinan tunggal di dalam genom kelapa sawit dan transkripnya diekspres dengan tinggi dalam tisu yang dikulturkan berbanding dengan tisu vegetatif. *Eg707* berkemungkinan memainkan peranan semasa somatik embriogenesis atau pada tahap awal perkembangan embrio. Jujukan klon EgHAD adalah seiras dengan putatif haloasid dehalogenase (HAD) superfamili hidrolase daripada monokot atau hidrolase fosfat daripada dikot. Walau bagaimanapun, hubungan filogenetik EgHAD adalah lebih dekat dengan tumbuhan monokot daripada dikot. *EgHAD* berkemungkinan merupakan ahli keluarga gen berbilang di dalam genom kelapa sawit. Ia diekspres dengan tinggi dalam daun dan meristem tetapi pengekspresannya adalah rendah di dalam akar, bunga betina, kalus embriogenik dan tidak embriogenik berbanding dangan kultur ampaian sel kelapa sawit.

Analisis berfungsi telah dijalankan di dalam padi melalui pengekspresan terlampau *Eg707* dan *EgHAD*, dibawa oleh gandaan konstitutif pendorong Cauliflower Mosaic Virus 35S. Konstruk telah dibuat melalui teknologi gateway dengan klonase (Invitrogen, USA). Tumbuhan transgenik yang mengekspreskan protein *Eg707* secara terkebuh mempunyai daun yang kecil, bergulung dan tegang, kurang tangkai dan berbiji benih kosong dan mengandungi kandungan klorofil keseluruhan yang tinggi. Fenotip ini dan kehadiran domain Xan-dh-C2 di protein Eg707 ini mencadangkan dengan kukuh penglibatannya dalam biosintesis ABA, terutamanya semasa somatik embriogenesis. Analisis berfungsi *Eg707* melalui kaedah gen

penyenyap berperantara-RNAi tidak berjaya kerana biji benih generasi T_1 gagal bercambah. Pengekspresan terlampau gen *EgHAD* dalam padi menghasilkan banyak akar lateral dan lebih tiler daripada padi jenis liar. Walau bagaimanapun, ia juga mengecilkan saiz tumbuhan, menghasilkan biji kosong dan banyak biji kecil yang tidak terdapat dalam pokok padi jenis liar. Penindasan ortolog *EgHAD* dalam padi tidak menunjukkan sebarang perubahan fenotip. *EgHAD* berkemungkinan merupakan protein metabolik yang terlibat dalam mekanisme kebuluran fosfat dan ekspresinya ini adalah penting untuk percambahan biji benih.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my main supervisor, Dr. Parameswari Namasivayam for her supports, guidance, advice and comments throughout this project. With her constant motivation and words of advice, has inspired me to become a better researcher and individual. I would not have completed my project without her help and guidance. Besides her, I would like to express my appreciation to my co-supervisor, Assoc. Prof. Dr. Suhaimi Napis for his giving me the opportunity to pursue my studies in the field of plant molecular biology and Dr. Ooi Siew Eng for being my friendly supervisory committee member and for her precious advice, comments and guidance whenever sought.

I would like to thank Ministry of Science, Technology and Innovation (MOSTI) under the Intensified Research for Priority Areas (IRPA) for funding this project Grant No. 01-02-04-0387-PR0015/07-02. I would also like to thank the Malaysian Palm Oil Board (MPOB) for providing the EST sequences and oil palm materials.

I would like to thank the Cantho University and Project 322, Vietnam for granting me study leave to study in Universiti Putra Malaysia, Malaysia for four years. Deepest gratitude is also goes to Assoc. Prof. Dr. Nguyen Bao Ve for all his help and encouragement to pursue this course. I appreciate all the support and motivation from my Vietnamese lecturers and colleagues in the Crop Science Department, College of Agriculture, Cantho University, Vietnam during my study in Malaysia.

Sincere thanks also go to Assoc. Prof. Dr. Norihan Saleh for allowing me to use her tissue culture laboratory facilities and Assoc. Prof. Dr. Ho Chai Ling for giving her

suggestions to my research. I also would like to thank the staff in Molecular Lab and the official staff in the Department of Cell and Molecular Biology, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, Malaysia for their help in laboratorial materials and in paper works, respectively. I also like to thank Dr. UmaRani Sinniah, from the Department of Crop Science, Faculty of Agriculture, Universti Putra Malaysia for her help in statistics and allowing me to use her laboratory facility.

My sincere appreciation also goes to Mr. Huynh Ky for his enthusiastic help, sharing and encouragement during this study. I would like to thank Mrs. Shahana for providing me Taipei rice seeds from her institute. I thank my friends for promptly helping me to find documents; Tung, Danh and Thien. My deep gratitude goes to Dr. Nguyen Phuoc Dang for giving his advices to pursue my PhD. course, Ms. Hariyanti Baharum and Mr. Yeap Swee Keong for helping me to translate the English abstract to Malay language. Also, thanks to my Vietnamese lecturer and friends in Malaysia, Dr. Le Ngoc Thach, Dr. Nghia, Mrs. Trang, Dr. Thanh, Thanh, Chi, Dang, Thu, Beh and all my friends at the Molecular Lab and Tissue Culture Lab for their encouragement and support. I also thank Mrs. Nancy, the officer of the Institute of Bioscience for her help in real time PCR techniques.

I am indebted to my parents, my sisters, brother and brothers in-law for their endless love and support. I also thank my relatives for all their supports. Last but not the least, I thank my beloved wife for her sacrifice living far from me and sharing my happiness and sadness during this study.

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirements for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Parameswari A/P Namasivayam, PhD
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Chairman)

Suhaimi b. Napis, PhD
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Member)

Ooi Siew Eng, PhD
Biological Research Division
Malaysian Palm Oil Board, Malaysia
(Member)

HASANAH MOHD GHAZALI, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 11 February 2010

DECLARATION

I hereby declare that the thesis is based on my original work except for equations and citations, which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UPM or other institutions.

_____

LE VINH THUC

Date: 2 December 2009

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2,4-D	-	2,4-dichlorophenoxyacetic acid
%	-	Percentage
α		Alpha
A. tumefaciens	-	*Agrobacterium tumefaciens*
aa		Amino acid
ANOVA	-	Analysis of variance
AS	-	Acetosyringone
BLAST	-	Basic Local Alignment Search Tool
bp	-	Base pair
BSA	-	Bovine serum albumin
CDD	-	Conserved domain database
cDNA	-	Complementary DNA
CI	-	Chloroform: isoamyl alcohol
cm	-	Centimeter
C_T	-	Threshold cycle
CTAB	-	Hexadecyl (or cetyl) trimethyl ammonium bromide
mm	-	Milimeter
DEPC	-	Diethylpyrocarbonate
DNA	-	Deoxyribonucleic acid
DNase I	-	Deoxyribonuclease I
dNTPs	-	Deoxynucleotides
EDTA	-	Ethylene diamine tetracetate
EST	-	Expressed Sequence Tag
EtBr	-	Ethidium bromide
g/L	-	Gram per liter
g	-	Gravitational acceleration
GAPDH	-	Glyceraldehyde-3-phosphate dehydrogenase
GUS	-	ß-Glucuronidase
HCl	-	Hydrochloric acid
HPT	-	Hygromycin phosphotransferase gene
Kinetin	-	6-Furfurylaminopurine
C-terminal	-	Carboxyl terminal
L	-	Liter
LB	-	Lauria-Bertani
M	-	Molar
mg	-	Milligram (s)
$MgCl_2$	-	Magnesium chloride

$MgSO_4$	-	Magnesium sulphate
min	-	Minute (s)
mL	-	Milliliter
mM	-	Millimolar
MOPS	-	3-(N-morpholino) propanesulfonic acid
MS	-	Murashige and Skoog
NaCl	-	Sodium chloride
kb	-	Kilobase
NAA	-	µ-Napthtalene acetic acid
NaOAc	-	Sodium acetate
NaOH	-	Sodium hydroxide
BSA	-	Bovine serum albumin
NCBI	-	National Center for Biotechnology Information
NH_4OAc	-	Ammonium acetate
Nos	-	Nopaline synthase gene terminator
^oC	-	Degree centigrade
OD	-	Optical density
ABA	-	Abscisic acid
ORFs	-	Open reading frames
PCI	-	Phenol: chloroform: isoamyl alcohol
x	-	Times
PCR	-	Polymerase Chain Reaction
Pfu	-	Plaque forming unit
pI	-	Isoelectric point
Poly (A)	-	Polyadenylated (mRNA)
rpm	-	Revolution per minute
PVP	-	Polyvinylpyrrolidone
N-terminal	-	Amino-terminal
qPCR	-	Quantitative PCR
R^2	-	Correlation coefficient
RNA	-	Ribonucleic acid
RNase	-	Ribonuclease
TBE	-	Tris-borate EDTA
RT-PCR	-	Reverse transcription – PCR
Jacq.	-	Jacquin
SA-AP	-	Streptavidine alkaline-AP
SDS	-	Sodium dodecyl sulfate
sec	-	Second
TAE	-	Tris-acetate-EDTA

TBE	-	Tris-borate-EDTA
T-DNA	-	Transferred-DNA
TE	-	Tris-EDTA
T_m	-	melting temperature
Tris	-	Tris [hydroxymethyl] aminomethane
Tris-HCl	-	Tris-hydrochloride
U	-	Unit
UTR	-	Untranslated region
Ubq		Ubiquitin
UV	-	Ultraviolet
V	-	Volt
v/v	-	Volume per volume
w/v	-	Weight per volume
µl	-	Microliter
µM	-	Micromolar
CaMV 35S	-	Cauliflower Mosaic Virus 35S