UNIVERSITI PUTRA MALAYSIA

TRANSCRIPTOMIC STUDY OF *GRACILARIA CHANGII* (GRACILARIALES, RHODOPHYTA) BY EXPRESSED SEQUENCE TAGS AND cDNA MICROARRAY APPROACH

TEOH SEDDON

FBSB 2009 12

TRANSCRIPTOMIC STUDY OF *GRACILARIA CHANGII* (GRACILARIALES, RHODOPHYTA) BY EXPRESSED SEQUENCE TAGS AND cDNA MICROARRAY APPROACHES

By

TEOH SEDDON

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

July 2009

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

TRANSCRIPTOMIC STUDY OF *GRACILARIA CHANGII* (GRACILARIALES, RHODOPHYTA) BY EXPRESSED SEQUENCE TAGS AND cDNA MICROARRAY APPROACH

By

TEOH SEDDON

July 2009

Chairman : Ho Chai Ling, PhD

Faculty : Biotechnology and Biomolecular Sciences

Gracilaria is one of the most extensively harvested seaweeds throughout the world due to its economic importance as an agarophyte for global agar production. In this study, *Gracilaria changii*, an indigenous seaweed species in Malaysia known for its high quality agar was chosen for transcription profiling. A total of 990 expressed sequence tags (ESTs) consisting of 766 tentative unique genes (TUGs) have been generated. These TUGs comprise 643 TUSs (tentative unique singletons) and 123 TUCs (tentative unique contigs). The putative identity of TUGs was identified by using Basic Local Alignment Search Tool X (BLASTX) algorithm and classified according to the functional groups in Kyoto Encyclopedia of Genes and Genomes (KEGG). The result showed that 198 TUGs (25.85%) have significant matches to the annotated proteins and 81 TUGs (10.57%) have significant matches to the unknown proteins in the non-redundant protein database in the GenBank; whereas the remaining 487 (63.58%) TUGs had non-significant matches or no matches to sequence from other organisms. Similar to animals and plants, *G. changii* showed preference for purine residues at -3 position and guanidine at +4 position at the

translational initiation signal. On the other hand, the 3' untranslated region of *G. changii* was found to have relatively less stringent polyadenylation process compared with plants and animals in producing mature mRNA. A cDNA microarray consisting of approximately 3,000 cDNA probes was constructed and used for the hybridization of cDNAs synthesized from *G. changii* samples cultured under conditions with and without light to understand the genetic acclimation of *G. changii* to light deprivation. The results suggested that genes related to photosynthesis, oxidative stress and sulfate metabolism were down-regulated during light deprivation. The cDNA microarray data were further verified by using real-time PCR. The results of the real-time PCR analysis of four genes encoding light-harvesting complex I polypeptide (DV962275.1), low molecular mass early light-inducible protein (DV964113.1), 14-3-3 protein (DV965610.1) and sonic hedgehog protein precursor (DV967367.1) supported the expression patterns demonstrated using cDNA microarray.

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

PENGAJIAN TRANKSRIPTOMIK TERHADAP *GRACILARIA CHANGII* (GRACILARIALES, RHODOPHYTA) DENGAN MENGGUNAKAN CARA TAGS JUJUKAN EKSPRES DAN MICROARRAY cDNA

Oleh

TEOH SEDDON

Juli 2009

Pengerusi : Ho Chai Ling, PhD

Fakulti : Bioteknologi and Sains Biomolekul

Gracilaria merupakan salah satu daripada rumpai laut yang paling banyak dituai di seluruh dunia disebabkan oleh kepentingan ekonominya sebagai agarofit untuk pengeluaran agar sejagat. Dalam kajian ini, *Gracilaria changii*, satu spesies rumpai laut asli di Malaysia yang terkenal dengan agarnya yang berkualiti tinggi telah dipilih untuk pemprofilan transkripsi. Sebanyak 990 tag turutan terekspres (EST) yang terdiri daripada 766 gen unik jangkaan (TUGs) telah dijanakan. TUG ini mengandungi 643 TUS (singleton unik jangkaan) dan 123 TUC (kontig unik jangkaan). Identiti jangkaan TUG ini dikenalpasti dengan menggunakan algoritma BLASTX dan diklasifikasikan mengikut kumpulan-kumpulan fungsian di “Kyoto Encyclopedia Genes dan Genomes” (KEGG). Keputusan menunjukkan bahawa 198 TUGs (25.85%) mempunyai padanan yang signifikan dengan protein yang telah dikenalpasti dan 81 TUGs (10.57%) mempunyai padanan yang signifikan dengan protein yang tidak dikenalpasti dalam pangkalan data protein tidak berulang (non-redundant protein) di GenBank; manakala 487 (63.58%) TUG yang lain tidak mempunyai padanan yang signifikan atau tiada padanan dengan jujukan daripada

organisma lain. Sepertimana di haiwan dan tumbuhan, *G. changii* mempunyai keutamaan terhadap purina di kedudukan -3 dan guanidin di kedudukan +4 pada isyarat permulaan translasi. Sebaliknya, bahagian hujung 3’ cDNA yang tidak ditranslasikan (3' UTR) *G. changii* didapati mempunyai keketatan yang agak kurang dalam proses poliadenilasi berbanding dengan tumbuhan dan haiwan dalam penghasilan mRNA yang matang. Satu cDNA mikroatur yang mengandungi kira-kira 3,000 prob cDNA telah dibina dan digunakan untuk penghibridan cDNA yang disintesis daripada sample *G. changii* yang dikultur di bawah atau tanpa cahaya bagi memahami penyesuaian genetik *G. changii* dalam keadaan tanpa cahaya. Hasil keputusan menunjukkan bahawa pengekspresan gen yang berkaitan dengan fotosintesis, stres oksidasi dan metabolisme sulfat berkurangan semasa ketiadaan cahaya. Data cDNA mikroatur telah dikenalpastikan selanjutnya dengan menggunakan kaedah tindak balas berantai polymeras masa nyata. Keputusan tindak balas berantai polymeras masa nyata ke atas empat gen terpilih iaitu ‘light-harvesting complex I polypeptide’ (DV962275.1), ‘low molecular mass early light-inducible protein’ (DV964113.1), ‘14-3-3 protein’ (DV965610.1) and ‘sonic hedgehog protein precursor’ (DV967367.1) menyokong corak pengekspresan gen yang dipamerkan dengan menggunakan mikroatur DNA.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my appreciation to my supervisor, Associate Professor Dr. Ho Chai Ling, for her patient and trust in me. Besides that, I would also like to thank her for her sincere guidance and help throughout my Master study. I also want to take this opportunity to thank Professor Phang Siew Moi and Professor Raha Abdul Rahim for being my supervisory committee members.

I wish to extend my gratitude to my friends from UPM: Swee Sen, Kok Ang, Chin Jit, Shu Shean, Swee Keong, Ky and Fung Chin for their advices and helps. Besides that, I would also like to thank Weng Wah and Chieh Wen for their technical advice during the course of this project, and members of Professor Phang Siew Moi's laboratory, especially Hui Yin and Mei Chea, for their help that seek to the completion of this project.

Last but not least, I am indebted to my family and my girl friend, Pey Jiun, for their encouragement, support and understanding. Their care and love have given me the strength to complete my study.

I certify that an Examination Committee has met on 10th of July, 2009 to conduct the final examination of Teoh Seddon on his Master of Science thesis entitled "Transcriptomic Study of *Gracilaria changii* (GRACILARIALES, RHODOPHYTA) by Expressed Sequence Tags and cDNA Microarray Approach" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the Master of Science.

Members of the Examination Committee are as follows:

Tan Soon Guan, PhD
Faculty of Biotechnology and Biomolecular Sciences
University Putra Malaysia
(Chairman)

Tan Wen Siang, PhD
Faculty of Biotechnology and Biomolecular Sciences
University Putra Malaysia
(Internal Examiner)

Suhaimi Napis, PhD
Faculty of Biotechnology and Biomolecular Sciences
University Putra Malaysia
(Internal Examiner)

Ismanizan Hj. Ismail, PhD
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
(External Examiner)

BUJANG KIM HUAT, Ph.D.
Professor/Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

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

Ho Chai Ling, PhD
Faculty of Biotechnology and Biomolecular Sciences
University Putra Malaysia
(Chairman)

Phang Siew Moi, PhD
Institut Sains Biologi
Universiti Malaya
(Member)

Raha Abdul Rahim, PhD
Faculty of Biotechnology and Biomolecular Sciences
University Putra Malaysia
(Member)

_____

HASANAH MOHD. GHAZALI, Ph.D.
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations 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 at any other institutions.

_____

TEOH SEDDON

Date: 14/08/2009

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ABC	ATP binding cassette
aRNA	Amplify ribonucleic Acid
ATP	2’-deoxy-adenosine-5’-triphosphate
BLAST	Basic Local Alignment Search Tool
bp	Base pair
cDNA	Complementary deoxyribonucleoside acid
CAP3	Contig Assembly Program 3
CTP	2’-deoxy-cytidine-5’-triphosphate
CO_2	Carbon dioxide
DEPC	Dietyl pyrocarbonate
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleoside acid
dNTP	Deoxynucleotides
E	Expectancy
EDTA	Ethylenediaminetetraacetic acid
EF-3	Elongation factor 3
EF-G	Elongation factor G
ELIP	Early light-inducible protein
EST	Expressed sequence tags
GTP	2’-deoxy-guanosine-5’-triphosphate
H^+	Hydrogen ion
H_2O	Water
H_2O_2	Hydrogen peroxide
KEGG	Kyoto Encyclopedia of Genes and Genomes
KCL	Potassium chloride
KO	KEGG orthology
kPa	Kilo Pascal
LiCl	Lithium Chloride
LOWESS	Locally weighted scatter plot smoother
Tris-Cl	Tris-chloride
MeV	Multiexperiment viewer

Mg^{2+}	Magnesium ion
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MIDAS	Microarray data analysis system
mRNA	Messenger Ribonucleic Acid
NADPH	Nicotinamide adenine dinucleotide phosphate
NaOH	Sodium Hydroxide
NCBI	National Center for Biotechnology Information
nr	Non-redundant
NUE	Near upstream element
OD	Optical Density
PAS	Polyadenylation signal
PCR	Polymerase Chain Reaction
RNA	Ribonucleic Acid
RNase	Ribonuclease
rRNA	Ribosomal ribonucleic acid
SDS	Sodium Dodecyl Sulphate
SPFH	Stomatin-prohibitin-flotillin-HflC/K
SSC	Sodium Chloride-Sodium Citrate buffer
TAE	Tris Acetate EDTA
™	Trade mark
TRE	T-rich element
TUCs	Tentative unique contigs
TUGs	Tentative unique genes
TUSs	Tentative unique singletons
UTP	2’-deoxy-urasil-5’-triphosphate
UTRs	Untranslated regions
UV	Ultraviolet

LIST OF SYMBOLS AND UNITS

cm	Centimeter
°C	Degree centrigrade
g	Gram
g	Gravitational force
h	Hour
µg/ml	Microgram per milliliter
µg	Microgramme
µl	Microliter
µM	Micromolar
µmol photons m^{-2} s^{-1}	Micromol photon per meter square per second
ml	Milliliter
mm	Millimeter
mM	Millimolar
min	Minute(s)
M	Molar
ng/µl	Nanogram per microliter
ng	Nanogramme
nm	Nanometer
nM	Nanomolar
ppt	Part per trillion
%	Percentage
Pfu	Plaque forming units
r.p.m	Revolution Per Minute
s	Second(s)
U/µl	Unit per microliter
V/cm	Volt per centimeter
v/v	Volume per volume
w/v	Weight per volume

CHAPTER 1

INTRODUCTION

The use of seaweed has a long history in human civilization. It has been tranditionally utilized as food additives, producers of phycocolloids, livestock feed, fertilizers and medicine (Hallmann, 2007). The ability of seaweeds to survive in various extreme or harsh environments may attribute to their highly adaptive life style and unique biochemical compounds. The discovery of phycocolloids had led to the application of seaweeds in a wide variety of industries. The seaweed industry is an ever-growing business. In year 2001, the total value of industrial products from seaweeds was US$ 590 million. The products from the seaweed industry have an estimated total value US$ 5.5-6 billion per annum (McHugh, 2003), and agar, which is uniquely produced by the red seaweeds (Rhodophyta) contributed a significant proportion of this value. *Gracilaria changii* (Xia *et* Abbott) Abbott, Zhang *et* Xia in the genus of *Gracilaria*, is one of the more abundant agarophytic rhodophyte found in Malaysia (Phang, 1994) which is known to produce high quality agar.

Despite their many usages and importance in industries, the systems biology of seaweeds is still largely unknown as only a few systematic studies have been carried out to understand the physiology of seaweeds at the molecular level. The first systematic study on seaweed was initiated by Lluisma and Ragan (1997) in which an expressed sequence tag (EST database) consisting of 200 ESTs was constructed from the marine red alga *Gracilaria gracilis*. Since then, the uses of EST approach have generated many seaweed ESTs from different species of seaweeds (Walker & Collet, 2005; Hallmann, 2007). Nevertheless, the number of ESTs from seaweeds in the EST

database is insignificant comparing to those from higher plants such as *Arabidopsis thaliana*, rice, wheat, grape and others.

Light is one of the major environmental factors that influences the growth and survival of submersed aquatic vegetation. Light deprivation caused by increased turbidity due to human activities such as urbanisation, deforestation, industrial pollution, dredging has been shown to affect the population of aquatic vegetation (Dennison *et al*., 1993; Longstaff *et al*., 1999) including seaweeds. The natural habitat of *G. changii* is located in the intertidal zone along densely populated coastlines in Malaysia. Therefore, it is important to understand the response of *G. changii* to light deprivation at the molecular level. This information may reveal the impact of water turbidity on the physiological adaptations of *G. changii*.

Functional genomic studies were made possible with the introduction of high-throughput molecular tools such as microarray and serial analysis of gene expression (SAGE) that have greatly enhanced the understanding of biological systems. Microarray experiments have been carried out on algae to study the cellular changes under different environmental conditions (Minoda *et al*., 2005; Collén *et al*., 2007). The functional genomic studies on *G. changii*, an important but less studied species using microarray will advance the understanding of its biology, which has important ecological and industrial implications.

The objectives of this study were:

1. To generate and analyse expressed sequence tags (ESTs) from *Gracilaria changii* cDNA library.

2. To fabricate a cDNA microarray for the global gene expression study of *G. changii*.
3. To examine the changes in gene expression profiles of *G. changii* under light deprivation by using cDNA microarray.
4. To validate the expression of selected cDNAs by using real-time PCR.

CHAPTER 2

LITERATURE REVIEW

2.1 An Introduction to algae

Algae are a diverse group of organisms from many different taxa. Basically they are organisms that contain pigment chlorophyll a and can carry out photosynthesis, besides land plants (Bhattacharya & Medlin, 1998). Algae consist of unicellular to multicellular organisms from prokaryotes to eukaryotes. The eukaryotic algae are among the earliest eukaryotes evolved from prokaryotes. Algae are classified into divisions based on the coloration and composition of their photosynthetic pigments. The four major taxonomic groups of algae are Cyanophyta, blue-green algae (prokaryotic cyanobacteria); Rhodophyta, red algae; Phaeophyceae, brown algae; and Chlorophyta, green algae (Prud'homme & Trono, 2001). Besides these criteria, the types of storage carbohydrates and the cell wall constituents are also used in the definition of algae groups (Inouye & Okamoto, 2005).

Similar to land plants, all eukaryotic algae contain photosynthetic organelles called chloroplasts. These chloroplasts are variable in nature among different lineages of algae. However, chroloplasts in all algae contain circular DNA and are fundamentally similar in structure to cyanobacteria, which represent reduced cyanobacterial endosymbionts (Bhattacharya & Medlin, 1995; Archibald, 2006; Yoon *et al*., 2002). This is referred as the primary endosymbiotic event, in which the origin of chloroplast was a result from the engulfment and retention of photosynthetic cyanobacteria by the ancient eukaryote, most probably a red alga (Bhattacharya & Medlin, 1998). The origin of photosynthetic plastids was previously

believed to happen several times, independently throughout evolutionary course giving rise to parallel lineages of photosynthetic organisms (Stiller *et al*., 2003). However, other molecular phylogenetic studies had suggested a contradicting hypothesis in which monophyletic origin of photosynthetic plastids was evidenced (Moreira *et al*., 2000; McFadden & Van Dooren, 2004).

2.2 Seaweeds

Seaweeds are marine benthic multicellular macroalgae that dominate the ocean vegetation. Multicellularity has some advantages in their physiological ecology (Lobban & Harrison, 1994). An important feature of being multicellular is to allow extensive development in three dimensional structures, which is beneficial in acquiring nutrients from a greater volume of water and, in resisting environmental stresses. Multicellularity also allows differentiation and specialization of tissue and their function, hence optimizing the efficiency of biological functions such as photosynthesis and structural scaffolding. Different types of seaweeds have differentiated to various degrees. Nevertheless, the diversity of cells in seaweeds are generally lower both physiologically and biochemically, than vascular land plants. Although they may have developed some of the structures that are equivalent to that of plants, seaweeds are not true vascular plants (Graham & Wilcox, 2000). This is because they do not have specialized vascular system such as xylems and phloems and, obviously do not have enclosed reproductive system like flowers or cones.

Seaweeds are plant-like oxygenic autotrophs, thus they share some similar structural and functional characteristics (Andersen, 1992). Plants have roots, while seaweeds have specialized holdfasts which act as anchors for attachment in the oceans. The

stem of seaweed is called a stipe. It provides support for the whole plant. The construction and physical properties of stipe vary in seaweeds. The leaves of seaweeds are called blades which are responsible in carrying out photosynthesis (Dring, 1982). In some species, the blades also support the reproductive structures of the seaweed. The one structure of plant that has no equivalent in seaweeds is the vascular system. Seaweeds do not need an internal conducting system as their constant submergence in water provides a high surface area to volume ratio for each individual cell to take up fluids, nutrients, and gases directly from the water (Denny & Gaylord, 2002; Druehl, 2000).

Nowadays, there are nearly ten thousand described species of seaweeds throughout the world. Most of them are red seaweeds, or rhodophyta (~6000 species), brown seaweeds, or phaeophyta (~2000 species) or green seaweeds, or chlorophyta (~1200 species). With all of them contain chlorophyll a; some have chlorophyll b or c; certain species have accessory pigments such as phycocyanin (blueish), phycoerythrin (reddish), carotenes (yellow-brown), and xanthophylls (brown) (Gantt, 1990). Other features used to classify seaweeds are cell wall composition, reproductive characteristics, morphology differences, and the chemical nature of their photosynthetic products (oil and starch).

2.3 Uses of Algae – An Overview

The algae, a highly diverse group of photoautotrophs, play a vital role in maintaining global productivity and biogeochemical cycling (Grossman, 2005). The broad diversity among the algae not only reflects on the phenotypes of the organisms, but also in the production of wide ranges of chemical compounds through unique

biosynthetic pathways. These diversified traits of algae make them highly useful for various commercial and industrial utilizations.

2.3.1 Human Food and Animal Feed

Seaweeds have been extensively used as food, particularly in the Orient since ancient time (McHugh, 2003). They can be served as a course or food additives. In the past, they are consummed as food mostly due to their availability and perhaps as delicacies. Nowadays, seaweeds gain more popularity for their nutritional values. They are rich in polysaccharides, proteins, minerals such as iodine, potassium, iron, magnesium and calcium, and almost all essential vitamins namely A, B, B2, B6, B12, C, E, nicotinate, biotin, folic acid and pantothenic acid (Dawczynski *et al*., 2007).

Seaweeds are also used as feed for aquaculture and livestocks. They contain microminerals and trace elements which enhance the nutritional values of conventional feed preparations (Becker, 2004). As in aquaculture, seaweeds are normally grown in a polyculture manner (Chopin *et al*., 2001). Seaweeds grown adjacent to aquaculture sites can be utilized directly as live feed for fish, mollusk, abalone and crustacean (Brown, 2002). In turn, the waste from fed aquacultures may provide nutrients for seaweeds to grow. Thus, this approach provides nutrients bioremediation capability which mutually benefits the co-cultured organisms.

2.3.2 Production of Phycocolloids

The most widely commercial uses of seaweeds in today world market is for phycocolloids (agar, carrageenan and alginate) manufacturing (Feizi & Mulloy,