ABSTRACT

Gracilaria is a red seaweed that has been cultivated worldwide and is commercially used for food, fertilizers, animal fodder and hydrocolloids. However, the high morphological plasticity and the lack of distinctive reproductive structures often lead to the misidentification in the traditional identification of *Gracilaria* species. Molecular markers are important especially in the correct identification of *Gracilaria* species with high economic value. Various types of molecular markers, for example random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphism (AFLP) have been used in the study of seaweeds but there are always limitations arising from each technique. Microsatellite markers or simple sequence repeats (SSRs), are the markers of choice as they have shown to be more polymorphic, relatively abundant, of multiallelic nature and co-dominant inheritance in the molecular study of plants.

The objective of this research is to mine the microsatellite markers from chloroplast genome and expressed sequence tag (EST) database of *Gracilaria* species deposited at the GenBank using available bioinformatics tools. The genomic SSR markers and EST-SSR markers obtained were then be used to assess their suitability for differentiating between different populations and species of *Gracilaria*. We also compared the variability of *cox*1 gene marker with the microsatellite marker that we developed and examined the polymorphism, number and pattern of SSRs from both genomic and ESTs of *Gracilaria tenuistipitata*. Most of the *Gracilaria* specimens were collected at various localities in Malaysia and some of the specimens were provided by the collaborators from other countries.

For the analysis on *G. tenuistipitata*, eight SSRs were obtained from the chloroplast genome of *G. tenuistipitata*. Two primer-pairs (GT5 and GT8) showed

polymorphisms on the specimens tested and the combinations of these two primer-pairs were able to differentiate *G. tenuistipitata* from the west coast of Peninsular Malaysia from populations facing the South China Sea. In addition, ten SSRs were obtained from the ESTs of *G. tenuistipitata* and the combined dataset of two primer-pairs (GT12 and GT18) generated four genotypes on the specimens tested and the populations from Kuah (Malaysia) and Pattani (Thailand) were grouped into two distinct clades.

For the analysis on *Gracilaria* species, one (primer-pair P3) out of 33 primer-pairs developed from the ESTs of *Gracilaria* species was able to distinguish between three different *Gracilaria* species, namely *G. changii*, *G. fisheri* and *G. manilaensis* which are morphologically indistinguishable. This marker can also differentiate the same species of *Gracilaria* from different populations, for example *G. changii* from Morib, Selangor has its unique allele that can be distinguished from other populations.

Comparison of two different molecular markers, primer-pair P3 and *cox*1 gene showed that *cox*1 gene is more variable than the microsatellite marker that we developed. Six haplotypes (C1 – C6) were obtained using *cox*1 gene while only three genotypes (M1 – M3) were obtained using primer-pair P3. Our study also showed that the number and polymorphism of SSR markers obtained from ESTs were higher than the genomic SSR markers but more different kind of motifs (mono-, di-, and tri-nucleotide) were observed in genomic SSR markers.

Development of molecular markers, particularly the microsatellite markers in distinguishing different populations and across related species, is essential to select valuable strains of the species for cultivation. Further studies in developing microsatellite markers from seaweeds with high economic value such as *Gracilaria* sp. will be most beneficial to the seaweed industry.

ABSTRAK

Gracilaria adalah rumpai laut merah yang ditanam di seluruh dunia dan digunakan secara komersial untuk makanan, baja, makanan haiwan dan hydrocolloids. Walau bagaimanapun, keplastikan morfologi yang tinggi dan kekurangan struktur pembiakan tersendiri sering membawa kepada pengenalan yang salah dalam spesies *Gracilaria* dengan menggunakan cara tradisional. Penanda molekul adalah penting dalam pengenalan spesies *Gracilaria* yang betul terutamanya bagi rumpai laut yang mempunyai nilai ekonomi yang tinggi. Pelbagai jenis penanda molekul seperti random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP) dan amplified fragment length polymorphism (AFLP) telah digunakan dalam kajian rumpai laut tetapi setiap teknik mempunyai kelemahannya yang tersendiri. Microsatellite markers atau simple sequence repeats (SSRs), adalah penanda pilihan kerana mereka lebih polimorfik, mempunyai bilangan yang banyak, bersifat multiallelic dan co-dominant dalam kajian molekul untuk tumbuhan.

Objektif kajian ini adalah untuk mecari penanda mikrosatelit dari genom kloroplas dan pangkalan data expressed sequence tag (EST) spesies *Gracilaria* yang disimpan di GenBank dengan menggunakan perisian bioinformatik yang sedia ada. Penanda SSR genomik dan penanda EST-SSR akan digunakan untuk menilai kesesuaian mereka untuk membezakan antara populasi dan spesies *Gracilaria* yang berbeza. Kami juga membandingkan kepelbagaian antara penanda *cox*1 gen dengan penanda mikrosatelit yang kami bentukan dan memeriksa polimorfisme, nombor dan corak SSRs dari genomik dan ESTs *Gracilaria tenuistipitata*. Kebanyakan spesimen *Gracilaria* dikutip di pelbagai kawasan di Malaysia dan specimen juga disumbangkan oleh kolaborator dari negara-negara lain.

Bagi analisis untuk *G. tenuistipitata*, lapan SSRs telah diperolehi dari genom kloroplas *G. tenuistipitata*. Dua pasangan primers (GT5 dan GT8) menunjukkan polimorfisme dalam spesimen yang diuji. Kombinasi kedua-dua pasangan primers dapat membezakan *G. tenuistipitata* dari pantai barat Semenanjung Malaysia dari populasi menghadap Laut China Selatan. Di samping itu, sepuluh SSRs telah diperolehi dari ESTs *G. tenuistipitata* dan gabungan dataset daripada kedua-dua pasangan primers (GT12 dan GT18) memberi empat genotip pada specimen yang diuji. Populasi dari Kuah (Malaysia) dan Pattani (Thailand) dikumpul dalam dua kled yang berbeza.

Bagi analisis untuk spesies *Gracilaria*, satu (pasangan primers P3) daripada 33 pasangan primers yang dibentukan dari ESTs spesies *Gracilaria* (*G. changii*, *G. gracilis* dan *Gp. lemaneiformis*) dapat membezakan antara tiga *Gracilaria* spesies yang morfologinya tidak dapat dibezakan, iaitu *G. changii*, *G. fisheri* dan *G. manilaensis*. Penanda ini juga dapat membezakan *Gracilaria* spesies yang sama tetapi dari populasi yang berbeza, sebagai contoh *G. changii* dari Morib, Selangor mempunyai allele yang unik yang dapat dibezakan dari populasi lain.

Perbandingan antara dua penanda molekul yang berbeza iaitu pasangan primers P3 dan *cox*1 gen menunjukkan *cox*1 gen lebih pelbagai daripada penanda mikrosatelit yang kami bentukan. Enam haplotaip (C1 - C6) diperolehi dengan menggunakan *cox*1 gen manakala hanya tiga genotip (M1 - M3) diperolehi dengan menggunakan pasangan primer P3. Kajian kami juga menunjukkan bahawa bilangan dan polimorfisme penanda SSR yang diperolehi dari ESTs adalah lebih tinggi daripada penanda SSR genomik tetapi lebih banyak jenis motif yang berbeza (mono-, di-, dan tri-nukleotida) dapat di cari dalam penanda SSR genomic.

Pembentukan penanda molekul, terutamanya penanda mikrosatelit dalam membezakan populasi dan spesis yang berkaitan, adalah penting untuk memilih jenis

spesies yang bernilai tinggi untuk penanaman. Penyelidikan lanjut dalam pembentukan penanda mikrosatelit daripada rumpai laut dengan nilai ekonomi yang tinggi seperti *Gracilaria* sp. adalah sangat bermanfaat kepada industri rumpai laut.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my utmost gratitude to my supervisor, Dr. Lim Phaik Eem, for granting me the opportunity to work on this project, and for her kindness, valuable guidance, and encouragement in the preparation and completion of this study. Special thanks to my co-supervisor, Prof. Dr. Phang Siew Moi for providing us with a good environment and facilities to complete this project, for her unselfish and unfailing support, and unrelenting patience throughout my project especially on manuscript writing.

I would also like to thank Mr. Lee Weng Wah, assistant Vice-President at ACGT Sdn. Bhd. for his kind assistance and advice in the bioinformatics part of this project. Mr. Lee has shared his priceless knowledge on bioinformatics specifically using Linux as operating system in the mining of microsatellite markers.

I would like to gratefully acknowledge Dr. David Edward from Institute for Molecular Biosciences and School of Land, Crop and Food Sciences, University of Queensland, Australia and his team for hosting my visit and giving me the opportunity to learn some bioinformatics techniques that are useful in my study. Special thanks to Prof. Khanjanapaj Lewmanomont, Assoc. Prof. Dang Diem Hong, Dr. Anchana Prathep, Dr. Danilo B. Largo and Miss Nurridan Abdul Han in providing some of the specimens used in this study.

In my daily work, I have friends and lab mates that work together with me, supporting me and giving their assistance whenever I need help. I would like to express my heartfelt thanks to Fiona Keng Seh Lin, for listening patiently to me, giving her moral and spiritual support, and being a good companion at all the time. My sincere thanks also go to Yoon Yen and Mei Chea, for giving their greatest assistance and

cooperation by readily providing me the needs for the research. I would say we are lucky to have such good seniors.

I am thankful to Poh Kheng and Chew Hock for keeping us entertained with their humorous stories and of course giving their hand with no complaint; Sze Wan for giving me suggestions and helped me to improve on article writing and Tan Ji for his assistance on solving any computer problems. You all have impressed me with the capability you have and the willingness to help each other when in need. I would also like to acknowledge all the members of the Algae Research Laboratory and staff of Institute of Ocean and Earth Sciences (IOES) for so much support throughout this research project.

I would like to thank University of Malaya for providing me the scholarship throughout my study and also funding from University of Malaya Research Grant Scheme (RG025-09SUS) and Postgraduate Research Fund (PS304-2009B, PS269-2010B and PV082-2011B).

Finally, I take this opportunity to extend my deep appreciation to my beloved family and friends for their love and blessings, understandings and faithful supports along this journey. This thesis wouldn't be completed without the help from anyone of you. Thank you!

Earning a PhD is like a long marathon; you can't stop and have to keep pace until you reach the finishing line.

"Struggling and suffering are the essence of a life worth living. If you're not pushing yourself beyond the comfort zone, if you're not demanding more from yourself - expanding and learning as you go - you're choosing a numb existence. You're denying yourself an extraordinary trip." ― Dean Karnazes, 2006

CONTENTS

LIST OF FIGURES **PAGE**

LIST OF TABLES **PAGE**

LIST OF APPENDICES **PAGE**

LIST OF ABBREVIATIONS

AFLP	amplified fragment length polymorphism
$AgNO_3$	silver nitrate
CAP3	contig assembly program version3
cDNA	complimentary deoxyribonucleic acid
cp DNA	chloroplast deoxyribonucleic acid
CsCl	cesium chloride
CTAB	cetyl trimethyl ammonium bromide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
EST	expressed sequence tags
EtBr	ethidium bromide
FAO	food and agriculture organization
LiCl	lithium chloride
MAGE	metaphor agarose gel electrophoresis
MISA	MIcroSAtellite search module
mt DNA	mitochondrial DNA
n	chromosome number
nrDNA	nuclear ribosomal DNA
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
RAPD	random amplified polymorphic DNA
rbcL	large-subunit of rubisco
rbcS	small-subunit of rubisco

LIST OF ABBREVIATIONS, continued.

rRNA	ribosomal ribonucleic acid
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RuBisCO	ribulose-1,5-bisphosphate carboxylase/ oxygenase
SNP	single nucleotide polymorphism
SSLP	simple sequence length polymorphisms
SSR	simple sequence repeat
SSU	small subunit
STR	short tandem repeats
Ta	annealing temperature
Tm	melting temperature
tRNA	transfer RNA
UHQ	ultra high quality
UPGMA	unweighted pair group method using arithmetic averages
URP	universal rice primer
UV	ultra-violet

LIST OF SYMBOLS AND UNITS

%	percent
℃	degree celsius
bp	base pair
µg	microgram
µL	microlitre
µm	micrometre
cm	centimetre
g cm^{-2}	gram per square centimetre
kb	kilobase
km	kilometre
kg	kilogram
M	molar
m	meter
mg	milligram
min	minute
mL	millilitre
mM	millimolar
ng	nanogram
nm	nanometre
pg	picogram
pmol	picomolar
rpm	revolutions per minute
U	unit