

**DEVELOPMENT AND CHARACTERIZATION OF  
SINGLE LOCUS DNA MARKERS FOR THE  
SEAWEEDS, *Kappaphycus* AND *Eucheuma*  
spp. IN SABAH, MALAYSIA**



**SALHAUDIN BIN MAILI MOHD RASLI**

**UMS**  
UNIVERSITI MALAYSIA SABAH

**BIOTECHNOLOGY RESEARCH INSTITUTE  
UNIVERSITI MALAYSIA SABAH  
2015**

**DEVELOPMENT AND CHARACTERIZATION OF  
SINGLE LOCUS DNA MARKERS FOR THE  
SEAWEEDS, *Kappaphycus* AND *Eucheuma*  
spp. IN SABAH, MALAYSIA**

**SALAHAUDIN BIN MAILI MOHD RASLI**



**UMS**

**THESIS SUBMITTED IN FULFILLMENT FOR THE  
DEGREE OF MASTER OF SCIENCE**

**BIOTECHNOLOGY RESEARCH INSTITUTE  
UNIVERSITI MALAYSIA SABAH  
2015**

## **DECLARATION**

The thesis was written by me and additional information has been properly cited in this thesis.

May 4, 2014

---

**SALHAUDIN BIN MAILI MOHD RASLI**

**MZ1211007T**



**UMS**  
UNIVERSITI MALAYSIA SABAH

## CERTIFICATION

NAME : **SALHAUDIN BIN MAILI MOHD RASLI**

MATRIC NO : **MZ1211007T**

TITLE : **DEVELOPMENT AND CHARACTERIZATION OF SINGLE LOCUS DNA MARKERS FOR THE SEaweEDS, *Kappaphycus* AND *Euचेuma* spp. IN SABAH, MALAYSIA**

DEGREE : **MASTER OF SCIENCE (BIOTECHNOLOGY)**

DATE OF VIVA VOCE : **NOVEMBER 12, 2014**

 **DECLARED BY**

**1. SUPERVISOR**  
Ms. Grace Joy Chin Wei Lie Signature

**2. CO-SUPERVISOR**  
Dr. Kenneth Francis Rodrigues \_\_\_\_\_

UMS  
UNIVERSITI MALAYSIA SABAH

## **ACKNOWLEDGEMENT**

In the name of Allah the most Gracious and the most Merciful. I would like to dedicate this work to my parents (Norsiah and Maili) and siblings (Siti, Ryen, Pian, Joe and Ali). I will not be able to repay you; for all that you have done for me all these years. I believe that I would not be able to reach this stage without their helps and prayers.

I would like to express my deepest appreciation to my dear supervisors, Ms. Grace Joy Chin Wei Lie and Dr. Kenneth Francis Rodrigues for their combined efforts, sparkling guidance and opinions throughout this research. I always believe that their love, patience for me and passion for research have brought me to the place where I am today. I am most grateful for their constant support and interest in this project. I would also like to thank Prof. Dr. Clemente Michael Wong Vui Ling (Director of Biotechnology Research Institute, UMS), Prof. Datin Seri Panglima Dr. Ann Anton, Prof. Datuk Seri Panglima Dr. Kamaruzaman Hj. Ampon, Assoc. Prof. Dr. Lee Ping Chin, Assoc. Prof. Dr. Jualang Azlan Gansau, Assoc. Prof. Dr. Zaleha Abdul Aziz, Assoc. Prof. Dr. Ivy Wong Nyet Kui, Assoc. Prof. Dr. How Siew Eng, Assoc. Prof. Dr. Vijay Kumar, Dr. Roziah Kambul, Mdm. Zarina Amin, Mdm. Roslina Jawan, Ms. Hartinie Marbawi, Ms. Cheong Bo Eng, Dr. Shafiqquzaman, Dr. Moh, Dr. Iqbal and Mr. Wilson Yong, since they are among the one who encouraged me to do my research with passions from the time I came to UMS.

I want to thank all the staff of the Biotechnology Research Institute, UMS for helping and encouraging me during my study. My special thanks goes to my friends, Syed Azman, Ruzzaini, Jordan, Stanly, Fernandes, Fauze, Yungchie, Zarifi, Afizah, Ashwini, Audrey Rose, Fui Fui, Azriah, Shafawati, Haslan, Jaclyn, Tamar, Raimi, Joseph, Zaidi, Ranjita, Rolando, Sadat, Yong, Vunyeey, Tyng, Hidayaty, Cheewei, Drake, Faez, Vidarita, Imran, Adam, Azian, Mony, Rudy, Christina, Yusdi, Ms. Roslina (Cocoa Biotechnology Research Center) and all the staffs of Yayasan UMS (Assoc. Prof. Dr. Suhaimi, Jessy, Jafson, Wan and Jason). To all my friends for helping me, thank you for being an important part of my life. This research was supported by PEMANDU (GPRL/SPS/1). It has been a long and wonderful journey for me and I would like to take this privilege to give my wholehearted thanks. My best wishes to all of you and may your future always shines.

**SALAHAUDIN MAILI MOHD RASLI**

MAY 4, 2014

## ABSTRACT

The red algae belonging to the genus of *Kappaphycus* and *Eucheuma* seaweeds are key sources of carrageenan. The genomic information of *Kappaphycus* and *Eucheuma* in public database at present is very limited. The paucity of such genomic information may become a hurdle for marker development. Different varieties of *Kappaphycus* and *Eucheuma* are known to have different agronomical value. Despite their high economic values, the genetic information of domesticated and wild varieties remains unclear. The present study was carried out to develop and characterize single locus DNA markers derived from three different partial small-insert genomic libraries as a solution for the molecular characterization of *Kappaphycus* and *Eucheuma* in Sabah, Malaysia. A genomic library was constructed and a total of 179 positive clones was selected and sequenced for their inserts. Primers were designed for 154 sequences and an array of 147 locus specific primers were selected and assayed against 13 congeners of *Kappaphycus* and *Eucheuma* seaweeds. Database searches and other bioinformatics analyses were performed on NCBI website and other online tools revealing a large proportion of 133 sequences with no BLASTn hits and several significant hits to known and unknown protein for BLASTx. A total of 117 primers amplifying single locus was observed and another 32 primers failed to produce clear and homogenous single locus amplification. Two primers amplified in all tested varieties, 18 primers amplified only with *Kappaphycus* and 44 primers amplified only in *Eucheuma*. The remaining 53 primers showed variable amplifications across the *Kappaphycus* varieties. Jaccard similarity coefficient revealed low genetic variations among the tested variety of *Kappaphycus* and *Eucheuma* seaweeds. Unweighted Pair Group Method with Arithmetic Mean (UPGMA) tree was used in clustering procedure based on the binary matrix scores. The clustering was found to be congruent, with some inconsistencies, with the previously reported phylogenetic tree reconstruction of the *Eucheuma* and *Kappaphycus* seaweeds by revealing four clusters including the newly found species of *Kappaphycus malesianus* from Sabah, Malaysia. A total of eight diagnostic markers were identified and selected for multiplex polymerase chain reaction (PCR) in the fingerprinting maps development for variety-specific identification to be used as rapid and reliable standard protocol for varietal identification of *Eucheuma* and *Kappaphycus* seaweeds. We demonstrated that the primer arrays were able to assign the varieties to their respective species and reveal inter- and intraspecific relationship between the tested varieties. This study also represents the first step in an experimental breeding programme for *Kappaphycus* and *Eucheuma*. In future, these DNA markers can be applied to the biotechnological improvement of *Kappaphycus* and *Eucheuma* seaweeds for interspecific or intergeneric hybrid development through protoplast fusion technology by using the single locus DNA marker as a verification tool and also for seaweeds variety protection purposes.

## **ABSTRAK**

### **PEMBANGUNAN DAN PENCIRIAN PENANDA LOKUS TUNGGAL DNA UNTUK RUMPAI LAUT, *Kappaphycus* DAN *Eucheuma* spp. DI SABAH, MALAYSIA**

*Alga merah jenis Kappaphycus dan Eucheuma adalah sumber utama untuk Karagenan. Maklumat genomik Kappaphycus dan Eucheuma dalam pangkalan data awam pada masa ini adalah sangat terhad. Kekurangan maklumat genomik ini boleh menjadi halangan kepada pembangunan penanda molekular. Pelbagai rumpai laut jenis Kappaphycus dan Eucheuma diketahui mempunyai ciri-ciri fenotip untuk pertanian yang berbeza. Walaupun mempunyai nilai ekonomi yang tinggi, maklumat genetik rumpai laut domestik dan liar masih tidak jelas. Kajian ini cuba untuk membangunkan dan mencirikan penanda DNA lokus tunggal diperolehi daripada tiga sumber separa kecil-insert perpustakaan genom yang berbeza sebagai penyelesaian untuk pencirian molekular Kappaphycus dan Eucheuma di Sabah, Malaysia. Sebuah perpustakaan genom telah dibina dan sebanyak 179 klon positif telah dipilih untuk proses penjujukan DNA. Sebanyak 154 lokus telah dikenalpasti sesuai untuk dijadikan rujukan untuk mereka primer, bagaimanapun, hanya sebanyak 147 primer lokus telah dipilih dan diuji terhadap 13 sampel rumpai laut jenis Kappaphycus dan Eucheuma. Carian pangkalan data dan analisis bioinformatik yang dilaksanakan pada laman web NCBI dan perisian dalam talian lain mendedahkan sebahagian besar daripada 133 jujukan DNA tanpa hit BLASTn dan beberapa hit kepada protein yang diketahui dan tidak diketahui untuk BLASTx. Sebanyak 117 primer yang menghasilkan lokus tunggal berjaya dikenalpasti dan 32 primer lain gagal untuk menghasilkan lokus tunggal yang jelas. Dua primer mampu diamplifikasi dalam semua varieti yang diuji, 18 primer diamplifikasi hanya dengan Kappaphycus dan 44 primer diamplifikasi hanya dalam Eucheuma. Sebanyak 53 primer menunjukkan amplifikasi yang berbeza-beza dalam kalangan rumpai laut jenis Kappaphycus. Pekali persamaan Jaccard mendedahkan variasi genetik yang rendah di kalangan rumpai laut Kappaphycus dan Eucheuma yang diuji. Pokok Unweighted Pair Group Method with Arithmetic Mean (UPGMA) telah dibina untuk proses pembahagian berasaskan kepada skor matriks binari. Kluster ini didapati bersesuaian dengan beberapa pembahagian yang tidak konsisten dengan pembinaan semula pokok filogenetik yang dilaporkan sebelum ini daripada rumpai laut jenis Eucheuma dan Kappaphycus dengan mendedahkan empat kluster termasuk spesies yang baru ditemui iaitu Kappaphycus malesianus dari Sabah, Malaysia. Sebanyak lapan penanda diagnostik telah dikenal pasti dan dipilih untuk multipleks tindakbalas berantai polimerase (PCR) untuk pembangunan peta cap jari untuk pengenalan varieti khusus telah berjaya dibangunkan untuk digunakan sebagai penanda pantas dan protokol standard yang boleh dipercayai untuk mengenalpasti identiti varieti daripada rumpai laut Eucheuma dan Kappaphycus. Kami menunjukkan bahawa koleksi primer dalam kajian ini dapat menentukan spesies untuk setiap varieti dan mendedahkan hubungan antara dan dalam varieti yang diuji. Kajian ini juga merupakan langkah pertama dalam program eksperimen pembiakan Kappaphycus dan Eucheuma. Pada masa akan datang, penanda DNA ini akan digunakan untuk membangunkan bioteknologi rumpai laut Kappaphycus dan Eucheuma untuk pembangunan hibrid antara spesis dan genus melalui teknologi gabungan protoplast dengan menggunakan locus DNA penanda tunggal sebagai alat pengesahan dan juga untuk tujuan perlindungan varieti rumpai laut.*

## TABLE OF CONTENTS

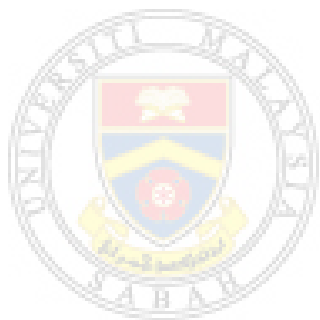
	Page
<b>TITLE</b>	i
<b>DECLARATION</b>	ii
<b>CERTIFICATION</b>	iii
<b>ACKNOWLEDGEMENT</b>	iv
<b>ABSTRACT</b>	v
<b><i>ABSTRAK</i></b>	vi
<b>TABLE OF CONTENTS</b>	viii
<b>LIST OF FIGURES</b>	xi
<b>LIST OF TABLES</b>	xiii
<b>LIST OF ABBREVIATIONS</b>	xv
<b>LIST OF APPENDICES</b>	xvii
<b>CHAPTER 1: INTRODUCTION</b>	1
1.1 Research background	1
1.2 Objectives	3
1.3 Significance of Study	4
<b>CHAPTER 2: LITERATURE REVIEW</b>	5
2.1 <i>Kappaphycus</i> and <i>Euचेuma</i> seaweeds	5
2.2 An Overview of Carrageenan Industry	9
2.2.1 Application of Carrageenan Derived from <i>Euचेuma</i> and <i>Kappaphycus</i>	12



2.3	Application of Genetic Molecular Markers in Seaweeds	14
2.3.1	Molecular Marker Type I and Type II	14
2.3.2	Mitochondrial, Chloroplast, Nuclear DNA marker	15
2.3.3	PCR-based Technique	18
2.3.4	Non DNA-based Markers	20
2.4	Construction of Genomic Library for Marker Discovery	22
2.4.1	Characterization of Single Locus DNA Marker	24
2.5	Application of Genome-wide Selection (GWS) for Breeding in Major Crops	27
2.6	Concluding remarks	28
<b>CHAPTER 3: MATERIALS AND METHODS</b>		30
3.1	Sampling of <i>Kappaphycus</i> and <i>Euचेuma</i> seaweeds	30
3.2	Genomic DNA Extraction from <i>Euचेuma</i> and <i>Kappaphycus</i>	31
3.3	Determination of Nucleic Acid Quality and Quantity	32
3.4	Construction of Partial Small-insert Genomic Library	32
3.4.1	Preparation of Competent Cell of <i>E. coli</i> TOP10	33
3.4.2	Ligation of Digested Genomic DNA with Plasmid pUC19	33
3.4.3	Transformation of Ligation Product into Competent cell of <i>E. coli</i> TOP10	34
3.4.4	Blue-white Screening for Recombinants Clones of <i>E. coli</i> TOP10	34
3.4.5	Colony PCR Screening for Recombinant Plasmid pUC19	35
3.4.6	Small Scale Preparation of Plasmid DNA of Selected <i>E. coli</i> TOP10 Transformant	35

3.5	Analysis of Nucleotide Sequence of Partial Small-insert Genomic <i>Kappaphycus</i> Seaweeds Libraries of <i>Eucheuma</i> and <i>Kappaphycus</i>	36
3.6	Bioinformatics Analysis	37
3.7	Primer design	38
3.8	Amplification of Designed Primers and Bands Scoring against 13 varieties of <i>Eucheuma</i> and <i>Kappaphycus</i> seaweeds	38
3.9	Analysis of Amplification Patterns of Designed Primers	39
<b>CHAPTER 4: RESULTS AND DISCUSSIONS</b>		40
4.1	Genomic DNA Extraction	40
4.2	Construction of Partial Small-Insert Genomic Libraries for <i>Kappaphycus alvarezii</i> , <i>Kappaphycus striatus</i> and <i>Eucheuma denticulatum</i> .	42
4.2.1	Insert Size Distribution and GC Content Distribution of Cleaned Inserts	44
4.2.2	Distribution of Simple Sequence Repeats	46
4.3	Functional Annotation for the Three Partial Small-insert Genomic Libraries	52
4.3.1	Similarity Searches using BLASTx against Non-redundant Protein Database (nr) in NCBI Database	52
4.3.2	Similarity Searches using BLASTx against Rhodophyta Seaweeds	57
4.3.3	Similarity Searches using using EST of and <i>Kappaphycus Eucheuma</i>	59
4.3.4	Similarity Searches using BLASTx against <i>Arabidopsis thaliana</i> .	65
4.4	Primer Design and Amplification	68
4.4.1	Primer Cross-species Amplification and Transferability	70
4.4.2	Statistical Analysis and DNA Fingerprint Map of <i>Kappaphycus</i> and <i>Eucheuma</i> Seaweeds	74

<b>CHAPTER 5: CONCLUSION</b>	89
5.1 Conclusions	89
5.2 Future Research	90
<b>REFERENCES</b>	93
<b>APPENDIX</b>	116



**UMS**  
UNIVERSITI MALAYSIA SABAH

## LIST OF FIGURES

	Page
Figure 2.1: <i>Kappaphycus</i> and <i>Eucauma</i> varieties in Malaysia.	6
Figure 2.2: Structural representations of iota-, kappa- and lambda-carrageenan.	11
Figure 2.3: The variability of three different classes of markers for different types of research.	16
Figure 2.4: The optimal marker that can be used for phylogeny, phylogeography, population genetic and mapping.	26
Figure 4.1: Electrophoretic analysis of total genomic DNA.	40
Figure 4.2: Construction of partial small-insert genomic libraries of KABA, KSGF and EDAD.	42
Figure 4.3: Example of colony PCR amplifications.	43
Figure 4.4: Size distribution of inserts size from three partial small-insert genomic libraries.	45
Figure 4.5: GC content distributions of the inserts from three partial small-insert genomic libraries.	47
Figure 4.6: Distribution of SSRs found in the three partial small-insert genomic libraries.	48
Figure 4.7: Distribution of SSRs motif found in the three partial small-insert genomic libraries.	48
Figure 4.8: Frequency of SSR motifs found in the three partial small-insert genomic libraries.	49
Figure 4.9: Statistic of the BLASTx hits annotated by nr database for the partial-small inserts genomic libraries of EDAD, KABA and KSGF.	52
Figure 4.10: Top hit species distribution using BLASTx similarity searches.	56
Figure 4.11: Proportion of queries producing significant BLASTx result against Rhodophyta.	57
Figure 4.12: Proportion <i>Kappaphycus</i> and <i>Eucauma</i> EST producing significant BLASTn result against the genomic library of ED, KA and KS.	59
Figure 4.13: Distribution of sequence derived from the genomic libraries producing significant BLAST result against the	60

EST database of *Eucheuma denticulatum* and *Kappaphycus alvarezii*.

Figure 4.14:	Distribution of sequence derived from the genomic libraries producing significant BLASTx result against Arabidopsis database.	65
Figure 4.15:	Total number of designed and tested primer derived small-insert genomic libraries of KABA, KSGF and EDAD.	68
Figure 4.16:	Example of single locus DNA marker amplifications of primer A (BHBA13), B (EHGF10) and C (BHAD02).	69
Figure 4.17:	Total number of amplified varieties tested with a panel of 117 markers derived from the small-insert genomic libraries of KABA, KSGF and EDAD.	70
Figure 4.18:	Transferability of marker derived small-insert genomic libraries of KABA, KSGF and EDAD.	73
Figure 4.19:	The UPGMA dendrogram based on Single locus DNA marker data showing the relationship between the tested varieties.	76
Figure 4.20:	The UPGMA dendrogram (consensus tree) based on Single locus DNA marker data showing the relationship between the tested varieties.	77
Figure 4.21:	Single locus PCR amplification of primers used in fingerprinting construction for multiplex set 1.	79
Figure 4.22:	Single locus PCR amplification of primers used in fingerprinting construction for multiplex set 2.	80
Figure 4.23:	Multiplex PCR amplification patterns of eight primers against 13 <i>Eucheuma</i> and <i>Kappaphycus</i> seaweeds.	82
Figure 4.24:	Fingerprinting map for varietal characterization of <i>Eucheuma</i> and <i>Kappaphycus</i> seaweeds.	83

## LIST OF TABLES

	Page
Table 2.1: List of countries in Southeast Asia involved in carrageenan industries and their production per year	12
Table 2.2: Commonly used nuclear, mitochondrial and plastidal marker for Rhodophyta.	17
Table 2.3: Classification and properties of different types of molecular markers	25
Table 3.1: List of <i>Kappaphycus</i> and <i>Eucheuma</i> varieties and location of sampling.	31
Table 3.2: Preparation of restriction enzyme double digestion.	34
Table 3.3: List of software used in bioinformatics analysis.	37
Table 4.1: Repeat analysis of the SSRs containing sequences.	50
Table 4.2: List of long SSRs containing sequences.	51
Table 4.3: Top BLASTx hit against nr database using Blast2go against non-redundant database (nr) in NCBI	54
Table 4.4: Distribution of BLASTx hits found in the three partial small-insert genomic libraries (BLASTx against Rhodophyta database).	58
Table 4.5: Distribution of BLASTn hits found in the three partial small-insert genomic libraries (BLASTn against <i>Eucheuma</i> and <i>Kappaphycus</i> EST database).	61
Table 4.6: Distribution of BLASTx hits found in the three partial small-insert genomic libraries (BLASTx against <i>Arabidopsis thaliana</i> ).	66
Table 4.7: Number of primers amplifying clear, strong bands and percent cross-amplification for three <i>Kappaphycus</i> and <i>Eucheuma</i> varieties across a panel of 13 varieties.	72
Table 4.8: Similarity matrix based on Single locus DNA marker data of 13 varieties of <i>Kappaphycus</i> and <i>Eucheuma</i> seaweeds calculated by Jaccard's coefficient.	75
Table 4.9: Potential specific primer for multilocus-based phylogenetic in <i>Kappaphycus</i> and <i>Eucheuma</i> seaweeds.	86
Table 4.10: Potential species- and variety-specific primers for <i>Kappaphycus alvarezii</i> .	87



## LIST OF ABBREVIATIONS

AFLP	- Amplified fragment length polymorphism
ANL	- Anonymous nuclear locus
ANMs	- Anonymous nuclear marker
BAC	- Bacterial artificial chromosome
BLAST	- Basic local alignment search tool
CMC	- Carboxymethylcellulose
cpDNA	- Chloroplast DNA
<i>Cox</i>	- Cytochrome c oxidase
CTAB	- Cetyltrimethylammonium bromide
DNA	- Deoxyribonucleic acid
dNTPs	- Deoxynucleotide triphosphates
DOF	- Department of fisheries
EDTA	- Ethylenediaminetetraacetic acid
EPIC	- Exon-primed intron-crossing
EPP	- Agriculture entry point projects
ESC	- European scientific committee
EST	- Expressed sequence tag
FAO	- Food and agricultural organization
FDA	- Federal drug administration
GC-MS	- Gas chromatography mass spectroscopy
GEV	- Genomic selection based on genomic breeding values
GWS	- Genome-wide selection
HPLC	- High performance liquid chromatography
IDT	- Integrated DNA technologies
ISSR	- Inter simple sequence repeats
ITS	- Internal transcribed spacer
LB	- Luria bertani
LC-MS	- Liquid chromatography mass spectroscopy
MAS	- Marker assisted selection
MCS	- Multi cloning site
MEGA 6	- Molecular evolutionary genetics analysis 6
MISA	- MIncroSATellite identification tool
NaCl	- Sodium chloride
NCBI	- National center for biotechnology information
NGS	- Next generation sequencing
NKEA	- National key economic areas
NMR	- Nuclear magnetic resonance
NPCL	- Nuclear protein coding loci
ORF	- Open reading frame
PacBio	- Pacific Biosciences
PCR	- Polymerase chain reaction
PVP	- Polyvinylpyrrolidone



QTL	- Quantitative trait loci
RAPD	- Randomly amplified polymorphic DNA
<i>rbcl</i>	- Ribulose-bisphosphate carboxylase
RedToL	- Red algal tree of life initiatives
RFLP	- Restriction fragment length polymorphism
RNA	- Ribonucleic acid
rpm	- Revolutions per minute
rRNA	- Ribosomal ribonucleic acid
SCAR	- Sequence characterized amplified regions
SMRT	- Single molecule real time sequencing
SNPs	- Single nucleotide polymorphisms
SSR	- Simple sequence repeat
UPGMA	- Unweighted pair-group method using arithmetic averages
X-gal	- 5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactopyranoside



UMS  
UNIVERSITI MALAYSIA SABAH

## LIST OF APPENDICES

	Page
Appendix A: Photo of <i>Eucheuma</i> and <i>Kappaphycus</i> Seaweeds used in this Study	116
Appendix B: Sequence Confirmation and BLAST Hit for <i>Kappaphycus malesianus</i> ( <i>Cox2-3</i> spacers).	117
Appendix C: Media and solutions	119
Appendix D: Restriction map of Plasmid pUC19	121
Appendix E: List of oligonucleotide primer derived from the partial small-insert genomic libraries (EDAD, KAGF and KSBA) used in PCR amplification analysis	122
Appendix F: Analysis of predicted coding region based on Open Reading Frame (ORF) (+3) of translated nucleotide sequences of SKGF10.	129
Appendix G: Analysis of predicted coding region based on Open Reading Frame (ORF) (- 2) of translated nucleotide sequences of SXBA04.	133
Appendix H: Sequence analysis with Blast2GO (Conesa <i>et al.</i> , 2005)	138
Appendix I : Amplification pattern of designed primers primer derived from the partial small-insert genomic libraries (EDAD, KAGF and KSBA) used in PCR amplification analysis	140
Appendix J: The NJ tree based on Single DNA marker data derived showing the relationship between the tested varieties. Group A (ED); B (KA/KM), C (KA) and D (KS).	146
Appendix K: Application of QiAxcel Advanced System for the multiplex set 1 and set 2	147

# CHAPTER 1

## INTRODUCTION

### 1.1 Research Background

The seaweed industry is one of the largest global industries with 7.4 % yearly average increments in production based on the report issued by Food and Agricultural Organization (FAO) in 2012. It is also estimated that the production of seaweed in the global market will reach 40 million tonnes by year 2020 (Gjedrem *et al.*, 2012). Owing to its great potential as an economically important crop for carrageenan production together with the increasing world demands on seaweeds, it has prompted forceful efforts to meet the current and future demands for carrageenan supplies in the global market. In Malaysia, seaweed industry has been recognized as one of the project under the Agriculture Entry Point Projects (EPP) in the 10<sup>th</sup> Malaysian plan. Currently, *Kappaphycus* spp. and *Eucheuma* spp. of the red algal genera Soliericeae (Rhodophyta, Gigartinales) have been given special attention for large scale cultivation in Malaysia for carrageenan supplies to the hydrocolloid industry. The interesting characteristic of the carrageenan has been the main subject of interest for biotechnological use intended for commercial purposes. *Kappaphycus* and *Eucheuma* are also frequently reported to have high nutrient compositions (Vairappan and Mikio, 2008; Matanjun *et al.*, 2008, 2010; Rajasulochana *et al.*, 2010; Abirami *et al.*, 2011) with valuable bioactive compounds that are pharmaceutically useful as an antibacterial (Vairappan and Mikio, 2008; Venkatesh *et al.*, 2011; Bibiana *et al.*, 2012), antifungal (Sridhar *et al.*, 2010), anticoagulant (Sun and Wu, 2007; de Araújo *et al.*, 2013) and antitumor (Yuan and Song, 2005; Yuan *et al.*, 2011). *Kappaphycus* spp. and *Eucheuma* spp. can also provide possible alternative for bioremediation (Rodrigueza and Montaño, 2007), bioethanol production (Khambhaty *et al.*, 2012) and synthesis of hydrogel for application in food industries (Distantina *et al.*, 2013).

Despite being of commercial potential, the taxonomy of *Eucheuma* and *Kappaphycus* are morphologically confusing with the existence of many phenotypic variants. Farmers have been relying on morphological characteristics for so many years which sometimes can lead to mixture of seaweeds in downstream processes due to mis-identification (Ganzon-Fortes, 2012). In addition, numerous reports about *Kappaphycus* and *Eucheuma* seaweeds on being an invasive plant in some coastal areas, such as Venezuela, Brazil and Hawaii (John, 2000; Barrios *et al.*, 2005; Conklin *et al.*, 2009; Castelar *et al.*, 2009). Conklin (2009) recommended rapid identification of these invasive seaweeds to be taken in order to combat the invasion of seaweed in Hawaiian coastal. Until recently, many investigators have focused their research on the phylogenetic study of seaweeds by employing different types of molecular markers to characterize *Eucheuma* and *Kappaphycus* seaweeds. Commonly used molecular markers, ribulose-bisphosphate carboxylase gene (*rbcL* gene), ribosomal ribonucleic acid (28S rRNA, 23S rRNA, ITS1, 5.8S rRNA and ITS2), mitochondrial cytochrome C oxidase (*cox1*, *cox2* and *cox2-3* spacer), which are mainly based on sequence comparative analysis of polymerase chain reaction (PCR) of targeted region or gene specific primer amplification were frequently reported in various articles (Zuccarello *et al.*, 2006; Conklin *et al.*, 2009; Zhao & He, 2011; Tan *et al.*, 2012; de-Barros Barretto *et al.*, 2012; Tan *et al.*, 2013). More recently, Liu *et al.*, (2011) has utilized inter simple sequence repeats (ISSR) to evaluate the genetic relationship between high carrageenan producing strains of *Kappaphycus alvarezii* while Randomly Amplified Polymorphic DNA (RAPD) and ISSR were used in combination for the genetic relationship study of *Kappaphycus alvarezii* in China (Zhao and He, 2011). These studies have proven that the utility of DNA markers as an alternative approach to aid in species identification and to complement morphological-based taxonomy.

In recent years, there has been an increase in the demand for molecular information pertaining to seaweeds for application in genomic studies of seaweed as well as for various applications, such as, molecular breeding of economically important seaweeds, especially, carragenophytes species (Zuccarello *et al.*, 2006). To date, there have been no fingerprinting developments for *Eucheuma* and *Kappaphycus* varieties for strain selection in Malaysia, except for the seaweed barcoding study carried out by Tan *et al.* (2012). In addition, little efforts have

been conducted to further characterize the genetic make-up of the existing *Kappaphycus* and *Eucheuma* varieties in Malaysia. There is no full genome sequence of either *Eucheuma* or *Kappaphycus* which has limited our understanding of the genomic information of both genera. Nonetheless, some transcriptomic data have been published for *Eucheuma* (Aspilla *et al.*, 2010; Hussein *et al.*, 2011) and *Kappaphycus* (Liu *et al.*, 2011). However, the data are incomplete. To provide clues about their genetic make-up, this research was designed to investigate the genetic variation of selected wild and locally cultivated strains of *Eucheuma* and *Kappaphycus* seaweeds in the vicinity of Sabah, Malaysia by utilizing randomly isolated and characterized Anonymous Nuclear Locus (ANL) as opposed to previous study which targeted known genome regions. For that reason, the development DNA-based molecular markers in *Kappaphycus* and *Eucheuma* seaweeds are vital in order to resolve the current problems of seaweed identification and fingerprinting as well as for future biotechnological applications.

## **1.2 Objectives of Study**

The objectives of this study are:

1. To develop partial small-insert genomic libraries of *Eucheuma denticulatum*, *Kappaphycus alvarezii* and *Kappaphycus striatus* seaweeds using randomly cloned inserts.
2. To characterize the molecular variation among local and wild seaweeds varieties in Sabah, Malaysia on the basis of single locus DNA markers.
3. To identify species- and variety-specific DNA-based molecular marker for fingerprinting in *Eucheuma denticulatum*, *Kappaphycus alvarezii* and *Kappaphycus striatus*.

### **1.3 Significance of Study**

The high potential of carrageenan in the industrial market has driven more cultivation of carragenophytes species in Sabah, Malaysia. This genetic study represents a preliminary genetic evaluation of the repetitively vegetated *Eucheuma* and *Kappaphycus* seaweeds by farmers in Sabah, Malaysia by monitoring their genetic diversity. This information will help in the selection of potential breeding candidates by providing simple, rapid and effective molecular tools for PCR-based identification of *Kappaphycus* and *Eucheuma* varieties with the development of unique and specific molecular markers for fingerprinting in *Kappaphycus* and *Eucheuma*. Moreover, this study will represent the first step towards Marker Assisted Selection (MAS) associated with high carrageenan producing varieties or varieties with agronomically desirable traits to increase the yield and quality of cultivated varieties of *Eucheuma* and *Kappaphycus* seaweeds in Malaysia. Besides, the development of an array of single locus DNA markers is likely to resolve or complements the limitation by other currently used marker systems in *Eucheuma* and *Kappaphycus* seaweeds, such as, chloroplast and mitochondrial derived markers. This will complement morphological-based classifications of the many varieties of *Eucheuma* and *Kappaphycus* seaweeds into a more refined classification. Early monitoring of the ecological impacts of *Eucheuma* and *Kappaphycus* seaweeds can also be conducted in future by evaluating the genetic characteristic of *Eucheuma* and *Kappaphycus* in Malaysian waters. Thus, this study is important to fill the gap of scientific knowledge in the seaweed industry for biotechnological applications in Malaysia.

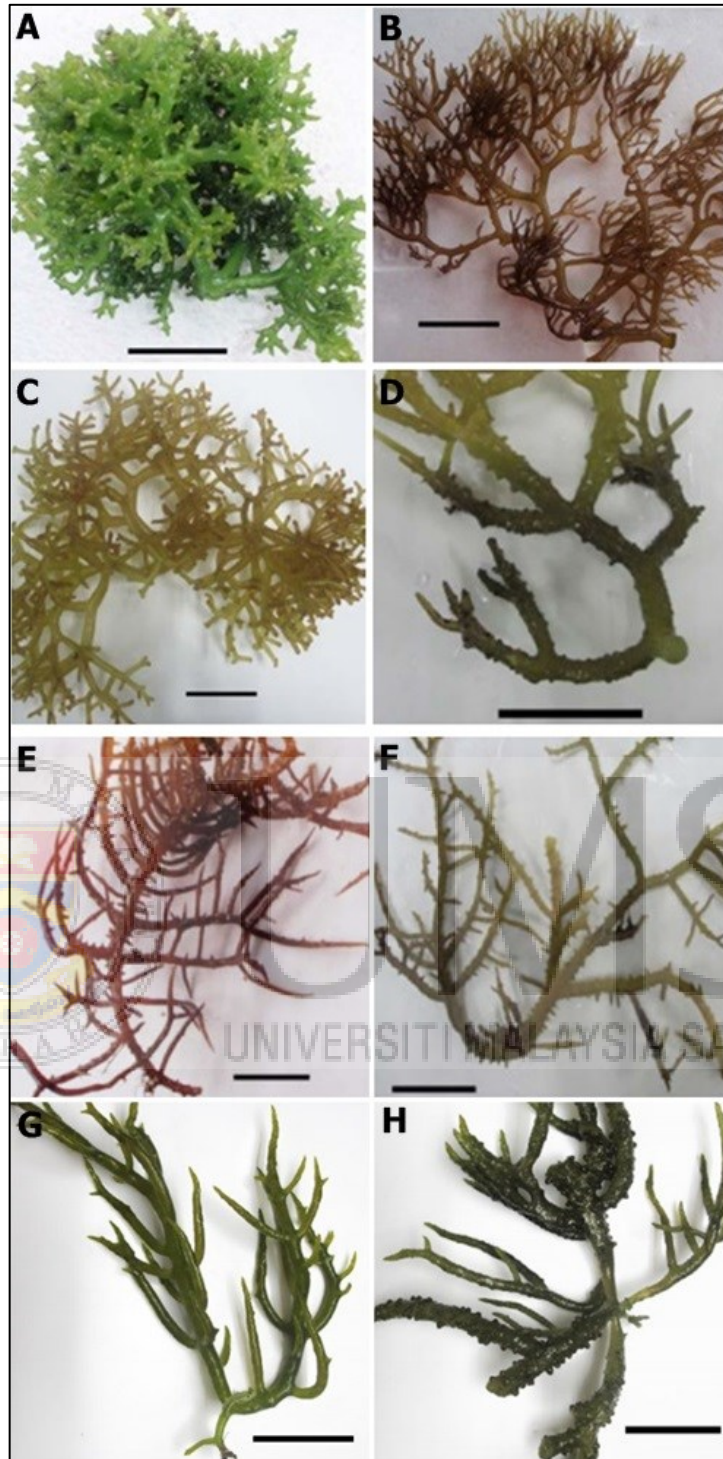
## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 *Kappaphycus* and *Eucheuma* Seaweeds

*Kappaphycus* and *Eucheuma* belong to the family of Rhodophyta and classified under the eukaryotic lineage. It is a renowned fact that Rhodophyta is one of the most diverse groups of seaweeds encompassing more than 6,000 species ranging from unicellular to multicellular organisms (Guiry and Guiry, 2013). Widely distributed across marine environment, Rhodophyta appears red in color. However, some colors are also commonly observed, such as, shades of green, brown and purple. For example (Figure 2.1), varieties with yellowish brown, pale brown, yellowish green and reddish brown were frequently observed among *Kappaphycus* and *Eucheuma* varieties found in Malaysia, causing confusion in morphological-based classification (Tan *et al.*, 2013). *Eucheuma* and *Kappaphycus* seaweeds are photosynthetic organisms, this red coloration in red algae is due to the present of phycobiliproteins on the surfaces of thylakoid which acts as light-harvesting complex and may include phycoerythrin (red), phycocyanin (blue) and allophycocyanin (blue–greenish) and were reported to respond to environmental conditions (Morisset and Kramer, 1984; Grossman *et al.*, 1993).

A study on the photosynthetic growth of red and green morphotypes of *K. alvarezii* in the Philippines showed that the over-expression of phycocyanin and allophycocyanin masked the red pigmentation by phycoerythrin (Aguirre-von-Wobeser *et al.*, 2001). Thus, suggesting an explanation for the different colorations in *Eucheuma* and *Kappaphycus* seaweeds. On the other hand, there are at least three types of photosynthetic pigments known for red algae, which are, chlorophylls, phycobiliproteins and carotenoids (Goodwin, 1974; Rowan, 1989; Dring, 1990; Grossman *et al.*, 1993; Naguit and Tissera, 2009). There are different types of carotenoids (Carotene, cryptoxanthin, zeaxanthin and lutein) found in *Kappaphycus* and *Eucheuma* seaweeds (Anderson *et al.*, 2006). Carotenoids are characterized by extensively conjugated polyene chain which makes carotenoids an



**Figure 2.1:** *Kappaphycus* and *Eucheuma* varieties in Malaysia. *K. alvarezii* (A and B), *K. striatus* (C and D), *E. denticulatum* (E and F) and *K. malesianus* (G and H). (scale bar: 5 cm)

**Source :** Tan *et al*, (2013, 2014)