DEVELOPMENT AND CHARACTERIZATION OF ANONYMOUS NUCLEAR MARKERS TO DETERMINE GENETIC VARIATION IN TWO GROUPERS SPECIES AND THEIR HYBRIDS

AHMAD ZAIDI BIN TANI

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITI MALAYSIA SABAH
2015

DEVELOPMENT AND CHARACTERIZATION OF ANONYMOUS NUCLEAR MARKERS TO DETERMINE GENETIC VARIATION IN TWO GROUPERS SPECIES AND THEIR HYBRIDS

AHMAD ZAIDI BIN TANI

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

BIOTECHNOLOGY RESEARCH INSTITUTE
UNIVERSITI MALAYSIA SABAH
2015

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS

JUDUL: DEVELOPMENT AND CHARACTERIZATION OF ANONYMOUS NUCLEAR MARKERS TO DETERMINE GENETIC VARIATION IN TWO GROUPERS SPECIES AND THEIR HYBRIDS

IJAZAH: SARJANA SAINS dalam bidang BIOTEKNOLOGI

Saya AHMAD ZAIDI TANI, Sesi Pengajian 2011-2015, mengaku membenarkan tesis Sarjana ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis ini adalah hak milik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/)

☐ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)

☐ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

☑ TIDAK TERHAD

Disahkan oleh,

NURULAIN BINTI ISMAIL
LIBRARIAN
UNIVERSITI MALAYSIA SABAH

Zaidi
(Tandatangan Penulis)

(Tandatangan Pustakawan)

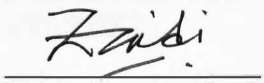
Tarikh: 31 Ogos 2015

K. Kenneth
(DR. KENNETH FRANCIS RODRIGUES)
Penyelia

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, excepts, equations, summaries and references, which have been duly acknowledged.

27 August 2015



Ahmad Zaidi Bin Tani
PB20118049

CERTIFICATION

NAME : **AHMAD ZAIDI BIN TANI**

MATRIC NO : **PB20118049**

TITLE : **DEVELOPMENT AND CHARACTERIZATION OF ANONYMOUS NUCLEAR MARKERS TO DETERMINE GENETIC VARIATION IN TWO GROUPERS SPECIES AND THEIR HYBRIDS**

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

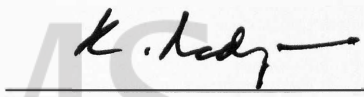
VIVA DATE : **23 APRIL 2015**

CERTIFIED BY

1. SUPERVISOR

Dr. Kenneth Francis Rodrigues

Signature



ACKNOWLEDGEMENT

I would like to express my gratitude to Allah S.A.W for granting me the opportunity for doing my Master in Science (Biotechnology). I am deeply indebted to my supervisor Dr. Kenneth Francis Rodrigues for entrusting this research to me and supervising me throughout the research. Thank you for stimulating suggestion and encouragement throughout the time of my research and thesis writing. To my former colleagues (Noor Hydayaty, Ong Wen Dee, Mohd Faiz and Senty), lab assistants (Vidarita, Mony and Azian) lecturers (Assoc. Prof. Dr. Vijay Kumar) and management team (Mdm Noor Asha); thank you for all of their help, support and interest. To my lovely family; my parents En Tani Bin Duet and Pn Betty Chin, my sisters and brother, Ibrahim, Zaiton and Normala, thank you for giving me your full support, love and understanding all this while till the completion of my study and this thesis.

Ahmad Zaidi Tani

23 April 2015

ABSTRACT

Grouper (Genus: *Epinephelus*) represent one of the high value fisheries resource in Malaysia and several inter-specific hybrids have been developed by the fish breeders. DNA based markers are ideal elucidation of the fundamental associated with inheritance and linkage in determinate hybrids. First generation markers (Microsatellite, RAPD, ISSR etc.) have limitations when applied to characterize intraspecific hybrids. This study set forth to overcome this limitation by developing the second generation of non-coding non informative anonymous nuclear markers (ncni-ANM) for subsequent genetic characterization of F1 grouper hybrids. Small-insert genomic DNA library of *Epinephelus lanceolatus* (EL) and *Epinephelus fuscoguttatus* (EF) were constructed via a shotgun approach following which they were annotated, subjected to a similarity searches using the blastn algorithms and deposited at the GenBank. Annotated genomic sequences of *E. lanceolatus* and *E. fuscoguttatus* were than used for designation of anonymous nuclear markers (ANM) and further categorized into being non-coding non-informative ANM (ncni-ANM), non-coding informative ANM (nci-ANM) or coding informative ANM (ci-ANM). Intraspecific DNA fingerprinting analysis was carried out using five *E. lanceolatus* and 12 *E. fuscoguttatus* broodstock, scored as binary matrix and calculated for genetic distance (GD), genetic similarity (GS) and cophenetic correlation index (CCC) via Jaccard's and Dice's coefficient. An unweighted pair-group method with arithmetic mean (UPGMA) phenogram was constructed based on the GD and CCC values to verify the genetic relationship among the grouper broodstock operational taxonomic units (OTUs). Subsequently, thirty two half-sib F1 hybrids obtained by crossing of female *E. fuscoguttatus* and male *E. lanceolatus* (EFEL) brooders were characterized using 34 male *E. lanceolatus* and 32 *E. fuscoguttatus* ncni-ANM in order to evaluate Mendelian segregation pattern based on genetic inheritance and segregation distortion analyses. A library of 127 genomic sequences was developed of which seven *E. lanceolatus* and 16 *E. fuscoguttatus* sequences could be mapped back to the *Dicentrarchus labrax* genome. Locus specific markers were developed using 60 *E. lanceolatus* and 67 *E. fuscoguttatus* genomic sequences. A total of 46 *E. lanceolatus* ANM were designed containing 34 ncni-ANM, 4 nci-ANM and 8 ci-ANM. While as in total of 42 *E. fuscoguttatus* ANM were designed containing 32 ncni-ANM, 3 nci-ANM and 7 ci-ANM. Polymerase chain reaction (PCR) amplification patterns of ncni-ANM in both *E. lanceolatus* and *E. fuscoguttatus* broodstocks revealed 11 *E. lanceolatus* and 10 *E. fuscoguttatus* ncni-ANM were reproducible. Among these 13 *E. lanceolatus* and *E. fuscoguttatus* ncni-ANM loci ranged in size from 153 to 177 bp and 157 to 624 bp respectively revealing variation across all OTUs through visualization of agarose gel electrophoresis patterns and binary matrix scoring. Clustering comparison based on CCC goodness-of-fit showed both species samples were best presented using Jaccard's coefficient. The phenogram constructed for the five *E. lanceolatus* OTUs showed two clusters with polyphyletic groups. The first cluster (C1) consisted of OTUs EL6924, EL5477 and EL140E with monophyletic group while as the second cluster (C2) consisted of OTUs EL4E4D and EL174A with

paraphyletic grouping. Four clusters of polyphyletic groups were revealed when a phenogram was constructed based on the 12 *E. fuscoguttatus* OTUs. The first and second clusters (C1 and C2) consisted of OTUs EF02X, EF002, EF009 and EF03X; and OTUs EF001, EF006, EF005 and EF006 respectively which showed monophyletic grouping. The third and fourth clusters (C3 and C4), consisted of OTUs EF007 and EF0010; and OTU EF003 which showed paraphyletic grouping. Genetic distances (≤0.500) of all five *E. lanceolatus* OTUs (EL6924, EL140E, EL434D and EL174A) and four *E. fuscoguttatus* OTUs (EF02X, EF003, EF005 and EF008) ranged from from 0.182-0.500 and 0.364-0.500 respectively; implying that high dissimilarities among OTUs were well characterized based on UPGMA. Applying the same ncni-ANM, 34 *E. lanceolatus* and 32 *E. fuscoguttatus* markers showed acceptable monohybrid segregation patterns in which 10 and 24 *E. lanceolatus*; 11 and 22 *E. fuscoguttatus* were grouped as 1:1 and 3:1 segregation pattern ratios respectively, thus proving the accordance of ncni-ANM to Mendelian segregation patterns. Analysis of segregation distortion analysis (x^2) proved that among the reported segregation loci, only six *E. lanceolatus* and 12 *E. fuscoguttatus* ncni-ANM were in accordance ($x^2 <$ 3.841) with Mendelian law while the remaining were classified as distorted loci. This is the first genomic library and the first report that affirmed the use of ncni-ANM markers in DNA fingerprinting and segregation inheritance study. Thus, the preliminary isolation of ncni-ANM of groupers in the study of groupers genotyping and inheritance is revealed and be made feasible and benefit the Malaysian aquaculture genetics research and industry.

ABSTRAK

PEMBANGUNAN DAN PENCIRIAN PENANDA NUKLEAR UNTUK MENENTUKAN VARIASI GENETIK DALAM DUA SPESIES KERAPU DAN HIBRID MEREKA

Kerapu (Genus: Epinephelus*) merupakan salah satu daripada sumber perikanan bernilai tinggi di Malaysia, dimana terdapat beberapa kacukan kerapu khas yang telah dihasilkan dan diusahakan oleh penternak ikan. Penanda DNA merupakan penjelasan yang ideal bagi asas berkaitan dengan perwarisan dan hubungan di dalam hibrid; namun penanda DNA generasi pertama (Mikrosatelit, RAPD, ISSR dll) mempunyai had apabila digunakan untuk mencirikan kacukan intraspesies. Kajian ini ditetapkan untuk mengatasi had tersebut dengan membangunkan generasi kedua penanda DNA non-coding non informative anonymous nuclear marker (ncni-ANM) untuk digunakan di dalam pencirian genetik kerapu hibrid F1. Perpustakaan DNA genom skala kecil* Epinephelus lanceolatus (EL) *dan* Epinephelus fuscoguttatus (EF) *telah dibina melalui pendekatan berikut shotgun carian persamaan jujukan menggunakan algoritma blastn dan seterusnya diserahkan kepada GenBank. Beranotasikan jujukan genom* E. lanceolatus *dan* E. fuscoguttatus*, penanda ANM telah berjaya dikategorikan kepada non-coding non-informative ANM (ncni-ANM), non-coding informative ANM (nci-ANM) atau coding informative ANM (ci-ANM). Analisis intraspesis cap jari DNA telah dijalankan menggunakan lima induk* E. lanceolatus *dan 12 induk* E. fuscoguttatus*, dijaringkan sebagai matriks binari dan jarak genetik (GD), persamaan genetik (GS) dan indeks korelasi cophenetic (CCC) dikira menggunakan pekali Jaccard dan Dice. Kaedah pasangan-kumpulan unweighted dengan aritmetik min (UPGMA) phenogram telah dibina berdasarkan nilai GD dan CCC bagi mengesahkan hubungan genetik di antara induk ikan kerapu. Sebanyak, 32 hibrid F1 separuh keluarga yang dihasilkan melalui pembiakan bersilang diantara* E. fuscoguttatus *betina dan* E. lanceolatus *jantan (EFEL) telah dicirikan menggunakan 34 penanda jantan* E. lanceolatus *dan 32 penanda betina* E. fuscoguttatus *ncni-ANM bagi menganalisa corak pengasingan berasaskan prinsip Mendel dan analisis herotan pengasingan. Perpustakaan 127 jujukan genom telah dibangunkan di mana tujuh jujukan* E. lanceolatus *dan 16 jujukan E. fuscoguttatus berjaya dipetakan semula kepada genom* Dicentrarchus labrax*. Penanda lokus khas telah dihasilkan berdasarkan 60 jujukan* E. lanceolatus *dan 67 jujukan* E. fuscoguttatus*. Sebanyak 46 penanda ANM* E.lanceolatus *telah direka mengandungi 34 ncni-ANM, 4 nci-ANM dan 8 ci-ANM. Manakala 42 penanda ANM* E.fuscoguttatus *telah Berjaya direka mengandungi 32 ncni-ANM, 3 IKN-ANM dan 7 ci-ANM. Corak amplifikasi ncni-ANM di dalam kedua-dua induk ikan* E. lanceolatus *dan* E. fuscoguttatus *mendedahkan 11 penanda* E. lanceolatus *dan 10 penanda* E. fuscoguttatus *ncni-ANM menunjukkan penghasilan semula. Antaranya 13 lokus* E. lanceolatus *dan* E. fuscoguttatus *ncni-ANM di antara saiz 153-177 bp dan 157-624 bp menunjukkan variasi melalui gel agarose elektroforesis dan jaringan matriks binari. Pengelompokan perbandingan berdasarkan nilai CCC, menunjukkan kedua-*

dua spesis terbaik dikelaskan dengan menggunakan pekali Jaccard. Phenogram untuk lima induk E. lanceolatus *membentuk dua kelompok dengan kumpulan polyfili. Kelompok pertama (C1) terdiri daripada EL6924, EL5477 dan EL140E iaitu kumpulan monofili manakala kelompok kedua (C2) terdiri daripada EL4E4D dan EL174A iaitu kumpulan parafili. Empat kelompok kumpulan polyfili telah dibina berdasarkan 12 induk* E. fuscoguttatus*. Kelompok pertama dan kedua (C1 dan C2) terdiri daripada EF02X, EF002, EF009 dan EF03X; dan EF001, EF006, EF005 dan EF006 dimana masing-masing yang menunjukkan kumpulan monofili. Manakala kelompok ketiga dan keempat (C3 dan C4), yang terdiri daripada EF007 dan EF0010; dan EF003 menunjukkan kumpulan parafili. Jarak genetik (≤0.500) di antara lima* E. lanceolatus *(EL6924, EL140E, EL434D dan EL174A) dan empat* E. fuscoguttatus *(EF02X, EF003, EF005 dan EF008) adalah di antara 0.182-0.500 dan 0.364-0.500; dimana ketidaksamaan yang tinggi telah berjaya dicirikan berdasarkan UPGMA. Selain itu, 34 dan 32 penanda* E. lanceolatus *dan* E. fuscoguttatus *juga menunjukkan corak pengasingan monomorfik di mana 10 dan 24 penanda* E. lanceolatus*; dan 11 dan 22 penanda* E. fuscoguttatus *telah dikelaskan sebagai nisbah perwarisan 1:1 dan 3:1, sekali gus membuktikan kecekapan ncni-ANM dalam menganalisa corak perwarisan Mendel. Analisis herotan pengasingan (x^2) membuktikan bahawa di antara pengasingan lokus yang dilaporkan, hanya enam lokus* E. lanceolatus *dan 12 lokus* E. fuscoguttatus *ncni-ANM diterima (x^2 <3,841) berdasarkan prinsip Mendel manakala bakinya dikelaskan sebagai lokus tidak berdasarkan prinsip Mendel; sekaligus menunjukkan pengasingan sekata. Ini merupakan perpustakaan genom dan laporan pertama yang mengesahkan fungsi penanda generasi kedua ncni-ANM di dalam analisis cap jari DNA dan kajian pengasingan warisan. Oleh itu, kepentingan ncni-ANM dalam genotyping dan kajian warisan kerapu telah dinyatakan dan dilaksanakan dimana memberikan manfaat kepada penyelidikan dan industri genetik akuakultur di Malaysia.*

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

µL	Microliter
ng	Nanogram
bp	Base Pair
cm	Centimetre
dNTP	Deoxyribonucleotide triphosphate
g	Graviti
min	Minute
mL	Mililitre
mM	Milimolar
$MgCl_2$	Magnesium chloride
U	Unit
ci-ANM	Coding infomrative-anonymous nuclear markers
nci-ANM	Non-coding informative-anonymous nuclear markers
ncni-ANM	Non-coding non-informative-anonymous nuclear markers
rpm	Rotation per minute
UV	Ultra violet

LIST OF SYMBOL

°C	Degree Celsius
%	Percentage

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

1.1 Research Background

Groupers are a carnivorous marine fish that inhabit the tropical and subtropical ocean worldwide. Groupers are farmed globally and represent high-value fisheries resources. Malaysia is one of the seven countries in Southeast Asia that produce high quantities of grouper breed. However, because of the high demand, small numbers of hatcheries produce and problematic issues in the production of groupers fry, the numbers of groupers production are tremendously decreasing.

The Hatchery of Borneo Marine Research Institute (BMRI), Universiti Malaysia Sabah (UMS) have conducted and successfully develop and produce approximately 10 different breeds of F1 hybrid groupers from a cross between various types of groupers from the genus of *Epinephelus.* The successful production of F1 hybrid crosses solve the aquaculture production. However, does not improve any of genetics study respectively.

During the past decade, genetic molecular markers (First generation markers) have been applied for various types of application in genetic study. The most common genetic markers used in the groupers study are microsatellite markers which used in the study of population genetics as well as mitochondrial DNA and various nuclear DNA markers which used in the study of strain and species identification. However for the past few years researcher have found that the competence of the existing genetic molecular markers is decreasing (Thomson *et al.*, 2010).