POPULATION GENETIC STUDIES OF MARBLE GOBY *Oxyeleotris marmoratus* (BLEEKER, 1852) IN MALAYSIA USING MICROSATELLITE AND MITOCHONDRIAL DNA MARKERS

RUZAINAH BT ALI @ JAAFAR

UNIVERSITI SAINS MALAYSIA

2008

POPULATION GENETIC STUDIES OF MARBLE GOBY *Oxyeleotris marmoratus* (BLEEKER, 1852) IN MALAYSIA USING MICROSATELLITE AND MITOCHONDRIAL DNA MARKERS

by

RUZAINAH BT ALI @ JAAFAR

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philoshopy

JUNE 2008

ACKNOWLEDGEMENTS

My profound gratitude to Allah s.w.t. who has given me the strength and health to finish this thesis and who has given me these wonderful people to work with

- my supervisor, Associate Professor Dr Siti Azizah Bt Mohd Nor:
  - whose immense knowledge is forever a wonder,
  - whose immeasurable patience in teaching, guiding, advising, correcting and recorrecting a green me is forever an awe,
  - whose endless humanity and hospitality when dealing with this particular student is forever admired

- my co-supervisor, Associate Professor Dr Sofiman Bin Othman and my field supervisor, Associate Professor Dr Patimah Bt Ismail who had dedicatedly given me their invaluable time, advice, guidance and encouragement throughout the completion of this thesis

- my parents especially my late mother Allahyarhamah Zaifah Mohd who had never doubted my ability even though many a time it is nonexistence, my father (Tn Haji Ali @ Jaafar Bin Khatib Mamat) and umi (Hajah Khadijah Muhammad) who had never ceased in convincing me that the sky is the limit

- and especially so to my husband, Mulyadi Harun who had never failed me especially with his wonderful way of going out of his way to be wonderful

- and to my children (Lutfi Danial, Lia Amirah and Lutfi Naim) whose time I had stolen and yet readily accepted the lies that I will make up the lost time

- not forgetting my present employer UniKL, UniKL MICET colleagues (especially Bioprocess and Biosystem Engineering Technology Department) and all my students whose understanding of the plight of this mere lecturer cum student stretches beyond my imagination

- and to my sisters, brothers and all those who had been very sincerely helpful to me, consciously and unconsciously

- Last but not least, special thanks to all my helpful friends at the School of Biological Sciences (Abang Amir, Emi, Mila, Fatimah, Ain, Nazia, Sue Yin, Sue Yee, Lim, Jam and Ghouse) and at Faculty of Medicine and Health Sciences, UPM (Vasu, Ita, Kak Wan, Kak Zila, Apai, Yaya, Zain, Farah, Hana and En. Zainan).

I am forever in debt. Thank you.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ………………………………………........... ii

TABLE OF CONTENTS ………………………………………………... iii

LIST OF TABLES ……………………………………………………… x

LIST OF FIGURES …………………………………………………... xvi

LIST OF PLATES …………………………………………………….. xix

LIST OF ABBREVIATION …………………………………………….. xx

LIST OF APPENDICES ………………………………………………… xxi

ABSTRAK ……………………………………………………………... xxii

ABSTRACT …………………………………………………………… xxiv

CHAPTER 1 - INTRODUCTION

1.1 Introduction …………………………………………………….. 1

1.2 Objectives …………………………………………………… 5

CHAPTER 2 - LITERATURE REVIEW

2.1 *Oxyeleotris marmoratus* …………………………………………… 6

2.1.1 Taxonomic Classification ………………………………… 6

2.1.2 Morphology and Habitat …………………………………… 7

2.1.3 Potential Candidate for Aquaculture Programme ………….. 8

2.2 Molecular Population Genetics ……………………………………. 9

2.3 Molecular Genetic Markers in Conservation ……………………… 11

2.4 Mitochondrial DNA ……………………………………………….. 12

2.4.1 Application of Mitochondrial DNA in Fish Genetics ……... 15

2.5 Microsatellite Markers …………………………………………….. 17

2.5.1 Microsatellite Characteristics ………………………………. 18

2.5.2 Evolution of Microsatellite ………………………………... 22

2.5.3 Application of Microsatellite in Fish Genetics ……………. 22

2.6 Mitochondrial DNA versus Microsatellite Analysis ………………. 24

CHAPTER 3 - MATERIALS AND METHODS

3.1 Sample Collection ………………………………………………. 27

3.2 Isolation of Genomic DNA - Mitochondrial and Microsatellite DNA 30

3.3 Mitochondrial DNA Analysis …………………………………... 30

3.3.1 PCR Amplification …………………………………….. 30

3.3.2 Purification of PCR Products for Sequencing ……………. 31

3.3.3 Automated DNA Sequencing of Purified Products ……… 31

3.3.4 Data Analysis …………………………………………….. 32

3.3.4.1 Phylogenetic Analysis – Parsimony and Distance Based Method 31

3.3.5 Population Study ……………………………………….. 33

3.3.5.1 Levels of population subdivision ……………… 33

3.4 Microsatellite Analysis …………………………………………. 35

3.4.1 Microsatellite Development by Randomly Amplified Microsatellite (RAM) protocol ……. 35

3.4.1.1 Detection of Microsatellite Repeat Motifs by RAM primers ….. 38

3.4.1.2 Purification of PCR Product of RAM Primers … 39

3.4.1.3 Cloning of the PCR products into the PCR 2.1 TOPO Vector ……... 39

3.4.1.4 Transformation of Competent Top 10 cells ……. 39

3.4.1.5 Plasmid Extraction ……………………………... 40

3.4.1.6 Automated Fluorescent Sequencing ……………. 40

3.4.1.7 Primer Design ………………………………….. 40

3.5 Population Study ……………………………………………... 41

3.5.1 PCR Amplification for microsatellite markers …………... 42

3.5.2 Detection of microsatellite markers ……………………… 43

3.5.3 Microsatellite Data Analysis ……………………………... 43

CHAPTER 4- RESULTS AND DISCUSSION – Mitochondrial DNA

4.1 DNA extraction ……………………………………………………. 45

4.2 PCR amplification of partial region of ND 5 and Cyt *b* gene ……... 46

4.3 Intrapopulation Phylogenetic Analysis ……………………………. 47

4.3.1 Cytochrome *b* Analysis …………………………………... 47

4.3.1(a) Sabah Population ………………………………. 48

4.3.1(b) Sarawak Population ……………………………. 49

4.3.1(c) Tasik Bera Population …………………………. 50

4.3.1(d) Tasik Chini Population ………………………… 51

4.3.1(e) Bachok Population ……………………………... 52

4.3.1(f) Pulau Keladi Population ……………………….. 52

4.3.1(g) Tanjung Demong Population …………………... 54

4.3.1(h) Tasik Kenyir Population ……………………….. 55

4.3.1(i) Tasik Chenderoh Population …………………... 56

4.3.1(j) Sungai Simpang Population …………………… 58

4.3.1(k) Sungai Melaka Population ……………………... 58

4.3.1(l) Pondok Tanjung Population …………………… 61

4.3.1(m) Tasik Banding population ……………………… 62

4.3.2 NADH-Dehidrogenase Subunit 5 (ND 5) Analysis ……….. 64

4.3.2(a) Sabah Population ………………………………. 64

4.3.2(b) Sarawak Population ……………………………. 64

4.3.2(c) Tasik Bera Population …………………………. 66

4.3.2(d) Tasik Chini Population ………………………… 67

4.3.2(e) Bachok Population …………………………….. 68

4.3.2(f) Sungai Pulau Keladi Population ………………. 69

4.3.2(g) Tanjung Demong Population …………………... 70

4.3.2(h) Tasik Kenyir Population ……………………….. 72

4.3.2(i) Tasik Chenderoh Population …………………... 72

4.3.2(j) Sungai Simpang Population …………………… 74

4.3.2(k) Sungai Melaka Population …………………….. 75

4.3.2(l) Pondok Tanjung Population …………………… 76

4.3.2(m) Tasik Banding population ……………………… 78

4.4 Interpopulation Analysis of total combined Haplotype …………… 80

4.4.1 Cytochrome *b* Analysis ………………………………….. 80

4.4.1.1 Neighbor Joining …………………………………. 80

4.4.1.2 Maximum Parsimony (MP) Analysis ……………. 81

4.4.2 NADH – Dehidrogenase 5 Analysis …………………….. 84

4.4.2.1 Neighbor Joining …………………………………. 84

4.4.2.2 Maximum Parsimony (MP) Analysis ……………. 84

4.5 Interpopulation Analysis – Mitochondrial DNA …………………. 86

4.5.1 Cytochrome *b* Analysis …………………………………... 86

4.5.1.1 Genetic Variability ……………………………… 86

4.5.1.2 Phylogenetic Analysis of Unique Haplotypes ….. 91

4.5.1.3 Analysis of molecular variance of Cyt *b* gene …. 96

4.5.1.4 Minimum Spanning Network …………………… 99

4.5.2 NADH – Dehydrogenase 5 (ND 5) Analysis …………….. 101

4.5.2.1 Genetic Variability ……………………………… 101

4.5.2.2 Phylogenetic Analysis of Unique Haplotypes ….. 108

4.5.2.3 Analysis of molecular variance of ND 5 gene ….. 113

4.5.2.4 Minimum Spanning Network …………………… 115

4.6 Discussion ………………………………………………………. 117

4.6.1 Genetic diversity ………………………………………... 117

4.6.2 Population Structure …………………………………….. 121

CHAPTER 5 - RESULTS AND DISCUSSION – Microsatellite

5.1 Screening for Microsatellite Regions ……………………………… 129

5.2 Cloning of the PCR products ……………………………………… 131

5.3 Plasmid Extraction ………………………………………………… 132

5.4 Automated Fluorescense Sequencing …………………………... 133

5.5 Microsatellite Characteristics ……………………………………… 133

5.6 Primer Designing ………………………………………………….. 134

5.7 Microsatellite banding Pattern Profiles of Selected Primers ……… 136

5.8 Microsatellite Variation …………………………………………... 140

5.9 Hierarchical AMOVA Analysis …………………………………… 147

5.10 Linkage Disequilibrium (LD) ……………………………………... 149

5.11 Cluster Analysis …………………………………………………… 150

5.12 Discussion ………………………………………………………… 152

5.12.1 Development of microsatellite markers ………………… 152

5.12.2 Genetic diversity and population structure …………….. 157

CHAPTER 6 - GENERAL DISCUSSION …………………………….. 167

CHAPTER 7 - SUMMARY AND CONCLUSION …………………... 177

REFERENCES ……………………………………………………….... 180

APENDIX A

APENDIX B

APENDIX C

APENDIX D

APENDIX E

APENDIX F

LIST OF PUBLICATIONS & SEMINARS

LIST OF TABLES

		Page
2.1	Examples of three categories of microsatellite sequences.	**20**
3.1	List of samples from different locations.	**28**
3.2	Description of RAM primers.	**38**
3.3	List of microsatellite primer designed.	**41**
3.4	Preparation for 6% non denaturing polyacrylamide gel.	**43**
4.1	Number of individuals successfully sequenced for the two mtDNA genes.	**47**
4.2	Distance matrix of partial Cyt *b* gene in Sabah population.	**48**
4.3	Haplotype frequency of partial Cyt *b* gene in Sabah.	**48**
4.4	Distance matrix of partial Cyt *b* gene in Sarawak population.	**49**
4.5	Haplotype frequency of partial Cyt *b* gene in Sarawak.	**49**
4.6	Distance matrix of partial Cyt *b* gene in Tasik Bera population.	**50**
4.7	Haplotype frequency of partial Cyt *b* gene in Tasik Bera.	**50**
4.8	Distance matrix of partial Cyt *b* gene in Tasik Chini population.	**51**
4.9	Haplotype frequency of partial Cyt *b* gene in Tasik Chini.	**51**
4.10	Distance matrix of partial Cyt *b* gene in Bachok population.	**53**
4.11	Haplotype frequency of partial Cyt *b* gene in Bachok.	**53**
4.12	Distance matrix of partial Cyt *b* gene in Pulau Keladi population.	**54**
4.13	Haplotype frequency of partial Cyt *b* gene in Pulau Keladi.	**54**

Page

4.14	Distance matrix of partial Cyt *b* gene in Tanjung Demong population.	**55**
4.15	Haplotype frequency of partial Cyt *b* gene in Tanjung Demong.	**55**
4.16	Distance matrix of partial Cyt *b* gene in Tasik Kenyir population.	**56**
4.17	Haplotype frequency of partial Cyt *b* gene in Tasik Kenyir.	**56**
4.18	Distance matrix of partial Cyt *b* gene in Tasik Chenderoh population.	**57**
4.19	Haplotype frequency of partial Cyt *b* gene in Tasik Chenderoh.	**57**
4.20	Distance matrix of partial Cyt *b* gene in Sungai Simpang population.	**59**
4.21	Haplotype frequency of partial Cyt *b* gene in Sungai Simpang.	**60**
4.22	Distance matrix of partial Cyt *b* gene in Sungai Melaka population.	**60**
4.23	Haplotype frequency of partial Cyt *b* gene in Sungai Melaka.	**61**
4.24	Distance matrix of partial Cyt *b* gene in Sungai Pondok Tanjung population.	**61**
4.25	Haplotype frequency of partial Cyt *b* gene in the Sungai Pondok Tanjung.	**61**
4.26	Distance matrix of partial Cyt *b* gene in Tasik Banding population.	**63**
4.27	Haplotype frequency of partial Cyt *b* gene in Tasik Banding.	**63**
4.28	Distance matrix of partial ND 5 gene in Sabah population.	**65**
4.29	Haplotype frequency of partial ND 5 gene in Sabah.	**65**
4.30	Distance matrix of partial ND 5 gene in Sarawak population.	**66**

Page

4.31	Haplotype frequency of partial ND 5 gene in Sarawak.	**66**
4.32	Distance matrix of partial ND 5 gene in Tasik Bera population.	**67**
4.33	Haplotype frequency of partial ND 5 gene in Tasik Bera.	**67**
4.34	Distance matrix of partial ND 5 gene in Tasik Chini population.	**68**
4.35	Haplotype frequency of partial ND 5 gene in Tasik Chini.	**68**
4.36	Distance matrix of partial ND 5 gene in Bachok population	**69**
4.37	Haplotype frequency of partial ND 5 gene in Bachok.	**69**
4.38	Distance matrix of partial ND 5 gene in Sungai Pulau Keladi population.	**70**
4.39	Haplotype frequency of partial ND 5 gene in Sungai Pulau Keladi.	**70**
4.40	Distance matrix of partial ND 5 gene in Tanjung Demong population.	**71**
4.41	Haplotype frequency of partial ND 5 gene in Tanjung Demong.	**71**
4.42	Distance matrix of partial ND 5 gene in Tasik Kenyir population.	**72**
4.43	Haplotype frequency of partial ND 5 gene in Tasik Kenyir.	**72**
4.44	Distance matrix of partial ND 5 gene in Tasik Chenderoh population.	**73**
4.45	Haplotype frequency of partial ND 5 gene in Tasik Chenderoh population.	**73**
4.46	Distance matrix of partial ND 5 gene in Sungai Simpang population.	**74**
4.47	Haplotype frequency of partial ND 5 gene in Sungai Simpang.	**75**

Page

4.48	Distance matrix of ND 5 gene in Sungai Melaka population.	**76**
4.49	Haplotype frequency of partial ND 5 gene in Sungai Melaka.	**76**
4.50	Distance matrix of partial ND 5 gene in Sungai Pondok Tanjung population.	**77**
4.51	Haplotype frequency of partial ND 5 gene in Sungai Pondok Tanjung.	**77**
4.52	Distance matrix of partial ND 5 gene in Tasik Banding population.	**79**
4.53	Haplotype frequency of ND 5 gene in Tasik Banding.	**79**
4.54	Polymorphic Sites for 16 unique mtDNA Cyt *b* haplotypes	**88**
4.55	Distribution of number of individuals per population per haplotype of Cyt *b* gene	**89**
4.56	Nucleotide diversity, number of haplotypes, haplotype diversity, number of polymorphic sites and neutrality test for partial Cyt *b* gene among populations of *O. marmoratus*.	**90**
4.57	Genetic distance among 16 unique haplotypes of partial Cyt *b* gene.	**92**
4.58	Number of character status for MP analysis of partial Cyt *b* gene.	**94**
4.59	Phylogenetic tree description for MP analysis.	**94**
4.60(a)	AMOVA results for hierarchical genetic subdivision for percentage of variation of Cyt *b* gene.	**96**
4.60(b)	AMOVA results for hierarchical genetic subdivision for F statistic of Cyt *b* gene.	**97**
4.61	Population differentiation based on pairwise F_{ST} values of Cyt *b*.	**98**
4.62	Polymorphic sites for 57 unique mtDNA ND 5 haplotypes	**102**

Page

4.63	Haplotype distribution of ND 5 gene among populations.	**106**
4.64	Nucleotide diversity, number of haplotypes, haplotype diversity, number of polymorphic sites and neutrality test for partial ND 5 gene among populations of *O. marmoratus*.	**107**
4.65	Number of character status for MP analysis of partial ND 5 gene.	**111**
4.66	Phylogenetic tree description for MP analysis.	**111**
4.67(a)	AMOVA results for hierarchical genetic subdivision for percentage of variation of ND 5 gene.	**114**
4.67(b)	AMOVA results for hierarchical genetic subdivision for F statistic value of ND 5 gene.	**114**
4.68	Population differentiation based on pairwise F_{ST} values of ND 5 gene.	**114**
5.2	Category, percentage of perfect and imperfect repeats for each RAM primer.	**133**
5.3	Microsatellite loci designed for *O. marmoratus.*	**135**
5.4	Allelic variability at ten microsatellite loci in 13 populations of *O. marmoratus*	**141**
5.5	Mean allelic variability among ten microsatellite loci in 13 populations of *O. marmoratus*	**146**
5.6	Fixation indices for each locus.	**148**
5.7(a)	AMOVA results for hierarchical genetic subdivision of microsatellite data on percentage of total variation.	**148**
5.7(b)	AMOVA results for hierarchical genetic subdivision of microsatellite data on F statistic and P value	**148**
5.8	Population Pairwise F_{ST} from haplotype frequencies among 13 populations.	**148**
5.9	Summary of significant linkage disequilibrium (exact test using a Markov chain; chain length100000;dememorization 1000)	**149**

LIST OF FIGURES

		Page
2.1	Homozygous and heterozygous microsatellite sequences in a diploid organism (Adapted from Moxon and Wills, 1999).	**19**
2.2(a&b)	Detection of Microsatellite marker.	**21**
3.1	Location map of *O. marmoratus* sampling sites in Malaysia.	**29**
3.2	Hierarchical division of the samples studied.	**34**
3.3	Flow chart showing the basic procedure utilised in Random Amplified Microsatellite (RAM) technique.	**37**
4.1	Total Genomic DNA Extracted Using Pure-Gene Tissue Kit II (BST TechLab).	**45**
4.2	The amplified 350bp single fragment of partial Cyt *b* gene as amplified by the PCR technique.	**46**
4.3	The amplified 1100bp single fragment of partial ND 5 gene as amplified by the PCR technique.	**46**
4.4	Dendrogram showing phylogenetic relationships among 37 haplotypes on the partial Cyt *b* gene generated through NJ method.	**82**
4.5	Dendrogram showing phylogenetic relationship among 37 haplotypes on the partial Cyt *b* gene generated through Maximum Parsimony (MP) method.	**83**
4.6	Dendrogram showing phylogenetic relationship among 74 haplotypes on the partial ND 5 gene generated through NJ method.	**85**
4.7	Dendrogram showing phylogenetic relationship among 16 unique haplotypes of the partial Cyt *b* gene generated through NJ method.	**93**
4.8	Dendrogram showing phylogenetic relationship among 16 unique haplotypes of the partial Cyt *b* gene generated through Maximum Parsimony (MP) method.	**95**
4.9	A minimum spanning network of sixteen mtDNA Cyt *b* haplotypes obtained from 13 *O. marmoratus* populations.	**100**

Page

4.10	Dendrogram showing phylogenetic relationship among 57 haplotypes of the partial ND 5 gene generated through NJ method.	**109**
4.11	Dendrogram showing phylogenetic relationship among 57 haplotypes of the partial ND 5 gene generated through Maximum Parsimony (MP) method.	**112**
4.12	A minimum spanning network of 16 mtDNA ND 5 haplotypes obtained from 13 *O. marmoratus* populations.	**116**
5.1	Discrete banding pattern amplified by each anchored primer: LR1, PCTA, PCTB, PCTC, PCTD and PCTE respectively.	**130**
5.2	Blue-white screening of a colony of LR1 Clones	**131**
5.3	Plasmids of RAM clones. Lane M is a high DNA mass ladder (1 KB) which was used to detect clones containing the inserts.	**132**
5.4a	Optimization of $MgCl_2$ concentration	**136**
5.4b	Workable test for LR15 primer (152 bp) - $(TAG)_3T(A)(TAG)_2(TAT)$	**137**
5.4c	Optimization of annealing temperature amplified by using primer LR15.	**137**
5.4d	Workable test for LR32 primer (119 bp - $(TCT)_3(TC)_4T(TCT)_3$	**138**
5.4e	Workable test for primer LR34 (124 bp- $(AAG)_{4(}G)_3$ $(AAG)_3$.	**138**
5.4f	Workable test for primer PCTC17	**138**
5.4g	Lane 1- 13 (Sample from Tasik Chenderoh, Perak individual 1 to 13) amplified by using primer PCTC17	**139**
5.4h	Lane 1- 13 (Sample from Pulau Keladi; individual 1 to 13) amplified by using primer PCTE53.	**139**

		Page
5.4i	Lane 1- 13 (Sample from Sarawak; individual 1 to 14) amplified by using primer LR32.	**139**
5.5	Neighbour Joining tree showing phylogenetic relationship of 10 microsatellite loci of *O. marmoratus.*	**151**

LIST OF PLATES

Page

3.1 A photograph of marble goby, *O. marmoratus* from Sarawak **28**

LIST OF ABBREVIATION

bp	Base pair
DNA	Deoxyribonucleic acid
dNTP	Dinucleotide triposphate
EDTA	Ethylenediamine tetra-acetic acid
EtBr	Ethidium Bromide
kb	Kilobase
mM	Milimolar
MW	Molecular weight
NaOH	Sodium hidroxide
NaCl	Natrium chloride
OD	Optical Density
rpm	Revolutions per minute
SDS	Sodium dodecyl sulfate
SSCP	Single-Strand Conformation Polymorphism
TBE	Tris borate EDTA
TE	Tris EDTA
µg	Microgram
µl	Microlitre
µM	Micromolar
V	Volts

LIST OF APPENDICES

Appendix A	Chromatogram for the partial ND 5 gene sequence of *O. marmoratus* of Sarawak population.
Appendix B	Chromatogram for the partial Cyt *b* gene sequence of *O. marmoratus* of Sabah population.
Appendix C	List of Microsatellites isolated from RAM Method.
Appendix D	Genetic Distance among 37 haplotypes based on Neighbour Joining for Cyt *b* gene of *O. marmoratus.*
Appendix E	Genetic Distance among 74 haplotypes based on Neighbour Joining for ND 5 gene of *O. marmoratus.*
Appendix F	Genetic distance among 57 unique haplotypes of ND 5 gene of *O. marmoratus*.

KAJIAN GENETIK POPULASI TERHADAP IKAN KETUTU *Oxyeleotris marmoratus* (BLEEKER, 1852) DI MALAYSIA DENGAN MENGGUNAKAN PENANDA MIKROSATELIT DAN MITOKONDRIA DNA

ABSTRAK

Variasi dan hubungan filogenetik ikan ketutu dari 13 populasi di Malaysia dikaji menggunakan 10 lokus mikrosatelit yang baru dibangunkan bergandingan dengan analisis turutan nukleotida gen mitokondria Cyt *b* dan ND 5 Lokus mikrosatelit dipencilkan melalui kaedah Amplifikasi Mikrosatelit Rawak (RAM) dinilai untuk kepelbagaian populasi dan sekurang-kurangnya lapan didapati bersifat polimorfik di dalam kesemua populasi yang diuji. AMOVA berhieraki terhadap kepelbagaian mikrosatelit menunjukkan pembezaan populasi secara signifikan di dalam jumlah kesemua 13 populasi tetapi tiada satu pun populasi antara kumpulan (F_{CT}) dapat dibezakan secara signifikan, i.e., tiada perbezaan genetik diantara kumpulan yang telah dikenalpasti terlebih dahulu mengikut struktur geografi iaitu Borneo, Barat dan Timur Semenanjung Malaysia. Dari 176 individu yang dikaji untuk gen Cyt *b* , 16 haplotip unik telah dikenalpasti manakala 57 haplotip unik bagi gen ND 5 juga telah dikenalpasti. Untuk kesemua kriteria kepelbagaian genetik (contohnya tahap polimorfik dan heterozigositi, kepelbagaian nukleotida dan haplotip) gen ND 5 menunjukkan nilai yang lebih tinggi. Heterogeneiti dalam taburan frekuensi haplotip diterjemahkan kepada nilai F_{ST} yang signifikan di dalam kesemua pembandingan populasi secara berpasangan. Analisis hierarki menunjukkan pembahagian variasi hampir sekata dan signifikan bagi setiap tahap hierarki bagi gen Cyt *b*. Pengstrukturan populasi di dalam sesuatu kumpulan adalah lebih ketara bagi gen ND 5. Kecuali bagi kumpulan Borneo, analisis NJ, MP and MSN berjaya mengkelompokan populasi kepada kumpulan yang telah ditentukankan; Barat dan

Timur Semenanjung Malaysia. Taburan populasi dihipotesiskan dipengaruhi oleh faktor sejarah asal usul dan geologi. Secara keseluruhan populasi yang dikaji masih mempunyai tahap variasi yang sihat. Kesimpulannya, analisis filogenetik daripada penanda molekul (mtDNA dan mikrosatelit) merupakan kaedah yang berkesan untuk mengenalpasti variasi dan hubungan filogenetik bagi spesies ikan ketutu di Malaysia. Kajian seumpamanya boleh dijadikan data asas bagi pemantauan genetik pada masa hadapan dan mungkin akan memberi sumbangan secara signifikan di dalam pembangunan program kacukan untuk pengeluaran dan pengurusan ikan ketutu di Malaysia.

POPULATION GENETIC STUDIES OF MARBLE GOBY *Oxyeleotris marmoratus* (BLEEKER, 1852) IN MALAYSIA USING MICROSATELLITE AND MITOCHONDRIAL DNA MARKERS

ABSTRACT

Marble goby from 13 populations in Malaysia were genetically characterised by examining the variability of 10 newly developed microsatellite loci coupled with the nucleotide sequences of mitochondrial Cyt *b* and ND 5 genes. Microsatellite loci isolated by the enrichment method of Random Amplified Microsatellite (RAM) were evaluated for population variability and at least eight were found to be polymorphic in all populations tested. A hierarchical AMOVA on microsatellite variability showed significant population differentiation within the total sample of 13 populations but none for among group comparisons (F_{CT}) were significant, i.e., no genetic differentiation between the defined geographical regions; Borneo, West and East Peninsular Malaysia. From the 176 individuals examined for the Cyt *b* gene, 16 unique haplotypes were generated and 57 unique ND 5 gene haplotypes were identified. For all criteria of genetic variability (for example, polymorphic and heterozygosity levels, nucleotide and haplotype diversities) the ND 5 gene analysis had higher values. The heterogeneity in haplotype frequency distribution translated into statistically significant F_{ST} value in all pairwise comparison for both genes. Hierarchical analysis showed the partitioning of variation was almost similar among each level for Cyt *b* gene. Structuring between populations within region was more obvious for ND 5 gene. With the exception of the Borneo group, NJ, MP and MSN analyses successfully cluster populations to their defined groups; the East and West Peninsular regions. Population distribution is hypothesized to be influenced by historical origin and geological factors. In general, the investigated populations still

retain healthy levels of variation In conclusion, molecular markers, mtDNA and microsatellites provided an efficient tool to identify variability and phylogenetic relationships of the marble goby in Malaysia. Such studies is intended as a baseline for future genetic monitoring and may contribute significantly in the development of breeding programmes for production and management of the marble goby species in Malaysia.

CHAPTER 1

INTRODUCTION

1.1 Introduction

The demand for fish as a rich and 'cheap' protein source has proportionately increased with the exponential growth of world population. To ensure continuous supply of the product, systematic management programmes have to be put in place. Our community cannot continue depending solely on traditional resources from freshwater and marine landings for fish protein supply. Malaysia has already stepped up efforts to develop the aquaculture sector to diversify the country's supply of protein-rich food. The 9th Malaysia Plan sees even greater emphasis on aquaculture sector development to augment existing wild catches. To maintain the quality of aquaculture product, it is necessary to carefully manage the genetics as well as evaluate the stocking effectiveness. Genetic improvement provides the capability of culturing a better quality animal in less time, with greater survival, and at less cost than animals removed from the wild.

The marble goby *Oxyleotris marmoratus* Bleeker 1852 (Teleostei : Gobiidae) is a highly valued species occurring in many parts of Southeast Asia including Thailand, Peninsular Malaysia, Singapore and Indonesia. In recent years, it has become a very popular gourmet especially among patrons of Chinese restaurants. This makes it a potentially good candidate for a systematic aquaculture proggrame to meet current demands. Therefore, it is important to obtain genetic knowledge of the brood stock to ensure high quality and sustainable aquaculture production.

In the last few decades, genetic variation at the molecular level (i.e. protein and DNA) has been widely used in population genetic studies *vis a vis* the measurement of genetic diversity and genetic relationship among individuals and populations (Skroch and Nienhuis, 1995). In addition, measurement of genetic diversity through molecular markers is relevant for assessment of ecological condition as it allows an estimation of important population parameters such as characterization of the geographic structure or connectivity of populations (Sivasundar *et al.*, 2001; Barroso *et al.*, 2005). Variation at the population level can provide an idea about different genetic classes, the genetic diversity among them and their evolutionary relationship with wild relatives (Paramanda *et al.*, 2005). Measures of genetic diversity are relevant in assessing extinction risks for populations while the genetic variability within population is extremely useful to gather information on individual identity, breeding pattern, degree of relatedness and disturbances of genetic variation among them (Schierwater *et al.*, 1994). Within the last decade, technological advancement has increased support for the use of genetics in determining population diversity (Wolff and Peters-van, 1993; Reilly and Ward, 1999; Sivasundar *et al*, 2001; Barroso *et al.*, 2005; Thai *et al.*, 2005; Li *et al.*, 2007). Therefore, development of molecular markers has become a major fundamental research activity in microorganism, plant and animal studies.

DNA markers are genetic variants detectable at the DNA level. Depending on their locations in the cell; they may be nuclear or mitochondrial. Nuclear DNA is found within the nucleus of the cell and is composed of two sources of DNA, maternal and paternal. Mitochondrial DNA (mtDNA) as the name suggests is contained in the mitochondria of the cell and are generally maternally inherited. The mitochondria are organelles located outside the nucleus in the cytoplasm of the cell.

These organelles are responsible for energy transfer and are basically the "powerhouses" of the cells.

These genetic variants are the product of single base mutations (substitutions, insertions, deletions) or larger alterations such as the insertion of a transposon (Griffiths *et al*., 1999). There are several advantages of DNA markers over visible (morphological) genetic markers. Firstly, it can normally distinguish heterozygotes from homozygotes. Secondly, each DNA marker usually exhibits greater variation within the population; with many individuals being heterozygous for two different sets of alleles of a particular DNA marker. Finally, many DNA markers can be found scattered throughout each chromosome. DNA markers that are known to be genetically linked to a trait of interest can be used for gene cloning, medical diagnosis, and for investigating trait introgression in plant and animal breeding programmes.

In this study, genetic analysis of several populations of the marble goby, *O. marmoratus* in Malaysia was conducted using two types of genetic markers; 1). mitochondrial NADH subunit 5 and Cytochrome *b* genes and 2). Nuclear genomic DNA of Simple Sequence Repeat (SSR) or also known as microsatellite markers.

Mitochondrial DNA (mtDNA) is widely used as molecular markers to assess variability in stock identification studies of fishes. For several reasons, it is one of the most studied portions of the genome in fishes both for population or evolutionary studies (Avise *et. al*., 1986; Seeb and Crane, 1999; Sivasundar *et. al* 2001). The major advantage of mtDNA analysis over many other molecular techniques is that its higher level of resolution provides a reliable method of examining relationships among closely related taxa. With mtDNA, differences between populations can be

assessed statistically by examining the variation in genotype frequencies between populations. MtDNA sequence differences could be indirectly measured by a technique called Restriction Fragment Length Polymorphism (RFLP) or directly through sequence analysis (Holmes *et al.,* 1999; Aurelle and Berrebi, 2001; Bernatchez, 2001; Koblmuller *et al*., 2005; Koblmuller *et al.,* 2007). The direct sequencing methods, although laborious, provides data in its most elemental and completes form. This method has now been made easier since the development of the PCR which functions to create many copies of the two complementary hypervariable portions of the region of the mtDNA molecule, using flanking primers. Direct sequencing of selected regions of genes without an intermediate cloning step allows rapid assessment of numerous individuals in population genetic studies (Innis, 1988).

Microsatellites are characterised by a core sequence which consists of a number of identical repeat sequences. Due its high level of polymorphism, microsatellites are now being increasingly used as the marker of choice in many studies studies involving closely related individuals. In this study, Random Amplified Microsatellites (RAMs) was used to develop microsatellite markers in the marble goby, *O. marmoratus*. RAMs is a dominant multiloci genetic marker system which can also be utilised to isolate microsatellite regions and develop microsatellite markers. The development of a 5' anchoring procedure allows the consistent anchoring of PCR primers at the 5' ends of microsatellite enabling amplification of two close and inverted simple sequence repeats and the region between them.

Samples were collected from thirteen locations in Malaysia representing natural and cultured populations. These populations which encompass the east and west coast regions as well as those separated by geographical distances were selected

to investigate factors affecting the genetic relationship of this species in Malaysia as well as their potential as broodstocks.

The markers developed in this study are the first set of microsatellites ever isolated for the marble goby, *O. marmoratus*. This population study of *O. marmoratus* from thirteen different populations in Malaysia based on MtDNA and microsatellite markers would allow us to gain insight into the genetic variation existing among the populations. The database obtained would benefit in the genetic identification, discrimination, assessment of levels of variation as well as preservation of this species genetic resources in order to support an adequate management programme and an efficient culture programme.

1.2 Objectives

The objectives of the present study are to ;

1. develop RAMs methodologies for the isolation and characterisation of microsatellite markers in *O. marmoratus*
2. assess the inter- and intra- genetic diversity of populations using microsatellite markers.
3. assess the inter- and intra-genetic diversity of populations using cytocrome *b* and NADH subunit 5 mtDNA genes.
4. determine phylogenetic relationships among populations based on the two marker types.
5. recommend strategies for broodstock selection based on the genetic variability and the relationship among populations.

CHAPTER 2

LITERATURE REVIEW

2.1 *Oxyeleotris marmoratus*

The marble goby, *O. marmoratus* is widely distributed in fresh water habitats throughout Asia; Mekong and Chao Praya basins, rivers and other water bodies in Malaysia, Singapore, Indochina, Philippines and Indonesia (Kottelat *et al*., 1993). This species is an economically important fish in the Southeast Asian region, especially in Malaysia, Singapore, Thailand and Indonesia (Rainboth, 1996). It is highly popular among consumers especially among the Chinese community due to its fine texture, tasty white flesh and is believed to have healing properties. In Malaysia, commercialisation of this species is not widely established. Demands are largely dependant on wild populations, fetching high prices in the market. The marble goby is locally known as ikan ketutu, hantu, ubi, belantuk to the Malays and 'Soon Hock' to the Chinese.

2.1.1 Taxonomic Classification

The marble goby belongs to the suborder Gobioidei in the order Perciformes which is comprised of about 268 genera and approximately 2121 species (Nelson, 1994) and have been variously classified into families and subfamilies in recent times (Akihito *et al.,* 2000). The taxonomic classification of marble goby is summarized below:

Class	: Actinopterygii (ray-finned fishes)
Order	: Perciformes (perch-like fish)
Family	: Eleotridae
Genus	: *Oxyeleotris* sp.
Species	: *marmoratus* or *marmorata*

2.1.2 Morphology and Habitat

This species is best identified in the field by its torpedo-like body shape with large snake-like head, symmetrical patterning on the dorsal surface and rounded, outstretched pectoral fins, two dorsal fins and rounded caudal fin and separated pelvic fins (Lim and Ng, 1990). According to the taxonomic description by Kottelat (2001), the marble goby has the following characteristics: total dorsal spines 7-7; anal spines 1; anal soft rays – 8, with 60 - 65 predorsal scales; without ocellus on caudal peduncle. It also has small eyes on its head and black fins with white spots.

It is a large, solitary, slow-moving fish (especially adults) which typically rests undisturbed at the bottom of quiet streams, canals, swamps, reservoirs and lakes (Lim and Ng, 1990) taking cover among rocks, woody debris or vegetation. It can grow to a maximum size of approximately 65 cm standard length (Kottelat, 2001). The largest gobioid fish may also reach 90 cm of total length (Zhong *et al*., 2004). It rarely moves, even when disturbed but is also predaceous. Although it primarily eats small fishes, it also takes crustaceans, aquatic insects, mollusks and crab which are usually oblivious to its presence (Ukkatawewat, 1979). Larval fish in culture ponds feed on cladocerans, rotifers, chironomids and brachiopods (Ambak *et al*., 2006). This fish displays a brown body with dark mottling giving a marbled appearance. Variation in colouration of its body is influenced by colour of the water and

environment of its habitat; in brackish water, the body colour is darker while, fish caught from crystal clear water, is clearer with visible spots on its body.

2.1.3 Potential Candidate For Aquaculture Programme.

The marble goby *Oxyeleotris* is a highly prized food fish in Malaysia and in many parts of Southeast Asia and therefore is a good candidate for a systematic aquaculture programme. In the year 2000 the combined aquaculture industries of Malaysia, Singapore and Thailand produced 277 tons of marble goby. This quantity represented about 74% of the total global aquaculture of gobioid fishes which valued at about $ 1.2 million USD. In Borneo, marble goby captured for markets in Singapore and also Japan may fetch prices of about $20 per lb ($10 per kg) (Grzimek's Animal Life Encyclopedia, 2005). Most of the fish sold are caught from their natural habitats. Very little biological information is available on this species except for a few reports on its culture (Tan and Lam, 1973; Cheah, 1994), rearing condition (Abol *et al.,* 2005), growth and feeding performance cultured in recirculating aquaculture systems (Ambak *et al*., 2006) and also effects of the different diets on growth and survival rate of the larval stage (Liem *et al*., 2000). In Vietnam, a study conducted in the Truong Dang Cove of Tri An Reservoir demonstrated that cove culture of marble goby was a prominent prospect ecologically, technologically as well as economically (Luong *et al*., 2005). To date there has been no systematic management effort conducted in Malaysia apart from the few aforementioned reports. In an allied study; the evolutionary aspect of gobioid fishes which included *O. marmoratus* was conducted based on the sequence analysis of mitochondrial cytochrome *b* gene by Akihito *et al.* (2000). The study established

the phylogenetic relationships of many morphological groups within the goboids living in diverse ecological habitats.

2.2 Molecular Population Genetics

Population genetics is the study of the factors that determine the genetic composition of a population. A major concern in the field of population genetics is to understand the causes of differentiation between populations across ranges of geographic distribution. These include the factors that determine the evolutionary forces in action, such as natural selection, genetic drift, mutation, recombination and gene flow (Halliburton, 2004) and to note the phylogenetic and biogeographic patterns among living organisms through geographical space and time (Silva and Russo, 2000). The knowledge produced by the use of theoretical and experimental tools has also proven useful for more applied purposes such as forensics, fisheries and conservation (Avise, 1994; Cipriano and Palumbi, 1999, Walsh *et al*., 2007).

The molecular era in population genetics began in 1966 when Lewontin and Hubby (1966) and Harris (1966) introduced the technique of protein electrophoresis or allozyme electrophoresis to population genetics. These two studies provided the impetus for a whole spectrum of investigation involving the union of molecular and population genetics. Starting from the 1980s, it became possible to manipulate and analyse yet another biological molecule, DNA and to apply these techniques to population genetics (and other fields of study). These rapidly gained terrain over allozymes due to the development of the Polymerase Chain Reaction (PCR) (Saiki *et al*., 1988) which enabled million fold amplification of genomic DNA. Initially, restriction enzymes were used to analyse easily isolated pieces of DNA such as

mitochondrial DNA (mtDNA) by the use of a technique called Restriction Fragment Length Polymorphism (RFLP). This was followed by direct DNA sequencing. Kreitman (1983) was the first to apply DNA sequencing techniques to population genetics. This led to a flood of similar surveys in various organisms. MtDNA sequencing has proven very useful for a wide range of phylogenetic studies (Holmes *et a*l., 1999; Wang and Race, 1999; Aurelle and Berrebi, 2001; Bernatchez, 2001, Koblmuller *et al*., 2007; Jondeung *et al*., 2007). However other molecular markers such as microsatellites and SSCP may be the methods of choice for populations genetic studies (Goldstein and Schlotterer, 1999; Hoarau *et al*., 2004; Thai *et al*., 2005; Li *et al*., 2007) due to their various other advantages.

The molecular and statistical tools for studying DNA variation are much more powerful than for protein variation. Since the 1980s, advances in molecular techniques, statistical methods and computer technology have combined to revolutionise the way population geneticists think and work. Nowadays, modern molecular techniques allow us to routinely compare the DNA sequences of dozens, or even hundreds, of individuals for coding as well as non-coding sequences.

As suggested by Silva and Russo (2000), in general the application of molecular techniques to population biology involves four main categories of biological problems including:

1) problems related to the analysis of genetic variation within individuals
2) problems involving variation within populations
3) problems concerned with population structuring

4) biological problems involving genetic variation above the species level.

2.3 Molecular Genetic Markers in Conservation

Information on the population genetic structure of a species should be a principal consideration in planning any conservation and management programme of wild and cultured organisms. The study of the genetic structure of populations through the analysis of molecular polymorphisms began with the pioneering work of Landsteiner (1931). The 1990's saw the rapid application of population genetics to conservation of threatened or endangered population (Tomaso *et al*., 1994; Hoole *et al*., 1998; Shaklee *et al*., 1999)**.** The genetic data obtained could provide essential information to preserve species genetic resources in order to support an adequate management programme and an efficient culture or captive programme. According to Karen *et al.* (2003), determining the degree of genetic differentiation between populations could assist in conservation decisions. Two populations that are genetically different should be conserved as two separate units, however if these two populations are not genetically divergent, conservation of only one may be sufficient for practical purposes (or budgetary constraint). The same concept can be applied to restocking resources. Therefore, wider application of molecular markers for assessment of genetic variability in wild, cultured as well as captive population should be emphasised. A vast array of molecular markers is available as tools in several areas of aquaculture; from broodstock selection and monitoring to quantitative trait loci mapping (QTL) (Myers *et al*., 2001; Usmani, 2003; Liu and Cordes, 2004). Genetic improvement through selective breeding provides the

capability of culturing a better quality animal in less time, with greater survival, and at less cost than animals removed from the wild.

As aforementioned genetic marker can be classified into two type, protein and DNA, each of which has its advantages and disadvantages. This is in parallel to the popularity of DNA analysis (both mitochondrial and nuclear) due to its greater sensitivity and increasing emphasis given to it in the last few decades. In some cases, several authors have compared populations using DNA methods in parallel with protein polymorphism, and populations which exhibit little or no differentiation utilising protein polymorphism often showed significant differences in DNA studies (Sanchez *et al*., 1996; Mjolnerod *et al*., 1997; Estoup *et al*., 1998).

2.4 Mitochondrial DNA

For several reasons, one of the most studied portions of the genome in fishes (for population or evolutionary studies) is the mitochondrial DNA or mtDNA (Avise *et. al*., 1986). It is an ideal genetic marker which has been successfully applied in many studies of population and evolutionary biology of fish (Verheyen and Ruber, 2000; von der Heyden *et al*., 2007). In a survey by Silva and Russo (2000), 54% of the conducted DNA – based studies used mitochondrial DNA, and this seems currently to be one of the most widely studied molecule as population-level genetic markers. This preference could be due to the fact that mitochondrial genes evolve, on average, faster than nuclear structural genes (Gillham, 1994) and this make them particularly suitable for population level studies. However the absolute rate of

mtDNA evolution among various organisms may vary (Martin *et al*., 1992; Bargelloni *et al*., 1994; Cantatore *et al*. 1994;).

Mitochondria is a spherical or elongated organelle found in the cytoplasm of nearly all eukaryotic cells. It contains genetic material and many enzymes important for cell metabolism, including those responsible for the conversion of food to usable energy. Mitochondria and chloroplast are believed to originate from ancient symbiont DNA. Both chloroplasts and mitochondria DNA replicate independently of the nuclear DNA. The size of the mtDNA in bony fishes is about 17.0 – 18.0 thousand base pairs (Ohno, 1974). Unlike nuclear DNA, mtDNA has no repetitive sequences, nor spacers or introns. However, their DNA is made up of numerous genes coding for respiratory and other functions. Its gene content appears to be conserved: There are two ribosomal RNA (rRNA) genes, 22 transfer RNA (tRNA) genes and 13 protein genes. The probability of extracting mtDNA is greater than for nuclear DNA because mtDNA molecules are present in hundreds to thousands of copies per cell compared to the nuclear complement of two copies per cell. Mitochondria DNA is generally maternally inherited. In consequence all offspring have mitochondrial genomes identical to their mother. There is little or no paternal contribution of mtDNA in most organisms and there is no known recombination between mitochondrial genomes (Avise and Smith, 1974). No active mechanism is known to be existed that incorporate diversity in mtDNA. Thus, it is unchanged over thousands of generations with the exception of random mutations. The evolutionary rate as well as the genetic differentiation of mtDNA among populations is thought to be approximately 5 to 10 times higher than that exhibited by nuclear genes (Birky *et*

al.,1989) accounting for greater sensitivity. This higher level of resolution provides a reliable method of examining relationships among closely related taxa.

MtDNA sequence differences can either be measured directly through restriction site analysis or by sequence analysis. Restriction site analysis is less expensive and allows rapid screening of many samples. The pattern of restriction site gain or loss among genotypes for a panel of enzymes can be used to construct phylogenetic trees of population or species (Avise *et al*., 1986). The direct sequencing method, although laborious, provides data in its most elemental and complete form. This method has now been made easier since the development of the PCR which functions to create many copies of the complementary hypervariable portions of the mtDNA region by using flanking primers. When adequate amounts of PCR product have been amplified to provide all the necessary information, sequencing reactions are performed. These chemical reactions use each PCR product as a template to create a new complementary strand of DNA, in which the nucleotide bases that make up the DNA sequence are labelled with dye. The strands created in this stage are then separated according to size by an automated sequencing machine that uses a laser to "read" the sequence, or order, of the nucleotide bases. Where possible, the sequences are determined on both strands of the double-stranded DNA molecule that characterize that particular sample (Melton and Sensabaugh, 2000). Direct sequencing of partial or complete genes allows rapid and precise assessment of numerous individuals in population genetic studies (Innis *et al*., 1988). Many softwares are available to statistically assess the mtDNA variation in genotype frequencies between populations.

2.4.1 Application of Mitochondrial DNA in Fish Genetics

In recent years, mtDNA have provided new information concerning the genetic variability of wild and cultivated populations of several fish species (Iguchi *et al*., 1999; Sivasundar *et al*., 2001; Kuriiwa *et al*., 2007). Billington and Hebert (1991) reviewed patterns of mtDNA variation in 40 fish species. Mitochondria DNA analyses have been employed in a number of research applications involving various aquatic organisms such as on salmon (Apostolidis *et al*., 1996: Shaklee *et al*., 1999); summer flounder (Jones and Quattro, 1999), Japanese *Mugilogobius* (Mukai *et al*., 2000), sea urchins (Palumbi and Wilson, 1990), horseshoe crab (Saunders *et al*., 1986), Penaeid shrimp (Palumbi and Brand, 1993; Hurwood *et al*., 2003), mussel (Edwards and Skibinski, 1987), common carp (Kohlmann *et al*., 2003) mahseer species (Nguyen *et al.*, 2006), and cyprinids (Perdices *et al*., 2004; Perdices *et al*., 2005). The results of these studies have among others proven population genetic structuring (Caccone *et al*., 1997; Chen *et al*., 2004; Barbanera *et al*., 2005), taxonomic differences (Smith and Bush, 1997), evolutionary relationship (Puterka *et al*., 1993) and identified and management stocks (Grewe and Hebert, 1988; Billington and Hebert, 1991) in the various organisms investigated.

The importance of mtDNA sequences to examine phylogenetic relationships and systematics among fishes over other techniques has been extensively reviewed (Kocher and Stepien, 1997). More recently, the trend has been to use PCR to amplify and sequence portions of specific fast evolving genes, for example, the Cyt *b* gene as it has a rate of base substitution that makes it particularly suitable for comparisons of

closely related species and genera (Whitmore *et al.,* 1994). An early example of this approach was the identification of differences in the Cyt *b* sequences of four different tuna *Thunnus* spp. (Bartlett and Davidson 1991). Numerous other molecular systematic studies of fishes have been published using the Cyt *b* gene, including studies of salmonid phylogenies (McVeigh and Davidson 1991; Patarnello *et al.,* 1994; Park *et al.,* 2000), relationships within the smelt genus *Osmerus* (Taylor and Dodson 1994), genetic identification of sharks (Heist and Gold 1998), and phylogenetic relationships among African cichlid fishes (Meyer *et al.,* 1990; Sturmbauer *et al.,* 1994), carps (Cyprinidae) (Gilles *et al.,* 1998; Tsigenopoulos and Berrebi, 2000), cods (Gadidae) (Carr *et al.,* 1999), and pangasiid catfishes (Pangasiidae) (Pouyard *et al.,* 2000). Other regions of the mtDNA molecule have also been used for molecular systematic analyses, including the D-loop or control region (Meyer et al. 1990; Shedlock *et al.,* 1992; Brown *et al.,* 1996; Faber and Stepien 1997; Stepien and Faber, 1998), ND2 and ND4 L genes (Broughton and Gold 2000), the ribosomal RNA (12S and 16S) genes (Bargelloni *et al.,* 1994; Tringali *et al.,* 1999; Sloss 2001), and cytochrome oxidase subunit I (Yokobori *et al.,* 1994; Carr *et al.,* 1999). More recent studies involve examination of two or more mitochondrial genes (Gilles *et al.,* 1998; Birstein and DeSalle 1998; Sloss, 2001; de Francisco and Galetti, 2005). The choice of mtDNA region examined depends upon the phylogenetic level of the hypothesis that is being tested. These levels range from examining intraspecific relationships (e.g., D-loop, ND 1, ND 3/4, and ND 5/6) (Miya and Nishida, 2000) as well as interspecific and intergenera relationships between closely related organisms through moderately evolving genes (e.g., cytochrome *b* to the slowly evolving 12S and 16S rRNA (Krzywinski, 2001) and Cytochrome Oxidase I genes for family level comparisons. However, levels of

evolutionary rates of these genes may sometimes differ in various organisms (Carvalho and Pitcher, 1995).

2.5 Microsatellite Markers

The term microsatellite refers to a class of co-dominant DNA markers which are inherited in a Mendelian fashion. Microsatellites are composed of tandemly repeated, simple DNA sequence motifs of as many as six nucleotides in length. These loci are commonly found throughout both prokaryotic and eukaryotic genomes and typically are highly polymorphic within species and populations (Jeremy and John, 2004). In addition, these co-dominant genetic markers are relatively easy to score and have high reproducibility and specificity. As such, microsatellites have become one of the most popular classes of molecular markers and are commonly employed to investigate the population genetics of a diverse range of organisms (Bruford and Wayne, 1993; Goldstein and Schlotterer, 1999; Castro *et al*., 2007). Due to their very high levels of polymorphism they have been frequently utilised to understand population genetic structure as well as genetic diversity within broodstocks. This knowledge has proven to be one of the most effective molecular tool towards an effective breeding programme (Carvalho and Pitcher, 1995). Nguyen *et al*., (2006) reported that the microsatellite markers provided a valuable molecular tool for detailed genetic analysis of stock structure in wild populations as well as of brood stock of cultured mahseer species. The data from their study may contribute significantly to the development of a sound management plan for mahseer species.

2.5.1 Microsatellite Characteristics

Microsatellites are alternatively known as simple sequences repeat (SSRs) (Jacob *et al*., 1991), short tandem repeats (STRs) (Craig *et al*., 1988), simple sequence length polymorphism (SSLPs) (Rassmann *et al*., 1991) or variable number of tandem repeats (VNTRs) (Innan and Tajima, 1997). These repetitive element consists of tandem reiterations of simple sequence repeats and are typically composed of two to four nucleotides such as $(AC)_n$ or $(GATA)_n$ where n lies between 5-50 (De Woody and Avise, 2000).

Due to its high mutation rate, there may exist many alleles of a single microsatellite locus in a particular population. These alleles differ in the number of repeats. For example, one allele may have seven repeats of CT motif and another allele may have eight repeats. An diploid individual who is homozygous for a locus will have the same number of repeats on both chromosomes, whereas a heterozygous individual will have different number of repeats on the two chromosomes. However, regions surrounding the microsatellite locus, called the flanking region, will have the same sequence. This is important because the flanking regions can therefore be used as PCR primers when amplifying microsatellite loci, and may sometimes be conserved across genera or even families. In Figure 2.1, the two lines represent the sequences on two homologous chromosomes in a diploid organism. For clarity, only one strand of each chromosome is shown (Moxon and Wills, 1999).

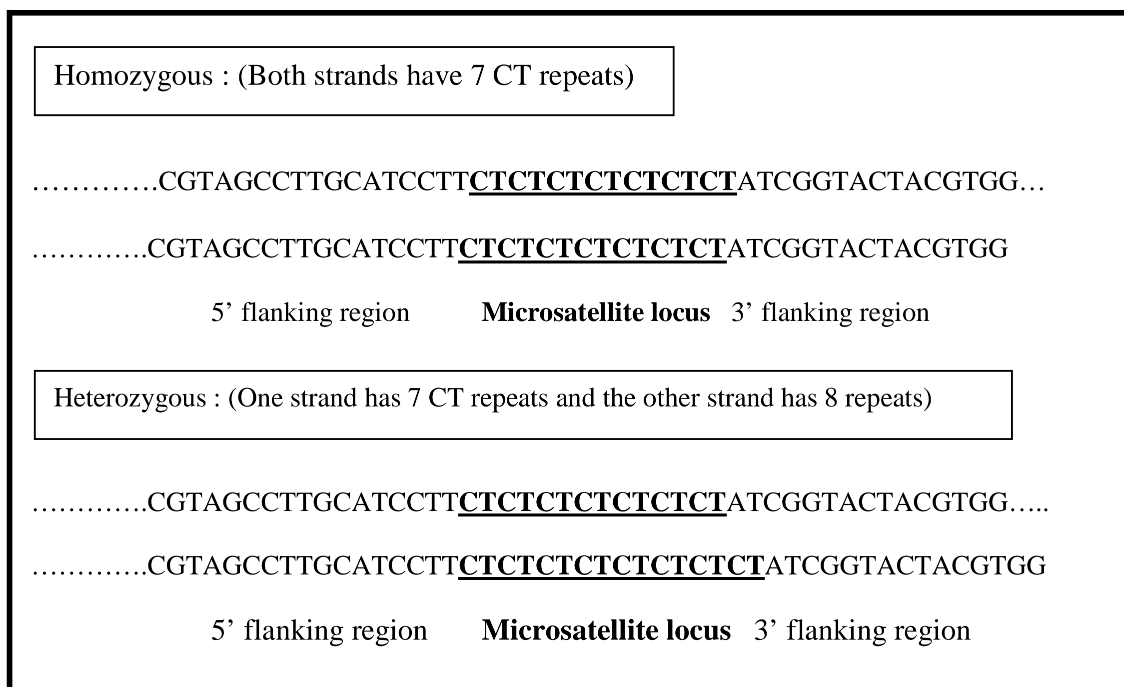


Figure 2.1 : Homozygous and heterozygous microsatellite sequences in a diploid organism (Adapted from Moxon and Wills, 1999).

If additional repeat is added to the existing sequence, then this slightly larger version can be passed on to offspring who will usually replicate it accurately. Over time, as animals in a population breed, they will recombine their microsatellites during sexual reproduction and the population will maintain a variety of microsatellites that is characteristic for that population and distinct from other populations which do not interbreed.

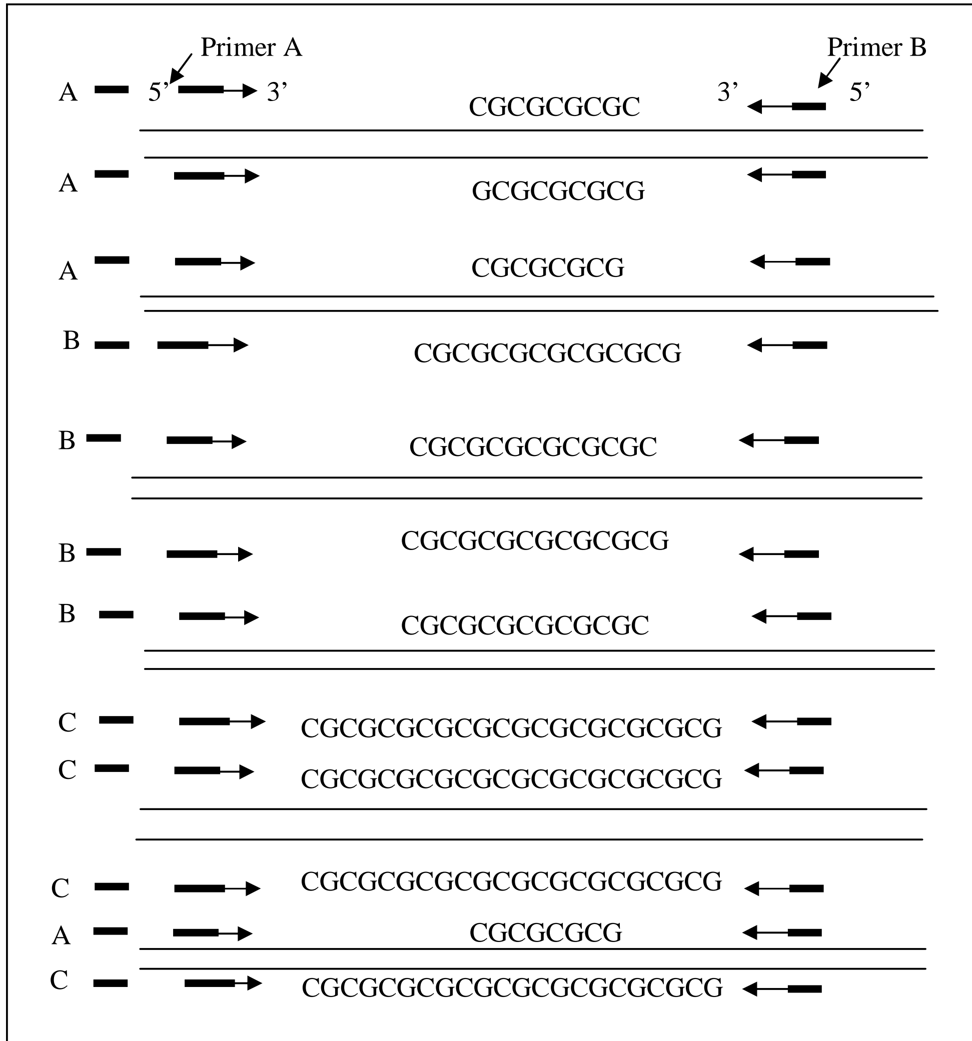
Microsatellites are classified into 3 categories, perfect, imperfect and compound as defined by Weber (1990). Perfect repeats are uninterrupted stretches of repeat units, while imperfect repeats have one to three intervening bases with repeat sequence on either side. Compound repeats consist of several different repeat types and are separated by less than three bases. Compound repeats are further subdivided

into perfect and imperfect sequences (Pongsomboon *et al*., 2000). The examples of these three categories is shown in Table 2.1 below:

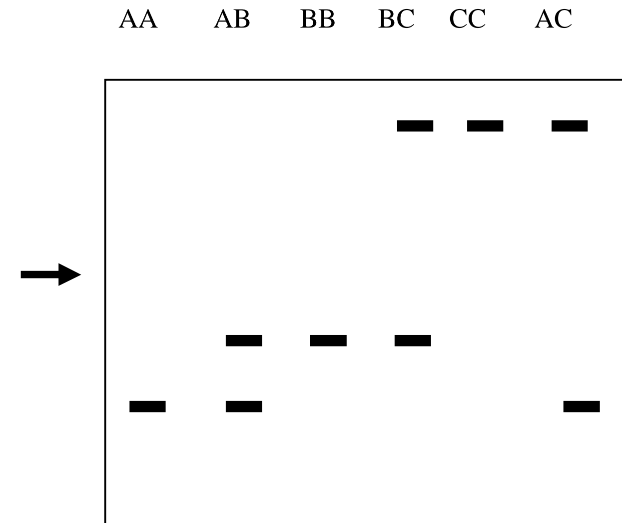
Table 2.1 : Examples of three categories of microsatellite sequences.

Microsatellite sequences categories	**e.g**
Perfect	CAACAACAACAACAACAACAACAACAA
Imperfect	CACACACACA**TG**CCACACACACATGCA
Compound (perfect)	GAGAGAGAGAGATCTCTCTCTCTCT
Compound Imperfect	GAGAGAGAG**G**GATCTCTCTCTCT**T**TCT

Populations may show variation in the number of repeats at particular microsatellite loci often producing an abundance of alleles distinguishable by molecular size. Microsatellite array are generally quite short, 20 to 300 bp, therefore microsatellites can be successfully isolated and characterized for variability of alleles using the polymerase chain reaction (Tautz, 1989; Weber and May, 1989). Figure 2.2 shows the detection of microsatellite variation by PCR primer recognition flanking sequences and how polymorphic amplified product can be identified on a polyacrylamide gel.



(a)



(b)

Figure 2.2a&b: Detection of Microsatellites.

Note – (a) Microsatellites are detected by amplifying the segment containing different numbers of repeats using the primers flanking the segments. Both chromosome homologs are shown. (b) The position of bands in the gel is determined by the molecular weight of the DNA. The longer the repeats the heavier the molecular weight of the segments. The smaller fragments move faster than the heavier fragments.

2.5.2 Evolution of Microsatellite

Microsatellites are among the fastest evolving DNA sequences, with length mutation rates ranging from 10^{-2} and 10^{-5} mutations per generation (Edwards *et al.* 1992; Weissenbach *et al.*, 1992; Weber and Wong 1993). Two mechanisms have been proposed, to explain the high mutation rates of microsatellite : 1) unequal crossing over in meiosis and 2) strand-slippage replication (Levinson and Gutman, 1987). From these two, strand-slippage replication appears to be the more predominantly accepted hypothesis (Schlotterer and Tautz, 1992; Strand *et al*., 1993; Weber and Wong, 1993). Strand-slippage is speculated to occur primarily during lagging strand synthesis of DNA strand upon dissociation of the polymerase complex (Schlotterer and Tautz, 1992). This slippage creates a transient bulge which upon DNA repair would be either removed or lead to the elongation of the repeat (Schlotterer and Tautz, 1992). This slippage frequently results in greater numbers of tandem repeats (Levinton and Gutman, 1987; Weber, 1990; Wierdl *et al*., 1997). Alternatively, the formation of a transient bulge in the template strand may also lead to the shortening of the repeat. As a consequence this slipped-strand mispairing results in the nascent strand having a different number of repeats from the template strand once DNA replication is complete.

2.5.3 Application of Microsatellite in Fish Genetics

Microsatellite marker variations have a number of advantages in aquaculture and fisheries research over other molecular markers (O'Reilly and Wright, 1995) and have a significant impact on the application of genetics for these two types of research. Its application range from simple population studies to quantitative trait

linkage analysis (Usmani, 2003**;** Qin *et al.*, 2007). As in the mtDNA previously described, the assay of microsatellite variation is based on the PCR technique, thus only small amounts of tissue, for example from fish scales or fin, are needed as a source of DNA. Edwards *et al.* (1991) examined the frequency of 5 microsatellite loci (tri and tetra repeats) on the X chromosome. They found that for either the tri or tetra microsatellite loci, any given repeat was found every 300 to 500 Kb. From this they estimated that for all the 44 possible unique trimeric and tetrameric repeats there were 400,000 loci or about 1 every 10 to 20 Kb. Of the class of loci examined by Edwards *et al.* (1991) 50% were polymorphic.

A number of studies have assessed the superiority and sensitivity of microsatellite markers over other markers in fish population genetics. For example, although allozymes and microsatellite showed similar patterns of differentiation of Irish and Spanish populations of Atlantic salmon, microsatellite loci showed higher levels of variation (Sanchez *et al*., 1996). In another study Atlantic salmon was shown to have low levels of genetic differentiation relative to other salmonid species using allozymes and mtDNA. However, with microsatellite, McConnell *et al.* (1995) were able to discriminate clearly between Canadian and European fish. Tessier *et.al*. (1995) found significant differences between wild and first-generation hatchery fish in Atlantic salmon using the same technique. In another study on cod, microsatellite loci provided evidence of population structure at finer geographical scale than that shown by other techniques (Ruzzante *et al*., 1996).

Microsatellite markers have also been utilised for genetic mapping (Weissenbach *et al.,*1992) and have been extensively used for linkage analyses in association with disease susceptibility genes (Robinson *et al.,*1996). In addition they have proven useful in the analysis of paternity and kinship (Queller *et al.,* 1993) and in the probability of sample identity at both the individual (Edwards *et al,* 1992) and population levels (Paetkau *et al.,*1995). Microsatellite variation has been used to study the amount of hybridisation between closely related species (Gottelli *et al.,* 2004; Roy *et al.,* 1994). In the study of entire populations microsatellites are also very useful to elucidate population diversity and structuring (Bruford and Wayne 1993). Comparison of levels of variation between species and populations have also proven useful in the assessment of overall genetic variation (Gottelli *et al.*, 2004; Paetkau and Strobeck 1995; Taylor and Dodson, 1994). They can be used to estimate effective population size (Allen *et al.*, 1995) and to gain insight into the degree of population substructure including both the amount of migration between subpopulations (Allen *et al.,* 1995; Gottelli *et al,*. 2004) and genetic relationships among various subpopulations (Bowcock *et al.,* 1994; Estoup *et al.*,1996; Forbes *et al.*, 1999).

2.6 Mitochondrial DNA versus Microsatellite Analysis

Over the last few decades mitochondrial DNA has had numerous utilisation in various organisms (Brown *et al*., 1982). The more slowly evolving regions tend to be useful for phylogenetic studies, while the more rapidly evolving regions may be used for population studies (Baker *et al*., 1993; Arnason and Gullberg, 1996). The lack of intermolecular recombination means that the mtDNA genome is inherited as a single locus. Since mtDNA is maternally inherited in most animals, discrepancies