



UNIVERSITI PUTRA MALAYSIA

**ISOLATION, CHARACTERIZATION AND GENETIC LINKAGE
MAPPING OF MICROSATELLITE MARKERS IN MYSTUS NEMURUS**

HOH BOON PENG.

FS 2006 11



**ISOLATION, CHARACTERIZATION AND GENETIC LINKAGE MAPPING
OF MICROSATELLITE MARKERS IN *Mystus nemurus***

By

HOH BOON PENG

**Thesis submitted to the School of Graduate Studies, Universiti Putra Malaysia, in
Fulfilment of the Requirement for the Degree of Doctor of Philosophy**

December 2005



To my dearest wife.....

YAN LI



Abstract of thesis presented to the Senate of University Putra Malaysia in fulfillment of the requirement for the degree of Doctor of Philosophy

**ISOLATION, CHARACTERIZATION AND GENETIC LINKAGE MAPPING
OF MICROSATELLITE MARKERS IN *Mystus nemurus***

By

Hoh Boon Peng

July 2005

Chairperson: Associate Professor Siti Shapor Siraj, PhD

Faculty: Science

The river catfish (*Mystus nemurus*) is of great potential importance as an alternative fish protein source, notably in Malaysia and Thailand. The seed supply is seasonal and inability to reproduce in captivity is a hindrance to mass produce this species. Molecular breeding and genetic mapping with the use of DNA based markers have paved ways in overcoming conventional breeding to improve important economic traits.

The main objectives of this study were to isolate and develop microsatellite markers for *Mystus nemurus*, and apply these markers on genetic linkage map for this species. Two methodologies were used to isolate the microsatellite loci in *M. nemurus*, namely Random Amplified Hybridizing Microsatellites (RAHMs) and 5' anchored Polymerase Chain Reaction (5' anchored PCR). A total of 236 repeat sequences were identified. Of these, 13 were developed from RAHMs and 223 were developed from 5' anchored PCR.



A total of 177 primer pairs were developed. Of these 64 have been screened for polymorphisms on 90 unrelated fishes collected from six locations (Kedah, Selangor, Perak, Johor, Terengganu and Sarawak), where each location consisted of 15 individuals. Forty four primer pairs were found to be polymorphic. The highest heterozygous marker was MnBp5-1-12a (0.8714) while the lowest heterozygosity was observed in MnBp8-1-25a (0.0000). A dendrogram was constructed from the 44 polymorphic loci. An additional 34 polymorphic markers, previously generated, were combined in order to construct another dendrogram, and it was found that the dendrogram stabilized at 50 polymorphic microsatellite loci.

Induced breeding was carried out to generate two known fish family groupings. The first family grouping was generated from the parents chosen randomly from the Terengganu broodstock. One hundred progeny were produced. The second family grouping was generated from the male chosen from Terengganu population and the female chosen from Pahang population. Fifty progenies were used for the analysis of Mendelian segregation and linkage mapping.

A total of 70 microsatellite markers isolated for this species were screened and evaluated for the Mendelian Inheritance ratio on the two fish families, mentioned of which 31 showed segregation in either one of the families. These markers were analyzed with contingency chi-square analysis. Of these, 13 pair markers were grouped into 8 groups. Strategy "Pseudo-testcross" was applied in the linkage mapping analysis. Analysis of linkage using LOD score was carried out and a total of 7 linkage groups



were generated from the 2 fish families under the LOD of 1.2. Additional analysis of LOD was performed on the family data of 44 dominant markers done previously. Three linkage groups were obtained at the LOD of 1.2 and fractionated into 6 groups when the LOD score was increased to 2.0.

This is the first map generated for *M. nemurus* which provides information on genetic linkage for this species. It is a stepping-stone for carrying out more intensive genetic mapping and QTL analysis on this species in the future. Eventually, marker assisted selection (MAS) or isolation of economically important genes could be made possible and these would benefit the Malaysian aquaculture industry

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Falsafah Kedoktoran

**PEMENCILAN, PEMBANGUNAN, DAN ANALISIS PEMETAAN PENANDA
MIKROSATELIT UNTUK *Mystus nemurus***

Oleh

HOH BOON PENG

Julai 2005

Pengerusi: Professor Madya Siti Shapor Siraj, PhD

Fakulti: Sains

Ikan baung (*Mystus nemurus*) adalah penting sebagai sumber protein ikan alternatif, terutamanya di Malaysia dan Thailand. Bekalan benih ikan adalah bermusim dan ketidakeupayaan spesies ini membiak dalam kurungan menyebabkan pengeluaran spesies ini tidak dapat dihasilkan secara besar-besaran. Pembiak kacukan molekul dan pemetaan genetik dengan penggunaan penanda genetik berasaskan DNA membuka jalan bagi memperbaiki kelemahan perbiakbakaan konvensional untuk mendapatkan trait-trait ekonomi yang lebih baik.

Objektif utama kajian ini ialah untuk memencilkan dan membangunkan penanda mikrosatelit untuk *M. nemurus*, dan mengaplikasikan penanda-penanda ini dalam analisis pemetaan genetik untuk spesies ini. Dua cara telah digunakan untuk memencilkan lokus mikrosatelit bagi *M. nemurus*, iaitu "Random Amplified Hybridization Microsatellites (RAHMs)" dan "5' anchored PCR". Sejumlah 236



jujukan mikrosatelit telah dikenalpasti. Daripada jumlah itu, 13 dapat dipencilkan daripada RAHMs manakala 223 dipencilkan daripada 5' anchored PCR.

Sebanyak 177 pasang primer spesifik telah direka, 64 daripadanya telah digunakan dalam penentuan polimorfisme ke atas 90 ikan yang disampel dari 6 lokasi (Kedah, Selangor, Perak, Johor, Terengganu dan Sarawak), di mana setiap lokasi mengandungi 15 individu ikan. Didapati 44 pasangan primer adalah polimorfik. Penanda yang mempunyai heterozigositi tertinggi ialah MnBp5-1-12a (0.8714), manakala heterozigositi yang terendah ialah MnBP8-1-25a (0.0000). Dendrogram dibina berasaskan kepada 44 lokus polimorfik tersebut. Penambahan 34 penanda polimorfik yang dibangunkan terdahulu digabungkan untuk membina satu lagi dendrogram. Didapati bahawa dendrogram ini adalah stabil pada jumlah lokus mikrosatelit polimorfik 50.

Pembiakbakaan aruhan dilakukan untuk menghasilkan dua kumpulan famili ikan. Kumpulan pertama dihasilkan daripada induk yang dipilih secara rawak dari stok Terengganu. Sebanyak 100 progeneri ikan dihasilkan. Untuk kumpulan kedua, induk jantan dipilih dari stok Terengganu dan induk betina dipilih dari stok Pahang. Sebanyak 50 progeneri telah digunakan untuk analisis segregasi Mendel dan pemetaan genetik.

Sejumlah 70 penanda mikrosatelit yang dipencilkan khusus untuk spesies ini telah disaring dan dinilai untuk Nisbah Pewarisan Mendel terhadap kumpulan dua famili ikan, yang mana 31 penanda menunjukkan segregasi dalam salah satu daripada

kumpulan famili tersebut. Penanda ini dianalisis dengan Analisis Kontengensi Khikuasa dua. Daripada ini, 13 pasang penanda dikumpulkan kepada lapan kumpulan. Strategi “Pseudo-testcross” diaplikasikan untuk analisis pemetaan rangkaian. Analisis rangkaian pemetaan dengan menggunakan skor LOD telah dilakukan dan sejumlah tujuh kumpulan rangkaian telah dihasilkan daripada dua kumpulan famili ikan apabila skor LOD ditentukan pada 1.2. Analisis tambahan LOD dilakukan terhadap data yang diperolehi daripada 44 penanda dominan yang telah dikaji sebelum ini. Tiga kumpulan rangkaian telah diperolehi pada LOD 1.2. Ini adalah peta genetik pertama bagi *M. nemurus* yang memberi informasi tentang rangkaian genetik untuk spesies ini. Kajian ini merupakan batu loncatan untuk menghasilkan pemetaan genetik yang lebih intensif dan analisis QTL bagi spesies ini pada masa hadapan. Akhirnya, Pemilihan Berbantuan Penanda (MAS) atau pemencilan gen yang berkepentingan ekonomi mungkin boleh direalisasikan dan ini akan menguntungkan industri akuakultur Malaysia.



ACKNOWLEDGMENT

I would like to thank my supervisors Assoc. Prof. Dr. Siti Shapor Siraj, Prof. Dr. Tan Soon Guan and Prof. Dr. Datin Khatijah Yusoff for their continuous support, direction, cooperation and interest in my research and their comments and suggestions on many aspects of this work.

I gratefully acknowledge the financial support from the National BioDirectorate grant 01-02-04-0074, the Ministry of Science, Technology and Innovative Malaysia, headed by Assoc. Prof. Dr. Siti Shapor Siraj. I am especially indebted to her for suggesting the problem and guiding me throughout the entire period of this project and for giving me the perfect environment to work in.

To Dr. Chong Lee Kim, thank you for providing me and sharing the data and made my research successful.

My thanks also goes to my fellow laboratory members, Dr. Sahar Usmani who is now in Yale University, USA and Mr. Chan Soon Choy, Miss Siti Nor Hajjar, Mr. The Chee Hong and Miss Ong Ching Ching, for their advice, friendship, helps and sharing. They had made my stay in the laboratory a wonderful experience. To Dr. Vijaya Kumar who is now in Universiti Sabah Malaysia, I specially appreciate for his comments and his help in using the PHYLIP analysis software. Additional thanks to Mr. Azmi Yaakub and Mr. Lee Kok Kuan for helping me with the fish breeding part of my research. To



Dr. Subha Bhassu, I appreciate for her advice and comments given to me throughout my entire research life.

Thanks also goes to my students, Chew Wee Chee, Norwahidayanty Kenasipan, Tham Poh Theam, and Khoo Chin Fui, who made teaching an exciting experience for me.

To Dad, Mum and my younger sister, Wei Nee, without their emotional support, love, and confidence, and trust in me I would never have been gone through all the hard times.

Last but not least, to my beloved wife, Yan Li, who had never complained for all the lonely minutes and seconds I had made her gone thru during my entire research life. Thank you for being so understanding and trust to me.

HOH BOON PENG

X



Table of Contents

	Page
ABSTRACT	III
ABSTRAK	VI
ACKNOWLEDGEMENTS	IX
DECLARATION	XI
LIST OF TABLES	XVII
LIST OF FIGURES	XVIII
LIST OF ABBREVIATIONS	XIX
CHAPTER	
1 INTRODUCTION	1
2 LITERATURE REVIEW	
2.1 <i>Mystus nemurus</i>	6
2.2 Genetic Marker Studies of <i>M. nemurus</i>	9
2.3 Microsatellites	9
2.4 Application of Microsatellites	12
2.5 Five Prime Anchored PCR	14
2.6 RAHM (Random Amplified Hybridization Microsatellites)	15
2.7 Genetic Linkage Mapping	17
3 MATERIALS AND METHODS	
3.1 Samples Collection	20
3.2 DNA Extraction	23
3.2.1 Standard Phenol-Chloroform DNA extraction	23
3.2.2 QIAamp DNA Mini Kit (Qiagen)	24
3.3 Five Prime Anchored PCR	25
3.3.1 Cloning	27
3.3.2 Plasmid Extraction	28
3.3.3 Sequencing of Plasmid DNA	29
3.3.3.1 Automated Sequencing	29
3.3.3.2 Manual Sequencing	29
3.4 RAHM (Random Amplified Hybridization Microsatellites)	30
3.4.1 RAPD amplification	31
3.4.2 Capillary Transfer of DNA to Nylon Membrane	31
3.4.3 Probe Labelling	32
3.4.4 Southern Blot Hybridization	32
3.4.5 Cloning	33
3.4.6 Replica plating and DNA fixing	33
3.4.7 Probe Labeling	33
3.4.8 Hybridization of Filters	34
3.5 Submission of DNA Sequences to GenBank	34
3.6 Designing of Microsatellie Primer Pairs	35



3.7	Characterization of Microsatellite Markers	36
3.7.1	Heterozygosity	37
3.7.2	Genetic Distance	37
3.8	Mendelian Inheritance Study	38
3.8.1	Generating of Known Family Groupings	38
3.8.2	Analysis of Allelic Inheritance	39
3.9	Genetic Linkage Mapping	40
3.9.1	Mapping Populations	40
3.9.2	Microsatellite Markers	40
3.9.3	Linkage analysis	41
4	RESULTS	
4.1	Isolation of Microsatellites	49
4.1.1	Isolation of Microsatellites by the RAHMs Enrichment Protocols	49
4.1.2	Isolation of Microsatellites by 5' Anchored PCR Protocol	56
4.1.3	Analysis of Basic Local Alignment Search Tool (BLAST)	71
4.2	Primer Design	75
4.3	Characterization of Microsatellite Primers	83
4.4	Genetic Distances and Dendrogram	99
4.5	Mendelian Inheritance	105
4.5.1	Generation of families	105
4.5.2	Medelian Segregation of Microsatellite Markers	106
4.6	Linkage Analysis	112
4.6.1	Chi-square Analysis for Marker-marker Linkage	112
4.6.2	Linkage Mapping Analysis and Map Construction	116
5	DISCUSSION	
5.1	Isolation of Microsatellites	130
5.1.1	Isolation of Microsatellites by 5' Anchored PCR	131
5.1.2	Isolation of Microsatellites by RAHMs	132
5.1.3	Microsatellite Repeat Unit Length	134
5.1.4	BLAST Analysis	135
5.2	PCR Amplification and Polymorphisms of Microsatellite Loci	136
5.3	Gel Electrophoresis	136
5.4	Characterization of Microsatellite Markers	137
5.4.1	Genetic Variability of Microsatellite Markers	137
5.4.2	Juxtaposed Microsatellite System	142
5.5	Mendelian Segregation Studies on the Microsatellite Markers	143
5.6	Linkage Analysis	145
5.6.1	Chi-square Analysis of Linkage	145
5.6.2	Analysis of Linkage	147
5.7	DNA Marker Technologies and Their Applications in Aquaculture Industry	151



6	CONCLUSION	153
	REFERENCES	156
	APPENDICES	
	BIODATA OF AUTHOR	185



LIST OF TABLES

Table	Page
3.1 Sampling location, size and year of collection of <i>Mystus nemurus</i>	21
3.2 List of anchored primers designed and optimized PCR annealing temperatures	26
3.3 Microsatellite primers and annealing temperature used in the mapping study. References given when available	43
4.1 Microsatellites of <i>Mystus nemurus</i> isolated by the RAHMs procedure.	51
4.2 Microsatellite repeats of <i>M. nemurus</i> isolated by 5' anchored PCR	58
4.3 Sequence of microsatellite primers designed for <i>M.nemurus</i> , primer sequences and the expected size of amplicon.	76
4.4 Microsatellite loci identified in <i>M. nemurus</i> with annealing temperature, observed allele sizes, number of alleles, observed and expected heterozygosities, and number of individuals screened for polymorphisms.	88
4.5 Chi-square and likelihood ratio test for HWE in the five populations studied.	94
4.6 Matrix of Nei's (1978) genetic distance coefficients.	101
4.7 Matrix of Nei's (1978) genetic distance coefficients calculated from (A) 34, (B) 50 and (C) 78 loci respectively.	101
4.8 Microsatellite loci genotype numbers among the F ₁ progeny and χ^2 values for the expected Mendelian segregation ratio.	110
4.9 Pairs of loci that deviated significantly from Mendelian segregations	113
4.10 Summary of statistics for the linkage map of the three mapping populations of <i>M.nemurus</i> .	118
4.11 Pairwise analysis of linkage with LOD score and recombination fraction	119



LIST OF FIGURES

Figure		Page
2.1	Morphology of <i>M. nemurus</i> (lateral view)	8
3.1	Malaysia map indicating the location of the sampling sites.	22
4.1	Comparison of the RAPD PCR profile and Southern Blot hybridization profile.	52
4.2	Colony hybridization of clones of the products of primer OPC15 obtained with the (AC) ₁₀ probe.	53
4.3	Microsatellite repeat sequence revealed by manual sequencing of clone Mnc23 obtained from cloning the products of RAPD primer OPC15.	54
4.4	Electrophoregram showing the (CA) _n microsatellite obtained from the RAHMs enrichment protocol.	48
4.5	Electrophoresis of the PCR products of the 5' anchored PCR protocol.	66
4.6	Plasmids of 5' anchored PCR clones.	67
4.7	(GA) _n & (CA) _n microsatellite repeat obtained from clone MnBp5-2-27	68
4.8	(GA) _n microsatellite sequences obtained from clone MnBp5-1-01.	69
4.9	(GA) _n and (GT) _n microsatellite sequences obtained from clone Mnbp5-1-02	70
4.10	BLAST search result of MnBp8-4-43.	72
4.11	BLAST search result of MnBp5-2-15.	73
4.12	BLAST search result of MnBp5-2-05.	74
4.13	Microsatellite profile of <i>M. nemurus</i> from the UPM population using primer Mnc23.	85
4.14	Microsatellite profile of <i>M. nemurus</i> from the Perak population using primer Mnc65b.	85



4.15	Microsatellite profile of <i>M. nemurus</i> from the Kedah population using Primer MnBp5-2-02b.	86
4.16	Microsatellite profile of <i>M. nemurus</i> from the Terengganu population using primer MnBp5-2-02b.	86
4.17	Microsatellite profile of primer Mnc441 <i>M. nemurus</i> from the Kedah population electrophoresed on 8% polyacrylamide gel stained with ethidium bromide.	87
4.18	Comparison between 4% Metaphor gel and 8% polyacrylamide gel electrophoresis of the microsatellite profile of primer Mnc340 from the Kedah population.	87
4.19	UPGMA dendrogram based on Nei's (1978) genetic distance calculated from 44 microsatellite markers.	102
4.20	UPGMA dendrogram based on Nei's (1978) genetic distance calculated from 34 (A), 50 (B) and 78 (C) microsatellite loci respectively.	103
4.21	UPGMA dendrogram based on Nei's (1978) genetic distance calculated from 34 (A), 50 (B) and 78 (C) microsatellite loci respectively (continued).	104
4.22	Examples of the genotypic profiles of the parents.	108
4.23	Genotypic profile of Primer MnBp5-1-30b in Family A.	108
4.24	Genotypic profile of Primer Mnc23 in Family A.	109
4.25	Genotypic profile of Primer MnBp5-4-20b.	109
4.26	Linkage groups formed from the contingency chi-square analysis.	114
4.27	Linkage map formed in Family A at LOD score 1.5.	123
4.28	Linkage map formed in Family B at LOD score 1.2.	124
4.29	Linkage map formed in Family C at LOD score 1.2.	125
4.30	Linkage map formed in Family C at LOD score 1.8.	126
4.31	Linkage map formed in Family C at LOD score 2.0.	128



ABBREVIATIONS

α	alpha
AFLP	Amplified Fragment Length Polymorphism
ATP	adenosine triphosphate
β	beta
bp	base pairs
Ci	Curie
dH ₂ O	distilled water
ddH ₂ O	deionized distilled water
dNTPs	deoxyribonucleotide
<i>g</i>	gravity
kb	kilobase
λ	Lambda
LOD	Logarithm of odds
LB	Luria Bertani
μ	micro
μ g	microgram
μ l	microliter
μ M	micromolar
MAS	Marker Assisted Selection
mM	millimolar
mtDNA	mitochondrial DNA
ng	nanogram



nmol	nanomole
OD	optical density
PCR	Polymerase Chain Reaction
pmol	picomole
QTL	Quantitative Trait Loci
RAHMs	Random Amplified Hybridizing Microsatellites
RAPD	Random Amplified Polymorphic DNA
rpm	revolution per minute
UV	ultra violet
V	volts



CHAPTER 1

INTRODUCTION

It has been estimated that the world's population will approach seven billion by the year 2010. Thus, the worldwide demands for food will also increase and will therefore, cause a shortfall of 19 million tonnes of fish and aquaculture productions (FAO, 1997). Consequently, the deficit in seafood production of the world is expected to increase.

One way of increasing the seafood supply is to increase marine fishing activities. However, the world's fish stocks are now struggling from overfishing. Almost two-thirds of marine fishes and other edible aquatic organisms from the Pacific and Atlantic Oceans had been fully exploited (reviewed by Liu and Dunham, 1998). The gap between the world's demand for fish and the ability of the ocean to provide sufficient food has widened. The solution to this problem lies in the development of sustainable aquaculture and hence, the introduction of more aquaculture species is an immediate necessity.

The river catfish, or Asian Red-tailed Catfish (*Mystus nemurus*; locally known as Ikan Baung) is of great potential importance as an alternative fish source to the world, especially to Southeast Asian countries. Its good flesh quality and taste, with high protein content but low in fat (Kamarudin *et al.*, 1987) makes it popular among consumers from this region. Therefore, this fish has been recognized as one of the favorite aquaculture species in Malaysia.



In some other countries, such as the United States, aquaculture products have become the third largest contributor to the trade balance of the country, and the channel catfish (*Ictalurus punctatus*) accounts for more than 50% of all the aquaculture production (Liu and Dunham, 1998). Similarly, *Mystus nemurus* has the potential to become the most important aquaculture product in this region, like the channel catfish is in the United States. However, most fish of this species available in the market currently comes from the various rivers of Malaysia. Their inability to reproduce on a large scale in captivity and the seasonal seed availability has led them to achieve high price in the market. Hence, commercialization and finding ways to increase production efficiency through genetic studies should be the immediate goal of the local aquaculture industry.

As traditional selective programs succeed, more effective selection programs need to be developed for traits that cannot be easily measured, such as disease resistance, growth, harvesting, oxygen tolerance, reproduction, and body mass. Therefore, emphasis should be on the genetic improvement of the species rather than on environmentally influenced characters. Marker Assisted Selection (MAS) and biotechnology offer great potential as alternatives to meet this challenge, but genetic linkage information of these traits must be obtained first. In other words, for the long term development and sustainability of aquaculture industry, it is important to construct a DNA marker-based genetic linkage map for application in selection programs through either the use of MAS, or isolation of economically important genes, in order to improve catfish broodstocks through biotechnology.

Genetic linkage mapping plays an important role for the purpose of understanding the organism studied. It reveals mechanisms of inheritance of phenotypes relevant to the markers. Correlation of recombination genome size with physical size of the genome makes it possible to estimate physical distances between markers. Comparative mapping among closely related species, particularly of catfish, will allow comparison of their genomic conservation and divergence, therefore, providing information on genomic evolution. The most important outcome of linkage mapping is that it reveals the physical distances among markers and the distances of markers to important traits which are useful in the isolation of the genes for the targeted traits.

The application of genetic markers in aquaculture research has increased dramatically in recent years. Several decades ago, protein level genetic markers such as isozymes were commonly used in the population characterization of fish, and identification of species or hybrids (Park and Moran, 1994). Later, with the discovery of various types of DNA level genetic markers, they have been widely accepted for use in various applications such as conservation, phylogenetic studies, construction of genetic linkage maps, studies of Quantitative Trait Loci (QTL) even in various breeding programs (Liu and Dunham, 1998). Markers such as Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLP), Random Amplified Polymorphic DNA (RAPD), Expressed Sequence Tagged sites (EST) and microsatellites have always been the favourite types of DNA markers in aquaculture research.

Many aquaculture species such as tilapia (Kocher *et al.*, 1998), rainbow trout (Young *et al.*, 1998), kuruma prawns (Li *et al.*, 2002), tiger shrimp (Wilson *et al.*, 2002),

