

**ISOLATION AND CHARACTERIZATION OF
MICROSATELLITE LOCI FROM WILD STRIPED
SNAKEHEAD (*CHANNA STRIATA*)**

YASMEEN SAMEER MAHMOUD SALAH

**DISSERTATION SUBMITTED IN FULFILMENT OF
REQUIREMENTS FOR DEGREE OF MASTER OF
BIOTECHNOLOGY**

INSTITUTE OF GRADUATE STUDIES

UNIVERSITY OF MALAYA

KUALA LUMPUR

2010

ABSTRACT

Channa striata (better known as Haruan in Malaysia), is a freshwater species that is well known as a protein source and also for its pharmaceutical value among wounded patients. Unfortunately stocks have declined over the past few years, thus an in-depth study is needed to rule out genetic irregularities that might be causing the slow extinction of the species.

This study was conducted for the purpose of studying the population diversity and genetic structure of *Channa striata* (striped snakehead) in seven different locations around the peninsular of Malaysia.

Microsatellite markers were isolated and identified using Random Amplified Microsatellite (RAM) primers. As a result, seven microsatellite primers were proven to be polymorphic and thereafter tested on 30 individuals from Johore, Kedah, Negeri Sembilan, Pahang, Penang, Selangor and Terengganu.

The primers chosen were not highly variable. As a matter of fact, most of the primers had 2-5 alleles, except for one that had 17. Observed heterozygosities ranged from 0.4341 to 0.7931. Expected heterozygosities however ranged from 0.4332 to 0.8635. There is no indication of inbreeding among individuals of each population, yet there is reasonable variation among the seven populations.

Clustering did not adhere to the geographical allocations, just as would have been expected initially. Negeri Sembilan branched off as an outgroup with not much similarities between the sampled individuals from other regions. Other unusual hybridizations and clusterings were observed; most probably due to human intervention and changes made to the water sources.

ACKNOWLEDGEMENTS

In the words of Werner von Braun a rocket physicist, astronautics engineer and space architect: “Research is what I'm doing when I don't know what I'm doing”. This stood true for me in the initial stages of my work, but as all of us, I had to learn and accomplish my objectives in a brief period of time.

First and foremost I would like to thank God for keeping me sane after being away from home for close to six years. As the times changed, he was my constant companion through it all.

To my family and especially my mom, this is probably more for you than it is for me. As I think she had hoped to see me fulfill my goals with more anticipation than I could ever have.

My gratitude goes out to my supervisor, Dr. Subha Bhassu, for believing in me even though I had no prior practical experience whatsoever. Also for pushing me and my colleagues constantly to finish up our dissertation, when we got too tired to do so.

An enormous thank you to Bharani Manoharan, who showed me the ropes around the lab, in the very short 4 months that I knew him. I owe my work to you and even though this is not much, I hope I can take whatever you taught me to another place and time, and achieve higher aspirations. I have never seen someone as patient and kind as you are. Hope your virtues take you places you belong and until we meet again...

I would also like to express my appreciation to CEBAR for allowing us to use their lab equipment from PCRs to the Genetic Analyzer. A special thanks to Teh Ser Huy, who went out of her way to make sure we knew what we are doing. Bless you.

Last but not least my lab mates and colleagues. Lab life would be no life without you guys. Nurriza Ab. Latif and Thamayanthi Nada Raja, it was good while it lasted and I wish you all the best with your work and further studies. Peter my savior, couldn't have done it without you and in your words: "God bless you my child".

TABLE OF CONTENTS

Contents	Page
ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	x
LIST OF ABBREVIATIONS	xi
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	6
2.1 Genomics	6
2.1.1 History overview	6
2.1.2 History of Population Genetics and Evolution	7
2.1.3 Applications of Genomics	8
2.1.4 Molecular Markers	9
2.2 Striped Snakehead (<i>Channa striata</i>)	35
2.2.1 Physical characteristics	36
2.2.2 Genetic Content	37
2.2.3 Geographical distribution	37
2.2.4 Ecological habitat	39
2.2.5 Reproduction	40
2.2.6 Breathing and organs associated	43
2.2.7 Skin	43
2.2.8 Diet	44

2.2.9 Dietary, pharmaceutical utilization and more	46
2.2.10 Infections and diseases	47
2.2.11 Wild and cultured	48
CHAPTER 3: METHODOLOGY	49
3.1 Samples	49
3.2 DNA extraction	51
3.3 Amplification of microsatellite regions using Random Amplified Polymorphic Primers (RAMs)	52
3.3.1 Primers	52
3.3.2 Polymerase chain reaction	53
3.3.3 Electrophoresis	53
3.4 Cloning	54
3.5 Colony PCR	55
3.6 Plasmid extraction	56
3.7 Sequencing	57
3.8 Sequence submission to GeneBank	57
3.9 BLAST analysis of microsatellite marker	58
3.10 Designing the microsatellite primers	58
3.11 Microsatellite primer optimization and amplification	59
3.12 Microsatellite analysis using Genetic Analyzer	60
3.13 Data analysis	61
CHAPTER 4: Results and Discussion	66
4.1 DNA extraction	66
4.2 RAMs primers used to construct microsatellite library	66
4.3 Cloning, sequencing & BLAST upload	68

4.4 Microsatellite primers	74
4.5 Polymorphism (PCR product and fragment analyzer)	75
4.6 Data analysis	85
4.6.1 Microsatellite primers	85
4.6.2 Genetic variation	86
4.6.3 Heterozygosity and inbreeding	89
4.6.4 Hardy-Weinberg Equilibrium (HWE)	92
4.6.5 Linkage disequilibrium	95
4.6.6. Genetic distances	97
4.6.7 Genetic differentiation	103
4.6.8 AMOVA	106
4.6.9 Population structure	107
CHAPTER 5: CONCLUSION	111
BIBLIOGRAPHY	117
APPENDIX A	154
APPENDIX B	166
APPENDIX C	173
APPENDIX D	175
APPENDIX E	177

LIST OF FIGURES

Figure	Page
Figure 1. Increase in the number of basepairs sequenced per 1\$US	32
Figure 2. Fishery genetics publications over the decades	33
Figure 3. Images of <i>Channa striata</i>	35
Figure 4. Taxonomy of <i>Channna striata</i>	36
Figure 5. Distribution of <i>C. striata</i> across the globe.	38
Figure 6. Distribution of <i>C. striata</i> in peninsula Malaysia.	39
Figure 7. Fertilized egg of <i>C. striata</i>	42
Figure 8. <i>C. striata</i> fry	42
Figure 9. Pie chart of <i>C. striata</i> diet in Bukit Merah Reservoir, Malaysia	45
Figure 10. Map of Malaysia showing regions of sample collection.	50
Figure 11. yT&A cloning vector	54
Figure 12. Optimization of RAMs primer BP14.	67
Figure 13. Gel image of plasmid extraction of RAM primer product T79113-7.	70
Figure 14. Pie chart showing a statistical overview of functional genes.	71
Figure 15. Range of repeats found in <i>C. striata</i> .	74
Figure 16. Polymorphism test of <i>C. striata</i> DNA using microsatellite primer BP6-2.	75
Figure 17. Electropherogram of an individual (ind. 1) from Johore using primer BP6-2.	76
Figure 18. Electropherogram of different a individual (ind. 2) from Johore using primer BP6-2.	76

Figure 19. Primer T79113-4 optimization attempt.	77
Figure 20. Gel image showing amplification of microsatellite primer T79113-17.	78
Figure 21. Amplification of T79113-7, showing false signs of polymorphism.	80
Figure 22. Fragment analyzer electropherogram of a <i>C. striata</i> individual from Kedah PCR product of microsatellite primer T79113-7.	81
Figure 23. Stutter bands shown in electropherogram of <i>C. striata</i> sample from Penang using microsatellite BP6-4.	82
Figure 24. Majority rule consensus tree (UPGMA) resulting from a summarization of 1000 trees.	101
Figure 25. UPGMA tree constructed using GDA software based on Nei's unbiased Genetic distance.	102
Figure 26. STRUCTURE image of the 7 microsatellites under study.	108

LIST OF TABLES

Table	Page
Table 1. Comparison of different markers	34
Table 2. Samples of <i>Channa striata</i> collected from 7 locations around peninsular Malaysia.	51
Table 3. List of RAMs primers optimized.	67
Table 4. Table showing functions (or predicted functions) of some of the RAMs primer products after running them through BLAST	72
Table 5. Polymorphic loci for each population.	84
Table 6. Null allele estimations obtained from MICRO-CHECKER.	85
Table 7. Characteristics of microsatellite primers isolated from <i>C. striata</i> .	87
Table 8. Numbers of alleles observed in each sample location and loci.	88
Table 9. Absolute values of inbreeding coefficient F_{is} .	90
Table 10. Overall population F_{is} .	91
Table 11. Probability values of HWE for each locations under study.	94
Table 12. Nei's Unbiased Measures of Genetic Identity and Genetic distance using software POPGENE	99
Table 13. Nei's Unbiased Measures of Genetic Identity and Genetic distance using software GDA v. 1.0	100
Table 14. Pairwise F_{st} values according to ARLEQUIN	105
Table 15. Molecular variance results obtained from AMOVA	106

LIST OF ABBREVIATIONS

μg	microgram
μl	microlitre
pmol	picamole
$^{\circ}\text{C}$	degree celsius
10X	ten times
1X	one time
AFLP	Amplified fragment length polymorphism
AMOVA	Analyses of molecular variance
ATP	Adenosine triphosphate
BLAST	Basic local alignment search tool
bp(s)	Basepair(s)
c	centi
cDNA	complementary DNA
D	Genetic distance
D-loop	Displacement loop
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxyribonucleotide triphosphate
dTTP	deoxythymidine triphosphate
ddH ₂ O	Double distilled water
D _N	Genetic distance
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
<i>E.coli</i>	<i>Escherichia coli</i>
EST	Expressed Sequence Tags
F _{is}	Inbreeding Coefficient
F _{st}	Fixation index
g(s)	Gram(s)
h or hr	hour
H _e	Expected heterozygosity
H _o	Observed heterozygosity
HWE	Hardy Weinberg equilibrium
IAM	infinite allele mutation model
K	Kilo
LB agar	Luria Bertani agar
LD	Linkage disequilibrium
m	meters
M	molar
MgCl ₂	Magnesium chloride
mRNA	Messenger RNA

mtDNA	Mitochondrial DNA
NCBI	National Center for Biotechnology Information
n	nano
n_a	Observed number of alleles
n_e	Effective number of alleles
OD	Optical density
p	Pico
PCR	Polymerase Chain Reaction
PHYLIP	Phylogeny Inference Package
PIC	Polymorphic information content
QTL	Quantitative trait Locus
QTN	Quantitative trait Nucleotide
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic Acid
rRNA	Ribosomal RNA
T_A	Annealing temperature
TBE	Tris-borate-EDTA buffer
tRNA	Transfer RNA
rpm	Rotations per minute
RRS	reduced representation shotgun
s	second
SMM	Stepwise mutation model
SNP	Single nucleotide polymorphism
SSM	Slipped strand mispairing
SSR	Simple sequence repeat
Std. Dev.	Standard deviation
UPGMA	Unweighted pair group method with arithmetic averaging
USA	United States of America
UV	Ultra violet
v	volt
VNTR	Variable number tandem repeats
X-gal	5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside