

**MINING AND VALIDATION OF
EST-MICROSATELLITES IN
FRESHWATER PRAWNS, *Macrobrachium rosenbergii***

SHAIRAH ABDUL RAZAK

**DISSERTATION SUBMITTED IN FULFILMENT OF THE
REQUIREMENT FOR THE DEGREE OF
MASTER OF BIOTECHNOLOGY**

**INSTITUTE OF BIOLOGICAL SCIENCE
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR**

2011

ABSTRACT

The current study conducted in purpose to illustrate the utility of EST-derived SSR in characterizing wild populations of *M. rosenbergii* in Malaysia's rivers. A novel set of EST-SSR generated from RNA of *M. rosenbergii* were validated in a full panel of 120 samples from four wild populations through Polymerase Chain Reaction (PCR). Seven EST-SSR loci were identified, characterized, and evaluated on 30 individuals each from the populations namely Negeri Sembilan, Kedah, Sarawak, and Terengganu. The average polymorphic informative content value (PIC) for these seven primers was found to be 0.6208 indicating considerable degree of polymorphism with number of alleles detected ranged from 4 to 10. The observed heterozygosity value count during multi-population analyses ranged from 0.5333 to 0.8333, whilst the expected ranged from 0.6288 to 0.7009. There was no linkage disequilibrium (LD) observed between all pairs of EST-SSRs loci. All loci, except for EST MR8 conformed to the Hardy-Weinberg equilibrium (HWE) suggesting factors violating the neutral expectation such as selection might have caused the locus to deviate from equilibrium. The F_{IS} index demonstrated no indication of inbreeding among individuals of each population. There was evidence that all populations assessed in this study are drawn from a single, large panmictic population; no genetic heterogeneity observed in population structure analysis, estimate of fixation index value in pairwise comparisons among the four localities revealed very low magnitude of differentiation (F_{ST} ranged between 0.00888 to the highest of 0.10644).

The results indicated that these polymorphic EST-SSR derived from *M. rosenbergii* would be useful for population genetic structure analysis and genetic diversity assessment in prawn populations as part of management policies of natural resources to ensure sustainability of wild broodstock for future development of prawn culture industries.

ABSTRAK

Kajian ini telah dijalankan dengan tujuan untuk memperlihatkan kegunaan EST-terbitan dari SSR dalam pencirian populasi liar *M. rosenbergii* di sungai-sungai di Malaysia. Set penanda EST-SSR baru telah dihasilkan dari RNA *M. rosenbergii*, dan 5 penanda ini seterusnya ditentukan dengan 120 sampel daripada empat populasi semulajadi melalui Tindakbalas Berantai Polimerase (PCR). Tujuh lokus EST-SSR loci telah dikenalpasti, dikelaskan, serta diuji ke atas setiap 30 individu daripada populasi Negeri Sembilan, Kedah, Sarawak, dan Terengganu. Purata nilai Kandungan Polimorfik Informatif (PIC) bagi kesemua tujuh primer ialah 0.6208, menunjukkan tahap polimorfisme yang tinggi, dengan bilangan alel yang dikesan berjulat antara 4 hingga 10. Nilai heterozigotiti yang diperoleh dalam analisis multi-populasi berjulat antara 0.5333 hingga 0.8333, sementara nilai dijangka berjulat antara 0.6288 hingga 0.7009. Tiada ketidakseimbangan rangkaian (LD) ditemui antara kesemua lokus EST-SSR. Semua lokus, kecuali EST MR8 memenuhi keseimbangan Hardy-Weinberg (HWE), dan ketidakseimbangan ini menunjukkan faktor melanggar jangkakan neutral seperti pemilihan mungkin mengakibatkan lokus tersebut melencong dari keseimbangan. Indeks F_{IS} menunjukkan tiada tanda berlakunya kacuk dalam antara individu bagi setiap populasi. Kesemua populasi yang telah dinilai dalam kajian ini berkemungkinan berasal daripada satu populasi besar yang membiak secara rawak; tiada keheterogenan dikesan dalam analisis struktur populasi, dan anggaran nilai indeks penetapan dalam bandingan-secara-pasangan antara keempat-empat lokasi mencatatkan magnitud perbezaan yang sangat rendah (F_{ST} berjulat antara 0.00888 sehingga 0.10644 bagi yang tertinggi).

Hasil kajian ini menunjukkan EST-SSR yang polimorfik yang didapati dari *M. rosenbergii* ini berguna untuk analisis struktur genetik serta penilaian kepelbagaian genetik bagi populasi udang, sebagai sebahagian dari dasar pengurusan sumber semulajadi bagi memastikan kelestarian benih udang liar untuk pembangunan industri kultur udang pada masa hadapan.

ACKNOWLEDGEMENTS

I thank Allah for giving me the strength to accomplish this task and being my constant companion to went through these times.

Special gratitude to my supervisor, Dr Subha Bhassu, for giving me a chance to fly further & the opportunity to join her lab community also entrusted me to finish part of her project. I wish you the best in times to come.

To Ellie, you are my savior. To Izzah, you are like my guardian angel and to Yasmin, special thanks for being the best Mentor ever, and Jothi my comrade, we finally did it! Not to be forgotten, the whole lab members for being there to support me all this while.

I am also grateful and thankful to my very supportive husband, parents, family, classmates and colleagues for their endless support and their company through thick & thin during my journey. There's more to come in the future, but there will be none if it's not for what I learned here today.

Thank You All

Shairah Abdul Razak

TABLE OF CONTENTS

PREFACE

Title page	i
Abstract	ii
Abstrak	iv
Acknowledgement	vi
Table of Contents	vii
List of Figures	xi
List of Tables	xii
List of Symbols and Abbreviations	xiii
List of Appendices	xv
1.0 INTRODUCTION	1
2.0 LITERATURE REVIEW	7
2.1 Population genetics	7
2.1.1 Molecular marker	7
2.1.2 Microsatellites (Simple sequence repeat, SSR)	8
2.1.3 Expressed Sequence Tags-derived Microsatellites (EST-SSR)	9

2.2 <i>Macrobrachium rosenbergii</i>	13
2.2.1 Nomenclature and taxonomy	13
2.2.2 The distribution and habitat description	16
2.2.3 Morphological and biological characteristics	18
2.2.4 Life cycle	22
2.3 The significance of studying genetic diversity of <i>Macrobrachium rosenbergii</i> 's wild population using EST-SSRs	23
3.0 METHODOLOGY	28
3.1 Materials	28
3.2 Methods	30
3.2.1 Detection of EST-microsatellite markers and primer design	30
3.2.2 Overview: validation of microsatellite loci	31
3.2.3 PCR conditions and gel electrophoresis	32
3.2.4 Fragment Analysis	33
3.2.5 Data Analysis	35
3.2.5.1 Identification and checking for scoring errors	35
3.2.5.2 Test for conformation to Equilibrium Expectations	36

3.2.5.3 Estimating Genetic Diversity	38
3.2.5.4 Measuring Sub-population differentiation	41
3.2.5.5 Inferring Population Structure	45
4.0 RESULTS	46
4.1 DNA Extraction	46
4.2 Microsatellite primers and preliminary polymorphism	46
Testing	
4.3 Determination of microsatellite allele sizes	51
4.4 Statistical data analysis	54
4.4.1 Error checking	54
4.4.2 Hardy-Weinberg Equilibrium and Linkage	54
Disequilibrium	
4.4.3 Characterization of EST-SSR loci isolated from	56
<i>M. rosenbergii</i>	
4.4.3.1 Polymorphic Information Content (PIC) and	56
Genetic variability	
4.4.3.2 Heterozygosity	57
4.4.4 EST-SSR loci for characterizing populations genetics	59
of four wild populations	

4.4.4.1 Genetic Diversity	59
4.4.4.2 Heterozygosity and Inbreeding	61
4.4.4.3 Genetic Differentiation	63
4.4.4.4 Population Structure	67
5.0 DISCUSSION	68
5.1 Microsatellite loci and preliminary polymorphism testing	68
5.2 Conformity to Neutral Expectations	72
5.3 Characterization of EST-SSR loci isolated from <i>M. rosenbergii</i>	74
5.4 Population Genetic Structure among populations from wild locations	75
6.0 CONCLUSION	80
REFERENCES	83
APPENDICES	91

LIST OF FIGURES

FIGURE	PAGE
Figure 2.1: Taxonomy of giant river prawn, <i>Macrobrachium rosenbergii</i>	13
Figure 2.2: Natural distribution of <i>M. rosenbergii</i>	17
Figure 2.3: Images of <i>Macrobrachium rosenbergii</i>	19
Figure 2.4: Three distinct morphotypes of mature male of <i>M. rosenbergii</i>	21
Figure 2.5: Life cycle of <i>M. rosenbergii</i>	22
Figure 3.1: The sampling locations of <i>M. rosenbergii</i> individuals	29
Figure 4.1: Gel images of DNA amplification of EST-SSR primers	48, 49
Figure 4.2: Pie chart showing distribution of 22 polymorphic EST-SSR primers according to functional classification	50
Figure 4.3: Electropherogram images for EST MR19, EST MR14	52, 53
Figure 4.4: UPGMA Phenogram	66
Figure 4.5: Graph showing value of $\ln P(D)$ (likelihood probability of $X K$) estimated for all number of inferred populations using Structure Ver 2.2	67

LIST OF TABLES

TABLE	PAGE
Table 3.1: Details on prawn individuals (genotypes) and sampling sites for assessment of genetic diversity	28
Table 3.2: List of primers with optimized annealing temperature (°C)	34
Table 4.1: Polymorphism screening for 30 successfully amplified EST-SSR loci	47
Table 4.2: Probability values of HWE for each locus per each studied populations	55
Table 4.3: Polymorphism assessment on microsatellites loci using four wild populations	56
Table 4.4: Summary of observed and expected heterozygosity (H_O and H_E) for each locus across four populations	58
Table 4.5: Summary of genetic diversity measures based on seven microsatellite loci in four populations	60
Table 4.6: Summary of observed and expected heterozygosity (H_O and H_E) for each locus across four populations	62
Table 4.7: Pairwise F_{ST} values between four populations	63
Table 4.8: Pairwise R_{ST} values between four populations	64
Table 4.9: Pairwise D_{est} values between four populations	65
Table 4.10: Values of correlation coefficient between matrices based on Mantel test	65

LIST OF SYMBOLS AND ABBREVIATIONS

A_e	effective number of alleles per locus
A_t	total number of alleles
BLAST	Basic local alignment search tool
bp(s)	Basepair(s)
cDNA	complementary DNA
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxyribonucleotide triphosphate
dTTP	deoxythymidine triphosphate
ddH ₂ O	Double distilled water
DNA	Deoxyribonucleic acid
DMSO	Dimethyl sulfoxide
EST	Expressed Sequence Tags
F_{IS}	Inbreeding Coefficient
F_{ST}	Fixation index
GFP	Giant Freshwater Prawn
H_E	Expected heterozygosity
H_O	Observed heterozygosity
HWE	Hardy Weinberg equilibrium
LD	Linkage disequilibrium
MAS	Marker-assisted selection
mRNA	Messenger RNA
N	effective sample size
NCBI	National Center for Biotechnology Information
OD	Optical density
PCR	Polymerase Chain Reaction

PIC	Polymorphic information content
QTL	Quantitative trait Locus
R_s	allelic richness
SSR	Simple sequence repeat
Std. Dev.	Standard deviation
T_A	Annealing temperature
UPGMA	Unweighed Pair-Group Method of Arithmetic Averages

LIST OF APPENDICES

APPENDICES

APPENDIX A Doyle & Doyle CTAB Procedure

APPENDIX B Characterization of EST-SSR primers according to its annotated protein classes and its functional classification based on its EST sequences

APPENDIX C Genotypic scoring for all individuals based on seven microsatellites loci

APPENDIX D Comparisons of linkage disequilibrium values for each locus pair combinations

APPENDIX E Genotypic frequencies of seven microsatellites loci for all four wild populations of *Macrobrachium rosenbergii*; Allelic frequencies of seven microsatellites loci for all four wild populations of *Macrobrachium rosenbergii*

APPENDIX F List of 22 polymorphic markers subjected for fragment analysis
