GENETIC VARIATION IN SEVERAL POPULATION OF MACROBRACHIUM ROSENBERGII DE MAN DEDUCED FROM SEQUENCING OF CYTOCHROME C OXIDASE I (COI) MITOCHONDRIAL DNA GENE.

Zuriani Binti Mat Nahi

Bachelor of Science with Honours (Resource Biotechnology) 2005
GENETIC VARIATION IN SEVERAL POPULATION OF *MACROBRACHIUM ROSENBERGII* DE MAN DEDUCED FROM SEQUENCING OF *CYTOCHROME C OXIDASE I* (COI) MITOCHONDRIAL DNA GENE.

ZURIANI BINTI MAT NAHI

This project is submitted in partial fulfillment of the requirements for the degree of Bachelor of Science with Honours (Resource Biotechnology)

Faculty of Resource Science and Technology
University Malaysia Sarawak
2005
# CONTENTS

<table>
<thead>
<tr>
<th>Content</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgement</td>
<td>iii</td>
</tr>
<tr>
<td>Abstract</td>
<td>1</td>
</tr>
<tr>
<td><strong>INTRODUCTION</strong></td>
<td></td>
</tr>
<tr>
<td>General Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Literature review</td>
<td></td>
</tr>
<tr>
<td><em>Macrobrachium rosenbergii</em></td>
<td>3</td>
</tr>
<tr>
<td>Polymerase Chain Reaction</td>
<td>6</td>
</tr>
<tr>
<td>DNA sequencing</td>
<td>6</td>
</tr>
<tr>
<td>Mitochondrial DNA</td>
<td>7</td>
</tr>
<tr>
<td><em>Cytochrome C Oxidase</em></td>
<td>8</td>
</tr>
<tr>
<td>Objectives</td>
<td>9</td>
</tr>
<tr>
<td><strong>MATERIAL AND METHODS</strong></td>
<td></td>
</tr>
<tr>
<td>Samples collection and preservation</td>
<td>10</td>
</tr>
<tr>
<td>DNA Extraction</td>
<td>11</td>
</tr>
<tr>
<td>Polymerase Chain Reaction</td>
<td>12</td>
</tr>
<tr>
<td>Purification</td>
<td>13</td>
</tr>
<tr>
<td>DNA sequencing</td>
<td>14</td>
</tr>
<tr>
<td>Sequence analysis</td>
<td>14</td>
</tr>
<tr>
<td>Flowchart of methodology</td>
<td>15</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

This project was performed in partial fulfillment for the Degree of Bachelor of Science. My sincere gratitude to the Faculty of Resource Science & Technology, UNIMAS and my supervisor, Mr. Yuzine Esa for the supportive, kindness, friendly, useful commitment in this research. A lot of thanks to all FRST’s lab assistance and UNIMAS for providing access facilities. Also thanks to Andy Kho, Jeffrine Rovie Ryan and to all my friends for their assistance. I would also like to thank members of my family for all the encouragement given.
Genetic Variation in several population of *Macrobrachium rosenbergii* de Man deduced from sequencing of *cytochrome c oxidase I* (COI) mitochondrial DNA gene

Zuriani binti Mat Nahi

Resource Biotechnology Programme
Faculty of Resource Science and Technology
Universiti Malaysia Sarawak

ABSTRACT

This study examines the genetic differentiation of *Macrobrachium rosenbergii* in several populations in Sarawak. The analyses were carried out using the 154 base pair *Cytochrome c oxidase I* (COI) mtDNA gene. The samples used in this study were collected from four sites in Sarawak; Sungai Samarahan, Batang Rajang, Batang Lupar and Sungai Roban. Phylogenetic analysis was done using two methods, Neighbor-joining and Maximum Parsimony. The result from phylogenetic analysis shows the existence of two subclusters of udang galah; one cluster consisted of samples from Sarikei, while another cluster consisted of samples from Sg.Roban, Btg.Lupar and Sg.Samarahan. However, further studies with higher number of samples from other population and longer sequences were needed to verify the results of this study. In general, this study indicated the potential of using COI mtDNA gene in detection of even small genetic variation, or synonymous substitution between samples.

Key words: *Macrobrachium rosenbergii*, genetic variation, *Cytochrome c oxidase I* (COI) mtDNA, DNA sequencing.

ABSTRAK

Kajian ini bertujuan untuk mendapatkan perbezaan genetik di antara *Macrobrachium rosenbergii* daripada beberapa populasi di Sarawak. Analisis genetik telah dibuat menggunakan kaedah penjujukan DNA pada 154 base pair gen mtDNA sitokrom c oksidase I (COI). Sampel yang digunakan dalam kajian ini diambil daripada empat kawasan di Sarawak iaitu Sungai Samarahan, Batang Rajang, Batang Lupar dan Sungai Roban. Analisis filogenetik dijalankan menggunakan dua kaedah iaitu Neighbour-joining dan Maximum Parsimony. Keputusan dari analisis filogenetik telah menunjukkan kehadiran dua kelompok kecil udang galah iaitu; satu kluuster mengandungi sampel dari Sarikei, sementara kluuster kedua mengandungi sampel dari Sg.Roban, Btg.Lupar dan Sg.Samarahan. Walaubagaimanapun, kajian lanjut perlu dilakukan dengan mengambil bilangan sampel yang lebih banyak daripada populasi-populasi lain untuk meningkatkan keyakinan terhadap keputusan yang diperoleh dari kajian ini. Secara amnya, kajian ini menunjukkan analisis jujukan DNA berpotensi dalam mengesan kehadiran variasi genetik yang kecil, atau substitusi sinonim bagi gen mtDNA sitokrom c oksidase I.

Kata kunci: *Macrobrachium rosenbergii*, variasi genetik, gen mtDNA Sitokrom oksidase I (COI), Penjujukan DNA.
GENERAL INTRODUCTION

The freshwater aquaculture industry has increased rapidly since the past ten years. This has concomitantly attributed to the pollution and over fishing affecting marine fisheries throughout the world. One of the freshwater organism commonly used for aquaculture's industry is the giant freshwater prawn (*Macrobrachium rosenbergii* de Man). The aquaculture of prawn has been well developed and they have greatly contributed to the socio-economic development of many countries by providing nutrition food, income and employment opportunities (Jayachandran, 2001).

Therefore, the study of genetic variation in organism such as 'udang galah' is important, particularly to develop suitable selective breeding program for genetic improvement that will increase the production efficiency, health management and product quality (James and Wetzel, 2001). Over the last three decades, many molecular techniques have been developed to analyze the genetic variability (genetic variation, population genetic and phylogenetics) in various animal taxa (Hoelzel, 1992). Among the techniques are Isozymes electrophoresis, DNA sequencing, Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), and microsatellites. The development of molecular techniques gives great opportunity to conduct studies on the genetics and diversity of freshwater aquaculture. DNA sequence can be the powerful tool for genetic characterization of population and species (Amos and Hoelzel, 1992). Essentially, mitochondrial DNA (mtDNA) possesses some characteristics that have made it attractive for studies of population structure, phylogenetic and conservation of endangered species (Avise, 1994). Thus this project
aimed to analyze the genetic variation in several population of *Macrobrachium rosenbergii* de Man using sequence analysis of COI mtDNA.

**LITERATURE REVIEW**

a) *Macrobrachium rosenbergii* (Udang Galah)

Giant freshwater prawns (*Macrobrachium rosenbergii* de Man) or locally known as ‘udang galah’ belong to the family Palaemonidae, it is related to crab and marine shrimp (Spott, 1981). The adult shrimp lives in freshwater environment such as big river and can also be found in brackish water. However, not in high salinity environment like sea-water (Kurian and Sebastian, 1976). *M. rosenbergii* has wide geographic distribution in the tropical and subtropical regions of the Indo-Pacific and is categorized as an important food source (Kurian and Sebastian, 1976). This species eats plant and animal materials such as pieces of fruits mollusc, small crustacean and phytoplankton. It becomes cannibalistic whenever extreme starvation (Ling and Merican, 1961).

The life of *M. rosenbergii* starts when fertilized egg attached to pleopods or appendage of the mother (James and Wetzel, 2001). Hatching (larvae) takes place in the brackish water of an estuary the mother entered after migrating downstream from a freshwater stream or river (Jayachandran, 2001). Larva prawns require brackish water during their early development but can consider lowering salinities environment as they mature (James and Wetzel, 2001). Transformation from larval prawn to postlarval prawn will increase the size and give tendency for prawn to move towards the freshwater. Prawns growth continued in freshwater through juvenile stage to adult (Kurian and
Sebastian, 1976). This prawn could reach a maximum size of 320mm (Kurian & Sebastian, 1976).

Previous researches of *M. rosenbergii* were done by Esa (1996) whereby he analyzed genetic polymorphisms of local population of udang galah (*M. rosenbergii*) in Malaysia, while Hedgecock *et.al* (1997) analyzed the genetic divergence and biogeography of natural population of *M. rosenbergii*. Ryan (2002) studied the genetic variation of *M. rosenbergii* using *Cytochrome c oxidase II (COII)* mtDNA gene from west coast of Sabah, Malaysia.
**Model system**

![Sample of *Macrobrachium rosenbergii*](image)

**Figure 1**: Sample of *Macrobrachium rosenbergii*

In this project, the prawn species used is *Macrobrachium rosenbergii*.

- **Family**: Palaemonidae
- **Genus**: *Macrobrachium*
- **Species**: *Macrobrachium rosenbergii*
- **Common Name**: Giant Fresh Water Prawn, Udang Galah
b) **Polymerase Chain Reaction (PCR)**

PCR is in vitro system for DNA amplification that employs the essential enzyme of cellular DNA replication, DNA polymerase, to selectively amplify a 'target' DNA region (Fox *et al.*, 1991). The key to this system is a pair of oligonucleotide primers which is single stranded DNA sequences of 20-30 nucleotides that serve as points of attachment for the polymerase (Kolmodin and Birch, 2002). The primer will bracket the region to be amplified: one primer is complementary to a sequence at the beginning of the target region, and the second is complementary to a sequence at the end of the target region on the anti parallel DNA strand. This PCR process requires a repetitive series of the three fundamental steps that defines one PCR cycle: double-stranded DNA template denaturation, annealing of two oligonucleotide primers to the single stranded template, and enzymatic extension of the primers to produce copies that can serve as templates in subsequent cycles (Fox *et al.*, 1991). As the cycles proceed, both the original template and the amplified targets serve as substrates for the denaturation, primer annealing and primer extension processes. Theoretically, every cycle doubles the amount of target copies. The advantages of PCR include it can allow one to detect (as opposed to characterize) the presence of particular gene sequences from extremely minute quantities of DNA (Fox *et al.*, 1991).

c) **DNA Sequencing**

DNA sequencing is the optimal method of population comparison both in terms of high resolution and of facilitations interpretation and used to determine the exact
order of bases in DNA (Hoelzel and Dover, 1991). The procedure uses DNA synthesis to produce copies of the target sequence much the same was as in PCR. However, instead of synthesizing multiple copies of the complete sequence, some of the copies are forced to terminate before reaching the end. In fact, the conditions are set such that a percentage of the copies ended at each of the base positions in the sequence. This chain termination method is accomplished by adding a small proportion of dideoxynucleotides (ddNTPs) to the standard DNA synthesis reaction, which contains deoxynucleotides (dNTPs). Dideoxynucleotides are chemically modified at one end so they can be added to a new chain, but nothing can be added to them. Thus, the replication process stops whenever one of these bases is added. The resulting copies, varying by only one base in length, are separated by electrophoresis on an agarose gel.

DNA sequencing is a powerful tool for characterization of population or genetic variation (Amos and Hoelzel, 1992). This tool also provides the greatest resolution for detecting genetic variation between individual and population, and also to elucidate the phylogenetics relationship between individual (Vitic and Strobeek, 1996). DNA sequence analysis has been used in study of Drosophila genome (Teresa and Thomas, 1996), and in analyses of mtDNA phylogeny of Gobiid Fishes, genus Tridentiger (Mukai et al., 1997).

d) **Mitochondrial DNA (MtDNA)**

Gene from mitochondrial genome are popular marker for population genetics studies and have been used in many researches (Meyer, 1994) on organism like as bats, crustacean (Teresa and Thomas, 1996) and human (Lutz et al., 1997; Pfeiffer et al.,
1997). In crustacean, studies on mtDNA have been done in *Artemia franciscana* (Velverde et al., 1994) and *Daphnia pulex* (Teresa & Thomas, 1996).

MtDNA possessed unique characteristics which make it attractive for studies of population structure, phylogenetic and conservation of endangered species (Avise, 1994). Some of these characteristics are 1) the mtDNA evolves at a rapid rate, approximately 2% per million years, and thus easily surveyed differences in mtDNA haplotypes within a species, 2) mitochondria are maternally inherited and are non-recombining, 3) only a single mtDNA genotype exist within an individual, 4) all animal have mitochondria which serve the same functions and have similar molecular characteristics (Whitmore, 1990; Hoelzel & Dover, 1991; Stepien & Kocher, 1997).

e) **Cytochrome C Oxidase I (COI)**

*Cytochrome c oxidase I (COI)* is the terminal catalyst in the mitochondrial respiratory chain and located in the inner membrane of mtDNA (Morlais & severson, 2002). This region of mtDNA is a useful marker for differentiating both the interspecific and intraspecific level of crustacean (Palumbi and Benzie, 1991). *COI* has been used to analyze phylogenetic relationship of Neopterygian fishes (Normark et al., 1991) and genetic structure of tautog (*Tautoga onitis*) population (Orbacz and Gaffney, 1996).
Objectives

Study of genetic variation is important for correct identification of parental stock and to develop suitable breeding program for genetic improvement. Besides that, information about genetic diversity still at the lower level for aquaculture especially for *M. rosenbergii* (in Malaysia). So, the objective of this study is to examine the levels of genetic diversity in several population of *Macrobrachium rosenbergii* from Sarawak using sequencing of cytochrome c oxidase I (COI) mtDNA gene.
Material and methods

1) Sample Collection and Preservation

Samples of *Macrobrachium rosenbergii* were collected from four locations in Sarawak (Sungai Samarahan, Sungai Rajang, Batang Lupar and Sungai Roban), shown in figure 2 and table 1. The samples were stored at -20°C fridge prior to DNA extraction.

![Sarawak Map](image)

**Figure 2**: Sampling location in four populations in Sarawak. P1: Sg. Samarahan, P2: Batang Rajang, Sarakei, P3: Btg. Lupar and P4: Sg. Roban.
Table 1: Collected samples and locations

<table>
<thead>
<tr>
<th>Population</th>
<th>Location</th>
<th>Number of individual (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population 1</td>
<td>MRSM (Sg. Samarahan, Kuching Sarawak)</td>
<td>4</td>
</tr>
<tr>
<td>Population 2</td>
<td>MRSR (Btg. Rajang, Sarikei, Sarawak)</td>
<td>9</td>
</tr>
<tr>
<td>Population 3</td>
<td>MRBL (Lingga, Btg. Lupar, Sarawak)</td>
<td>15</td>
</tr>
<tr>
<td>Population 4</td>
<td>MRR (Sungai Roban, Sarawak)</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>44</td>
</tr>
</tbody>
</table>

2) DNA extraction

Total DNA was extracted from muscle tissue using a modified CTAB method (Grewe et al., 1993). Tissue sample were transfer into 1.5 ml microcentrifuge tube containing 700μl CTAB. 5 μl Proteinase K was then added. The sample was incubated in water bath at 60°C until completely dissolved, which is around 3 hours. 600 μl Chloroform isoamyl alcohols (24:1) was subsequently added and mixed for 2 minutes before centrifugation at 13,000 rpm for 10 minutes.
The upper aqueous phase containing DNA was transferred to new tube, and an equal volume of absolute ethanol was added. The sample was centrifuged for about 10 minutes at 13,000 rpm. The supernatant was removed from tube and approximately 600μl cold 70% ETOH and 25μl 3M NaCl was added, followed by centrifugation at 13,000 rpm for 10 minutes. Finally, the pellet was air-dried before resuspended in 50μl of distilled water (ddH₂O).

3) Polymerase Chain Reaction (PCR)

All 44 samples of *Macrobrachium rosenbergii* were used for DNA amplification. A 550bp of the *Cytochrome c oxidase I* (COI) mitochondrial gene fragment were used for amplification. Thermal cycle amplification was performed in 50 μl reaction volume containing 34.74 μl sterilized water (ddH₂O), 5.0 μl 10X PCR buffer (Promega), 1.0 μl of dNTP, 3.0 μl MgCl₂, 0.26 μl *Taq* DNA polymerase (Promega), 2.0 μl DNA template and 2.0 μl of each primer. Details of *Cytochrome c oxidase I* (COI) primers used for this analysis is given below;

**COIf**: 5’-CTGCAGGAGGAGGAGGAYCC-3’ (forward)

**COIl**: 5’-CCAGAGATTAGAGGGAATCAGTG-3’ (reverse)

(Palumbi et al., 1991)

Cycle parameter were 2 min at 94 °C for initial denaturation, 1 min at 94°C for denaturation, 1 min at 50-54°C for annealing, 2 min at 72 °C for elongation and 10 min at 72 °C for final elongation. GeneRuler™ 100bp DNA Ladder was used as a standard size
marker. PCR product was visualized by using 1.0% agarose gel (containing ethidium bromide) for 45 min at 90V.

4) Purification

All PCR products were purified before sequence analysis. Purification was done by a gel excised method following the manufacturer's instruction (Fermentas) since PCR amplification produced multiple bands. The whole PCR reaction, (about 47μl) was resolved in agarose gel and the desired band was excised, followed by purification using the Fermentas purification kit. The excised gel was filled with binding solution and incubated for 5 minutes at 55°C to dissolve the agarose. 5 μl 'resuspended silica powder suspension' was added prior to a second incubation at 55°C. After 5 minutes the solution was centrifuged (quick spin) to form a pellet, and the supernatant was removed. The DNA pellet was subjected to a second wash using concentrated washing buffer. Then, the supernatant was discarded and a clean tissue was used to dry the pellet. Approximately 36μl of sterile distilled water was added and following by incubation for 2 minutes and then centrifugation for 2 minutes. Then, approximately 32μl of the liquid (purification product) was transferred to another labeled microcentrifuge tube and store in -20°C before being sent for sequence analysis. About 3μl of the purification products was checked using agarose gel electrophoresis to ensure the sufficiently recovery of the PCR product was obtained.
5) **DNA sequencing**

The cycle sequencing reaction was performed in a programmable thermal cycler (BiometraT-Personal). Cycle sequencing reaction was done for 25 cycles involving a denaturation process at 96°C for 10 sec, annealing process at 55°C and extension at 60°C for 4 min. The sequencing was performed on an ABI Prism® 377 automated DNA sequencer.

6) **Sequence Analysis**

CLUSTAL X (1.81) (Thompson *et al.*, 1997) software was used for multiple alignments of DNA sequences. CHROMAS software (version 1.45) was used to view and display DNA sequence result. Distance matrix was calculated using the Phylogenetic Analysis Using Parsimony (PAUP*) program version 4.0b10. Besides this, MEGA (Molecular Evolutionary Genetic Analysis software version 2.1 (Kumar *et al.*, 2001) was utilized for phylogenetic analysis using Neighbour-Joining (N-J) (Saitou and Nei, 1987) and maximum Parsimony (MP) method with bootstrap analysis of 1000 replication. DNA sequence Polymorphism (DNASP) version 3.53 (Rozas and Rozas, 2001) was used to investigate gene flow ($N_m$) and population structure ($F_s$) between populations.
Sample collection and preservation

DNA extraction

Polymerase Chain Reaction (PCR)

Purification of PCR product

DNA Sequencing

1) Population structure
   [Gene Flow (Nm),
   population structuring (Fst)]

2) Phylogenetic relationship among population

Figure 3: Flowchart of methodology
RESULTS AND DISCUSSION

DNA Extraction

Genomic DNA was extracted from all samples (54 samples) from Sungai Samarahan, Sungai Rajang, Batang Lupar and Sungai Roban. Successful extractions were achieved because most samples were fresh samples that were only preserved in -20°C freezer for a short period of time. Previous result (Kadri, 2003) showed that samples that were kept in freezer for storage produced better extraction result compared to samples that were preserved in ethanol. Figure 4 below showed an example of the extraction results.

![Genomic DNA](image.png)

**Figure 4:** DNA Extraction. Lanes 1, 3, 4 & 6 shows some of the presence of bright extracted DNA bands. Lanes 2 & 5 indicate unsuccessful extraction result. Lane 7 represents GeneRuler™ 100bp DNA Ladder (Fermentas) as a standard size marker.

All successful extractions showed the presence of high molecular weight DNA.

Two unsuccessful results (no band or smearing) are shown in lane 2 and lane 5.
Extraction result also shows the presence of smears band as indicated in Figure 4 (lanes 6). This smear might appear due to several reasons; 1) The presence of RNA in the extraction. 2) Contaminant from equipment which not proper sterile 3) Contaminant from the protein which not completely removed from nucleic acids. The presence of contamination during extraction protocol might contribute to these problems. All glassware, plastic-ware, buffers solution and bench surface are some of the potential sources of the contamination. In order to prevent contamination, it is important to make sure that all glassware, plastic-ware and buffers solution used in DNA isolation are autoclaved or sterilized. The bench surface where isolation is carried out should be cleaned with detergent and a set of clean instruments and glove should be used for each isolation.

A few samples of *Macrobrachium rosenbergii* did not completely dissolved after 3 hours at 60°C due to high amount of tissue used. A longer time was needed to ensure the remaining tissue was completely dissolved and the vortexing of the tube was done for every 15 minutes during incubation. Therefore the amount of tissue was reduced for the subsequent extraction procedures to ensure that the tissue was completely dissolved. Additionally for undissolved tissue after 3 hour of incubation, about 5 μl of Proteinase K was added into the mixed suspension. Proteinase K is one of the endopeptidase (protein-digesting enzyme) which catalyzes the cleavage of peptide protein within the cell. It enhances the degradation of protein into smaller fragment. Proteinase K works best at temperature between 40-60°C. It will denature if the incubation temperature is above 70°C (Grange et al., 1991).
Since there was no compatible size marker provided for the size estimate of isolated DNA, GeneRuler™ 100bp DNA Ladder was used as a standard size marker. A single and bright band was appeared above the range of the GeneRuler™ 100bp DNA Ladder.

**Polymerase Chain Reaction (PCR)**

Amplification of *Cytochrome c oxidase I* (forward & reverse) primers were done at annealing temperature range from 50°C to 54°C. All the amplification product failed to produce a single band. At the annealing temperatures between 50°C to 54°C, PCR amplification assay produce double bands (shown on Figure 5, lanes 3-5) with the extra band slightly higher than the expected PCR product. At annealing temperature below 50°C, result shows the presence of multiple bands (result not shown). No amplification product was seen using the annealing temperature of more than 54°C. However, amplification failures also occurred at the annealing temperature of 54°C and below (shown on lane 2, Figure 5).
Multiple bands or nonspecific product probably occurred due to several factors such as unsuitable amount of DNA template used. In this research, I used 2µl of DNA but still managed to obtain PCR product. I have tried using lesser amount of DNA template but this failed to produce amplified products. However, the uses of higher amount of DNA template usually yielded in multiple bands of PCR product. The concentration of MgCl$_2$ might also take place in problem of multiple bands. According to Kidd & Ruano (1995), higher concentration of MgCl$_2$ stabilizes double-stranded DNA and prevents complete denaturation of the product at each cycle but reducing the amplification yield. Additionally, high concentration of Taq DNA polymerase also