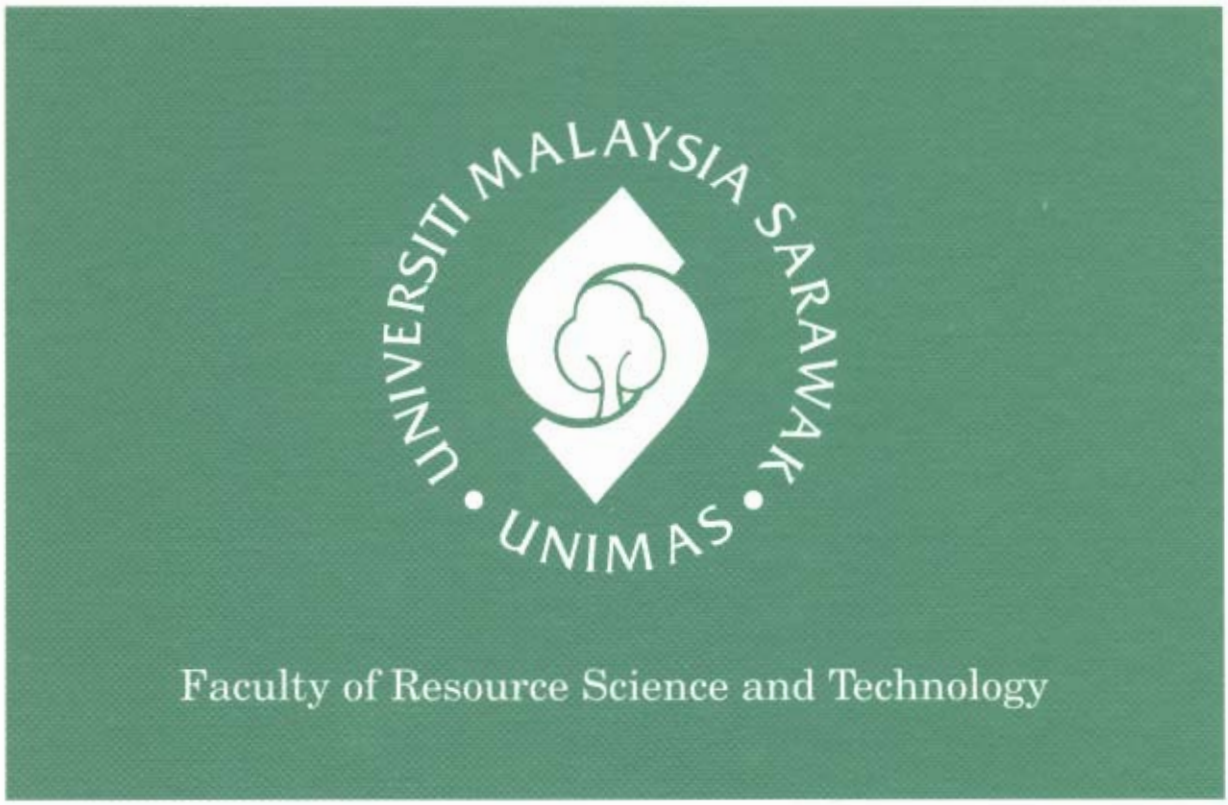

POPULATION GENETIC ANALYSIS OF *OREOCHROMIS NILOTICUS* FROM BALUNG RIVER, SABAH MALAYSIA USING SEQUENCING ANALYSIS OF *CYTOCHROME C OXIDASE* I (COI) MITOCHONDRIAL DNA FRAGMENT.

Sulinda Hassan

QH
441
S949
2004

Bachelor of Science With Honours
(Resource Biotechnology)
2004

Pusat Khidmat Maklumat Akademik
UNIVERSITI MALAYSIA SARAWAK
94300 Kota Samarahan

P.KHIDMAT MAKLUMAT AKADEMIK
UNIMAS
1000126464

Faculty of Resource Science and Technology

POPULATION GENETIC ANALYSIS OF *Oreochromis niloticus* FROM BALUNG RIVER, SABAH, MALAYSIA USING SEQUENCING ANALYSIS OF *CYTOCHROME C OXIDASE* I (COI) MITOCHONDRIAL DNA FRAGMENT.

SULINDA HASSAN

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

2004

POPULATION GENETIC ANALYSIS OF *Oreochromis niloticus* FROM BALUNG RIVER, SABAH, MALAYSIA USING SEQUENCING ANALYSIS OF *CYTOCHROME C OXIDASE* I (COI) MITOCHONDRIAL DNA FRAGMENT.

SULINDA HASSAN

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
2004

POPULATION GENETIC ANALYSIS OF *Oreochromis niloticus* FROM BALUNG RIVER, SABAH, MALAYSIA USING SEQUENCING ANALYSIS OF *CYTOCHROME C OXIDASE* I (CO1) MITOCHONDRIAL DNA FRAGMENT.

Sulinda binti Hassan

Biotechnology Resource Program
Faculty of Resource Science and Technology
University Malaysia Sarawak

ABSTRACT

A study on the genetic variation of *Oreochromis niloticus* (Nile tilapia) populations from Malaysia was done using partial sequencing of *cytochrome c oxidase* I (COI) mtDNA gene. Samples were collected from various sites along Balung River, Tawau, Sabah. Phylogenetic relationship inferred using neighbour-joining (NJ) and maximum parsimony (MP) methods generally divided samples into two major clades (Cluster A and Cluster B) with inconsistent distribution of the samples. One group consisted of 14 samples while other clades clustered seven samples of *O. niloticus* both from three different sites along Balung River, Sabah. Phylogenic analysis suggests that parental stocks of samples might be originated from two different mtDNA lineages, as reflected by both methods. Low distance values between haplotypes sequences (0.000 to 0.087) and low nucleotide diversity, P_i (0.000 to 0.082) indicated a very close genetic relationship between samples, enforcing their taxonomic status as belonging to a single taxon (*O. niloticus*). However, the high level of gene flow (N_m=10.50) and low population structuring (F_{st}=0.05) among samples was probably because the close distances between sites.

Key words: Genetic variation, *Oreochromis niloticus*, partial sequencing, *cytochrome C oxidase I* (COI), Phylogenetic relationship.

ABSTRAK

Kajian mengenai variasi genetik di dalam populasi Oreochromis niloticus (Tilapia Nil) dari Malaysia telah dijalankan dengan menggunakan penjujukan separa DNA gen mtDNA cytochrome c oxidase I (COI). Sampel diperolehi dari tiga lokasi di sepanjang Sungai Balung, Tawau, Sabah. Hubungan filognetik yang diperolehi dengan menggunakan keadah pokok filogenetik iaitu neighbour-joining (NJ) and maximum parsimony (MP) secara umum telah membahagikan sampel kepada dua kumpulan yang utama (Cluster A dan Cluster B) dengan taburan sampel yang tidak sekata. Kumpulan pertama mengandungi 14 sampel manakala kumpulan satu lagi mengandungi tujuh sampel O. niloticus, masing-masing dari tiga kawasan yang berlainan sepanjang Sungai Balung, Sabah. Analisis filogenetik menyarankan kemungkinan stok asal O. niloticus di Sungai Balung mungkin berasal daripada dua mtDNA 'lineages' yang berbeza seperti yang diperolehi daripadakedua-dua kaedah. Nilai jarak genetik yang rendah antara 'haplotype'(0.000 hingga 0.087) dan kepelbagaian nukleotida yang rendah, P_i (0.000 hingg 0.082) menunjukkan hubungan genetik yang rapat diantara sampel sekaligus menyokong sampel dalam status taxonomi yang sama (O. niloticus). Walau bagaimanapun, nilai aliran gen yang tinggi (N_m=10.50) dan struktur populasi yang rendah (F_{st}=0.05) diantara sample mungkin disebabkan oleh jarak diantara satu kawasan dengan kawasan yang lain yang agak berdekatan.

Kata kunci: Variasi genetik, Oreochromis niloticus, sebahagian penjujukan, cytochrome C oxidase I (COI), Hubungan filogenetik.

INTRODUCTION

Over three decades, molecular techniques have become available to screen large numbers of individual's animals for examining their genetic variability at the molecular level (Amos and Hoelzel, 1992). Molecular markers are used to measuring genetic diversity and differentiation in natural, managed and breeding populations, and estimating rates of gene flow or migration. Other than that, it is also can be used for studying phylogeny or taxonomy and genetic mapping (Glaubitz and Moran, 1997). Among the popular markers used for molecular studies includes isozymes electrophoresis, restriction fragment length polymorphisms (RFLPs), amplification fragment length polymorphisms (AFLPs) and DNA sequencing. DNA sequencing technique offers the most powerful tool in examining population structure but it was also the most expensive molecular methods (Amos and Hoelzel, 1992).

The greatest resolution of genetic variation is by DNA sequencing (Amos and Hoelzel, 1992). Apart from that, DNA sequencing offers the most optimal method for population comparison both in terms of high resolution and facilitating interpretation (Hoelzel and Green, 1992), thus provides informative data for studying population structure and phylogenetic analysis. With these characters, a combination of Polymerase Chain Reaction (PCR) method and DNA sequencing approaches enabled the relationships study among both close and distance relatives (Stepien and Kocher, 1997). For an instant, DNA sequencing analysis of amplified mitochondrial DNA (mtDNA) control region have been used in genetic studies of fish in the *Galaxias vulgaris* complex, a group of freshwater-limited galaxii fishes endemic to South Island, New Zealand (Waters and Wallis, 2000).

Besides, DNA sequence data can be used for phylogenetic reconstruction. Genetic distances and clustering algorithms emerged along with the development of molecular techniques, which describes the degree of similarity or genetic relatedness among pairs of taxa and summarize this information data in a "tree" construction (Stepien and Kocher, 1997). Tree construction using sequences from one species, *Oreochromis niloticus*, in this study, produced genetic divergence relationship, gene flow and population structuring.

Oreochromis niloticus (Nile tilapia) was chose to test the population genetic analysis along Balung River, Sabah. *O. niloticus* is an introduced species originated from Africa but has established themselves in Balung River, Sabah since their introduction. The introduction of *O. niloticus* into this country began as early as 1920s with initial stocks imported from Indonesia and Taiwan (Esa and Rahim, 2003). The main purpose for introduction was to replace cichlid, *Oreochromis mosambicus* which grow slower and not favored by locals. Since then, *O. niloticus* was extensively stocked into ponds and hatcheries throughout Malaysia by government (Fisheries Department) and private sectors (aquaculture purposes), until they being replaced by new strains of Tilapia hybrids. Nowadays, *O. niloticus* can be found in almost every water body, from brackish water to the fast flowing streams in the highland area (Esa and Rahim, 2003).

Prior to the important role not only as a food source but as a source of outcome in fish trading, some study need to conduct in order to conserve our local introduced species. According to Tan (1996), it is expected that freshwater aquaculture will take a major part in the fisheries industry, as environmental hazards and over-fishing have depleted the sources of marine fisheries worldwide. Freshwater aquaculture contributed about 80,000

metric tones to the total fish production in Malaysia and is rapidly developed. The rapid growth of culture techniques for freshwater fish such as pond-culture system, concrete-tank system and pond-cage system (Litis *et al.*, 1997) suggests that daily protein diet from freshwater fish is becoming increasingly important worldwide.

Molecular study on the establishment of *O. niloticus* populations in natural river systems in Malaysia has never been done, although that could possess negative impact (indirectly through ecological competitions) with indigenous species (Esa and Rahim, 2003). Therefore, the study was conducted at a first attempt to investigate the genetic diversity and population structure of *O. niloticus* population in natural system. Realizing the importance of our freshwater fish resource, a comprehensive study on their systematic, particularly among the important cichlids group, are essential for appropriate management and conservation of our existed freshwater fish diversity.

OBJECTIVES

To examine population structure of *O. niloticus* in Balung River using DNA sequencing of *cytochrome c oxidase* I (COI) mtDNA fragment.

Nile tilapia (Oreochromis niloticus)

Tilapia is the common name for almost 70 species of porch-like fishes (family *Cichlidae*) native to the fresh waters of tropical Africa (Trewavas, 1982; Stiassny, 1991). They include the mouthbrooding genera *Sclrotherodon and Oreochromis,* and substrate spawning Tilapia. According to Trewavas (1982) and Stiassny (1991), the Nile tilapia (*Oreochromis niloticus*) was one of the first fish species cultured (more than 3,000 years ago). *O. niloticus* has been introduced species into nearly every tropical and subtropical country in the whole world in order to boost the development of freshwater aquaculture. FAO Fisheries Statistics (1997) stated that these fishes are estimated almost 659,000 metric tones per year for world production which the majority consists of the species. Tilapias have been farmed during the last half-century throughout the tropical and semi-tropical country. Nowadays, all commercially important tilapia outside of Africa belong to the genus *Oreochromis,* and more than 90% of all commercially farmed tilapia outside of Africa are Nile Tilapia. Less commonly farmed species are Blue Tilapia (*Oreochromis aureus*), Mozambique Tilapia (*Oreochromis mossambicus*) and the Zanzibar Tilapia (*Oreochromis urolepis hornorum*).

Model system

Oreochromis niloticus also known as Nile tilapia. It is widely distributed in Malaysia and was found abundances in places like Serting River in Negeri Sembilan, Balung River in Tawau, Sabah and Ba'kelalan River in Sarawak.

FIGURE 1: Diagram of *Oreochromis niloticus* (Nile Tilapia)

Family : Cichlidae

Genus : *Oreochromis*

Species : *Oreochromis niloticus*

Local name : Nile tilapia

Physical Characteristics

O. niloticus has strong vertical bands. Mature *O. niloticus* has grey or pink pigmentation in the throat region. However, coloration is often an unreliable method of distinguishing tilapia species because environment, state of sexual maturity, and food source greatly influence color intensity (Popma and Masser, 1999).

Specialty

O. niloticus is a fish endemic to Africa. They are a popular culture and experimental animal because they grow fast, are efficient users of natural aquatic foods, will consume a variety of supplement feeds, they have an herbivorous nature, and are resistant to disease and handling (Popma and Masser, 1999). Popma and Masser (1999) also stated that when kept in captivity, tilapias reproduce easily and will tolerate a wide range of environmental condition. These attributes, along with relatively low input costs, have made tilapia the most widely cultured freshwater fish in tropical and subtropical countries.

MATERIALS AND METHODOLOGY

1) Sample collection and preservation

Twenty four samples of *O. niloticus* were collected from four different stations along Balung River, Tawau, and Sabah (Table 1). All samples were preserved in 70% ethanol and stored at -20° in freezer prior to DNA extraction and genetic analysis. The samples sites, coding number (abbreviations), GPS (Global Positioning System) references and number of individuals are shown in Table 2.

TABLE 1: Sample collection map of *Oreochromis niloticus* in Tawau, Sabah

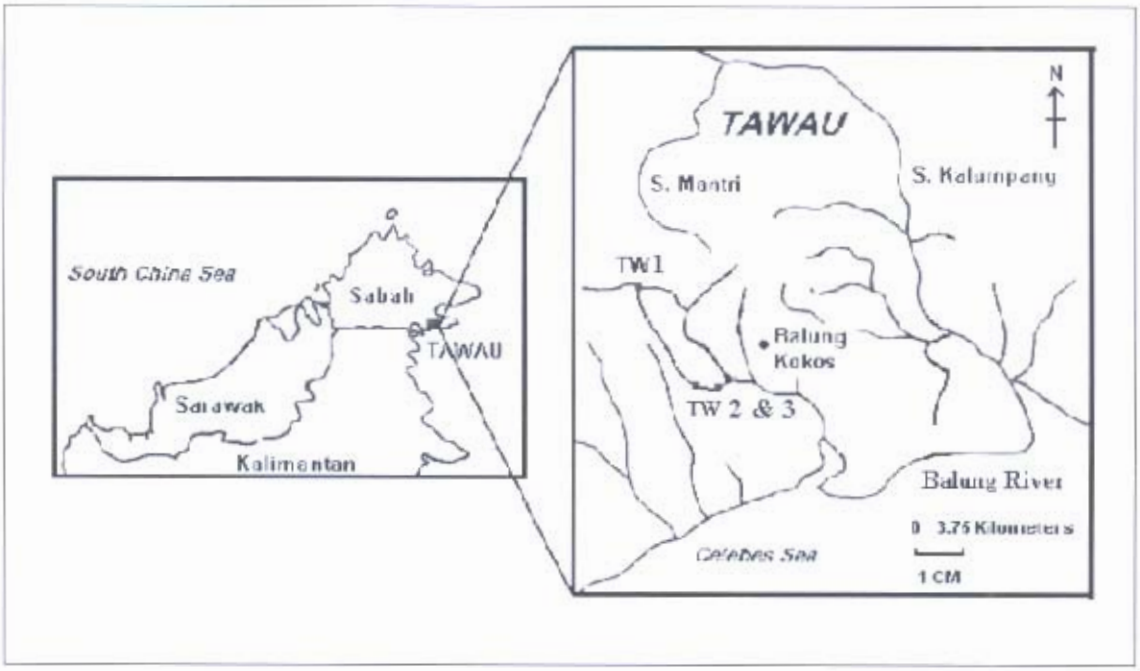

TABLE 2: Samples distribution of *Oreochromis niloticus* in Tawau, Sabah

Sites	Sample size	Coding number	GPS references
TW1	14	TW1(1)-TW1(14)	N 04° 27' 49.8'' E 118° 02' 58.3''
TW2	5	TW2(15)-TW2(19)	N 04° 22' 29.1'' E 118° 06' 20.6''
TW3	5	TW3(20)-TW3(24)	N 04° 22' 25.8'' E 118° 06' 29.2''

Total of samples: 24

TW - *Oreochromis niloticus* from Balung River, Tawau, Sabah.

TW1 - Hulu Sungai Balung I and II, Tawau

TW2 - Sungai Balung, Tawau (Kebun)

TW3 - Sungai Balung Tawau (Bawah Jambatan)

2) DNA Extraction

For extraction method, total genomic DNA for each samples of *O. niloticus* were extracted from muscle tissue using a modified CTAB protocol (Grewe *et al.*, 1993). About 0.2 cm^3 of tissue sample were dissected and resuspended in 700 µl CTAB with 10µl Proteinase K and digested at 65° C for 1 hour. The digested tissues were mixed with 600µl of chloroform-isoamyl alcohol (24:1) followed by centrifugation for 10 min at 13, 000 rpm using. Then, 600µl of supernatant (upper layer contains DNA) was taken into a new tube, precipitated with equal volume of 100% ethanol, and centrifuged again for 10 min at 13 000 rpm. The resulting pellet was washed with 70% ethanol and 25µl Natrium Acetate / Natrium Chloride and air-dried. Finally, dried DNA pellet was resuspended in 100µl of sterile distilled water. The extracted DNA was stored at -20°C until use for further analysis.

3) Polymerase Chain Reaction (PCR) and Purification

The Polymerase Chain Reaction (PCR) method was used to amplify a 550 base pair (bp) region of *cytochrome c oxidase* I gene (COI). The region of the COI gene was amplified using the following primers (Palumbi *et al.*, 1991):

Forward primer: COIf: 5'-CCTGCAGGAGGAGGAGAYCC-3'

Reverse primer: COIe: 5'-CCAGAGATTAGAGGGAATCAGTG-3'

PCR reaction was prepared in a laminar flow cabinet to avoid contamination. Thermal cycle amplification was performed in 50µl reaction volume. The content of reaction mixtures for PCR and their appropriate amount are described in Table 3:

TABLE 3: Polymerase Chain Reaction Mixtures contents

Component	Component volumes (µl)
Sterile ultra pure water	31.75
10X Reaction Buffer	5.00
$MgCl_2$ (25mM)	3.00
PCR Nucleotide Mix (10mM)	1.00
Upstream Primer (10µM)	2.50
Downstream Primer (10µM)	2.50
Taq DNA Polymerase	0.25
Template DNA	4.00
Final Volume	50

DNA amplification was performed in a programmable thermal cycler (Biometra T-personal). The PCR amplification was started by initial denaturation followed by 30 cycles involving denaturation step, annealing, primer extension and final elongation step. The cycle parameters are described as below:

A. Initial denaturation	95°C for 5 minute
B. Strand denaturation	95°C for 45 second
C. Annealing	47°C for 1 minute 30 second
D. Primer extension	72°C for 1 minute 30 second
E. Final elongation	72°C for 10 minute

PCR amplification products were then observed using 1.0% agarose gel electrophoresis containing Ethidium bromide. A mixture of 3µl of DNA PCR product and 0.6µl of Gel Loading Dye was loaded into each well. A 1 Kb Plus Ladder (Fermentas) was used as a standard size marker. The products were run for approximately 30 minute at 90V and then photographed under UV light.

PCR products were subsequently purified to clean them from extraneous salts, primer dimer, small fragment of DNA and other non-specific products prior to sequencing. PCR Purification Kit (Fermentas and Promega) were used for purifying PCR products.

4) DNA Sequencing

The best purified products were selected for sequencing according to the visibility of clear DNA band appeared during visualization. The purified products were sent in a 20-25µl volume in PCR cubical tubes. Sequencing reaction was constructed using the Big-Dye® Terminator v3.0 Cycle Sequencing Kit with thermal cycling performed in T-personal Combi Thermocycler. Sequencing reaction consisted of 35 cycles of 96°C (10 sec), 55°C (5 sec), 60°C (4 min), 4°C hold and proceed to Ethanol / Sodium Acetate precipitation and then run on an ABI PRISM ® 377 DNA Sequencer. Rapid thermal ramp is 1°C/sec.

7) DNA Sequence Analysis

Automated DNA sequence chromatogram results were checked and edited using CHROMAS (Version 1.45) program to display fluorescence-based DNA sequence analysis results. Multiple sequence alignment for forward reactions of *cytochrome c oxidase* I (COI) partial sequence of *O. niloticus* was done using CLUSTAL X program (Version 1.81; Thompson *et al.*, 1997), and subsequently aligned by eye. 460-480 nucleotide sites were utilized for this analysis. Molecular Evolutionary Genetic Analysis (MEGA) Version 2.1 (Kumar *et al.*, 2001) was used to reconstruct neighbour joining (NJ) tree (Saitou & Nei 1987) while Phylogenetic Inference Package (PHYLIP) Version 3.6a3 (Felsenstein, 2002) was used to reconstruct parsimony tree and converted it into MEGA format. Kimura 2-parameter distance (1980) was used, base on equal base frequencies and unequal ratio of transition to transversion (Ti:Tv). MEGA program was also utilized to examine other relevant analyses such as distance matrix. One sample of Biawan fish (*Helostoma temminckii*) from Bakong, Sarawak was included as an outgroup to root the phylogenetic trees inferred from the *cytochrome c oxidase* I (COI) sequence. Phylogenetic confidence was estimated by bootstrapping (Felsenstein, 1985) with 1000 replicate data sets. The level of genetic diversity between, within and among populations and geographical regions were quantified by pair-wise estimates of nucleotide diversity (Pi, Jukes and Cantor, 1969); level of gene flow (N_m) and F-statistic analysis (F_{st}) from Hudson *et al* (1992). The parameters were analyzed using DNA Sequence Polymorphism (DNAsp) Version 3.53 (Rozas and Rozas, 1999).

RESULTS

1) DNA Extraction

Twenty one samples of *Oreochromis niloticus* collected from Balung River, Sabah were successfully extracted and indicated positive results while only three samples showed negative results. Table 4 show the number of successfully extracted DNA samples from different population sites.

TABLE 4: Diagram show the total number of samples and the successfully extracted DNA for PCR

Sites	Number of Samples	Number of successfully extracted DNA	New coding number
TW1- Hulu Balung I	14	13	TW1 (1) -TW1(13)
TW2- Kebun	5	4	TW2(14)-TW2(17)
TW3- Bawah Jambatan	5	4	TW3(18)-TW3(21)
Total number of samples	24	21	21

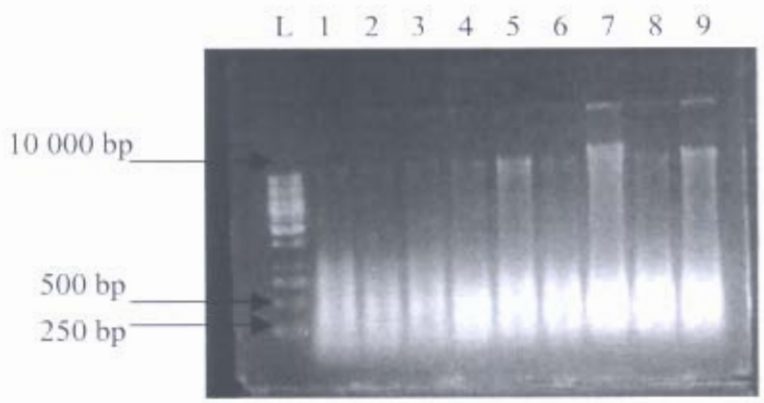

FIGURE 2: Polaroid picture of DNA extraction [L=1 kb Marker (Fermentas); Lane 1-9 represents the successfully extracted DNA which are TW1 (1) - TW1 (9)]. TW stands for sample sites of different station of *Oreochromis niloticus* from Balung River, Tawau, Sabah. The incubation temperature was 65°C.

2) Polymerase Chain Reaction (PCR) Amplification

Successful PCR amplification is shown by the appearance of a single bright band at about 550 bp in length. The bright band shows that the product is in large quantity while a dull band shows small amount of amplification product is small. The estimated size of the PCR amplified *cytochrome c oxidase* I (COI) mtDNA was about 550 bp. All 24 samples were successfully amplified.

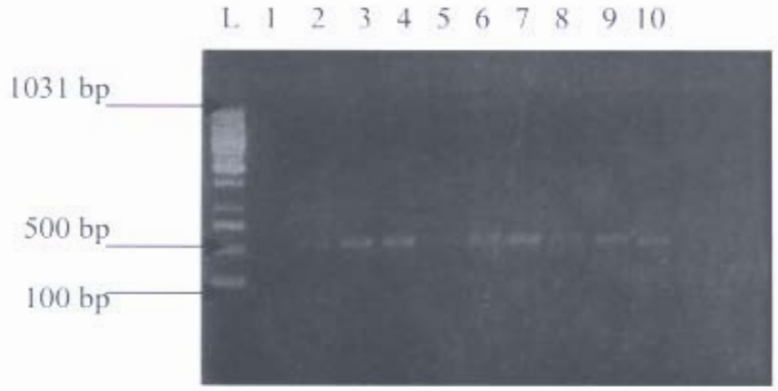

FIGURE 3: Polaroid picture of amplified PCR products [L=Mass Ruler™ DNA Ladder Low Range]. Lane 1 represents the Negative Control while Lane 2-10 shows the successfully amplified PCR products which are TW stands for sample sites of different stations of *Oreochromis niloticus* from Balung River, Tawau, Sabah.The annealing temperature used for amplification was 47°C.

3) Purification of PCR products

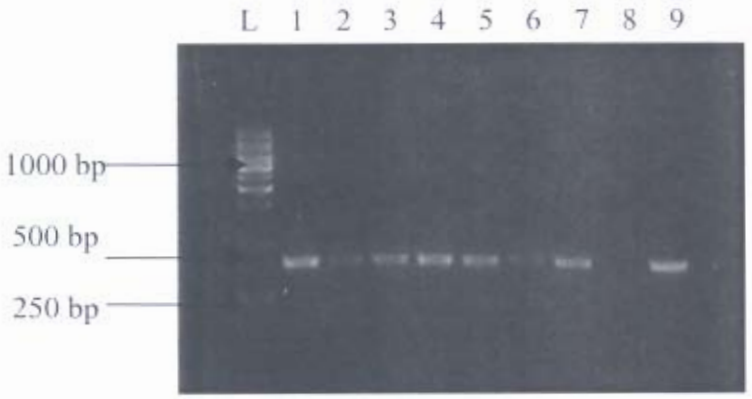

Figure 4: Polaroid picture of purified PCR products where L= 1 kb Marker; Lane 1 – 7 and Lane 9 shows the purified PCR products represent TW1(1) - TW1(7) and TW1(9). Sample coding (TW) stands for sample sites from four different sites of *Oreochromis niloticus* from Balung River,Tawau, Sabah.

4) Genetic Analysis and Population Structure

TABLE 5. F-statistic analysis of population structuring, F_{st} (below diagonal) and level of gene flow, N_m (above diagonal) between populations of *Oreochromis niloticus* from Balung River, Sabah.

	TW1	**TW2**	**TW3**
TW1	-	4.20	5.50
TW2	0.14	-	4.72
TW3	0.10	0.12	-

TABLE 6. Calculation of nucleotide diversity, P_i (JC) (Jukes & Cantor 1969) within and among populations and geographical regions of *Oreochromis niloticus* in Tawau, Sabah.

Study site	Nucleotide diversity, Pi (JC)
TW1	0.03
TW2	0.03
TW3	0.04
TW1 vs. TW2	0.03
TW1 vs. TW3	0.04
TW2 vs. TW3	0.03
Total	0.03

From the data obtained in Table 5, TW1 against TW3 exhibited the highest value of gene flow (N_m=5.50) thus the lowest F-statistics analysis of population structuring (F_{st}=0.10) while the lowest N_m value (4.20) obtained from TW2 against TW3 with the highest population structure value (0.14). The N_m values between samples sites were relatively high, with a range between, N_m=4.20 (TW1 vs. TW2) to N_m=5.50 (TW1 vs. TW3). Likewise, for F_{st} value, the value between samples sites were low, with a range between, F_{st}=0.10 (TW1 vs. TW3) to F_{st}=0.14 (TW1 vs. TW2). According to the data indicated in the Table 5, we can say that the gene flow, N_m values are relatively high with low population structuring value.

The genetic diversity within, between and among sites were generally low. All sites exhibited low within nucleotide diversity (TW1=0.04, TW2 and TW3=0.03) with a total P_i=0.03. Table 6, the nucleotide diversity (sequence divergence) values (P_i) between sites were low, ranging from P_i=0.03 (TW1 vs. TW2 and TW2 vs. TW3) to P_i=0.04 (TW1 vs. TW3). This probably showed little genetic variation among samples in all sites.

Pusat Khidmat Maklumat Akademik
UNIVERSITI MALAYSIA SARAWAK
94300 Kota Samarahan

TABLE 7: The pair-wise distance value (above diagonal) and nucleotide diversity (below diagonal). Both calculations incorporated Kimura 2-parameter distance (1980) method. Abbreviations refer to individuals in populations identified in Table 4.

	1 TW1(1)	2 TW1(2)	3 TW1(3)	4 TW1(4)	5 TW1(5)	6 TW1(6)	7 TW1(7)	8 TW1(8)	19 TW1(9)	10 TW1(10)
1	-	0.064	0.024	0.015	0.082	0.017	0.013	0.071	0.017	0.028
2	0.060	-	0.083	0.073	0.013	0.077	0.073	0.013	0.075	0.087
3	0.024	0.078	-	0.020	0.073	0.026	0.017	0.069	0.013	0.000
4	0.015	0.069	0.019	-	0.070	0.011	0.002	0.059	0.011	0.024
5	0.077	0.012	0.069	0.067	-	0.074	0.070	0.006	0.070	0.077
6	0.017	0.073	0.026	0.011	0.070	-	0.009	0.063	0.017	0.033
7	0.013	0.069	0.017	0.002	0.067	0.009	-	0.059	0.009	0.022
8	0.067	0.013	0.065	0.056	0.006	0.060	0.057	-	0.061	0.073
9	0.017	0.071	0.013	0.011	0.066	0.017	0.009	0.058	-	0.015
10	0.028	0.082	-	0.024	0.072	0.032	0.021	0.069	0.015	-
11	0.015	0.069	0.019	0.011	0.067	0.013	0.009	0.058	0.013	0.024
12	0.011	0.069	0.013	0.004	0.067	0.006	0.002	0.056	0.006	0.017
13	0.069	0.017	0.067	0.058	0.011	0.064	0.058	0.004	0.060	0.071
14	0.013	0.069	0.017	-	0.067	0.009	0.002	0.056	0.009	0.022
15	0.069	0.011	0.067	0.058	0.002	0.062	0.058	0.002	0.060	0.071
16	0.011	0.069	0.013	0.004	0.067	0.006	0.002	0.056	0.006	0.017
17	0.011	0.069	0.013	0.004	0.067	0.006	0.002	0.056	0.006	0.017
18	0.069	0.017	0.067	0.058	0.010	0.064	0.058	0.004	0.060	0.071
19	-	0.061	0.024	0.015	0.078	0.017	0.013	0.067	0.017	0.028
20	0.069	0.011	0.067	0.058	0.002	0.062	0.058	0.002	0.060	0.071
21	0.017	0.071	0.019	0.013	0.068	0.017	0.011	0.0.058	0.013	0.023

TABLE 7: Continued

	11 TW1(11)	12 TW1(12)	13 TW1(13)	14 TW2(14)	15 TW2(15)	16 TW2(16)	17 TW2(17)	18 TW3 (18)	19 TW3(19)	20 TW3(20)	21 TW3(21)
1	0.015	0.011	0.073	0.013	0.073	0.011	0.011	0.073	0.000	0.073	0.017
2	0.073	0.073	0.017	0.073	0.011	0.073	0.073	0.017	0.064	0.011	0.075
3	0.020	0.013	0.071	0.018	0.071	0.013	0.013	0.071	0.024	0.071	0.020
4	0.011	0.004	0.062	0.000	0.062	0.004	0.004	0.061	0.015	0.062	0.013
5	0.073	0.070	0.011	0.071	0.002	0.070	0.070	0.011	0.082	0.002	0.072
6	0.013	0.006	0.068	0.009	0.066	0.006	0.006	0.068	0.017	0.066	0.017
7	0.009	0.002	0.062	0.002	0.062	0.002	0.002	0.061	0.013	0.062	0.011
8	0.061	0.059	0.004	0.059	0.002	0.059	0.059	0.004	0.071	0.002	0.061
9	0.013	0.006	0.064	0.009	0.064	0.006	0.006	0.064	0.017	0.064	0.013
10	0.024	0.017	0.075	0.022	0.075	0.017	0.017	0.075	0.028	0.075	0.024
11	-	0.006	0.064	0.009	0.064	0.006	0.006	0.064	0.015	0.064	0.013
12	0.006	-	0.062	0.002	0.062	0.000	0.000	0.061	0.011	0.062	0.006
13	0.061	0.058	-	0.062	0.002	0.062	0.062	0.004	0.073	0.006	0.061
14	0.009	0.002	0.059	-	0.062	0.002	0.002	0.062	0.013	0.062	0.011
15	0.061	0.058	0.006	0.059	-	0.062	0.062	0.006	0.073	0.000	0.064
16	0.006	0.000	0.058	0.002	0.058	-	0.000	0.061	0.011	0.00662	0.006
17	0.006	0.000	0.058	0.002	0.058	0.000	-	0.061	0.011	0.07362	0.006
18	0.060	0.058	0.004	0.058	0.006	0.058	0.058	-	0.073	0.006	0.061
19	0.015	0.011	0.069	0.013	0.069	0.011	0.011	0.069	-	0.073	0.017
20	0.061	0.058	0.006	0.059	0.000	0.058	0.058	0.006	0.069	-	0.064
21	0.013	0.006	0.058	0.011	0.060	0.006	0.006	0.058	0.017	0.060	-

From Table 7, sequence divergence, P_i ranging from 0.000 to 0.082 while the pair-wise distance value, ranging from 0.000 to 0.087. The lowest value (P_i=0.000) were between TW1 (12) vs. TW2 (16), TW1 (12) vs. TW2 (17), TW2 (16) vs. TW2 (17) and TW2 (15) vs. TW3 (20) with the highest value (Pi=0.082) was between TW1 (2) vs. TW1 (10). Meanwhile, the lowest value of pair-wise distance is 0.000 [TW1 (1) vs. TW3 (19), TW1 (10) vs. TW1 (3), TW1 (12) vs. TW2 (16), TW1 (12) vs. TW2 (17), TW2 (15) vs. TW3 (20) and TW2 (18) vs. TW2 (16) while the highest value (0.087) was between TW1 (10) vs. TW1 (2).

5) Phylogenetics Analysis

Phylogenetic analysis, both neighbour-joining (NJ) and maximum parsimony (MP) (1000 replicates) trees were constructed using MEGA (Molecular Evolutionary Genetics Analysis) Version 2.1 software (Rozas and Rozas, 1999). In this analysis, the multiple sequence alignment for forward reactions of cytochrome c oxidase I (COI) mtDNA gene sequences was done using CLUSTAL X (1.81) program including one sample of *Helostoma temminckii* from Bakong, Sarawak as an outgroup.

Both neighbour-joining (NJ) (Figure 5) and maximum parsimony (MP) (Figure 6) methods divided samples into two major clades (Cluster A and Cluster B) with strong bootstrap values (100/99 for neighbour-joining and 99/99 for maximum parsimony). The top group (Cluster A) consisted of 14 samples from all three different sites, while Cluster B grouped seven samples of *O. niloticus* also from all three sites. Thus, phyogenetic analyses suggest the existence of two mtDNA lineages in population of *O. niloticus* in Balung River.

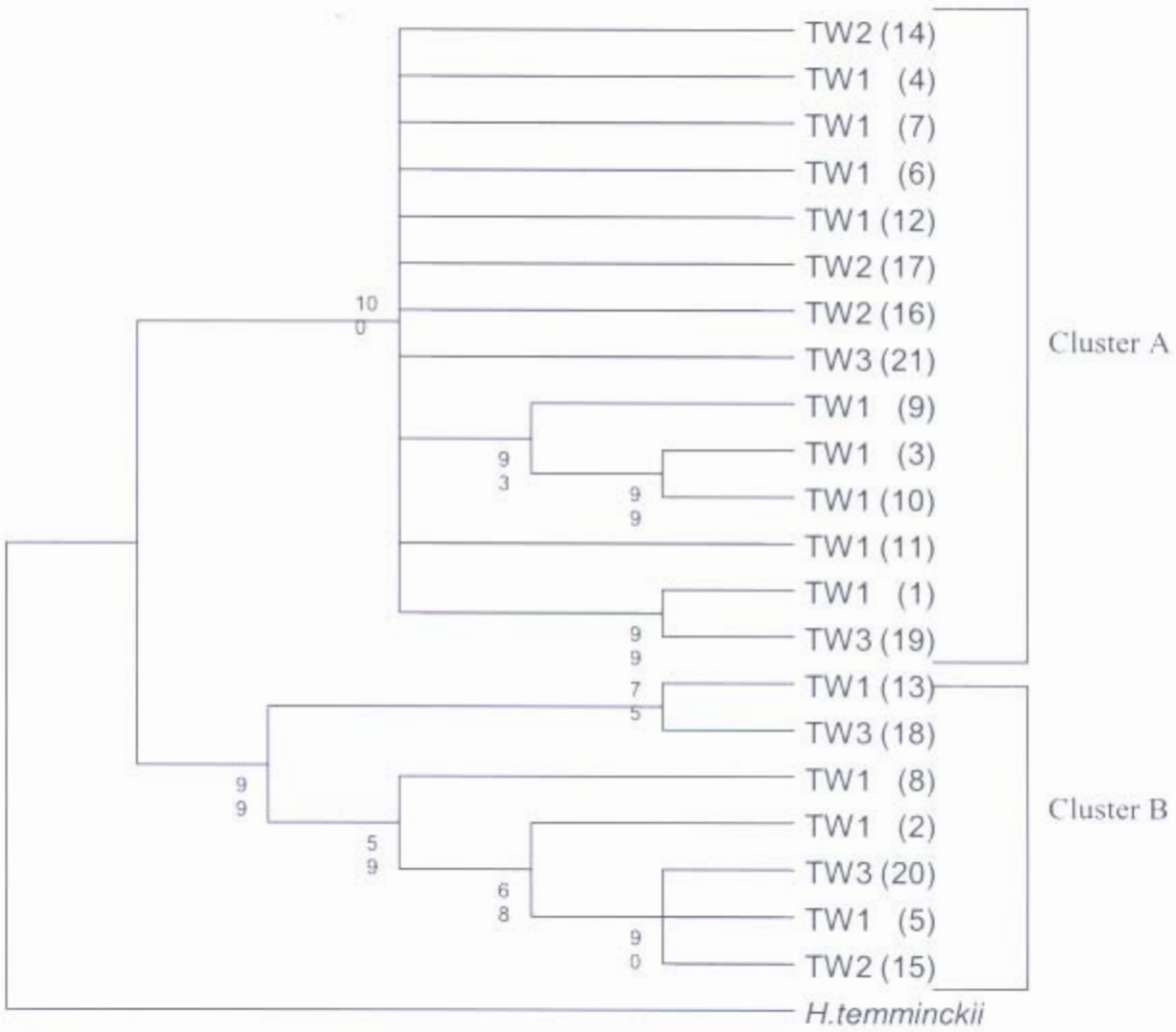

FIGURE 5: Neighbour-joining consensus tree (MEGA) of the *Oreochromis niloticus* inferred from *cytochrome c oxidase* I (COI) mtDNA gene. Abbreviations refer to individuals in populations identified in Table 4. The tree was rooted with a sequence of *Helostoma teminckii* from Bakong, Sarawak. Kimura 2-parameter distance (1980) with 1000 replications was used. Numbers at nodes indicate the bootstrap values (above 50) in percentage.

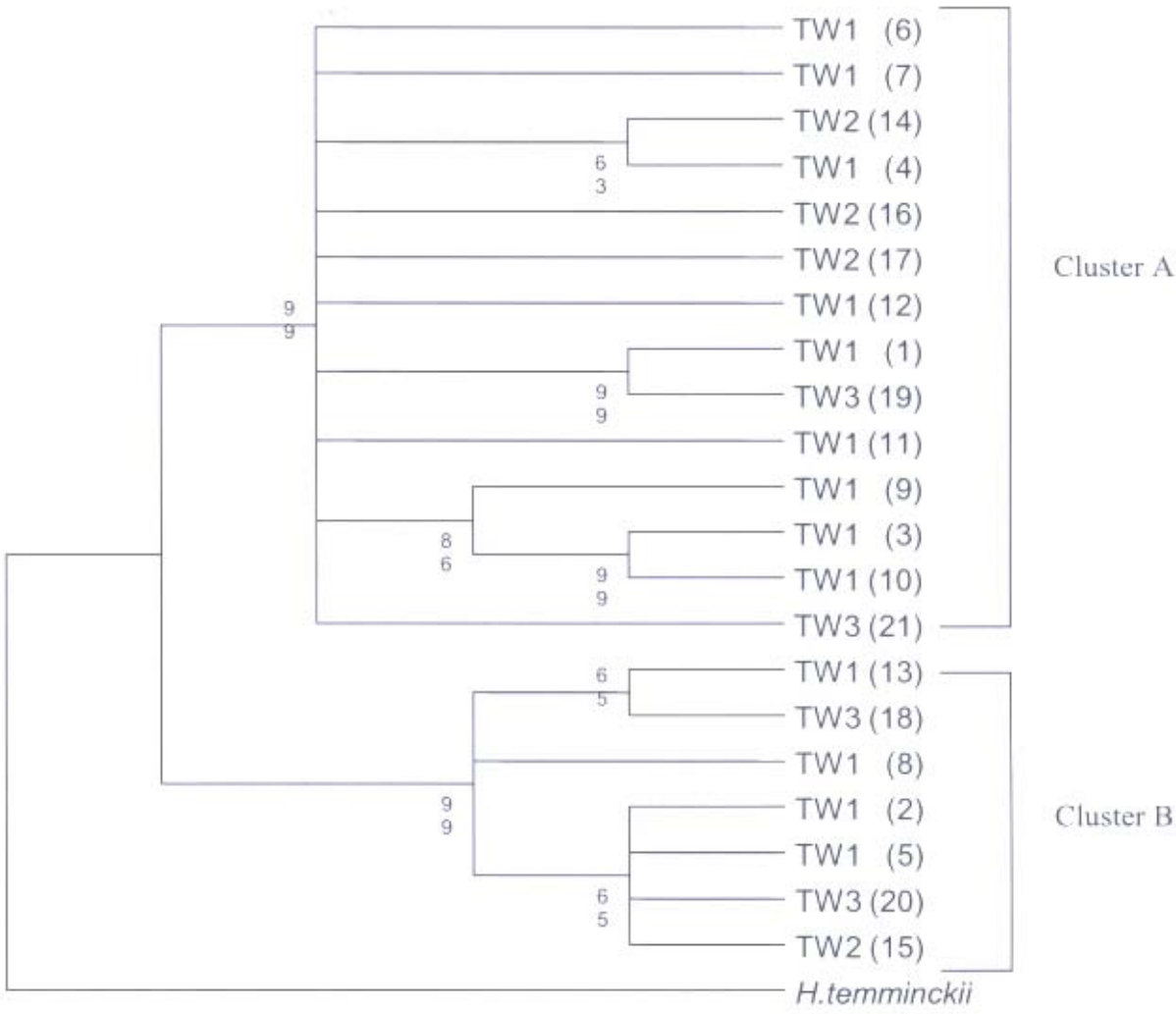

FIGURE 6: Maximum Parsimony consensus tree (MEGA) of the populations of *Oreochromis niloticus* inferred from *cytochrome c oxidase* I (COI) mtDNA. Abbreviations refer to individuals in populations identified in Table 4. The tree was rooted with a sequence of *Helostoma teminckii* from Bakong, Sarawak. Kimura 2-parameter distance (1980) with 1000 replications was used. Numbers at nodes indicate the bootstrap values (above 50) in percentage.

DISCUSSION

Apart from determination of population genetic structure, the objective of the research is to determine the utility of PCR-DNA sequencing of amplified *cytochrome c oxidase* I (COI) in clarifying the genetic structure of *O. niloticus* along Balung River, Tawau, Sabah. *O. niloticus* was introduced into Balung River for aquaculture or reared in fish ponds (Esa, Personal comment). This fish accidentally escaped from hatcheries ponds during flood seasons into Balung River. Since then, they had established themselves in the river.

Twenty one out of 24 samples from various stations of *Oreochromis niloticus* (See Table 2) were successfully extracted and indicated positive results. The majority results of genomic DNA extractions performed yielded a small amount of DNA. During tissue digestion, longer time needed in water bath, which was more than 1 hour to ensure that all the tissues were totally digested. Addition of Proteinase K during tissue digestion was needed if the tissues are hardly to digest. Because it is unusually stable (Burrell, 1993) and its ability to digest native proteins, it has been found in considerable use in procedures where the inactivation and degradation of proteins is required, particularly during purifications of nucleic acids. To facilitate the digestion process, the tube contained tissue should be vortexes using vortex machine.

For preservation, most of the samples were preserved in -20°C. The unsuccessfully extracted DNA might due to the contaminating proteins and RNA. These factors can in some cases inhibit the PCR reaction (Hoelzel and Green, 1992). Some of the successfully extracted DNA only showed the positive results after several extractions (See Figure 2).

PCR amplification was carried out after the extraction protocol was successfully performed. PCR is a sensitive technique which can lead to failure of getting the product. There were many problem associated with PCR and the most common problems are there were no PCR product detected, smearing or multiples bands of double-stranded PCR and the faintness of bands during observation under UV light. Contamination is one of the caused of the problems apart from contaminated solutions or equipment with non-target DNA. The contamination of solutions and equipments might be caused by the exposure to air-borne contaminant and laboratory surface. To avoid the contamination, all the reaction mixtures for PCR should be prepared in safety laminar flow cabinet. The equipments used for PCR dilution protocol especially pipettes should be designated for different procedures. All the pipettes should not come in contact with any amplified DNA at all.

Beside the carefully usage of solutions and equipment, there were also some of the techniques should be taken seriously during loading steps into well. To confirm the presence of contamination, control reaction should always be performed. DNA template should not be taken greedily because too much template of DNA will cause smearing or multiple-bands of double-strand of PCR product. To obtain the presence of PCR product, the DNA needs to perform an electrophoresis using 0.5% agarose gel together with 0.8µl Ethidium bromide solutions. For electrophoresis, 3µl of PCR product were added with 0.6µl GLD. Then, after approximately 30 min running using around 90V, the PCR product should be observed under UV light in order to confirm the best result for the next step; purification (See Figure 3).