PHYLOGENETIC ANALYSIS OF MYSTUS SPECIES USING 16S rRNA IN BATANG KERANG, BALAI RINGIN, SARAWAK.

Goh Pei Tian

Bachelor of Science with Honours (Aquatic Resource Science and Management) 2015
PHYLOGENETIC ANALYSIS OF MYSTUS SPECIES USING 16S rRNA IN BATANG KERANG, BALAI RINGIN, SARAWAK.

Goh Pei Tian (36144)

This dissertation is submitted in partial fulfilment of the requirement for the degree of Bachelor of Science with Honours in Aquatic Resource Science and Management

Aquatic Science Department
Faculty of Resource Science and Technology
University Malaysia Sarawak
2015
ACKNOWLEDGEMENT

Prima facie, I am grateful that with the spiritual strength given through my religion, I am able to complete this final year project smoothly and on schedule. I would like to express my deepest appreciation toward everyone that helps me throughout this study. First of all, I would like to thanks my supervisor, Dr Fazimah bt Aziz for her patient, caring, advice, guidance, dedications and sharing her pearls of wisdom until the end of this final year project. Other than that, I would like to express my appreciation on her hard work for giving comments to improve this thesis. Next, I would also like to immense my grateful toward my co-supervisor, Dr Khairul Adha A. Rahim who also supports the completion on this thesis. I would also like to dedicate my deepest acknowledgement toward Prof Madya Dr Ramlah Zainuddin by lending her laboratory for us to complete our final year project. Next, I would like to give my warmest and most respectful appreciation toward the staff and laboratory assistant of Department of Aquatic Science especially Mr Zaidi Ibrahim, Mr Zulkifli Ahmad and Mr Richard Toh for lending their helping hands regarding the field work. Not to forget, Mr Huzal Irwan Husin, lab assistant for Department of Zoology by giving his warmest help especially in lab work. I would also like to extent my appreciation to post graduate students, Ms Elvy, Ms Farah and Mr Amirul for their helpful guidance and provide their insight and expertise that greatly assisted in this final year project. Besides that, I would like to dedicate my most genuine and deepest thanks to my father, Mr Goh Bon Kiong, mother, Mrs Wong Bee Lan and also my sisters by giving me infinite support, confidence and love to accomplish this final year project successfully. Last but not least, big warmest thanks to all my lab mate, course mate and friend by giving me support, advice, encouragement and friendship. I would also place on record, my sense of appreciation to the one who directly or indirectly have lent their helping hand for me.

I
DECLARATION

I, Goh Pei Tian, hereby to declare that the works and data presented in this thesis are my own work and have been generated by me as the result of my own original research. I have not copied from any other students’ work or from any other source except where due to reference or acknowledgement.

Goh Pei Tian
Date: 28 MAY 2015
Aquatic Resource Science and Management Programme
Department of Aquatic Science
Faculty of Resource Science and Technology
University Malaysia Sarawak
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgement</td>
<td>I</td>
</tr>
<tr>
<td>Declaration</td>
<td>II</td>
</tr>
<tr>
<td>Table of Content</td>
<td>III</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>V</td>
</tr>
<tr>
<td>List of Tables and Figures</td>
<td>VI</td>
</tr>
<tr>
<td>Abstract</td>
<td>VII</td>
</tr>
<tr>
<td>1.0 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2.0 Literature Review</td>
<td>3</td>
</tr>
<tr>
<td>2.1 <em>Mystus</em> species</td>
<td>3</td>
</tr>
<tr>
<td>2.1.1 Taxonomy, Morphology and feeding behaviour</td>
<td>3</td>
</tr>
<tr>
<td>2.1.2 Distribution</td>
<td>4</td>
</tr>
<tr>
<td>2.1.3 Morphology characteristics of some <em>Mystus</em> species</td>
<td>4</td>
</tr>
<tr>
<td>2.1.4 Aquaculture status of <em>Mystus nemurus</em></td>
<td>5</td>
</tr>
<tr>
<td>2.2 Phylogenetic analysis</td>
<td>6</td>
</tr>
<tr>
<td>2.3 Mitochondrial DNA 16s rRNA gene</td>
<td>6</td>
</tr>
<tr>
<td>2.4 Polymerase Chain Reaction (PCR)</td>
<td>8</td>
</tr>
<tr>
<td>2.5 Primer</td>
<td>9</td>
</tr>
<tr>
<td>2.6 Basic Local Alignment Search Tool</td>
<td>10</td>
</tr>
<tr>
<td>2.7 Phylogenetic tree</td>
<td>10</td>
</tr>
<tr>
<td>2.7.1 Tree building method</td>
<td>11</td>
</tr>
<tr>
<td>2.7.2 Tree reliability method</td>
<td>12</td>
</tr>
<tr>
<td>2.7.3 Outgroup</td>
<td>13</td>
</tr>
<tr>
<td>3.0 Materials and method</td>
<td>14</td>
</tr>
<tr>
<td>3.1 Fish sample collection</td>
<td>14</td>
</tr>
<tr>
<td>3.2 Measure and preservation</td>
<td>15</td>
</tr>
<tr>
<td>3.3 Preparation of 0.5M EDTA (Ethylenediaminetetraacetic acid) with pH 8</td>
<td>15</td>
</tr>
<tr>
<td>3.4 DNA extraction</td>
<td>16</td>
</tr>
<tr>
<td>3.5 Gel electrophoresis processes</td>
<td>18</td>
</tr>
</tbody>
</table>
3.5.1 Gel preparation 18
3.5.2 Gel loading activity 18
3.5.3 Getting gel electrophoresis result 19

3.6 Polymerase Chain Reaction 19
3.6.1 Preparation of PCR cocktail or mastermix 19
3.6.2 PCR thermocycler 21

3.7 Purification of PCR Product and DNA Sequencing 22

3.8 Sequence analysis 23

3.9 Phylogenetic analysis 23

4.0 Result 24
4.1 Gel electrophoresis’ result for DNA extraction 24
4.2 Optimising Polymerase Chain Reaction Product 25
4.3 Gel electrophoresis’s result for purified PCR Product 26
4.4 Basic Local Alignment Search Tool (BLAST)’s result 28
4.5 Sequence alignment 29
4.5.1 Information of aligned sequences 29
4.5.2 Percentage of nucleotide base composition 30
4.5.3 Pairwise Distance 33

4.6 Phylogenetic analysis 35
4.6.1 Maximum likelihood (ML) 35
4.6.2 Neighbour Joining (NJ) 37
4.6.3 Maximum Parsimony (MP) 39

5.0 Discussion 41
5.1 Troubleshooting in DNA extraction 41
5.2 Optimizing and Troubleshooting in Polymerase Chain Reaction (PCR) 41
5.3 Sequence divergence 43
5.4 Phylogenetic analysis 44
5.5 Taxonomic Confusion among Mystus spp. 45
5.5.1 Mystus gulio versus Mystus rhegma 45
5.5.2 Mystus nemurus versus Hemibagrus nemurus 47

6.0 Conclusion and recommendation 50

7.0 References 51
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
<td>6</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>16S ribosomal ribonucleic acid</td>
<td>6</td>
</tr>
<tr>
<td>mtDNA</td>
<td>mitochondrial Deoxyribonucleic Acid</td>
<td>6</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
<td>8</td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
<td>10</td>
</tr>
<tr>
<td>UPGMA</td>
<td>Unweighted pair group method with Arithmetic mean</td>
<td>11</td>
</tr>
<tr>
<td>OUT</td>
<td>Operational Taxonomic Units</td>
<td>11</td>
</tr>
<tr>
<td>NJ</td>
<td>Neighbour Joining</td>
<td>11</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
<td>11</td>
</tr>
<tr>
<td>MP</td>
<td>Maximum Parsimony</td>
<td>12</td>
</tr>
<tr>
<td>GPS</td>
<td>Global Positioning System</td>
<td>14</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetraacetic acid</td>
<td>15</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris Acetate EDTA</td>
<td>18</td>
</tr>
<tr>
<td>MEGA</td>
<td>Molecular Evolutionary Genetic Analysis</td>
<td>23</td>
</tr>
</tbody>
</table>
List of Tables and Figures

Table 3.1. Component for PCR mastermix 20
Table 3.2. Primer for 16S rRNA and its sequences 20
Table 3.3. Profile of PCR thermocycler 21
Table 4.1. Thermocycler profile proposed by Jondeung, Sangthong and Zardoya (2007) 25
Table 4.2. The range of thermocycler profile tried for optimisation 26
Table 4.3. The comparison result between BLAST and identified species 29
Table 4.4. Origins and Genbank Accession number of 16S rRNA sequence used in this study 30
Table 4.5. Nucleotide base composition for wild samples only 31
Table 4.6. Nucleotide base composition for wild samples and samples from Genbank 32
Table 4.7. Pairwise genetic distance in percentage for fish samples from Balai Ringin and Genbank 34

Figure 3.1. The study site, Balai Ringin 14
Figure 3.2 Sampling location at Batang Kerang, Balai Ringin, Sarawak 14
Figure 4.1. Agarose gel (1%) showing genomic DNA extraction from Mystus spp. 24
Figure 4.2. Agarose gel (1.0%) showing the PCR purification product from DNA extraction product 27
Figure 4.3. Agarose gel (1.0%) showing the PCR purification product from DNA extraction product 27
Figure 4.4. Maximum Likelihood (ML) tree for Mystus spp 36
Figure 4.5. Neighbour Joining (NJ) tree for Mystus spp 38
Figure 4.6. Maximum Parsimony (MP) tree for Mystus spp 40
Phylogenetic Analysis of *Mystus* spp. using 16S rRNA in Batang Kerang, Balai Ringin, Sarawak

**Goh Pei Tian**

Aquatic Resource Science and Management Program  
Faculty of Resource Science and Technology  
University Malaysia Sarawak

**ABSTRACT**

Southeast Asian region has one of the highest diversity of freshwater fishes. *Mystus* spp. is one of the famous freshwater fish species but there were less research was done on this particular species. This species is regularly misidentified with other family member as they have similar appearance. Therefore this study not only investigated the phylogenetic relationship among the *Mystus* spp. by using the 16s rRNA sequence but also constructed phylogenetic trees for the *Mystus* spp. that can be found in Batang Kerang, Balai Ringin, Sarawak. The muscle tissues of the fishes were taken and their DNA was extracted by using Wizard Genomic DNA Purification Kit (Promega). The extracted DNA was undergoing PCR. The PCR product was observed by using gel electrophoresis and the DNA sequences were analysed using MEGA 6.06 software. Overall, the phylogenetic tree constructed showed the monophyletic relationship of the identified *Mystus* spp. that captured at Batang Kerang. The method used to analyse in this study were Neighbour Joining, Maximum Likelihood and Maximum Parsimony. The methods showed highly similar topography tree with similar bootstrap value. This showed that, the *Mystus* spp. at Batang Kerang is highly and closely related. The phylogenetic trees constructed also successful classified *Mystus* spp that capture at Balai Ringin into respective clades.

Key words: *Mystus* spp., phylogenetic analysis, 16S rRNA, freshwater species, Sarawak

**ABSTRAK**


Kunci: *Mystus* spp., analisis filogenetik, 16S rRNA, ikan tawar, Sarawak.
1.0 INTRODUCTION

Southeast Asian region has one of the highest diversity for freshwater fishes in the world this included Malaysia. There were more than 400 freshwater species have been described in Peninsular and more than 350 freshwater species in Borneo (Esa et al. 2012). Freshwater fishes in Malaysia especially fishes from family Bagridae are the main sources of protein for local residence and important sources of livelihood to them especially in fisheries and aquaculture (Chong et al. 2000). One of the famous species that can be found in Bagridae family is Mystus spp. or known locally as ikan baung.

Based on Ghada et al. (2012) research, they found that there are eight species of Mystus that can be found in Malaysia which are, Mystus gulio, Mystus nigriceps, Mystus planiceps, Mystus micracanunthus, Mystus wyckii, Mystus nemurus, Mystus wolffii and Mystus bimaculatus. They also reported that M. nemurus is the most famous and the largest group that can be found locally.

The study site Batang Kerang, Balai Ringin, Sarawak has been selected in this study because of its diverse freshwater fish species such as the family Anabantidae, Bagridae, Channidae, Clariidaem and Cyprindae (Rahim et al. 2009). Balai Ringin has two different types of river: brown water river and black water river (Rahim et al. 2009). The species flora and forest distribution provided an important habitat for fishes (Rahim et al. 2009).

Comprehensive study on freshwater fishes phylogenetic analysis is limited especially the study on Mystus spp. Previous studies were based on morphology and specified to a particular species only (Darsan, Mahanta & Kumar 2013). Besides that, Mystus spp. considered as new species of catfish as it always being misidentified with
other family member because they shared some similar phenotypic characteristic. Furthermore, phylogenetic analysis on *Mystus* spp. is rarely documented by researcher especially in Malaysia although *Mystus* is one of the common species that can be found in Malaysia freshwater. Darsan, Mahanta and Kumar (2013) stated that the phylogenetic relationships among the Bagridae family are not very clear, especially within the genus *Mystus*. This statement also been supported by Ferdous (2013), as he reported that there was insufficient phylogenetic study has been done on *Mystus* but new species continuously been introduced. The phylogenetic analysis of *Mystus* spp was done by Ferdous (2013) by using cytochrome b but the bootstrap value for supporting the monophyletic status of *Mystus* spp. was low. Additional from him, the research done by Hardman did not support the monophyly of *Mystus* spp.

Therefore the phylogenetic analysis research was done by comparing the 16S rRNA gene sequence of the *Mystus* species that found in the Batang Kerang, Balai Ringin, Sarawak. Esa et al. (2012) also reported that, the development of molecular marker has reinforced the genetic studies on fishes and the invention of Polymerase Chain Reaction (PCR) has greatly enhance the examining and investigation of genetic variation in natural population.

Although there are many researchers that studied phylogenetic relationship for many freshwater species, the phylogenetic relationship for *Mystus* species is still lacking. Thus, the aim of this research is to study the phylogenetic relationship of *Mystus* species. The objective of the research was:

1. To classify *Mystus* spp. based on the sequence analysis of 16s rRNA and examine the relationship among the species.
2.0 LITERATURE REVIEW

2.1 Mystus species

2.1.1 Taxonomy, Morphology and feeding behaviour

*Mystus* spp. belongs to family Bagridae which are classified or categorized under the order Siluriformes. According to Radhakrishnan (2006), the fishes in family Bagridae have an elongated, slender and compressed body; widely separated nostril and the barbel are well developed. The fish’s teeth for this family are either in pre maxillaries, mandible or vomer and the gill membranes are free from isthmus. He also added that, the Bagridae family also have forked caudal fins with intact lateral line and strong spine pectoral fins (Radhakrishnan 2006). The *Mystus* species is first discovered by Scopoli (1777). Based on Ng (2012) the genus of *Mystus* has 30 valid species while 21 species can be found in Southeast Asia. Members of *Mystus* are mostly inhabited freshwater but out of this, two species in South and Southeast Asia are known to live in estuaries and coastal marine, which are *Mystus gulio* and *Mystus wolffii* (Ng 2012). He also proposed the third species, *Mystus velifer* which is a new species that also inhabited the estuaries.

*Mystus* spp. can be identified based on these characteristics: prolong cranial fontanel that reach the base of the occipital process, extremely long adipose fins and barbel, the first gill arch has 11-30 gill rakers, the number of vertebrae reach 37-40, abdominal and caudal part can be measure evenly (Darsan *et al*. 2011). Mohsin and Ambak (1983) reported that, *Mystus* spp. have constrained head with median fontanel and have four pairs of barbell with longer upper jaws. Ghada (2012) reported that, this species are bottom feeder which feed on different range of food items such as other smaller bony fish,
crustaceans, benthic invertebrate and some undigested materials. Most species are carnivorous.

2.1.2 Distribution

Ghada (2012) reported that Mystus spp. is widely distributed in Southeast Asia region ranging from Indonesia to the Asian mainland or from East Indies to Asiatic mainland such as Peninsular Malaysia, Indochina and Thailand. According to Chong et al. (2000), this species can be found in variety habitat such as brackish water, at the mouth of the river to upstream of water or in the freshwater habitat like ponds, lakes, swamps, canals, drains, mining pool and paddy field. The currently Mystus spp that can be found in Malaysia are M. bimaculatus, M. nemurus, M. planiceps, M. wolffii, M. micracanthis, M. wyckii, M. gulio, and M. nigriceps (Ghada 2012).

2.1.3 Morphology characteristics of some Mystus species

Rahim et al. (2009) proposed that in Batang Kerang, Balai Ringin M. micracanthis is one of the Mystus spp. that can be found abundantly. Roberts (1989) stated that M. micracanthis has a narrow head, longer adipose fins than dorsal, moderately large eye and a pair of long barbels with maxillary barbell extending beyond end of the caudal. Other than that, M. micracanthis also has a brownish colour which the tympanum area is a bit darker and has a triangular midpenduncular spot (Robert 1989).

According to Ghada (2012), M. nemurus is one of the most popular Mystus spp. found in Southeast Asia therefore knowing the morphology characteristics are very important. Mohsin and Ambak (1983) pointed that, M. nemurus’s head is wider with slightly longer upper jaw, the adipose fins is shorter than dorsal with eye reaching nasal barbel, maxillary barbels reach the end of the anal fin and the caudal fin deeply forked with
pointed or produced upper lobe. The colour of *M. nemurus* is dark brown with well-defined dark and thin midaxial streak.

According to the local villager, other species that might be found in Batang Kerang, Balai Ringin is the *M. gulio*. Radhakrishnan (2006) said that *M. gulio* has elongated, compressed and slender body with depressed head where the upper surface of the head is not smooth and granulated, the maxillary barbels is black and extend to the base of the pelvic fins. The *M. gulio* has small adipose fin where the base is smaller than the base of anal fin, the fins is more to brownish in colour (Radhakrishnan 2006).

2.1.4 Aquaculture status of *Mystus nemurus*

In Malaysia, the freshwater aquaculture industry only grows rapidly in the year of 1995 where the freshwater cage culture is usually located at mining pools and reservoirs (Kechik 1995). According to Hoh *et al.* (2013) the river catfish *Mystus nemurus* has a great potential to become the alternative fish protein source in South East Asian region. This fish species has been known as one of the famous aquaculture species in Malaysia as it has high commercial value and good flesh quality with high protein content but low in fat value. However, most of the fish available in the market currently comes from various rivers of Malaysia (Hoh 2006). *Mystus nemurus* can be considered as the ‘new’ indigenous species in Malaysia aquaculture but specific information about this species is still lacking especially the feeding behaviour of larval stage (Ghada *et al.* 2012). Although they have high aquaculture value but it is difficult for them to produce in large scale as the seed availability are seasonal which has cause farmer to face some difficulty for mass production (Hoh 2006). According to Hoh *et al.* (2013) there are some specific studies on
the genetics, nutrition and diseases of this species in order to increase their ability to reproduce and lead to increase in species density.

2.2 Phylogenetic analysis

Phylogeny is description on evolutionary history of taxa while phylogenetic is the study of the evolutionary relationships but the relationship can only be reliable by excluding some circumstances (Harrison & Langdale 2006). Additional from Gregory (2008), phylogeny can be defined as a diagrammatic description or taxonomic grouping of organism that might be the in same descent. Therefore, phylogenetic analysis is to deduce or estimate this relationship in terms of relativeness between descent and ancestor (Brinkman & Leipe 2001). Besides that, Brinkman and Leipe (2001) said that the evolutionary history obtained from the phylogenetic analysis is usually illustrated in a branches tree-like diagram that clearly showed the estimated pedigree of the inherited relationship among organism which known as phylogenetic tree. According to Schreiber (2008), he defined phylogenetic tree as a statement about the evolutionary relationship between a set of corresponding characters of one to several organism which may share the same ancestor. Phylogenetic sometime called as cladistics which has a meaning of a set of descendant from a single ancestor. The word “clade” is derived from Greek word means branch.

2.3 Mitochondrial DNA 16S rRNA gene

DNA (deoxyribonucleic acid) is the hereditary material that passed from generation to next generation which made from gene sequence. Eukaryotic cell contains at least one copy of nuclear genome in its nucleus but one cell contains hundreds to thousands of mitochondria. Mitochondrial DNA is found in the cell’s mitochondria. mtDNA is a circular
chromosome with about 17,000 bases pair which most code for replication activity (Minnich & Bowling 2011). According to Muhammad (2011), mtDNA is an intraspecific, micro-evolutionary and highly conserved marker that employed extensively in molecular genetic as information guide for phylogenetic analysis.

16S rRNA is a component of prokaryotic ribosome. 16S rRNA sequences were first discovered by Carl Woese in year 1978 and until now there are over thirty additional sequences had been completed (Noller et al. 1985). Noller et al. (1985) said that, the list of sequence covers the phylogenetic spectrum, chloroplasts and mitochondria which includes the eubacterial, archaeabacterial and eukaryotic kingdom where these sequences allowed sequence analysis in determine the higher order structure. There are several benefit of using 16S rRNA gene sequence for phylogenetic analysis such as 1) large enough with 1500 base pair for information purposes, 2) often exist as multigene family or operons, and 3) the function never change and random sequence change are more accurate after measure of time or after evolution (Janda & Abbott 2007). According to Devereux and Wilkinson (2004), nearly 60 000 16s rRNA sequences are currently available in the Ribosomal Database Project. Besides that, they also listed the importance of ribosomal RNAs such as 1) important elements for protein synthesizing which is the basic components that present in all primary kingdoms, 2) most highly conserved cellular molecules, 3) contain sufficient sequence variability so that the relationship between close related groups can be obtained, 4) the abundance of rRNA made them readily obtained, 5) the conserved nature of 16s rRNA enabled the primers to be designed in nearly full length, 6) creation of PCR made rRNA genes even more accessible for sequencing (Devereux & Wilkinson 2004).
2.4 Polymerase Chain Reaction (PCR)

Every cellular organism has the ability to replicates its own DNA. Polymerase chain reaction (PCR) is scientific creations that magnify a piece of DNA or a particular DNA sequence from thousands to millions of copies (Mullis 1993). PCR is a popular and common technique used in variety of medical and biological research because it is quick, inexpensive and simple. Furthermore, only a tiny amount of original DNA is required and even if the source DNA is relatively poor in quality, the PCR still able to carry out. PCR was developed in year 1984 by the Kary Mullis, American biochemist who received the Nobel Prize and the Japan Prize for her discovery of PCR in year 1993 (Joshi & Despande 2010).

The PCR process only involved the desire DNA to be amplified. In the PCR, there were three major steps which are: denaturation, annealing and extension. In denaturation, DNA is denatured at high temperature or slice into two pieces of single-stranded DNA (Joshi & Despande 2010). Palumbi (1996) proposed that, denaturation process involved using heat to stop all the enzymatic reaction. In annealing, primers anneal to the DNA template in a low temperature and this is the most important phase. This is because if the primers bind to the correct targeted positions in the template, then there will be a probability that the expected synthesis product will be obtained (Palumbi 1996). The primer will tried to discard until close toward the suitable temperature which after that more and more primer molecule will find their perfect complement and start to anneal (Joshi & Despande 2010). In extension, the enzyme is allowed to work, where the targeted DNA segment will be synthesized, the annealed primers extension occurs to create a complimentary copy strand of DNA (Palumbi 1996).
2.5 Primer

Based on Palumbi (1996), to direct the specific synthesis of particular segment of DNA, an oligonucleotide which is a small number of joined nucleotide in a short stretch of nucleotide DNA is required where this oligonucleotide is known as primer. According to Stephenson (2012), a primer is a single-stranded and short piece of DNA that anneals or attaches to its corresponding sequence on the template. In order to provide a starting site for DNA synthesis, the paired primers must bind with either side of the targeted DNA segment. The primers used in PCR are normally in pair, a forward primer and a reverse primer where these two is designed to anneal with each other (Stephenson 2012).

A good or ideal primer will only pair and anneals to targeted DNA sequence in the sample but not the other sequences. Borah (2011) said that a good primer must have characteristics like: a) the optimal primer length for PCR primer is about 18-22 base pair which this length is sufficient for specificity but also enough for binding to template, b) the melting temperature for primers is in the range of 52-58°C as at those temperature the DNA will dissociate into single-stranded and become stable, c) the primer annealing temperature cannot be too high or too low as high temperature cannot produce sufficient primer while low temperature may cause high number of base pair mismatches, d) the guanine and cytosine content in primer should be 40%-60%. From the report of Laboratory for Environmental Pathogens Research (2004), the primer should be avoid to have more than three guanine or cytosine nucleotides at the 3’ end and the primer should not be self-complementary or complementary to any other primer when the reaction occur.
2.6 Basic Local Alignment Search Tool (BLAST)

According to Bal and Hujol (2007), Basic local Alignment Search Tool or BLAST is short was a database search tool or software tool that launched and supported by National Center for Biotechnology Information (NCBI). The database were estimated consist of 540 million sequence in year 2004. The primary purpose of BLAST was to search in the database for local or global alignment that had similar isolated regions to find the relationships among the sequence interested by using the heuristic algorithm (Bal & Hujol, 2007). BLAST is very useful for biologist especially in searching for more information about the DNA sequence obtained and to understand the existing similarity with the sequence in database.

2.7 Phylogenetic tree

A phylogenetic tree is a graphic illustration used to illustrate the evolutionary relationships among organism (Vens & Blockeel 2007). Saitou and Nei (1987) said that, the standard in constructing phylogenetic tree were based on the principle of analyse all possible topologies or certain amount of topologies that are closer to the true tree and the one that showed the smallest amount of evolutionary change will be chosen as the final tree. According to Cho (2012), the tips of the phylogenetic tree are the DNA sequences of the species that examining. The internal node of the tree represent the DNA sequence of ancestor for the organism we observing and the branches of the tree is to connect the DNA sequence to another DNA sequence and finally the root is the DNA sequence of the common ancestor of all the species we examining (Cho 2012). The entire element above will made up a topology. Based on Chuang (2014), the phylogenetic tree construction methods can be classified into two categories which are character based methods and
distance based methods. Character-based method is design based on discrete characters from the molecular sequence of individual taxa while distance based method is designed to measure distance or precisely the degree of differences among sequence of organisms and such distance will be used to frame or construct distance matrix (Chuang 2014). Distance methods include Fitch-Margoliash, Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Neighbour Joining and Minimal Evolution while character based methods include Maximum Parsimony, Maximum Likelihood and Bayesian Method (Cho 2012). For this research only Neighbour Joining, Maximum Likelihood and Maximum Parsimony will be chosen to construct the phylogenetic tree.

2.7.1 Tree building method

Neighbour joining (NJ) was considered as one of the most popular algorithms in constructing phylogenetic trees where it focus on the similarity of molecular information in organism which initially proposed by Saitou and Nei (Vens & Blockeel 2007). Saitou and Nei (1987) defined ‘neighbour’ as a pair of neighbour Operational Taxonomic Units (OTU) that were related through a single interior node in a tree which is unrooted and bifurcating. According to Pearson, Robin and Zhang (2008), neighbour joining can be used to reconstruct large phylogenies because of its high computational speed and high accuracy. This method start by calculating the dissimilarity between a pair of sequence then a pairwise distance will be produced which afterward it will assign each individual to its own cluster and then the most similar clusters will be joined until the last cluster that remain (Vens & Blockeel 2007).

Maximum likelihood (ML) method is normally used to estimate, find or calculate the phylogenetic tree which involves using the topology and branch lengths of the tree
This method creates all possible trees and statistically evaluates which tree is the most possible. The founder of the maximum likelihood method, Felsenstein, stated that this method is for deducing phylogeny from homologous DNA sequences and has a firm statistical basis, which is very useful in recovering correct tree topologies for computer simulation research (Yang 1993). William and Moret (2003) said that likelihood is calculating the probability of the observed data in a given model and maximum likelihood is to estimate the phylogenetic tree data that are most probable. Besides, maximum likelihood tree construction involves two main tasks: first, find the maximum edge lengths of the particular topology, and next find the tree topology that maximizes the likelihood (William & Moret 2003).

According to Yang (1996), Maximum Parsimony (MP) is one of the methods that proposed by Edward and Cavalli-Sforza, then further strengthened by Camin and Sokal. In order to perform a site-by-site analysis, MP will first reconstruct the nucleotide sequence data. Next, the minimum number of nucleotide changes or substitutions will be calculated to illustrate the site pattern (Yang 1996). Then, the number of different substitutions will be summed up to give a parsimony score for the tree topology. Finally, the topology that gives the smallest differential number will be taken as the estimated values for the phylogeny (Yang 1996). However, this method also has the weakness as some had criticized it as being statistically unsound and not suitable for counting the evolution rate (Yang 1996).

2.7.2 Tree reliability method

Peng (2014) said that the purpose of having tree reliability method is to assess the reliability or how much can be trusted for the phylogenetic tree that had been constructed or to compare the reliability of the tree construct with another phylogenetic tree. There are
two methods can be used to test the tree reliability which are Bootstrapping and Jacknifing. Bootstrapping was chosen to be used in this research. Bootstrap was first purposed by Felsenstein which had been widely used and provide examine for the confidence of each clade of a phylogenetic tree which based on the proportion of bootstrap tree (Efron, Holloran & Holmes 1996). Holmes (2003) also said that bootstrapping not only to measure how consistently the data get supported but also give information on the stability of the tree topology which high bootstrap value mean the certain clade are close enough to agree that there are belong to same group.

2.7.3 Outgroup

Historically, outgroup served to root the unrooted networks and/or to estimate ancestral states (Luo et al. 2010). Nowadays, the criteria of outgroup has been studied extensively and expected to be very useful for rooting the phylogenetic tree (Huelsenbeck, Bollback & Levine 2002). According to Huelsenbeck, Bollback and Levine (2002), outgroup is very important and necessary to add in a phylogenetic tree as it can determine the root of the tree. However the quality of the rooting provided by outgroup should also depend on the sample of the outgroup species and their phylogenetic proximity to the ingroup (Huelsenbeck, Bollback & Levine 2002). Outgroup can be defined as a group of species or a single species that is not included or not related in to the group of species that under study (Michu, 2007). Based on Luo et al. (2010) viewed, selection of outgroup is often arbitrary or depends on the obscure relationship between outgroup and ingroup taxa. The selection or choice of outgroup is very critical as it can directly influence the topology of the ingroup tree. There are several criteria in choosing outgroup such as: 1) should be outside but have relationship with ingroup, 2) prefer to be the sister group with the ingroup, 3) should contain more than one taxon (Luo et al. 2010)
3.0 MATERIALS AND METHODS

3.1 Fish Sample Collection

The wild samples used in this research were from Batang Kerang, Balai Ringin, Sarawak. Sampling was done according to Rahim et al. (2009) studied, where the fish was caught using 3 layers gill net. The coordinate was taken by using Global Positioning System, GPS (Garmin GPSmap 62S). The gill nets were placed at the selected stations early in the morning and the specimens were collected at the noon. Fish samples were also obtained from local fish market (Balai Ringin market). The sampling was conducted on 11 October 2014. According to Rahim et al. (2009), Balai Ringin faced high water seasons during that period therefore more fish and diverse species can be found.

Figure 3.1. The study site, Balai Ringin
(Source: Google Map, 2014)

Figure 3.2. Sampling location at Batang Kerang, Balai Ringin, Sarawak.
(Source: Google Earth, 2015)
3.2 Measure, preservation and identification

Fish samples were measured for the total length, standard length and body depth by using the measuring board (Wildco 118). The weight was measured by using portable weight balance (SHIMADZU ELB 3000). The samples were morphologically identified by using the keys provided by Inger and Chin (1962), Mohsin and Ambak (1983), Robert (1989), Kottelat et al. (1993) and Radhakrishnan (2006). Some muscle tissues were taken out from the fish and were preserved with 95% of absolute ethanol in the sample bottle (Esa et al. 2012). The sample bottle was labelled carefully with name of collector, day and date, venue of sampling and coordination, sample number and purposes. The dissected fish was preserved in 70% ethanol inside a big sample bottle. The fish tissues were stored in refrigerator with -20˚C before undergo genetic analysis.

3.3 Preparation of 0.5M EDTA (Ethylenediaminetetraacetic acid) with pH 8

The EDTA was prepared according to the recipe provided by Dolan DNA Leaning Center (2005). The EDTA has the molarity of 0.5M and pH of 8. Before the preparation, a schott bottle has autoclaved (Hirayama autoclave HVE-50) beforehand as it will be used to fill in the diluted EDTA. First, a 1 litter beaker was placed on the magnetic stirrer plate (Boeco Germany MSH 300) and the magnetic stirrer was also place inside the beaker. The measured 900ml of distilled water was poured into the beaker. The power of magnetic stirring plate was turn on and the speed of stirring was gradually increase from low to high until a vortex appeared in the distilled water. 186.10g of crystallised EDTA was measured by using the digital balance (AND company limited FX-3000i). The weighted EDTA was pour slowly into the beaker. The mixture was allowed to mix and dissolve for at least 1 hour.