MOLECULAR APPROACH TO THE IDENTIFICATION AND PHYLOGENETIC RELATIONSHIP AMONG ACTINOPTERYGII USING MITOCHONDRIAL DNA 12S rRNA, 16S rRNA AND CYTOCHROME B GENES IN PENINSULAR MALAYSIA

by

SITI FARAH QURAISHIA BINTI SAMRAH

Thesis submitted in fulfillment of the requirements

for the degree of

Master of Science

August 2018

ACKNOWLEDGEMENTS

Thanks to Merciful God, Allah the Almighty for giving me the opportunity in completing my MSc study.

My appreciation extends to several people who have both physically and morally contributed in this journey. First of all, my heartfelt gratitude goes to my family, especially my mom. Her ceaseless psychological support made me wanting to pursue MSc and be more determined to complete it. My sincere appreciations go to my father and sister for their support starting from the beginning till end.

I would also like to express my deepest appreciation and thankful to my main supervisor, Dr. Nur Haslindawaty Abd Rashid for her expert guidance, advice and knowledge throughout my Master of Science (Forensic Science) journey. I am also grateful to my co-supervisor, Professor Dr. Zafarina Zainuddin and Mr. Sundarajulu Panneerchelvam for their endless help and guidance.

It is my privilege to acknowledge Malaysia Fisheries Department particularly Mr. Sallehudin Jamon from Fisheries Research Institute. I would like to also thank Prof. Emeritus Dr. Mohd. Azmi Ambak and Dr. Seah Ying Giat from Universiti Malaysia Terengganu for their contribution in pre-identification of fish species.

This research study is funded by the university's Short Term Grant Scheme (304/PPSK/61313109). Special thanks to MyBrain15 for supporting the tuition fees for two and half semesters. Also, I would like to express my gratefulness to USM Fellowship for sponsoring the remaining tuition fees and monthly allowances as well. Last but not least, thanks to all forensic lab staffs, lab colleagues and fellow postgraduate friends who may have indirectly contribute to this study.

TABLE OF CONTENTS

LIST OF TABLES **PAGE**

LIST OF FIGURES **PAGE**

LIST OF SYMBOLS AND ABREVIATIONS

A	Adenine
$NH_4(SO_4)_2$	Ammonium sulfate
T_m	Annealing temperature
cytb	Cytochrome b
C	Cytosine
°C	Degree celcius
dNTPs	Deoxynucleotide triphosphates
DNA	Deoxyribonucleic acid
ddH_2O	Double deionized water
EDTA	Ethylenediaminetetraacetic acid
g	Gram
G	Guanine
$MgCl_2$	Magnesium chloride
ML	Maximum likelihood
MP	Maximum parsimony
µL	Microliter
mg	Milligram
mg/mL	Milligram per milliliter
mL	Milliliter
ng/µL	Nanogram per microliter
NJ	Neighbor-joining
nDNA	Nuclear DNA
%	Percent
PBS	Phosphate buffered saline

pmol	Picomole
PCR	Polymerase chain reaction
rpm	Revolution per minutes
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
T	Thymine
TBE	Tris borate EDTA
UV	Ultraviolet
U/µL	Unit per microliter
V	Volt

PENDEKATAN MOLEKULAR BAGI PENGENALPASTIAN DAN HUBUNGAN FILOGENETIK ANTARA ACTINOPTERYGII MENGGUNAKAN GEN-GEN MITOKONDRIA DNA 12S rRNA, 16S rRNA DAN SITOKROM B DI SEMENANJUNG MALAYSIA

ABSTRAK

Kaedah morfologi telah lama digunakan dalam menentukan dan mengklasifikasikan spesies ikan. Walaubagaimanapun, kaedah ini tidak memberikan informasi mengenai genetik sesuatu spesies yang menyebabkan kurangnya fungsi kaedah ini dalam kajian filogenetik. Kajian ini telah menggunakan gen-gen mitokondria DNA (mtDNA) iaitu 12S rRNA, 16S rRNA dan sitokrom b bertujuan untuk pencirian, mengkaji polimorfisme dan hubungan genetik antara spesies ikan laut. Hubungan genetik antara spesies ikan laut terpilih ditentukan melalui pohon-pohon filogenetik menggunakan kaedah jarak, *Neighbor-joining* (NJ) dan kaedah karakter, *Maximum parsimony* (MP) dan *Maximum likelihood* (ML). Keputusan menunjukkan gen 12S rRNA mempunyai jujukan yang lebih terpelihara berbanding gen-gen 16S rRNA dan sitokrom b. Hasil penelitian juga menunjukkan *Nemipterus nemurus* mempunyai jumlah polimorfisme tertinggi yang mencadangkan spesies ini mungkin telah mengalami perubahan peristiwa genetik terbesar serta mempunyai pencapahan genetik yang paling jauh dari susur galur keturunan. Analisa dari pohon filogenetik menggunakan gen-gen 12S rRNA, 16S rRNA dan sitokrom b telah menghasilkan corak pengelompokan topologi yang serupa dengan sedikit perbezaan dalam nilai bootstrap. Beberapa spesies seperti *Liza macrolepis* dan *Liza vaigiensis* telah mempamerkan sifat morfologi yang seiras, tetapi menunjukkan hubungan genetik yang jauh. Walaubagaimanapun, spesies-spesies seperti *Pampus argenteus* / *Pampus chinensis* dan *Drepane longimana* / *Drepane punctata* telah memperlihatkan

persamaan dari segi morfologi dan secara konsistennya telah ditemui dalam klad yang sama dalam semua topologi disokong dengan nilai bootstrap yang tinggi (97-100%). Spesies yang berbeza morfologi seperti *Megalaspis cordyla* dan *Caranx sexfasciatus* telah mempamerkan kelompok bersama dalam semua topologi dan memiliki nilai persamaan jujukan DNA yang tinggi. Kajian ini juga menunjukkan bahawa kaedah gabungan jujukan gen-gen memberikan resolusi yang lebih tinggi bagi semua pohon filogenetik. Secara keseluruhan, penemuan-penemuan dalam kajian ini menunjukkan bahawa kaedah molekular menggunakan gen-gen mitokondria DNA berjaya diaplikasikan kepada spesies ikan laut dalam kelas Actinopterygii di Malaysia.

MOLECULAR APPROACH TO THE IDENTIFICATION AND PHYLOGENETIC RELATIONSHIP AMONG ACTINOPTERYGII USING MITOCHONDRIAL DNA 12S rRNA, 16S rRNA AND CYTOCHROME B GENES IN PENINSULAR MALAYSIA

ABSTRACT

The morphological based method has been utilised to identify and classify fish species. However, this method is lacking genetic information which makes it less useful in phylogenetic studies. The present study utilises mtDNA markers namely 12S rRNA, 16S rRNA and cytochrome b genes to characterize, identify the polymorphism and genetic relationship among marine fish species. The genetic relationship among the selected marine fish species was inferred through phylogenetic trees using Neighbor-joining (NJ), Maximum parsimony (MP) and Maximum likelihood (ML) methods. The result showed that 12S rRNA gene was more conserved than 16S rRNA and cytb genes based on the polymorphisms reported. The finding also revealed that *Nemipterus nemurus* exhibits the highest number of polymorphisms indicating that this species might have experienced the most changed events and thus possessing the most genetic deviation from the common shared ancestry. All the phylogenetic trees that were constructed based on 12S rRNA, 16S rRNA and cytochrome b genes showed a similar clustering with slight differences in their bootstrap values. Some species with greatly morphological resemblance such as *Liza macrolepis* and *Liza vaigiensis* however, showed a genetically distant relationship. Nevertheless, another morphological resemblance species namely *Pampus argenteus* / *Pampus chinensis* and *Drepane longimana* / *Drepane punctata* were consistently observed in all topologies with high bootstrap support (97-100%). Having a distinctive morphological, the *Megalaspis cordyla* and *Caranx sexfasciatus* occupied a highly similar nucleotide

sequences and showed a consistent clustering in all topologies. This study also revealed that the combined genes provide a better resolution for all phylogenetic trees. Overall, these findings showed that the molecular based method using mtDNA genes were successfully applied to study the diverse Malaysia marine fish species in class Actinopterygii.

CHAPTER 1

INTRODUCTION

This thesis studies the diverse class Actinopterygii found in Malaysia's marine waters. Class Actinopterygii has been recognised as the largest vertebrates (Jonna, 2004) and known to be commercially important as they provide about 85% of total seafood sources in Malaysia (Chowdhury and Yahya, 2012). Due to a great consumer demand, it has caused a threat in stock decline which had eventually increases the risk of fraudulent substitution or mislabeling in the processed food products. Therefore, investigating their phylogeny and evolutionary history are important, whereby, accurate fish species identification and proper management of the fish stock will ensure their sustainabilities (Murthy *et al*., 2001). Genetic markers such as mitochondrial DNA 12S rRNA, 16S rRNA, cytochrome b and cytochrome oxidase I genes have been used frequently in identification of species including in fish studies (Li and Orti, 2007; Kochzius *et al*., 2010; Cawthorn *et al*., 2012; Wang *et al*., 2017). The ability of mitochondrial DNA markers to correctly identify species especially those closely related species have made these markers to be more utilised in species identifications (Boonseub *et al*., 2009). The topographies of conserved and rapidly evolving regions have proved that these markers are useful for resolving phylogenetic clade at various categorical levels. The application of these genes give advantages to the present study on class Actinopterygii in Malaysia which has been rather questionable and requires a scrutiny from many aspects including molecular identification and phylogenetic studies.

1.1 Systematics

The science of systematics based on the classification of organisms according to their evolutionary relationships. Systematics involves three main components which are recognized as a basic unit (species), species classification in a hierarchic system and placing into some broader context (Schuh, 2000).

1.2 Fish systematics

Fishes are aquatic vertebrates that possess general physical morphologies such as gills, scales and fins with variety of colors, forms and sizes. Klappenbach (2016) described the Ostracoderms as the most primitive fish identified with a notochord, no jaw bones and teeth that had appeared about 510 million years ago. Agnatha is also among the oldest fish consisting lamprey and hagfish which are known to be the only survival ancient fishes nowadays (Klappenbach, 2016). Fishes are very diverse, comprising cartilaginous fishes and bony fishes that can be further classified into lobe-finned fish and ray-finned fishes.

1.2.1 Chondrichthyes (cartilaginous fishes)

Chondrichthyes or cartilaginous fish is an ancient group of animals, had experiencing changes in 100 million years (Icelandic Fisheries, 2015). The cartilaginous fishes include sharks, skates, rays and chimaeras. The cartilaginous fish constitutes around 850 species (Klappenbach, 2016). The main features of cartilaginous fishes include possessing skeleton entirely made of cartilage, breathing through spiracles and have no gills (Kennedy, 2011). Cartilaginous fish is classified under vertebrates possessing four types of features which include caudal, anal, dorsal and ventral fins which are upheld by girdles of the skeleton inside their body (Nicholson, 2015).

1.2.2 Osteichthyes (bony fishes and tetrapods)

Osteichthyes comprises all bony fishes while tetrapods constitutes over 42,000 species with more than 28,000 living species in class Actinopterygii and 24,000 species in class Sarcopterygii (Diogo *et al*., 2008). Based on Encyclopaedia Britannica (2014), general characteristics of Osteichthyes include ossified (true) bone, presence of swim bladder and gills. Osteichthyes is a superclass comprising subdivisions of Sarcopterygii and Actinopterygii which are referred as lobe-finned and ray-finned fishes, respectively.

1.2.2(a) Sarcopterygii (lobe-finned fishes)

The word 'Sarcopterygii' originated from Greek meaning flesh fin and known to be the ancestor of all tetrapods, as their fins resemble the shape of ancestral tetrapod limbs (Zimmer, 2017). As described by Murphy (2005), the main features of Sarcopterygii are muscular pelvic and pectoral fins which are joined to the body by a single bone. Sarcopterygii is further classified into two types which are Dipnoi (lungfish) and Coelacanths (Actinistia). The lungfish is the only genera that persisted today and can be found in the area commonly experiencing seasonal droughts, with the ability to breathe air (Edwin, 2014). Meanwhile, two species under Coelacanths known as *Latimeria chalumnae* and *Latimeria menadoensis* are listed as both critically endangered and vulnerable in the Red List by the International Union for Conservation of Nature (IUCN, 2015).

1.2.2(b) Actinopterygii (ray-finned fishes)

The existence history of Actinopterygii, was started during Devonian Era and diversed during Carboniferous Era (Jonna, 2004). Initially, this group evolved from freshwater and later had broadened into marine water (Klappenbach, 2016). The class Actinopterygii is the most successful and the largest fish group, constituting half of all living vertebrates (Weitzman, 2015). Actinopterygii are adapted to various habitats from cold Artic and Antartic oceans to desert hot springs, but often limited to one habitat throughout their life (Weitzman, 2015). According to Ax (2003), this group makes up the largest aquatic jawed vertebrates. The main characteristic of Actinopterygii is that they possess lepidotrichia or fin rays (Jonna, 2004). The fins are connected to the internal skeletons which are held by bony spines and forms web-like or ray fins (Jonna, 2004).

The major subdivision in Actinopterygii known as Teleost encompasses over 23,000 species of total 24,000 in class Actinopterygii (Jonna, 2004). Malaysia seawater shows a great diversity in fish species as there are approximately 1,352 marine fishes recorded (Froese and Pauly, 2017). Due to highly diverse and presence in a vast number, it is vital to have an appropriate and systematic identification system to ensure an accurate characterisation and documentation of fish species could be implemented.

1.3 Species identification

There are many methods that had been used for identification and characterisation of organisms which includes morphological based method (Kim, 2002; Ouattara *et al*., 2014), macro-molecular study of hemoglobin (Espinoza *et al*., 1999) and protein based method (Asensio and Montero, 2008; McHugh and Arthur, 2008). However,

application of those methods failed to effectively discriminate closely related species (Bellis *et al*., 2003). Introduction of DNA markers for species identification provide a more accurate and robust, while increasing sensitivity involving those incomplete features of species. Friedheim (2016) suggested the requirement of a combination method to produce an accurate identification of species.

1.3.1 Morphological based method

Morphological method involves identification and classification of organisms solely based on their physical characteristics which are inherited from their ancestor with no genetic relationship information is acquired. Morphological characteristics such as body shape, size, fin ray count and color are normally used in identification based on this method. However, this method is susceptible to misinterpretation of evolution information of each species since only appearance is taken into account instead genetic information which may have deviated without obvious changes to physical traits.

Teletchea (2009) specified that morphological method cannot be applicable in species identification as they may have undergone physical changes such as color and shape when processed as food products. Kochzius *et al*. (2010) has stated that the difficulty to authenticate the processed foods and traded wildlife species using morphological method is due to the loss of morphological traits which will be troublesome and complicate the identification of species using morphological methods. It has also been reported that fraudulent substitution with less expensive species had contributed to the difficulty in the species identification due to the partially or completely loss of physical characteristics (Pepe *et al*., 2005). According to Armani *et al*. (2011), in China, the most exploited fish species (Chinese fish) has been used to

replace the expensive Italian fish species since both species showed similarity in their size and shape (Armani *et al*., 2011).

Morphological identification is also unsuitable to be used for confirmation of organisms due to its unspecified and restricted characters observed (Zhang, 2011). In addition, intraspecific variation that exhibited between fish species may also contribute into the weakness of assessment based on physical characteristics and morphological differentiation (Teletchea, 2009). As reported by Ouattara *et al*. (2014), the African catfish genus Chrysichthys share a high appearance resemblance among catfish species therefore leads to difficulty in identification and differentiation of the species. Hence, the knowledge and experiences acquired by taxonomists are the main factors influencing identification of fish species based on morphological appearance. Besides, study by Ko *et al*. (2013) has revealed a misidentification of fish larvae based on morphological method conducted by different taxonomists. This study also showed poor discrimination was obtained through morphological method compared to DNA barcoding (Ko *et al*., 2013). Therefore, those limitation factors have led to lack of accuracy and inconsistency of identification using morphological method.

1.3.2 Protein based method

Protein based method is a classical method used to assess fish authentication mainly based on separation of specific protein using electrophoresis procedures, which include capillary electrophoresis, immunoassay systems, high performance liquid chromatography and isoelectric focusing methods (Teletchea, 2009). Proteins such as haemoglobin and transfererrin have been used as molecular genetic markers (Okumus and Ciftci, 2003). A few years later, alloenzyme was used to identify the genetic variation at the level of enzymes (encoded directly by DNA) (Okumus and Ciftci,

2003). This method is more reliable for testing of fresh, frozen tissue or lightly processed foods. However, this method is unsuitable for heavily processed foods claimed for fraudulent and species misidentification due to protein degradation. Moreover, protein is thermally unstable and can easily be modified when exposed to higher temperature which would cause loss of biological information shortly after death (Cawthorn *et al*., 2012). Study by Mackie *et al*. (1999) shows that the intense heat during processing and drying are able to destroy the biochemical properties and structural integrity of a protein.

One of the protein based method that proved to be useful even in heat sterilised products is enzyme linked immunosorbent assay (ELISA) which has been used for several fish species identification (Carrera *et al*., 1997). However, this immunoassay is ineffective in differentiating closely related species and requires development of species specific antibodies (Woolfe and Primrose, 2004; Hellberg and Morrissey, 2011). General limitations of the protein based method are due to very laborious and time consuming protocol, which has raised the need for an alternative method based on the application of DNA markers.

1.3.3 DNA based method

DNA profiling also known as DNA analysis is a technique employed to assist in the identification of individual based on uniqueness of DNA sequence (Easteal *et al*., 1991). DNA fingerprint technique was proposed by a geneticist, Alec Jeffreys when he discovered the unique length of variable number of tandem repeats (VNTRs) between individuals in paternity and immigration disputes (Inman and Rudin, 1997).

Application of DNA analysis in identification of fish species has been practiced since almost 30 years ago (Ward *et al*., 2005). Since then, the identification based on DNA analysis is widely used for species characterisation and delineation. Taxonomic classification based on DNA method offers forthright basic protocols through few steps including DNA extraction, amplification by Polymerase Chain Reaction (PCR) and DNA sequencing (Tautz *et al*., 2003). According to Finizio *et al*. (2007), DNA based method provides a good platform for the authentication of processed food due to several advantages over protein which are heat labile, reproducible and accessible in almost every cells of organism. Due to the relative stability features of DNA as compared to other biological molecules, DNA has become an essential tool applied in identification of body or body parts from mass disasters, anthropology and ancient historical studies (Inman and Rudin, 1997). In case of lack or loss morphological characteristics such as changes in color and shape in processed foods, DNA offer the best choice for species identification due to its unchanged genetic information.

Application of DNA based method delivers as a relatively simple, specific and sensitive method especially in those involving poorly defined morphological characteristics such as newly hatched fish larvae and largely differed juvenile and adults (Boonphakdee and Sawangwong, 2008). In addition, this method is also applicable for fisheries research and control as well as in clarifying adulteration or substitutions in the marketplace (Cawthorn *et al*., 2012). DNA based method can be performed either by using nuclear DNA (nDNA) or mitochondrial DNA (mtDNA) to identify species and resolve species relationship.

1.4 Nuclear DNA

Antoni von Leeuwenhoek (1632-1723) was the first person who discovered nucleus which was later described by Robert Brown in 1831 (Mazzarello, 1999). Nucleus is the largest membrane bound organelle involved in activity of the cell, reproduction, synthesis of protein and also carries the cell's genetic information (Bailey, 2017). Unlike mtDNA, nDNA consists mostly of noncoding gene and present either between genes of unknown functions or between genes with some stretches of noncoding gene (Krawczak and Schmidtke, 1998). The nuclear ribosomal RNA (rDNA) genes comprise hundreds of tandem repeats distributed on different chromosomes. The nuclear rDNA genes also possesses a coding gene, the intergenic spacers (IGSs) which include the external transcribed spacer (ETS), internal non-transcribed spacer (NTS) and the transcription unit; 3 rDNA genes (18S, 5.8S and 28S).

1.4.1 Applications of nuclear DNA markers

The nuclear DNA (nDNA) genes have been used frequently in species identification and evolutionary studies of organisms due to their slow rate of evolution. Meyer (1993) has suggested that the nDNA such as 18S rRNA and 28S rRNA genes are evolving 100 times slower than the mtDNA as a whole. Due to this feature, these genes were proposed as standard genetic marker or DNA barcoding gene in species identification of invertebrates (Tautz *et al*., 2003; Vences *et al*., 2005) since their amplification were challenging in vertebrate samples. Dabert (2006) stated that the 18S rRNA gene is the most frequently used for phylogenetic study to resolve all taxonomic categories since it provides informative characters in evolutionary studies due to small size of the gene (1.8 Kb). However, Hofmann *et al*. (2015) encountered difficulties in distinguishing a closely related fish species using this gene although it showed highly conserved at

species level. Meanwhile, the conserved region of 5.8S rRNA gene showed relatively few informative characters which made this gene less utilised in phylogenetic studies (Hillis and Dixon, 1991; Rokas *et al*., 2002; Dabert, 2006).

The internal transcribed spacers (ITS) include ITS1 and ITS2 located in the ribosomal nDNA found to be evolved more rapidly than the 18S rRNA and 28S rRNA genes (Dabert, 2006). Nevertheless, sequences intra-specificity to this gene has restricted application in phylogeny inferences. The feature of fast evolving has made nuclear DNA genes to be more appropriate for confirmation of species status rather than species identification (Dabert, 2006).

1.5 Mitochondrial DNA

Animal mitochondrial are abundantly present in the cell (Mandal *et al*., 2014) and are in circular in shape (Ladoukakis and Zourus, 2017). Mitochondria play important role as the main source of energy production (Abhyankar *et al*., 2009), respiration, self-destruction of a cell, aging and genetic disease (Mandal *et al*., 2014). Due to its vital role, the genome of mitochondria is expected to be highly conservative (Brown *et al*., 1979).

Mitochondrial DNA (mtDNA) genome is an alternative genome other than nuclear DNA which exists in most organisms and possesses their own organelle-specific DNA systems involving replication, transcription and translation (Yang *et al*., 2014). The animal mtDNA genome is generally a small, circular single and double-stranded DNA molecule with approximately 17 Kb of length (Figure 1.1) (Ladoukakis and Zourus, 2017). In different animal species, mtDNA showed a slightly vary in size

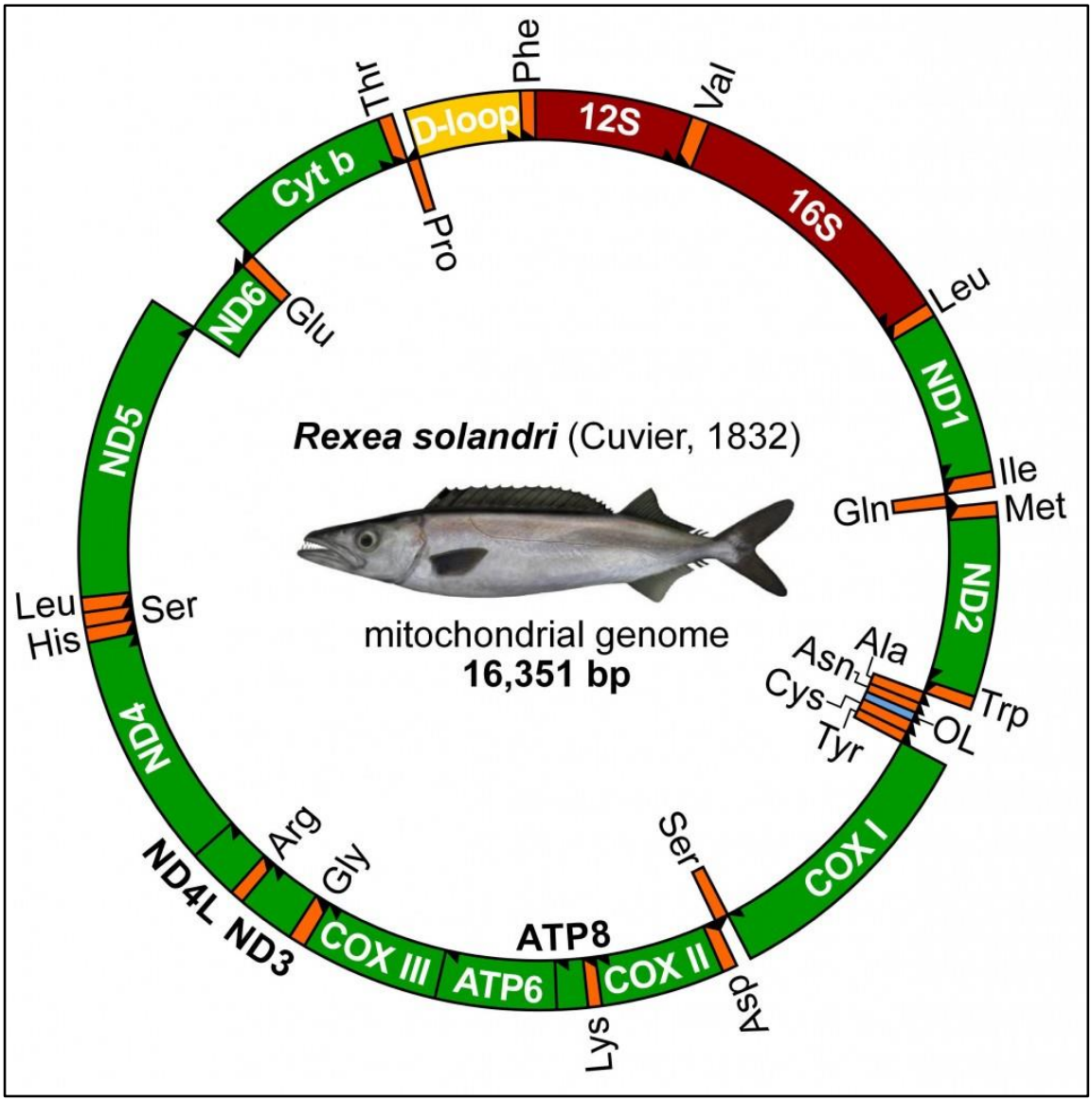

Figure 1.1 Diagram showing the representative animal mtDNA genome (Source:www.molecularfisherieslaboratory.com.au/next-generation-sequencing-aids-gemfish-fishery/)

and arrangement of the genes. An example of the largest animal mtDNA observed is in scallop in which their size can be up to 39.3 Kb (Gjetvaj *et al*., 1992). According to Yamauchi *et al*. (2004), animal mtDNA has a simple and uniform organisation consisting of 37 genes that include two ribosomal genes (12S rRNA and 16S rRNA), 22 transfer RNA (tRNA) genes and 13 protein coding genes that are involved in oxidative respiration (Abhyankar *et al*., 2009).

MtDNA can be divided into two main regions, comprising a small control region and more than 94% representing a coding region (Hoong and Lek, 2005). The animal control region is also known as displacement loop (D-loop) which consists of three domains such as central domain, the extended termination associated sequences (ETAS) domain and the conserved sequence block (CSB) domain (Arif *et al*., 2011; Panday *et al*., 2014). The central domain is a conserved region while both ETAS and CSB are rapidly evolving regions (Arif *et al*., 2011).

Mitochondrial DNA coding region can be classified into protein and non-protein genes. The protein coding region comprises 13 genes include seven subunits of NADH dehydrogenase (ND1, ND2, ND3, ND4, ND4L, ND5 and ND6), cytochrome b, three subunits of cytochrome c oxidase (COI, COII and COIII) and two subunits of ATP synthase (ATPase 6 and ATPase 8) (Meyer 1993; Wada and Yokohama, 2004). The 13 protein coding genes are responsible in the oxidative respiration process (Linacre and Tobe, 2011). Meanwhile, the non-protein coding regions include two ribosomal RNAs (small 12S and large 16S) and 22 transfer RNA (tRNA) genes (Abhyankar *et al*., 2009). As stated by Klug and Cummings (1997), both rRNAs and tRNAs genes do not encode for any protein but are directly involved as ribosome's constituent especially tRNAs. Moreover, both rRNAs and tRNAs also aid

in regulation of mitochondrial replication and transcription of D-loop (Abhyankar *et al*., 2009).

1.5.1 Mitochondrial DNA (MtDNA) markers

The mtDNA genes have been used frequently as molecular markers for identification of various species (Robba *et al*., 2006; Nuraini *et al*., 2012; Imtiaz *et al*., 2016), phylogenetic studies (Seah *et al*., 2008; Rahim *et al*., 2016), authentication of processed foods (Pepe *et al*., 2005; Kesmen *et al*., 2012) and wildlife investigation (Rajput *et al*., 2013; Panday *et al*., 2014). Owing features such as no recombination process (Martins *et al*., 2003), higher copy number in each cell as compared to nDNA (Branicki *et al*., 2003; Yang *et al*., 2014), exclusively maternal inherited (Carr *et al*., 2008) and exhibiting higher evolutionary rate about ten times more rapid than nDNA (Boonphakdee and Sawangwong, 2008) have made mitochondrial more preferable as genetic marker.

Arif and Khan (2009) suggested that the order of conservative mtDNA genes starts with 12S rRNA as the most conserved region then followed by 16S rRNA, cytb and D-loop. This implies that the 12S rRNA gene harbors the most conserved sequence and D-loop serves as the fastest evolving gene. The conserved region is useful as molecular marker for different purposes such as species identification, delimitation and population distribution studies. The different evolutionary rates occupied by genes are important in resolving taxonomic at categorical levels (Arif and Khan, 2009; Arif *et al*., 2011).

1.5.2 Mitochondrial DNA ribosomal genes

Animal mitochondria consist of small and large ribosomal subunits (Pel and Grivell, 1994; Koc *et al*., 2001). The small subunits enclose 12S rRNA and 30 ribosomal proteins whereas the large subunits contain 16S rRNA and 50 ribosomal proteins (Cahill *et al*., 1995; Koc *et al*., 2001). The mitochondrial ribosomal RNA (rRNAs) are encoded by the mitochondrial DNA itself, while the proteins ribosomal are encoded by nuclear DNA then later transported into the ribosome (Sylvester *et al*., 2004). The mtDNA ribosomal genes are non-protein coding genes which are important for cellular energy homeostasis and to ensure the function of control region or D-loop during replication and transcription of mitochondria (Abhyankar, 2009). The ribosomal RNA genes in animal mitochondria exhibits both conserved and variation regions also shares similarity including structures roles in organisms although there are countless interspecific and intraspecific variations (Yang *et al*., 2014). The length mutations are more frequent in ribosomal RNA than in protein coding genes. Hence, the species-specific primers of mtDNA rRNA have been applied broadly in numerous species including mammals, birds and shrimps (Yang *et al*., 2014).

1.5.2(a) MtDNA 12S rRNA gene

The mtDNA 12S rRNA gene takes up 1/16 of whole mitochondrial genome with nucleotide size in vertebrates ranging from 819-975 bp (Meyer, 1993; Rasmussen and Morrissey; 2008). The 12S rRNA gene located between $tRNA^{Phe}$ and $tRNA^{Val}$ (Prosdocimi *et al*., 2011). This gene has been characterised as highly conserved region and exhibits a slower evolution rate than other genes in mtDNA (Arif and Khan, 2009; Dudu *et al*., 2011). Its conserved feature contributes to its effectiveness as a universal primer to amplify various organisms such as insects and avian (Gupta *et al*., 2008;

Yang *et al*., 2014) as well as wildlife species (Rajput *et al*., 2013; Arif *et al*., 2011). The mtDNA 12S rRNA has also been regarded as the most powerful tool in more recent lineage evolutionary (Dudu *et al*., 2011) and in tracing moderate to even earlier divergence history (Patwardhan *et al*., 2014). It is also useful in resolving the phylogeny at various taxonomic levels, such as family, genera and species (Dudu *et al*., 2011). According to Mandal *et al*. (2014), due to its highly conserved region, this gene has been used to resolve the genetic diversity of insect up to phyla levels. It is also applied to expound the phylogeny of diverse organisms including laniatorid (Sharma and Giribet, 2009), scleractinian corals (Chen *et al*., 2002), fish species (Li and Orti, 2007) and mammalian (Douzery and Catzeflis, 1995).

The 12S rRNA gene also exhibits a significant intra-specific and inter-specific mutations which are useful for species distinction, population and evolutionary studies of marine fish species (Jin *et al*., 2013). Due to these characteristics, mtDNA ribosomal gene has been proposed as suitable genetic marker for food authentication (Comesana *et al*., 2003; Rasmussen and Morrissey, 2008; Khallaf *et al*., 2017). It also offers an advantages in identification across diverse taxonomic order such as Perciformes, Siluriformes, Characiformes and Osteoglossiformes (Comesana *et al*., 2003; Ardura *et al*., 2010; Khallaf *et al*., 2017).

Cespedes *et al*. (2000) applied PCR-RFLP of 12S rRNA to differentiate a closely related flatfish species due to possible fraudulent substitutions of *Solea solea* with *Reinhardtius hippoglossoides*. Nevertheless, the highly conserved region and low sequence divergence of 12S rRNA may lower the resolution in species delimitation of certain closely related species. Cawthorn *et al*. (2012) has studied a various species of order Actinopterygii and revealed that the 12S rRNA gene is unable to differentiate a particular closely related species of the same genus. Due to its insufficient nucleotide

variations, they concluded that the 12S RNA gene is not suitable to be used independently to avoid erroneous result. This is in concordant with other previous studies done by Arif and Khan (2009), Arif *et al*. (2011) and Mandal *et al*. (2014) stated that the 12S rRNA gene is more feasible for higher categorical levels.

1.5.2(b) MtDNA 16S rRNA gene

The 16S rRNA gene is located between $tRNA^{Val}$ and $tRNA^{Leu}$ (Prosdocimi *et al*., 2011) occupying 1/10 of the whole mitochondrial genome with nucleotide size ranging from 1571-1640 bp in vertebrates (Meyer, 1993; Rasmussen and Morrissey; 2008). According to Meyer (1993), the 16S rRNA region possesses more length mutations than the 12S rRNA, hence creating a difficulty in aligning sequences of distantly related species. However, this gene showed an outstanding performance by producing an accurate and specific result for closely related species with similarity nearly 99% due to lower intraspecific variation compared to protein coding genes, thus, making its suitable for phylogenetic studies (Baharum and Nurdalila, 2012).

Baharum and Nurdalila (2012) reported that the stability and specificity of 16S rRNA gene as a molecular marker in species identification, as it offers the most direct approach and is successfully applied in species description as well as in phylogenetic studies of bacteria, marine animals and plant (Baharum and Nurdalila, 2012). Kochzius *et al*. (2010) has mentioned that the 16S rRNA gene is the most widely used genetic marker for fish species identification, fisheries and seafood control and ichthyoplankton identification.

However, Kochzius (2010) reported low resolution of 16S rRNA gene contributes to the failure of distinguishing closely related species of gurnard fish (*Chelidonichthys lucernus*, *Eutrigla gurnardus* and *Trigloporus lastoviza*) and flatfish

species (*Pleuronectes platessa* and *Platichthys flesus*). The lower mutation rate in 16S rRNA gene as compared to COI and cytb genes is the principal factor influencing species differentiation (Kochzius, 2010). Wang *et al*. (2017) found a higher inter-specific variation compared to intra-specific variation in 16S rRNA gene when it is used as DNA barcoding for cutlassfishes. Meanwhile, Ferri *et al*. (2009) mentioned that the presence of insertion and deletion in this gene can complicate the species identification and taxonomic studies.

1.5.3 Cytochrome b (Cytb) gene

Cytochrome b (cytb) gene is one of the mtDNA protein coding genes located between $tRNA^{Thr}$ and $tRNA^{Glu}$ (Talbot and Shields, 1996; Pereira, 2000). The sequence size for cytb region is 1140 bp in length (Linacre and Lee, 2016). The cytb subunit is encoded by mtDNA unlike other protein coding subunits which are encoded by the nucleus (Anderson *et al*., 1981; Blakely *et al*., 2005). This gene is responsible for the production of cytochrome b protein which is part of the electron transport chain in oxidative phosphorylation process to generate ATP (adenosine triphosphate) energy (Linacre and Lee, 2016).

The cytb region has been reported to exhibit a rapid evolutionary rate compared to ribosomal genes which makes it as a powerful marker for evolutionary history study involving lower taxonomic levels such as families, genera and species (Arif *et al*., 2011). Other than that, this gene is also widely used as marker for biodiversity analysis, conservation management and planning (Arif and Khan, 2009). For species delimitation and phylogenetic construction, cytb gene showed more accurate resolution compared to the COI gene (Nicholas *et al*., 2012). Hence, Cawthorn *et al*. (2012) has recommended cytb gene to be used in discrimination of closely related

species. The cytb gene has been used also for species detection from unknown samples (Hsieh *et al*., 2005), raw and processed food products (Meyer, 1994; Murugaiah *et al*., 2015) which are in form of comminuted, mixed, degraded and old specimen. Jain *et al* (2007) has reported reliability and specificity of cytb gene for meat authentication from grounded and mixed samples especially in those involving issues related to fraudulent, allergies and religious purposes.

The cytb gene exhibits both variable and conserved regions which make it to be frequently used in systematic and phylogenetic studies among fish family Cichlidae (Farias *et al*., 2001). As stated by Meyer (1994), due to the extensive utilisation, cytb gene probably has become the standard phylogenetic marker with the assumption that it is sufficiently variable and conserved to resolve respective deep phylogeny among distantly related species and population ambiguity (Meyer, 1994). Although the cytb gene has been suggested as the most powerful phylogenetic marker in resolving closely related species, however, it produced irresolute result for the deeper phylogeny and evolutionary (Patwardhan *et al*., 2014). With the statement that no gene can work independently without a glitch, Meyer (1994) has listed several drawbacks of cytochrome b gene as genetic marker which include base compositional biases that would limit the actual variation and dearth of variation at the first and second codon positions resulting in information deficiency at deeper phylogeny. The functionality of cytb gene is also lineage-dependent (Patwardhan *et al*., 2014), as different lineages of organism show different evolutionary rate of cytb gene, for an example rapid evolution is seen in salamanders than in frogs (Graybeal, 1993; Meyer, 1994).

1.5.4 Cytochrome oxidase I (COI) gene

Cytochrome oxidase I (COI) also known as DNA barcoding, is a protein coding gene found in mtDNA and involved in electron transport chain (Meyer, 1993). This gene has been suggested as the universal marker for species identification by Paul D. N. Herbert in 2003 (Bucklin *et al*., 2011) due to its rapid and accurate identification using standard universal gene sequence (Hebert and Gregory, 2005; Chauhan and Rajiv, 2010). This region is widely applied for the identification of many animal species due to the fact of highly conserved DNA and low degree of variation.

DNA barcoding is a novel technology which uses a universal region of COI gene based on the assumption that the genetic variation between species exceeds within species. As stated by Vences *et al.* (2005), species identification through DNA barcoding fulfills five criteria such as consisting conserved sequence to allow amplification, able to discriminate species, able to amplify under variable conditions, can be sequenced and carries enough phylogenetic information for species determination (Vences *et al*., 2005). Rapid variation in the COI gene is observed between species than within species as compared to cytb gene and has been suggested to be suitable for animal species (Zhang and Hanner, 2012) and processed food identifications (Bucklin *et al*., 2011). However, this gene shows to be poorly resolved for phylogenetic study (Bucklin *et al*., 2011).

1.5.5 Other mtDNA genes

There are other mtDNA genes that are also utilised as genetic marker such as control regions or D-loop and nicotinamide adenine dinucleotide (NADH) dehydrogenase. The control region or D-loop is involved in controlling the replication and transcription of mitochondria (Abhyankar *et al*., 2009) and is also responsible for the mitochondrial

genome size variances in organisms (Pereira, 2000). The control region is also referred as the most variable region and the fastest evolving mtDNA region (Jiang *et al*., 2011). In contrast to 12S rRNA and 16S rRNA genes (Linacre and Tobe 2011), this gene is rarely applied for species identification but is frequently used in intra-species distinction (Clifford *et al*., 2004; Zhang *et al*., 2006). Previous studies have shown that the D-loop region has been used in different fields including wildlife investigation, population structure and genetic diversity studies of Chinese water deer (Hu *et al*., 2006) and black muntjac (Wu *et al*., 2006). However, Panday *et al*. (2014) has stated that this region is less informative for animal species identification due to its back and parallel mutations which limits its functions as genetic marker. Besides that, the insertion and deletion in this region also complicates the feasibility of this gene in resolving deeper phylogeny thus not preferable to be applied in phylogeographic and population studies (Vila and Bjorklund, 2004).

The NADH regions such as ND5 and ND6 have been suggested to be supportive for phylogenetic construction particularly for fishes (Miya and Nishida, 2000). The combined ND5 and ND6 genes have been implemented in phylogenetic study of family Leiognathidae and the result had showed a completely resolved phylogenetic tree supported with high bootstrap value (Ikejima *et al*., 2004). Jiang *et al*. (2011) used ND4 together with control region, cytb and COI genes to study the genetic relationship among five Asian freshwater turtle species. Meanwhile, Mayer *et al*. (2007) has sequenced 900 bp of the ND1 gene from 534 bats of Western Palearctic regions to study their phylogeny and diversity. Politov *et al*. (2000) has recommended ND1 subunit as the most effective marker for differentiating among Palearctic coregonid fish species. However, this gene showed to be more prominent at inter-specific level and less suitable for intra-specific level as the variation is less definite.

Urantowka *et al*. (2017) has stated that the problems in phylogenetic analyses inferred from single genes of ND2, ND6 and cytb are due to disagreement of methods and severe variances accounted from the phylogenetic trees.

1.6 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was invented by Kary Mullis in 1984 and later he was conferred for Nobel Prize in chemistry (Joshi and Deshpande, 2010). Bartlett and Stirling (2003) stated that the original concept of PCR is a combination of several components which are referred to the synthesis and uses of oligonucleotides to catalyse a new and specific DNA region copies using DNA polymerase. The PCR technique allows scientists from many fields to amplify DNA into millions copies of DNA starting from a minute sample (Hadidi *et al*., 2003) and poor DNA quality (Erlich, 1989; Joshi and Deshpande, 2010). The PCR technique gathers the interest of scientists from different fields to implement due to its simple, rapid, low cost (Joshi and Deshpande, 2010), highly sensitive and specific results (Hadidi *et al*., 2003; Garibyan and Avashia, 2013). This technique also has benefitted the genomic and evolutionary studies as well as in forensic fields (Hadidi *et al*., 2003). The basic concept of PCR method involves denaturation of DNA target, annealing of forward and reverse primers and elongation of nucleotides (Erlich, 1989). The PCR product can be detected through staining methods either by chemical dye such as ethidium bromide or fluorescent dyed primers (Garibyan and Avashia, 2013).

1.7 DNA sequencing

DNA sequencing is defined as a method used to determine the order of nucleotides in a DNA strand (Griffiths, 2015). DNA sequencing benefits various studies such as molecular biology, genetics, forensic science, archaeology and biotechnology (Franca *et al*., 2002). As stated by Griffiths (2015), the sequenced nucleotides are important to

understand the function of particular gene as well as in evolutionary history (Griffiths, 2015).

The development of sequencing method began in February 1977 by Maxam-Gilbert (Guzvic, 2013). Sanger sequencing method was introduced by Frederick Sanger and his colleagues (Guzvic, 2013) when they successfully sequenced the first DNA genome of bacteriophage ϕX174 or PhiX (Christina, 2015; Heather and Chain, 2016). Maxam and Gilbert developed DNA sequencing based on chemical degradation, which therefore limited their applications and became less popular as it relies on toxic chemicals (Obenrader, 2003). This is in contrast to Sanger method which is based on the selective incorporation of chain terminator dideoxynucleotides by DNA polymerase during in vitro DNA replication (Obenrader, 2003). The Sanger method is more frequently used in molecular field and eventually has become a standard method in DNA sequencing (Obenrader, 2003).

1.8 Significance of the study

Fish species represents the most reported group of aquatic metazoan from freshwater or marine water (Bucklin *et al*., 2011). Huge number of marine fish species had showed diversity in shapes, colors and sizes. Traditionally, fish identification was made based on morphological features which includes osteological characters (Luther, 1977), eyes and retinas (Miyazaki and Kobayashi, 2015), fin rays and gill rakers (Uiblein and Gouws, 2015) which can only be performed by an expert taxonomist. The recognition of fish species is not only of interest for taxonomist and scientists, but is also required for studies of natural history and ecology, fisheries management and authentication of food products. Therefore, a simple and standardised method is needed to provide an easier and appreciative method for all individuals from various fields. Other concerns based on morphological characterisations are due to high diversity and morphology

plasticity of fish species which may contribute to difficulty or false identification of the species since species with plasticity occupies diverse developmental stages (Hebert *et al*., 2003; Pires and Marinoni, 2010). In addition, identification based on morphological features requires a complete structure of species to be correctly identified. Therefore, damage in morphological characteristics of fish species directly contributes to difficulty and complicated species identification due to incomplete information displayed by the species (Armani *et al*., 2011). Moreover, small intra-specific variations that are observed in some species also restrict the identification through physical features since similar morphological appearances will be observed (Teletchea, 2009). Due to these arise issues, the taxonomy solely based on the morphological method is problematic and prone to misidentification as well as incorrect classification. With the confusion from variety common names of each fish species for each state, the identification of fish species has become more complicated. Phylogenetic studies of marine fishes in Malaysia, is not yet scrutinised. Previous studies mostly focused on specific species, genus or family only rather than on the large class Actinopterygii. Introduction of molecular methods has enlightened the traditional identification method by providing of unbiased information for each species. Current study uses three mtDNA genes namely 12S rRNA, 16SrRNA and cytochrome b as genetic markers to characterise the fish species through phylogenetic relationship within Class Actinopterygii.

1.9 Objectives

1.9.1 General objective

To study marine fish species using mitochondrial DNA marker genes

1.9.2 Specific objectives

1. To investigate the selected marine fish species through partial sequence of mitochondrial DNA 12S rRNA, 16S rRNA and cytochrome b genes.
2. To identify the nucleotide polymorphisms observed in the fish species using BioEdit software.
3. To infer phylogenetic relationships among the studied fish species based on mtDNA 12S rRNA, 16S rRNA and cytochrome b genes.