MOLECULAR IDENTIFICATION OF FISH LARVAE IN MERBOK RIVER THROUGH DNA METABARCODING

NORLI FAUZANI BINTI MOHD ABU HASSAN ALSHARI

UNIVERSITI SAINS MALAYSIA

2020

MOLECULAR IDENTIFICATION OF FISH LARVAE IN MERBOK RIVER THROUGH DNA METABARCODING

by

NORLI FAUZANI BINTI MOHD ABU HASSAN ALSHARI

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

July 2020

ACKNOWLEDGEMENT

Praise be to Allah, the Most Gracious and Most Merciful, for His blessing which keep me moving on to finish my study. This thesis became a reality with the kind support and help from many wonderful people whom without their supports, none of this would be possible. I would like to extend my immeasurable appreciation to the following:

My sincere, utmost respect and heartfelt gratefulness for my both supervisors, Dr Mohd Ghows bin Mohd Azzam and Prof. Dr Siti Azizah binti Mohd Nor for their continuous support of my M.Sc. study and research, for their patience, never-ending guidance, understanding, encouragement and motivation for this long journey. May Allah shower His constant blessings on both of you and your beloved ones in return. This thesis would not have been accomplished without the support from Universiti Sains Malaysia for providing the research fund (1001/PBIOLOGI/870018) and scholarship (Program Pembiyaan MyMaster Kementerian Pendidikan Tinggi Malaysia) for personal financial assistance during my study.

My special thanks also go to Dr Alex Foong Choon Pin and Dr Sebastien Lavoué for their invaluable advices, spared times and helps for my master study despite their own responsibilities. The journey of my study is incomplete without these precious Lab 308 sisters and brothers, who coloured my plain life, shared the laughs and tears together for almost four years moments in the lab. I will miss all of you. To my favourite partner in crimes, Kak Noorul, Zu, Mel and Danial thank you for being the reasons of my happy moments. Million thanks to Daus, Abang Fadli, Abang Jam, Kak Adel, Kak Nurul, Fong, Kak Zafirah and Suzye, and best friends (Fariza, Kak Nadia, Fatin and Kak Idah), not to forget Kak Fatimah and Azali, my analysis ‘sifu’, En. Rahman and En. Haris from FRI Batu Maung and FRI Kampung Aceh without

whom I may not understand how to do my works properly, and everyone that might have cross my path. Thank you everyone for all your invaluable advice and patience while helping me through my thesis.

My profound love to my parents, Mohd Abu Hassan Alshari Bin Samsuri and Joraidah Binti Abd. Ghafar and to my siblings, Kakak and Adib. Your waiting, supports and sacrifices during this master journey will be in my heart forever, and I love you. To my three fur brothers, Awang, Miki and Leo, thank you for being my most favourite stress reliever. Finally, this thesis will not be completed without so much love from everyone that is important in my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENT…… ii

TABLE OF CONTENTS……... iv

LIST OF TABLES…… viii

LIST OF FIGURES……... x

LIST OF SYMBOLS AND ABBREVIATIONS…… xii

LIST OF APPENDICES…… xiv

ABSTRAK…….. xv

ABSTRACT…… xvii

CHAPTER 1 INTRODUCTION…… 1

1.1 General introduction……... 1

1.2 Problem statement…….. 5

1.3 Objectives……... 6

CHAPTER 2 LITERATURE REVIEW…… 7

2.1 The Merbok Estuary: its importance in economy and ecosystem services……... 7

2.2 Inter-relationship of fish larvae (ichthyoplankton) and the mangrove ecosystem…….. 12

2.2.1 Habitat and distribution of fish larvae……. 13

2.2.2 Identification of larval fishes…….. 15

2.2.2(a) Importance of fish larval identification…….. 15

2.2.2(b) Identification tools for fish larvae…….. 16

2.3 The mitochondrial DNA (mtDNA) as a taxonomic tool……. 18

2.4 DNA Metabarcoding…….. 20

2.4.1 Introduction of DNA metabarcoding…….. 20

2.4.2 Application of DNA metabarcoding……... 23

2.4.3 Challenges of DNA metabarcoding…… 24

CHAPTER 3 MATERIALS AND METHODS……. 28

3.1 Establishment of 12S rRNA gene database for fishes of Merbok River……... 28

3.1.1 Sample collection, identification and preservation……. 28

3.1.2 Genomic DNA extraction……... 30

3.1.3 Amplification of 12S rRNA and cytochrome oxidase subunit I……... 31

3.1.4 Sequence analyses of COI and 12S rRNA genes……. 33

3.2 Molecular identification of fish larvae through DNA metabarcoding…… 33

3.2.1 Fish larvae sample collection…….. 33

3.2.2 Genomic DNA extraction of fish larvae…….. 38

3.2.3 Genomic DNA purification……. 39

3.2.4 Metabarcoding through COI and 12S rRNA genes……. 40

3.2.5 Next-generation sequencing (NGS) and bioinformatics analysis……... 41

3.2.5(a) NGS data filtering…….. 44

3.2.5(b) *De novo* assembly…….. 45

3.2.5(c) Taxonomic assignments……. 46

3.2.5(d) Diversity analysis……... 46

3.3 Verification of metabarcoding larval species identification using species-specific primers…… 48

3.3.1 Primer design…….. 48

3.3.2 PCR amplification of 15 species-specific primer pairs…… 50

3.3.3 Sequence analysis: Nucleotide alignment…….. 51

CHAPTER 4 RESULTS…….. 52

4.1 Establishment of Merbok River Mitochondrial 12S rRNA gene database…….. 52

4.2 Identification of Merbok River fish larvae species using DNA metabarcoding approach……. 52

4.2.1 CTAB DNA extraction and purification……. 52

4.2.2 Mitochondrial DNA (mtDNA) amplification……. 53

4.2.3 Next-generation sequencing (NGS) statistics of fish larvae… 55

4.2.4 *De novo* assembly……... 56

4.2.5 Taxonomic assignment of fish larvae in Merbok River using DNA metabarcoding approach…… 57

4.2.5(a) Species diversity of fish larvae in St1…… 60

4.2.5(b) Species diversity of fish larvae in St2……. 62

4.2.5(c) Species diversity of fish larvae in St3……. 66

4.2.5(d) Species diversity of fish larvae in St4……. 69

4.2.5(e) Overall species diversity of fish larvae in Merbok River……. 73

4.2.6 Comparison of larval fish diversity among four stations along Merbok River…… 77

4.2.7 Fish larvae identification based on newly designed primer…. 79

4.2.7(a) The newly designed primers of COI for species verification……. 79

4.2.7(b) Taxonomic verification of fish larvae……. 83

CHAPTER 5 DISCUSSION……... 85

5.1 Adult fish sequences as a database for larval fish identification…….. 85

5.2 Larval fish species occurrence and diversity in Merbok River… 85

5.3 The COI and 12S rRNA gene markers for species identification…….. 92

5.4 Species-specific primer pairs to validate DNA metabarcoding data……. 94

5.5 Contrasting morphological with molecular data of fish diversity in Merbok River……... 97

5.6 The advantages and disadvantages of pooling samples for DNA metabarcoding…… 99

CHAPTER 6 CONCLUSION……. 101

REFERENCES……... 103

APPENDICES

LIST OF PROCEEDINGS

LIST OF TABLES

		Page
Figure 3.1	List of 22 fish species that were morphologically identified based on previous studies and without 12S rRNA reference database…..	29
Table 3.2	PCR amplification conditions for 12S rRNA and COI genes…….	32
Table 3.3	Sampling stations and description of the sampling locations…….	34
Table 3.4	Primer pairs used for mitochondrial DNA amplification using REPLI-g Mitochondrial DNA Amplification Kit……....	41
Table 3.5	List of samples for NGS and bioinformatic analysis……..	32
Table 3.6	List of 15 fish species randomly selected from DNA metabarcoding result for specie-specific primer design…….....	49
Table 3.7	PCR conditions used to amplify all 15 species-specific primer pairs……..	50
Table 4.1	Concentration (ng/uL) and purity (A_{260}/A_{280} ratio) of the purified genomic DNA using Wizard® SV Gel and PCR Clean-Up System (Promega, USA)……..	53
Table 4.2	The concentration (ng/µL) and purity (A260/280) of amplified mitochondrial DNA (mtDNA)…….	55
Table 4.3	Number of raw reads, number of trimmed reads, average GC contents after quality trimming of samples from St1, St2, St3 and St4…….	56
Table 4.4	Summary of assembled scaffolds statistics using MEGAHIT (v 1.0.2)…….	56
Table 4.5	Number of scaffolds annotated to COI and 12S rRNA genes that were assigned to fish larvae species with ≥90% similarity identity based on NCBI database……..	58
Table 4.6	Family, species and scaffold abundance of fish larvae identified in St1 of Merbok River…….	62
Table 4.7	Family, species and scaffold abundance of fish larvae identified in St2 of Merbok River…….	65
Table 4.8	Family, species and scaffold abundance of fish larvae identified in St3 of Merbok River…….	68

Table 4.9	Family, species and scaffold abundance of fish larvae identified in St4 of Merbok River…….	71
Table 4.10	The species occurrence at four stations along the Merbok River based on combined metabarcoding analysis of COI (▲) and 12S rRNA (◊) genes…….	74
Table 4.11	Comparison of diversity index, species evenness and species richness in St1, St2, St3 and St4 samples……..	77
Table 4.12	List of newly designed species-specific primer pairs with description of its primer length, product size, GC content (%) and melting temperature (°C). The primers were designed using the online software Primer3Plus following standard criteria for primer design. The successfully amplified primer pairs were marked as '√'…….	80
Table 4.13	List of identified fish larvae species from each sample based on newly designed primer pairs…….	84

LIST OF FIGURES

		Page
Figure 2.1	Division of land uses in Merbok area, which shows the coastal area of Straits of Malacca, Merbok River estuary and its small tributaries, mangrove forest areas, human activities areas, agriculture areas and aquaculture area (Ismail *et al*., 2018)……	11
Figure 2.2	Complete mitochondrial genome of fish (*Atule mate*, family Carangidae) (Li et al., 2016). The circular mitochondrial genome is made up of 37 genes: 12S rRNA and 16S rRNA genes are the two mitochondrial ribosomal RNA genes; ND 1-6 indicate gene of the NADH dehydrogenase subunits 1-6; COX1, COX2 and COX3 are cytochrome oxidase subunit I, II and III; cyt*b* is cytochrome *b*; ATP8 and ATP6 are ATP synthase subunit 8 and ATP synthase subunit 6; 22 transfer RNA (tRNAs); and D-loop or control region…….	20
Figure 2.3	Workflow of DNA metabarcoding method……	23
Figure 3.1	Sampling locations along the Merbok River (Kedah) namely, St1: Sungai Lalang; St2: Sungai Semeling; St3.1: Sungai Keluang; St3.2: Sungai Teluk Wang; St4.1: Sungai Gelam and St4.2: Sungai Terus……..	36
Figure 3.2	(a) Hand scoop nets with 500 µm mesh size used in (b) conducting the samples collection……	37
Figure 3.3	Flowchart of library preparation for next-generation sequencing (NGS)…….	43
Figure 3.4	The raw data provided in FASTQ file format that include the detailed read sequences and the read quality information……..	44
Figure 3.5	Flowchart of the bioinformatics analysis pipeline…….	47
Figure 4.1	Total genomic DNA extracted from fish larvae samples of the six sampling stations (From left lane (Lane 1: 1 kb Plus DNA ladder); to the right lane (Lane 2-7: pooled fish larvae DNA sample from St1 to St4.2 after purification))……...	53
Figure 4.2	Mitochondrial DNA (mtDNA) enrichment step using REPLI-g Mitochondrial DNA Amplification Kit (QIAGEN, Germany) (From left lane (Lane 1: 1 kb Plus DNA ladder); to the right lane (Lane 2-7: amplified mtDNA))……..	54
Figure 4.3	Number of scaffolds assigned to COI, 12S rRNA, other mtDNA genes and unclassified scaffold……..	57

Figure 4.4 Assignment of fish larvae species of each family through DNA metabarcoding method based on COI and 12S rRNA scaffolds from combined St1, St2, St3 and St4 of Merbok River…….. 59

Figure 4.5 Assignment of fish larvae to species based on COI and 12S rRNA scaffolds in St1……... 61

Figure 4.6 Assignment of fish larvae to species based on COI and 12S rRNA scaffolds in St2……... 64

Figure 4.7 Assignment of fish larvae to species based on COI and 12S rRNA scaffolds in St3……... 67

Figure 4.8 Assignment of fish larvae to species based on COI and 12S rRNA scaffolds in St4……... 70

Figure 4.9 (a) Dendrogram of Bray-Curtis similarity showing community similarity of larval fish species at all four stations of Merbok River, and (b) Bray-Curtis similarity plotted by two-dimensional non-metric multidimenstional scaling (NMDS)…… 78

Figure 4.10 The successfully amplified nine species-specific primer pairs: a) From left lane to the right lane is M (P5), sample St2-P5 to St4-P2, followed by M (P2) and L; b) from left lane is L, M (P14), sample St1-P14 to St4-P14]; c) from left lane is sample St2-P4 and St3-P4, M (P4) to M (P12), followed by two empty wells, sample St2-P6 to St4-P7, L, sample St4-P12 and St3-P10; and d) from left lane is sample St2-P17 to St4-P17, L and M (P17); where M is negative control, P5, P2, P14, P4, P6, P7, P10, P12 and P17 are the primer pairs, and L (ladder) is 100 bp DNA ladder…….. 82

LIST OF SYMBOLS AND ABBREVIATIONS

°C	Degree Celcius
µL	Microlitre
mg/L	Milligram per litre
ppt	Part per thousand
×G	Times gravity or relative centrifugal force (RCF)
12S rRNA	12S ribosomal ribonucleic acid
BLAST	Basic Alignment Search Tool
BOLD	Barcode of life data system
COI	Cytochrome oxidase subunit I
CTAB	Cetyltrimethyl ammonium bromide
DO	Dissolved oxygen
DoF	Department of Fisheries
eDNA	Environmental DNA
E-value	Expect value of hits
FAO	Food and Agriculture Organization
FRI	Fisheries Research Institute
GenBank	Genetic sequence database
min	minute
NCBI	National Centre for Biotechnology Institute
NGS	Next generation sequencing
NMDS	Nonmetric-multidimensional scaling
PCR	Polymerase chain reaction
SAL	Salinity

TEMP	Temperature
TURB	Turbidity
WD	Water depth

LIST OF APPENDICES

APPENDIX A — Descriptions of the parameters used for *de novo* assembly using MEGAHIT (v.1.0.2) assembler.

APPENDIX B — Bioinformatics tools used in the bioinformatics analysis and the description of the tools used for each analysis.

APPENDIX C — Workflow of the whole process of fish larvae identification using DNA metabarcoding method.

APPENDIX D — Representatives of 22 fish species collected in the study to complete 12S rRNA reference database. 1) *Scomberoides tol*, 2) *Plotosus canius*, 3) *Cynglossus lingua*, 4) *Cynoglossus bilineatus*, 5) *Arius venosus*, 6) *Plicofollis platystomus*, 7) *Nemapteryx caelata*, 8) *Nibea soldado*, 9) *Secutor ruconius*, 10) *Leiognathus brevirostris*, 11) *Scatophagus argus*, 12) *Otolithes ruber*, 13) *Pomadasys kaakan*, 14) *Brevitrygon walga*, 15) *Johnius amblycephalus*, 16) *Johnius belangerii*, 17) *Batrachomoeus trispinosus*, 18) *Siganus javus*, 19) *Scomberoides commersonnianus*, 20, *Carangoides praeustus*, 21) *Strongylura strongylura* and 22) *Platycephalus indicus*.

PENGECAMAN MOLEKULAR LARVA IKAN DI SUNGAI MERBOK MELALUI METAPENGEKODAN PALANG DNA

ABSTRAK

Kajian ini bertujuan untuk menjelaskan kepelbagaian dan taburan larva ikan di Sungai Merbok melalui pengecaman molekular larva ikan menggunakan pendekatan metapengekodan palang DNA. Di dalam fasa pertama kajian ini, satu pangkalan data molekular untuk spesies ikan Sungai Merbok telah dilengkapkan. Ini telah dijalankan melalui pengembangan rujukan pangkalan data yang telah wujud bagi gen 12S RNA ribosom (12S rRNA) untuk 120 spesies yang telah didokumentasi melalui kaedah pengecaman morfologi di dalam kajian-kajian terdahulu, bersama dengan penggabungan pangkalan data sitokrom oksidase subunit 1 (COI). Di dalam fasa kedua, satu kajian metapengekodan palang telah dijalankan terhadap sampel larva ikan yang didapati dari empat kawasan (stesen) (St1, St2, St3 dan St4), mewakili zon hulu sehingga ke zon hilir sungai yang mewakili kecerunan saliniti rendah ke saliniti tinggi. DNA genom larva ikan telah diekstrak, diamplifikasi dan diperkaya dengan menggunakan kit REPLI-g Mitochondrial DNA. Penjujukan generasi seterusnya (NGS) ke atas genom lengkap telah dilakukan menggunakan platform Illumina MiSeq, disusuli analisis bioinformatik. Analisis BLAST terhadap himpunan *de novo* perancah genom untuk empat sampel ini telah mengesan sebanyak 89 spesies larva ikan, 65 genera dan 41 famili daripada 1071 dan 943 jumlah perancah dimiliki oleh masing-masing gen COI dan 12S rRNA. Melalui jumlah perancah, *Oryzias javanicus* dan *Oryzias dancena* telah menunjukkan kelimpahan tertinggi dan diikuti oleh spesies yang lazim terdapat di kawasan muara seperti *Lutjanus argentimaculatus, Lutjanus malabaricus* dan lain-lain. Terdapat beberapa pengecaman spesies yang tidak

bertindih di antara data metabarkod dan data morfologi yang dicadangkan antara beberapa faktor, disebabkan variasi musim bertelur dan pengaruh persampelan (spesies komersial berbanding liar). Untuk mengesahkan ketepatan data metapengekodan palang DNA, pencetus spesifik spesies gen COI telah direka bentuk bagi 15 spesies yang dipilih secara rawak di kalangan 89 spesies larva ikan. Walau bagaimanapun, pasangan pencetus tersebut gagal untuk mengecam spesies yang disasarkan. Sebaliknya, kesemua pencetus yang telah berjaya diamplifikasi mengecam tiga spesies lain yang juga dikenalpasti menggunakan kaedah metapengekodan palang DNA (*Oryzias javanicus*, *Decapterus maruadsi* dan *Pennahia macrocephalus*). Menariknya, dua lagi spesies yang tidak dikenalpasti oleh metapengekodan palang DNA juga telah dikenalpasti di dalam langkah pengesahan ini iaitu *Ambassis marianus* dan satu spesies yang mempunyai padanan terdekat kepada *Carangoides chrysophrys*. Selain itu, satu spesies udang (*Acetes sibogae*) juga telah dikenalpasti di dalam langkah ini. Semua percanggahan ini boleh dikaitkan dengan kesilapan pengecaman di dalam rujukan pangkalan data dan kontaminasi. Walaupun terdapat pelbagai cabaran, perkembangan kaedah metapengekodan palang DNA untuk larva ikan Sungai Merbok dapat memudahkan pengenalpastian kawasan bertelur spesies ikan dan pola migrasi di dalam muara Sungai Merbok. Hal ini akan membolehkan pendekatan pemuliharaan dan pengurusan perikanan di Sungai Merbok dijalankan secara efektif yang boleh menjadi sistem model kepada kawasan lain. Sebagai kesimpulan, pengenalpastian larva ikan menggunakan kaedah metapengekodan palang DNA perlu diteruskan di kawasan lain untuk pengenalpastian taksonomi yang lebih mendalam.

MOLECULAR IDENTIFICATION OF FISH LARVAE IN MERBOK RIVER THROUGH DNA METABARCODING

ABSTRACT

This study aimed to elucidate the diversity and distribution of fish larvae in Merbok River through molecular identification by the DNA metabarcoding approach. In the first phase of this study, a molecular database for the fish species of Merbok River was completed through the expansion of the existing reference database of 12S ribosomal RNA (rRNA) gene for the 120 documented species morphologically identified in previous studies in combination with existing Cytochrome Oxidase Subunit 1 (COI) database. In the second phase, a metabarcoding study of fish larvae samples obtained from four stations (St1, St2, St3 and St4), representing the upper to lower zone of a lower to higher salinity gradient was conducted. The genomic DNA of the fish larvae were extracted, amplified and enriched by using REPLI-g Mitochondrial DNA kit. Next generation sequencing (NGS) of this whole genome was done using the Illumina MiSeq platform, followed by bioinformatics analysis. A BLAST analysis on the *de novo* assembled scaffolds of the four samples detected a total of 89 fish larvae species, 65 genera and 41 families from 1071 and 943 total genome scaffolds belonging to COI and 12S rRNA genes, respectively. From the total number of scaffolds, *Oryzias javanicus* and *Oryzias dancena* showed the highest abundance and followed by other estuarine species commonly found such as *Lutjanus argentimaculatus, Lutjanus malabaricus* and others. Considerable non-overlapping of species identity between the metabarcoding and morphological data is suggested to be due to variation in spawning season and sampling biasness (commercial compared to wild species) among other factors. To validate the accuracy of the DNA metabarcoding

data, species-specific primers were designed for the COI gene of 15 randomly selected among the 89 fish larvae species. However, the primer pairs failed to identify the targeted species. Instead, all the successfully amplified primers identified two other fish larvae species that had also been identified using DNA metabarcoding method (*Oryzias javanicus*, *Decapterus maruadsi* and *Pennahia macrocephalus*). Interestingly, another three fish species that were not identified by DNA metabarcoding were also identified in this validation step namely *Ambassis marianus* and an unknown species with the closest match to *Carangoides chrysophrys*. Besides that, a shrimp species (*Acetes sibogae*) was also identified at this step. All these discrepancies could be attributed to the misidentified reference sequences in the database and contamination. Despite the challenges, the development of the DNA metabarcoding approach for the fish larvae diversity of the Merbok River could facilitate identification of fish species spawning sites and migration pattern in the Merbok estuary. This would allow for effective conservation and management approach of fisheries in Merbok River which could be a model system for other sites. In conclusion, identification of fish larvae should be continued at other sites using DNA metabarcoding method for deeper taxonomic level identification.

CHAPTER 1

INTRODUCTION

1.1 General introduction

Fish larvae communities, similar to adult fish communities, not only provide understanding of the ecological conditions (Moser and Smith, 1993, Moser, 1996) but also provide information for management of fisheries resources (Kidwai and Amjad, 2001). For example, the identification of fish larvae and their distribution data could inform on fish reproductive biology, spawning seasons, nursery grounds of the fish during the larval stages and the migration routes of the fishes (Kidwai and Amjad, 2001, Jaxion-Harm *et al*., 2012, Deepananda and Arsecularatne, 2013, Frantine-Silva *et al*., 2015). These data can in turn facilitate ecological monitoring, assessment of environmental impacts and, ultimately, development of management and conservation plans for the habitat and its inhabitants (Ooi and Chong, 2011, Frantine-Silva *et al*., 2015).

Mangrove forests are important ecological ecosystems for the survival of many aquatic and terrestrial organisms. They act as nursery and feeding grounds for many commercial fishes and shrimps, not only of marine and brackish water origins, but also freshwater species (Chong *et al*., 1990, Sasekumar *et al*., 1992, Somerfield *et al*., 1998, Baran and Hambrey, 1999, Yusoff *et al*., 2006). Peninsular Malaysia is surrounded by large water bodies with the Straits of Malacca in the west and South China Sea in the east bordering a total coastline of 4800 km (Mazlan *et al*., 2005) supporting a high diversity of mangrove forests. These forests were estimated to cover a total area of 695,000 hectares (ha) in Malaysia at the turn of the century (Romañach *et al*., 2018). However, since then, 12% has been lost due to anthropogenic activities and the area

has decreased to 575,000 ha (Romañach *et al*., 2018). Of these, 97,000 ha are found in Peninsular Malaysia where one of the mangrove forests, the Merbok Permanent Forest Reserve (MPFR) (Chong, 2007) is located.

The MPFR with an area of 4883 ha size is utilized by abundant aquatic fauna including fishes, as nursery grounds as described in previous studies (Mansor *et al*., 2012a, Mansor *et al*., 2012b). This high diversity supports the socio-economy of the local residents mainly involved in artisanal and recreational fisheries, aquaculture and agricultural activities. The water flows from the neighbouring residential, agricultural and aquaculture areas, however, threatens its water quality and, consequently, the distribution and abundance of aquatic organisms (Mansor *et al*., 2012b, Mansor *et al*., 2012c). To evaluate the impacts of such pollution generated by high anthropogenic activities, the Merbok estuarine ecosystem health and its inhabitants, like in other areas, should be closely monitored through various approaches.

One of the important approaches to address the above is knowledge on the community structure of fishes since they are common bio-indicators of ecological conditions of the river (Schiemer, 2000, Deepananda and Arsecularatne, 2013). Based on the importance of the mangrove areas in the fisheries industry, many studies have been done to identify adult fishes inhabiting mangrove estuaries (Chong *et al*., 1990, Sasekumar *et al*., 1992, Sasekumar and Chong, 1998, Faunce and Serafy, 2006, Blaber, 2007). In comparison, there have been limited studies conducted on ichthyoplankton (eggs and fish larvae) and even fewer studies have been done in Malaysia (Ooi, 2012, Neal *et al*., 2012, Muhamad and Ab Rahim, 2014, Rezagholinejad *et al.*, 2016a), with most of them based on morphological examination. Ichthyoplankton and its identification is important especially for fish stocks management and understanding the spawning and migration patterns of fish (Van der Lingen and Huggett, 2003,

Richardson *et al*., 2007, Ko *et al*., 2013). Some related studies, including Ara et al. (2011a), Ara *et al*. (2013), Arshad *et al*. (2012), Ooi and Chong (2011), and Rezagholinejad *et al*. (2016a) have been reported to examine larval fish assemblages, distribution, composition and diversity of fish larvae in seagrass beds of Southwestern Johor, Pendas River estuary, Matang Mangrove Forest Reserve, and mangrove estuary in Marudu Bay. This traditional approach, based on morphology, typically permits identification only up to family level, and at best genus level, due to the lack of comprehensive larval fish identification keys. The methodology is tedious and challenging even to an experienced taxonomist and often prone to misidentifications (Ooi and Chong, 2011, Frantine-Silva *et al*., 2015).

Although the identification of ichthyoplankton still largely relies on traditional morphological keys, molecular techniques are emerging to assist the identification process (Lewis *et al*., 2016). The DNA barcoding approach that was introduced by Hebert *et al.* (2003a) facilitates the identification of many species. This method is based on species level variation at the mitochondrial cytochrome oxidase subunit 1 (COI) gene and has been widely successful in identifying most animal specimens up to species level, including identifying ichthyoplankton (Pegg *et al*., 2006, Victor *et al*., 2009, Hubert *et al*., 2010, Ko *et al*., 2013, Lewis *et al*., 2016, Azmir *et al*., 2017, Collet *et al*., 2018). A prerequisite to link these minute specimens to correct assignment is the need of a reference database. To this end, the generation of a large DNA barcode database based on adult fish species would function as a good reference for ichthyoplankton identification.

The physical separation of individual ichthyoplankton specimens from bulk samples is tedious, more so the taxonomic identification of each specimen. To overcome these issues, concomitant with the development of advanced molecular

technologies and sequencing platforms, the traditional DNA barcode method has evolved into a more advanced technique called DNA metabarcoding. DNA metabarcoding is the barcoding of DNA or eDNA (environmental DNA) that allows for simultaneous identification of many taxa within the same (environmental) sample through next generation sequencing (NGS) tools and rigorous bioinformatics analyses. It utilizes the same reference databases as traditional DNA barcoding but on bulk samples (Taberlet *et al*., 2012b, Cristescu, 2014). The method introduced by Taberlet *et al*. (2012b) allows rapid and easier species identification of bulk samples or environmental samples. Since its development, many studies have utilized the DNA metabarcoding to facilitate simultaneous species level identification of numerous specimens. For example, Maggia *et al.* (2017) and Mariac *et al*. (2018) successfully determined the species composition of several thousand specimens of ichthyoplankton from the Amazon basin.

This study will focus on the diversity and distribution of larvae in the Merbok River, of the MPFR, a mangrove estuarine area that supports numerous commercial fishes. Considering the threat of overharvesting and degradation of habitat due to coastal erosion and conversion of the lands to aquaculture and agricultural sites (Chong, 2006, Jusoff, 2008), it is vital that more active steps are taken to manage the area for sustainable fisheries to the local community. While regulations are in place to manage the adult fishes, very little is known on the larvae diversity and its distribution. This information is vital to fisheries managers to understand the species that are utilizing the area and also the locations they inhabit as their nursery grounds. With this knowledge, fisheries managers can take measures to protect the specific sites. Thus, in complement with the management efforts on the adult fishes, a more comprehensive and holistic management strategies could be implemented.

Adopting the DNA metabarcoding method, the molecular identification (at the species level) of fish larvae in Merbok River is conducted using two mitochondrial genes, the 12S ribosomal RNA (12S rRNA) and the cytochrome oxidase subunit 1 (COI). These two genetic markers have been widely used in previous metabarcoding studies of fishes (Miya *et al*., 2015, Sato *et al*., 2017, Yamamoto *et al*., 2017, Maggia *et al*., 2017, Mariac *et al.*, 2018). Hence, it is hoped that this study will provide useful information on the species diversity of fish larvae utilizing Merbok River and their nursery sites. Such studies have direct application in the management and conservation of the fishery resources and the results are useful as bio-indicator to assess the ecological health of the MPFR. The COI marker is now established as the standard DNA barcode marker in fish. Since the reference database for 12S rRNA is still limited compared to the COI gene, this would first be developed prior to the metabarcoding experimental analysis.

1.2 Problem statement

There have been a few studies on the identification of fish larvae in the estuaries in Malaysia. Furthermore, the documented ones have been almost completely based on morphological identification of the fish larvae (Ara *et al*., 2011a, Ooi and Chong, 2011, Arshad *et al*., 2012, Ara *et al*., 2013, Rezagholinejad *et al*., 2016a). It is well known that identifying such minute organisms to species level is time- and labour-consuming. Previous studies in Merbok River, have been done on the diversity of adult fish (Mansor *et al*., 2012a, Mansor *et al*., 2012b, Mansor *et al*., 2012c) but there is a limited amount of information on fish larvae. Understanding the fish larvae diversity and distribution in an estuary are important to understand how species are utilizing the

area and the locations they inhabit as their nursery grounds. Based on these knowledge gaps, this study is aimed at achieving the following objectives.

1.3 Objectives

The objectives of this study are:

1) To update a molecular identification database for fish in the Merbok River
2) To determine the DNA metabarcoding method for diversity and distribution studies of fish larvae in Merbok River

CHAPTER 2

LITERATURE REVIEW

2.1 The Merbok Estuary: its importance in economy and ecosystem services

Mangrove forests are found in tropical and subtropical regions of the world (Polidoro *et al*., 2010, Giri *et al*., 2011, Gajdzik *et al*., 2014) and occupy the inter-tidal zone between terrestrial and estuarine ecosystems. Malaysia boasts some of the largest mangrove areas within Asia lining the coasts of Peninsular Malaysia and Malaysian Borneo (Sabah and Sarawak) with a total of 695,000 hectares (ha) in the 1970s (Romañach *et al*., 2018). The Merbok estuary which lies in north-west Peninsular Malaysia supports one of the most important mangrove forests in Malaysia known as Sungai Merbok Permanent Forest Reserve. The Merbok estuary lies between latitude 100° 20' 57.33" and longitude 5° 40' 53.47" facing the Straits of Malacca, and between latitude 100° 30' 24.56' and longitude 5° 42' 13.46" in the upper reaches (Mansor *et al*., 2012b), encompassing 4085 ha of mangrove forests (Mohd *et al.*, 2008).

The Merbok River (Sungai Merbok) with a length of 35 km is the main river that connects the small rivers or tributaries within these forests and further inland. It flows through agricultural, fish and shrimp farming areas (aquaculture), and also residential areas (Figure 2.1) (Ong *et al*., 1991, Mansor *et al*., 2012b). The freshwater segment of this river is only a few kilometers long as the seawater intrudes until 30 km of its length. The freshwater discharge into the estuary originates from the small streams along the Merbok River (Ong *et al*., 1991). This ecosystem is highly diverse in terms of its natural floral resources, such as fuel wood, timber, charcoal and resins as well as faunal diversity, similar to other mangrove ecosystems (Jusoff, 2008, Barbier *et al*., 2011). It also provides many ecosystem services such as water purification, carbon sequestration, and contribute to the economy and education

through fisheries, aquaculture, farming tourism, recreational and research activities (Ong *et al*., 1991, Barbier *et al*., 2011). Other important ecological functions of the mangroves are for coastal protection, erosion control and habitat for food, shelter and protection of its aquatic inhabitants (Hookham *et al*., 2014). The mangrove ecosystem is the first defence for protection of the coastline against natural disasters such as tsunamis, coastal erosion, storms and floods (Jusoff, 2008, Mohd *et al*., 2008), as highlighted by the devastating tsunami of 2004.

The role of the Merbok mangroves, similar to other mangroves in supporting fisheries resources is widely known. It is an important habitat for numerous aquatic organisms, many of which are of economic importance such as fishes, shrimps and mollusks (Rönnbäck, 1999, Manson *et al.*, 2005, Chong, 2006). These residents depend on the habitat for food, shelter, protection and as a spawning area like in other mangrove ecosystems (Potter *et al*., 1990, Low *et al.*, 1994, Potter and Hyndes, 1999, Latiff and Faridah-Hanum, 2014). Approximately 80% of commercial fish species are directly or indirectly dependent on mangrove ecosystems around the world (Polidoro *et al*., 2010). Previous studies of the Merbok estuary have recorded a combined total of 120 fish species encompassing marine, marine-estuarine, estuarine and freshwater species (Mansor *et al*., 2012a, Mansor *et al*., 2012b).

The marine fish inhabitants utilize the mangrove ecosystem at various stages and duration of their life cycle. Several families such as Carangidae (e.g.: *Carangoides coeruleopinnatus,* and *Scomberoides tol*), Serranidae (e.g.: *Epinephelus coiodes*) Sillaginidae (e.g.: *Sillago sihama*) and Terapontidae (e.g.: *Terapon jarbua*) are referred to as occasional marine visitors (Day *et al*., 1989, Mansor *et al*., 2012a, Mansor *et al*., 2012b) because only a small proportion of their overall population use estuaries (Potter *et al*., 1990, Whitfield, 1999). Fish from family Mugilidae (e.g.: *Liza*

subviridis, Liza tade and *Liza vaigiensis*) Sciaenidae (e.g.: *Dendrophysa russelii, Johnius belangerii* and *Johnius borneensis*) and Gobiidae (*Acentrogobius audax, Periophthalmus schlosseri* and *Paratrypauchen microcephalus*), are examples of marine-estuarine taxa. They are considered as marine migrants that often use inshore areas and estuaries during the juvenile stage. Adults may also be found here (Potter *et al*., 1990, Whitfield, 1999, Mansor *et al.*, 2012b).

Estuarine residents refer to species that live within estuaries to complete their life cycles (Potter and Hyndes, 1999, Whitfield, 1999, Blaber, 2008). These include species from families Eleotridae (*Butis butis* and *Butis gymnopomus*), Belonidae (*Strongylura strongylura*) and Latidae (*Lates calcarifer*) which are also economically important to the Merbok area (Mansor *et al.*, 2012b). Freshwater species refers to those species that habitually spawn in freshwater, but occasionally migrate into estuaries under favourable conditions (Day *et al*., 1989, Chong, 2007). Some examples of the native freshwater fish species that have been recorded in the Merbok River are from family Eleotridae (*Oxyeleotris marmorata*) and Bagridae (*Hemibagrus nemurus*) (Mansor *et al*., 2012a). The invasive freshwater species from family Cichlidae (*Oreochromis mossambicus*) has also been recorded in Merbok River (Mansor *et al*., 2012b). Most of these fish species as well as invertebrates inhabit the Merbok estuary during at least a single stage of their life cycle (Faezah and Farah, 2011, Hookham *et al*., 2014). Being a mangrove area, it is also a nursery ground for most of its inhabitants including shrimps, crabs and fishes (Ashton and Macintosh, 2002, Mumby *et al*., 2004). Robertson and Blaber (1993), and Blaber (2000) proposed three factors to account for the role of mangroves as an important nursery site; 1) a refuge from predators, 2) an abundance of food resources and 3) a shelter from physical disturbances (Manson *et al*., 2005).

The mangroves ecosystem also helps to increase the biomass of commercially important coral reef fishes in nearby islands (Mumby *et al.*, 2004) and in the case of the Merbok estuary, the surrounding islands of Pulau Bunting, Pulau Songsong and Pulau Bidan. Furthermore, the Merbok area is also important to non-aquatic organisms (Latiff and Faridah-Hanum, 2014). Among these are the occasional or seasonal visitors such as migratory birds en route on their winter journey to warmer areas in the Southern Hemisphere. However, while being an important region of biodiversity and provider of ecosystem services, the anthropogenic activities would (and has) ultimately negatively impact the size, environmental conditions and water quality of the Merbok estuary. Malaysia like other nations is experiencing a loss in its mangroves by 1% or 1282 ha per year since 1990 (Omar *et al*., 2012, Kanniah *et al*., 2015) and current record has estimated its decline to 575,000 ha (Romañach *et al*., 2018).

As in other regions, the major loss of the mangrove forests have been attributed to land conversion mainly for agriculture and aquaculture activities, infrastructure development and natural causes such as coastal erosion (Polidoro *et al*., 2010, Latiff and Faridah-Hanum, 2014). These disturbances and changes would negatively impact the faunal and floral communities that occupy the mangrove ecosystems, such as the reduction of fisheries catch (Manson *et al*., 2005, Jusoff, 2008). Although natural forces are among the reasons for the destruction of mangrove ecosystems, of late, the major causes are due to development of aquaculture farms in coastal areas, especially shrimp aquaculture (Rönnbäck, 1999, Barbier *et al*., 2011). Furthermore, unsustainable forestry practices, illegal harvests, water disposal and pollution from aquaculture and agricultural activities in the nearby farms, and herbicide impacts have also contributed to the degradation of the mangrove ecosystem of Merbok (Chong and Sasekumar, 2002, Chong, 2007).

Figure 2.1 Division of land uses in Merbok area, which shows the coastal area of Straits of Malacca, Merbok River estuary and its small tributaries, mangrove forest areas, human activities areas, agriculture areas and aquaculture area (Ismail *et al*., 2018).

2.2 Inter-relationship of fish larvae (ichthyoplankton) and the mangrove ecosystem

Knowledge on the community structure of fish larvae is very important since they are known to be excellent indicators of ecological conditions of their habitats (Schiemer, 2000). Fish larval communities, similar to adult fish communities provide understanding of the ecological and health conditions of a particular habitat (Moser and Smith, 1993, Moser, 1996) which could be applied for management of fisheries resources (Kidwai and Amjad, 2001). Each fish species has its own reproductive strategy and most fishes are oviparous (Kelso and Rutherford, 1996, Moser, 1996, Ma *et al*., 2012). These oviparous species produce thousands to millions of planktonic eggs and larvae, and release them into the water column for external fertilization to occur within 24 hours (Houde and Solomons, 2009, Osse and Van den Boogaart, 1995). The fertilized eggs are hatched and the fish larvae are released into the water column (Osse and Van den Boogaart, 1995).

The eggs and fish larvae are known as ichthyoplankton (Neira *et al*., 1998, Termvidchakorn and Hortle, 2013). Osse and Van den Boogaart (1995) defined a larva as the period of fish after hatching prior to becoming a juvenile, and then to adults. They are unable to swim freely and have to drift with the water currents (Moser, 1996, Chong, 2007). They are small free-living fish larva and subsequently develop characteristic meristic and morphometric traits and reach reproductive maturity into adults (Neira *et al.*, 1998). Different egg morphological characteristics such as size and shape of the eggs, size and number of oil globules and pigmentation are useful for the identification of fish eggs (Kendall Jr and Ahlstrom, 1984). Morphological diagnostic characteristics of larvae are explained in 2.2.2(b). However, due to their

minute size, such morphological characteristics may not be efficient in differentiating among species.

2.2.1 Habitat and distribution of fish larvae

Fish ecology is one of the most dynamic areas of ichthyology. The understanding on how environmental factors influence the distribution and abundance of fishes is important for fisheries management, protecting aquatic ecosystems, and water quality survey (Moyle and Cech, 2004). Information on the distribution and abundance of ichthyoplankton of an ecosystem could facilitate capture fisheries management and also permit locating shoals of fish and their breeding grounds in the area (Manickasundaram et *al*., 1987). Distribution of fish eggs and larvae vary both temporally and spatially (Kelso and Rutherford, 1996), and also differ among species (Houde and Solomons, 2009). The patterns of distribution of ichthyoplankton often depend on the egg and larval buoyancy, and the behaviour of the fish larvae. Furthermore, the distribution of marine larval invertebrates and fish larvae in the estuarine is also controlled by the habitat condition and environmental factors such as temperature, salinity, dissolved oxygen (DO), pH and turbidity (Whitfield, 1999, Kingsford et *al*., 2002, Moyle and Cech, 2004).

High abundance of foods in the mangrove estuaries especially planktons and low predation pressure play important roles as the biological and physical factors for the abundance and high survival rate of fish larvae (Houde and Solomons, 2009, Ooi and Chong, 2011). It is well known that mangrove estuaries and seagrass beds are among the most favoured habitats for fish including those that are not permanent residents of the estuarine areas. These habitats attract many fish species including

freshwater, estuarine-dependent, coral-reefs as well as pelagic fishes to reside at the estuarine (Nagelkerken *et al*., 2000a, Nagelkerken *et al*., 2000b, Dahlgren *et al*., 2006). Adeel and Pomeroy (2002) estimated that approximately 90% of marine organisms migrate and spend some portion of their life cycles within the mangrove ecosystems.

Many fish species that are found in mangrove estuaries are however, commonly known to be euryhaline, where the adult occurs in marine waters during one phase of their life history (Blaber and Milton, 1990). Adult fish and the early-life stages of many species do not share the same habitat. Larvae and juvenile fishes inhabit mangrove areas as one of their nursery grounds, in addition to seagrasses, estuaries and mudflats (Ara *et al*., 2013, Igulu *et al*., 2014). Most estuary-dependent marine fishes enter these habitats during the juvenile and postflexion larval stage (Bell *et al.*, 1984, Whitfield, 1999, Ooi and Chong, 2011). Distributions of fish larvae within the nursery areas differ among species and most larvae will remain at the upper 200 m of the water column (Houde and Solomons, 2009).

The physical features of the mangroves trees and their surroundings for example the prop roots, pneumatophores and fallen litter of the mangrove trees, soft mud and shallow water provide suitable nursery habitats and refugial space for these juvenile fishes and invertebrates such as decapods (Sheridan and Hays, 2003, Chong, 2007). Jaxion-Harm *et al*. (2012) noted that mangroves, seagrass beds and coral reefs are attractive for many economically important marine fish communities and the loss of these habitats may affect the fisheries resources. Verweij *et al.* (2008) in their findings support a study done by Mumby *et al*. (2004) that tropical coral reef fishes migrate into mangroves and seagrasses during larval and juvenile period before returning to the coral reefs.

2.2.2 Identification of larval fishes

2.2.2(a) Importance of fish larval identification

Fishes are one of the main indicators in the management of the aquatic environment and consequently the commercial fisheries they support. The identification of adult fish alone is not enough to conclude the overall condition of the fisheries resources and the aquatic environment. This should be complemented by data on early life stages which is one of the best ways to provide beneficial information, knowledge and understanding for a better fisheries management (Arshad *et al.*, 2012). Since many fish species utilize the mangrove estuary as spawning, nursery and feeding ground, including at the Merbok estuary (Tzeng and Wang, 1992, Ooi and Chong, 2011, Mansor *et al.*, 2012b), studies on the distribution of the early life stages of a species could facilitate in the understanding of these features (Arshad *et al.*, 2012, Silva *et al.*, 2012, Ávila-Simas *et al*., 2014) as well as understanding the migration routes of its early life stages (Ko *et al.*, 2013). This could assist in strategizing areas for protection and management.

Silva *et al*. (2012) stated that knowledge about the spawning and nursery sites are very important as it could be applied to monitor the fish populations and fishery resources. Monitoring of ichthyoplankton in an ecosystem could assist rapid discovery of the changes that occur on the quantity and biomass of the fisheries stock (Richardson *et al.*, 2007). Knowledge of timing, duration, mode and dispersal of larvae could improve the understanding of the patterns and processes of fish populations, communities and the aquatic ecosystem (Webb *et al.*, 2006). Ichthyoplankton data is also a powerful tool to determine the distribution and diversity of rare and cryptic species in an ecosystem (Richardson *et al*., 2007). Such data is useful guide to propose the establishment of protected areas in order to protect the

native fisheries resources (Ko *et al.*, 2013). However, studies on ichthyoplankton especially in Malaysia estuarine waters are still lacking, due to several factors, the most significant being incomplete taxonomic keys and limited expertise for identification of these minute organisms (Ooi and Chong, 2011).

2.2.2(b) Identification tools for fish larvae

The keys for morphological identification are only available and efficient on certain life stages or gender that makes the identification of fish larvae very challenging (Hebert *et al*., 2003a, Frantine-Silva *et al*., 2015). Each life stage requires a different taxonomic key for species level identifications and for some taxa the keys are incomplete or inaccurate (Hatzenbuhler, 2015). Traditional identification tool using morphological characteristics are time and labour consuming as well as requires high level of taxonomic expertise to achieve the challenging tasks (Hebert *et al*., 2003a, Ooi and Chong, 2011, Azmir *et al*., 2017). Hatzenbuhler (2015) stated that accurate identification of larval fishes requires well-trained taxonomists that are familiar with fish larvae, but even then, there are often conflicting opinions among the experts which reduces the taxonomic certainty. Damaged specimen characteristics can also be a hindrance to accurate identification and taxonomic resolution (Hatzenbuhler, 2015). In addition to that, traditional morphological identification tool often misses cryptic taxa, which are common in many groups, and therefore remains misidentified (Hebert *et al.*, 2003a).

Traditional morphological fish larval taxonomic keys, include the myomeres, meristic counts and morphometric measurements, pigmentation, head spination, and fin formation (Leis and Carson-Ewart, 2000). However, morphological characters are

often not efficient for identification of larval fish especially for those rare and cryptic species (Matarese *et al*., 2011, Ko *et al*., 2013), since most share similar meristics counts, and overlap in morphometrics measurements (Victor *et al*., 2009). Larval fish are often fragile and easily broken during handling, especially during sample collection and preservation. As a result, many morphological characteristics such as fins, body structure and pigmentation pattern which are important for the identification process might be damaged and lost (Valdez-Moreno *et al.*, 2010). The limitations in the traditional tool for identification have led to development of other complementary techniques. An increasingly popular approach is through the use of molecular tools, specifically referred to as DNA barcoding method (Hebert *et al*., 2003a) (further details as given in the Section 2.3 below). Many studies have now revealed the efficacy of this technique for species identification and distinguishing morphologically similar species (Ward *et al.*, 2008, Packer *et al*., 2009).

DNA barcoding refers to a molecular method that is used as molecular species tags based on short, standardized gene regions, typically the mitochondrial cytochrome oxidase subunit 1 gene for animals (Hebert *et al*., 2003a, Hebert and Gregory, 2005). This method has successfully identified bird species (Hebert *et al*., 2004), adult fishes (Hebert et al., 2003a, Ward et al., 2005, Ivanova et al., 2007, Zhang and Hanner, 2011) as well as ichthyoplankton (Pegg *et al*., 2006, Webb *et al*., 2006, Victor *et al*., 2009, Hubert *et al*., 2010, Ko *et al*., 2013, Hubert *et al*., 2015, Lewis *et al*., 2016, Azmir *et al*., 2017). DNA barcoding complements the traditional taxonomic identification by providing a faster method to identify organisms to their species level, especially those that are uncertain, cryptic species and also helps to distinguish morphologically similar species (Matarese *et al*., 2011). Shao *et al*. (2002) observed that molecular tools could confirm fish egg identity according to their species. Among all the molecular

techniques developed, DNA barcoding has become the golden standard to support the traditional morphological identification and to confirm the larval fish species as it is the most rapid and efficient method (Ko *et al.*, 2013).

2.3 The mitochondrial DNA (mtDNA) as a taxonomic tool

Animal mitochondrial DNA (mtDNA) is an important and efficient genetic marker used in many molecular studies such as for species identification (Ward *et al*., 2005, Webb *et al.*, 2006, Victor *et al*., 2009, Mat Jaafar *et al*., 2012, Ardura *et al.*, 2013, Azmir *et al.*, 2017), phylogenetic and evolutionary studies (Dalziel *et al*., 2006, Cheng *et al*., 2011, Adibah, 2017). A survey conducted by Teletchea (2009), revealed that mtDNA genes were the most widely used genetic markers in molecular studies at that time. The mtDNA consists of 37 genes encoding two mitochondrial ribosomal RNAs (rRNAs) (i.e. 12S rRNA, and 16S rRNA), 13 mitochondrial protein coding genes (i.e. cytochrome oxidase subunit I (COI), II (COII) and III (COIII), cytochrome *b* (cyt*b)*, ND 1-6, 4L, ATP synthase subunit 8 (ATP8), ATP synthase subunit 6 (ATP6)) and 22 transfer RNA (tRNAs) genes (Boore, 1999). The mtDNA also contains the D-loop, or the control region which plays an important role in the replication of DNA and transcription of RNA (Taanman, 1999). The mtDNA genes have several advantages over nuclear DNA that make them widely selected as markers for species identification. The mitochondrial genome of fish showing the orientation of the genes is shown in Figure 2.2 below.

Teletchea (2009) listed three main reasons of the mtDNA genes usage in species identification. Firstly, the multiple copies of mtDNA in the cell increases the probability of fragment amplification within the mtDNA genome compared to within the nuclear genome. Next, the circular genome of mtDNA in most animal species of

16 to 20 kilobases (kb) long is maternally inherited, haploid and does not undergo recombination. These characteristics make the study using mtDNA easier and more straightforward. Lastly, the evolution rate of the mtDNA is faster compared to single nuclear DNA genes. Brown *et al* (1979) and Kochzius (2009) also stated that the mtDNA markers evolve five to 10 times faster than the nuclear DNA and thus, enables the identification of closely related species. Besides, Teletchea (2009) also suggested that a suitable DNA marker for species identification should be sufficiently variable between species (particularly among closely related ones) and display either low or no-intraspecific value.

The standard DNA barcoding gene as proposed by Hebert *et al.* (2003b) is based on the cytochrome c oxidase subunit I (COI) covering an approximately 650 base pairs (bp) region. The amplified length should be longer than 500 bp in order to achieve high resolution species identification (Hebert et al., 2003b, Coissac et al., 2012). Successful identification of animals including marine and freshwater fishes (Ward *et al*., 2005), birds (Hebert *et al*., 2004) and Lepidopteran (Hajibabaei *et al*., 2006) have been documented. The mitochondrial COI gene was chosen as a standard barcoding marker for animals based on several criteria. The COI gene appears to be the most conserved mtDNA protein-coding region compared to other genes, and it also has a short sequence length (400-800 bp) which results in easier and faster amplification (Kress and Erickson, 2008). Based on these advantages, the COI gene region has become the universal DNA barcode to identify animal species, albeit with some exceptions. The immense data globally collected is deposited in a DNA barcoding database called Barcoding of Life Database (BOLD). Voucher sequences in BOLD are linked to all major sequencing websites such as GenBank, MitoFish, etc. The COI target gene has been expanded to the more advanced technique of

metabarcoding (Hebert *et al*., 2003a, Ratnasingham and Hebert, 2007, Ratnasingham and Hebert 2013) as described below.

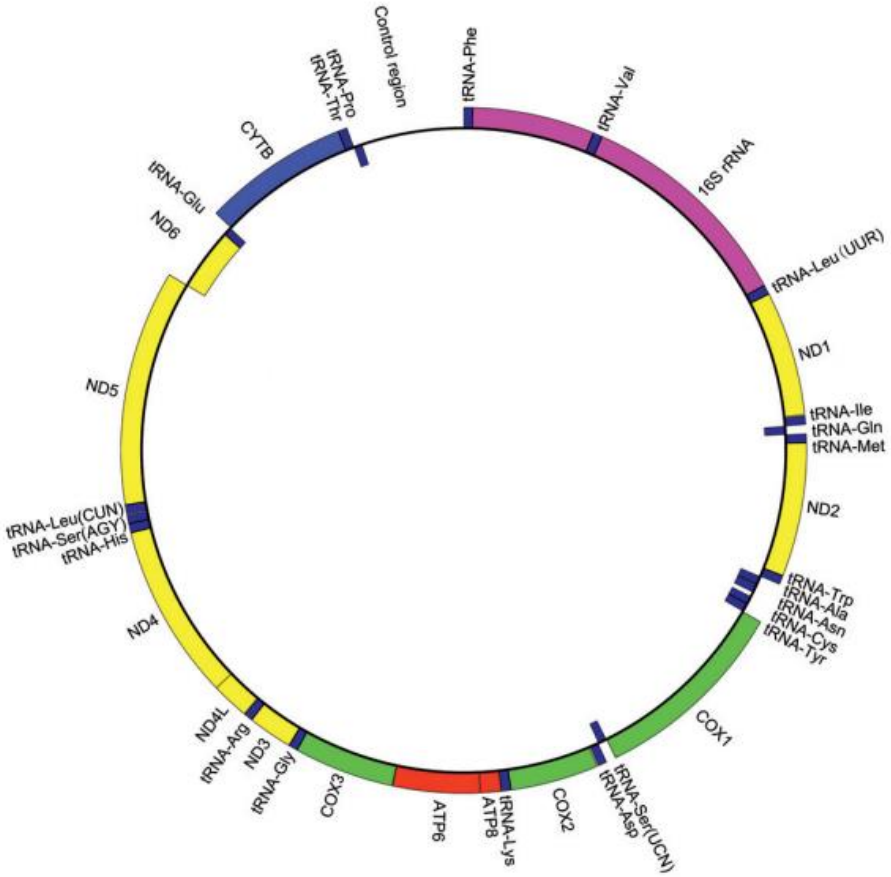

Figure 2.2 Complete mitochondrial genome of fish (*Atule mate*, family Carangidae) (Li et al., 2016). The circular mitochondrial genome is made up of 37 genes: 12S rRNA and 16S rRNA genes are the two mitochondrial ribosomal RNA genes; ND 1-6 indicate gene of the NADH dehydrogenase subunits 1-6; COX1, COX2 and COX3 are cytochrome oxidase subunit I, II and III; cyt*b* is cytochrome *b*; ATP8 and ATP6 are ATP synthase subunit 8 and ATP synthase subunit 6; 22 transfer RNA (tRNAs); and D-loop or control region.

2.4 DNA Metabarcoding

2.4.1 Introduction of DNA metabarcoding

Arnot *et al*. (1993) was the first research team to introduce DNA-based method for species identification which was later advanced and standardized by Hebert *et al*. (2003a) through a method called DNA barcoding. Through this effort, the Consortium

for the Barcode of Life (CBOL, http://barcoding.si.edu), that supports the development of DNA barcoding (Hebert *et al*., 2003a, Valentini *et al*., 2009, Cristescu, 2014) was formed. As earlier mentioned, the term DNA barcoding approach relies on the use of a standardized DNA region as a tag for rapid and accurate species identification (Valentini *et al*., 2009). Currently, DNA barcoding is widely used not only as a complementary tool for taxonomic identification, but also used for biodiversity assessment, diet analysis and food safety (Valentini *et al*., 2009, Taylor and Harris, 2012, Clare *et al*., 2014, Chin *et al*., 2016). However, DNA barcoding using Sanger sequencing method can identify only a single specimen at a time and is time-consuming and logistically impossible to identify thousands of larvae specimens which is normally found at spawning or nursery sites (Taberlet *et al*., 2012b, Maggia *et al*., 2017, Mariac *et al*., 2018). Recently, further advancement of the traditional DNA barcoding has been developed which is able to address this. This new approach is referred to as DNA metabarcoding.

DNA metabarcoding as introduced by Taberlet *et al*. (2012b) can be generally defined as the rapid and cost-effective identification method of multiple species from bulk samples containing entire organisms (e.g. fish eggs, larvae) or damaged specimens that are difficult to morphologically identify or from a single environmental sample containing fragmented specimens and possibly degraded DNA (Taberlet *et al.*, 2012a)) collected from soil, water or faeces. Conventional DNA barcoding relies on the standard Sanger sequencing, also known as the first-generation sequencing, is restricted to sequencing a single gene from a single specimen in a single run (Ji *et al*., 2013). DNA metabarcoding, has no limitation on the number of specimens that it can identify simultaneously. This approach relies on the Next Generation Sequencing (NGS) technologies that can generate thousands to millions of individual sequence

reads in parallel, from fragmented library of a specific genome or from mixture of specimens DNA at a relatively low cost (Shokralla *et al*., 2012). Thus, it is an ideal method for high-throughput species identification especially for biodiversity assessment and ecological studies (Cristescu, 2014, Lobo *et al*., 2017, Coissac *et al*., 2012, Andersen *et al*., 2012). Therefore, DNA metabarcoding has raised DNA-based identification to another level through its efficiency in distinguishing individuals in bulk and environmental samples (eDNA) to species level. The general workflow of the DNA metabarcoding is shown in Figure 2.3 below.

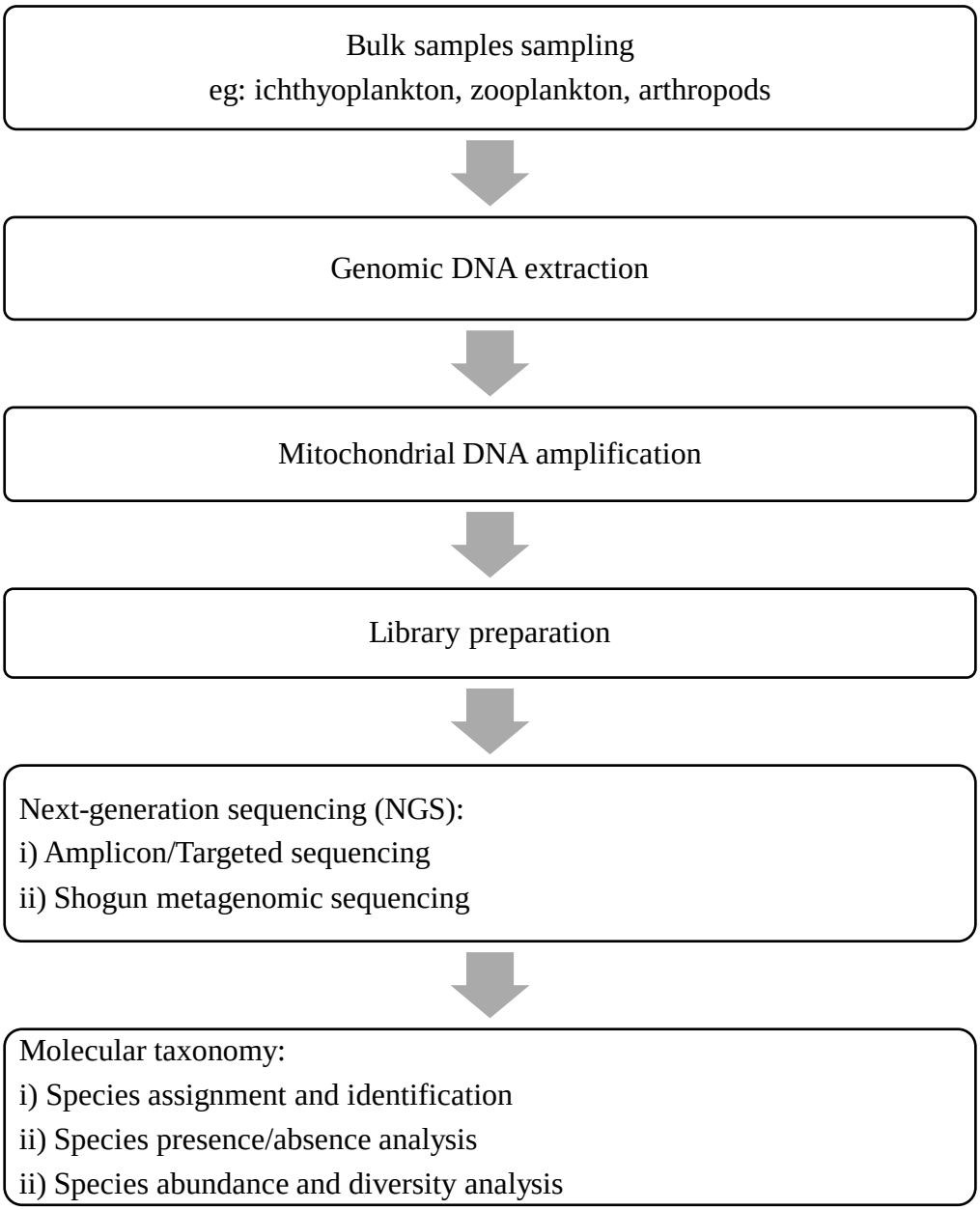

Figure 2.3 Workflow of DNA metabarcoding method.

2.4.2 Application of DNA metabarcoding

Although the DNA metabarcoding approach is still relatively new, its potential application especially for biodiversity monitoring and community structure of various organisms has been successfully illustrated through several studies. DNA metabarcoding has been applied to identify and quantify the freshwater fish larvae of

Amazon river basin (Maggia *et al*., 2017, Mariac *et al*., 2018), identification of ichthyoplankton (fish eggs and larvae) at the Mogi-Guacu River in the northeast region of Sao Paulo, Brazil (Nobile *et al*., 2019) and quantitative assessment of fish larvae at the Irish and Celtic Seas (Ratcliffe *et al*., 2020). The technique has also been applied on taxa other than fish larvae. These include protists (Pawlowski *et al*., 2014, Geisen *et al*., 2015), fungi (Geml *et al*., 2014, Korpelainen *et al*., 2016), nematodes (Dell'Anno *et al*., 2015), arthropods (Douglas *et al*., 2012, Beng *et al*., 2016) and zooplankton (Corell and Rodríguez-Ezpeleta, 2014, Bucklin *et al.*, 2016, Clarke *et al*., 2017). In addition, DNA metabarcoding has also successfully provided data from faeces and stomach contents for diet analysis of organisms (Pompanon *et al*., 2012).

Furthermore, this tool also has also been applied to analyse DNA released or isolated from environmental samples (eDNA) (Taberlet *et al*., 2012a, Creer *et al.*, 2016, Deiner *et al*., 2017). For example, to study the whole diversity of fish in an aquatic environment in both freshwater and marine fish communities based on water samples (Thomsen *et al*., 2012, Miya *et al*., 2015, Yamamoto *et al.*, 2017, Sato *et al*., 2017, Smith, 2017). The eDNA analysis has also been applied on soil samples to get better insight on the plant diversity (Epp *et al*., 2012), invertebrate and vertebrate communities (Bienert *et al.*, 2012, Andersen *et al*., 2012) that are present in or on the soil.

2.4.3 Challenges of DNA metabarcoding

Despite the advantages of DNA metabarcoding, there are also several issues associated with it. The challenges are attributed to several factors; types of sources, sampling area, technical (laboratory) and bioinformatics analyses (Deiner *et al*., 2017).