CHAPTER 1: INTRODUCTION

Channa striata, locally known as "Haruan", is one of the Asian snakeheads (genus *Channa*) belongs to the family Channidae. This commonly recognized snakehead murrel extensively inhabits a wide range of habitat ranging from water bodies of rivers, swamps, ponds, canals and lakes to land of rice fields and is distributed across its natural distribution range of southern Asia, southern China, Indochina and the Sunda Islands (Mohsin & Ambak, 1983; Lee & Ng, 1994; Hossain *et al*., 2008). This freshwater carnivore species is able to tolerant to adverse environment due to its hardiness characteristic and its unique property of air-breathing capabilities assisted with suprabranchial chamber – an air-breathing organ (Chandra & Banerjee, 2004) which is not found in other freshwater fish families (Munshi, 1962; Ishimatsu & Itazawa, 1981; Ishimatsu & Itazawa, 1993). As a locally agreeable flavor food (Hossain *et al*., 2008) and locally recognized medical purpose in the post-operative to enhance wound healing as well as to reduce postoperative pain and discomfort (Mat Jais *et al*., 1997), the local market demand on *C. striata* is greatly expanding due to its commercialized value. Population genetic structure of *C. striata* is urgently needed to provide information on brood-stock management and also could add spatial conservation value on this species.

To date, the exact taxonomic status at the level of species of the genus of *Channa* remains unclear as *C. striata* was placed as one of the "species complexes" under this mentioned genus (Courtenay & Williams, 2004). This had raised the interest to investigate the phylogenetics relationship of the snakehead particularly the origin of the Asian snakehead and its evolution towards formation of different *Channa spp*. Previous research

study showed that the occurrence of the Asian snakehead in southern Asia particularly in South-East Asia was mostly due to phylogeography influence (Adamson *et al*., 2010) and the population structure of *C. striata* surveyed in Perak, Malaysia was greatly explained by historical Pleistocene event (Jamaluddin *et al*., 2011). These had strongly suggested that the historical coalescence of freshwater taxa was greatly influenced by the phylogeographic structure of the historical land mass during ancient period of climatic changes. Specifically, the Pleistocene event that happen approximately 250k years ago was encountered with multiple episodes of glaciation and deglaciation with the accompanied lowering and rising in sea-water level. Historical Pleistocene glaciation with the maximum lowering sea-water level is greatly affecting the land mass configuration and thus the biodiversity of fauna where the Peninsular Malaysia, island of Borneo, southern Thailand, southern Indo-China, Sumatera and Java are once interconnected (Voris, 2000). Migration and gene flow of fauna between the Peninsular Malaysia and the island of Borneo accomplished through the land bridge by the formation of Sunda River until the submersion of this river system during the last Pleistocene Epoch with rising of sea-water level (Mohsin & Ambak, 1983; Dodson *et al*., 1995). This phenomenon was evidenced by the genetic divergence pattern of *Hampala macrolepidota* (Ryan & Esa, 2006), *Tor tambroides* (Esa *et al*., 2008) and *Barbonymus schwanenfeldii* (Kamarudin & Esa, 2009) with the witness of non-significant genetic differentiation and overlapping haplotype distribution pattern between the Peninsular Malaysia and the island of Borneo. On top of that, within the Peninsular Malaysia, previously the populations of freshwater fish species were divided into three main geographical division by Mohsin & Ambak (1983, 1991) which are a North-West division ranging from northern region towards west way until Taiping, a North-East and Central division which covered the area after Taiping southward

before reaching Kelang and Kuala Rompin, and a Southern division ranging from Kelang and Kuala Rompin to the remaining southern region.

Previous study on *C. striata* by Adamson *et al*. (2010) had investigated the high divergence pattern of this species across geographically wide range of South-East Asia, thus indicated the possible existence of putative cryptic species in this sympatric population. Besides, another study by Bhassu & Salah (accepted manuscript) on wild *C. striata* populations across Peninsular Malaysia revealed that Negeri Sembilan population might be a possible cryptic species using microsatellite markers. Hence, in this study, we attempt an advance species identification approach known as "DNA barcoding" which was invented in the year of 2003 (Hebert *et al*., 2003a) to investigate the presence of unidentified cryptic species in this potential freshwater species among its allopatric populations across Malaysia. This molecular species discrimination technique focus on a standardized barcode region – mitochondrial DNA (mtDNA) of cytochrome *c* oxidase subunit 1 (COI) (Hebert *et al*., 2003a) as a molecular species tag and had a number of successful records in revealing putative cryptic species (Hebert *et al*., 2004b; Huang *et al*., 2007). In this DNA barcoding approach, while using the subset of the same Peninsula's *C. striata* populations in previous microsatellite study (Bhassu & Salah, accepted manuscript) and additional specimens sampled from Sarawak population, we included additional 26 freshwater species in addition to *C. striata* and also examined multiple specimens within each species in order to have greater statistical support and more accurate estimation of variation present within and among freshwater fish taxa in Malaysia. Furthermore, while revealing potential cryptic species of *C. striata*, in order to have likely extent of con-specific variation, we included this representative species which had sampled across its natural geographical range throughout Malaysia; otherwise, the intra-species divergence is

more likely to be underestimated due to phylogeographical divergence influence. Another study of Jamaluddin *et al*. (2011) on *C. striata* in the state of Perak, Malaysia had revealed that the historical Pleistocene event play a dominant role in population divergence pattern in this species. Their finding together with the evidence of the similar historical coalescence in other freshwater taxa (Ryan & Esa, 2006; Esa *et al*., 2008; Kamarudin & Esa, 2009), had raised up the interest to further study this Pleistocene impact on *C. striata*. Thus, we included *C. striata* population from Sarawak as a representative from the island of Borneo to examine the population's genetic divergence pattern induced by the effect of historical Pleistocene event and also to examine the effect exerted on the populations after the recent separation of the island from the mainland of Peninsular Malaysia by the South China Sea. For this molecular divergence pattern to be studied, mitochondrial DNA was selected as a marker of choice due to its nature of slow evolution rate and thus is more appropriate to be chosen for the study of historical dispersal pattern of *C. striata*. Additional morphometric data were also included in examining phenotypic divergence pattern of *C. striata* throughout Malaysia inferrered using Truss Network Measurement (Bhassu & Rashid, 2009). While investigating the possible hidden sister taxa of *C. striata* in Malaysia, the aim of this study is to examine the population divergence pattern of *C. striata* in order to provide spatial conservation status and assessment according to spatial conservation value and needs.

Objectives:

The specific objectives for this study are listed as follow:

(1) To examine genetic and morphological variation in natural populations of *Channa striata*

(2) To investigate the presence of cryptic species in *Channa striata* across Malaysia using molecular data

(3) To propose spatial conservation management planning for *Channa striata*

Hypothesis:

The hypothesis for the study is:

If variation is present among natural population of *Channa striata* in Malaysia, then patterns of molecular and morphological divergence will be similar.

Publication:

The following are the papers that had been submitted for publication in ISI journal:

(1) Song, L.M., Adamson, E.A.S., Munian, K., Rashid, Z.A., Othman, R.Y. & Bhassu, S. (2011) DNA barcoding and divergence of freshwater fish species of Malaysia. Manuscript submitted to *GENE*.

(2) Song, L.M., Hassan, R., Rashid, Z.A. & Bhassu, S. (2012) Characterisation of Asian snakehead murrel *Channa striata* (Channidae) in Malaysia; insight from molecular data and morphological approach. Manuscript submitted to *JOURNAL OF FISH BIOLOGY*.

Publication in proceeding (28^{th} – 30^{th} September 2011: 9^{th} Malaysia Genetics Congress):

(1) Song, L.M., Munian, K., Adamson, E.A.S., Rashid, Z.A. & Bhassu, S. (2011) DNA barcoding of freshwater fishes of Malaysia. *Ninth Malaysia Genetics Congress*, Kuching, Sarawak, Malaysia.

CHAPTER 2: LITERATURE REVIEW

2.1: VARIATION IN NATURAL POPULATION

2.1.1: INTRODUCTORY VARIATION

Each organism in nature has its unique imprint which could be differentiated from their siblings as well as other organism in the inter-breeding population of the same species. This unique imprint is known as "variation" that exists in natural population which will increase survival fitness of an organism. Previous study had showed the existence of variation in natural population ranging from invertebrate of insecta (Bosio *et al*., 2005), to vertebrate of bird (Grant *et al*., 1976), mammal (Mullen *et al*., 2009), and to marine and freshwater bodies fish (Fujio & Kato, 1979; Ambak *et al*., 2006). Fishes which inhabit the water bodies of environment are well studied for their divergence pattern due to their population structures which are greatly affected by divergence forces such as ecology, biogeography of the sea level and man-induced alteration of the environmental niches (Fuijo & Kato, 1979). Genetically, variation can be explained by variance of genetic component, briefly, mutation on DNA sequence; on the other hand, phenotypically, such variation which is observable will be an effect of genetic component coupled with additive environmental condition.

2.1.2: SOURCE OF VARIATION

Traits are being passed through from the parental line to offspring every generation. An error in terms of base pair will occur during inheritance of certain gene. This error is known as "mutation" which is one of the evolutionary processes that causes intra-species variation and at certain extant coupled with other evolution factors which will eventually end up with formation of a new species. The complexity of variation is mainly due to it is an additive effect of various divergence factors, including mutation, habitat ecology, gene flow and different trait selections that could leave a genotypic or phenotypic imprint on an organism. In some cases, phenotypic variability is not correlated with genetics as environmental condition is most likely to affect the trait development which need not be associated with underlying genetic differences.

Biogeography, being one of the variation sources, had a significant impact on population variation. Isolated populations which are separated by physical barrier such as mountains are evolving under different selection pressure separately. Different selection pressure as induced by different environmental forces to differentially select respective adaptive trait in isolated populations will lead to variation in two populations. In addition, different mutation rates that acted on to separate population's gene pool will diversify the genotypic and/or phenotypic between two isolated populations. Ambak *et al*. (2006) study had proven that geographical barrier of mountain which separate East and West of Peninsular Malaysia had a significant impact on population genetic distance as compared to those populations on the same side of that mountain. Biogeography is sometimes known as the initial step in allopatric speciation because the two allopatric populations who share the same common ancestral evolve separately to a certain extent in which interbreeding

between these two populations could not produce fertile offspring after many generations with the removal of physical barrier.

Furthermore, ecology may play a vital role in arising variation in ecological trait of an organism. Study of Grant *et al*. (1976) on *Geospiza fortis* implied that ecology heterogeneity in habitat is a population diversifying force which leads to biodiversity of a certain species. Their study had proven that the morphological variation of *G. fortis* in two different habitats was associated with ecological difference. Indeed, food supply on different ecological niches could speed up the process of variation. Schneider *et al*. (1999) had proposed that the lizard morphology in two different habitats was selected under different selective force due to different predation risk. Thus, variation in certain trait will occur for survival purpose. Another example of feeding behavior induced divergence was proposed by Hebert & Gregory (2005) on different coloration of forewings of moth as the larvae of *Simyra henrici* feed on grasses whereas *Acronicta* species is a tree-feeding moth. This finding suggested that natural selection had driven the shift in the coloration of the forewings due to their different feeding habits.

The two divergent populations in geographical distant habitat will maintain their given level of variation through natural selection (Grant *et al*., 1976). Natural selection will act on different habitat to select different adaptive character for survival. Hence, certain adaptive trait that is ecologically important will be maintained from generation to generation. This could explain the misidentification of a species by taxonomist as the two populations of same species which inhabit distinct environment will have significant phenotypic variation as environmental condition had speed up phenotypic variability with little genetic variation. .

2.1.3: ASSESSMENT OF VARIATION

Understanding of population variation of certain species will provide informative data in assisting conservation planning as it indirectly explain the biodiversity of certain species. Assessment of population variation had long been studied using various techniques which successfully clarified population variation.

The way to assess phenotypic variation is generally based on morphometric measurement of measurable character and meristics which rely on countable quantitative features. This is a conventional method for the taxonomist in classifying certain species based on the variation in morphology (Maharadatunkamsi & Maryanto, 2002; Maryanto *et al*., 2005).

The trend of assessment of population genetic variability started from survey of protein polymorphism using enzyme electrophoresis starting from the year of 1960 (Lewontin & Hubby, 1966; Nei, 1975). This method had been applied in fish population study to investigate the variation among species of marine fish by Fujio & Kato (1979). However, there are some obstacles in this protein-based technique such as the nature of redundancy of the genetic code and the failure of electrophoresis in detecting amino acid replacement (Lynch & Crease, 1990).

Starting from the year of 1980,, the advance of DNA technologies have benifited population biologist..With the advance of introduced polymerase chain reaction technique and informative restriction site of the DNA sequence, henceforth, surveys of genetic variability become possible in terms of sequence variation (Lynch & Crease, 1990). This led to the development of random amplified polymorphic DNA (RAPD) markers to investigate population variation. Bardakci & Skibinski (1993) have provided evidence of the successful RAPD markers in discriminating tilapia species of the genus *Oreochromis* and several subspecies of *Oreochromis niloticus*. Their finding is a head of the previous protein electrophoresis which could not discriminate subspecies of *O. niloticus* (Seyoum, 1990; Seyoum & Kornfield, 1992). However, with the dominant character of RAPD, this methodology is facing some drawback due to the impossibility in making any determination in a given homozygous or heterozygous individual (Ambak *et al*., 2006).

Microsatellite and mitochondrial DNA are next targeted markers in classifying genetic structure of a population. Microsatellite is a candidate marker in the studied of population diversity due to its nature of non-coding DNA fragment which is fast evolving. The quicker evolution rate of this marker making assessment of biodiversity and divergence at population level can be assesed significantly (Bhassu & Rashid, 2009; Salah & Bhassu, accepted manuscript). Mitochondrial DNA approach is particular adopted as it is a candidate marker in describing a history lineage of an organism due to it slow evolving capability compared to microsatellite marker. It is well studied in phylogenetics as it is able to differentiate closely related species which is separated after a short evolutionary period. In 21st century, studies of population genetic structure using mitochondrial genome become more popular which speculate the variation observed in haplotypes sequence

(Hogg & Hebert, 2004; Bosio *et al*., 2005; Huang *et al*., 2007; Esa *et al*., 2008; Kamarudin & Esa, 2009).

Hebert *et al*. (2003a) has introduced a molecular based technique of species classification, known as "DNA barcoding" to discriminate animal kingdom above species taxa using 5' end of mitochondrial cytochrome *c* oxidase subunit 1 (COI). DNA barcoding system are now universally established for several groups of animals, such as birds (Hebert *et al*., 2004a), cowries (Meyer & Paulay, 2005), spider (Barrett & Hebert, 2005), and several arrays of Lepidoptera (Hebert *et al*., 2004b). Besides, this molecular approach is also being established to other groups of organisms, including plants (Kress *et al*., 2005), macroalgae (Saunders, 2005), fungi (Summerbell *et al*., 2005), protists (Scicluna *et al*., 2006) and bacteria (Sogin *et al*., 2006). Therefore, it is understood that DNA barcoding is a desirable tool in describing intra-species variation while remaining the two species as different species.

2.1.4: PHYLOGEOGRAPHY

Historical Pleistocene event with multiple episodes of glaciation and deglaciation with the accompanied lowering and rising of sea-water level is greatly affecting the land mass configuration (Voris, 2000). Since freshwater fish inhabiting river ecosystem, thus the Pleistocene glaciation with maximum lowering sea-water level had affected the divergence pattern of freshwater species. During the Pleistocene scenario which happen approximately 250k years ago, Peninsular Malaysia, island of Borneo, southern Thailand, southern Indo-China, Sumatra and Java were interconnected and were once part of the submerged Sunda Shelf (Mohsin & Ambak, 1983). The formation of land bridge which is known as Sunda River is allowing gene flow and migration of the fish between the Peninsular Malaysia and the island of Borneo through the interconnected land bridge (Mohsin & Ambak, 1983; Dodson *et al*., 1995). Refer **Figure 2.1** for illustration of sea-level on depth contour of 120m below present level. The island of Borneo is separated from the Peninsular Malaysia by the formation of South China Sea with the rising of sea-water level during the late Pleistocene deglaciation. Since the Peninsular Malaysia and the island of Borneo were once interconnected through Sunda River, hence Pleistocene event explain most of the historical coalescence of freshwater species. The population divergence pattern of *Hampala macrolepidota* (Ryan & Esa, 2006), *Tor tambroides* (Esa *et al*., 2008), *Barbonymus schwanenfeldii* (Kamarudin & Esa, 2009) and *Channa striata* (Jamaluddin *et al*., 2011) are evidenced by Pleistocene event.

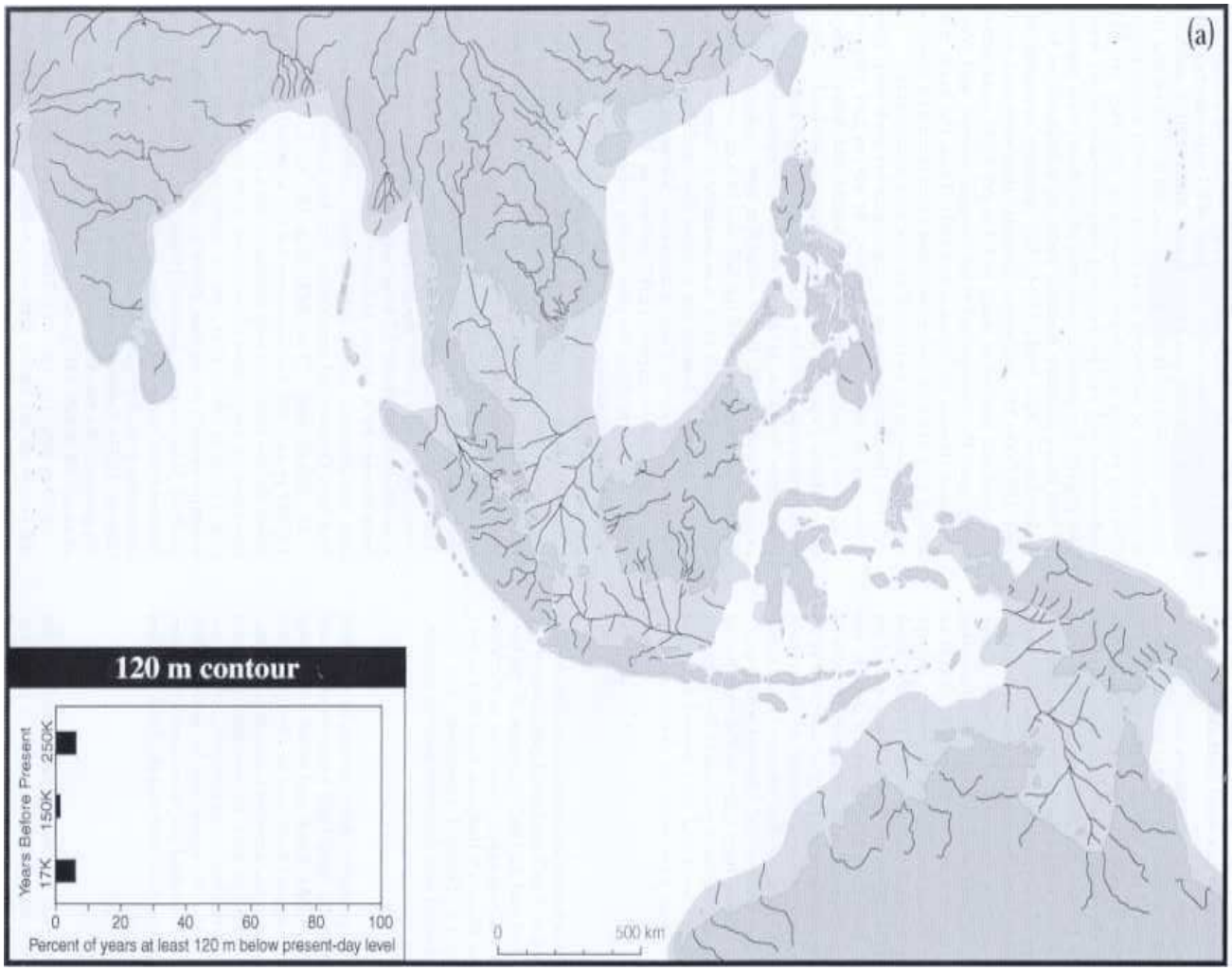

Figure 2.1: Map of Pleistocene sea level for Southeast Asia and Austral-Asia with depth contour of 120m below present level. Figure cited in Voris (2000).

2.2: FRESHWATER FISH

2.2.1: STATISTICS ON FRESHWATER FISH

Fishes had been recognized as the major food protein supplement for worldwide human consumption and its richness in diversity had been extensively studied. There are approximately 32 000 fish species (FishBase: www.fishbase.org, accessed on 27 March 2011) described globally. Freshwater fishes of Malaysia comprise approximately 626 species among the total of 1840 described fish species in Malaysia (FishBase: www.fishbase.org, accessed on 27 March 2011). In 2000, fisheries provided more than 15% of total animal protein to the global food supply which consist of 35 million people. The supply of animal protein from fish had an estimated first sale value of about US$81 billion (FAO 2002). Furthermore, in the year of 2008, the production of freshwater fishes was dominant to other aquatic groups with a production of 28.8 million tones, which contribute to more than 50% of the total aquatic animal production (FAO 2010). The population of the fish is currently declining mainly due to over harvesting and habitat destruction. Consequently, conservation planning to conserve the endangered species as well as to maintain the biodiversity richness is urgently needed in the current status.

2.2.2: TAXONOMIC CLASSIFICATION OF FISH

The diversity of fishes ranging from ancient jawless species (Agnatha: hagfish and lampreys) through to cartilaginous fishes (Chondrichthyes: chimaeras, sharks and rays) and to old and modern bony fish (Osteichthyes: coelacanths, eels, carps, tunas, flatfishes, salmonids, seahorses, etc) (Ward *et al*., 2005).

The genera of the fishes constitute into families. For instance, *Barbichthys* belongs to the family Cyprinidae together with other genera such as *Balantiocheilus*, *Hypsibarbus*, *Barbonymus*, *Hampala* and *Amblyrhynchicthys*. Some other fish families include Clariidae, Cobitidae, Siluridae, Sisoridae, Notopteridae, Palanoeidae and Soleidae. Families are then constituted into orders. For example, the three families Cyprinidae, Clariidae and Cobitidae all belong to the order Cypriniformes. Other important orders are Siluriformes, Beloniformes, Gasteroateiformes, Perciformes and Osteoglossiformes. Orders are further grouped into classes. Some of the example of classes are Chondrichthyes (the cartilaginous fish class), which include sharks, rays and jawfish, and Osteichthyes (the bony fish class). Finally, classes are grouped into phyla. All fish species in the world belong to the same phylum which is Chordata (the vertebrates with a backbone), which also includes amphibians, reptiles, birds and mammals. Phylum Chordata is belongs to kingdom Animalia. Kingdom Animalia is under domain Eukaryota. The general classification of fishes starting from the common domain of Eukaryota to different classes of fish is illustrated in **Figure 2.2**.

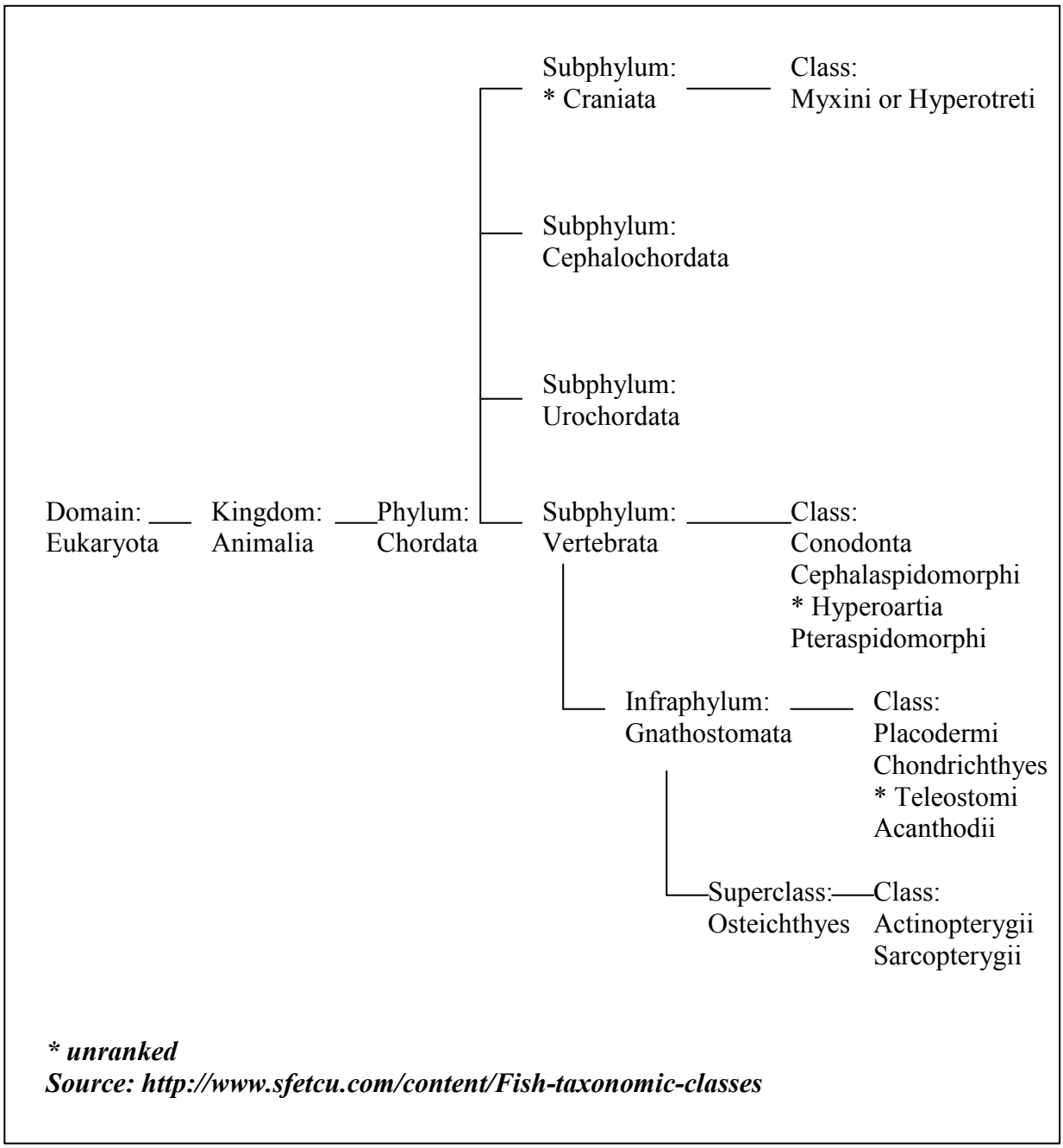

Figure 2.2: Taxonomic classification of fishes from the common domain of Eukaryota up to different classes whereby diversification of fish species started from each class.

2.2.3: *Channa striata*

The systematic of the Asian snakeheads genus of *Channa* still remain unclear as there are different numbers of species under this genus being reported which are in the study of Li *et al*. (2006) who reported 26 species and Ambak *et al*. (2006) who reported 27 species. These had consequently claims that there is a potentially hidden diversity of unknown species under *Channa* genus and the phylogenetics information of snakeheads are actively studied for the interest of its origin. The occurances of Asian snakeheads in Asia were predicted to be arrived from India via the Indian "land-bridge" (Hedges, 2003; Karanth, 2006). In stead of the "Out-of India" theory and others hypotheses of the origin of Asian snakeheads in southern Asia (Datta-Roy & Karanth, 2009), the study of Adamson *et al*. (2010) reported that the more suitable hypothesis to explain the evolutionary history of snakeheads should be "Out-of-Asia-into-Africa".

Channa striata, being one of the freshwater Asian snakeheads belongs to the family Channidae (snakehead fishes) which is also a commercially important fish in Southeast Asia (Wee, 1982; Dey *et al*., 2005). This species was commonly named as "snakehead murrel" and was taxonomically classified as shown in **Figure 2.3**.

As described in the given name - "snakehead", this freshwater fish owning unique head shape which is a large head reminiscent of a snake's head. The basic structure of this snakehead murrel comprises of a sub-cylindrical body shape with rounded caudal fin, dark at its side and dorsal surface mottled with a combination of black, ochre and white on the belly (Fishbase - Spread sheet of Morphology Data of *Channa striata*: Identification keys). See **Figure 2.4** for image of *C. striata*. It is commonly known as "Haruan" in Malaysia and

is indigenous to many tropical countries (Mohsin & Ambak, 1983). It is a carnivorous species that is able to tolerant to adverse environment due to its hardiness characteristic and air-breathing capabilities assisted with suprabranchial chamber – an air-breathing organ (Chandra & Banerjee, 2004) which is not found in other freshwater fish families (Munshi, 1962; Ishimatsu & Itazawa1981; Ishimatsu & Itazawa1993). It can be found in wide ranges of habitats including rivers, swamps, ponds, canals, rice fields and lakes across its native geographical range of southern Asia, southern China, Indochina and the Sunda Islands (Mohsin & Ambak, 1983; Lee & Ng, 1994; Hossain *et al*., 2008). See **Figure 2.5** for distribution map of *C. striata*.

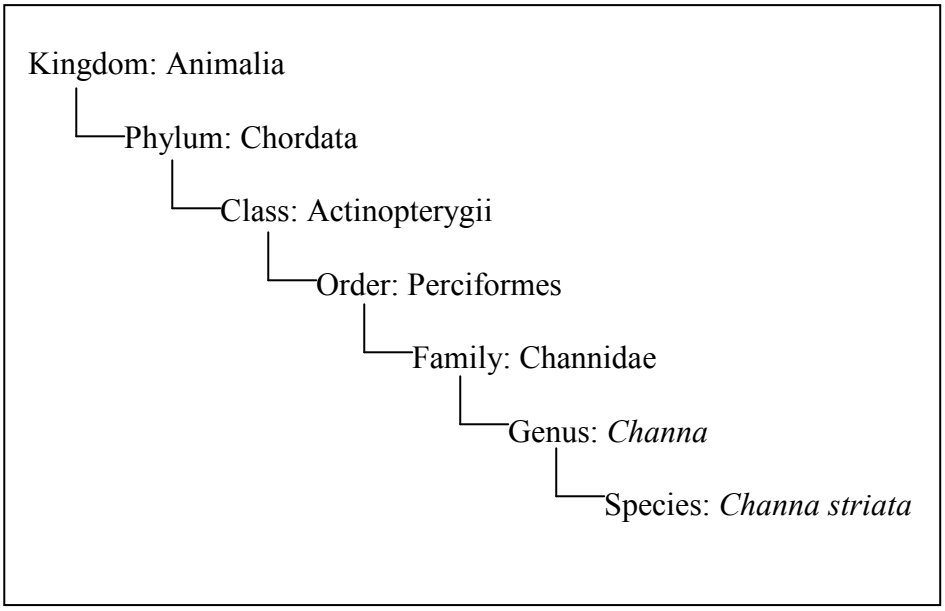

Figure 2.3: Taxonomic classification of *C. striata*.

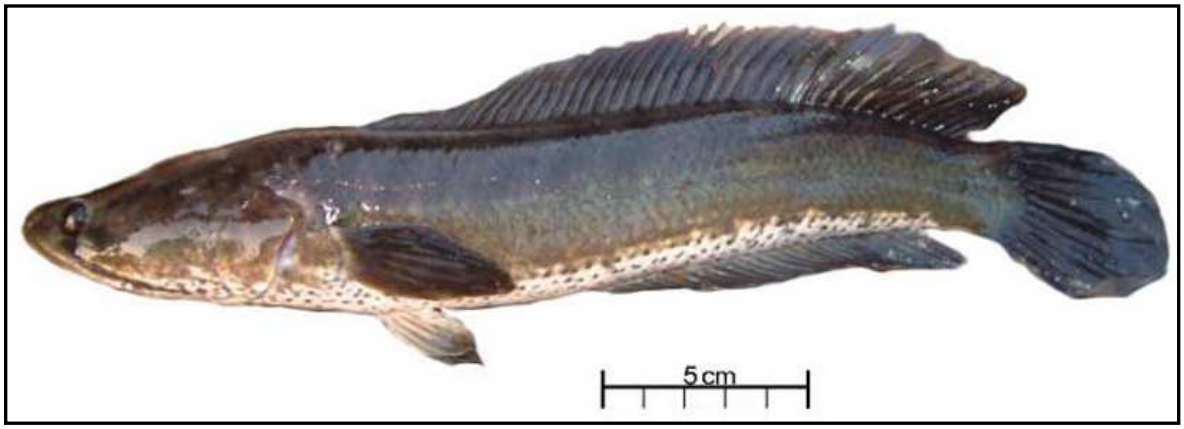

Figure 2.4: Image of *Channa striata*. Image obtained from Department of Fisheries Malaysia, Freshwater Fisheries Research Division, FRI Glami Lemi, Jelebu, 71650 Titi, Negeri Sembilan, Malaysia; www.pppat.gov.my; ffrc@pppat.gov.my.

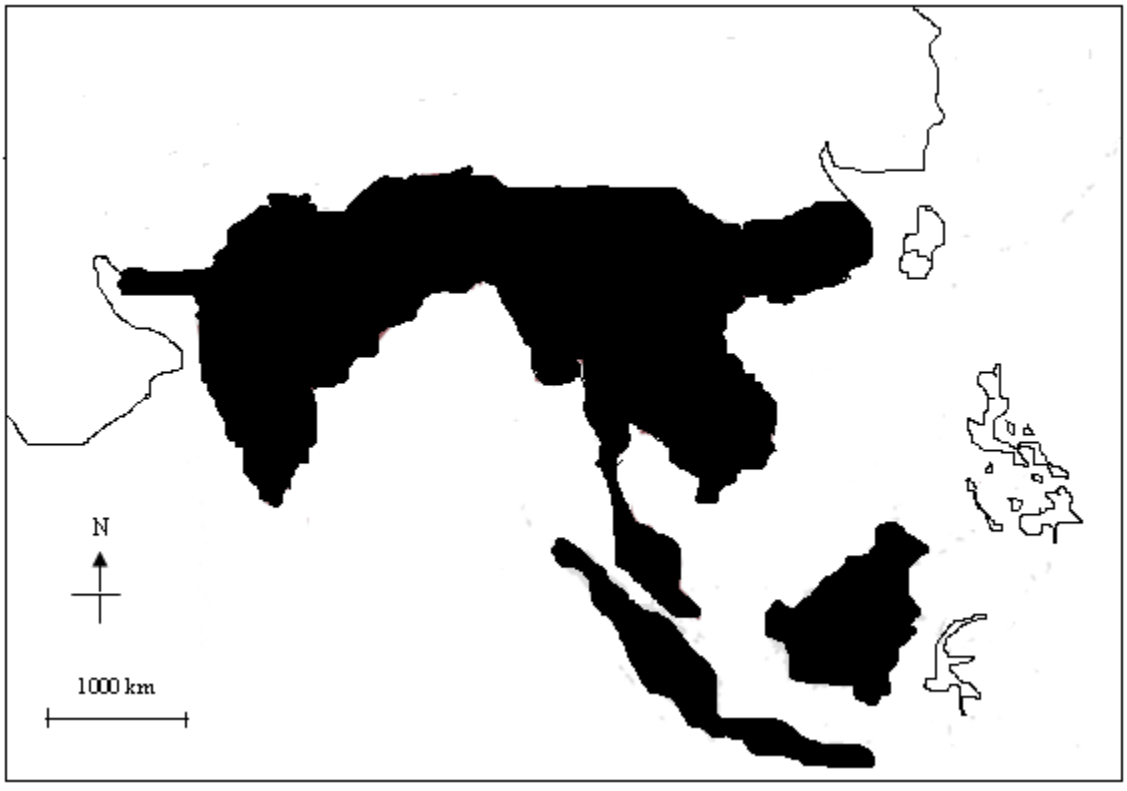

Figure 2.5: Distribution map of *Channa striata* in Asia with darken regions indicate its natural geographical ranges.

The annual capture and aquaculture of *C. striata* is greatly expanding due to the local market demand (See **Figure 2.6**). This extensively studied fish is one of the protein sources having high essential amino acid and fatty acid content which have a potential medication value. The major amino acid in *Channa spp.* such as aspartic acid and glutamic acid are recognized to have wound healing effect (Chyun & Griminger, 1984). Nevertheless, the amino acid glycine found in *Channa spp.* is known be acted together with other amino acid as well as the two amino acid rich in *Channa spp.* (aspartic and glutamic acid) to produce a polypeptide which is responsible in promoting growth and healing (Heimann, 1982). In terms of fatty acid, the arachidonic acid level in *C. striata* is unexpectedly higher than its con-generic *Channa spp.*, indicating that *C. striata* can promote inflammatory recovery as arachidonic acid play a key role in initiating blood clotting (Bowman & Rand, 1980). Due to these high essential wound healing components; *C. striata* had been chosen as an important diet for medication purpose in Asia Pacific region including Malaysia (Mohsin & Ambak, 1983). It is traditionally used by patients in the post-operative to enhance wound healing as well as to reduce postoperative pain and discomfort (Mat Jais *et al*., 1997).

In Malaysia, *C. striata* is widely consumed perhaps due to its boneless flesh and its agreeable flavour (Hossain *et al*., 2008). The product of *C. striata* using its extract had been commercialized. Palm oiled based cream of "Haruan" extract is one of the commercialized products. This *C. striata* cream had experimentally proved to promote wound healing by increasing tensile strength (Baie & Sheikh, 2000). In fact, orally administered *C. striata* extract had scientifically proven to improve osteoarthritis (Tan *et al*., 2004).

Previous study on *C. striata* had implied that physical barrier and biogeography structure had significant impact on genetic structure and hence was great explained historical coalescence of this candidate species (Jamaluddin *et al*., 2011). One of the studies (Adamson *et al*., 2010) had investigated high divergence pattern of this species across geographically wide range of South-East Asia. The study showed that *C. striata* had a highest level of intra-specific divergence among the investigated *Channa spp*. The deep divergence pattern was revealed by sympatric groups in Mekong River Basin suspecting the present of cryptic species. Another population study by Salah & Bhassu (accepted manuscript) using nuclear microsatellite marker had revealed unexpected possible cryptic taxa of *C. striata* in Negeri Sembilan population across Peninsular Malaysia. All these evidences will raise the interesting issue concerning on this species to further investigate its divergence pattern and to monitor population changes in Malaysia.

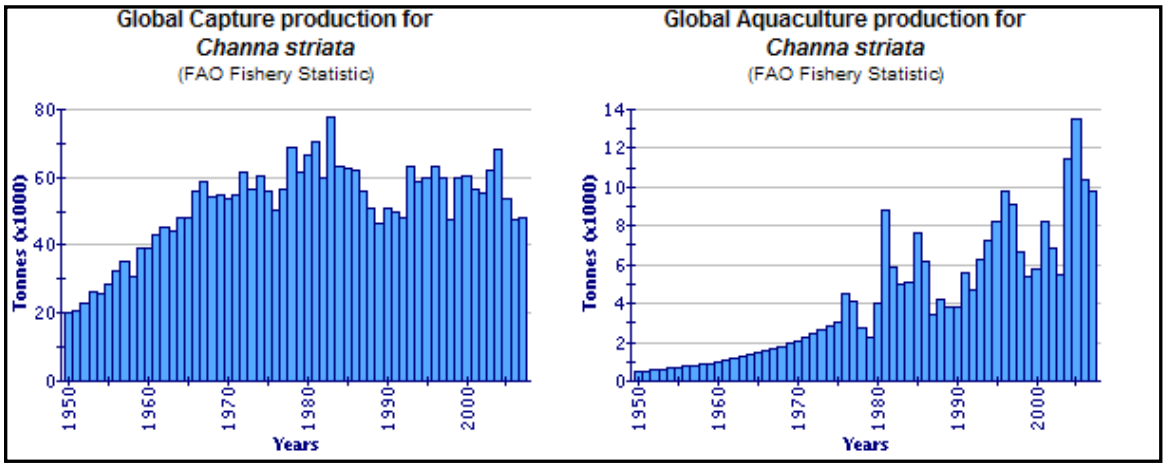

Figure 2.6: Statistic for the global capture and global aquaculture of *Channa striata* from the year of 1950 to 2007 (FAO 2010).

2.3: MORPHOLOGICAL APPROACH

2.3.1: TRADITIONAL TAXONOMIC CLASSIFICATION OF SPECIES

Morphology based technique of species classification had traditionally used to differentiate species using different taxonomy keys such as morphometric characters which is one of the major keys in determining animal's systematics and growth variability (Kováč & Copp, 1999). Morphometric of certain organism is mainly based on anatomy features of the organism being analyzed. In other words, understanding of the biodiversity richness of the organism as well as the knowledge of phylogeography should be taken into account as these will act on to the phenotype of an organism. It is agreed upon that this traditional method of species identification is time consuming and labor intensive as it require a statistical adjustment to the measurement of certain character associated by allometric growth. Additional transformation of the measurement is necessary to eliminate the component of body size that is possible to constitute in variability in further statistical analysis. Transformation of the size dependent data to eliminate the size effect on the shape of organism had been proposed by Lleonart *et al*. (2000).

To date, morphological classification of species which is based solely on morphology is unstandardized due to the difficulties in fixing any character to be considered in measurement. This probably due to the presence of overlapping features across different taxa (Pop *et al*., 2003), environmental stasis whereby environment stress acting on the phenotype and sexual dimorphism in which organism will have different size or different coloration unique to different sex. Indeed, taxonomic systems may vary among different taxonomist when using different taxonomic key for classification (Huang *et al*., 2007). Previously, Beddard, 1895 defined families of earthworm based on male pores in

section XVIII. However this study is discontinued by subsequent researcher (Michaelsen, 1900; Stephenson, 1930) as the pore location is not consistent within and among families of earthworm. Furthermore, identification of earthworm using male pore in male genitalia (Tsai *et al*., 2000; Shen *et al*., 2003) is limited to adult stage.

Due to the enormous effect yielded by morphometric discrimination, Hebert *et al*. (2003a) had introduced a DNA based method of species identification, so called – DNA barcoding, making the possibility of discovery of new species by focusing on animal in all life stage (Hebert *et al*., 2004b; Huang *et al*., 2007). However, with the advance of DNA technology, traditional morphometric measurement is still applied as one of the ways to study the phenotypic variation due to this variability could not suitable to be measured using DNA methodology. Combination of both morphology and genotypic approach will provide a better understanding of population structure (Bhassu & Rashid, 2009).

2.3.2: MULTIVARIATE ANALYSIS

As described in the term "multivariate", multivariate analysis refers to the statistical analysis of data sets which consist of more than one variable. Factor analysis is always takes into account in multivariate analysis. It is a statistical method analyzing how underlying factor influence the observable variables. The variation in some observable measurement is influence by an unobservable common factor. Factor analysis is attempts to examine the correlation between the underlying variance by grouping the highly correlated measurement by a single common factor. It is carried out in the purpose of summarize a data sets of measurement into a smaller number of factors without loosing any informative data (DeCoster, 1998). The amount of variance accounted for each common factor is described by the eigenvalue.

2.4: GENETIC APPROACH

2.4.1: DNA BARCODING

In the previous year, Manwell & Baker (1963) were first used starch gel electrophoresis to identify species. This traditional technique had been employed more than 40 years ago and had been modified by using single gene sequence analysis of ribosomal DNA to investigate evolutionary relationships at a high level (Woese & Fox, 1977). In the late 1970s and 1980s, species identification was further advanced by adopting mitochondrial DNA approaches as molecular systematics (Avise, 1994). Recently, studies of Tautz *et al*. (2002, 2003) lead to successive DNA-based taxonomic system.

Recent developments in DNA-sequencing technology had generated DNA barcoding which is a process introducing the possibility of using variations in short sequences of DNA as labels for species (Waugh, 2007). DNA barcoding is aimed to provide accurate and automated species identifications using molecular species tags based on short and standardized gene regions (Hebert *et al*., 2003a; Hebert & Gregory, 2005). Sequencing entire genomes of animals is an enormous undertaking and time consuming process. A more practical approach was suggested which settle on the sequence of a single gene that is found in all animal life.

The DNA barcoding paradigm had its conceptual birth in Canada (Hebert *et al*., 2003a), which subsequently became the first country establishing a national research network which assembly the DNA barcodes on a large scale. The aim of this national network is to continue work on animals while extending the DNA barcode paradigm to the remainder of eukaryotic life, including fungi, plants and protists. The first network who barcode all biodiversity within national boundaries was The Canadian Barcode of Life Network who representing and involves researchers and funding support from a broad range of institutions across the country (Hanner *et al*., 2005).

DNA barcode was defined as a short sequence of nucleotides isolated from an appropriate part of an organism's genome that is used to identify species. Species were differentiated by the means of intraspecific variation in this selected sequence is relatively less than that observed interspecifically. Particularly, the mitochondrial cytochrome *c* oxidase subunit 1 (COI) gene has been employed as a possible DNA marker for species together with a number of studies in a variety of taxa have been carried out to examine its efficiency (Waugh, 2007). Each species is believed to have its unique nucleotide sequence at COI exhibiting only slight intraspecific K2P divergence. Within-species DNA sequences need to be more similar compared to sequences in different species in order for a barcoding approach to species identification to be succeed (Hubert *et al*., 2008).

The study of Hebert *et al*. (2004a) and those studies from other groups of animals imply that a standard screening threshold of sequence differences was 10X average intraspecific difference for the group under study and this threshold level could speed up the discovery of new animal species. Study of Hebert *et al*. (2003b) on 13,320 congeneric species pairs showed that intraspecific variation was usually less than 1% and rarely more than 2%, while mean interspecific divergence was 11.3% which is 10X of the intraspecific variation.

DNA barcoding is a successful molecular approach through the selection of a suitable segment of DNA. The selected segment must have mutation rate that is slow enough so that intraspecific variation is minimised but sufficiently rapid to highlight interspecific variation (Hebert *et al*., 2003a). It must also be relatively easy to be collected and should have as few insertions or deletions as possible to facilitate sequence alignment.

The success of DNA barcoding has further aided by establishment of the Consortium for the Barcode of Life who devoted to develop DNA barcoding as a global standard in taxonomy (CBOL: http://barcoding.si.edu) and by development of the Barcode of Life Data System (BOLD: http://www.barcodinglife.org) which is a global search tool for DNA barcode (Ratnasingham & Hebert, 2007).

2.4.2: DNA BARCODING AND DISCOVERY OF CRYPTIC SPECIES

Cryptic species, also known as sibling species, is the species that looks morphologically alike but genetically distinct. In the case of cryptic complex, DNA sequencing outcompete morphology as it rely on the gene fragment.

The biological species concept (BSC) defined species as a group of interbreeding populations isolated from the other group and a new species are formed when they are reproductively isolated (Bickford *et al*., 2006). The strong evidence in separating species is the changes in more than one characteristic of an organism together with some independent data such as molecules, morphology and mating signals (Mayr, 1963). However, speciation does not always accompanied by morphological changes as cryptic species is differentiated by nonvisual mating signals and it processes under selection that promotes morphological stasis. Cryptic species taxa are usually visually non-detective by taxonomist as this taxa often distiguished by different in mating pheromones or mating cells in which changes in these signals need not involve morphological change (Bickford *et al*., 2006). According to Mayr (1963), sibling species are normally more developed in chemical senses compared to vision senses. Hence, the occurrence of sibling species is rarest in taxa that depend on vision for mating (Bickford *et al*., 2006). Such taxa is frequently exist in marine organism that more rely on chemical signals for gamete recognition (Palumbi, 1994) and mate choice (Stanhope *et al*., 1992) which rarely leave a visualable morphological imprint (Bickford *et al*., 2006).

The successful finding on cryptic species was demonstrated by Hebert *et al*. (2004b) on *Astraptes fulgerator*, a skipper butterfly "10 species in one" which eventually discovered hidden new species. Furthermore, Huang *et al*. (2007) suspected that one of the species investigated, *Eisenia fetida*, was actually represent the presence of unrecognized sibling species (Huang *et al*., 2007). Hence, DNA barcoding approach using mitochondrial COI fragment outperform traditional taxonomy that rely on morphology when dealing with cryptic species discovery.

2.4.3: DNA BARCODING AND CONSERVATION BIOLOGY

Freshwater fishes are the cheap source of major food protein supplement. Overfishing coupled with environment degradation as well as human disturbance had currently causes depletion in freshwater fish's population and its genetic diversity (Chong *et al*., 2010). The main objective to conserve the diverse of aquatic genetic resources is to maintain ecological equilibrium. Freshwater fishes often had implication on conservation priority due to its lower effective population sizes or lower inter-population migration rates in freshwater environment as compared to marine fish and are more prone to extinction (Lakra *et al*., 2007).

Species conservation may benefit from the approach of DNA barcoding. In addition to species identification, DNA barcoding is indirectly aim to provide alternative route to protect endangered species by preventing illegal transportation of protected species as the exact taxonomic status of certain organism is possible to be recovered even from a tissue sample. Generally, taxonomic units such as species or subspecies will judge conservation decision. Hence, the precise taxonomic status of a population is essential. The National Bureau of Fish Genetic Resources (NBFGR) has integrated DNA barcode for Indian marine fish species facilitating phylogeny inference in determining status of species and subspecies as well as species distribution (Lakra *et al*., 2007).

Barcoding all species in a taxonomic group is necessary on demand to obtain a reference barcode library assigning unidentified organism to its correct species. If all the species of a certain genus are protected, it is valuable to a conservation biologist to protect certain organism that came from within that genus (Barcoding for Endangered Species Conservation - http://www.barcoding.si.edu/PDF/CRES%20Meeting%20Report%20-%20Final.pdf). The checklist published in CITES (Convention on International Endangered Species of Wild Fauna and Flora- http://www.cites.org/) act as a reference list in deciding conservation status of an organism. Any identified species through DNA barcode that shows any matches to the protected species listed in CITES will be protected. Besides, the search tool of BOLD is able to assign the organism to certain collection sites which act as a reference of the geographical distribution of that species. Consequently, international and national agencies who work to maintain lists of endangered species are able to protect and preserve natural habitats of that endangered species.

Various strategies on fish conservation had been implemented including conservation of ecosystems and natural habitats in designating specific areas as protected sites which is known as in situ conservation. Ex situ conservation on the other hand refer to maintain live animals outside their natural environment or through cryopresavation of germplasm where by the existing allelic diversity of rare species can be conserved for future use (Lakra *et al*., 2007).

2.5: MITOCHONDRIAL GENOME

2.5.1: MITOCHONDRIA

Mitochondria are intracellular organelles that involve in cellular metabolism. They arose as free-living bacteria that colonized proto-eukaryotic cells. Other functions of mitochondrial include production of ATP for cellular energy, metabolize fats and participate in apoptosis.

Mitochondrial DNA (mtDNA) is the DNA located in the mitochondrial. In animal, mtDNA is generally a small (15-20kb) genome containing 37 genes which encodes 13 protein subunit of the enzymes of oxidative phosphorylation, two rRNAs of the mitochondrial ribosome, and the 22 tRNAs required for the translation of the proteins encoded by mtDNA (Boore, 1999). Refer **Figure 2.7** for organization of mitochondrial genes illustrated from vertebrate mitochondrial genome.

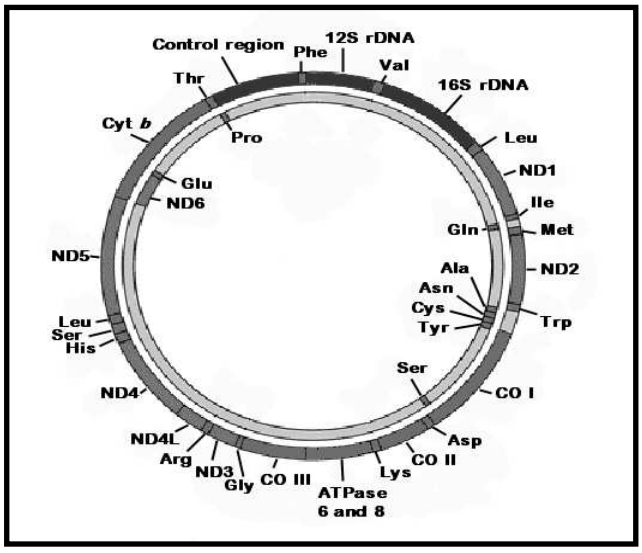

Figure 2.7: Diagrammatic representation of the "conserved" vertebrate mitochondrial genome (Pereira, 2000).

2.5.2: DISCOVERY OF NUMT IN MITOCHONDRIA

Nuclear mitochondrial pseudogenes (numts) are nonfunctional copies of mtDNA in the nucleus. They can be easily coamplified with orthologous mtDNA using conserved universal primers (Song *et al*., 2008). Nuclear integration of mtDNA that give rise to numts, is a widespread phenomenon that has been reported in many eukaryotic clades (Bensasson *et al*., 2001; Richly & Leister, 2004).

Numt come in many sizes, from all types of mtDNA sequences, and bear varying degrees of similarity to their mitochondrial counterparts (Zhang & Hewitt, 1996). Once mtDNA fragments become incorporated into the nuclear genome, they immediately are exposed to different modes of evolution (lower mutation rates due to nuclear DNA repair, a distinct genetic code, and the possibility of recombination), which will influence the divergence patterns between the two sequences (Lopez *et al*., 1994; 1996; 1997).

Numts was predicted that they integrated into the nuclear genome during the repair of chromosomal breaks by nonhomologous recombination (Blanchard & Schmidt, 1995; 1996). Factors influencing the escape of mtDNA from mitochondrial include the action of mutagenic agents and other forms of cellular stress that can damage mitochondria or their membranes (Thorsness & Weber, 1996). It has been postulated that the random insertion of mtDNA into nuclear genomes, associated with such mitochondrial stresses, could be a cause of cancer or of aging (Corral *et al*., 1989; Hadler *et al*., 1998; Shafer *et al*., 1999).

Numts have been checked for fish but the result showed that they absent in such organism (Bensasson *et al*., 2001). Result of the studies of Antunes & Ramos (2005) on *Fugu rubripes*, *Tetraodon nigroviridis*, and *Danio rerio* showed that mitochondrial pseudogenes were present in fish genome and their result is significant enough to reject the hypothesis of absence or rarity of numts in fishes. Antunes & Ramos (2005) data analysis using TBLASTN found that numts homologous to each 13 protein-coding genes indicated that the 13 mitochondrial protein coding genes including COI, had detected in the three fish genome (*Fugu rubripes*, *Tetraodon nigroviridis*, and *Danio rerio*) in their studies.

The accidental co-amplification of numts was due to several reasons including the abundance of numts in the species or individuals being studied, primers used for the PCR reaction (Collura & Stewart, 1995; Zhang & Hewitt, 1996; Sorenson & Quinn, 1998) and which DNA extraction protocol (Wallace *et al*., 1997) and tissue source are used (Sorenson & Quinn, 1998; Greenwood & Paabo, 1999). Several solutions have been suggested to avoid numts co-amplification, such as RT-PCR, long PCR, and mtDNA enrichment (Bensasson *et al*., 2001).

There are several characteristics of numts. The size of numts can be highly variable and not necessarily smaller than 700bp size of DNA barcodes (Richly & Leister, 2004). Song *et al*., (2008) study stated that a large number of these numts have unusually high numbers of point mutations, suggesting that nuclear integration of mtDNA would result in random accumulation of nucleotide changes. Song *et al*., (2008) study also suggested that numts can be indentified by examining the sequence on the basis of in-frame stop codons, indels and nucleotide composition due to different compositional bias between mtDNA and nuclear DNA (Bensasson *et al*., 2001).

2.5.3: MITOCHONDRIA AS MOLECULAR SPECIES TAG

Mitochondrial DNA (mtDNA) offers several advantages over nuclear DNA. According to Drake's observation (Drake *et al*., 1998), the rate of DNA mutation had negative correlation to the genome size. Hence, nuclear DNA undergoes relatively slow mutation compared with mtDNA since nuclear DNA is relatively longer than mtDNA. Consequently, nuclear genome would require a much longer nucleotide sequence than is necessary with mtDNA in order to provide a barcode capable of discriminating species. mtDNA has been widely employed in phylogenetic studies due to its evolution rate is more rapid than nuclear DNA which will lead to the accumulation of differences between closely related species (Brown *et al*., 1979; Moore, 1995; Mindell *et al*., 1997). The differences between populations which only been separated for brief periods of time can be discriminated as a result of the rapid evolution rate of sequence in mtDNA (Hebert *et al*., 2004a).

The success of mtDNA sequences in phylogenetic studies is due to several characteristic. Mitochondrial genome reveal a compact gene packing, consisting little noncoding intergenic nucleotides and some nucleotide overlapping between genes encoded in opposite strands (Cantatore & Saccone, 1987).

Besides, lack of recombination event occurs in mtDNA (Clayton, 1982, 1992; Hayashi *et al*., 1985). mtDNA are mainly maternally inherited (Kondo *et al*., 1990; Gyllestein *et al*., 1991) and it does not undergo recombination with every generation as the nuclear genome does. Therefore, the sequence of the mtDNA remains relatively constant from generation to generation. Any changes in mitochondrial genome within a family will only occur due to random mutation in which this mutation will be inherited by subsequent generation. This will enable mtDNA an interesting subject for studying evolution (Tamara Golden, 2008).

Furthermore, perhaps due to repair inefficiency, mtDNA posses faster sequence evolution as compared to nuclear sequences (Brown *et al*., 1979). In addition, mtDNA are present in multiple copies in a cell (Michaels *et al*., 1982; Robin & Wong, 1988). Each mitochondrion contains several such circular mtDNA and, therefore, several complete sets of mitochondrial genes. Nevertheless, each cell has several mitochondria. Thus, the mitochondrion offers a relatively abundant source of DNA when sample tissue is limited (Waugh, 2007).

Stoeckle *et al*. (2003) suggest that an appropriate target gene for DNA barcoding is conserved enough to be amplified with broad-range primers and divergent enough to allow species discrimination. Study of Hebert *et al*. (2003a) had selected a segment near the 5'-terminus of the mitochondrial COI gene as the barcode region for members of the animal kingdom. Refer **Figure 2.8** for the diagram of the region of 5' end of mitochondrial COI gene. Some of the previous works (Hebert_*et al*., 2003a; Hebert *et al*., 2003b; Hebert *et al*., 2004a; Hogg & Herbert, 2004; Lorenz *et al*., 2005; Smith *et al*., 2005) had successfully differentiated different species indicating that a

DNA barcode derived from an approximately 700 base pair segment of the mitochondrial COI gene representing highly efficient system for the species identification and discovery of animal life.

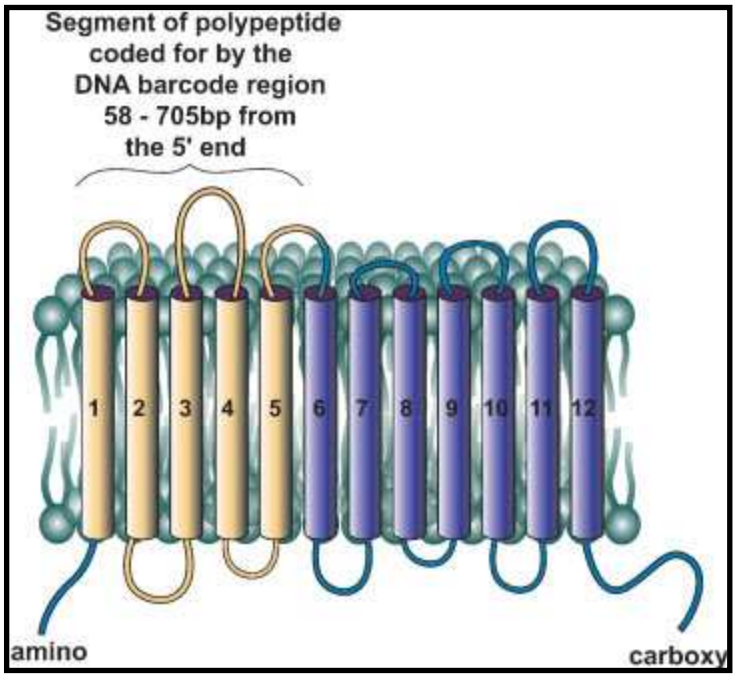

Figure 2.8: Diagrammatic representation of selected 5' end of mitochondrial cytochrome *c* oxidase subunit 1 (COI) gene as the barcode region (Waugh, 2007).

2.6: CYTOCHROME *C* OXIDASE SUBUNIT I (COI)

2.6.1: INTRODUCTION OF COI

Cytochrome *c* oxidase is a large transmembrane protein found in the mitochondrion, which is highly conserved for all species that employ oxidative phosphorylation for metabolism. It functions as the terminal electron acceptor in the respiratory chain by catalysing the reduction of oxygen to water and pumping protons across the membranes of the cristae (Research Collaboratory for Structural Bioinformatics (RCSB) Protein Databank (PDB) - Cytochrome c oxidase website: http://www.bioc.rice.edu/~graham/CcO.html) . The protein comprises several subunits of nuclear origin and three subunits synthesised in the mitochondrion which are known as subunits I, II and III. COI is the catalytic subunit of the enzyme predominantly imbedded in the membrane of the mitochondrial crista. Previous research found that intraspecific variation in this gene is generally <10% of that observed between species. Moreover, insertions and deletions are rarely found in this gene (Blaxter, 2004).

2.6.2: COI AS DNA BARCODE

Various gene regions including COI, cytochrome *b* or even nuclear gene have been employed for species-level biosystematics (Hajibabaei *et al*., 2007). However, DNA barcoding based on adoption of a 'global standard' and an approximately 700-base fragment of the 5' end of the mitochondrial COI gene has gained designation as the barcode region for animals (Hebert *et al*., 2003a). Recent studies associated with CBOL have generally selected a 648bp segment of the COI gene, starting from the 5' end, to generate a suitable barcode (Barcode of Life website: http://www.barcodinglife.org/).

Past phylogenetic work has often focused on mitochondrial genes encoding ribosomal (12S, 16S) DNA, but their application in broad taxonomic analyses is not encouraged by the presence of insertions and deletions (indels) that greatly complicate sequence alignments (Doyle & Gaut, 2000). Due to functional constraints hold some nucleotide positions constant and intraspesific diversity exists at others, the survey of nucleotide diversity needs to be more comprehensive. The impact of functional constraints can be reduced by focusing on protein-coding gene due to most shifts at the third nucleotide position of codons are weekly constrained by selection because of their four-fold degeneracy. Indels are rarely occurs in nature in protein coding gene since most of the indels lead to a shift in the reading frame. Therefore, the 13 protein-coding genes in animal mitochondrial genome are the better target as molecular species tag (Hebert *et al*., 2003a).

COI does have several important advantages over other mitochondrial region to serve as a DNA barcode region. The universal primers used for this gene are very robust and hence enabling recovery of its 5' end from representatives of most of the animal phyla (Folmer *et al*., 1994; Zhang & Hewitt, 1997). Besides, as common with other protein-coding genes, the rate of molecular evolution is about three times greater than that of 12S and 16S rDNA due to its third-position nucleotides show a high incidence of base substitutions (Knowlton & Weigt, 1998). Furthermore, the evolution of this gene is rapid enough to allow the discrimination of closely allied species as well as phylogeographic groups within a single species (Cox & Hebert, 2001; Wares & Cunningham, 2001). Besides, due to the rate of changes in COI amino acid sequence occur more slowly than those in cytochrome *b* or any other mitochondrial gene (Lynch & Jarrell, 1993), COI gene is more likely to provide deeper phylogenetic insights than alternatives such as cytochrome *b* (Simmons & Weller, 2001). Consequently, if cytochrome *b* served as a barcode region, by examining amino acid substitutions, it may be possible to assign any unidentified organism to a higher taxonomic group such as phylum or order before examining nucleotide substitutions to determine its species identity (Hebert *et al*., 2003a).