# PHYLOGEOGRAPHY AND GENETIC VARIATION OF THE BIGHEAD CATFISH, *Clarias macrocephalus* (GÜNTHER, 1864) FROM PENINSULAR MALAYSIA AND MEKONG RIVER BASIN BASED ON MITOCHONDRIAL AND MICROSATELLITE MARKERS

**NAZIA BINTI ABDUL KADAR**

**UNIVERSITI SAINS MALAYSIA**

**2014**

# PHYLOGEOGRAPHY AND GENETIC VARIATION OF THE BIGHEAD CATFISH, *Clarias macrocephalus* (GÜNTHER, 1864) FROM PENINSULAR MALAYSIA AND MEKONG RIVER BASIN BASED ON MITOCHONDRIAL AND MICROSATELLITE MARKERS

**by**

**NAZIA BINTI ABDUL KADAR**

**Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy**

**August 2014**

## ACKNOWLEDGEMENTS

First and foremost I'm very thankful to Allah s.w.t. for blessing me with good health and strength from the beginning of my research until submitting my dissertation. I'm forever grateful to the best supervisor ever, Prof. Siti Azizah Mohd Nor, for her priceless guidance, patience and motivation to complete my dissertation. Not forgetting invaluable ideas that I received from Assoc. Prof. Khoo Khay Huat from Universiti Sains Malaysia, Dr. Thuy Nguyen from NACA, Dr. Geoff Chambers from Victoria University of Wellington, New Zealand and Dr. Eleanor Adamson from Natural History Museum, London, during my studies.

I would like to thank The Ministry of Science, Technology and Innovation, Malaysia for funding this project under E-Science Grant (305/PBIOLOGI/613207) and Universiti Sains Malaysia, Penang, under Postgraduate Research Grant Scheme (1001/PBIOLOGI/843002). I'm thankful to Mr. Azhar from Pusat Pengembangan Akuakultur, Jitra, Kedah, and Dr. Bui Min Tham from College of Aquaculture and Fisheries, Can Tho University, Vietnam, for their assistance in obtaining samples for my research. My appreciation also goes to my labmates from Lab 308, Pau, Suzana, Katie, Farhana, Kak Adel, Su Yin, Jamsari, Lim, Lutfi, Danial, Faisal, Amirah, Balkhis, Hasmah, Dilla, Zaza, Elham and Layla for their help during my studies.

My sincere thankfulness goes to my beloved parents, Abdul Kadar and Hamidah, and my younger brother Muhammad Arsyad, for their continuous support, help and motivation throughout my studies. My special thanks go to my dearest husband, Arrif, for his love, care and encouragement especially during my writing. Finally, I'm thankful to everyone that was involved directly or indirectly in completing my doctoral dissertation.

**TABLE OF CONTENTS**

## LIST OF TABLES

## LIST OF FIGURES

\

## LIST OF PLATES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AMOVA | Analysis of molecular variance |
| Cyt b | Cytochrome *b* |
| D-loop | Control region |
| dNTP | Dinucleotide triphosphate |
| FDR | False discovery rate |
| HWE | Hardy-Weinberg equilibrium |
| IAM | Infinite allele mutation |
| LD | Linkage disequilibrium |
| Ma | Million years ago |
| MP | Maximum parsimony |
| MSN | Minimum spanning network |
| NJ | Neighbor-joining |
| *PIC* | Polymorphism information content |
| SAMOVA | Spatial analysis of molecular variance |
| SSM | Single stepwise mutation |
| SSR | Single sequence repeat |
| LGM | Last glacial maximum |

## LIST OF SYMBOLS

| | |
|---|---|
| $F_{CT}$ | Variance among groups |
| $F_{ST}$ | Variance within population |
| $F_{SC}$ | Variance among populations within group |
| $N_m$ | Gene flow estimates |
| H | Number of haplotypes |
| V | Number of variable sites |
| *h* | Haplotype diversity |
| $\pi$ | Nucleotide diversity |
| $H_O$ | Observed heterozygosity |
| $H_E$ | Expected heterozygossity |
| $A_R$ | Allelic richness |
| $N_A$ | Number of allele |
| $F_{IS}$ | Inbreeding coefficient |
| $H_S$ | Gene diversity |
| *r* | Frequency of null alleles |

**FILOGEOGRAFI DAN VARIASI GENETIK KELI BUNGA, *Clarias macrocephalus* (GÜNTHER, 1864) DARI SEMENANJUNG MALAYSIA DAN LEMBANGAN SUNGAI MEKONG BERDASARKAN PENANDA MITOKONDRIA DAN MIKROSATELIT**

## ABSTRAK

Penanda mitokondria dan mikrosatelit telah digunakan untuk mengkaji variasi genetik dan hubungan filogenetik di kalangan populasi ikan keli, *Clarias macrocephalus* dari Semenanjung Malaysia dan dua rantau lembangan sungai Mekong iaitu Kemboja dan Vietnam. Untuk analisis mitokondria, 332 individu dari 19 populasi telah dibuat penjujukan untuk gen Cyt *b* (609 bp) manakala 382 individu dari 19 populasi untuk gen D-loop (479 bp). Kedua-dua gen mengesan kebarangkalian hibrid di antara *C. macrocephalus* dan lain-lain spesies atau kehadiran spesies yang belum pernah direkod. Kajian seterusnya ke atas jujukan *C. macrocephalus* yang jelas mendedahkan 39 haplotip untuk gen Cyt *b* dan 47 haplotip untuk jujukan D-loop. Kepelbagaian haplotip berjulat antara $h$=0 hingga 0.908 ± 0.048 (Cyt *b*); $h$=0 hingga 0.892 ± 0.044 (D-loop) manakala kepelbagaian nukleotida, antara $\pi$=0 hingga 0.008 ± 0.005 (Cyt *b*); $\pi$=0 hingga 0.012 ± 0.007 (D-loop). Nilai-nilai tersebut paling tinggi dalam populasi Kemboja. Kedua-dua gen mendedahkan aliran gen yang tinggi terutamanya antara populasi bersebelahan. Walau bagaimanapun, gen D-loop menunjukkan penstrukturan genetik yang lebih tinggi untuk populasi dalam rantau berbanding dengan gen Cyt *b* yang berevolusi dengan perlahan. Pada fasa seterusnya dalam kajian ini, lapan penanda mikrosatelit polimorfik untuk *C. macrocephalus* telah dibangun menggunakan kaedah

penghibridan terpilih terubahsuai. Penanda ini seterusnya digunakan untuk menilai variasi genetik 393 individu *C. macrocephalus* dari 15 populasi dan juga spesies *Clarias* yang terpilih. Kebanyakan lokus sangat polimorfik kecuali lokus *NCm-H2* yang mempunyai nilai *PIC* terendah. Bilangan alel dalam populasi *C. macrocephalus* adalah antara 36 hingga 111. Dua populasi berada dalam keseimbangan HWE, dua populasi menunjukkan lebihan heterozigus manakala selebihnya menunjukkan lebihan homozigus kerana alel nol dalam beberapa lokus dan pembiakbakaan dalaman pada beberapa populasi. Nilai pasangan $F_{ST}$ yang kecil tetapi sangat ketara telah diperolehi bagi semua populasi berbanding dengan analisis mitokondria yang menunjukkan beberapa nilai yang tidak ketara. Tiada populasi yang menunjukkan bukti kejadian cerutan. Selain daripada beberapa populasi, variasi genetik dari lingkungan sederhana kepada tinggi telah diperhatikan di mana populasi Sungai Mekong paling pelbagai. Analisis selanjutnya menunjukkan dua unit evolusi, iaitu Semenanjung Malaysia dan Lembangan Sungai Mekong. Kajian ini menunjukkan bahawa *C. macrocephalus* berasal dari Indo-China kemungkinan Kemboja sebelum memasuki Semenanjung Malaysia. Kombinasi kedua-dua jujukan mitokondria dan penanda mikrosatelit telah menghasilkan maklumat yang terperinci mengenai variasi genetik dan struktur populasi *C. macrocephalus*. Ini adalah penting untuk pemilihan induk ikan untuk program pembiakbakaan serta pemuliharaan populasi liar *C. macrocephalus.*

**PHYLOGEOGRAPHY AND GENETIC VARIATION OF THE BIGHEAD CATFISH, *Clarias macrocephalus* (GÜNTHER, 1864) FROM PENINSULAR MALAYSIA AND MEKONG RIVER BASIN BASED ON MITOCHONDRIAL AND MICROSATELLITE MARKERS**

## ABSTRACT

Mitochondrial and microsatellite markers were utilised to investigate the genetic diversity and phylogenetic relationships among catfish, *Clarias macrocephalus* populations from Peninsular Malaysia and two Mekong River Basin regions, namely Cambodia and Vietnam. For mitochondrial analyses, 332 individuals from 19 populations were sequenced for Cytochrome *b* (609 bp) gene while 382 individuals from 19 populations were analysed for the D-loop (479 bp) gene. Both genes detected possible hybrids between *C. macrocephalus* and other species or presence of previously undocumented species. Further analyses on unambiguous *C. macrocephalus* sequences revealed 39 haplotypes for Cyt *b* gene and 47 haplotypes for D-loop sequences. Haplotype diversity ranged from $h$=0 to 0.908 ± 0.048 (Cyt *b*); $h$=0 to 0.892 ± 0.044 (D-loop) and nucleotide diversity ranged from $\pi$=0 to 0.008 ± 0.005 (Cyt *b*); $h$=0 to 0.012 ± 0.007 (D-loop). These values were highest in the Cambodian populations. Both genes revealed high gene flow especially between adjacent populations. However, D-loop gene demonstrated higher genetic structuring of populations within region compared to the slower evolving Cyt *b* gene. In the next phase of the study, eight polymorphic microsatellite markers for *C. macrocephalus* were developed using the modified selective hybridization method. These markers were further utilised to assess the genetic variation of 393 individuals of *C.*

*macrocephalus* from 15 populations as well as cross-amplification in selected *Clarias* species. Most of the loci were highly polymorphic except for locus *NCm-H2* that harboured low *PIC* value. Number of alleles in *C. macrocephalus* populations ranged from 36 to 111 alleles. Two populations were in HWE, two populations showed heterozygous excess while the rest showed homozygous excess due to null alleles in several loci and inbreeding in several populations. Low but highly significant pairwise $F_{ST}$ values were obtained for all the populations compared to mitochondrial analyses that displayed several non-significant values. None of the populations showed evidence of bottlenecks. Apart from a few populations, moderate to high genetic variation were observed, where the Mekong River populations was the most diverse. Further analyses revealed two evolutionary units, which were Peninsular Malaysia and the Mekong River Basin. This study suggests that *C. macrocephalus* has originated from Indo-China possibly Cambodia before invading Peninsular Malaysia. Combination of mitochondrial and microsatellite markers has provided detailed information on the genetic variation and population structure of *C. macrocephalus*. This is important for selection of broodstocks for a breeding programme as well as conservation of wild populations.

# CHAPTER 1

# INTRODUCTION

## 1.1 Introduction

*Clarias macrocephalus* or locally known as 'keli bunga' is a popular food fish in Malaysia, Thailand and the Philippines due to its tender and delicate taste (Somnuek *et al*., 2009). Since this catfish is well distributed across Asia and is economically important, it has been subject to culturing activities in several regions of Southeast Asia for local consumption with varying success (Thalathiah, 1998; Poompuang and Na-Nakorn, 2004). According to FAO (2000), the annual production of Asian *Clarias* species in 1999 was 150,000 metric tonnes/year. In Malaysia, the fry of *C. macrocephalus* was successfully produced in the mid-1980s by the Freshwater Fish Research Centre, Batu Berendam (Thalathiah *et al*., 1988; Thalathiah *et al.*, 1990) but however had low survival rate and the yield never exceeded 2 tonnes/ha/year and could not match the demand of local consumers. Catfish production in Malaysia which are sold as live or frozen product at USD 6.32/kg, amounted to 14,693.42 metric tonnes or USD 11.4 million during the year 2004 (Department of Fisheries, 2004) and 81,041 metric tonnes in 2009 (Anon, 2011).

On the other hand, Na-Nakorn *et al.* (1998) reported that 80% of Thailand's farmers successfully cultured the hybrids of *C. macrocephalus* and *C. gariepinus* for commercial purpose since the late 1980s. The annual production of *Clarias* in Thailand was at 82,000 metric tonnes which valued at USD 52 million (Department of Fisheries, 2003). Of these, Na-Nakorn *et al*. (2004) reported that in Thailand 90% of total *Clarias* production with 50,000 metric tonnes/year involved hybrids. The

hybrid technology became widely utilised among local farmers to produce hybrids that matched the high qualities of 'keli bunga' and soon the hybrids dominated the total *Clarias* production (Thalathiah, 1998).

*Clarias macrocephalus* also known as 'keli bunga' or 'keli kampong' in Malaysia is found to be generally limited to the northern region of Peninsular Malaysia in paddy fields, irrigation canals, stagnant pools or streams (Mohsin and Ambak, 1983; Lee *et al.,* 1993). They are air-breathing catfish and tolerant of harsh environment and consequently can move to adjacent habitats using their pectoral fins for spawning, feeding or seeking shelter (Ali, 1993; Pouyaud *et al*., 2009) during the dry season. However, wild populations of *C. macrocephalus* are depleting because of habitat destruction, over-fishing and competition from the alien African catfish, *C. gariepinus* and its hybrids (Wiecaszek and Krzykawski, 2010; Vidthayanon and Allen, 2013). Therefore, genetic assessment is vital in order to conserve the remaining pure and wild populations of *C. macrocephalus.*

According to Doveri *et al*. (2008), genetic markers such as mitochondrial and nuclear markers are considered powerful tools to discover genetic uniqueness of individuals, populations or even species. Several researchers have stated that mitochondrial DNA (mtDNA) markers have been informative in discriminating genetic structure at large scales over time, such as between major phylogeographic lineages (Ketmaier *et al*., 2004; Yang *et al.,* 2012). The evolutionary rate as well as the genetic differentiation of mtDNA among populations are thought to be approximately 5 to 10 times higher than that exhibited by nuclear genes (Tzeng *et al.,* 2007; Zhao *et al.,* 2008; Yu *et al.,* 2010) accounting for greater sensitivity. This

higher level of resolution provides a reliable method of examining relationships among closely related taxa.

MtDNA is a powerful tool for tracking ancestry as it is maternally inherited and has been used to track the ancestry of many species back hundreds of generations (Tzeng *et al*., 2007; Nwafili *et al.*, 2009; Yu *et al*., 2010). These molecular markers have been employed in a number of research applications involving various aquatic organisms such as in phylogeographic studies (Ketmaier *et al.*, 2004; Yang *et al.,* 2012), population genetics (Tzeng *et al.*, 2007; Sah *et al.*, 2011) and species taxonomy and hybrid identification (Kyle and Wilson, 2007; Wouters *et al*., 2012).

The choice of the mtDNA region examined depends upon the phylogenetic level of the hypothesis that is being tested. These levels range from examining intraspecific relationships (control region or D-loop and NADH dehydrogenase subunit 5/6 genes) (Miya and Nishida, 2000; Yu *et al.,* 2010; Lee *et al*. 2011) as well as interspecific and intergenera relationships between closely related organisms through moderately evolving genes (Cytochrome *b*- Cyt *b*) (Ketmaier *et al.*, 2004; Sah *et al*., 2011; Yang *et al.,* 2012) to the slowly evolving 12S and 16S ribosomal RNA (rRNA) (Nwafili *et al.,* 2009) and Cytochrome Oxidase I (COI) genes (Spies *et al.,* 2006; Kim *et al*., 2011; Pereira *et al*., 2013) for family level comparisons. Nevertheless, levels of evolutionary rates of these genes may sometimes differ in various organisms (Carvalho and Pitcher, 1995). There are however limitations in relying on only a single type of gene. For instance, the phylogenies and population structures derived from mtDNA data may not reflect those of the nuclear genome due to gender-biased migration (Savereide, 2012) or introgression (Chow and Kishino,

1995; Aboim *et al*., 2010) and therefore nuclear markers such as microsatellites are often utilized to complement.

Microsatellite markers have been widely used and have become the marker of choice for fish population studies (Ha *et al.,* 2009; Aldenhoven *et al.,* 2010; Tian *et al*., 2013). Microsatellites consist of multiple copies of tandemly arranged simple sequence repeats (SSRs) that range from 1 to 6 base pairs and is assumed to be evenly distributed in the genome on all chromosomes (Liu and Cordes, 2004). The marker is codominant in inheritance and is highly sensitive in detecting genetic variability within and between populations (Nasren *et al*., 2009; Langen *et al.* 2011; Hoban *et al*., 2013). Alam and Islam (2005) in their study used eight microsatellite markers to investigate the genetic structure of Indian major carp species (*Catla catla* Hamilton) in Bangladesh and revealed that the genetic variation of the hatchery population was lower than the river populations. Such information is essential for management of the populations in order to maintain their genetic quality. This marker is also useful in studies of parentage assignment, genome mapping, kinships, and stock structure (Chen *et al.,* 2012; Luo *et al*., 2013; Tian *et al*., 2013).

In aquaculture, microsatellite could be used for estimation of relatedness between potential breeding pairs in parentage assignment (Jeong *et al*., 2007; Schreier *et al.,* 2012; Luo *et al*., 2013). Such information would serve as an additional tool while carrying out selection programme. It can also be used in pedigree analysis to minimize the unwanted loss of effective population size over the course of selection. Most of the performance traits are QTLs (quantitative trait loci) and their analysis should greatly assist in selection programme. These QTLs could be mapped to its relative position by constructing a linkage map. This is conducted by

assigning polymorphic DNA markers such as microsatellite to chromosome positions based on family segregation (Moen *et al*., 2009; Hoh *et al*., 2013; Zhang *et al*., 2013). Waldbieser *et al*. (2001) investigated 293 polymorphic microsatellite loci in channel catfish (*Ictalurus punctatus*) and concluded that seven loci were closely linked to the sex-determining chromosome region. However, primer design can be problematic, associated with *de novo* isolation from species that are being examined for the first time due to the very high substitution rate associated with non-coding regions compared with coding regions (Zane *et al*., 2002). As a result, microsatellite markers are typically species-specific with limited cross-species amplification. Therefore, new markers are needed when initiating a study on a new organism. A few microsatellite markers are already available for *Clarias macrocephalus* (Na-Nakorn *et al*., 1999; Sukmanomon *et al.*, 2003; Sukkorntong *et al.*, 2008) but more are needed for a comprehensive population study. Thus far, microsatellite markers have been developed for many organisms ranging from common carp (Ji *et al*., 2012), Chinook salmon (Naish and Park, 2002), sturgeon (Zeng *et al.*, 2013), catfish (Sukmanomon *et al.*, 2003; Yue *et al*., 2003) and frogs (Eterovick *et al*., 2011).

The current study was focused on determination of the genetic variation of wild populations of *C. macrocephalus* that were obtained from available localities throughout Southeast Asia, focusing on Peninsular Malaysia, Cambodia and Vietnam since the studied catfish species are well distributed in these regions and are in high demand by local people. Both mtDNA and nuclear markers have been found to be very powerful in assessing population genetics of many fish species (Nwafili *et al*., 2009; Sah *et al.*, 2011; Tian *et al*., 2013). The information obtained from this study will provide baseline data of *C. macrocephalus* populations that are depleting from the wild. Successful long-term management of wild stock, preservation and

conservation could be carried out by understanding the population history and current genetic constituent of the studied species.

## 1.2 Objectives

The objectives of the current study were;

1. To investigate population diversity, population structuring and related historical demographic events of *C. macrocephalus* in three regions of Southeast Asia using mitochondrial Cytochrome *b* and D-loop genes.

2. To develop novel microsatellite markers for *C. macrocephalus* species and cross-amplification in other *Clarias* species.

3. To assess genetic variation of *C. macrocephalus* populations in the three Southeast Asian regions utilising the newly developed microsatellite markers.

Two different types of molecular markers namely mitochondrial Cyt *b* and D-loop (control region) and nuclear microsatellites were utilised to assess the genetic variation of the studied species. This thesis contains three working chapters. Both mitochondrial Cyt *b* and D-loop genes were utilised to investigate the genetic structuring of wild populations and demographic events as described in Chapter 3. Chapter 4 describes the development of eight novel microsatellite markers that were successfully isolated from *C. macrocephalus.* Chapter 5 reports on the investigation of genetic variation and population structure of wild populations based on the newly developed microsatellite markers.

# CHAPTER 2

# LITERATURE REVIEW

## 2.1 Biogeography of Peninsular Malaysia and the Mekong Basin

Malaysia is located near the equator and lies on the Sunda Shelf, between latitude $2^0$30'N and longitude $112^0$30'E. The Sundaland which comprises of the Malay Peninsula, Borneo, Java and Sumatra is one of the major biodiversity hotspots with more than 15,000 endemic plant species, 770 bird species, 380 mammal species, 450 reptiles species, 240 amphibians species and 1,400 species of freshwater fishes (Myers *et al*., 2000; Brooks *et al*., 2002; Corlett, 2009a). The backbone of Peninsular Malaysia, the Titiwangsa Mountains running from the Thai border southwards to Negeri Sembilan, with a length of 480 km from north to south effectively divides Peninsular Malaysia into the east and west coast regions. Thus, the eastern states of Kelantan, Terengganu and Pahang are isolated from western states of Perlis, Kedah, Penang, Perak, Selangor, Negeri Sembilan and Melaka (Mohsin and Ambak, 1983; Lee *et al.,* 1993). Other natural barriers include lakes and rivers such as Lake Temenggor, Lake Kenyir, Perak River and Pahang River. Previous studies have reported that the Titiwangsa Mountains act as natural barrier and restricts gene flow between populations of east and west such as studies on the marble goby, *Oxyeleotris marmoratus* (Ruzainah, 2008), the Cyprinidae, *Barbonymus schwanenfeldii* (Kamarudin and Esa, 2009), the freshwater terrapin, *Batagur baska* (Nor Karmila, 2009), the climbing perch, *Anabas testudineus* (Jamsari *et al.,* 2010) and the striped snakehead, *Channa striata* (Siti Balkhis *et al.*, 2011; Rahim *et al.*, 2012; Tan *et al.*, 2012).

The Mekong River is the longest river in Southeast Asia and has about 1,200 recorded fish species (Ngamsiri *et al.*, 2007). It runs 4,800 km from Tibetan plateau through China, Myanmar, Laos, Thailand, Cambodia and Vietnam to the South China Sea (Figure 2.1). The Mekong basin catchment area covers 795,000 $km^2$ with a mean discharge volume of 15,000 $m^3/s$ (Phan *et al*, 2009). After flowing through Phnom Penh, the Mekong River is linked to the biggest lake in Cambodia, the Tonle Sap Lake, by Tonle Sap River (120 km long). Tonle Sap is the largest freshwater body in Southeast Asia and extends over 300 km from northwest Cambodia to the Mekong River at Phnom Penh. Only about 20-30% of the Tonle Sap Lake waters originate from the Tonle Sap drainage as a result of its unique hydrological system. During the flood season (June to October), the rise of the Mekong River forces the Tonle Sap River to flow backwards into the Tonle Sap Lake. Then the Mekong divides into the main Mekong River and the Bassac River, which flows through the Mekong Delta of Vietnam to the South China Sea. Thus, the basin has an annual inflow of Mekong waters (57%) during the wet season when the Mekong water level increases and the lake empties again during the dry season. This lake acts as a vital fish breeding ground and flood mediator for the Mekong River (Matsui *et al*., 2006).

This ecosystem is believed to be one of the most productive inland waters in the world, with a high abundance of fish where flooded forests and rice fields offer shelter and breeding grounds for fish and other aquatic animals. Migration of different fish species between Tonle Sap Lake and the Mekong River is extensive and diverse. A study by So *et al*. (2006) based on microsatellites on the sutchi catfish, *Pangasianodon hypophthalmus* in the Mekong River showed that the populations has high genetic diversity. Furthermore, several divergent haplotype groups were also detected thus relating to the Pleistocene climate fluctuations. As an

example, populations of the bronze featherback *Notopterus notopterus* from the Mekong River and Tonle Sap were examined using mtDNA control region by Takagi *et al.* (2006) and it was found that the lake population was genetically separated from the river population. In contrast, a study by Ngamsiri *et al*. (2007) found that the Mekong giant catfish, *Pangasianodon gigas* from Thailand and Cambodia was genetically the same.

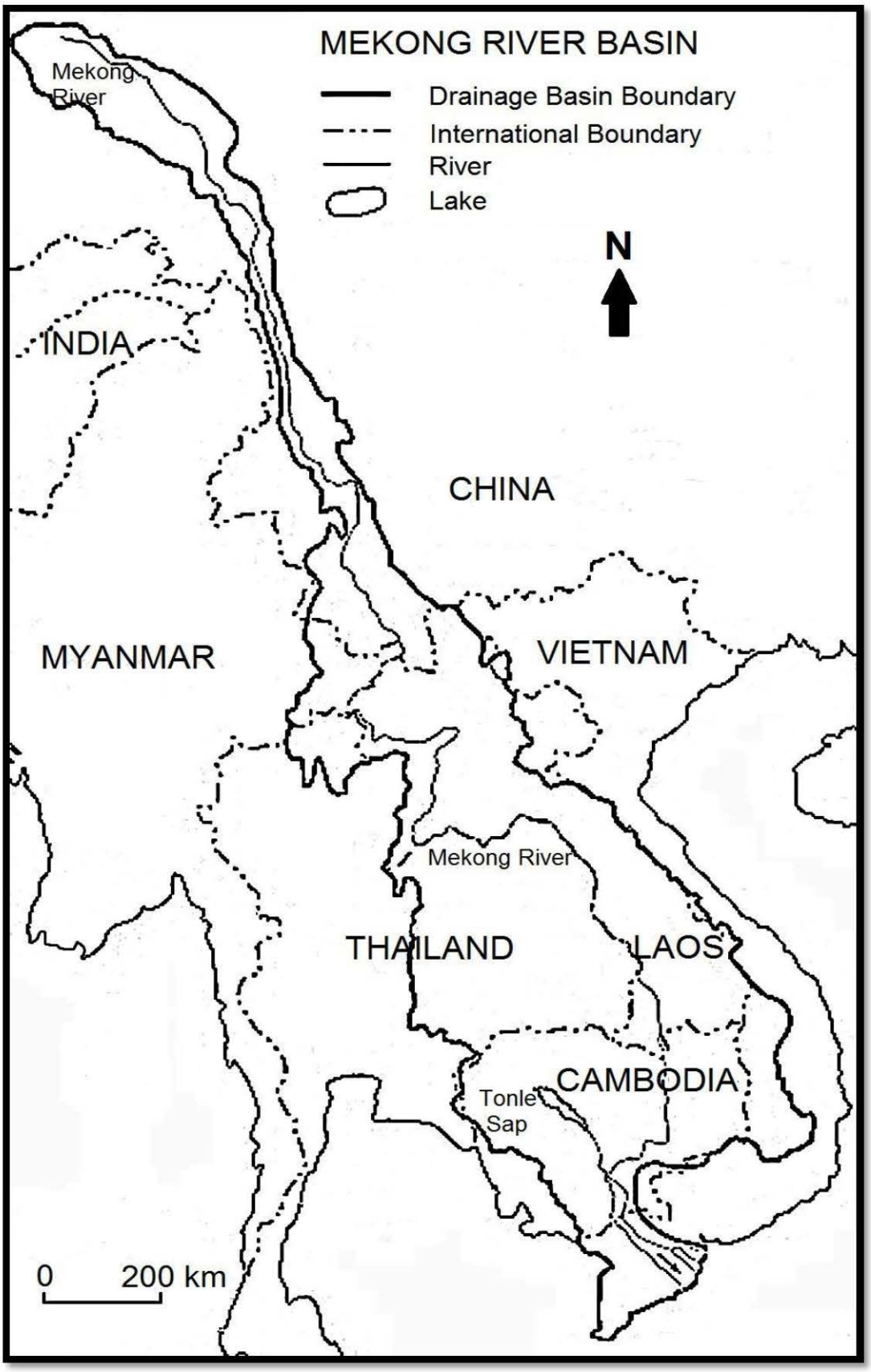


Figure 2.1: Map showing the Mekong River system (modified from Matsui *et al*., 2006).

The distribution of species or populations and their genetic differentiation depends on biological, environmental and historical factors. Several studies have shown that the genetic structure of freshwater fish populations was influenced by fluctuations of sea level and alternate wet and dry seasons during the past (Yang *et al.*, 2009; Lukoschek *et al*., 2011). The last marine regression (approximately 110 meters below actual sea level) which is dated 20,000 years before present led to the disappearance of the South China and Java Sea. During this time, the exposed Sunda Shelf act as a land bridge mass between Indo-China, the Greater Sunda Island and also the Malay Peninsula thus creating possible connections between river drainages (which are at the present time disconnected) and a possible dissemination of freshwater ichthyofauna (Sathiamurthy and Voris, 2006) (Figure 2.2). By contrast, the transgression (approximately 6 meters above actual sea level) was responsible of the disappearance of many lowland areas and the decreasing of many freshwater populations, excepting in refugial areas such as large river systems and highlands (Pouyaud *et al.,* 1998). According to Kottelat (1989), 44% from 263 fish species from the Malay Peninsula also occur in the Mekong, 47% in the Chao Phraya and 66% in Borneo, Sumatra and Java which was assumed to be associated with dispersal during the Pleistocene Epoch.

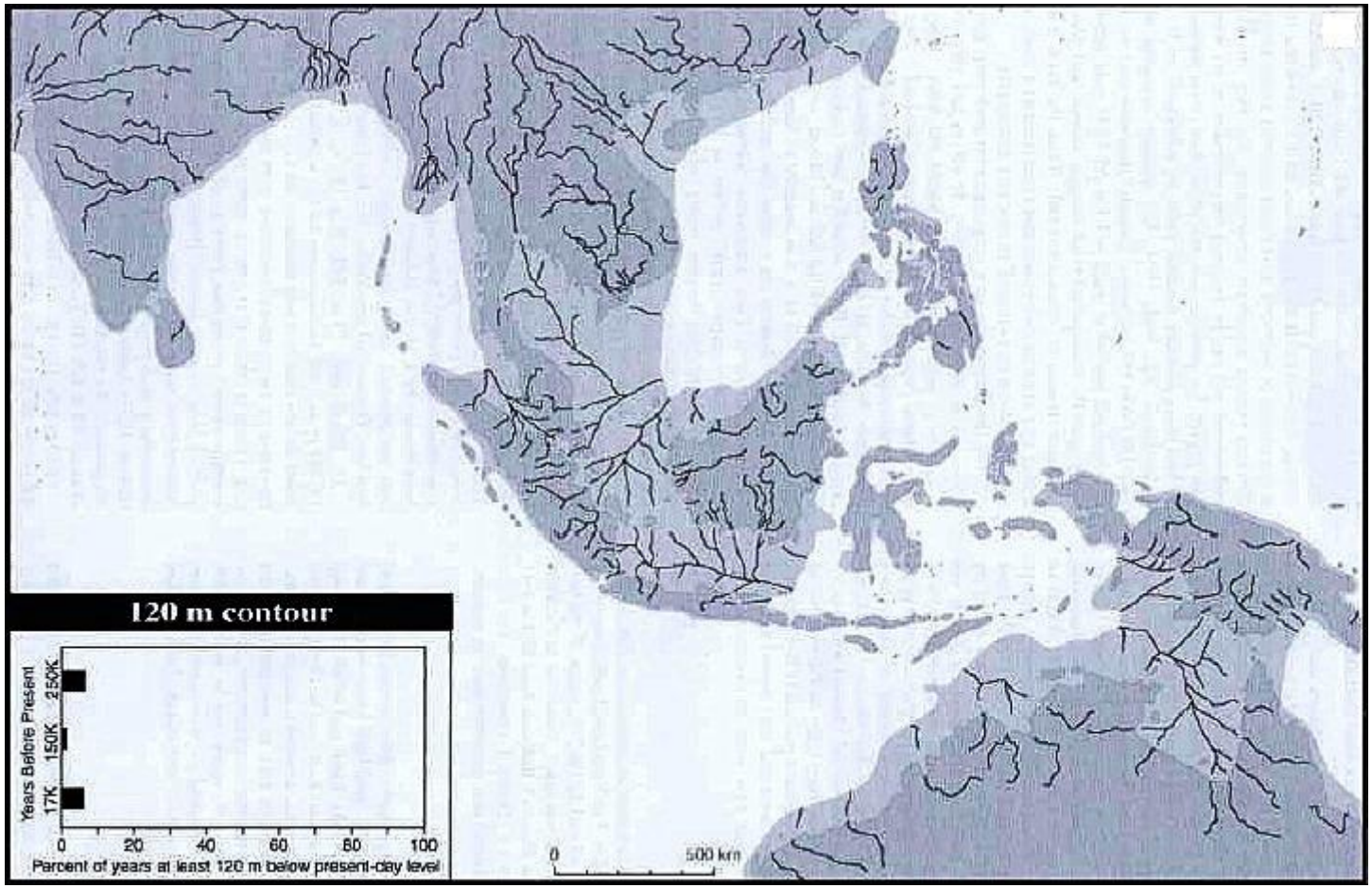

Figure 2.2: Map showing ancient river systems (modified from Voris, 2000).

## 2.2 Evolutionary history and distribution of catfish species

Catfishes are one of the most diversified groups of fish with 3093 species in 478 genera and 36 families. The family Clariidae or walking catfish has 113 species in 16 genera, three of them are endemic to Asia, 12 endemic to Africa, and one (*Clarias* Scopoli, 1777) is present on both continents, the latter being represented by 56 species (Ferraris, 2007). However, it is impossible to evaluate the evolutionary biogeography of catfishes as available data is very limited. According to Briggs (1970), 64% of freshwater catfish species occurs in Central and South America, 19% in Africa, 15% in Eurasia and Southeast Asia and only 2% in North America. Catfish species diversity is lower in Europe and North America as their existence is related to Pleistocene glaciations that caused extinction.

The origin of catfish is assumed to have occurred before the splitting of Gondwanaland in the late Mesozoic when Africa and India separated, followed by East Antarctica and Australia, New Zealand and then West Antarctica and South America (Teugels, 1996). Agnese and Teugels (2005) reported that Clariidae originated 50 Ma (million years ago). However, recent lineages are believed to colonise Africa and Southeast Asia independently from Asian origin about 15 Ma. Colonization of Africa was enabled through terrestrial connections and / or brackish water bridges within the Arabian Plate during the Lower Miocene (Otero and Gayet 2001). In spite of the relatively recent colonization of Southeast Asia, *Clarias* has achieved remarkable diversity in species, morphology and habitat. According to Pouyaud *et al.* (2009), the oldest fossils belonging to the genus *Clarias* were discovered in Oman on the Arabian Plate originating from the Oligocene period, (30

Ma) (Otero and Gayet 2001). However they suggested that Asian *Clarias* started its speciation at 30 Ma.

The presence on the Arabian Plate of neurocranial bone remains of Clariidae belonging to *Heterobranchus* or to *Clarias* on the site of Taqah (30–31 Ma) and the record of *Heterobranchus* in East African Lower Miocene (16–23 Ma) suggested that the ancestor of African Clariidae has an Arabian origin and colonized Africa during the Lower Miocene (16–20.5 Ma). *Clarias* diversification at the Eocene–Oligocene boundary in Asia coincided with the collision of the Indian and Burmese Plates (Hall 2002). The majority of the Asian *Clarias* that is present on the Sunda Shelf (Sumatra, Borneo and Java) appeared about 20 Ma. Pouyaud *et al*. (2009) reported that *C. macrocephalus, C. punctatus, C. batrachus* and *C. fuscus* are mainly found in Bangladesh, Vietnam and Thailand. During the Last Glacial Maximum (LGM) period, the Sundaland region which encompasses the Sunda Shelf, Asian mainland (Myanmar, Thailand, Laos, Cambodia, Vietnam, the Malay Peninsula and Singapore), Sumatra, Java and Borneo was an exposed landmass and was crucial for *Clarias s*pecies dispersal (Voris, 2000; Bird *et al.,* 2005; Pouyaud *et al*., 2009). As reported by Mohsin and Ambak (1983) and Ali (1993), endemic *Clarias* species from Thailand such as *C. macrocephalus* and *C. batrachus* are mainly found in Northern states in Peninsular Malaysia most probably as this region is adjacent with Thailand.

## 2.3 *Clarias* species

The body forms observed in Clariid species represent the adaptations to habitat as catfish do not follow orthogenetic series in morphological and osteological evolution (Agnes and Teugels, 2005). As reported by Ng (1999), head shape in

*Clarias* species does not change significantly with ontogeny neither the degree of mouth closure or barbel articulation (due to the inflexibility associated with the heavy ossification of the neorocranium) and can be reliably used as a diagnostic character for species discrimination (Ng and Kottelat, 2008). With the exception of a revision of *Clarias* Scopoli, 1777 (Teugels, 1986) and *Heterobranchus* Geoffroy Saint-Hilaire, 1809 (Teugels *et al.*, 1990), no reliable keys are available to identify other representatives of this family (Teugels *et al.*, 1999). *Clarias* Scopoli, 1777 is the largest genus in the Old World catfish family Clariidae, with 48 species (Teugels, 1986; Ng, 2004) distributed in Africa and Asia.

Most *Clarias* species are found in Africa (Teugels, 1986) and about 18 species are found in Southeast Asia. *Clarias insolitus* from Barito River drainage in southern Borneo (Ng, 2003a) and *C. nigricans* (Ng, 2003b) from Mahakam River drainage in eastern Borneo have been described recently. *C. insolitus* can be differentiated from other *Clarias* in having a long and thin anterior fontanel and hypertrophied sensory canal pores on the head and body that are easily visible to the naked eye. While *C. nigricans* can be differentiated by having a narrow snout, dark violet grey coloration with small white spot on flanks, a narrow head and large serrations on the anterior edge of the pectoral spine (Ng, 2003b).

Teugels *et al.* (2001) depicted eight Asian catfish; *Clarias meladerma* (Bleeker, 1846), *C. macrocephalus* (Günther, 1864), *C. intermedius* (Teugels *et al.*, 2001), *C. batrachus* (Linnaeus, 1758), *C. punctatus* (Valenciennes, 1840), *C. nieuhofii* (Valenciennes, 1840), *C. leiacanthus* (Bleeker, 1851a) and *C. teijsmanni* (Bleeker, 1857). All Asian *Clarias* species show regular pattern for neuromasts displacement on the flanks.

To date ten *Clarias* species have been documented in Malaysia namely *C. batrachus* (Linnaeus, 1758), *C. gariepinus* (Burchell, 1822), *C. nieuhofii* (Valenciennes, 1840), *C. leiacanthus* (Bleeker, 1851a), *C. teijsmanni* (Bleeker, 1857), *C. macrocephalus* (Günther, 1864), *C. anfractus* (Ng, 1999), *C. batu* (Lim and Ng, 1999), *C. planiceps* (Ng, 1999) and *C. sulcatus* (Ng, 2004). However, *C. batrachus*, *C. macrocephalus* and the exotic African catfish *C. gariepinus* are the most studied species in this family as they are widely used in aquaculture, aquarium fish trade, studies on biochemistry, behaviour, diseases and others (Teugels, 1996; Ng and Kottelat, 2008; Manuel *et al.*, 2014).

*Clarias macrocephalus* has a round and broad occipital process compared to *C. batrachus* that has angular and narrow shaped occipital process. The distance from the dorsal fin base to tip of occipital process is seven to eight times in length of head in *C. macrocephalus* and five to six times in *C. batrachus* (Mohsin and Ambak, 1983). On the other hand, *C. gariepinus* has a head which is rectangular and pointed in dorsal outline and the distance between the occipital process and the base of the dorsal fin is short (Teugels, 1986). *Clarias batu* and *C. sulcatus* are endemic in Pulau Tioman and Pulau Redang. The genus *Clarias* from Southeast Asia is one of the problematic groups of Clariidae. In a phylogenetic study of Clariids, Agnese and Teugels (2001a, 2005) concluded that the current systematics of the Clariid catfishes requires a review. However, more Clariid species need to be studied before introducing a new nomenclature.

## 2.4 *Clarias macrocephalus*

### 2.4.1. Taxonomic and nomenclature of *Clarias macrocephalus*

The focus of this project, *C. macrocephalus* (Günther, 1864) (Plate 2.1) also known as bighead catfish is an economically important air-breathing catfish in Southeast Asia. The standard classification of this fish according to the Integrated Taxonomic Information System (ITIS) is as below.

Kingdom: Animalia

Phylum: Chordata

Subphylum: Vertebrata

Superclass: Osteichthyes

Class: Actinopterygii

Subclass: Neopterygii

Infraclass: Teleostei

Superorder: Ostariophysi

Order: Siluriformes

Family: Clariidae

Genus: *Clarias*

Species: *macrocephalus*

Plate 2.1: Dorsal view of *Clarias macrocephalus*. Universal colour wheel is included for colour comparisons.

### 2.4.2 Morphological characteristics of *Clarias macrocephalus*

The morphological characters of *Clarias macrocephalus* were first described by Teugels (1986, 1996) and Teugels *et al.* (1999). This air-breathing catfish has an extremely short and rounded occipital process and a very high dorsal fin (10% of the standard length). The body is elongated with long dorsal (without any spine) and anal fins, no adipose fin, head dorsally depressed, elongated neural spines and strong venomous spine at pectoral fin, a broad terminal mouth with four pairs of barbels and eyes with free orbital margin and located dorsolaterally (Teugels, 1986). According to Mohsin and Ambak (1983), the body is muddy black at the dorsal and lateral surfaces and whitish at the ventral side. The pelvic fin is whitish and the rest of the fins are blackish. The males can be identified by the presence of elongated conical shaped urogenital papillae while females have an oval or round opening (Mollah and Tan, 1982). They have four pairs of barbels. The maxillary barbels extend to the middle of the pectoral fin base while nasal barbels reach more than two thirds the distance of the gill opening.

### 2.4.3 Habitat and distribution

Clariidae occur naturally in North, Central and South America, Africa, Eurasia, Southeast Asia, Japan and Australasia (Teugels, 1996). According to several authors, *C. macrocephalus* is an introduced species in Peninsular Malaysia (Mohsin and Ambak, 1983; Froese and Pauly, 2011) although this is disputed by other researchers and database (eg. IUCN Redlist; Ali, 1993; Vidthayanon and Allen, 2013). The species is also threatened by aquaculture and (through hybridization and competition) by escaped hybrids (Na-Nakorn 2004) across the northern parts of its range, but this is not thought to affect the Malaysian populations.

In Peninsular Malaysia, this species is known as ‘keli bunga’ or ‘keli kampong’ and have been reported in Perlis, Kedah, Perak and Terengganu (Mohsin and Ambak, 1983; Lee *et al*., 1993). They are widespread in rice fields, irrigation canals, stagnant pools, ditches and also streams. Normally, they remain at the bottom of water body except for a few occasional trips to the surface to gulp air and feed on aquatic insects, shrimps and small fishes. They have both gills and aborescent organs (Teugels and Adriaens, 2003) that enable them to breathe in atmospheric air and to survive in hypoxic environments such as swamps, dried pools and rainforests (Pouyaud *et al.,* 2009). Therefore this catfish is able to be buried in mud during dry seasons and can move on land for few hundred meters using their pectoral spines and by making sinuous movements with their body. They move to one habitat to another to breed, to search for food and to find shelter (Ali, 1993).

## 2.5 *Clarias macrocephalus* in aquaculture

The African catfish, *C. gariepinus* and its hybrid with the local *C. macrocephalus* have become very popular among farmers due to their hardiness, easy to culture, high growth rates, and availability of fry (Kechik, 1995). The catfish that is widely cultured now is the hybrid between *C. batrachus*, which is indigenous, and *C. gariepinus*, an exotic African catfish which was introduced in the early 1980s. The catfish is an important protein source and commercially cultured freshwater fish for consuming (Marte, 1989). Southeast Asian countries like Malaysia, Cambodia and Vietnam has been practising rice-fish farming as early as 1928 and this has become the main source of freshwater fish supplies especially for *C. macrocephalus* and *Channa striata* (Halwart and Gupta, 2004) .

In Malaysia, the genus *Clarias* is one of the most important cultured fish groups. Their production in 2009 amounted to 81,041 metric tonnes which was dominated by the exotic African catfish (Anon, 2011). *C. macrocephalus* has been harvested in low cost rice-fish farming system in North Kerian, Perak, Malaysia, that uses natural reproduction of the wild fish (Ali, 1993). Fish from rice fields has become an extra income for rice farmers (Tan *et al*., 1973; Ali, 1990) and the harvested fish are exported to Thailand and Singapore. Popular catfish species in Malaysian aquaculture are keli kayu (*C. batrachus*), keli bunga (*C. macrocephalus*), African catfish (*C. gariepinus*), patin (*Pangasius sutchii*) and baung (*Mystus nemurus*). However the production of rice-field fishes has been deteriorating due to introduction of the double-cropping system and also widespread use of pesticides and herbicides (Ali, 1990). *Clarias macrocephalus*, although more superior in terms of meat quality constitutes only a small amount to this value, primarily because of its slow growth rate and low resistance to diseases. Concerns are that wild populations are depleting because of habitat and prey competition with *C. gariepinus* as well as through overfishing (*C. macrocephalus* is preferred by consumers). They are sold commercially as live priced at RM20/kg or frozen. This species is also now documented in the IUCN Redlist of threatened species in 2014 (Vidthayanon and Allen, 2013).

## 2.6 Application of molecular markers in population studies

Population genetics which display the distribution of genetic variability in a population is influenced by the evolutionary processes of mutations, migration, selection and random drift (Hansen 2003; Mojekwu and Anumudu, 2013). Genetic data is important for effective management and conservation of a species, including

in wild and aquaculture stocks of fish. The Polymerase Chain Reaction (PCR) technique became a revolutionary tool in investigation of genetic variation of fish populations since its development over the years (Ferguson *et al*., 1995). To date, many PCR-based molecular techniques are available such as DNA sequencing, DNA barcoding, randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), microsatellites genotyping, single nucleotide polymorphism (SNP) and expressed sequence tag (EST) markers to examine stock structure besides other approaches such as tagging, morphometrics and meristics, cytogenetics and many more. The marker of choice for a particular research is largely dependent on the expertise, facilities and available funding. Molecular markers such as mitochondrial and microsatellite markers are widely used in population studies to investigate the phylogeography and population genetics in organisms including in fish as will be described below.

### 2.6.1 Mitochondrial DNA (MtDNA)

Mitochondrial DNA has many advantages in various types of genetic analyses due to its maternal transmission, rapid rate of evolutionary changes, transmission without recombination, and haploid inheritance (Avise, 1994; Briolay *et al*., 1998; Liu and Cordes, 2004). Mitochondrial DNA in most animals range from 16 to 18 kb and encodes 13 proteins, 2 ribosomal RNAs, 22 transfer RNAs and a regulatory region known as the control region in vertebrates or the displacement loop (D-Loop) in invertebrates (Wilkinson and Chapman, 1991). The size of mtDNA in teleost fish range between 16,000 to 19,000 base pairs and in channel catfish for example is 16,497 base pairs in length (Waldbieser *et al.*, 2003).

Vertebrate species studies has shown that sequence divergence accumulates quickly in mitochondrial than in nuclear DNA (Brown, 1985). Therefore, mutation rate in mtDNA is faster due to the lack of repair mechanisms during replication, and smaller effective population size as it is maternally inherited. As its results of its rapid rate of evolution, mitochondrial markers are beneficial in revealing relationships among closely related species (Chauhan and Rajiv, 2010). However, the rates of evolution vary among mtDNA genes. D-loop is the segment where the replication and the transcription of the molecule is started and therefore evolve rapidly, cytochrome *b* has moderate mutation rate while 16S rRNA is the slowest evolving gene.

Various studies utilising mtDNA genes in genetic variability assessment have been conducted and such data is important and can indicate the life histories and degree of evolutionary isolation (Okumus and Ciftci, 2003). Mutation rates of each gene vary. For instance, cytochrome *b*, cytochrome c oxidase subunit I-III are moderately conserved while ATPase 6/8, NADH subunit 1-6/4L are more variable (Miya *et al*., 2006). However, the two rRNA genes are highly conserved and therefore more advantageous in phylogenetic studies (Ortí and Meyer, 1997). Cytochrome *b* gene is most commonly used in phylogenetics and phylogeography of fish as well as population studies (Rahim *et al*., 2012; Yang *et al.*, 2012) while the control region or D-loop is used to detect genetic variation and population structure (Lee *et al.* 2011; Terencio *et al*., 2012).

On the other hand, cytochrome c oxidase subunit 1 (COI) has been vastly used in DNA barcoding to differentiate closely related species (Hebert *et al.*, 2002; Pereira *et al.*, 2013). More recently, taking advantage of the different mutational

rates, many researchers have utilised combination of several genes for a more holistic study (Yang *et al*., 2010). Thus, mtDNA has proven to be an effective marker for investigating stock structures (Kochzius, 2009), identification of fish species (Ward *et al.*, 2009; Lago *et al*., 2012) or broodstock (Senanan *et al*., 2004), determination of species origin (Hardman *et al*., 2005), detection of introgression of genome (Na-Nakorn *et al*., 2004; Wouters *et al*., 2012) and tracking of released animals (Mohindra, 2007). However, the major disadvantage of this marker is that the data may not be complete as reflected by nuclear marker if gender-biased migration, selection or introgression occurs in a population (Chow and Kishino, 1995).

### 2.6.2 Microsatellite markers

Microsatellites also known as simple sequence repeats (SSRs), variable number tandem repeats (VNTR) and short tandem repeats (STR), are tandem repeats of one to six bases, found in both coding and noncoding regions in all prokaryotic and eukaryotic organisms. Microsatellites are mainly dinucleotides (30 to 67%) and the rest are mostly trinucleotides and tetranucleotides (Li *et al.*, 2001). The mutation rates are between $10^{-2}$ to $10^{-6}$ per locus per generation (Ellegren, 2000), which is typically explained by polymerase slippage during DNA replication, results in the differences in number of repeat units (Tautz, 1989).

Microsatellite has numerous alleles and is highly polymorphic and has become the marker of choice in fish population genetic studies. The polymorphisms obtained provide vital information and could be applied to identify species, strain or hybrids (Wouters *et al*., 2012; Agbebi *et al*., 2013) in population analysis, conservation and management of fish stocks (Alam and Islam, 2005; Luo *et al*., 2012). Microsatellite loci with higher number of alleles per locus (>20) is suitable for