DNA BARCODING OF MALAYSIAN MARINE FISH, PHYLOGENETICS AND PHYLOGEOGRAPHY OF THE SNAPPERS (PERCIFORMES: LUTJANIDAE)

ADIBAH BINTI ABU BAKAR

UNIVERSITI SAINS MALAYSIA
2017

DNA BARCODING OF MALAYSIAN MARINE FISH, PHYLOGENETICS AND PHYLOGEOGRAPHY OF THE SNAPPERS (PERCIFORMES: LUTJANIDAE)

by

ADIBAH BINTI ABU BAKAR

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

November 2017

ACKNOWLEDGEMENT

In the Name of Allah, Most Gracious, Most Merciful.

First and foremost I want to thank my advisor Dr. Darlina Md Naim. It has been an honor to be her first Ph.D. student. I appreciate all her contributions of time, ideas, and funding to make my Ph.D. experience productive and stimulating. I am also thankful to my supervisory committee, Prof. Siti Azizah Mohd Nor (Universiti Sains Malaysia, USM) and Prof. Wei-Jen Chen (National Taiwan University, NTU) for their time, interest, helpful comments and insightful questions.

The members of the LAB 308 have contributed immensely to my personal and professional time during my research. The group has been a source of friendships as well as good advice and collaboration. I am especially grateful for the fun group of LAB 308: Danial Hariz, Jamsari (the first two who had invited me into the lab), Dr. Adelyna, Dr. Su Yin, Dr. Nazia, Dr. Evan, Dr. Alex, 'Sifu' Lim, Fong, Ana Farhana, Siti Hasmah, Balkhis, Faisal, Lia Juliana, Nurul Farhana, Wani, Ayesha, Aliah, Kak Nurul (SEAFDEC), Kak Masazurah (FRI, Batu Maung), new batch 2015; Zu, Mel, Abg Fadhli, Hadi, Nabilah, alongside numerous bachelor students as well as rotation of industrial trainings students who have come through the lab.

I gratefully acknowledge the funding sources that made my Ph.D. work possible. I was funded by the Skim Latihan Akademik Bumiputera (SLAB) from Universiti Pendidikan Sultan Idris (UPSI) and Ministry of Education, Malaysia for four years. My work was also supported by Research University Grant (1001/PBiology/815087) and Delivering Excellence Grant (1002/PBIOLOGI/910317).

Lastly, I would like to thank my family for all their love and encouragement. For my parents who raised me with a love of science and supported me in all my pursuits. And most of all for my loving, supportive, encouraging, and patient husband Syazwan Saidin whose faithful support during stages of this Ph.D. is so appreciated.

I would like to dedicate this valuable journey to my two precious bundle of joy, Muhammad Luqman Al-Hakim and Nur Aisyah Al-Humaira.

Thank you.

Adibah Abu Bakar
Universiti Sains Malaysia (2012-2017)

TABLE OF CONTENTS

CHAPTER 3 – GENERAL MATERIALS AND METHODS

CHAPTER 4: FIRST LARGE-SCALE DNA BARCODING ASSESSMENT OF COMMERCIAL MARINE FISHES IN MALAYSIAN WATER: AN APPLICATION FOR SUSTAINABLE FISHERY MANAGEMENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BI	Bayesian Inference
BOLD	Barcode of Life Data system
bp	Base pairs
COI	Cytochrome oxidase subunit 1
cyt b	Cytochrome b
DoF	Department of Fisheries
EGR	Early Growth Response
EP	Eastern Pacific
EPB	Eastern Pacific Barrier
FAO	Food and Agriculture Organization
FRI	Fisheries Research Institute
GenBank	Genetic sequence database
IO	Indian Ocean
IOP	Isthmus of Panama
IP	Indo-Pacific
I*ss*	Index of substitution saturation
I*ss.c*	Critical index of substitution saturation
K2P	Kimura-2-Parameter
Max	Maximum
$MgCl_2$	Magnesium chloride
Min	Minimum
min	Minute
ML	Maximum likelihood

Mya	Million years ago
NJ	Neighbour-Joining
No.	Number
PCR	Polymerase chain reaction
RH	Rhodopsin
SE	Standard error
sec	Second
sensu	In the sense of
tmrca	The most recent common ancestor
TTE	Terminal Tethyan Event
WA	Western Atlantic

LIST OF SYMBOLS

n	Sample size
µL	Microliter
^{0}C	Degree Celsius
E	East
M	Molar
mg/ml	Milligram per milliliter
mM	Millimole
N	North
P	Probability value
R^2	Regression value
rpm	Rotation per minute
s	Transition
U	One unit of taq polymerase
U/µL	Unit of taq polymerase per microliter
v	Transversion

PENGKODAN DNA IKAN MARIN DI MALAYSIA, FILOGENETIK DAN FILOGEOGRAFI BAGI IKAN 'SNAPPER'

(PERCIFORMES: LUTJANIDAE)

ABSTRAK

Kajian ini dilaksanakan dengan menggunakan teknik Barkod DNA yang menawarkan potensi yang baik sebagai alat pengecaman untuk klasifikasi tangkapan ikan di Malaysia. Secara keseluruhan, kajian ini berjaya untuk menkodkan sebanyak 107 spesies, 69 genus, 36 famili dan 10 order ikan komersial Malaysia. Juga didapati bahawa, ikan jenahak, ikan bulu, ikan bayan dan ikan biji nangka dikesan mengalami penspesiesan 'cryptic' atau terdiri daripada spesies yang tidak diketahui. Hasil daripada kajian ini menunjukkan bahawa klasifikasi secara morfologi sahaja didapati tidak selalunya menjurus tepat kepada spesies individu atau kumpulan spesies. Pengenalan konvensional sehingga peringkat spesies yang menggunakan ciri morfologi didapati sukar bagi genus ini terutama bagi kes-kes berkaitan spesimen juvana dan dewasa yang mempamerkan variasi warna. Menggunakan gen COI, nilai perbezaan genetik yang tinggi (K2P = 6.1%) diperolehi antara kumpulan *L. lutjanus*, LL1 dan LL2, serta analisis seterusnya menggunakan jujukan COI dan *cyt b* menunjukkan tapak penggantian nukleotid diagnostik bagi setiap kumpulan. Oleh itu, pada masa ini, kajian ini mendedahkan bahawa wujud satu leluhur yang mengandungi takson yang tidak dikenali bagi 'kompleks ikan jenahak berbaris kuning'. Sejumlah 3612 bp jujukan yang selanjar dengan kombinasi dua gen mitokondria (654 bp gen COI dan 1116 bp gen cyt b) dan dua gen nuklear (897 bp gen RH dan 945 bp gen EGR1) telah digunakan untuk membina semula kerangka filogenetik Lutjanidae yang komprehensif. Hasil dari kajian semasa juga

menunjukkan bahawa ‘lutjanids’ menunjukkan distribusi monofiletik tidak resiprokal dalam rantau Pasifik Timur (EP), Indo-Pasifik (IP), Lautan Hindi (IO) dan Atlantik Barat (WA). Dengan penggunakan penentukuran berasaskan fosil sebagai kekangan terhadap model jam molekul bagi analisis biogeografi dalam kajian semasa, didapati bahawa ‘lutjanids’ WA, EP dan IO adalah berasal dari leluhur Indo-Pasifik.

DNA BARCODING OF MALAYSIAN MARINE FISH, PHYLOGENETICS AND PHYLOGEOGRAPHY OF THE SNAPPERS (PERCIFORMES:LUTJANIDAE).

ABSTRACT

This study, implemented DNA Barcoding which offers great potential as a reliable identification tool to classify catches in Malaysia. Overall, this study had successfully barcoded a total of 107 species, 69 genera, 36 families and 10 order of commercial Malaysian fishes. It is found that, the snappers, threadfin fishes, parrot fishes and goatfishes sampled were detected to either experience cryptic speciation or consist of unknown/undescribed species. Conventional identification to species level using morphological characters were found to be difficult for this genera especially in cases where juvenile and adult specimens exhibit variation in colouration. Using COI gene, deep genetic divergence (K2P = 6.1%) values was obtained between LL1 and LL2 groups of *L. lutjanus* and the subsequent analysis of both COI and *cyt b* sequences revealed diagnostic nucleotide substitution sites exclusively to each group. Thus, at present, this study exposed that at least one lineage represents a currently unrecognized taxon of the 'yellow-lined snapper complex'. A total of 3612 bp aligned sequences corresponding to the combinations of two mitochondrial genes (654 bp of the COI gene and 1116 bp of cyt b gene) and two nuclear genes (897 bp of the RH gene and 945 bp of the EGR1 gene) were also employed to reconstruct a comprehensive phylogenetic framework of the Lutjanidae. Results from current study did indicate that lutjanids presented a non-reciprocal monophyletic distribution within Eastern Pacific (EP), Indian Ocean (IO), Indo-Pacific (IP) and Western Atlantic (WA) regions. Using a fossil-based calibration to constraint the relaxed

molecular clock model for biogeography analysis in current research, it is found that WA, EP and IO lutjanids derived from Indo-Pacific lineages.

CHAPTER 1

INTRODUCTION

The marine fisheries sector plays an important role in the Malaysian economy, contributing to the national Gross Domestic Product (GDP), employment and foreign exports and representing a rich source of protein for Malaysians. Located on the edge of the "coral triangle" and home to an estimated >2000 species of marine fishes (Allen, 2008), the marine waters surrounding the Malay coastlines that support these activities are some of the most biodiverse regions in the world.

Approximately 200-300 species of marine fishes are landed in the major Malaysian landing sites, with an average of 50-100 species being displayed for sale daily in fish markets (DoF, 2014). Additional species may appear seasonally, with certain species predominating market landings during the monsoons, while other permanent resident species of estuaries, bays and reef areas are landed throughout the year (DoF, 2014). The diversity of wild species harvested and variety of fisheries operations in the country makes assembling accurate detailed catch data challenging.

This thesis covers three areas of investigation in which each is focused on resolving specific issues. The objectives of current study are as follows:

1) Large-scale DNA Barcoding assessment of commercial marine fishes in malaysian water: An application for sustainable fishery management.
2) Phylogenetic analysis of *Lutjanus* species (Pisces: Lutjanidae) in Malaysian fisheries catch.
3) Phylogenetic and historical biogeography analyses of the family Lutjanidae, using multi-gene approach and fossil-calibrated tree.

Finally, overall discovery was summarized and concluded in Chapter 7 along with description of future recommendations.

CHAPTER 2

LITERATURE REVIEW

2.1 Fish diversity in the marine realm

A considerable amount of literature has been published on the importance of marine sphere. Most research to date has tended to focus on the economic and ecological values of marine biota especially as a source of food and indicator of environmental health. Furthermore, publications on major threats to marine biodiversity which include overharvesting, habitat degradation, pollution, global warming, biological invasions and anthropogenic stressors (Costello *et al.*,, 2010) are also increasingly gaining coverage. Still, much of the literature until today, fails to identify the current number of marine species in the ocean. Although scientists have estimated the ocean to contain approximately 2.2 million species (Mora *et al*., 2011), it is actually challenging to measure the incredible diversity that lies beneath the waves.

In recent years, scientists have been making serious headway in trying to understand the marine diversity. One of the most extensively studied groups of organism in the marine realm is fish (Mora *et al*., 2003). Fish represent a keystone in present-day monitoring of environmental health of marine ecosystems (Thomsen *et al.*, 2012). More than 50,000 available species names of fishes have been documented, with over 31,000 of them currently regarded as valid species. Eschmeyer (2010) reported that new marine species are being catalogued at a rate of about 100–150 per year. For easy access to broad information on fish, there are

several online websites which can be used. The two most referred sites are the Catalog of Fishes (Eschmeyer and Fong, 2016) and FishBase (Froese and Pauly, 2016).

2.2 Assessing the diversity through conventional methods

Great effort has been devoted to the study of fish species identification for several decades. Traditionally, external morphological features, including body shape, pattern of colours, scale size and count, number and relative position of fins, number and type of fin rays, or various relative measurements of body parts (Strauss and Bond, 1990) are utilized for classification. However, major drawbacks from only using morphological criteria for fish identification is the limited characters for differentiation purposes in certain taxa (Callejas and Ochando, 2001). Teletchea (2009) also informed that even with whole specimens, there might be only small interspecific variations. Although many improvements have been made to comprehend such weakness especially by adding more phenotypic characters, morphological ambiguities still persist during species recognition research.

Examples of identication issues regarding dependency on morphological characters are not uncommon. In 2002, Iff At, demonstrated that number of gill rakers, can be used to differentiate morphologically similar species of Korangi Creek mullets. However, in a later study, Lindsey (1981) deduced that gill raker characteristics is highly influenced by environment through her experiment with coregonids. She discussed that though gill-raker count is less subject to phenotypically induced variation from environmental influences than are most

morphological characters, it is actually not insusceptible. She advised that when employing even relatively stable characters such as gill raker counts, one should be aware of the influence of the biological effect that exists with them. Besides the usage of gill rakers, analysis of otoliths is an alternative to describe fish species (Pierce and Boyle, 1991; Granadeiro and Silva, 2000). Otoliths are commonly referred to as "earstones" or "fish ear bones". However, the main limitation of this tool are it is destructive, meaning that the extraction of otoliths kills the fish, and otoliths can easily break during extraction/ manipulation. Moreover, otolith analysis is also very difficult because of the concave form of the otoliths and overall variability of its shape.

Vecchione *et al.* (2000) reported that there are many factors that affect fish identification. The most significant factors include experience level of identifiers, reliable taxonomic references, distinction of morphological characters, condition of specimens, life stages of specimen and cryptic speciation. Based on these, they suggested that a more uniform system should be developed. It is notably known that erroneous identification used in analysis or publication can seriously affect future inferences. As species identification is of paramount importance especially to monitor biodiversity (Vecchione *et al.*, 2000), researchers have attempted to improvise conventional methods for identifying fish species without relying exclusively on morphological features.

2.3 Molecular approach for marine fish identification

Knowlton (1993) pointed out that it is no surprise that scientists took the opportunity provided by the advancement of molecular methods to clarify many ambiguities in conventional taxonomy. Problems in morphological diagnosis are usually associated in identification of the early life stages such as eggs and larvae. Furthermore, large phenotypic plasticity and sexually dimorphic species as well as cryptic species that are widely distributed in marine systems also contribute to the complications.

At the onset, molecular methods used in species identification, including fish, were based on the separation and characterization of specific proteins using electrophoretic techniques, such as isoelectric focusing (IEF) (Rehbein, 1990) and capillary electrophoresis (CE) (Kvasnička, 2005), high performance liquid chromatography (HPLC) (Hubalkova *et al.*, 2007) and even immunoassay systems, such as EnzymeLinked ImmunoSorbent Assay (ELISA) (Asensio and Montero, 2008). These techniques have been widely reviewed (Mackie *et al.*, 1999; Civera, 2003; Moretti *et al.*, 2003; Hubalkova *et al.*, 2007). Even though most of the mentioned methods are of considerable value in certain instances, they are not suitable for routine sample analysis because proteins lose their biological activity after animal death (Telechea, 2009) and their presence and characteristic depend on the cell types (Asensio and Montero, 2008). This has now caused attention to turn towards DNA as a source of information. As an alternative to protein analysis, DNA-based identification methods have currently been explored and extensively developed (Teletchea *et al.*, 2005).

There are several significant advantages of DNA over protein analysis (Telechea, 2009). The most significant are; firstly, DNA is more resistant and thermo stable than proteins and even though DNA might be altered by various processing stages (example by canning or heating), it is still possible to PCR-amplify small DNA fragments (with adequate information to allow identification) and secondly, DNA could potentially be extracted from any substrate because it is present in almost all cells of an organism. A variety of DNA-based methods are potentially available for use in fish species identification nowadays. However, these vary in their range of applications, complexity and costs. In addition, tremendous advances in molecular biology have now rendered possible the identification of any species using DNA in virtually any kind of organic substrate, such as muscle, fin or blood (Lockley and Bardsley 2000; Teletchea *et al.* 2005).

Two primary purposes of DNA studies are used in marine fish taxonomy (Telechea, 2009); firstly, to delineate species and populations and secondly, to define higher categories and relationships of taxa. Numerous documentions on DNA molecular studies have facilitated differentiation between populations that were otherwise indistinguishable morphologically as well as species discovery that upon further careful investigation, may actually have demonstrable morphological differences. For example, using cytochrome oxidase I (COI) gene, Ward *et al.* (2008) discovered that 2 out of 15 fish species shared between North Atlantic and Australasia have deep intraspecific divergence (2.75% and 7.44%). Employing the same gene, Zemlak *et al.* (2009) reported that populations of commercial fish with inshore distribution in South Africa and Australia have high levels of genetic divergence (mean 5.10%). He estimated that one third of the 1,000 shared species

between these two regions include cryptic taxa. Thus, the use of molecular taxonomy is a necessary complement to conventional approaches.

2.3.1 Efficacy of mitochondrial markers for species identification

Tanya and Kumar (2010) reported that with DNA markers, it is hypothetically possible to observe and exploit genetic variation in the entire genome of organism. Some of the commonly used techniques and markers used are allozyme analysis, restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), microsatellite typing, single nucleotide polymorphism (SNP), expressed sequence tag (EST) markers and sequence analyses of nuclear and mitochondrial markers.

A comparative study by Telechea (2009) found that the most studied DNA markers for identification are mitochondrial genes. He attributed three major reasons for the biasness in favour of mitochondrial DNA genome (mtDNA) to nuclear DNA. Firstly, due to the multiple occurrence of mtDNA inside a cell, it is more likely to amplify a fragment within this genome rather than within the nuclear genome. Secondly, this small circular genome (16 kb in most vertebrate species) exhibits maternal inheritance in most animal species, is haploid, and does not undergo recombination (characteristics that make its study easier and more straightforward). Lastly, mtDNA generally evolves much faster than nuclear DNA and thus enables even closely related species to be distinguished and identified. He also recommended that a suitable DNA marker for identification at the species level should be

sufficiently variable between species (particularly the closest ones) and display either low or no-intraspecific value.

Earlier studies indicate that the interspecific genetic divergence established through species specific diagnostic molecular markers can provide precise knowledge on phylogenetic relationships and also resolve taxonomic ambiguities (Chauhan and Rajiv, 2010). In fact, mitochondrial DNA sequences top the chart to be useful in the past for species-specific identification of many fishes generally on tuna (Chow and Inoue, 1993), billfish (Finnerty and Block, 1992), snappers (Chow and Inoue, 1993; Hare *et al.*, 1998), Myctophidae (Suneetha and Dahle, 2000) and grey mullets (Murgia *et al.*, 2002). Other examples employing similar DNA genes such as the usage of sequence variation in the mitochondrial control region (D-loop) of the short-jaw cisco, *Coregonus zenithicus* (Reed *et al.*, 1998), identification of *Astyanax altiparanae* (Teleostei, Characidae) in the Iguacu River, Brazil on the basis of mitochondrial DNA and RAPD markers (Prioli *et al.*, 2002) and determination of two species of Siberian sturgeon, *Acipenser baeri* and *A. stellatus* using multiple mitochondrial genes (D-loop, *cytb* and ND5/6 genes) (Doukakis *et al.*, 1999). Given these points, the efficacy of mtDNA markers for species identification is irrefutable and their utility as a critical tool in similar investigations have continued (Tanya and Kumar, 2010; Zhang and Hanner, 2012).

2.3.2 Era of DNA barcoding

In 2003, Hebert *et al.* proposed standardizing the diverse techniques used in species identification with a new approach- the DNA barcoding system. This system is

analogous in practice to a supermarket barcode aimed at identifying products and is based on a single sequence: a 648-bp portion of the mitochondrial gene of cytochrome c oxidase I (COI) for the same identification purpose, but in animals. The major goals of this program are to provide molecular identification of organisms using standardized DNA region (DNA barcode) and to create a dedicated database that would be more taxonomically accurate and would have more rigorous regulations for entry of specimen data compared with the existing databases, such as GenBank. Initial responses to this DNA barcoding proposition however, have ranged from enthusiasm, especially from ecologists (Janzen, 2004), to criticisms, chiefly concerning the identification of closely related species using a single gene (Lipscomb *et al.*, 2003; Mallet and Willmott, 2003; Moritz and Cicero, 2004). Figure 2.1 shows the general process involved in DNA barcoding.

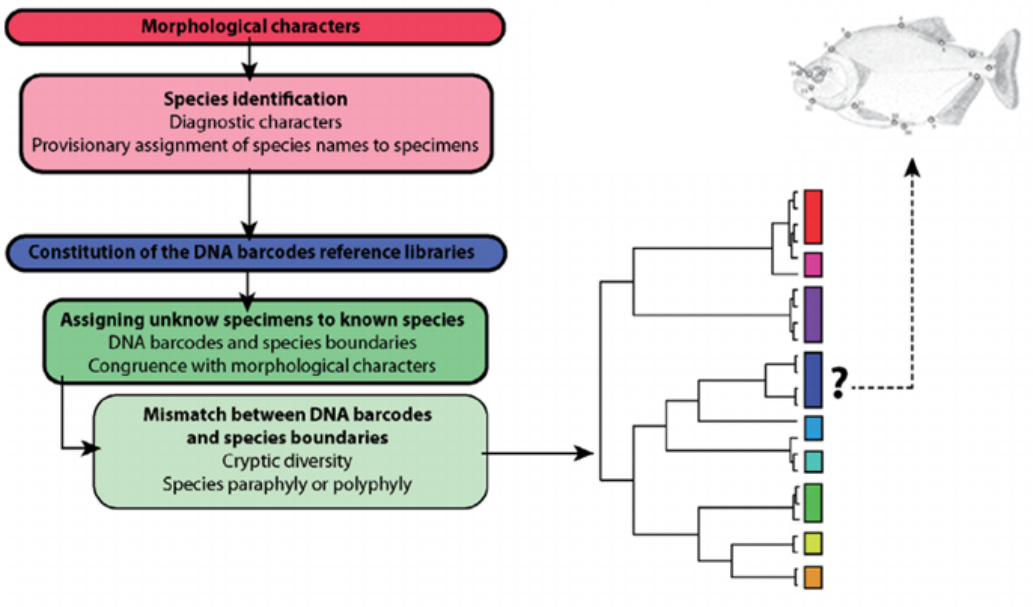

Figure 2.1 Steps of the DNA barcoding process. Edited from http://www.barcodeoflife.org/content/about/what-dna-barcoding.

Being among the most studied marine groups, fishes are currently barcoded within two global campaigns, FISH-BOL (http://www.fishbol.org) and SHARK-BOL (http://www.sharkbol.org) (Ward *et al.*, 2005). One of the early studies on barcoding marine life are discrimination of 207 fish species from Australia (Ward *et al.*, 2005), research on fishes from Pacific Canada (Steinke *et al.*, 2009), North Atlantic (Ward *et al.*, 2008) and determination of fish larvae from the Great Barrier Reef (Pegg *et al.*, 2006). Additionally, DNA barcodes has been shown to be a powerful tool in discriminating various marine fishes with 98% success rate (Rodulovici *et al.*, 2010). However, one major disadvantage of DNA barcodes is incongruence due to potential cryptic species or species complexes (deeply divergent intraspecific clusters) or in cases of hybrids, recent radiation, taxonomic over-splitting or morphological misidentification (shared haplotypes) (Ward *et al.*, 2008).

The campaign described above involves the use of its own data system which is The DNA Barcode of Life Data Systems (BOLD; http://www.boldsystems.org) (Ratnasingham *et al.*, 2007) which provides a unifying procedure for data acquisition, storage and analysis. Data stored in BOLD consist of sampling details with GPS coordinates, images, taxonomic information, DNA barcodes, primer sequences, electropherogram 'trace' files, and even detailed laboratory operations for specimens processed at the Biodiversity Institute of Ontario (BIO, http://www.biodiversity.uoguelph.ca). Above all, this database is freely accessible and all data can be downloaded after publication or analyzed directly in BOLD with distance-based methods. Furthermore, future taxonomic updates of submitted specimens can also be done. These attributes make BOLD a more advantageous tool to use when dealing with DNA barcodes than GenBank which is notoriously known

to host high percentage of erroneous data (Harris, 2003). As has been known, GenBank also includes thousands of fish DNA sequences but nearly all past records lack any explicit connection to vouchers. Due to this, cases of sequence disorder among supposedly conspecific individuals, which may reflect specimen misidentifications, cannot be easily resolved (Ruedas *et al.*, 2000; Pleijel *et al.*, 2008). Likewise, the raw sequence data from which GenBank submissions derive are rarely archived and normally no critical evaluation of the reported nucleotide base calls are done. Owing to such problems, existing GenBank data are of limited utility for molecular diagnostic applications and should be used with caution.

Since its debut, DNA barcoding initiatives have seem to discover unknown biodiversity (Barber and Boyce, 2006; Concepcion *et al.*, 2008; Zemlak *et al.*, 2009; DiBattista *et al.*, 2011; Hubert *et al.*, 2012). A variety of improvements have been implemented to its method with a view to making the system more relevant and user friendly. Rodulovici *et al.*, (2010) pointed out that it will serve multiple applications for marine life such as identification of larvae, invasive species, cryptic species, new species, illegal trade of protected species, stock management, biodiversity assessments, ecosystem monitoring, revisions of certain taxa, inference of phylogenetic relationships, phylogeographic and speciation patterns. Furthermore, its use can be expanded to provide insights into the processes driving speciation in marine systems (Teske *et al.*, 2011).

2.4 Marine diversity and genetic discontinuity within Indo-Pacific

The Indo-Pacific, which encompasses large areas of the tropical Indian and Pacific oceans, houses an incredible array of marine biodiversity (von der Hayden *et al*., 2014). This marine biogeographic province shelters an astounding diversity in terms of species richness especially in the central region bounded by the Philippines, East Sabah, Indonesia, and Papua New Guinea (Briggs, 1999; Mora *et al*., 2012) also known as the Coral Triangle (Figure 2.2).

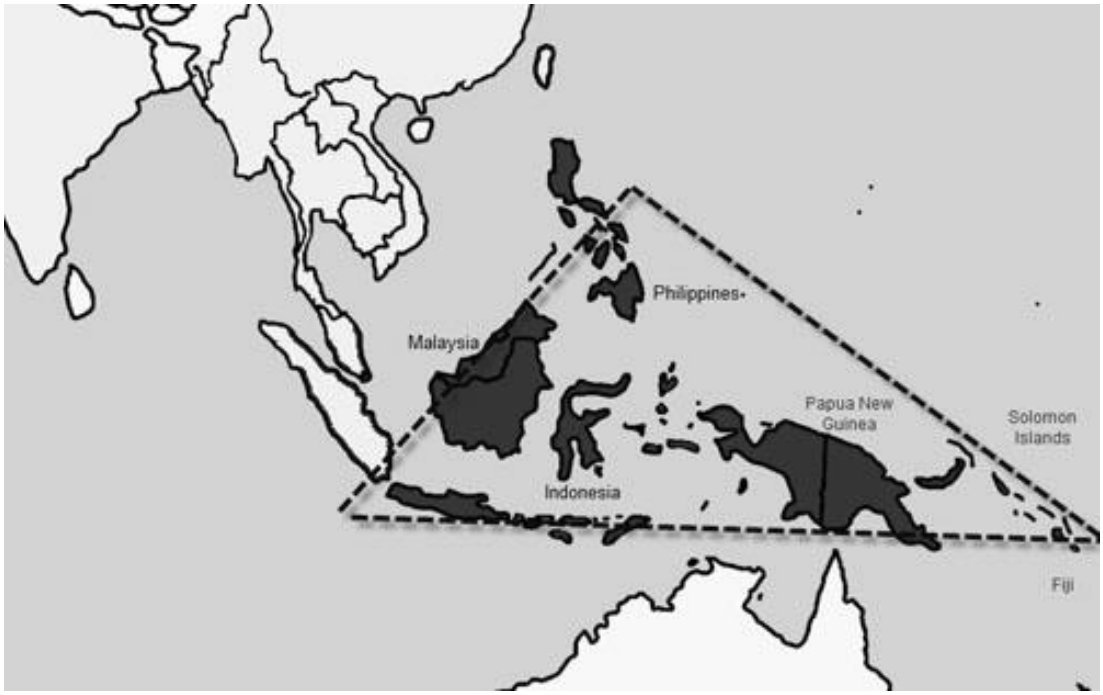

Figure 2.2 Indo-Pacific region. Edited from web source: www.leisurepro.com

Numerous molecular phylogenetic and population genetic studies on different marine organisms have revealed genetic discontinuities between the Indian and Pacific Oceans within this Indo-Pacific zone. It is hypothesized that these observations are attributed to sea-level changes during glaciations in the Pliocene and Pleistocene which formed temporary land barrier in the area (McMillan and Palumbi

1995; Williams and Benzie 1998; Duda and Palumbi 1999; Nelson *et al.,* 2000; Kochzius *et al.,* 2003; Froukh and Kochzius 2008; Timm *et al.,* 2008). Such structuring of populations and even speciation processes have also been documented to occur in the seas within the Indo-Pacific which had undergone historical partitioning (Barber *et al.,* 2002, 2006; Sugama *et al.,* 2002; Lourie *et al.,* 2005; Knittweis, 2008).

William *et al.* (2002) mentioned that there is a continuous route for gene exchange between tropical organisms in the Indian and Pacific Oceans. The main pathway is through the Indonesian through flow, which is responsible for a large transfer of warm surface water from the tropical northwest Pacific along the Makassar Strait, into the Flores and Banda Sea before entering the Timor Sea and the Indian Ocean (Gordon and Fine, 1996). Sea levels have been postulated (Voris, 2000) to be substantially lower during the previous epochs than at present and there is an increasing body of phylogeographical evidence to support this. Chappell and Shackleton and Opdyke (1973) had hypothesised that during the height of the last glaciation event about 18,000 years ago, sea levels are thought to have been about 130 m lower than present day levels. During that period, the Indonesian through flow is thought to have been greatly reduced, and the Torres Strait (which provides a tropical marine connection between western and eastern Australia) was completely closed by a land bridge between New Guinea and Australia, thereby greatly decreasing the opportunity for genetic exchange between the two oceans. Henceforth, multiple glaciations resulted in repeated periods of isolation between tropical marine faunas in the Pacific and Indian Oceans provide an operative mechanism for population differentiation and incipient speciation.

Many compelling evidence regarding the overlaying genetic breaks throughout Indo-Pacific region have been deduced through continuous molecular research. These include high phylogeographic discontinuity reported for barramundi (*Lates calcarifer*) on either side of the Torres Strait (Chenoweth *et al.*, 1998), between false clown anemonefish (*Amphiprion ocellaris*) populations from the western edge of the Sunda Shelf (West coast of Sumatra, Malacca Strait) and those from the rest of the Indo-Malay archipelago, including South China Sea, Sunda Strait, Bali Strait, Sulu Sea and Sulawesi Sea (Nelson *et al.*, 2000), phylogeographic structure of pelagic fish species with migrating adults, like the Indian scad mackerel (*Decapterus russelli*) in the Indo-Malay archipelago (Perrin and Borsa, 2001) and pronounced geographic structure among populations of snapper (*Pristipomoides multidens*) sampled from the seas between northern Australia, Indonesia and West Papua (Ovenden *et al.*, 2004). Molecular evidence as such show that the process of evolutionary radiation may be different between the sea and terrestrial environments (Palumbi, 1992; Knowlton, 2000). Colborn *et al.* (2001) proposed that the distinction of speciation mechanisms in marine system may be due to the size and connectivity of marine habitats and to the high potential for dispersal in a transglobal aquatic medium. Though vicariant separations due to habitat discontinuities (and corresponding allopatric divergences) are a foundation of terrestrial speciation, boundaries and causes of habitat discontinuities for oceanic species is still largely indeterminate. However, with the use of molecular data and rapidly improving genetic technologies, detailed insights into this aspect could be attainable.

2.5 Cryptic speciation in Indo-Pacific

Tropical marine systems, such as the Indo-Pacific have been highlighted as potentially rich in cryptic species given that they are among the most species-rich habitats (Bickford *et al.*, 2007). von der Heyden (2014) defines cryptic speciation as the occurrence of two or more reproductively isolated species, but which are classified as one recognized species. This is due to the absence of phenotypic characters that can distinguish them or probable sharing of a large degree of morphological traits. Several studies have revealed the occurrence of cryptic species in the Indo-Pacific fishes (Colborn *et al.*, 2001; Hubert *et al.*, 2012) and such speciation were conjectured to be attributed by the relatively stable marine environment or to the homeostatic qualities of marine communities on evolutionary time scales (Knowlton, 1993).

The number of cryptic species, taxa that cannot be distinguished morphologically but are genetically distinct, may be vastly underestimated in the marine environment (Knowlton, 2000). Due to high frequency of cryptic species discovery which have been majorly uncovered with DNA sequence data, Bickford *et al*. (2007) suggested that molecular data should be incorporated in the research of alpha taxonomists. Moreover, identification and description of cryptic species have significant associations with conservation and natural resource security and management. Similarly, investigating novel mechanisms of speciation, conservation planning with new data on cryptic species, and projecting taxonomic, regional and global diversity indices are worthwhile avenues for future research.

von der Hayden *et al*. (2014) highlighted the importance of an accurate species inventory that should include validation of undescribed cryptic species in order to understand issues of ecological functions. They proposed that cryptic species may also play a significant role in managing exploited species. Therefore, by revealing cryptic species (either as sibling species or as species flocks), rare insights into the processes driving speciation in marine environments, especially where species live sympatrically without evidence of strong vicariant barriers can be elucidated (von der Heyden *et al.*, 2014; Briggs and Bowen, 2013).

2.6 Phylogenetic study of snappers

2.6.1 Introduction to snappers

The most studied marine fish families in the Indo-Pacific are groupers (Serranidae) and snappers (Lutjanidae) followed by surgeonfishes (Acanthuridae) and rabbitfishes (Siganidae) (Colin, 2012). Snappers are known to be large top-predatory fish that inhabit tropical coastal ecosystems worldwide (Espana, 2003). They are economically important for fisheries as food (Bullis and Jones, 1976; Polovina and Ralston, 1987) and are one of the major species caught in tropical fisheries (Polovina and Ralston, 1987).

Most of the snappers can be recognized by their distinctive head profiles which consist of moderately long snouts, fairly moderate to large mouths, and possess strong jaws with large canine-like sharp teeth without incisiform or molariform teeth (Figure 2.3). Snapper commonly feeds on fish, crustaceans,

molluscs and squid but there are several species that consume plankton. They are predatory fish and is so named from their tendency to snap or bite swiftly at food. Most snappers aggregate in large schools that swim close to the bottom over reef or rocky ground but there are a few species that foray into freshwater to find food.

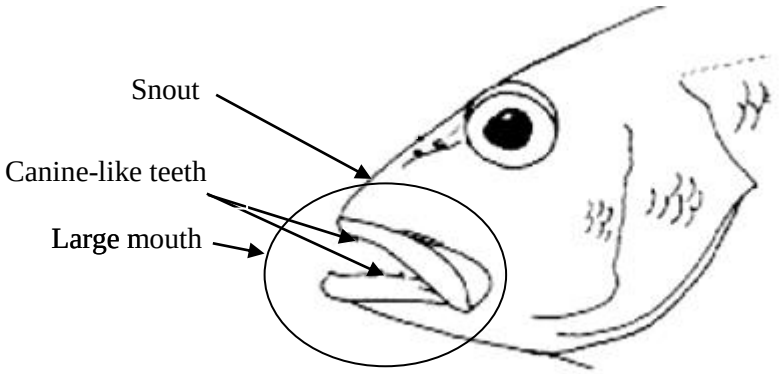

Figure 2.3 Snappers head profile adapted from Allen (1985).

Snappers have a two-stage life cycle which consists of a pelagic (open water) phase and a demersal (bottom oriented) phase. During the pelagic stage, eggs and larvae of snappers are moved and dispersed by currents. At this particular time, the larvae recruit or settle to bottom habitats as the demersal stage begins. Once settled, juveniles and adults become connected with a particular reef or structure in a specific area. Snappers are known to be “sedentary”, because they will remain in such area for most of their adult life. Various adaptations have been used by these fish to overcome the large loss of larvae during the pelagic stage. Some of the slow-growing, long-lived species reproduce several times during their lives while others gather in areas during spawning. Such simultaneous release of larvae gives the fish a better chance

for survival against predators during their early stages. As these fish mature, their large body size can help them escape from predators and become more efficient hunters. Generally, species of snapper can grow up to a meter in length.

There are about 110 species of snappers in the world that make up the *Lutjanidae* family worldwide (retrieved from FishBase). The *Lutjanidae* family can be divided into 4 subfamily which are the *Lutjaninae*, *Etelinae*, *Apsilinae* and *Paradicichthyinae* (Figure 2.4). There are about 77 species under *Lutjaninae* which are the typical snappers, 19 species under *Etelinae* (referred as jobfishes), 12 species under *Apsilinae* (referred as fusiliers) and 2 species under *Paradicichthyinae*. In Malaysia, snappers are represented by 10 genera and 49 species which includes 33 species under *Lutjaninae*, 11 species under *Etelinae*, 3 species under *Apsilinae* and 2 species under *Paradicichthyinae* (Chu *et al.*, 2013).

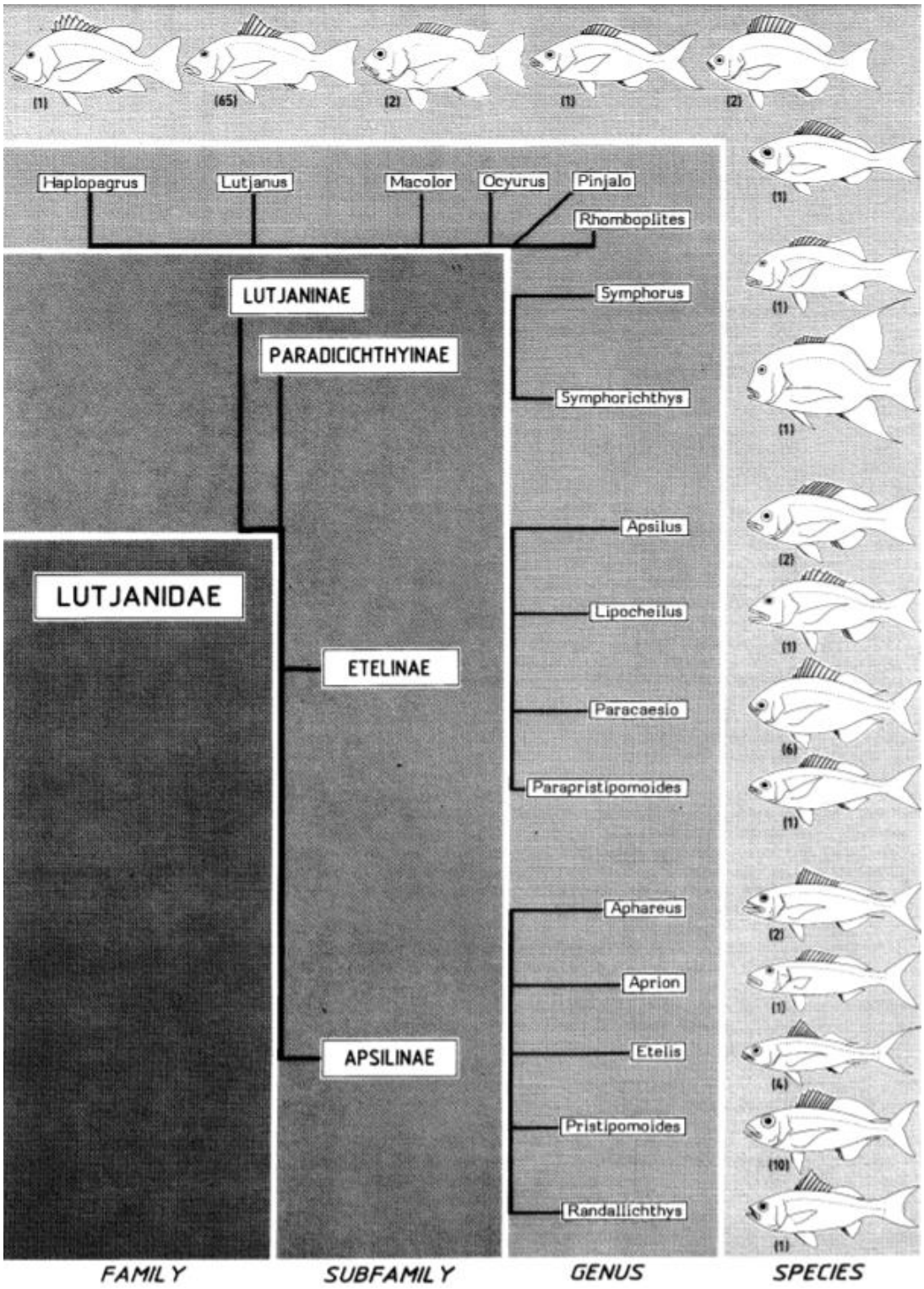

Figure 2.4 Overview of Family Lutjanidae. The number of species in each genus is given in parenthesis. Sources: Allen (1985).

2.6.2 Evolutionary relationships of Lutjanidae

There has been much debate in the literature as to the relationships of 1. species under Lutjanidae and 2. between the families Lutjanidae and Caesionidae. The former issue is due to the morphological similarity (Marko *et al.*, 2004; Zhang *et al.*, 2004) and the ability to hybridize (Domeier and Clarke 1992; Loftus, 1992) among lutjanid taxa causing misidentification while the latter involves argument of whether Caesionidae should be placed under Lutjanidae based on phenotypic sharing of the characteristic adductor mandibulae of both families. These issues have been raised due to uncertainties and ambiguites in the current available taxonomical data. As molecular techniques for phylogenetic analysis have evolved rapidly over the years, employment of them might help in solving both issues.

Recently, phylogenetic relationships of several western Atlantic and Indo-Pacific species of lutjanids have been explored using molecular data.The relationships of 14 western Atlantic species of snappers were investigated using the 12S ribosomal RNA and cytochrome b (*cytb*) genes by Sarver *et al.* (1996). Lee and Tsoi (1988) determined systematic relationships of 10 lutjanid species by isozymic analysis of fixed allelic differences of 18 loci. Zhang *et al.* (2004) utilized amplified fragment length polymorphism AFLP analysis to identify larvae of 11 species of Indo-Pacific lutjanids while Zhou *et al.* (2009) employed partial fragments of the 16S rRNA gene to infer relationships between 6 species of *Lutjanus* found off China. In another study, Zhu *et al.* (2006) used a 316 bp segment of the *cytb* gene to infer the relationships among 10 Indo-Pacific and 9 western Atlantic species of *Lutjanus.* Most recent studies are investigations by Miller and Cribb (2007), Gold *et al*. (2011) and Gold *et*

al. (2015), involving research pertaining to the relationships of Lutjanidae species using mitochondrial genes. Overall, these researches have focused on the systematics of lutjanids but based on a narrow range of sampling area and low number of gene markers. Thus, future studies should incorporate wider geographic ranges, employing an increased number of mitochondrial genes as well the inclusion of nuclear to resolve the relationships of this important family of fishes. Inclusion of these elements will enable researchers to present a more complete picture on the evolution of the Lutjanidae along with tracing the origin and the role vicariance has played in the evolution of this group.

2.6.3 Relationships between Lutjanidae and Caesionidae

The systematics status of Lutjanidae and Caesionidae have been the subject of much debate among marine fish taxonomists and remains contentious. Johnson (1980) proposed a separation between the Lutjanidae and Caesionidae based on adult morphology. According to FishBase, the subfamily Caesioninae belongs to the family Caesionidae, and the family Lutjanidae is divided into four subfamilies (Etelinae, Lutjaninae, Paradicichthyinae, Apsilinae) (Guo *et al*., 2007). However, in contrast to Johnson (1980), Allen (1985), reported that several lutjanid taxa possess intermediate features between Lutjanidae and Caesionidae for example, species under the genus *Macolor* (*Macolor niger* and *M. macularis*). Moreover, retention of many Caesionidae characters are seen in Lutjanidae taxa especially for the external colouration features (Carpenter, 1987). However, Gold *et al.* (2011) argued such trophic characteristics are usually exposed to convergent evolution and it is still debatable whether external colouration is reliable for assessing phylogenetic

relatedness. Leis (1987) had also proposed that the caesionids should be treated as members of the Lutjanidae but with very weak rationale; they suggested that both families should be synonymised only because the larvae of the two families are indistinguishable.

The taxonomic position of the Caesionidae was again called into question by Miller and Crib (2007). However, their hypothesis was constructed by molecular findings using 16S ribosomal RNA and cytochrome b mitochondrial DNA. They reported that caesionids should be nested within the subfamily Lutjaninae, making Caesionidae a synonym of Lutjanidae. However, with only mitochondrial genes used, they proposed that a more comprehensive investigation should be conducted with the addition of nuclear genes to resolve the phylogenetic dispute of caesionids placement with the lutjanids. Henceforth, with the increasing availability of molecular markers, integrating multi-locus datasets can be employed for follow- up research to resolve complex phylogenetic problems that single or few-locus studies have been incapable to answer with certainty (Barley *et al*., 2010).

2.6.4 Indo-Pacific as centropomine origin of Lutjanidae

Biogeography and barriers to dispersal have played several roles in the divergence of lineages of reef-associated fishes (Hanel *et al*., 2002; Wesneat and Alfaro, 2005; Floeter *et al*., 2008; Fessler and Wesneat, 2007). Recent phylogeographic studies have reported that there is an influence of porous hydrological barriers towards marine fish speciation (Drew and Barber, 2012). Historically, three barriers have separated the circumtropical belt into three major marine realms; the Indo-Pacific,

Atlantic and Eastern Pacific (Cowman and Bellwood, 2013) (Figure 2.5). Several studies have postulated that putative vicariant events had generated a basal Indo-Pacific origin of transisthmian marine organisms likely to occur during the late to middle of Miocene period (Tringali *et al.,* 1999; Teske *et al.,* 2007).

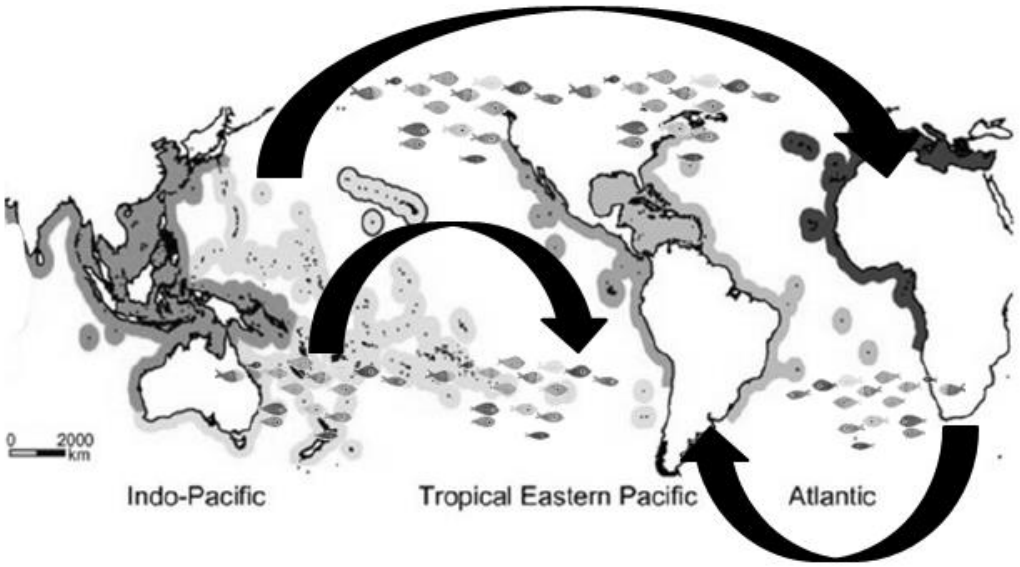

Figure 2.5 Three major marine regions for marine fish dispersal; the Indo-Pacific, Atlantic and Eastern Pacific areas. Edited from Kulbicki *et al*., (2013).

The Miocene era was reported as a time of global warming and sea grass expansion (Brasier, 1975). Such environment would have greatly increased suitable habitat for lutjanids to disperse from the Indo-Pacific to the eastern Pacific and through the gateway into the western Atlantic. Moreover, this era was found to be a fairly stable period tectonically (Duque-Caro, 1990), which could have allowed lutjanids and other structure-associated fish to colonize the various islands and reefs within the three regions. To understand the complex history of barrier formation in the marine tropics, researchers currently outlines family-level partitioning of marine fishes on a large geographical scale through molecular-time calibration approach (Cowman and Bellwood, 2013). Using such approach, they can identify and estimate