UTILISATION OF MOLECULAR MARKERS FOR DETERMINING MANAGEMENT STRATEGIES OF THE BLOOD COCKLE, *Anadara granosa*

by

FARHANA MOHD GHANI

Thesis Submitted in Fulfillment of Requirements for the Degree of Master of Science

School of Biological Sciences
Universiti Sains Malaysia
Pulau Pinang, Malaysia

July 2011

ACKNOWLEDGEMENT

'On no soul doth Allah place a burden greater than it can bear.'
Surah Al-Baqarah: Verse 286.

First and foremost, praise to Allah, the most gracious and the most merciful, who made all things possible in this journey of life.

I owe my deepest gratitude to my supervisor, Assc. Prof. Dr Siti Azizah Mohd Nor for the support, advices and guidance throughout my research and thesis. This thesis would not have been completed without her patience and knowledge. One simply could not wish for a better or friendlier supervisor. I would also give my gratitude to Fishery Research Intitute (FRI) and Universiti Sains Malaysia (RU-PGRS) for funding this project, Ministry of Science, Technology and Innovation (MOSTI) for the scholarship award and the guidance from Assc. Prof Dr Khoo Khay Kuat and Dr Thuy Nguyen.

A special thanks goes to Jam, Leen, Zai and Mira for their vital encouragement, understanding and assistance. Without them, this research would not be completed successfully. Besides them, I would like to thank the rest of my lab members including Pau, Lim, Sue, Katie, Kak Naz, Su Yin, Lutfi, Abdullah, Om (Dr) Icin, Kak Adel and Balkhis for the help and constant reminders. To all my beloved friends and those that have not been mentioned here, I appreciate their support and sharing of joy.

Last, but not least, I thank all my family members especially my mother and father, for their patience and nonstop absolute support in pursuing my interest. 'There is no friendship, no love, like that of the parent for the child' - Henry Ward Beecher.

TABLE OF CONTENTS

CHAPTER 3: MATERIALS AND METHODS

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF GRAPH

LIST OF ABBREVIATIONS

bp	basepair
DNA	Deoxyriboncleic acid
dNTP	Dinucleotide triphosphate
EDTA	Ethylenediamine tetra-acetic acid
EtBr	Ethidium bromide
g	Gram
HCl	Hydrochloric acid
kb	Kilobase
L	Litre
M	Molar
mM	Milimetre
cm	Centimeter
mM	Milimolar
NaCl	Sodium chloride
NaOH	Sodium hydroxide
ng	Nanogram
OD	Optical density
rpm	Rotation per minute
SDS	Sodium dodecyl sulphate
TBE	Tris-borate-EDTA
TNES-Urea	Tris-sodium chloride-EDTA-SDS-Urea
U	Unit
ρM	Picomolar
µM	Micromolar
µg	Microgram
µL	Microlitre
UV	Ultra violet
V	Volt

LIST OF APPENDICES

LIST OF PUBLICATIONS & SEMINARS

1.1 Farhana, M.G., Devakie, M.N., Chee, S.Y., & Siti Azizah, M.N. Population studies of blood cockles, *Anadara granosa* from southern Peninsular Malaysia. (Proceedings of the Asian Pacific Aquaculture Conference, 3rd-6th November 2009, Kuala Lumpur, Malaysia). GENETICS IN AQUACULTURE: ORAL PRESENTATION.

1.2 Farhana, M.G., Devakie, M.N. & Siti Azizah, M.N. Study on the Genetic Structure of *Anadara granosa* Spats in Malaysia Based on Cytochrome Oxidase 1 Gene Sequence. (Proceedings of the 7th IMT-GT UNINET & the 3rd Joint International PSU-UNS Conferences, 7th-8th October, 2010, Hatyai, Songkhla, Thailand). BIODIVERSITY: POSTER PRESENTATION.

1.3 Farhana, M. G. & Siti Azizah, M. N. Population Studies of Blood Cockles (Spats), *Anadara granosa* From Western Peninsular Malaysia. (Proceedings of the 4th Bio Colloquium, 15th December 2010, Penang, Malaysia). BIODIVERSITY: ORAL PRESENTATION.

PENGGUNAAN PENANDA MOLEKUL UNTUK MENENTUKAN STRATEGI PENGURUSAN KERANG DARAH, *Anadara granosa*, DI MALAYSIA

ABSTRAK

Satu kajian genetik populasi kerang darah, *Andara granosa*, di bahagian pantai barat Semenanjung Malaysia telah dijalankan untuk menentukan struktur genetik dan variasi temporal antara dua kelas tahun (2007/2009 dan 2008/2009) spesis ini. Di samping itu penanda genetik mikrosatelit novel telah dibangunkan. Kajian populasi adalah berdasarkan gen DNA Sitokrom Oksidase Subunit I (COI mtDNA) mitokondria yang telah digunakan sebelum ini untuk menerangkan struktur populasi kerang dewasa, *A. granosa*. Tinjauan genetik telah dijalankan ke atas tujuh populasi benih kerang dan enam populasi kerang dewasa dari utara ke selatan pantai barat Semenanjung Malaysia. Populasi telah dikelaskan kepada tiga kumpulan: utara, tengah dan selatan. Penjujukan DNA 149 sampel mendedahkan struktur populasi yang signifikan antara populasi selatan daripada yang lain (Φ_{ST} = 0.217, P < 0.05). Semua populasi juga mempamerkan nilai kepelbagaian haplotip yang tinggi (h = 0.563-0.933) tetapi nilai kepelbagaian nukleotida yang rendah (π = 0.001-0.013), menunjukkan perkembangan populasi yang baru. Analisis temporal yang dijalankan antara kelas tahun mendedahkan ketiadaan stuktur genetik yang signifikan (Φ_{CT} = 0.018, P = 0.709), mengisyaratkan pola taburan yang konsisten. Kajian ini telah menyediakan maklumat lanjut tentang demografi sejarah, corak penyebaran benih dan kedinamikan populasi spesis ini.

Pembangunan penanda genetik mikrosatelit kerang darah telah berjaya mengasingkan 82 klon positif mengandungi ulangan mikrosatelit dengan kebanyakan (91.4%) adalah hanya dinukleotida. Berdasarkan aras polimorfisme dan berat molekul yang tidak bertindih, 8 penanda digunakan seterusnya ke atas populasi terpilih. Aras polimorfik adalah antara H_o = 0.103-0.833 dan H_e = 0.773-0.937. Namun, satu alel nol serta ketidakseimbangan rangkaian yang signifikan telah dikesan pada lima pencetus genetik. Penanda ini berguna untuk pengurusan populasi kerang, pemetaan genom dan pemilihan berasaskan penanda genetik untuk menyumbang kepada masa hadapan perikanan dan pengurusan akuakultur.

UTILISATION OF MOLECULAR MARKERS FOR DETERMINING MANAGEMENT STRATEGIES OF THE BLOOD COCKLE, *Anadara granosa*

ABSTRACT

A population genetic study of the blood cockle, *Anadara granosa*, on the west coast of Peninsular Malaysia was conducted to define the genetic structuring and determine temporal variation between year classes (2007/2008 and 2008/2009) of this species. In addition, novel microsatellite markers were developed. The population investigation was based on the Cytochrome Oxidase subunit I mitochondrial DNA (COI mtDNA) marker that had previously used to infer population structuring of adult *A. granosa*. The genetic survey was done on seven spat populations and six adult populations involving populations from north to south of the west coast Peninsula Malaysia. Populations were defined into three regions: north, central and south. DNA sequencing of 149 samples revealed significant genetic structure of the south populations from the rest (Φ_{ST} = 0.217, $P < 0.05$). All populations exhibited high haplotype diversity (h = 0.563-0.933) but low nucleotide diversity (π = 0.001-0.013) indicating recent population expansion. Temporal analysis between year classes revealed no significant genetic structure between both cohorts (Φ_{CT} = 0.018, P = 0.709) signalling consistent larval dispersal pattern over time. This study has provided further insight on the historical demography, dispersal pattern and population dynamics of this species.

Development of microsatellite markers for the blood cockle successfully isolated 82 positive clones containing microsatellite repeats with the majority of these (91.4%) being dinucleotide repeats. Based on the levels of polymorphism and

non-overlapping molecular weights, 8 markers were further utilised on selected population. Levels of polymorphism were found to be between H_o = 0.103-0.833 and H_e = 0.773-0.937. However, one null alleles and significant linkage disequilibria were detected for five primers. These markers could be useful for blood cockle population management, genome mapping and marker-assisted selection to provide future groundwork for fisheries and aquaculture management

CHAPTER 1

INTRODUCTION

1.1 General Introduction

The blood cockle, *Anadara granosa*, Linnaeus 1758 is by far the most important aquaculture species along western Peninsular Malaysian coastline, harvested for its nutritional value (FAO, 2010). The high commodity value of cockles has made it one of the most cultured species in the aquaculture sector of the South East Asia region. They are widely cultured on mudflats and therefore the growth is affected by environmental factors such as the nature of the substrate, salinity and slope of bed (Broom, 1985). Despite their extensive culture in Malaysia, Indonesia and Thailand, the supplies have not met the market demands.

The culture of *A. granosa* has been practiced in Malaysia since 1948 and has developed into the most important and organized aquaculture industry in the country (Pathansali & Soong, 1958). Cockle aquaculture can be nurtured using simple technology with low labour inputs in the rearing process with no costs of feeds, being filter feeders (Lucas, 2003). Thus, this makes it one of the favourite species cultured. Malaysia's production of blood cockles in 2007 had increased from 45, 674 metric tonnes in 2006 to 49, 620 metric tonnes with increasing production value of RM 62.3 million from RM49.24 million. Nevertheless, the production of cockles in 2007 was low compared to the production in 2005 with 59, 520 metric tonnes but with lower profit value of RM51.18 million (Department of Fisheries, 2010). The higher production profit over lower production indicated high market demand for the species. Hence, to increase supply, a proper management plan is needed. Besides

environmental aspects (nutritional and water quality), genetic aspects are important to increase production. The reproductive potential of any population can be maximised if the genetic pool are properly managed (Tave, 2003). However, there is limited genetic information concerning genetic structuring of this species, to-date.

Managing cockle stocks will be challenging if individuals occur in discrete patches. Population distinction will complicate harvest management in the natural population if the availability of suitable environment (habitat and oceanographic attributes) or life history is unknown thus limiting local recruitment (Elfstrom *et al.*, 2005). Populations that are isolated indicate that they are most likely independent, self-recruiting and locally adapted and therefore should be managed as individual units (Elfstrom *et al.*, 2005; Johnson, 2000). Current knowledge on the larval dispersal patterns of the cockles is also poorly understood in Malaysia. . Although, there is some available data on the genetic structuring of adult population (Chee *et al.*, 2007; Chee *et al*., 2008a; Chee *et al*., 2008b; Chee *et al*., 2011a), the origins of naturally recruited spat, the impact of translocations on natural population structure and genetic variability are unknown.

Studies conducted by Ng (1986) in three states namely Penang and Perak in the north and Selangor in central Peninsular Malaysia suggested that population sub-division does exist as there were different growth and mortality parameters for cockles obtained from different culture plots. Nevertheless it was difficult to interpret the dissimilarities at that stage but population sub-division was expected due to the different ecological features. Determination of population correlation must use an indirect method because movement of larvae in the ocean currents is impossible to be quantified (Elfstrom *et al.*, 2005). To measure population correlation, the amount of

gene flow among populations must be estimated and this can be investigated by utilizing molecular markers. Two approaches are widely utilised to investigate the degree of isolation and population structuring of numerous species (Lundy *et al*., 1999; Nguyen *et al*., 2006a). One is the analysis of mitochondrial DNA (mtDNA) sequencing and secondly, variability of microsatellite marker. MtDNA have several advantages in resolving genetic structure. The maternal mode of inheritance with lack of genetic recombination retains the phylogenetic information of accumulated mutations. Rapid evolution of mtDNA can expose the effect of recent subdivision for population genetic study. Furthermore, the variation in mtDNA that is highly sensitive to isolation of populations could greatly contribute to information of population structure (Johnson, 2000; Peng *et al.*, 2006).

The nuclear microsatellite or simple tandem repeat (STR) markers consist of tandemly repeated mono- to hexanucleotide motifs dispersed throughout the genome. They are usually characterized by high degree of polymorphism compared with those of other molecular markers. Microsatellite polymorphism is marked as allelic length differences due to the different numbers of repeated units present in the alleles and is easily assayed by PCR amplification (Sampaio *et al*., 2005). Since STRs contains high variability, they are very powerful genetic markers with applications that span over a wide area from forensic DNA studies to population genetic and conservation (Jarne & Lagoda, 1996; Zane *et al*, 2002; Diniz *et al*., 2007). However, there are no available microsatellite markers for this species and therefore its development is one of the focus of the project with the aim utilising them in future aquaculture and conservation studies.

The spatial genetic variation of adult cockles along the western Peninsula Malaysia coastline using RAPD-PCR and COI mtDNA sequencing methods have been reported by Chee *et al*. (2007), Chee *et al*. (2008a), Chee *et al*. (2008b), Chee (2009) and Chee *et al*. (2011a). These studies involved only adult cockles with no research on cockle seeds. However, they have provided a foundation for further studies in addressing population subdivision of the Malaysian cockle.

Objectives

With the above consideration, the objectives of this study were:

(i) To further define the population structuring of *A. granosa* in western Peninsular Malaysian coastline based on COI mtDNA sequencing of cockle seeds (spats).

(ii) To determine temporal variation of *A. granosa* sampled in western Peninsular Malaysian coastline based on COI mtDNA sequencing.

(iii) To develop novel microsatellite markers for *A. granosa* for future studies.

CHAPTER 2

LITERATURE REVIEW

2.1 *Anadara granosa* (Linnaeus, 1758)

2.1.1 Taxonomy and Species Identification

Anadara granosa or its synonym, *Tegillarca granosa* (Poultiers, 1993; FAO, 2010) is an important source of protein in the South East Asian countries including Malaysia (Broom, 1985) and has been a source of food since prehistoric times (FAO, 2010). It is often referred to as the bloody cockle because it possesses the red blood pigment, hemoglobin. Locally known as 'kerang', it is widely distributed and is also known as 'granular arc' in Australia, 'hoy kreng' in Thailand and 'si ham' in China (Broom, 1985; Faulkner, 2009). It is harvested on an intensive commercial basis in Malaysia and Thailand, making it the most important anadarinid harvested (Broom, 1985).

Anadara granosa belongs to the family Arcidae which comprises of 142 species (BSF, 2010). Commercially important species in the genus *Anadara* includes *A. granosa* in Malaysia and Thailand, A. *subcrenata* in Japan, *A. broughtoni* in South Korea, *A. nodifera* in Thailand, *A. satowi* in China and South Korea and *A. antiquate* in the Philippines (Broom, 1985). In Malaysia, the centre of production is in the west coast of Peninsular Malaysia because of the presence of extensive tidal mudflats.

The taxonomic classification of *Anadara granosa* is presented below:

Domain: Eukaryota

Kingdom: Animalia

Subkingdom: Bilateria

Phylum: Mollusca

Class: Bivalvia

Order: Pteriomorpha

Superfamily: Arcoidea

Family: Arcidae

Subfamily: Anadarinae

Genus: Anadara

Species: *Anadara granosa* (Linnaeus, 1758)

Morphological characteristics of this species have been described by Quayle & Newkirk (1989), Carpenter (2002) and FAO (2010) (Figures 2.1 & 2.2 and Plate 2.1). The shell of adult *A. granosa* is heavy and sturdy with one valve slightly larger than the other. The anterior end of the shell is short and rounded while its posterior ends is longer and angled. There are sculptures of 15 to 20 radial ribs crossed by fine prominent concentric lines between ribs with wide interstices at each valve. Ribs are solid and distinctly wrinkled, bearing regular rectangular nodules. The ribs never bifurcate and the periostracum is thin and smooth. The ribs and taxodont hinges also contain numerous chevron-shaped teeth on both sides of the umbone. The hinge is straight while umbones are prominent. The umbones are more or less centrally placed. No pallial sinus is found in *A. granosa* and the posterior muscle and scar are larger than the anterior. Having no siphons, the inhalant flow enters the anterior ventral shell area and the exhalant flow goes out just below the posterior adductor muscle. The foot is well developed thus enabling the cockle to travel easily through its habitat. *Anadara granosa* shell colour is white with yellowish brown periostracum.

2.1.2 Distribution, Habitat and Biology

Anadara granosa is widely distributed in the Indo-West Pacific, from East Africa to Polynesia; Japan in the north and northern, and eastern Australia in the south (FAO, 2010). The centre of concentration of *A. granosa* in South East Asia is in the Malacca Straits between west Malaysia and Indonesia.

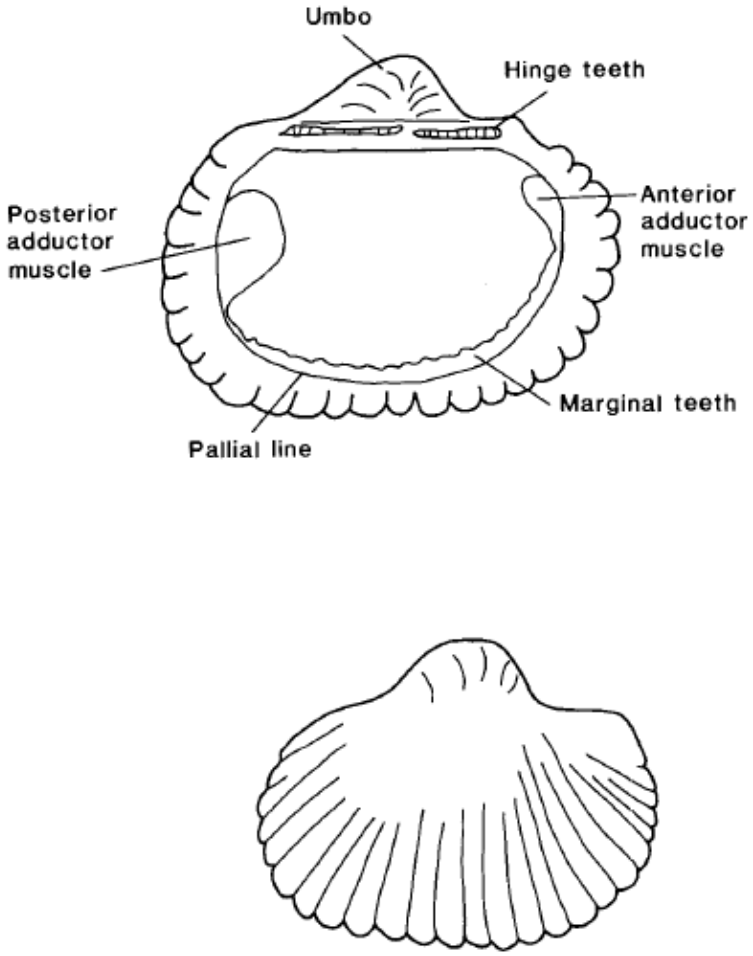

Figure 2.1: Features of *A. granosa* shell (Quayle & Newkirk, 1989)

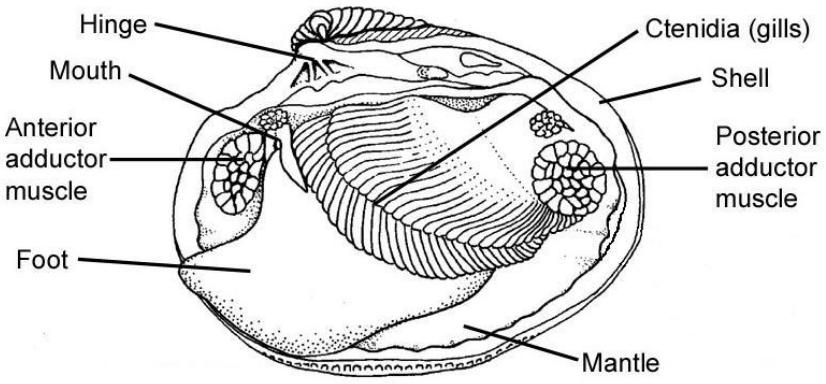

Figure 2.2: General anatomy of bivalve (Hosie, 2010)

Plate 2.1: Adult and spat specimens of *Anadara granosa* (Linnaeus, 1758)

This species occur in estuarine mud flats bordered by mangrove forests, at lower tidal levels and into the subtidal where the habitat is very soft, almost liquid mud, up to 75 cm deep, usually rich in organic material. The typical habitat in West Malaysia has salinity between 18 and 30 ppt with temperature in the range of 25 to 30° C. Waters overlying the cockle beds are nearly always turbid, due to the silty bottom and phytoplankton production (Quayle & Newkirk, 1989).

The breeding cycle of blood cockle exhibits a definite seasonality although some spawning occurs throughout the year. Spawning cues depend on the seasonal salinity fluctuations and period of rainfall. *Anadara granosa* reach sexual maturity at about 25mm in shell length (Broom, 1985). According to the Fisheries (Cockles Conservation and Culture) Regulations 2002, cockle spat are defined as cockle that is four millimeters or less measured in a straight line across the widest part of the shell; cockle seeds are defined as individuals that are more than four millimeters but less than twenty five millimeters while adult cockles are twenty-five millimeters or more. The growth rates of cockle are determined by three most important environmental factors which are the nature of substrate, salinity and slope of bed (Broom, 1985).

2.1.3 Status of *A. granosa* Aquaculture in South East Asia

Gathering cockles, *A. granosa* from either natural or cultured beds is an important activity in the Chinese fishing communities in north and central Peninsular Malaysia. The greatest development of this industry is in Perak, where about 1,200 ha of the foreshore are under cockle culture. Harvesting begins when the cockles have attained a marketable size of 24-30 mm. The gear used is a long-handled close-set wire scoop, usually operated by one person, who stands in a boat, extends the scoop as far as his reach allows, and draws it through the mud with a gentle, rocking motion, trapping the cockles, which are then deposited in the boat (Pathansali & Soong, 1958).

Cockles are also extremely popular in Thailand. Cockle consumption exceeds the local production every year. This is largely overcome by importing both cockles of commercial size (adults) as well as seeds from Malaysia. Malaysian cockles are usually cultivated in mud in the intertidal zone with salinity of around 10-32 ppt. In 1996 and 1997 the production was around 20 to 21 thousand tonnes (Tookwinas & Kittiwanich, 1999). In 2002, the production of blood cockles was ranked 5th in Asia and the Pacific with 78, 712 tonnes and was placed in the top 10 species of mollusks produced (Sugiyama *et al*., 2004). In 2004, it contributed to a total export of USD 525, 546, 000 worth of fishery products (Pawiro, 2004). The global production for *A. granosa* also increased from 10, 000 tonnes in 1958 to 72, 000 tonnes in 2008 but with decreased production in some period of time (Figure 2.3).

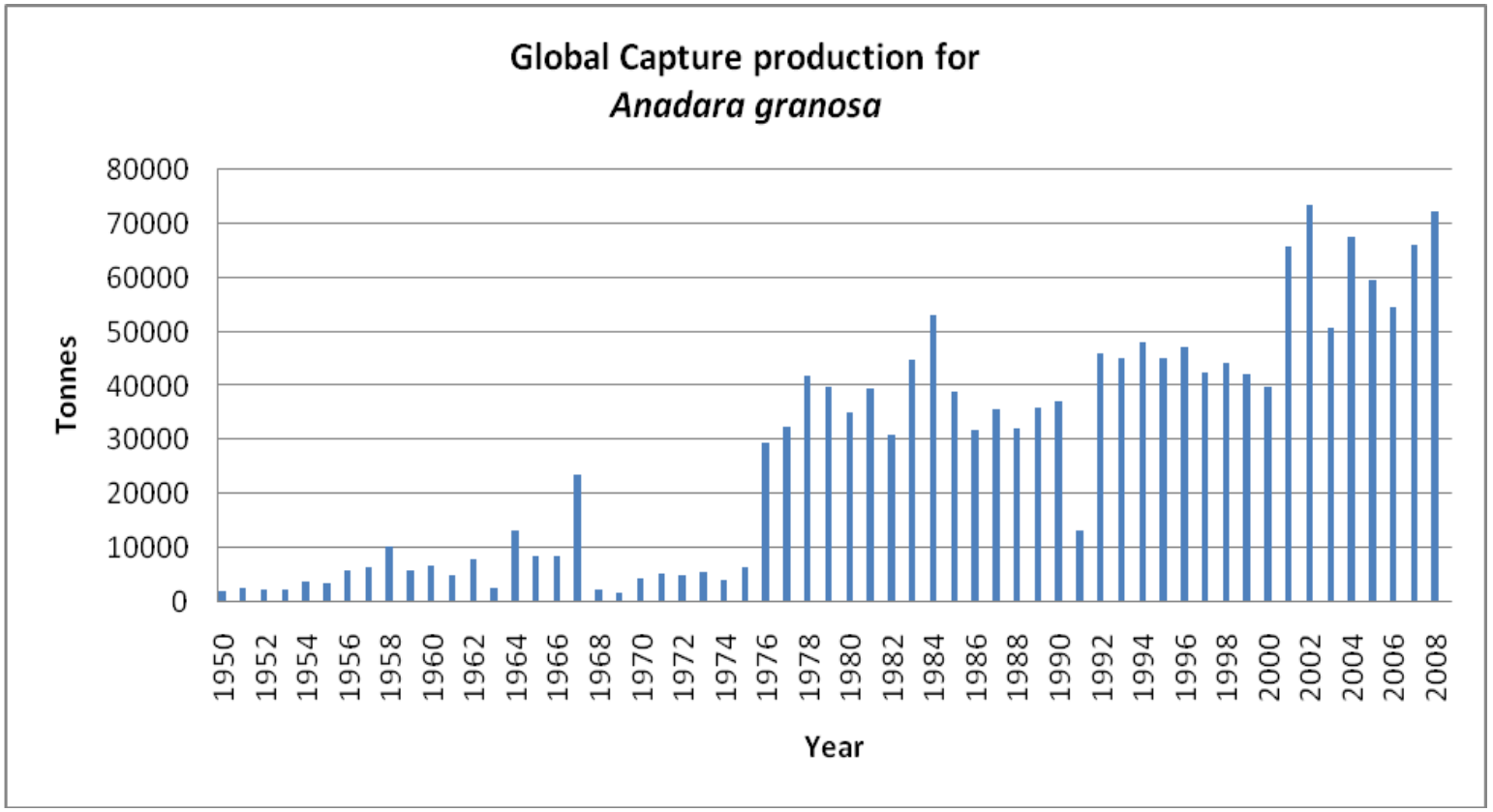

Figure 2.3: Global capture production for *A. granosa*. (Adapted from FAO, 2010).

2.2. Mitochondrial DNA (mtDNA) in Population Genetics Study

2.2.1 MtDNA In General

Mitochondria play an essential role in metabolism, apoptosis, illness, and aging (Boore, 1999; Cao *et al*., 2006). These organelles are the sites for the production of ATP by oxidative phosphorylation and possess their own double-stranded circular mitochondrial DNA (mtDNA) (Kucuktas & Liu, 2007). The size of mtDNA genome varies among different organisms, but in animal, this extrachromosomal genome is small and typically 16 000 – 20 000 base pairs (bp) (Boore, 1999; Kucuktas & Liu, 2007). Studies of the whole animal mtDNA reveal that the majority of the invertebrate mitochondrial genome is made up of nearly the same number of genes as in vertebrates (Pereira, 2000). Animal mitochondrion typically contains 37 genes coding for; 13 proteins, 2 ribosomal RNAs and 22 transfer RNAs (Boore, 1999) and a noncoding control region also known as the D-loop (Pereira, 2000) (Figure 2.4).

However, the gene arrangement in mollusks in general and bivalves in particular are found to exhibit an atypical mitochondrial gene order when compared with other metazoans. MtDNA gene arrangement is highly variable among bivalves (Serb & Lydeard, 2003). Based on a study on *Mytilus edulis* (blue mussel) by Boore *et al*. (2004), it contains 37 genes that are typically found in mtDNA but with an extra gene, *trnM* in place of *atp8.* In another study by Wu *et al*. (2009), based on the analysis of three scallop species (Bivalvia: Pectinidae), high level of genomic variation and a diversity of tRNA gene sets lacking *atp8* gene was observed as in most bivalves (Figure 2.5).

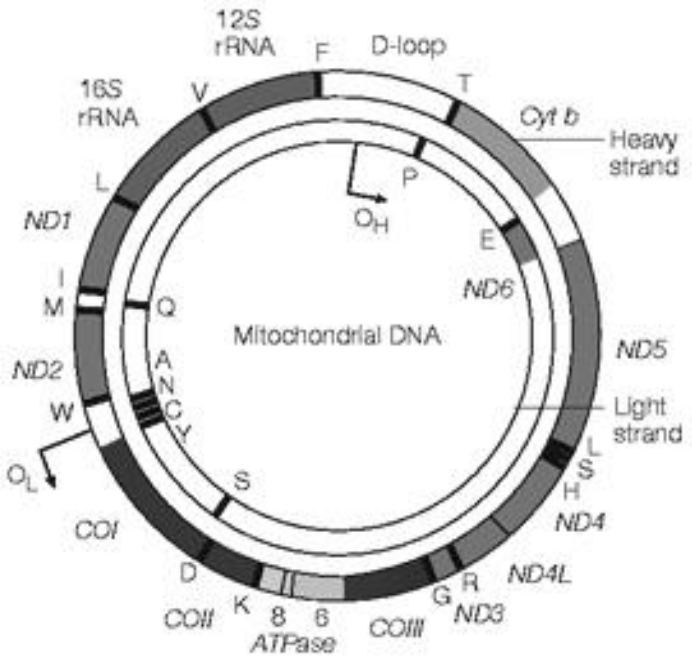

Figure 2.4: Diagrammatic representation of mitochondrial DNA (Adapted from Taylor and Turnbull, 2005).

The outer circle of mitochondrial DNA represents the heavy strand and the inner circle,the light strand. The genes are as follows (anti clockwise-top): small ribosomal RNA (12S rRNA), large ribosomal RNA (16S rRNA), NADH subunit 1 (ND1), NADH subunit 2 (ND2), Cytochrome oxidase subunit 1 (COI), Cytochrome oxidase subunit 2 (COII), ATPase subunit 8 (ATPase 8), ATPase subunit 6 (ATPase 6), Cytochrome oxidase subunit 3 (COIII), NADH subunit 3 (ND3), NADH subunit 4L (ND4L), NADH subunit 4 (ND4), NADH subunit 5 (ND5), NADH subunit 6 (ND6), Cytochrome b (Cyt b). The displacement loop (D-loop), or non-coding control region, contains sequences that are vital for the initiation of both mtDNA replication and transcription, including the origin of heavy-strand replication (shown as O_H). The origin of light-strand replication is shown as OL. According to the International Union of Biochemistry and Molecular Biology (IUBMB), the nomenclature of amino acids are as follows: Phenylalanine (F); Valine (V); Leucine (L); isoleucine (I); Methionine (M); Tryptophan (W); Aspartic acid (D); Lysine (K); Glycine (G); Arginine (R); Histidine (H); Serine (S); Threonine (T); Glutamine (Q); Alanine (A); Asparagine (N); Cysteine (C); Tyrosine (Y); Proline (P) and Glutamic acid (E).

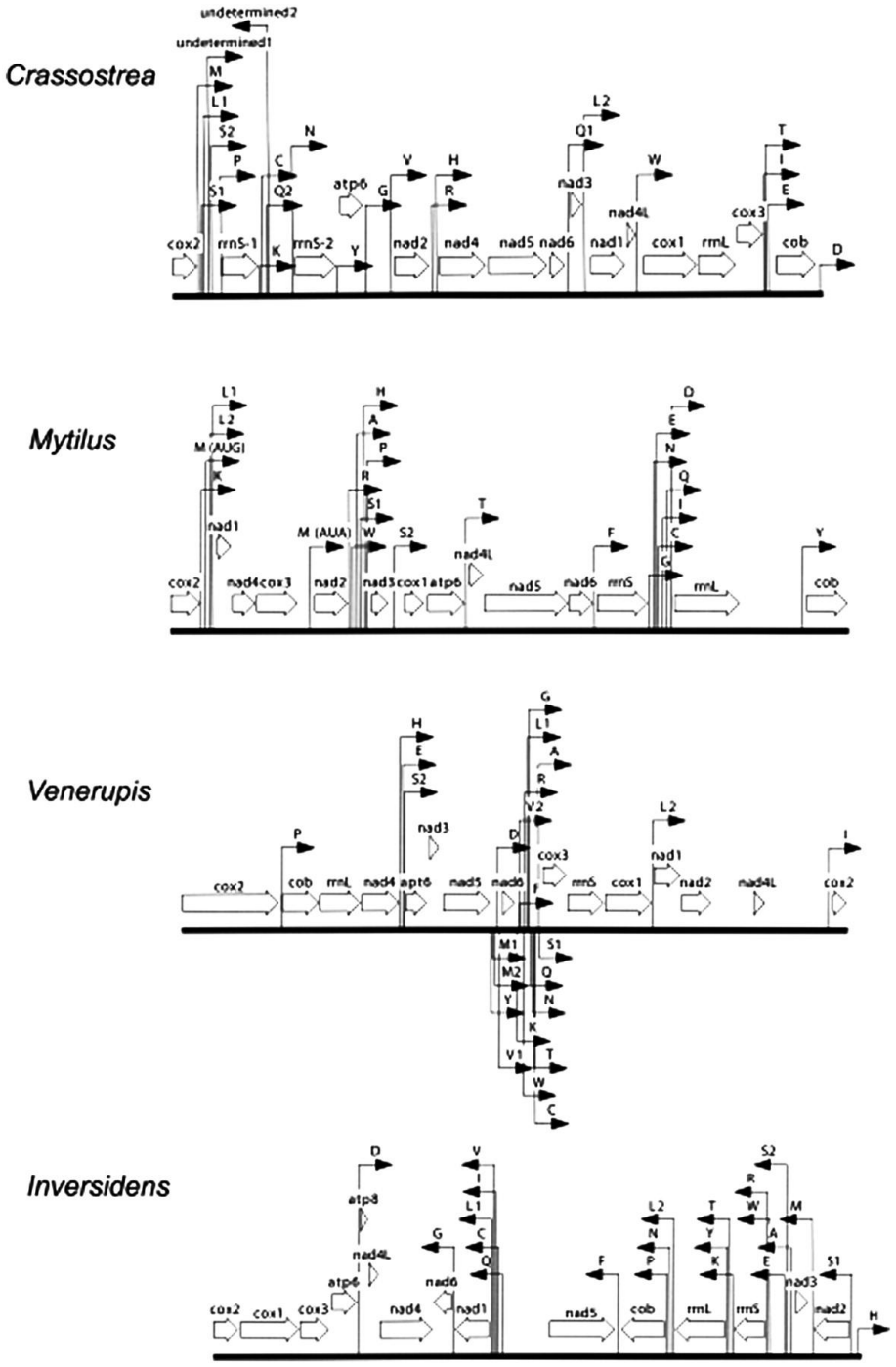

Figure 2.5: Mitochondrial gene order of four bivalve representatives (*Crassostrea*, *Mytilus*, *Venerupis* and *Inversidens*). The direction of transcription is depicted by arrows (Serb & Lydeard, 2003).

The mtDNA pattern of inheritance is distinctively different from those of nuclear genomes (Xu, 2005). MtDNA is haploid and generally maternally inherited (Avise, 1994) but some paternal transmission have been reported during fertilization (Elson & Lightowlers, 2006). Nevertheless, transmission of paternal mtDNA occurs in animals at a very low rate amounting to 10^{-4} to 10^{-3} of an individual's mtDNA (Kondo *et al*., 1990).

An exception to the typical maternal inheritance of animal mtDNA is found in three bivalve lineages (i.e., the orders Mytiloida, Unionoida and Veneroida), which possess an unusual system termed doubly uniparental inheritance of mtDNA (DUI) (Doucet-Beaupré *et al*., 2010). This inheritance pattern is governed by two types of mtDNA, the F type and the M type. The F type mtDNA acts like the typical animal mtDNA, while the M type is transmitted through the sperm and establishes itself only in the male gonads. Thus, these two mtDNA types have separate transmission routes; one through the female lineage, and the other through the male lineage (Xu, 2005). Since the male-dependent DUI of mtDNA are related to the male gonads, association of DUI in *A. granosa* mtDNA genome analysis could be prevented by using only mantle tissue for DNA extraction.

2.2.2 Application of mtDNA Markers in Population Genetics Studies

During the last few decades, mtDNA analyses have become a powerful tool for evolutionary and population studies in animals. They have been used to provide an understanding of population structure and gene flow, hybridization, biogeography and phylogenetic relationships (Moritz *et al*., 1987; William *et al*., 2004). This non-Mendelian inherited marker is now the most widely used marker in aquaculture and

fishery (Liu, 2007). For example they are used for investigation of stock structure (Liu & Cordes, 2004), conservation of population (Ferguson, *et al.*, 2006) and management in aquaculture and fisheries (Çiftci & Okumus, 2002). MtDNA has also been an efficient marker for addressing questions relating to evolutionary relationships and determining population genetic structure among and within bivalves for example in the scallop, mussel and giant clam (Mahidol *et al*., 2006; Nuryanto *et al*., 2007; Zanatta and Murphy, 2008; Divya *et al*., 2009). It has also been utilized to investigate indigenous capture and aquaculture species in Malaysia including *Lates calcarifer* (Asian sea bass) (Norfatimah *et al.*, 2009), *Nemipterus hexodon* (ornate threadfin bream) (Lim, 2009), *Rastrelliger kanagurta* (Indian mackerel) (Siti Azizah *et al.*, 2008), *Channa striata* (sneakhead murrel) (Jamsari *et al.*, 2011a) and *Clarias macrocephalus* (broadhead catfish) (Nazia *et al.*, 2010).

MtDNA gene sequences typically evolve 10 to 30 times faster than nuclear genome from generation to generation (Lane, 2009). The fast evolution of mtDNA is mainly because mitochondria DNA does not have repair enzymes to edit errors occurring during the replication process (Clayton, 1982). Thus, mtDNA has a high level of transitions and transversion with high rate of small length mutation (Cann & Wilson, 1983) leading to high amount of variation in mtDNA sequences (Liu, 2007). As a result, they can reveal the effects of more recent historical subdivision (Johnson, 2000).

Since mtDNA is haploid and follows maternal inheritance, the effective population size (N_e) for mtDNA analyses is only one quarter that for nuclear genes (Johnson, 2000; Mahidol, 2006; Kartavtsev & Lee, 2006). The variation for mitochondrial genes within populations is expected to be lower, and the divergence

between populations are higher when compared to nuclear gene for equivalent mutation rates and selection pressure (Latorre *et al*., 1992). Hence, the variations in mtDNA are more sensitive to the effects of isolation (Johnson, 2000) and increased sensitivity to genetic drift and bottleneck effect (Bernatchez *et al*., 1989; Mahidol, 2006; Cao *et al*., 2009).

The cytochrome oxidase 1 mitochondrial DNA (COI mtDNA) gene has been widely used as a marker in molecular systematics (Medina & Walsh, 2000; Stoeckle, 2003; Tang *et al*., 2003; Gompert *et al*., 2006). In fact, a fragment of the 5' end of the COI gene is the primary marker for the 'Barcoding of Life' project, specifically utilizing a 640 base pair nucleotide sequence as the unique identification code for all species to identify specimens and the discovery of new species (Hebert *et al*., 2003a; Moritz & Cicero, 2004). Even though the sequence information gathered for DNA barcoding is not sufficient to thoroughly address population-level matter, COI mtDNA can provide an early insight into the patterning of genomic diversity within a species. This particular sequence can provide a first signal of the degree and nature of population divergences and will facilitate comparative studies of population diversity in many species (Hajibabaei *et al*., 2007). The phylogenetic relationships of *A. granosa* with other members of genus *Anadara* has been documented by Chee *et al*. (2011b) using the barcoding gene.

2.3 DNA Microsatellites in Population Genetics Analysis

2.3.1 Microsatellites in General

Microsatellites or also known as simple sequence repeats (SSRs), short tandem repeats (STRs), sequence tagged microsatellite site (STMSs) and variable

number of tandem repeats (VNTRs) (Mace & Godwin, 2002). They are tandemly repeated motifs of one to six bases (Zane *et al*., 2002) that recently have been the spotlight in molecular works and proving to be a most powerful marker (Nunome *et al*., 2006). Microsatellite applications span over different areas such as genome mapping, biomedical diagnosis of diseases, parentage analysis, biological relatedness of organisms or populations, population genetics, conservation biology, aquaculture, molecular anthropology, studies of human evolutionary history and forensic studies (Jarne & Lagoda, 1996; Zane *et al*., 2002; Nunome *et al*., 2006; Dong *et al*., 2006; Jerry *et al*., 2006; Nguyen *et al*., 2006a; Diniz *et al*., 2007; Mirabello *et al.*, 2008; Li *et al.*, 2009; McInerney *et al.*, 2009; Bhargava & Fuentes, 2010).

In the late 80's, the world of genome mappers and behavioral ecologists was enlightened by the discovery of polymerase chain reaction (PCR) -based microsatellite genotyping. Microsatellites have since dramatically influenced both genome mapping and behavioral ecology in the subsequent years. The effect of microsatellite analysis is intensely demonstrated by the growing number of genes that have been mapped, cloned and characterized by genetic mapping (Schlötterer, 1998). Microsatellites also have a significant impact on population genetic studies in recent years and data on their molecular dynamics are on the increase (Jarne & Lagoda, 1996; Nguyen *et al.*, 2007; McInerney, 2009). However, Even though microsatellites were first identified in 1981, the mechanism of microsatellite evolution is still unclear (Schlötterer, 1998; Noor *et al*., 2001; Zane *et al*., 2002).

Microsatellites can be found in all prokaryotic and eukaryotic genomes, distributed over the euchromatic part of the genome (Schlötterer, 1998). They are abundant and have a uniform distribution throughout the euchromatic region of the

genome (Mace & Godwin, 2002), highly polymorphic and evolve faster than mitochondrial or nuclear gene (Mirabello *et al*., 2008). In general, microsatellites have a relatively high mutation rate (10^{-2}–10^{-6}) as compared to point mutations in coding genes (Li *et al*., 2007). The majority of microsatellites, are believed to evolve neutrally in higher organisms, where there is no selection pressure on the number of repeats (Kashi & Soller, 1999; Ellegren, 2004).

Microsatellites are often found in noncoding regions since the nucleotide substitution rate is higher than in coding regions (Zane *et al*., 2002). Nevertheless, some microsatellites exist in the promoter regions and may be sites for protein binding. About 10-15% of microsatellites reside within coding regions (Serapion *et al*., 2004). The presence of microsatellites in coding region leads to the emergence of repetitive patterns in protein sequence (Katti *et al*., 2001). The number of repeats in these microsatellites has an effect on transcription and the degree of protein binding (Kashi & Soller, 1999). Several human neurodegenerative diseases such as mental retardation have been found to be connected with mutations occurring at microsatellite loci within or near protein coding genes (Ashley & Warren, 1995).

Microsatellites are classified into mono-, di-, tri- and tetranucleotides microsatellites and so on based on the repeat composition. They can be further differentiated by the specific composition of their core sequence which is simple microsatellites (containing only one type of repeats) and composite microsatellites (containing more than one type of repeats) (Liu, 2007). Composite microsatellites can be categorized as perfect repeats without any interruption, while imperfect repeats with interruption by non-repeat bases and compound repeats with two or more repeat run present adjacent to each other (Weber, 1990). The majority of

microsatellites found in many species are dinucleotides followed by mono- and tetranucleotide repeats while trinucleotide repeats are least dominant (Schug *et al*., 1998; Ellegren, 2004). The most common dinucleotide repeat type is $(CA)_n$, followed by $(AT)_n$, $(CT)_n$ and the least, $(GT)_n$ (Toth *et al*., 2000; Ellegren, 2004).

The abundance of microsatellite varies among organisms and is mainly due to the different efficient mismatch repair system which are found to be 10^{-5}-10^{3}- fold less stable (Chambers & MacAvoy, 2000). Newly arisen mutations are often corrected by cellular repair processes (Eisen, 1999) and mutations that are caused by strand slippage are repaired by a three-enzyme system comprising the enzymes mutL, mutS and mutH (Sreenu *et al.,* 2006). Springer *et al.* (2004) affirmed that some genomes lack these enzymes thus facilitate genome evolution of organisms. Furthermore, microsatellite repeats containing longer units evolve faster than those containing shorter units (Weber & Wong, 1993) which are attributed to the relatively inefficient repair of larger mismatched segments (Chambers & MacAvoy, 2000).

The variation in the number of repeats among different genotypes provides the basis for polymorphism (Schlötterer, 1998; Liu, 2007). Their high polymorphism is a result of their hypermutability and result in the accumulation of various forms in the population of a species (Liu, 2007). Microsatellites containing a large number of repeats tend to be more polymorphic (Calabrese & Sainudiin, 2005). Microsatellites are inherited as codominant markers (Liu, 2007) and are stable over a couple of generations (Schlötterer, 1998). Polymorphism of microsatellite loci can be high between species, and between individuals and populations within a species (Maguire *et al*., 2000).

2.3.2 Microsatellite Mutational Mechanism

Although microsatellites evolution is still unclear, the predominant mutational mechanism hypothesized for microsatellite variability is through a process known as DNA slippage (Levinson & Gutman, 1987; Schlötterer, 1998; Ellegren, 2004; Liu, 2007; Anmarkrud *et al*., 2008). Other microsatellite mutational mechanisms proposed include unequal crossing over, nucleotide substitutions or duplication events (Bull *et al*., 1999). DNA slippage is a transient dissociation of the replicating DNA strands followed by misaligned reassociation. It involves a gain or reduction of one or more repeat units (Ellegren, 2004).

During DNA replication, the two DNA strands dissociate momentarily. In non-repetitive DNA, the DNA strands reassociate the same way they were before the slippage event, with matching base pairs on the opposing strands. However during the formation of repetitive DNA, since there are so many possible matching base pair alignments, sometimes the strands realign differently, forming an unmatched loop on one of the strands. When the two DNA strands completely disassociate and begin replication, the DNA strand which had the loop will contain a longer microsatellite than the opposing strand. In slippage event, the microsatellite on the template strand will always have the same length before and after the event. The microsatellite on the replicating strand will be shorter if the loop is on the template strand; and if the loop is on the replicating strand, then the microsatellite on its side will be longer (Calabrese & Sainudiin, 2005) (Figure 2.6). Only the small fraction that was not repaired ends up as microsatellite mutation events because most of these primary mutations are corrected by the mismatched repair system (Schlötterer, 1998; Ellegren, 2004).

Replication slippage also occurs during PCR amplification of microsatellite sequences *in vitro* (Ellegren, 2004; Webster & Hagberg, 2007; Kelkar *et al*., 2010). The feature of such amplifications is the presence of 'stutter bands' which are minor products that differ in size from the main product by multiples of the length of the repeat unit (Hauge & Litt, 1993; Murray *et al*., 1993). Such slippage occurs because the *Taq* polymerase slippage rate increases with the number of repeat units and is inversely correlated with repeat unit length (Schlötterer, 1998; Shinde *et al*., 2003; Ellegren, 2004). In addition, microsatellite arrays that contain dinucleotide repeats have a higher probability of producing 'stutter bands' than arrays that contain tri-and tetranucleotide repeat (Chambers & MacAvoy, 2000).

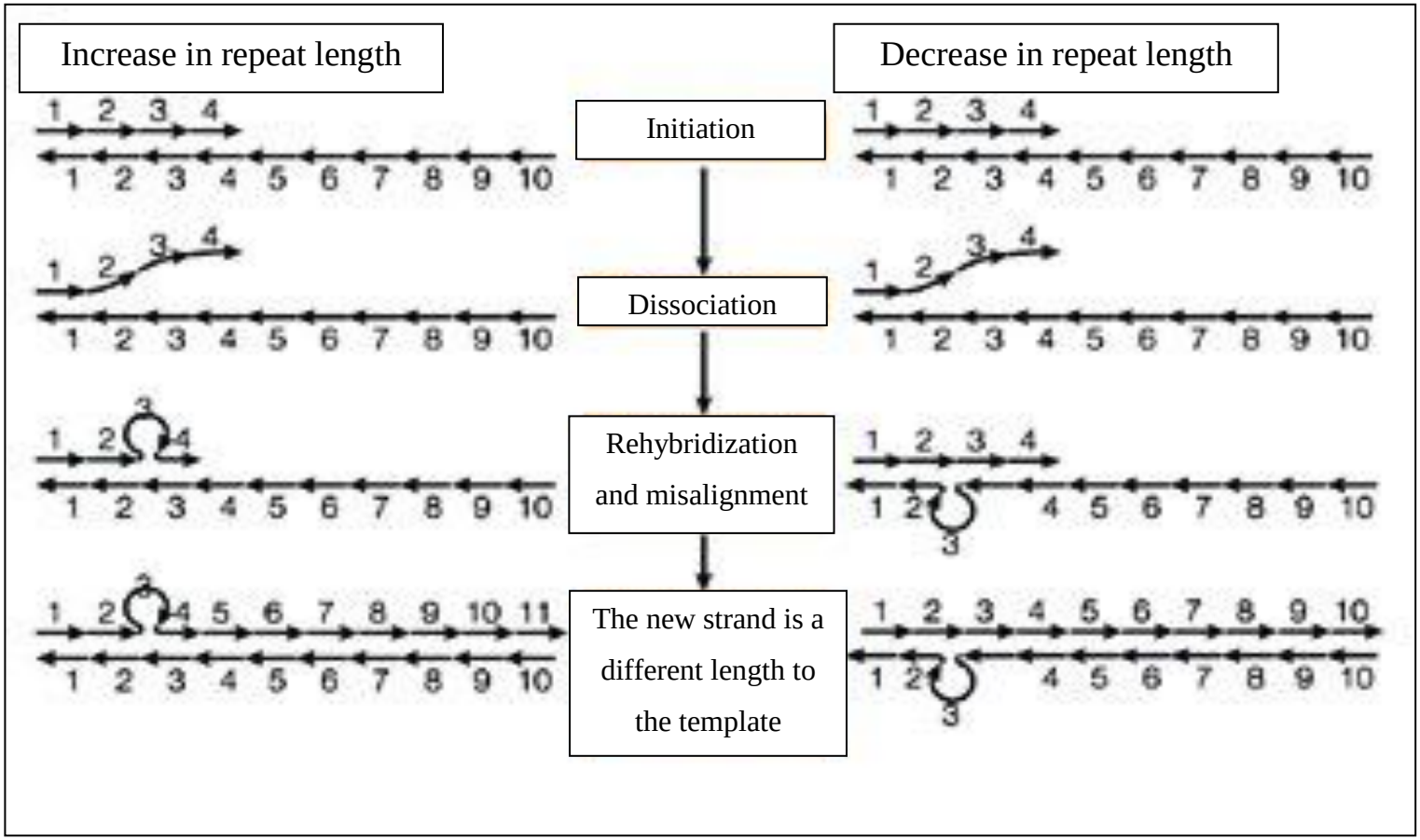

Figure 2.6: Replication slippage. (Adapted from Ellegren, 2004).

2.3.3 Development of Microsatellite Markers & Its Application in Fisheries and Aquaculture

PCR and gel electrophoresis analysis of microsatellite markers only require small amount of DNA (Ritschel *et al*., 2004). Due to the simplicity of detection through PCR and their ease of transferability and reproducibility they have emerged to become the marker of choice in contemporary biology (Mace & Godwin, 2002). However, one major drawback in the utilization of microsatellite markers is that they need to be specifically isolated *de novo* from most species to be analysed. This is due to the fact that microsatellites are usually found in non coding regions (Zane *et al*., 2002) which are highly variable among species. However, cross amplification may be possible in closely related species (Brown *et al*., 2005; Nguyen *et al*., 2007; Díaz-Viloria *et al*., 2008; McInerney *et al*., 2009). Microsatellites can be isolated directly from total genomic DNA libraries or from libraries enriched for specific microsatellites (Maguire *et al*., 2000). Additionally, microsatellite development can be based on DNA sequence information deposited in databases (Ritschel *et al*., 2004).

According to O'Brien (1991), molecular markers can be divided into Type I marker (markers associated with genes of known function, coding region) and Type II marker (markers associated with anonymous genomic sequence, non coding region). Type I markers are harder to develop and show lower levels of polymorphism because of functional constraint (Liu, 2007). Microsatellite usually represent Type II markers as they are derived from enriched libraries of non coding region (Liu, 2007) where the rate of nucleotide substitution is high. Development of microsatellite is both time consuming and costly. However, once developed, the