CHAPTER 1

INTRODUCTION

The genus of *Macrobrachium* is one of the most diverse genera of freshwater Crustacea and contains about 240 described species (Wowor *et al.*, 2009). Among these, Divu *et al*. (2008) reported that the *Macrobrachium rosenbergii* which is also known as the giant river prawn is one of the most widely cultured freshwater prawn species all over the world with Asia accounting more than 98% of the global production.

This giant freshwater prawn can be found in coastal river systems from the western part of Pakistan to the eastern part of Vietnam, across South East Asia and to the southern part of Papua New Guninea and northern Australia (de Bruyn *et al*. 2004). *Macrobrachium rosenbergii* inhabits a wide variety of areas such as rivers, streams, lakes, ponds, canals and swamps as well as estuarine areas (Wowor *et al*., 2009).

Macrobrachium rosenbergii is one of aquaculture species that is highly-demanded with a high economic value due to its' delicate flavour. However, the wild stocks of this freshwater prawn are not enough to supply consumer demand. Therefore, it is believed that the prawn culture industry can increase prawn production so that it can fulfill the consumer order.

In hatcheries, domestication selection is an unavoidable process that normally occurs due to the natural selection on fitness and reproduction traits under a human-

controlled environment. Domestication is essential for good aquaculture stocks because it has been proven to improve the reproductive success of the Nile tilapia and increase the growth rate of rainbow trout (Osure and Phelps, 2006). However, Osure and Phelps (2006) also stated that positive selection for one trait may have a negative effect on another trait. This is because domestication results must consider genotype–environmental interactions where one strain may perform best in one environment while another strain being better in another setting.

Chareontawee *et al*. (2007) stated that many prawn cultures have low productivity most probably due to genetic deterioration in the hatchery. In 2008, Dixon *et al.* figured out that a number of analyses of various aquaculture species have shown that the genetic diversity typically declines over successive generations of domesticated stocks. For example, traits such as the survival rate and growth rate have shown to decrease in domesticated stocks of aquaculture species. The loss of genetic variation in domesticated strains is a common phenomenon that could be explained through inbreeding, negative selection, insufficient effective breeding number in the hatcheries, or a combination of these factors (Alam and Islam, 2005).

Selective breeding programs can be exploited to improve the poor performance of hatchery strains. Before it can be implemented, the programs require an analysis of the genetic structure and diversity of the selected strains to determine the levels of inbreeding and genetic variation. Therefore Yue *et al*. (2009) suggested that in order to ensure a successful selective breeding project, it is important to understand the genetic diversity of wild populations so that a high allelic and gene diversity founder population can be set up. It is also important to monitor the genetic structure of the stocks for stock enhancement, management for sustainable yield, preservation of genetic diversity and to ensure that the

genetic variation in economically traits has not been impacted (Barman *et al.*, 2003; Dixon *et al*., 2008).

Investigations of genetic variability and inbreeding, species and strain identification, parentage assignments and the construction of high-resolution linkage maps for aquaculture species can be done rapidly by the application of genetic markers (Li *et al*., 2007a; Divu *et al*., 2008). Genetic markers such as the random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), restriction fragment length polymorphisms (RFLP), mt-DNA sequencing and microsatellites are widely used by researchers to study the genetic structure of diverse aquaculture species such as shrimp, carps, flounder and mussels (Barman *et al.*, 2003; Liu and Cordes, 2004; Wang *et al*., 2004; Liu *et al*., 2005a; MacAvoy *et al*., 2008). However, microsatellites have become the markers of choice for a wide range of applications in genetic mapping and genome analysis because they are easy to clone and characterize besides displaying multiple alleles that are highly polymorphic among individuals (Barker, 2002; Wuthisuthimethavee *et al*., 2003).

Microsatellite which was discovered in the early 80s are a form of repetitive DNA that is abundant across the genomes (Yang *et al*., 2008). It is the short stretches of DNA in which a motif of one to six base pairs is repeated up to 60 times (Barker, 2002). Microsatellites have been widely utilized in the study of genetic variations and genetic structure of cultivated and wild stocks of many fishery species because of the promising future and more variable results can be generated compared to other markers (Yang *et al*., 2008). In this research, 228 samples from eight different locations were analysed to examine the genetic structure using 10 selected microsatellites primers.

1.1 Objective of Study

The objective of this research is to characterize and compare the genetic diversity of wild and domesticated broodstocks of *Macrobrachium rosenbergii* based on microsatellite marker.

CHAPTER 2

LITERATURE REVIEW

2.1 *Macrobrachium rosenbergii*

2.1.1 Taxonomy and distribution of *Macrobrachium rosenbergii*

Macrobrachium rosenbergii (Figure 2.1) is a freshwater prawn under the group of decapod crustaceans. *Macrobrachium rosenbergii* is in the phylum of anthropoda i.e. under the kingdom of animalia. *Macrobrachium* includes around 200 species and is the largest genus in the family of Palaemonoidiae (Valencia and Campos, 2007). Other species that are in the same genus are *M. nipponense*, *M. malcomsoni*, *M. formosense*, and *M. americannum* (Holthuis and Ng, 2009).

Figure 2.1: *Macrobrachium rosenbergii*

Figure 2.2 shows the taxonomic hierarchy of this species. Previously, the giant freshwater prawn was known as *Palaemon rosenbergii* until 1959 when the present scientific name i.e *Macrobrachium rosenbergii* was introduced and accepted world-wide (Food Agriculture Organization of United Nations, 2002). *Macrobrachium rosenbergii* is indigenous to the Indo-Pacific region in which it can be found mainly in Malaysia, Thailand, Philippines, Indonesia, Vietnam, India, Sri Lanka and Myanmar (Tayamen, 2001). However due to its economic importance and the development of aquaculture, the species can now be found in almost all regions in the world (Valencia and Campos, 2007).

This giant freshwater prawn can be found in highly diverse areas such as in circum-tropical marine, estuarine and freshwaters (de Bruyn *et al*., 2004). Tayamen (2001) stated that some of the prawns can also be found 200 km from the sea such as in lake or paddy field because the prawns are able to move upstream by migration. Thus, this can be an explanation to the wide distribution of the prawns.

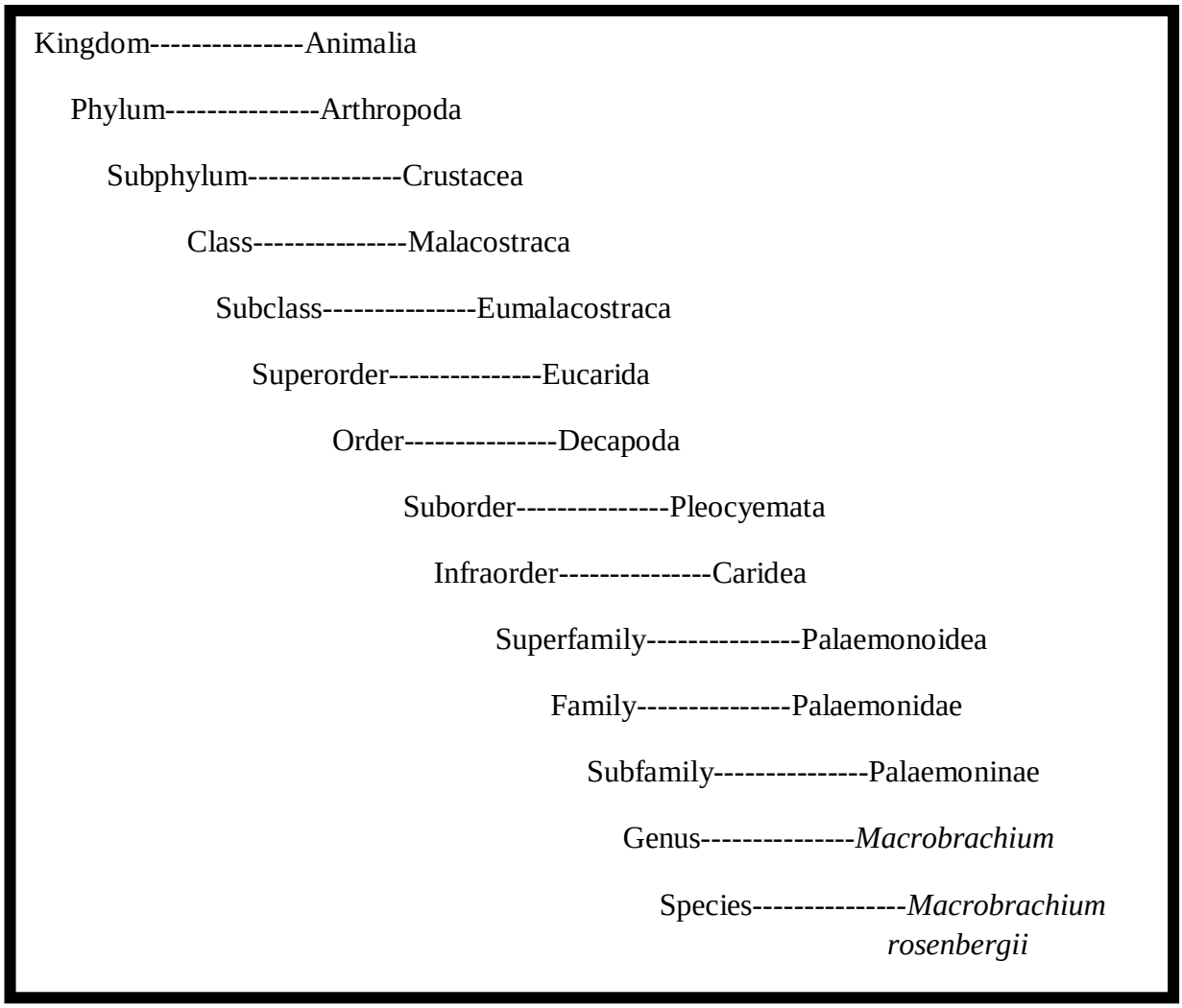

Figure 2.2: The hierarchy of *Macrobrachium rosenbergii*. Retrieved April, 1, 2010 from the Integrated Taxonomic Information System on-line database, http://www.itis.gov

2.1.2 Morphological characteristics of *Macrobrachium rosenbergii*

M. rosenbergii can be distinguished easily from other species in the genus by looking at the body size. It is the largest species in the genera where *M. rosenbergii* males can reach up to 320mm in total body length, whilst the females can grow up to 250 mm (Wowor and Ng, 2007; Holthuis and Ng, 2009). Generally, the body colour of *M.*

rosenbergii is greenish to brownish grey. However, according to Brown *et al*., (2009) for larger specimens, the colour may be bluish or sometimes darker. The main parts of *M. rosenbergii* are shown in Figure 2.3. From the figure, it can be seen that the body of the prawn is divided into two i.e. head or cephalothorax and tail.

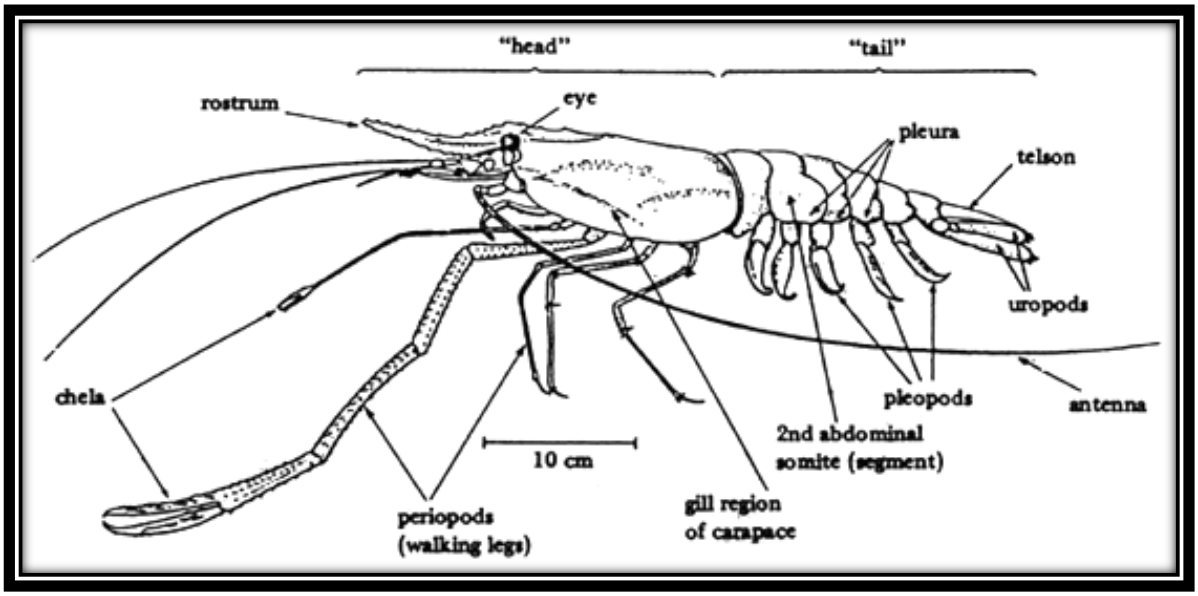

Figure 2.3: The morphology of *Macrobrachium rosenbergii*. Retrieved March, 19, 2010 from http://www.fao.org/docrep/field/003/AC277E/AC277E02.htm

The head part of *M. rosenbergii* consists of the eyes, rostrum and five pairs of legs. Its long rostrum is one of the main characteristics that was used by the researchers to differentiate *M. rosenbergii* from other *Macrobrachium* species. *Macrobrachium rosenbergii* has a very long rostrum compared to other species with 11-14 dorsal teeth and 8-10 ventral teeth. The legs or periopods which are also situated in the head segment play important roles in *M. rosenbergii* life cycle. The first and second periopods both have pincers (chelae) at the end of each periopods. The pincers are useful for capturing food and mating as well as for fighting. The second periopods are longer than any other periopods and are much stronger with numerous spines. The other three periopods which are slender

and much shorter than the second one are called as walking legs. These three periopods does not have pincers but simple claws (Food Agriculture Organization of United Nations, 2002; Brown *et al*., 2009).

The tail or the abdomen part of *M. rosenbergii* consists of six abdominal somites with each having a pair of pleopods. Pleopods are different from periopods as they are used mainly for swimming while the periopods are used while moving on the ground. In females, the pleopods function as the holder of the clusters of eggs. The tail fan which is at the end of the abdomen consists of one telson and two uropods. Their role is to help the prawn moves backward. Male and female prawns can be differentiated based on their sizes. In general, male prawns are larger in body size compared to the females. The head of the males are also bigger than the females. Gravid females are easy to recognize due to its large abdomen (Tayamen, 2001; Brown *et al*., 2009;).

2.1.3 Life cycle of *Macrobrachium rosenbergii*

Macrobrachium rosenbergii is a unique species. Even though *M. rosenbergii* inhabits freshwater, it can tolerate a wide range of salinities ranging from 0 to 25 ppt because the larval and post larval phases are spent in brackish water (Cheng and Chen, 1998). There are four phases in the freshwater prawn life cycle. The first phase is the egg phase followed by the larvae, post-larvae and the final phase i.e. adult. The general life cycle of *M. rosenbergii* is shown in Figure 2.4.

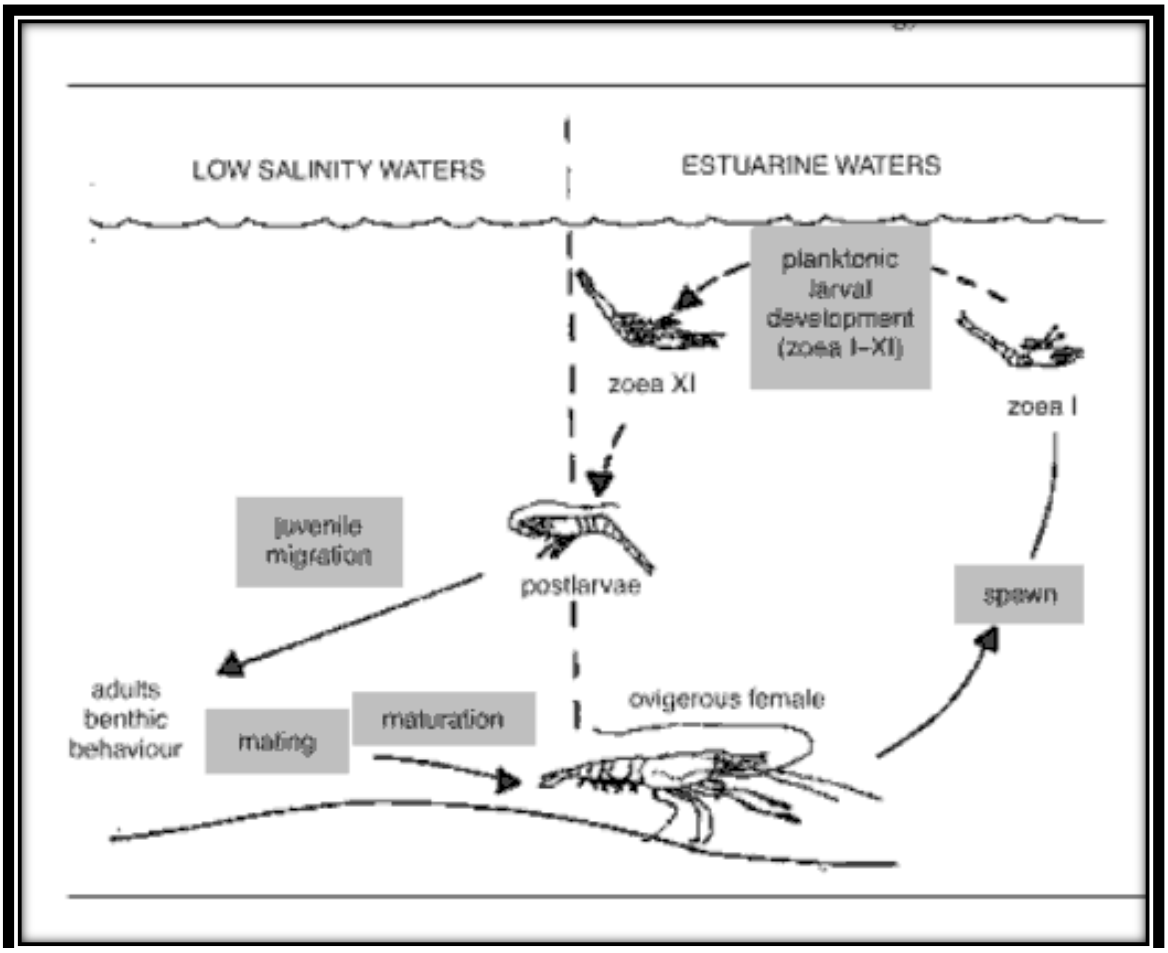

Figure 2.4: The reproduction cycle of *Macrobrachium rosenbergii* (Brown *et al*., 2009)

Macrobrachium rosenbergii females may spawn three to four times a year and produce up to 12,000 eggs each time. The colour of the eggs is typically bright orange. However, it will fade to pale grey and darken to slate grey when they are about to hatch. The egg incubation process normally takes about 19 days at 26 to 28°C (Bardach *et al*., 1972). In a manual for culturing freshwater prawn published by Food Agriculture Organization of United Nations (2002), it is stated that mating can only be successful after the females completed their pre-mating moult.

Under natural conditions, mating can occur throughout the year, although there are sometimes peaks of activity related to the environmental conditions. For example, Rao (1991) reported that the wild *M. rosenbergii* females from India showed a peak of

reproductive activity during the summer months i.e. in August to October. In Israel, O'Donovan *et al*. (1984) study revealed that 90% of *M. rosenbergii* females collected in the ponds during the summer were observed to carry eggs. Both studies were cited in Cavalli *et al*. (2001).

Berried females usually will migrate from freshwater to brackish water which is essential for larval survival and its early development. In this area, the hatched larvae are dispersed by rapid movements of the abdominal appendages of the mother and undergo metamorphose into post-larvae (de Bruyn *et al*., 2004). Previous studies reported that the larvae can complete their larval stage for as little as 16 days or longer depending on the water temperature and other factors (Food Agriculture Organization of United Nations, 2002).

Postlarvae exhibit good tolerance to a wide range of salinities, which is a characteristic of the freshwater prawns. They begin to migrate upstream into freshwater conditions within one or two weeks after metamorphosis. From this point, they are known as juveniles i.e. miniature of adult prawns and they mainly crawl rather than free-swimming. According to Bardach (1972), the juveniles are believed to crawl slowly upstream until two to three months later when they reach pure freshwater. Some of the young *M. rosenbergii* may travel over 60 km upstream from the sea.

2.2 Production of *Macrobrachium rosenbergii* in Malaysia

Macrobrachium rosenbergii has been known as one of the most commercial important aquaculture species. Asia has been recognized as the major producer of this

species. However, this giant freshwater prawn has been cultured not only within Asia but also in regions that are far distant from its natural distribution.

It is known that the broodstock of *M. rosenbergii* originated from Malaysia. The broodstock of the species was brought to Anuenue Fisheries Research Center in Honolulu, Hawaii where Takuji Fujimura and his team successfully developed a protocol to commercialize the culture of this prawn in 1965 (Nhan *et al*., 2009; Wowor and Ng, 2007). Since then, *M. rosenbergii* broodstocks have been disseminated to many other regions including North and South America, Africa, Europe and parts of Asia. As in year 2000, FAO reported that *M. rosenbergii* has been cultured in at least 43 countries across five continents (Divu *et al*., 2008).

Even though *M. rosenbergii* originated from Malaysia, it is not the major aquaculture species that is being cultured in Malaysia. In fact, based on the production percentage, it is the least cultured freshwater species (Figure 2.5) (Department of Fisheries Malaysia, 2009). Although the production is relatively small, *M. rosenbergii* has the highest retail price among other aquaculture species sold in the Malaysian market (Figure 2.6). The retail price of this prawn can reach up to RM 25/kg or more depending on the consumers demand. According to New (2005), *M. rosenbergii* is said to be increasingly popular among consumers of all races in Malaysia. Thus this might be one of the main factors that make the price of this giant freshwater prawn high compared to other freshwater species and remained as one of the most important commercial species.

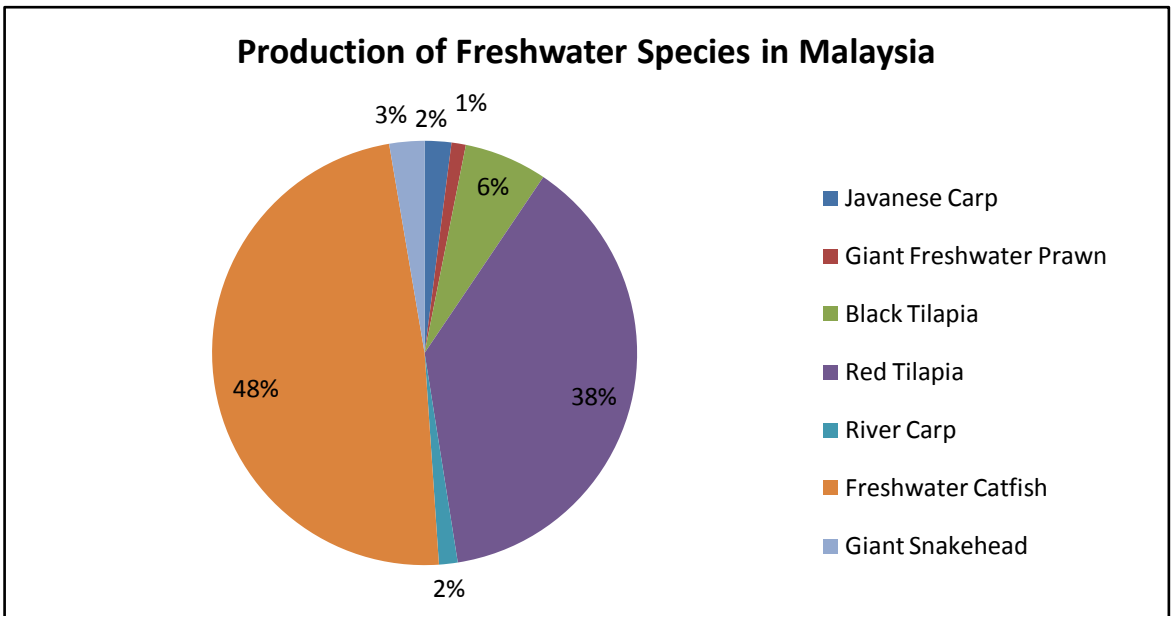

Figure 2.5: Average production percentage of freshwater species in Malaysia from year 2003 to 2007 (Department of Fisheries Malaysia, 1997-2008)

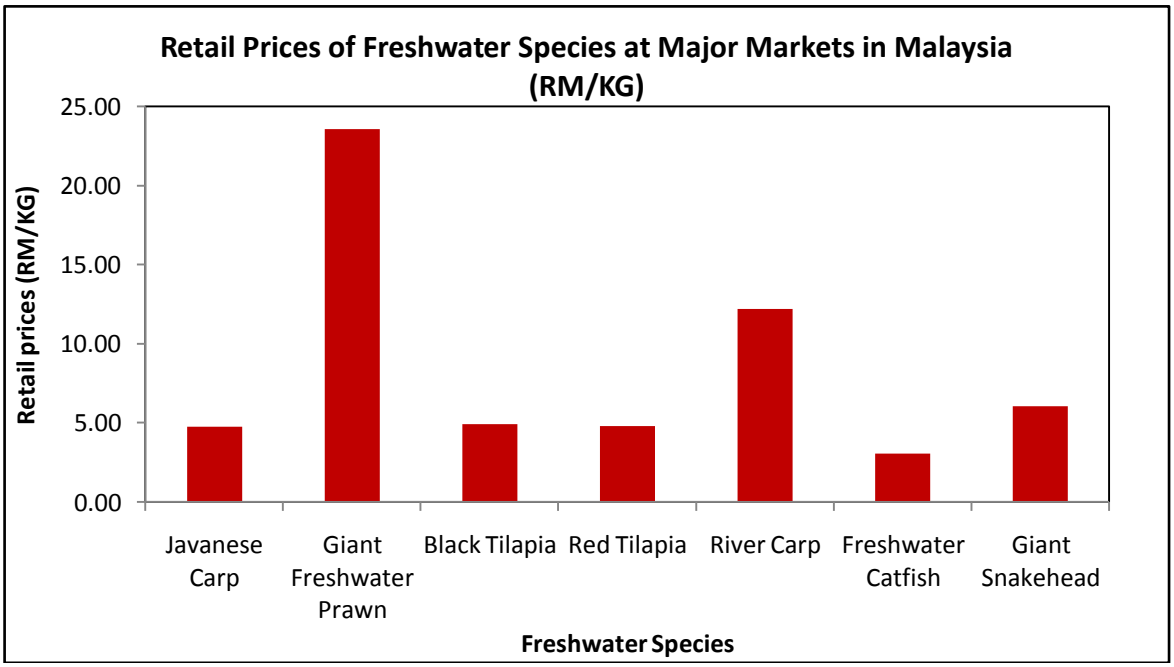

Figure 2.6: Average retail price for freshwater species sold in Malaysia from year 2003 to 2007 (Department of Fisheries Malaysia, 1997-2008)

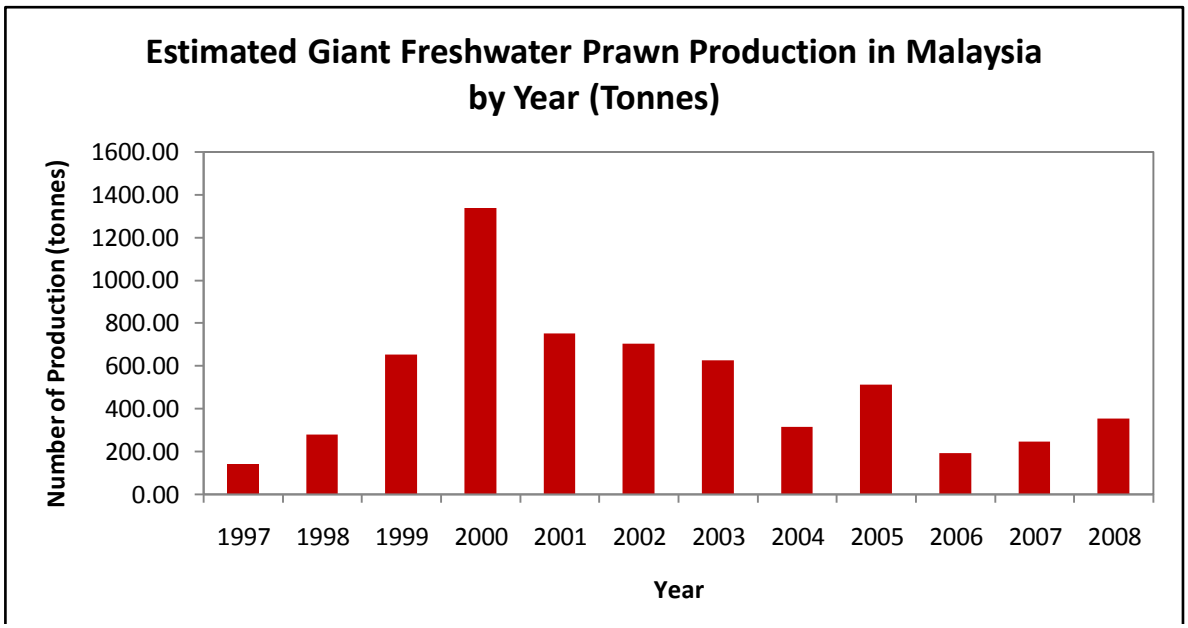

Figure 2.7: Trends in *Macrobrachium rosenbergii* (giant freshwater prawn) production in Malaysia from 1997 to 2008 (Department of Fisheries Malaysia, 1997-2008)

Malaysia is known as one of the top ten producers of *M. rosenbergii* (New, 2005). Data provided by Department of Fisheries Malaysia denotes that the production of *M. rosenbergii* in Malaysia expanded rapidly from 1997 to 2000 (Figure 2.7). However, in year 2001 the production began to decrease until 2004 when the production dropped tremendously. There is no information or study that has been done to investigate the reason of the sharp decrement. In 2005, the production of *M. rosenbergii* rose however it fells again in the following year. Starting from the year 2006, the production of this freshwater species has been slightly increasing each year. This is most probably due to the continuous recent market expansion.

2.3 Molecular Markers

Generally there are two types of molecular markers i.e. mitochondrial DNA (mtDNA) and nuclear DNA markers (Arif and Khan, 2009). The application of molecular markers in genetic studies is reliable since it can provide consistent results for rapid identification among species, levels of genetic variability, levels of gene flow and also the underlying factors that contribute to fitness (Esa *et al*., 2008).

Since its development in the early 1990s, the molecular markers could offer a method to assist the selection of individuals in breeding programs thus, enhance the domestication process to a shorter time (Garcia and Benzie, 1995). Recently, molecular markers are known to be the most reliable method to assess the genetic status of the hatchery populations against the known levels of variation in wild populations (Alam and Islam, 2005). A summary of advantages and disadvantages of each type of molecular markers are illustrated presented in Table 2.1.

Allozymes are electrophoretically distinguishable protein variants (Maguolas, 1998). It is a classical assay that had been widely used in studying genetic variability in populations (Zhang *et al*., 2002). According to Li *et al*. (2004), even though the resolution power of allozyme markers has allowed the genetic discrimination of hatchery and wild populations, it has not allowed for the effective detection of changes in genetic composition of hatchery strains. This might be due changes in the nucleotide sequence that do not correspondingly change the encoded polypeptide.

Liu and Cordes (2004) explained that the level of detectable variation in allozymes may also be reduced because of polypeptide changes that may not alter the mobility of the protein in the electrophoretic gel. Therefore, molecular marker systems with greater levels

of capability for the detection of polymorphism and a better solution for the assessment of genetic variations are very much needed.

Restriction fragment length polymorphisms (RFLPs) is one of the early molecular markers developed on the basis of the use of bacterial restriction enzymes that cut the DNA at sites with specific nucleotide sequences. This technique was developed to visualize the differences at the level of the DNA structure (Montaldo and Meza-Herrera, 1998). Another type of molecular marker is the random amplified polymorphic DNA (RAPD). The marker is analyzed by using PCR to amplify segments of nuclear DNA. Single primer that attaches to both strands of DNA is used and low annealing temperatures increase the likelihood of amplifying multiple regions that represent a particular locus.

RAPD is a simple and inexpensive technique (Arif and Khan, 2009). However its major limitation is the inability to differentiate between homozygote and the heterozygote; thus it is classified as a dominant type. Previous studies have demonstrated that the restriction fragment length polymorphisms (RFLPs) and random amplified polymorphic DNA (RAPD) are useful for genetic studies especially in the area of genetic relationships and diversities (Barman *et al*., 2003; Wuthisuthiwethavee *et al*., 2003).

AFLP or amplified fragment length polymorphism is a multi-locus technique which is also a PCR-based marker that combines with the restriction endonuclease digestion practice (Wang *et al*., 2004). These markers are dominant and almost similar to the RAPD. However, it is more sensitive than the RAPD due to the multi-locus fingerprinting technique thus producing high number of polymorphic fragments.

Arif and Khan (2009) described that the strength of AFLP compared to the RAPD is its reproducibility due to the use of restriction digestion of DNA, specific adaptors and high annealing temperatures for selective amplification. This marker had been successfully and widely applied to study genotyping, population differentiation and genetic diversity in

many aquaculture species such as catfish, kuruma prawn, penaeid shrimp, Japanese flounder and pearl oyster (Liu *et al.*, 1998; Li *et al.*, 2003; Wang *et al*., 2004; Liu *et al*., 2005a; Yu and Chu, 2006).

Microsatellites are nuclear DNA type markers and similar to other nuclear DNA markers, it also uses PCR as the basis of analysis. Jarne and Lagoda (1996) stated that microsatellites are the most powerful Mendelian markers. Microsatellites represent the ideal molecular markers because they have multiple alleles that are highly polymorphic among individuals. Microsatellite loci are highly abundant and dispersed evenly throughout the eukaryotic genomes. They are also inherited in a co-dominant fashion, and are fast and easy to assay.

Microsatellites have various applications in genetic studies such as genetic diversity, parentage and kinship assessment, monitoring genetic structure, etc. Moreover, they can also serve as the sequence-tagged sites for anchoring in genetic and physical maps (Wuthisuthiwethavee *et al*., 2003).

Mitochondrial DNA (mtDNA) is also a molecular marker. However it is a different type of markers than the mentioned above. mtDNA as the name suggests is not a nuclear marker and follows a non-Mendelian mode of inheritance (Liu and Cordes, 2004). It has several characteristics different from the nuclear marker as it occurs in multiple copies in each cell, is transmitted uniparentally as well as does not recombine (Magoulas, 1998). Thus, this marker is widely used for maternal and paternal lineage studies. Erhardt and Weimann (2007) stated that the analyses by mtDNA are important on domestication process because it allows issues such as threats and conservation between domestication and wild species to be characterized.

Table 2.1: Summary of advantages and disadvantages of available molecular markers.

Molecular markers	Advantages	Disadvantages	Reference
Allozyme markers	• Cheap • Genetic interpretation is straightforward	• Require large quantity of tissue sample • Not effective for genetic changes detection	Magoulas, 1998; Zhang *et al*., 2002; Li *et al*., 2004
AFLP (amplified fragment length polymorphism)	• High levels of polymorphism • High resolution	• Not transferable between labs, species and crosses	Wang *et al*., 2004; Liu *et al*., 2005; Ning *et a*l., 2007
RAPD	• Quick • Cheap • Low experimental requirements	• Specific characteristics • Low reproducibility • Detection of polymorphism is not very high	Barman *et al*., 2003; Liu *et al*., 2005a
Microsatellites	• High number of polymorphic fragments • Reliable and reproducible bands	• The development of the markers are tedious and costly • The presence of stutter bands and null alleles	Magoulas, 1998
mtDNA	• Highly-conserved	• Limited numbers of available markers	Liu and Cordes, 2004

2.4 Microsatellite Markers

Microsatellites are short stretches of repetitive DNA that are abundant across the genomes of both prokaryotes and eukaryotes. Since they are very abundant, this type of marker is sufficient enough to be readily developed and used for various research projects (Wright and Bentzen, 1994). It core sequences consist of two to six base pairs lengths and repeated up to 100 times (Zhivotovsky and Feldman, 1995; Barker, 2002). A tandem repeat is often formed from simple dinucleotide for example $(GA)_n$ where *n* is the repeated number of the dinucleotide.

Generally, it has been discovered that mammals such as humans and rats, plants, birds and fishes have their certain characteristics of microsatellite motifs that occur frequently in their genomes. For example, the repeat motifs of (CA)n are rare in plants. However, (AA)n and (AT) are very common. On the other hand, the fish genome consists of more (CA)n and (CG)n repeat motifs (Yang *et al*., 2008). Sekar *et al*. (2009) reported that this structure which was discovered in the early 1980s is also known with other names such as the short or simple tandem repeats (STRs), simple sequence repeats (SSRs), simple sequence length polymorphisms (SSLPs) and variable number tandem repeats (VNTRs).

Microsatellites are the co-dominant markers present in nuclear DNA and located in both coding and non-coding region (Ariff and Khan, 2009; Yue *et al*., 2009). They are neutral, highly polymorphic with up to dozens of alleles at each locus and inherited in a Mendelian fashion (Yue *et al*., 2009; Jerry *et al*., 2006; Zhivotovsky and Feldman, 1995). Thus, the microsatellites provide more information especially in population and pedigree studies compared to other dominant markers. Microsatellites which exhibit high levels of allelic variation are preferable in many research contexts, for example, for species that shows low overall levels of variation with conventional markers.

Microsatellites are also useful to study populations that are inbred or geographically proximate where the genetic differentiation may be limited (Wright and Bentzen, 1994).

Another attractive point of the microsatellites is it can be easily assayed using the PCR method. Thus, only small amounts of DNA from any tissue such as scale, muscle, pleopods, etc. are needed for analysis (Esa *et al*., 2008; Wright and Bentzen, 1994). Ward (2000) believed that this attribute is very important especially for species or populations that are being protected where non-destructive sampling is required for genetic analysis (Esa *et al*., 2008). Microsatellite markers are also highly reproducible (Yang *et al*., 2008). This type of molecular marker can provide a relatively low cost for the user with rapid research progress when it is analysed in a multiplex fashion (Li *et al*., 2007a).

There are several types of microsatellites which can be described by the number of nucleotide in the repeat motifs (Table 2.2). Dinucleotide is the most abundant repeat loci found in the genome and usually used in linkage studies or building genetic maps. Tetranucleotide repeat loci i.e. easier to interpret had been used as a forensic DNA marker (Walsh *et al*., 1996). Sekar *et al*. (2009) mentioned that microsatellite can also be described by the term perfect, imperfect and compounds. Perfect microsatellites refer to an uninterrupted stretch of the identical repeats while imperfect microsatellites are repeat sequences with one or more interruptions. A repeat sequence which is made up of two or more adjacent tandem repeats is described as compound microsatellite.

Table 2.2: Types of microsatellite DNA based on their number of repeat nucleotides

Types of microsatellite	Number of repeat nucleotide
Dinucleotide	2
Trinucleotide	3
Tetranucleotide	4
Pentanucleotide	5
Hexanucleotide	6

2.4.1 Application of microsatellites marker

Microsatellites can be used in analyzing genetic diversity of wild and cultured populations such as mud carp, black tiger shrimp, common carp and Pacific abalone (Li *et al*., 2004; Li *et al*., 2007a; Li *et al*., 2007b; Yang *et al*., 2008). Fine genetic variations can be detected by microsatellite DNA markers due to their high level of polymorphism. Recently, there are major concerns in studying the genetic diversity of cultured stocks of aquaculture species since they show lower genetic diversity compared to the wild stocks. Species or strain can also be identified using microsatellites marker (Liu and Cordes, 2004).

The markers have also been used extensively for parentage assignments and pedigree tracing in prawn lines and hatchery populations due to it codominance characteristics (Divu *et al*., 2008; Yang *et al*., 2008). According to Jerry *et al*. (2006), reliable pedigree data is crucial to determine the success of a breeding program which is based on family selection. Hence, it is crucial to know the relatedness between individuals in order to minimize inbreeding and the loss of genetic variation in cultured stocks (Arif and Khan, 2009).

Microsatellites have been selected as the marker of choice since it can also be used for individual identification and kinship inference (MacAvoy *et al*., 2008). Its high allelic diversity characteristic provides microsatellites with best resolving power in determining pedigree relationships among individuals (Jerry *et al*., 2006). Microsatellite markers are more preferable in genetic studies compared to other molecular markers because of its high level of heterozygosity and transferability across strains and species (Ning *et al*., 2007).

Microsatellite markers have also been suggested to be used for traceability in aquaculture species. Traceability schemes are important in aquaculture industry as they provide information on the origin and the production of chain of food products to consumers. For example, in a case of disease detection or toxins in market fish traceability schemes, it allows for tracing of fish back to the farm of origin. According to Hayes *et al*. (2005), DNA markers i.e. microsatellites and single nucleotide polymorphisms (SNPs) have been introduced as one of the traceability methods. Microsatellite marker is more preferable because it is highly informative and can verify results from other traceability scheme methods such as chemical marking and external tagging.

2.4.2 Limitations of microsatellite markers

Even though microsatellite markers display many characteristics that are advantageous for genetic studies, there are also some limitations to their use that need to be understood and addressed before the markers can be applied (MacAvoy *et al*., 2008). One of the major concerns is the genotyping errors. There are several factors that could

contribute to this problem. One of the most frequently reported is the occurrence of null alleles.

Null alleles are alleles which fail to amplify during the PCR. It is usually because of mutations that occurred within a primer site (Wagner *et al,* 2006; Oosterhout *et al*., 2004). There are two types of genotyping problems caused by null alleles. According to Wagner *et al*. (2006) the first type of genotyping error occurs if an individual is a homozygote where the genotyping will fail. The other type of genotyping problem takes place if an individual is a heterozygote with one null allele. In this case, the observed genotype will be indistinguishable from a true homozygote resulting to 'false homozygote' (Oosterhout *et al*., 2004).

It is common observation during the PCR process that short alleles amplify more efficiently than the larger alleles. This is known as large allele dropout or short allele dominance (Oosterhout *et al*., 2004). The differential amplification of size-variant alleles are sometimes called as 'partial null alleles' since they can be made visible by loading more sample or adjusting contrast (Dakin and Avise, 2004). Another problem or genotyping error in microsatellite data occurs when inconsistent DNA template quality or low template quantity is used in the PCR. This may result in an allele failing to amplify due to stochastic sampling error or also known as allelic dropout (Oosterhout *et al*., 2004).

Stutter or shadow band is an additional PCR product observed due to slippage during the PCR-amplification. Referring to a report by Shinde *et al*. (2003), slippage means the extension of a primer-template complex containing a loop in either the primer or template strand by a DNA polymerase. It may occur either in the active site of the enzyme or before the substrate binds to the enzyme. Stutter products are often formed in dinucleotide loci and usually differ from the original PCR product by multiples of the

repeat unit length. Hence, it is difficult to discriminate between the homozygotes and heterozygotes (Oosterhout *et al*., 2004).

Null alleles usually affect the interpretation of microsatellites data especially in the parentage analysis. Parents and offspring must share at least one identical allele at every locus. Thus, even the occurrence of null alleles at very low frequencies may eliminate potential parents as possible candidates (Wagner *et al*., 2006).

Genotyping errors in microsatellite data may also cause deviations from Hardy-Weinberg proportions, in particular a heterozygote deficiency which potentially bias the population genetic analyses (Oosterhout *et al*., 2004). To solve the problem of null alleles, researchers usually eliminate the primer at the affected locus by redesigning the primer or develop new primer for alternate loci that do not have null alleles. However both solutions are not preferable for many researchers who seek to apply microsatellite in their studies since it involves additional time and expense (Wagner *et al*., 2006).

2.5 Genetic characterization of wild and domesticated stocks

Genetic studies are becoming very increasingly important in recent years. This is due to the distribution and population sizes for many species of plants and animals which have been heavily affected by the humans activities. Therefore, to save populations from extinction, artificial propagation or the transfer of individuals are introduced and has become a common practice. However, little is known about how this practice can affect the distribution of genetic variation (Zhao *et al*., 2005). Genetic studies are important to protect the genetic resource bank, conserving genes for significant characteristics and restraining the genetic decline (Li *et al*., 2007b). Genetic

variation of endangered species should be retained as much as possible to enhance the chance for their recovery (Yue *et al*., 2004).

Genetic resources such as genetic variation of a species can be exploited for selective breeding programs. Selective breeding seems to be a promising approach in order to produce strains that are more resistant to diseases (Liu *et al*., 2006). Genetic study is important in selective breeding programs so that good quality breeding stocks can be identified. Low quality of breeding stocks can lead to low productivity for example in China, some pearl mussel farms had to be closed due to the poor quality of breeding stocks (Li *et al*., 2009a).

CHAPTER 3

MATERIALS AND METHODOLOGY

3.1 Sample collections

Four wild and four domesticated samples of giant freshwater prawn, *M. rosenbergii* obtained from selected areas in Malaysia were analysed in this study. The samples of wild *M. rosenbergii* were collected from Sarawak, Kedah, Terengganu and Negeri Sembilan. Domesticated giant freshwater prawn samples were obtained from the hatcheries located in Negeri Sembilan, Kedah and Thailand. Details of sample locations are presented in Table 3.1 and Figure 3.1.

Table 3.1: Details of *M. rosenbergii* samples locations

Population abbreviation	Population name	Location	Longitude/Latitude	Population type	Sample size
THAI	Thailand	Thailand	(the location of the hatchery was not revealed)	Domesticated	30
NAPFRE	National Prawn Fry Production and Research Centre	Pulau Sayak, Kedah	$5^0$40'00"N $100^0$20'06"E	Domesticated	30
HatA	Mun's Aquaculture	Pantai, Negeri Sembilan	2°48'11" N 101°59'30" E	Domesticated	30
HatB	Wong's Hatchery	Kuala Kelawang, Negeri Sembilan	$2^0$56'49"N $102^0$05'07"E	Domesticated	30
SRWK	Sarawak River	Sungai Serian Sarawak	$1^0$50'05"N $113^0$54'06"E	Wild	30
TRGN	Terengganu River	Sungai Penarik Terengganu	$5^0$37'48"N $102^0$48'36"E	Wild	30
NSBL	Timun River	Sungai Timun/Linggi Negeri Sembilan	$2^0$28'29"N $102^0$02'05"E	Wild	24
KDH	Kedah River	Sungai Muda, Kedah	$5^0$43'01"N $100^0$31'46"E	Wild	24

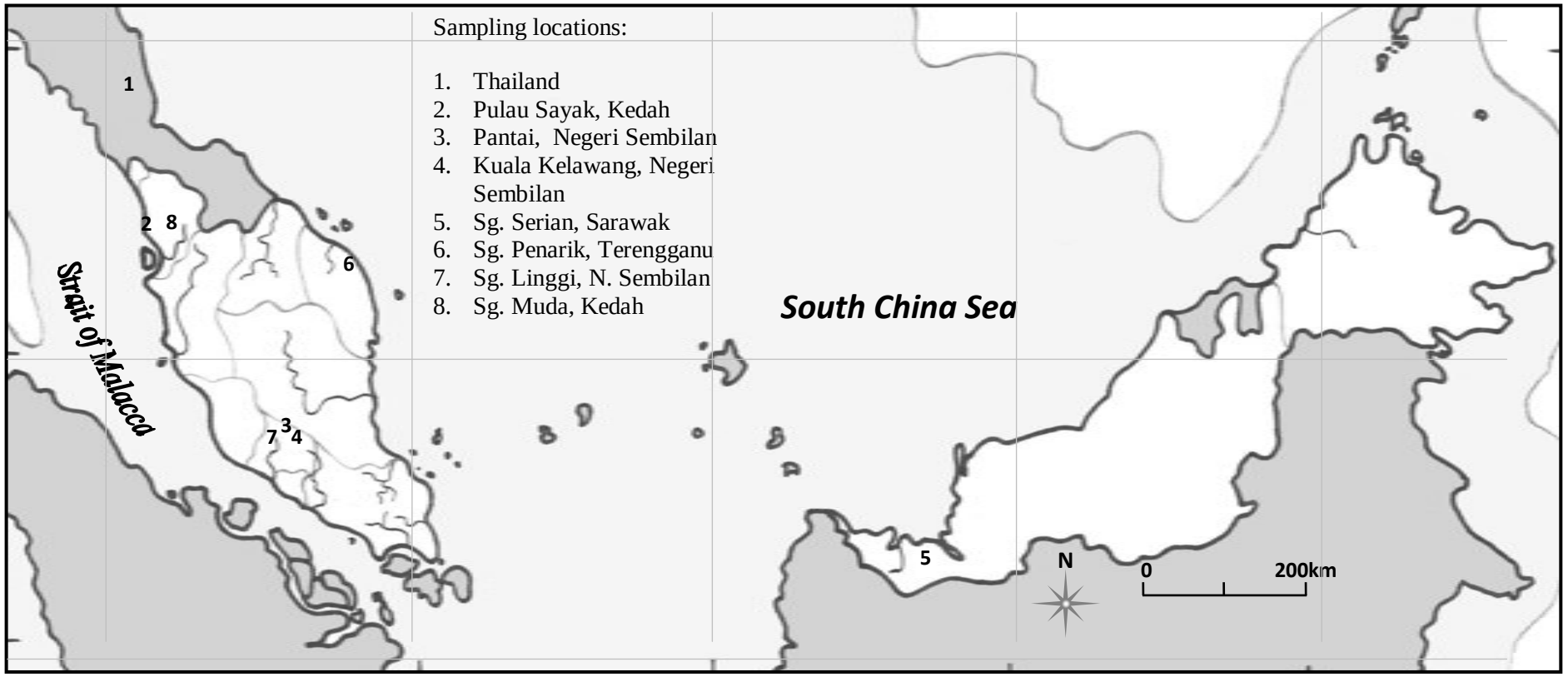

Figure 3.1: Sampling locations for giant freshwater prawn (*Macrobrachium rosenbergii*) in Malaysia

3.2 Genomic DNA extraction

Genomic DNA from tissue of *M. rosenbergii* was extracted using a GF-1 Tissue Extraction Kit (Vivantis). Extraction protocols were done according to the manufacturer's instruction with slight modifications. 30 – 40 mg of *M. rosenbergii* tissue samples were weighed and placed into the 1.5 ml micro-centrifuge tube. The tissue was mashed using sterilized toothpicks until it breaks into small pieces. 250 µl of Buffer TL and 20µl of Proteinase K were added to the sample and mixed thoroughly by pulse-vortexing to obtain a homogenous solution. 12 µl of Lysis Enhancer were added and mixed immediately. Tissue samples were incubated overnight at 65°C in a waterbath. During incubation, tissue samples were occasionally mixed to ensure a thorough digestion.

On the following day, 20 µl of RNase A was added to the sample followed by the process of mixing and incubating at 37°C for 15 minutes. Then the sample was added with 600 µl of Buffer TB and vortexed to obtain a homogeneous solution followed by an incubation of 10 minutes at 65°C. After the incubation, 200 µl of absolute ethanol was added to the sample and immediately mixed to prevent any uneven precipitation of nucleic acid due to the high local ethanol concentrations. 600 µl of the tube mixture was gently transferred into a column assembled in a clean collection tube and centrifuged at 5000 x *g* for one minute. The flow-through was discarded and the remainder of the original mixture was then processed in the same manner.

Next step, the column was washed with 750 µl Wash Buffer and centrifuged at 5000 x *g* for one minute and the flow through was discarded. Column washing process was repeated once again. After that the column was centrifuged at 10 000 x *g* for one minute to remove all traces of ethanol. Finally the column was placed into a clean 2.0 ml micro-centrifuge tube and 120 µl of preheated elution buffer was added directly onto

the column membrane and allowed to stand at room temperature for three hours. Then it was centrifuged at 5000 x *g* for one minute to elute the DNA. The extracted DNA was then stored at -20°C.

3.3 Primer amplification

10 microsatellite primers were used in this study (Table 3.2). Eight out of the 10 primers had been designed and published by Charoentawee *et al.* (2006). PCR amplification was carried out in 10 µl reaction mixture that included $MgCl_2$ ranging from 1 to 1.5 mM, 1X Promega reaction buffer, 0.25 mM each of dATP, dCTP, dGTP and dTTP, 1 unit of Promega *Taq* Polymerase, 100 mM of each forward and reverse primer, template DNA and deionized distilled water. Forward primers of each marker were fluorescently labelled at 5' end with 6 FAM (NHK Biosciences).

A master mix containing all the mixture components except for the DNA sample for PCR reactions was prepared. 2 µl of template DNA was pipetted into labelled PCR tubes. Then 8 µl of prepared master mix was added to the tubes and tapped gently. PCR amplification was performed by using a Labnet multigene thermocycler with following protocol:

Step	**Temperature**	**Time**
Initial denaturation	94°C	5 mins
1. **Denaturation**	94°C	40 s
2. **Annealing**	48 - 62°C	40 s
3. **Extension**	72°C	40 s
Repeat step 1 – 3 for 35 cycles		
Final extension	72°C	7 mins

Table 3.2: Sequences of the 10 pairs of microsatellite markers

Locus	Primer sequences (5'-3')	Repeat sequence	Annealing temp. (°C)	Size range (bp)	Accession no.
Mbr 1	F:CCCACCATCAATTCTCACTTACC R:TCCTTTTCACATCGTTTCCAGTC	(GA)24	60	272-320	DQ019863
Mbr 3	F:CAACTCTATGTTTCGGCATTTGG R:GGGGAATTTTACCGATGTTTCTG	(AG)14	62	232-284	DQ019865
Mbr 4	F:CCACCTACCGTACATTCCCAAAC R:CGGGGCGACTTTTAGTATCGAC	(GT)29	62	288-326	DQ019866
Mbr 5	F:CAAGGCTCGTGTCTCTTGTTTC R:GCTTGTACTTGTTCAGCTTTTGC	(AG)25	62	286-328	DQ019867
Mbr 7	F:ATAAAAGAGTCGCCAAATGAGCA R:ATTGGGAATTGTTGACCTCCAAG	(TGC)16	62	274-286	DQ019869
Mbr 8	F:AACCAGCCGACTTAGACTGTGC R:CGCCATTTGCGTCTATCTCTTAC	(AGC)6(AG)5AA(AG)4	62	256-266	DQ019870
	F:TTTCAGGCTATATCAAGCAACAG R:ATGACGATGATGAGGAATGAAGC	(ATG)3A(ATG)4	60	n/a	DQ019871
Mbr 10b	F:ATGACGATGATGAGGAATGAAGC R:TTTCAGGCTATATCAAGCAACAG	(ATG)3A(ATG)4	62	n/a	DQ019871
UVC 807	F:TACGTGATTCGAGGCATGAG R:CTACCGGGACTAGTGGAACG	(GA)27	55	n/a	Unpublished
UVC 817	F:ATGGCCAAGATGAAAGATGC R:CTGTCTGTACCGCAGTCGAA	(CT)20	57	n/a	Unpublished

Amplification products were resolved electrophoretically on 4% Metaphor gels to ascertain that the amplification had been successful and to estimate the concentration. Molecular weights were estimated using a 20bp DNA ladder.

3.4 Gel electrophoresis

PCR products were electrophoresed on 4% Metaphor agarose gel. This gel was prepared by adding 4 g of Metaphor agarose into 100 ml 1X TBE buffer in a 500 ml conical flask. During the adding, the flask was gently shaken to prevent formation of clots. The mixture was heated in a microwave oven until a completely dissolved mixture was obtained. The heated mixture was then poured into the gel tray and will be allowed to solidify for 2 hours. When the gel had solidified, the comb was removed and the gel was placed into an electrophoresis tank. The tank was filled with 1X TBE buffer until the gel had completely submerged. 3 µl of PCR products were mixed with 1.5 µl loading dye and loaded into each well of the gel. A 20bp ladder was used to determine the size of the PCR product.

Electrophoresis was carried out at 70 V with 150 Amp for 2 hours and 30 minutes. After electrophoresis, the gel was removed and submerged in a container with ethidium bromide solution for 20 minutes. Then the gel was photographed using the Alpha Imager Gel Documentation System.

3.5 Fragment analysis

PCR products were diluted into 10X dilution by adding 0.1 µl of the mixture with 0.9 µl of ddH_2O. 10 µl of Hidi was added into the tube followed by 0.2 µl of Genscan 500 LIZ ladder. The mixture was mixed by pipetting it several times followed by vortexing for a few seconds. The tubes were quickly spun using the Denville, Scientific Inc. brushless microcentrifuge. The mixture was then heated for 5 minutes at 95°C in ABI Gene Amp PCR System 2400. After heating, the tubes were kept in ice for exactly 5 minutes. The whole mixture was loaded into the analysis plate and covered with septa. The analysis plate was tapped to make sure that no bubble exists. The plate was placed into ABI 3100 Genetic Analyzer. Results obtained were analyzed using the GeneMapper 3.5 software (Applied Biosystems)

3.6 Statistical Analysis

Micro-Checker was used in this study to check for scoring errors in the microsatellite data since factors such as DNA degradation, low DNA concentrations and primer-site mutations may result into null alleles, stutter bands and large allele dropout during the amplification process by PCR (Oosterhout *et al.*, 2004). CONVERT software version 1.31) was used to convert raw data from excel sheet into formats that can be read by other population genetic computer programs such as GENEPOP, POPGENE, ARLEQUIN, GDA, STRUCTURE AND PHYLIP (Glaubitz, 2004, obtained from http://www.agriculture.purdue.edu/fnr/html/faculty/Rhodes/ Students%20and%20Staff/glaubitz/software.htm).

Observed number of alleles (A), effective number of alleles (N_e), observed heterozygosity (H_o) and expected heterozygosity (H_e) were analyzed using the POPGENE version 1.31 (Yeh *et al*., 1999, retrieved from http://www.ualberta.ca/~fyeh/). Genotypic disequilibrium between pairs of loci and test for deviation from Hardy-Weinberg equilibrium (HWE) were calculated by the software of GENEPOP version 4.0 (Rousset, 2008; available at http://genepop.curtin.edu.au). Levels of significance for this test were adjusted using the false discovery rate control (FDR) which is an alternative to the Bonferroni correction procedure (Benjamini and Hochberg, 1995)

GDA or Genetic Data Analysis (Lewis and Zaykin, 2001, available at http://lewis.eeb.uconn.edu/lewishome/software.html) was employed to estimate the genetic distance between populations computed by Nei (1978). Analysis of Molecular Variance (AMOVA) which is implicated in ARLEQUIN version 3.11 was used to evaluate the population pairwise F_{ST} value and F_{IS} value (Excofifier *et al*., 2005, downloaded from http://lgb.unige.ch/arlequin/).

In this study, population structure was inferred by a programme named STRUCTURE (Falush *et al*., 2003, retrieved from http://pritch.bsd.uchicago.edu/software). A consensus tree was constructed using a set of programmes in PHYLIP 3.63 software (Felsenstein, J. 2005, obtained from http://evolution.genetics.washington.edu/phylip.html).

CHAPTER 4

RESULTS

4.1 Microsatellite amplification and band scoring

Ten different primers which are fluorescently labeled were used to amplify the microsatellite loci in wild and domesticated stocks of *M. rosenbergii*. PCR conditions used for the amplification are as in Section 3.3 with the annealing temperature for each primer as listed in Table 3.2. The PCR products were run on 4% metaphor agarose to check on the success of the amplification. Confirmation could also be done on 1% Agarose gel. However, 4% metaphor was used because sometimes when the fragment analysis has failed, manual scoring can be done to obtain the allele size. Figure 4.1 to 4.10 demonstrate the banding profile of *M. rosenbergii* from selected populations using the specific primers.

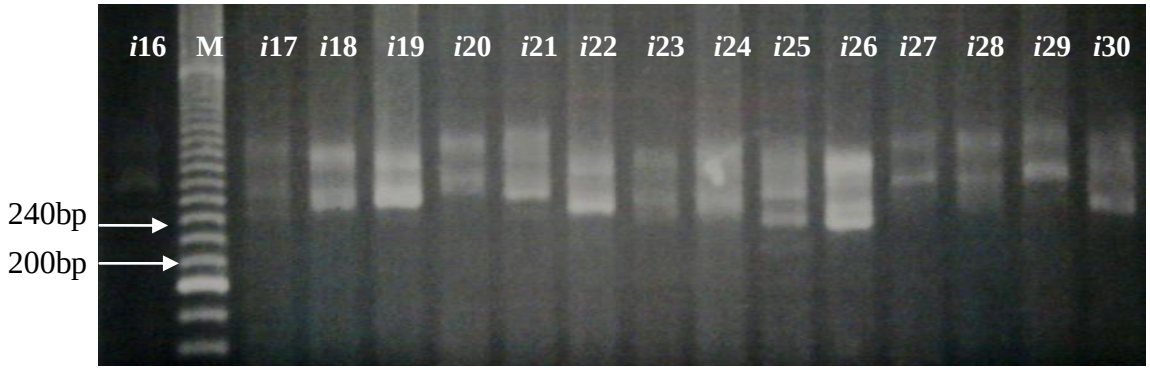

Figure 4.3: Microsatellite banding profile of *M. rosenbergii* (sample TRGN) using primer pair *Mbr*-1 obtained on 4% Metaphor agarose gel. Lane M: 20 bp DNA ladder.

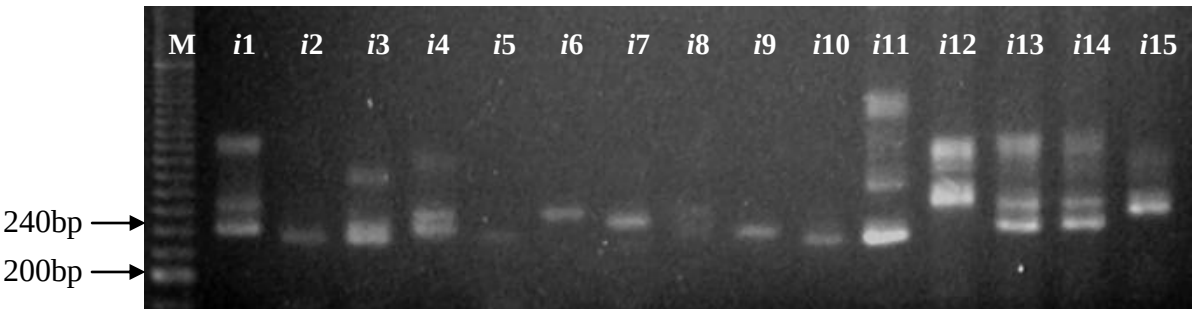

Figure 4.2: Microsatellite banding profile of *M. rosenbergii* (sample THAI) using primer pair *Mbr*-3 obtained on 4% Metaphor agarose gel. Lane M: 20 bp DNA ladder.

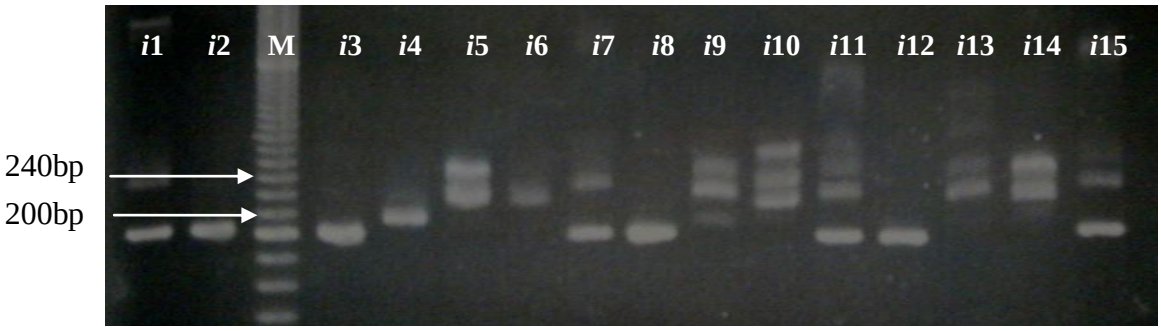

Figure 4.3: Microsatellite banding profile of *M. rosenbergii* (sample HatB) using primer pair *Mbr*-4 obtained on 4% Metaphor agarose gel. Lane M: 20 bp DNA ladder.

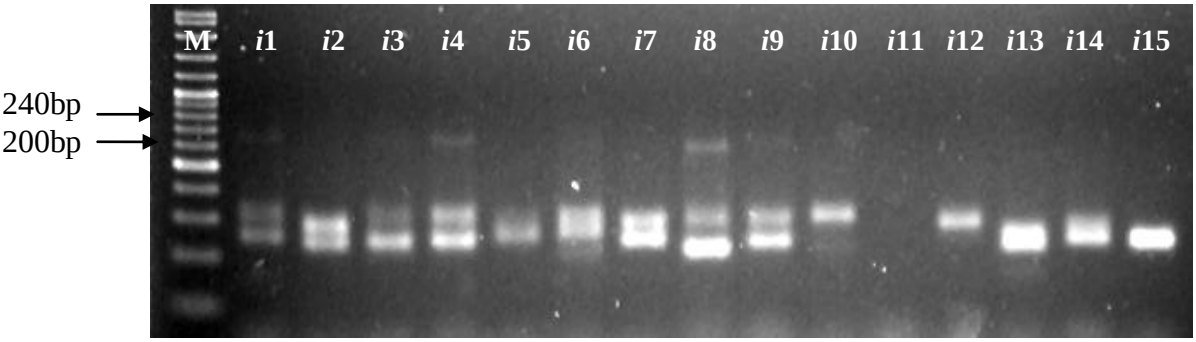

Figure 4.4: Microsatellite banding profile of *M. rosenbergii* (sample SRWK) using primer pair *Mbr*-4 obtained on 4% Metaphor agarose gel. Lane M: 20 bp DNA ladder.

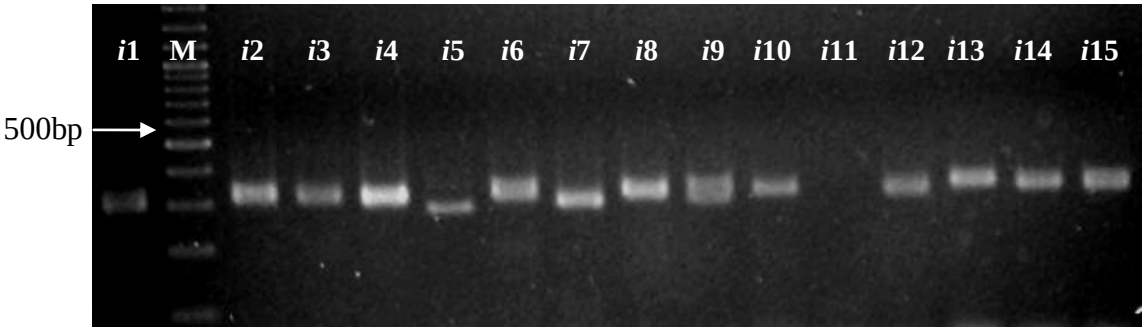

Figure 4.5: Microsatellite banding profile of M. rosenbergii (sample SRWK) using primer pair *Mbr*-5 obtained on 4% Metaphor agarose gel. Lane M: 100 bp DNA ladder.

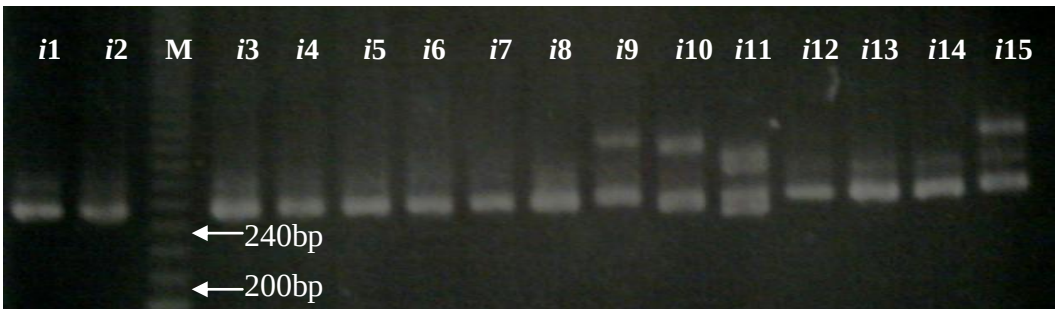

Figure 4.6: Microsatellite banding profile of *M. rosenbergii* (sample NAPFRE) using primer pair *Mbr*-7 obtained on 4% Metaphor agarose gel. Lane M: 20 bp DNA ladder.

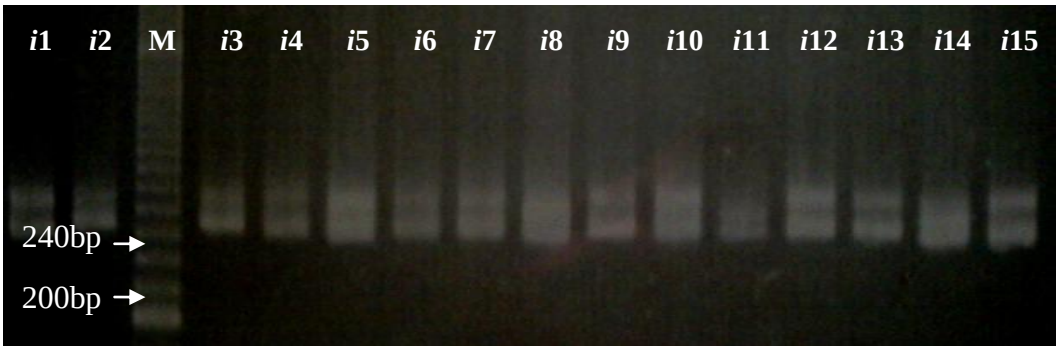

Figure 4.7: Microsatellite banding profile of *M. rosenbergii* (sample HatA) using primer pair *Mbr*-8 obtained on 4% Metaphor agarose gel. Lane M: 20 bp DNA ladder.

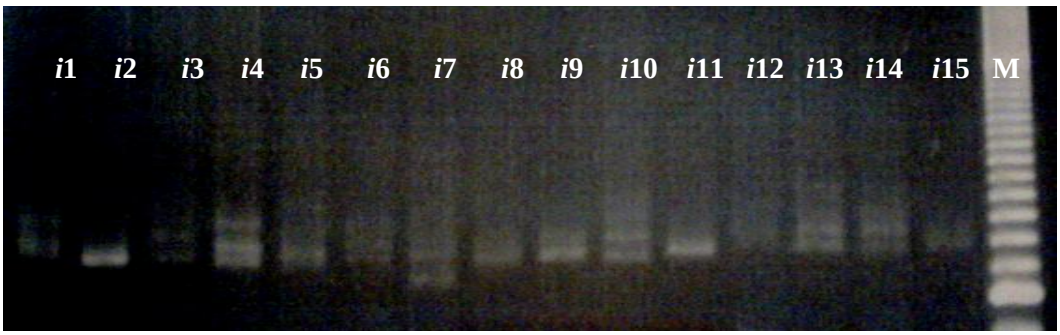

Figure 4.8: Microsatellite banding profile of *M. rosenbergii* (sample TRGN) using primer pair *Mbr*-10a obtained on 4% Metaphor agarose gel. Lane M: 20 bp DNA ladder.

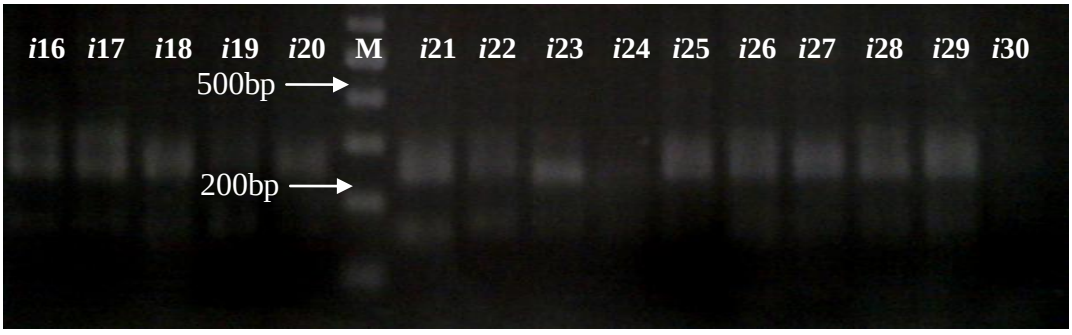

Figure 4.9: Microsatellite banding profile of *M. rosenbergii* (sample HatB) using primer pair *Mbr*-10b obtained on 1% Agarose gel. Lane M: 100 bp DNA ladder.

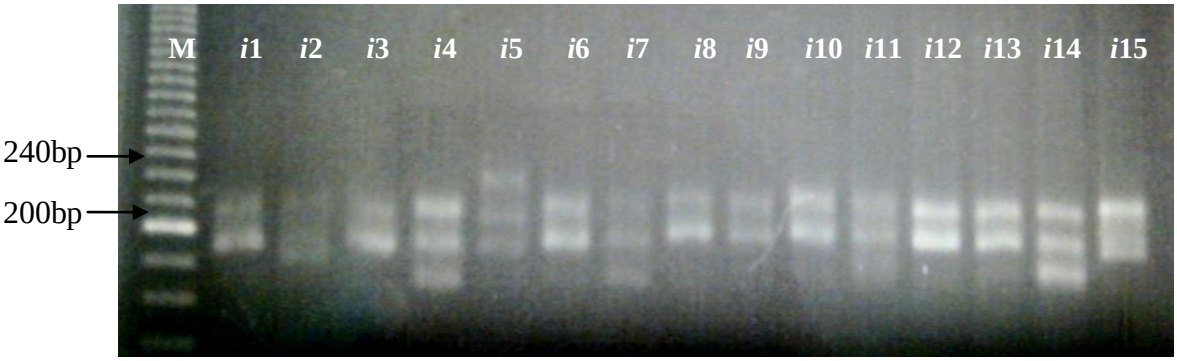

Figure 4.10: Microsatellite banding profile of *M. rosenbergii* (sample THAI) using primer pair *UVC*-807 obtained on 4% Metaphor agarose gel. Lane M: 20 bp DNA ladder.

Based on the gel electrophoresis, not all amplifications were successful. Some of the samples did not show any band on the gel. However, this does not mean the amplification failed because it might be due to technical error while loading the sample into the well. The sample was analyzed with ABI 3100 Genetic Analyzer. The result obtained from Genetic Analyzer showed that the amplification was successful with the present of peaks in the electropherogram even though the sample did not show any band on the gel (Fig. 4.11). Amplifications were repeated for samples that did not showed any peak after analyzed with ABI 3100 Genetic Analyzer and no band on the gel.

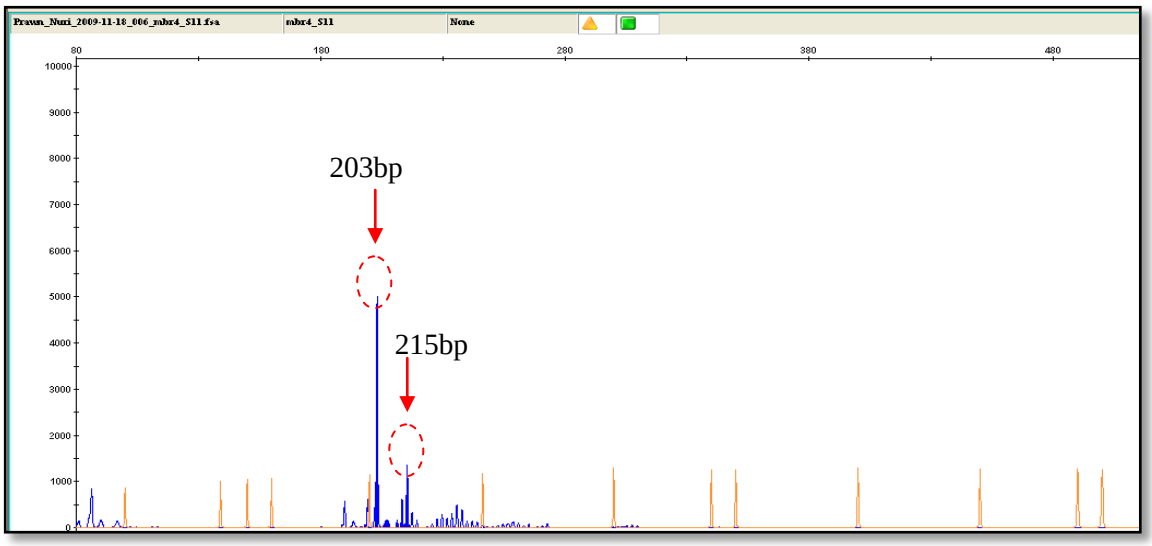

Figure 4.11: Electropherogram showing sample from SRWK (*i*11) successfully amplified with primer *Mbr*-4 even though Figure 4.4 shows there was no band obtained from agarose gel photo.

All PCR products in this study were analyzed using ABI 3100 Genetic Analyzer. This automated genetic analyzer was used instead of conventional scoring method because it provides several advantages such as easy and precise fragment sizing (Li *et al*., 2009b). Agarose gel photos show that some of the samples yielded multiple bands and smeared for example, Figure 4.1. This is difficult to determine the allele size especially when scoring was done using conventional method. However, the scoring process was not affected by the multiple bands and smearing when PCR products were analyzed using ABI 3100 Genetic Analyzer.

Figure 4.12 shows electropherogram obtained from the genetic analyzer of same sample (TRGN) and same primer (*Mbr*-1) as shown in Figure 4.1. In this case, scoring was done by seizing the value of the highest peak from the electropherogram. The highest peak value was chosen because it indicates the true allele for the sample. Other peaks that are lower and usually occur immediately before and after the real peak are known as stutter peaks (Gilder, 2004).

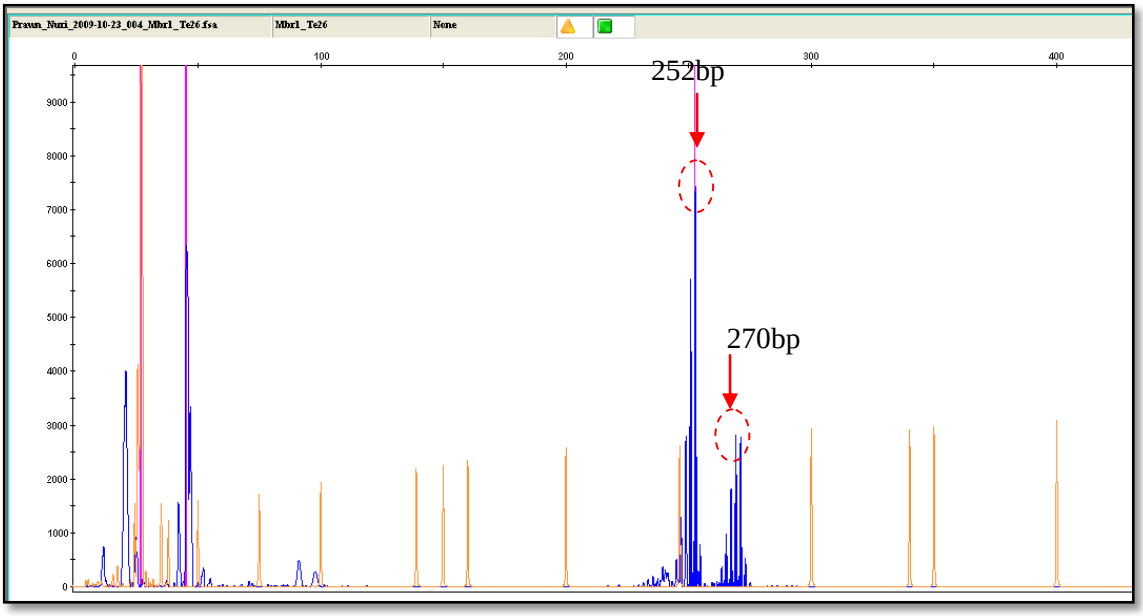

Figure 4.12: Electropherogram showing sample from TRGN (*i*26) amplified with primer *Mbr*-1 containing two alleles with the sizes of 252 bp and 270 bp.

Based on the gel images, it can be seen that a number of the samples showed very bright bands while others seemed to appear very faint. Despite of its advantages, there is a precaution that needs to be considered before the samples can be loaded into the genetic analyzer. It is very crucial to dilute the PCR products because it can affect the electrophoresis capillary effectiveness. The dilution can be made by referring to the gel images. The brighter the band is the more it should be diluted. In this study, a standard of 10X dilution was used. However, for faint bands no dilution was done prior to the loading.

Figure 4.13 shows an electropherogram obtained from the genetic analyzer. Generally, the electropherogram consists of a few peaks. The orange peaks are usually referring to the standard while the blue coloured peaks indicate the PCR product. The peak of PCR product can be in several colours such as red, green or yellow depending on the dye colour used to label the primer. In this study, 6FAM dye was used to label all primers thus the PCR products peak is seen in blue. The X axis of the electropherogram shows the allele size or the PCR product size while the Y axis denotes the height of the allele size.

Figure 4.14 is an example of electropherogram with the PCR product being sufficiently diluted. On the other hand, Figure 4.15 shows an electropherogram where the sample is not diluted sufficiently. From the electropherogram it can be seen that a pink line appears together with the blue line i.e. the PCR product's peak. The pink line indicates that the sample loaded into the capillary is concentrated. While doing this study, 107 or 5.13% of the samples failed to show their allele sizes or peaks when using the ABI 3100 Genetic Analyzer (Fig. 4.16). Therefore, these samples were scored manually and compared the size obtained with the peaks or alleles size seized by Genetic Analyzer of other samples within the population.

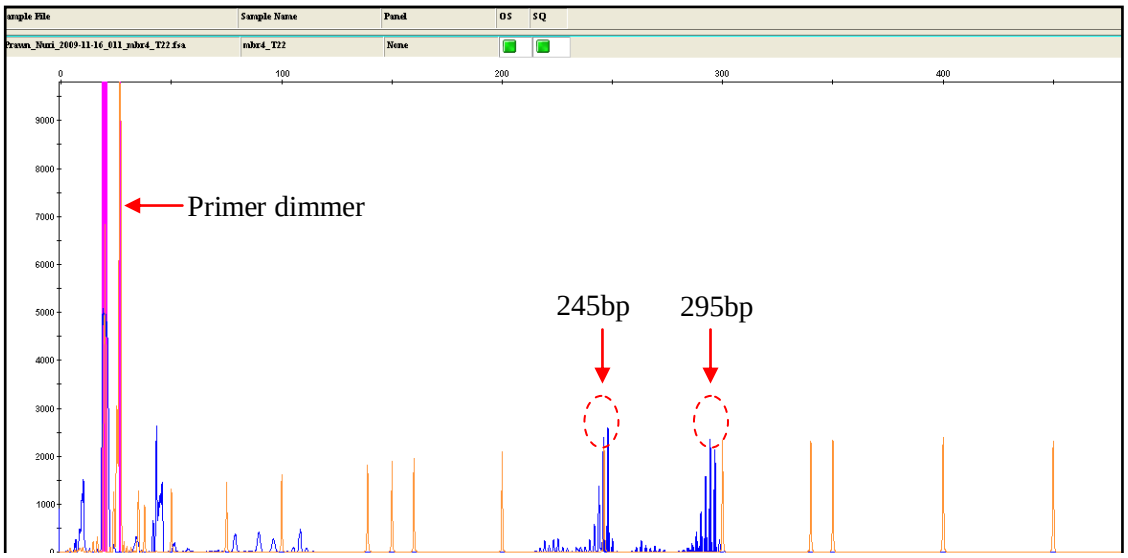

Figure 4.13: An example of electropherogram showing a result of a PCR product amplified with fluorescently-labeled primer that were loaded onto an ABI 3100 sequencer. The electropherogram shows sample from THAI (*i*22) amplified with primer *Mbr*-4 containing two alleles with the sizes of 245 bp and 295 bp.

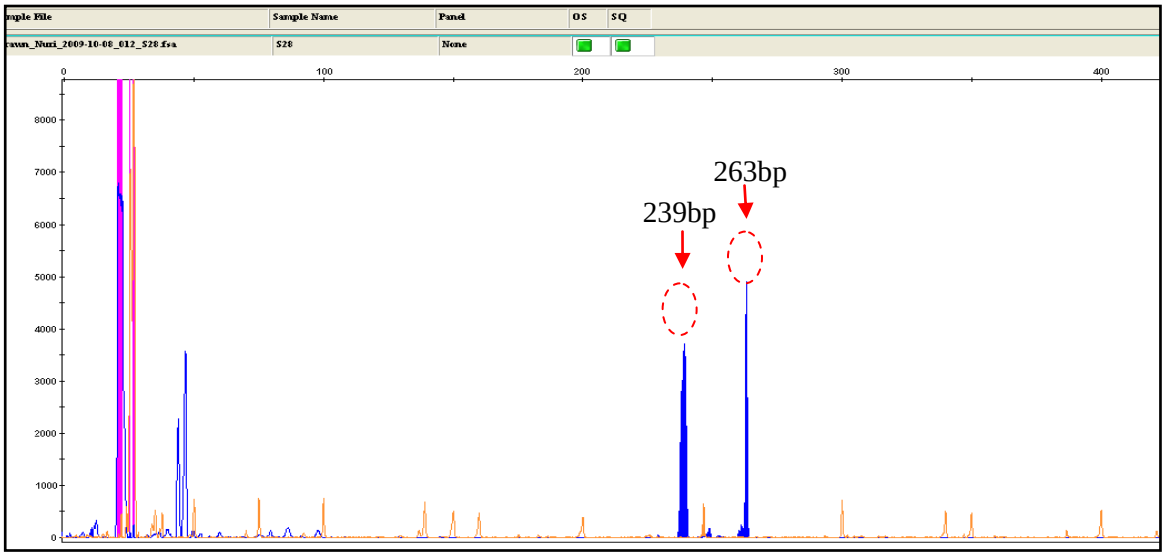

Figure 4.14: Electropherogram of amplified product of sample SRWK (*i*28) with primer *Mbr*-8.

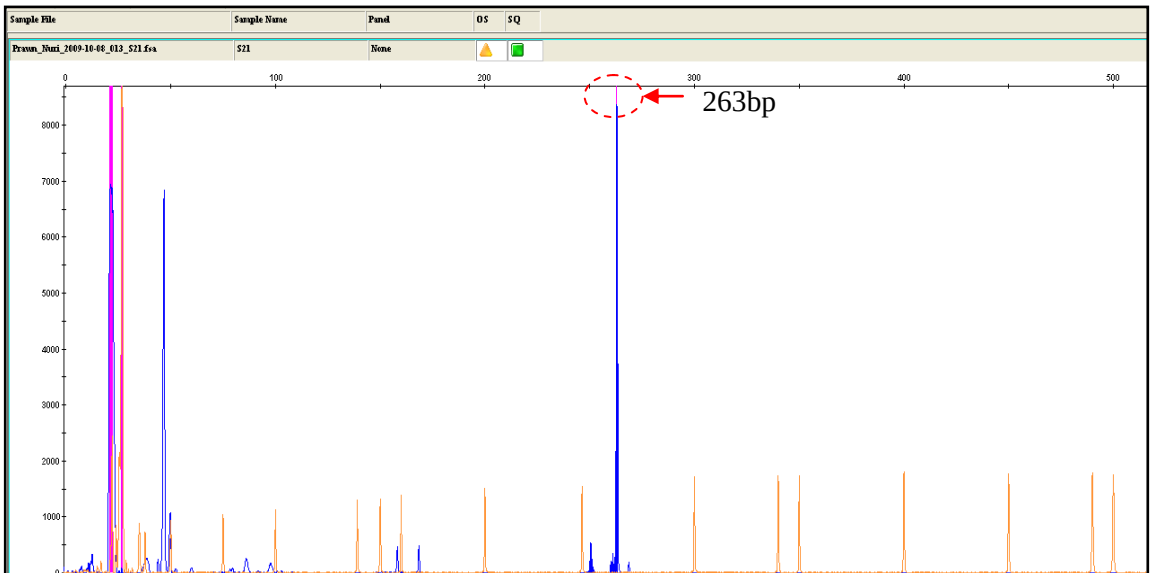

Figure 4.15: Electropherogram of amplified product of sample SRWK (*i*5) with primer *Mbr*-8.

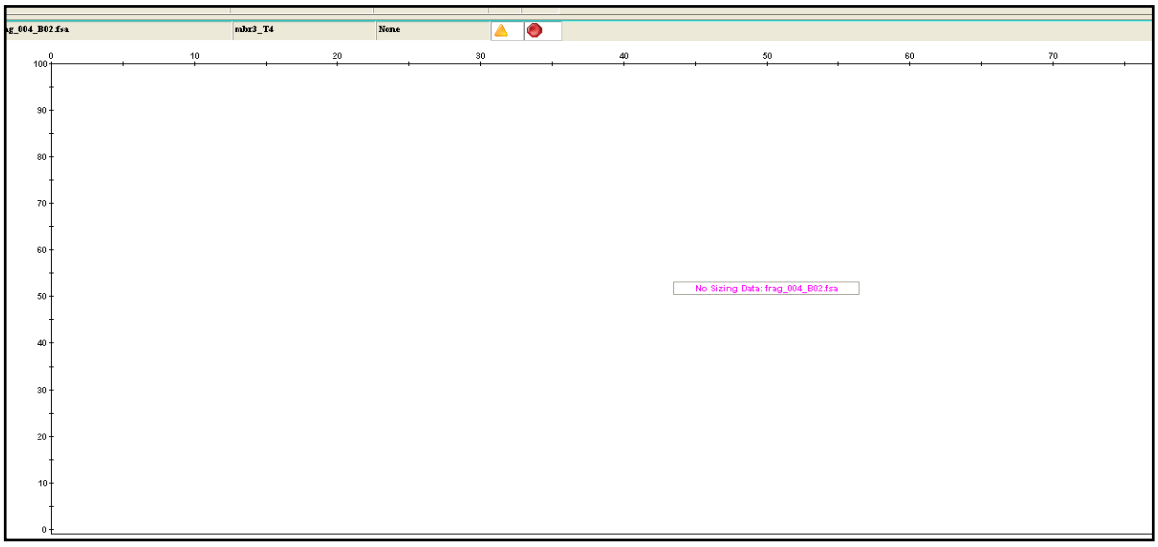

Figure 4.16: Electropherogram of amplified product of sample THAI (*i*4) with primer *Mbr*-3.

4.2 Null alleles

The MicroChecker which was employed to test for genotyping errors revealed no evidence for large-allele dropout or any stutter band scoring of any of the 10 loci.

However, six of the loci i.e. *Mbr*-1, *Mbr*-3, *Mbr*-4, *Mbr*-5, *Mbr*-10a and *UVC*-807 demonstrated the presence of null alleles.

4.3 Genetic variability within populations

All ten microsatellites loci were polymorphic with *Mbr-5* as the most variable locus with 15.6 alleles. This is followed by the primers *Mbr*-1 and *Mbr*-3, both with 14.5 alleles, *UVC*-817 (14.375), *UVC*-807 (14.25), *Mbr*-4 (12.25), *Mbr*-10a (9.75), *Mbr*-8 (9.25), *Mbr*-10b (7.25) and *Mbr*-7 (7.0). A total of 161 alleles were detected over all the ten loci. Table 4.1 demonstrates the number of alleles observed as well as the effective number of alleles for all populations. High levels of genetic variation of *M. rosenbergii* were exhibited in both domesticated and wild populations with the average number of alleles per locus varying from 9.0 (HatA) to 13.1 (KDH sample).

Observed heterozygosities were lower than the expected heterozygosities in all populations (Table 4.2 and 4.3). Observed heterozygosities across loci ranged between 0.58 (HatA) to 0.80 (NSBL) while the expected heterozygosities ranged between 0.76 (HatA) to 0.86 (NSBL). In domesticated populations, the mean observed heterozygosity for the ten loci was highest in NAPFRE with 0.71, followed by HatB (0.65), THAI (0.64) and HatA (0.58). NSBL showed the highest mean observed heterozygosity among wild populations with H_o=0.80. This is followed by TRGN (0.68), KDH (0.63) and SRWK (0.61).

Table 4.1: Summary of observed number of alleles and effective number of alleles in eight *M. rosenbergii* stocks.

		***Mbr*-1**	***Mbr*-3**	***Mbr*-4**	***Mbr*-5**	***Mbr*-7**	***Mbr*-8**	***Mbr*-10a**	***Mbr*-10b**	***UVC*-807**	***UVC*-817**	**Mean**
THAI	***A***	12.00	16.00	7.00	21.00	5.00	12.00	7.00	9.00	18.00	18.00	**12.50 ± 5.52**
	N_e	8.53	8.49	3.18	14.06	2.49	3.39	4.39	3.77	12.59	12.41	**7.33 ± 4.46**
NAPFRE	***A***	19.00	18.00	14.00	9.00	8.00	8.00	9.00	6.00	18.00	14.00	**12.30 ± 4.88**
	N_e	14.17	11.25	5.57	5.23	3.33	3.25	3.32	2.48	11.76	6.04	**6.64 ± 4.20**
HatA	***A***	11.00	4.00	12.00	12.00	11.00	7.00	8.00	5.00	11.00	9.00	**9.00 ± 2.91**
	N_e	7.20	2.60	6.12	6.59	3.78	4.23	3.19	2.20	4.57	5.96	**4.65 ± 1.75**
HatB	***A***	13.00	13.00	15.00	11.00	7.00	9.00	11.00	6.00	10.00	8.00	**10.30 ± 2.87**
	N_e	8.33	9.14	11.04	5.63	2.50	4.46	5.61	3.20	5.14	5.31	**6.03 ± 2.68**
SRWK	***A***	14.00	12.00	14.00	18.00	7.00	11.00	10.00	7.00	11.00	17.00	**12.10 ± 3.73**
	N_e	7.47	5.10	3.64	8.82	2.44	4.64	5.47	2.94	6.00	12.24	**5.88 ± 2.97**
TRGN	***A***	14.00	16.00	10.00	22.00	3.00	11.00	13.00	9.00	19.00	8.00	**12.50 ± 5.56**
	N_e	11.11	5.63	4.10	12.68	2.41	4.71	6.12	3.89	11.54	3.26	**6.54 ± 3.78**
NSBL	***A***	18.00	18.00	12.00	15.00	6.00	8.00	11.00	8.00	14.00	18.00	**12.80 ± 4.52**
	N_e	13.55	10.57	8.53	9.14	2.80	4.47	5.76	3.97	10.67	10.38	**7.98± 3.54**
KDH	***A***	15.00	15.00	14.00	17.00	9.00	8.00	9.00	8.00	13.00	23.00	**13.10 ± 4.79**
	N_e	8.73	8.66	8.17	8.00	3.41	5.38	3.69	3.69	8.79	16.94	**7.55± 4.01**

*note: THAI – Thailand, NAPFRE – *National Prawn Fry* Production and Research *Centre, HatA – Mun Aquaculture, HatB – Wong Aquaculture*, SRWK – Sarawak River,
TRGN – Terengganu River, *NSBL* – Timun River, *KDH – Kedah River*
A – Observed number of alleles; Ne – Effective number of alleles

Table 4.2: Summary of heterozygosity statistics for all 10 loci in the domesticated *M. rosenbergii* populations.

	THAI			**NAPFRE**			**HatA**			**HatB**		
	H_o	H_e	F_{is}	H_o	H_e	F_{is}	H_o	H_e	F_{is}	H_o	H_e	F_{is}
Mbr-1	0.6667	0.8977	0.26068	0.7333	0.9452	0.22714	0.4333	0.8757	0.50943	0.5667	0.8949	0.37077
Mbr-3	0.5333	0.8972	0.40967	0.7333	0.9266	0.21137	0.1667	0.6254	0.73684	0.7667	0.9056	0.15570
Mbr-4	0.3000	0.6972	0.57388	1.0000	0.8345	-0.20249	0.4333	0.8508	0.49498	0.6667	0.9249	0.28262
Mbr-5	0.7333	0.9446	0.22667	0.4000	0.8226	0.51801	0.7000	0.8627	0.19124	0.4333	0.8362	0.48603
Mbr-7	0.8667	0.6085	-0.43482	0.8667	0.7119	-0.22204	0.9333	0.7480	-0.25309	0.8000	0.6096	-0.31943
Mbr-8	0.7000	0.7169	0.02404	1.0000	0.7040	-0.443092	0.9667	0.7768	-0.25963	0.8333	0.7864	-0.05762
Mbr-10a	0.4667	0.7853	0.40988	0.6333	0.7107	0.11057	0.7667	0.6983	-0.09975	0.9667	0.8356	-0.16000
Mbr-10b	0.6667	0.7469	0.10906	0.6000	0.6068	0.01136	0.7000	0.5548	-0.26743	0.8000	0.6989	-0.14757
UVC-807	0.5333	0.9362	0.43449	0.5667	0.9305	0.39509	0.3667	0.7944	0.54265	0.2000	0.8192	0.75900
UVC-817	0.9000	0.9350	0.03808	0.5333	0.8486	0.37550	0.3667	0.8463	0.57095	0.5000	0.8254	0.39834
Mean	**0.6367 ±0.1829**	**0.8166 ±0.1207**	**0.205163**	**0.7067 ±0.1999**	**0.8041 ±0.1156**	**0.098142**	**0.5833 ±0.2672**	**0.7633 ±0.1081**	**0.216619**	**0.6533 ±0.2289**	**0.8137 ±0.0967**	**0.176784**

*note: THAI – Thailand, NAPFRE – *National Prawn Fry* Production and Research *Centre, Mun – Mun Aquaculture, Wong – Wong Aquaculture*
H_o= Observed heterozygosity; H_e= Expected heterozygosity; F_{is}: absolute value of inbreeding coefficient

Table 4.3: Summary of heterozygosity statistics for all 10 loci in the wild *M. rosenbergii* populations.

	SRWK			**TRGN**			**NSBL**			**KDH**		
	H_o*	H_e*	F_{is}	H_o*	H_e*	F_{is}	H_o*	H_e*	F_{is}	H_o*	H_e*	F_{is}
Mbr-1	0.2000	0.8808	0.77592	0.3000	0.9254	0.67956	0.7500	0.9459	0.21068	0.5000	0.9043	0.45238
Mbr-3	0.8333	0.8175	-0.01969	0.4667	0.8362	0.44611	0.8750	0.9246	0.05479	0.4167	0.9034	0.54410
Mbr-4	0.7667	0.7379	-0.03975	0.9333	0.7689	-0.21830	0.7500	0.9016	0.17117	0.6667	0.8963	0.26030
Mbr-5	0.6333	0.9017	0.30120	0.7000	0.9367	0.25596	0.7500	0.9096	0.17857	0.6250	0.8936	0.30514
Mbr-7	1.0000	0.6006	-0.68441	0.9000	0.5949	-0.52632	0.7500	0.6569	-0.14523	0.8333	0.7216	-0.15869
Mbr-8	0.5667	0.7977	0.29319	0.9000	0.8011	-0.12581	1.0000	0.7926	-0.26897	0.5000	0.8316	0.40389
Mbr-10a	0.8000	0.8311	0.03801	0.9333	0.8508	-0.09878	0.9583	0.8440	-0.13886	0.6667	0.7447	0.10680
Mbr-10b	0.6667	0.6712	0.00685	0.8000	0.7554	-0.06017	0.7083	0.7642	0.07456	0.7500	0.7447	-0.00730
UVC-807	0.1000	0.8475	0.88377	0.6667	0.9288	0.28571	0.5417	0.9255	0.41998	0.5833	0.9051	0.36048
UVC-817	0.5667	0.9339	0.39731	0.2000	0.7051	0.71981	0.9167	0.9229	0.00687	0.7917	0.9610	0.17934
Mean	**0.6133 ±0.2785**	**0.8020 ±0.1046**	**0.19524**	**0.6800 ±0.2709**	**0.8103 ±0.1095**	**0.135777**	**0.8000 ±0.1372**	**0.8588 ±0.0938**	**0.056356**	**0.6333 ±0.1358**	**0.8506 ±0.0845**	**0.244644**

*note: SRWK – Sarawak River, TRGN – Terengganu River, *NSBL* – Timun River, *KDH – Kedah River*
H_o= Observed heterozygosity; H_e= Expected heterozygosity; F_{is}: absolute value of inbreeding coefficient

Alleles frequencies for each locus in all populations were determined using the POPGENE programme. Allele frequency for the ten microsatellites varied among the different populations. Figure 4.17 illustrates the allele frequency of each population at loci *Mbr*-1 and *Mbr*-3. 12 alleles were detected in THAI, 19 in NAPFRE, 11 in HatA, 13 in HatB, 14 in SRWK, 14 in TRGN, 18 in NSBL and 15 in KDH at locus *Mbr*-1. From the figure, only one allele was identified to appear in all populations.

At locus *Mbr*-3, the highest total number of alleles were observed in NAPFRE and NSBL with 18 alleles. This is followed by THAI and TRGN (16), KDH (15), HatB (13), SRWK (12) and HatA (4). Eventhough samples from NAPFRE and NSBL shared the same total allele number at loci *Mbr*-3 only 11 of the alleles were detected at both samples. Similarly, in sample THAI and TRGN, only 6 out of 16 alleles were seen in both samples.

Allele frequencies of locus *Mbr*-4 and *Mbr*-5 are depicted in Figure 4.18. For locus *Mbr*-4, two alleles were discovered in all samples. One allele is observed in all domesticated populations but not in wild populations with the allele frequency between 0.0167 and 0.0500. At this locus, samples from HatB showed the highest total allele number i.e. 15 alleles while the lowest was obtained from the sample THAI i.e. 7 alleles.

There is no allele was detected to be found in all eight populations for locus *Mbr*-5. However, five alleles were recorded in seven of the populations. All the five alleles appeared in all wild populations but did not exist in at least one of the domesticated populations. At the same locus, it is discovered that four alleles were present in all populations but not in NAPFRE and HatB. The lowest allele number at this locus was 9 which was recorded in NAPFRE population whereas the highest was in population TRGN population with 22 alleles.

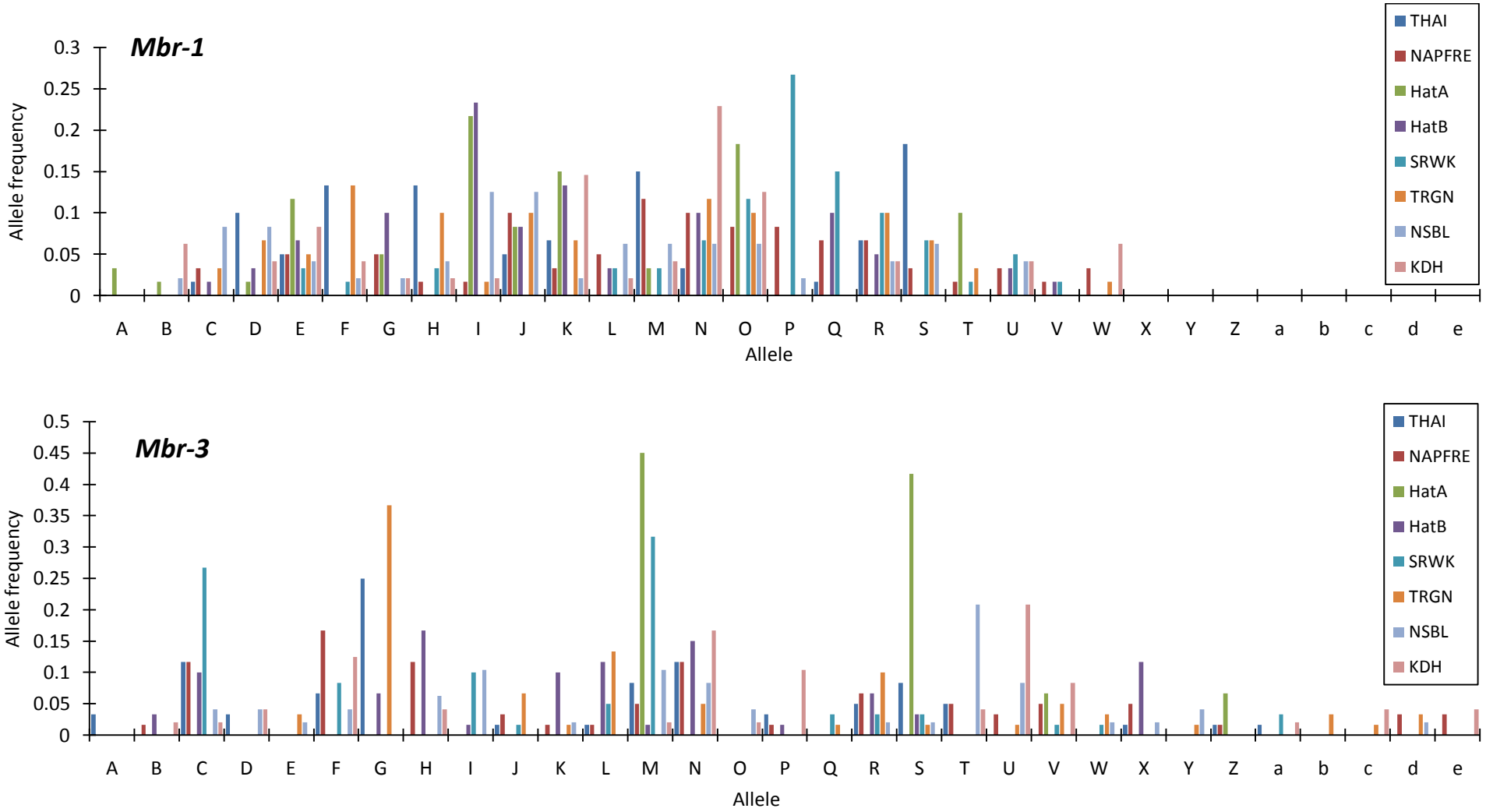

Figure 4.17: Allele frequencies at loci *Mbr*-1 and *Mbr*-3 for eight populations

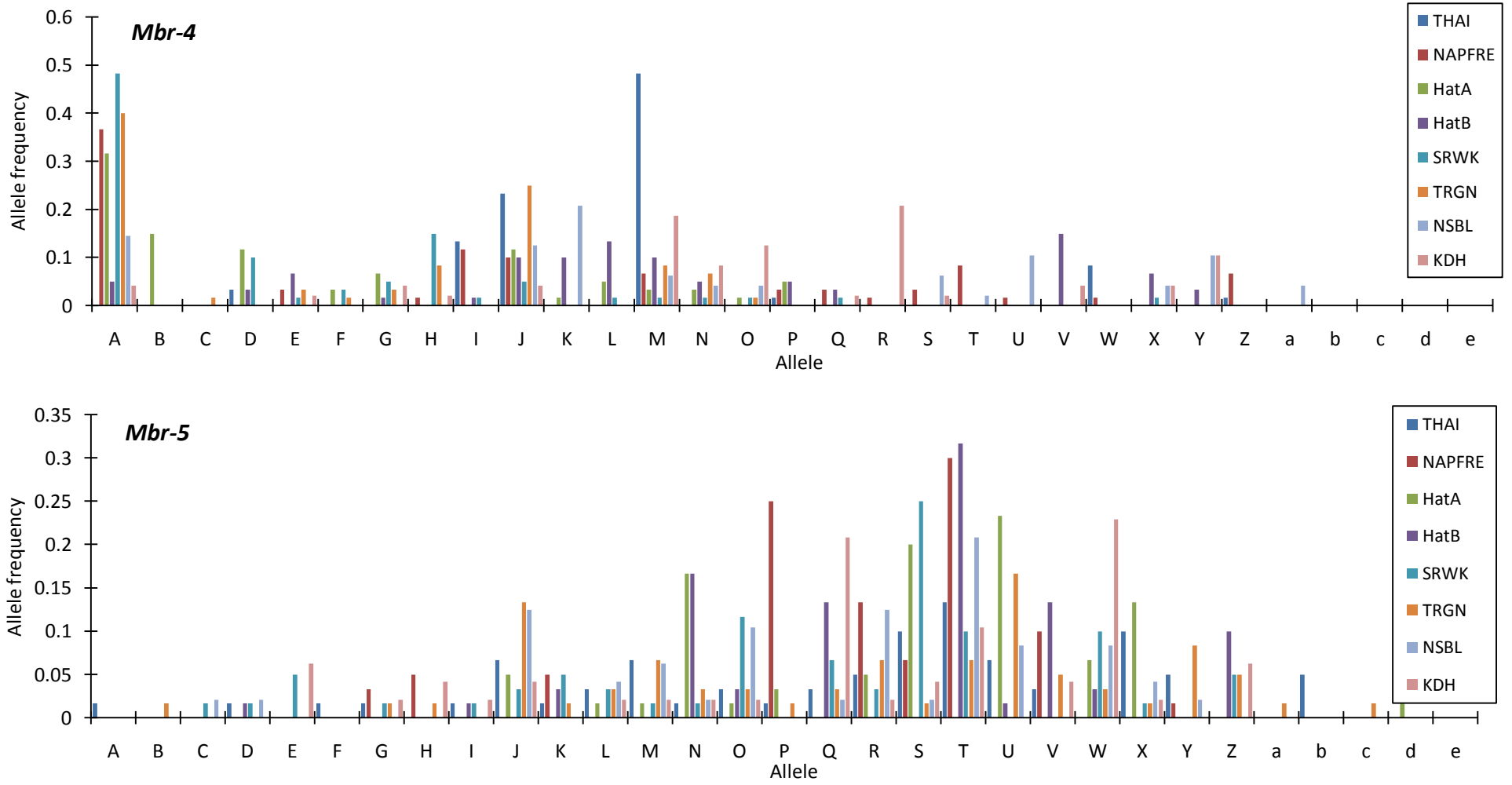

Figure 4.18: Allele frequencies at loci *Mbr*-4 and *Mbr*-5 for eight populations

At locus *Mbr*-7, an allele with high frequency (0.45-0.60) existed in all populations (Fig. 4.19). The lowest allele number (3) was observed in a wild population i.e. TRGN while the highest allele number (11) was detected in a domesticated population i.e. HatA. Two alleles were found to appear in all populations at *Mbr*-8. At the same locus, another two alleles were also detected at all populations except for HatA. THAI population showed the highest total allele number with 12 alleles followed by SRWK and TRGN (11) and HatB (9). NAPFRE, NSBL and KDH shared the same total allele number i.e. 8. HatA population has 7 alleles and is the lowest allele number compared to other populations.

Figure 4.20 showed the allele frequencies for *Mbr*-10a and *Mbr*-10b in all populations. The highest allele number at *Mbr*-10a was detected in TRGN population with 13 alleles. This is followed by HatB and NSBL (11), SRWK (10), NAPFRE and KDH (9), HatA (8) and THAI (7). Two alleles were present in all populations at this locus. At locus *Mbr*-10b, two alleles were found in all populations. A different allele was observed to display a high frequency of between 0.2708 and 0.5833 in all populations except NAPFRE. However, another allele was detected with the allele frequency 0.5833 in NAPFRE but showed lower frequency of 0.0167 and 0.05 in other seven populations.

Allele frequencies at locus *UVC*-807 and *UVC*-817 in all populations are demonstrated in Figure 4.21. Population TRGN exhibits the highest allele number i.e. 19 at locus *UVC*-807. The lowest allele number at the same locus is 10 which was recorded in population HatB. Allele frequencies for this locus range between 0.0167 and 0.4000. At *UVC*-817, an allele was found in all wild samples but not in any of the domesticated samples. The allele frequencies are between 0.0167 and 0.0667. Eight to 23 of the total allele number were recorded at this locus with the highest was identified in population KDH. Total allele number for other populations are 18 (NSBL), 17 (SRWK), 14 (NAPFRE), 10 (THAI), 9 (HatA) and 8 (HatB and TRGN).

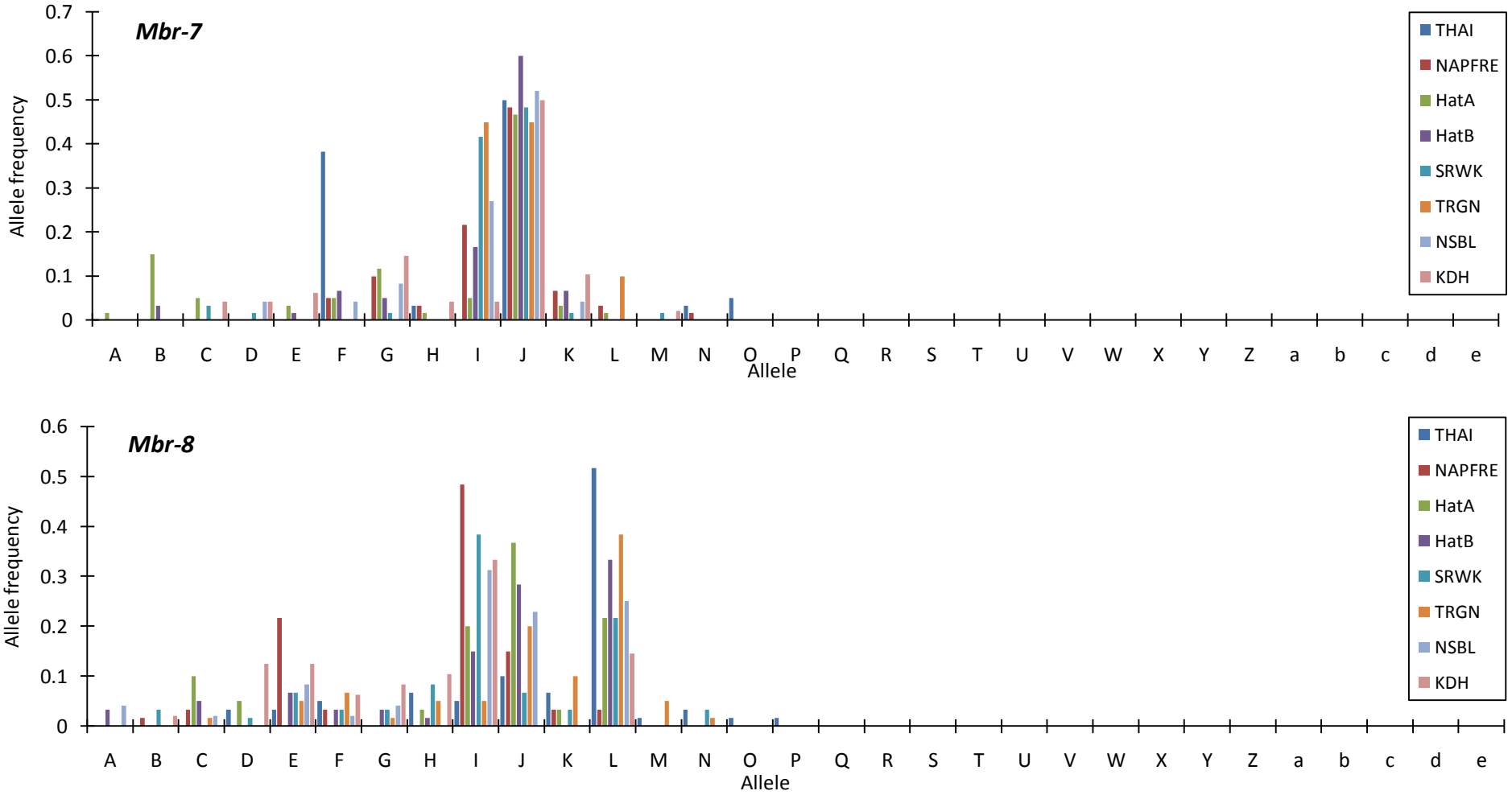

Figure 4.19: Allele frequencies at loci *Mbr*-7 and *Mbr*-8 for eight populations

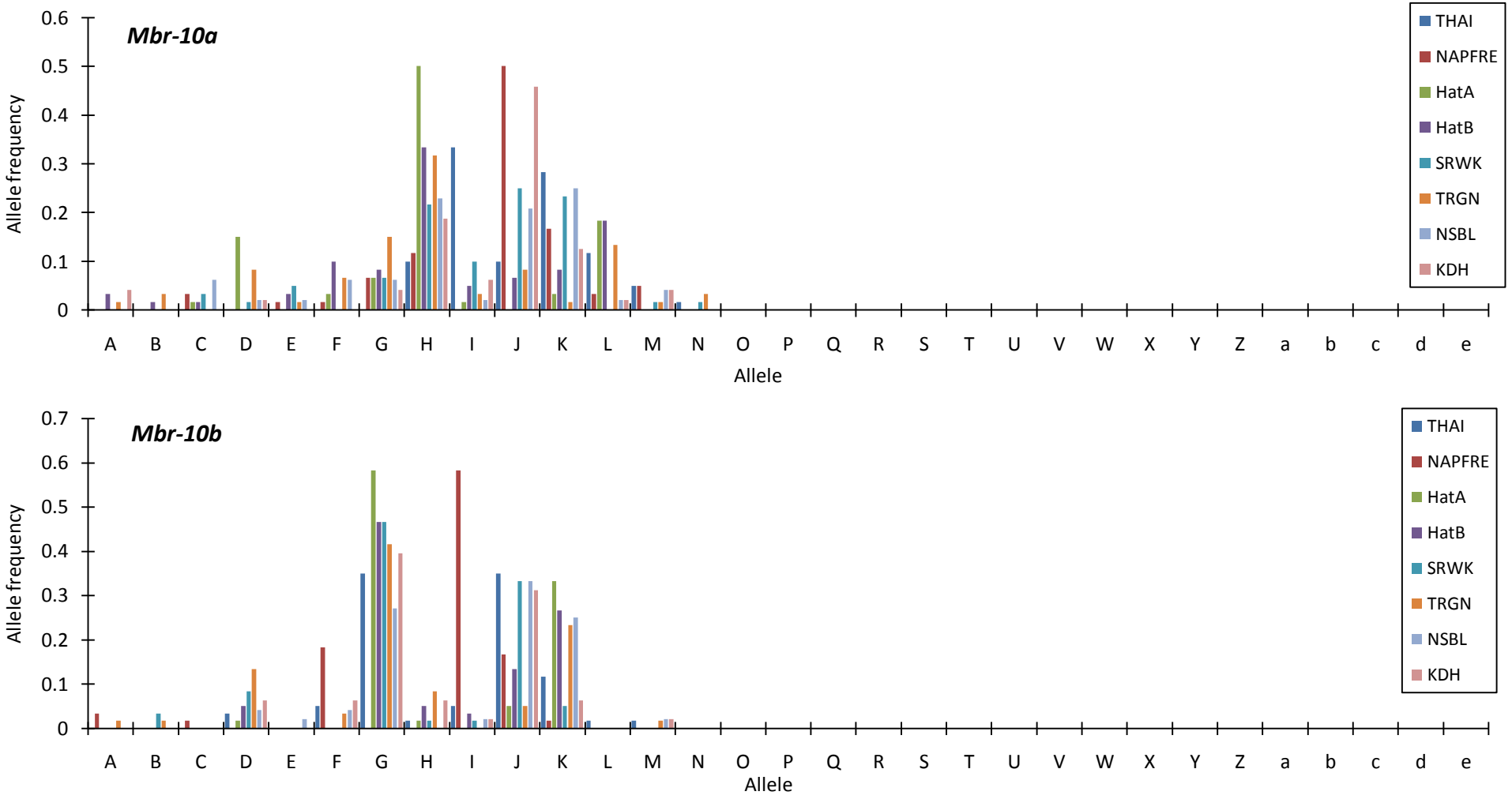

Figure 4.20: Allele frequencies at loci *Mbr*-10a and *Mbr*-10b for eight populations

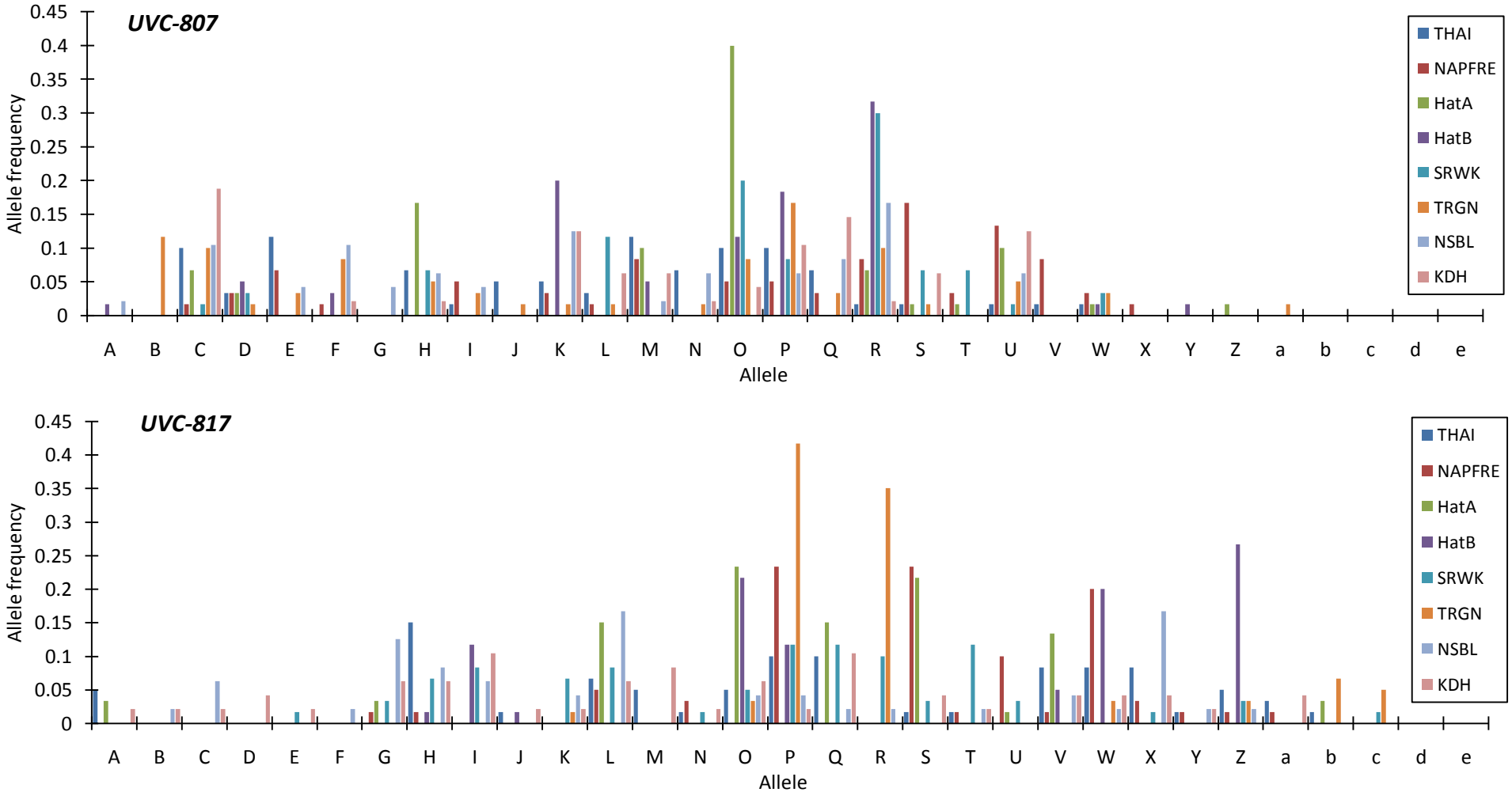

Figure 4.21: Allele frequencies at loci *UVC*-807 and *UVC*-817 for eight populations

4.4 Deviation from Hardy-Weinberg Equilibrium and linkage disequilibrium

Hardy-Weinberg equilibrium (HWE) was tested by an exact test of *P*-values using the Markov chain method. The results are demonstrated in Table 4.4. Of 80 tests (ten loci, eight populations), 20 were in the HWE (P>0.05). Another 60 (75%) revealed significant deviation from the HWE. However, after correction using false discovery rate procedure (FDR), only 59 tests were significant to deviation of HWE.

Mbr-1, *Mbr*-3, *Mbr*-5 and *UVC*-807 were observed to be significantly departed from the HWE in all domesticated and wild populations. At locus *UVC*-817, only one sample (THAI) was conformed to be in the HWE and the rest were departed from the HWE. Two samples were not deviated from the HWE at *Mbr*-4 whereas three samples at *Mbr*-10a. Departures from HWE were not observed in four samples at loci *Mbr*-7 and *Mbr*-8. At locus *Mbr*-10b six out of the eight samples were detected to be in HWE.

Table 4.4: Estimation of the exact *P*-Values for each population by the Markov chain method.

Locus	THAI	NAPFRE	HatA	HatB	SRWK	TRGN	NSBL	KDH
***Mbr*-1**	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
***Mbr*-3**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
***Mbr*-4**	0.0000	0.1259*	0.0000	0.0000	0.5940*	0.0125	0.0268	0.0000
***Mbr*-5**	0.0000	0.0000	0.0037	0.0000	0.0000	0.0010	0.0000	0.0000
***Mbr*-7**	0.0000	0.0060	0.1555*	0.6970*	0.0000	0.0000	0.5449*	0.1628*
***Mbr*-8**	0.5131*	0.0000	0.1574*	0.1730*	0.0002	0.0489*	0.0536*	0.0000
***Mbr*-10a**	0.0000	0.1357*	0.0264	0.0171	0.0003	0.2599*	0.0044	0.1417*
***Mbr*-10b**	0.5873*	0.8720*	0.2855*	0.0084	0.0729*	0.1370*	0.1909*	0.0232
***UVC*- 807**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
***UVC*-817**	0.1927*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0021	0.0000

note: THAI – Thailand, NAPFRE – *National Prawn Fry* Production and Research *Centre, HatA – Mun Aquaculture, HatB – Wong Aquaculture,* SRWK – Sarawak River, TRGN – Terengganu River, *NSBL* – Timun River, *KDH – Kedah River*
*Not significant at P>0.05 (adjustment by FDR)

Examination of linkage disequilibrium was conducted using Fisher's exact test by adopting the GENEPOP version 4.0. Across all populations, 26.19% or 11 out of 42 pairs of loci showed highly significant genotypic linkages. The linkages of other 31 loci were not significant. The results are shown in Table 4.5. Adjustment of the P-value was made using FDR. After the adjustment, 12 out of 42 pairs of loci were found to be highly significant genotypic linkages (Appendix C).

Table 4.5: Linkage disequilibrium analysis of 10 pairs of microsatellite loci before adjustment using FDR.

Locus pair			**Chi2**	**df**	**P-value**
Mbr-1	&	*Mbr-3*	∞	16	0.0000**
Mbr-1	&	*Mbr-4*	∞	16	0.0000**
Mbr-3	&	*Mbr-4*	31.628536	16	0.011173
Mbr-1	&	*Mbr-5*	17.834828	16	0.333657
Mbr-3	&	*Mbr-5*	19.179481	16	0.259466
Mbr-4	&	*Mbr-5*	∞	16	0.0000**
Mbr-1	&	*Mbr-7*	14.278936	16	0.577942
Mbr-3	&	*Mbr-7*	20.134852	16	0.214208
Mbr-4	&	*Mbr-7*	15.316789	16	0.501570
Mbr-5	&	*Mbr-7*	24.876292	16	0.072036
Mbr-1	&	*Mbr-8*	22.423613	16	0.130033
Mbr-3	&	*Mbr-8*	17.729926	16	0.339947
Mbr-4	&	*Mbr-8*	∞	16	0.0000**
Mbr-5	&	*Mbr-8*	11.731260	16	0.762263
Mbr-7	&	*Mbr-8*	21.507685	16	0.159811
Mbr-1	&	*Mbr-10a*	35.103621	16	0.003847
Mbr-3	&	*Mbr-10a*	∞	16	0.0000**
Mbr-4	&	*Mbr-10a*	38.489766	16	0.001288
Mbr-5	&	*Mbr-10a*	∞	16	0.0000**
Mbr-7	&	*Mbr-10a*	9.938060	16	0.869843
Mbr-8	&	*Mbr-10a*	21.603415	16	0.156465

Mbr-1	&	*Mbr-10b*	∞	16	0.0000**
Mbr-3	&	*Mbr-10b*	∞	16	0.0000**
Mbr-4	&	*Mbr-10b*	22.444879	16	0.129399
Mbr-5	&	*Mbr-10b*	23.672745	16	0.096900
Mbr-7	&	*Mbr-10b*	18.077047	16	0.319405
Mbr-8	&	*Mbr-10b*	33.220624	16	0.006908
Mbr-10a	&	*Mbr-10b*	∞	16	0.0000**
Mbr-1	&	*UVC-807*	23.658135	16	0.097242
Mbr-3	&	*UVC-807*	27.021099	16	0.041249
Mbr-4	&	*UVC-807*	15.982050	16	0.454214
Mbr-5	&	*UVC-807*	13.752325	16	0.617159
Mbr-7	&	*UVC-807*	7.275896	16	0.967570
Mbr-8	&	*UVC-807*	30.411232	16	0.015983
Mbr-10a	&	*UVC-807*	18.617496	16	0.288996
Mbr-10b	&	*UVC-807*	21.258318	16	0.168793
Mbr-1	&	*UVC-817*	33.423972	16	0.006491
Mbr-3	&	*UVC-817*	∞	16	0.0000**
Mbr-4	&	*UVC-817*	∞	16	0.0000**
Mbr-5	&	*UVC-817*	9.269611	16	0.901896
Mbr-7	&	*UVC-817*	6.716254	16	0.978350
Mbr-8	&	*UVC-817*	35.062198	16	0.003897

**Highly significant P < 0.01

4.5 Genetic differentiation among populations

Genetic distance estimations were computed by Nei (1978) using the GDA software. The results are given in Table 4.6. The greatest genetic distance was observed between HatA and NAPFRE (*D*=0.9333). This indicates that these two groups are the most diverse genetically. On the other hand, the shortest genetic distance was detected between KDH and NSBL (*D*=0.2087). Both are wild populations. Among the wild and domesticated populations, the genetic distance of NSBL was very close to HatB

(*D*=0.216). Among the domesticated samples, HatB and HatA had the shortest distance (*D*=0.3987) while the greatest were between HatA and NAPFRE (*D*=0.9333).

Table 4.6: Genetic distance (*D*) among eight populations.

	THAI	**NAPFRE**	**HatA**	**HatB**	**SRWK**	**TRGN**	**NSBL**	**KDH**
THAI	****							
NAPFRE	0.7705	****						
HatA	0.6413	0.9333	****					
HatB	0.4160	0.6237	0.3897	****				
SRWK	0.5364	0.4212	0.4114	0.4101	****			
TRGN	0.4705	0.7011	0.4815	0.3400	0.3597	****		
NSBL	0.3401	0.3842	0.4233	0.2160	0.2257	0.3514	****	
KDH	0.4243	0.4286	0.6338	0.3919	0.3585	0.6149	0.2087	****

note: THAI – Thailand, NAPFRE – *National Prawn Fry* Production and Research *Centre, HatA – Mun Aquaculture, HatB – Wong Aquaculture*, SRWK – Sarawak River, TRGN – Terengganu River, *NSBL* – Timun River, *KDH – Kedah River*
*Significant at $P < 0.05$

Analysis of AMOVA indicated that 7.65% of the variation was due to the differences among populations and 17.56% due to the differences among individuals within populations. The remaining 74.79% of the variation was due to the differences within individuals (Table 4.7).

Table 4.7: AMOVA results

Source of variation	Sum of squares	Variance components	Percentage of variation
Among populations	168.543	0.3377	7.65
Among individuals within populations	1067.725	0.7753	17.56
Within individuals	753.000	3.3026	74.79
Total	1989.268	4.4156	

Analysis of pair wise genetic differentiation between all populations using AMOVA approach showed that the F_{ST} values were between 0.0311 and 0.1438 with all being significant (P<0.05). This indicates there are genetic heterogeneities among all the populations studied. The highest F_{ST} value was between HatA and NAPFRE (0.1438). Populations between KDH and NSBL showed the lowest differentiation with F_{ST} value of 0.0311. Based on the result obtained (Table 4.8), the level of genetic differentiation in all populations were moderate.

Classifying individuals in a sample into populations are usually useful in population genetic studies (Pritchard *et al*., 2000). In this study, STRUCTURE version 2.2 was used to assign individuals to populations on the basis of their genotypes and at the same time estimate the population allele frequencies. The average proportion of membership of each sample in each cluster are presented in Table 4.9.

Table 4.8: Pairwise genetic differentiation (F_{ST}) among eight populations of *M. rosenbergii* based on ten microsatellites loci.

Location	**THAI**	**NAPFRE**	**HatA**	**HatB**	**SRWK**	**TRGN**	**NSBL**	**KDH**
THAI	****							
NAPFRE	0.1117	****						
HatA	0.1127	0.1438	****					
HatB	0.0716	0.0988	0.0807	****				
SRWK	0.0892	0.0778	0.0863	0.0741	****			
TRGN	0.0793	0.1074	0.0950	0.0625	0.0678	****		
NSBL	0.0542	0.0627	0.0793	0.0382	0.0421	0.0571	****	
KDH	0.0654	0.0692	0.1044	0.0623	0.0611	0.0867	0.0311	****

note: THAI – Thailand, NAPFRE – *National Prawn Fry* Production and Research *Centre, HatA – Mun Aquaculture, HatB – Wong Aquaculture,* SRWK – Sarawak River, TRGN – Terengganu River, *NSBL* – Timun River, *KDH – Kedah River*
*Significant at $P < 0.05$

Table 4.9: Average proportion of membership of each *M. rosenbergii* sample in each of eight genetic clusters identified by STRUCTURE version 2.2.

Samples	Clusters							
	1	2	3	4	5	6	7	8
THAI	0.030	0.030	0.016	0.045	0.059	0.015	0.779	0.026
NAPFRE	0.031	0.017	0.015	0.014	0.044	0.034	0.017	0.828
HatA	0.012	0.893	0.019	0.010	0.016	0.028	0.011	0.012
HatB	0.219	0.012	0.588	0.028	0.017	0.019	0.080	0.037
SRWK	0.069	0.026	0.016	0.098	0.072	0.641	0.033	0.046
TRGN	0.039	0.029	0.027	0.819	0.018	0.016	0.034	0.019
NSBL	0.517	0.030	0.081	0.090	0.172	0.048	0.038	0.025
KDH	0.165	0.021	0.057	0.019	0.567	0.033	0.107	0.032

Figure 4.22 summarize the estimations of Q i.e. the membership coefficients for each individual in each clusters. Individuals from this study are represented by the single vertical line broken into different coloured segments. These coloured segments correspond to the predefined populations.

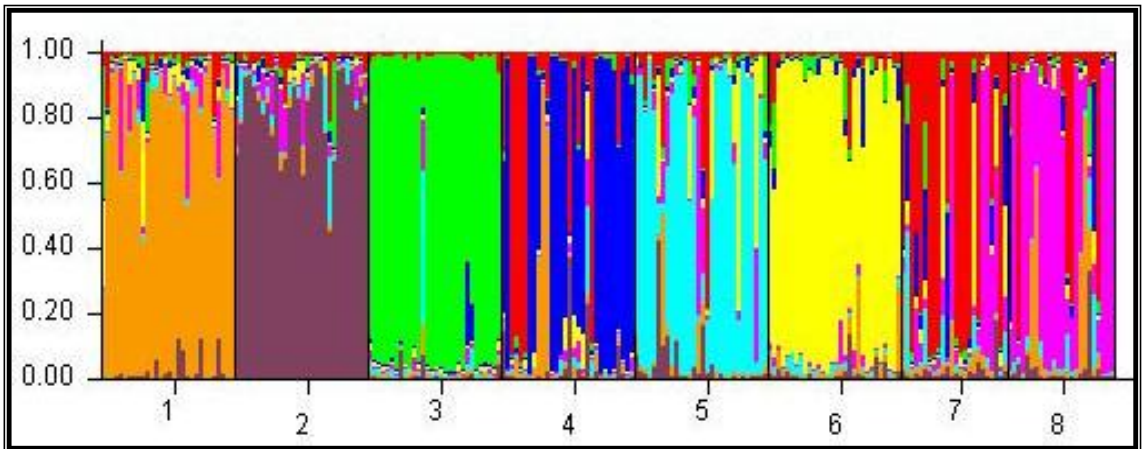

Figure 4.22: Proportional membership (Q) of each individual of *M. rosenbergii* in the eight clusters identified by STRUCTURE version 2.2. The numbers (1-8) stands for the predefined populations. (1-THAI; 2-NAPFRE; 3-*HatA; 4-HatB; 5*-SRWK; 6-TRGN; 7-*NSBL; 8-KDH)*

A consensus tree was constructed using the PHYLIP version 3.63 (Fig. 4.23). Values at the node represent the times that the particular node occurred in 1000 trees generated by bootstrapping the original allele frequencies. The UPGMA dendogram depicted a close clustering of KDH, HatB and NSBL where HatB clustered together with NSBL and KDH formed another sub-cluster to the first branch. SRWK and NAPFRE populations formed another cluster to this major branch which indicates their close genetic relationship. The TRGN and THAI populations formed the second independent branch. Another domesticated sample i.e. HatA formed the third discrete branch indicating that it is clearly separated from all the other populations. These results revealed that the genetic divergence between HatA and other populations is very high. However, the support is very low.

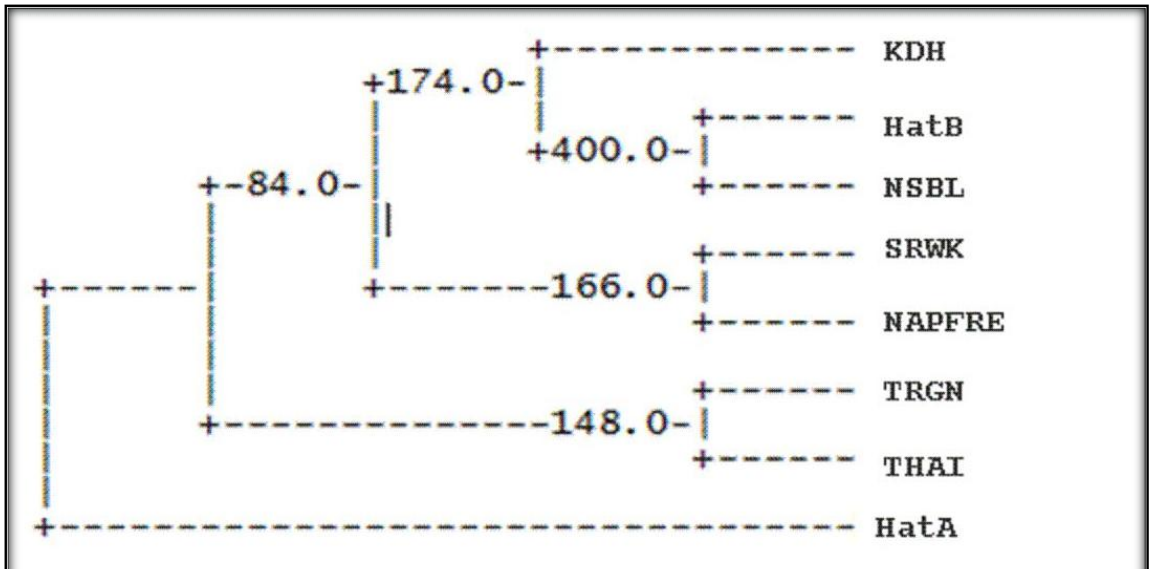

Figure 4.23: UPGMA dendogram illustrating relations between eight populations of *M. rosenbergii.*

CHAPTER 5

DISCUSSION

Macrobrachium rosenbergii can be found in almost all river systems in Malaysia. With the growing demand of this organism, culture practices are now being implemented. So far, domestications of this species are done in selected farms in Perak, Kedah and Negeri Sembilan. However, to date, there are no records on the status of domestication for freshwater prawns, even when this species have gone through a closed lifecycle since 1969.

A national survey was conducted on the distribution of the wild caught prawns and also the activity of possible movement of *M. rosenbergii* wild stocks from selected location into hatcheries and pond to grow vice versa. It is therefore important to know the effect of this gene flow and how the domestication in farms has affected the genetic diversity. This would give us a general idea on the stocks used in the genetic improvement programme.

Domesticated type can be obtained through a selective breeding programme or cross population breeding, where high genetic variation in the original stock is the basis. Nevertheless, hatchery populations are more prone to changes in their genetic make-up than the wild populations most probably due to their smaller size. Thus inbreeding and genetic drift which cause reduction in genetic variation is very common. Therefore, it is

essential to monitor any changes in the genetic variation and structure of the hatchery or domesticated population with respect to the wild populations.

The hactheries screened in this study are Mun Aquaculture (HatA) i.e. located in Pantai region and Wong Aquaculture (HatB) i.e. located in Kuala Kelawang region, Negeri Sembilan. One population was obtained from Kedah (NAPFRE) i.e. a research centre in Pulau Sayak (National Prawn Fry Production and Research Centre) and the other domesticated sample was obtained from an unknown hatchery in Southern Thailand (THAI).

In Malaysia, there are a number of rivers where wild stocks could be collected. However, in this study the samples used were collected from Sungai Serian, Sarawak (SRWK); Sungai Setiu, Terengganu (TRGN); Sungai Linggi, Negeri Sembilan (NSBL) and Sungai Muda, Kedah (KDH). The populations were selected mainly due to several reasons. NSBL was chosen because two of the domesticated populations are also situated in Negeri Sembilan. Similarly, the KDH population was chosen as the NAPFRE (one of the domesticated stocks) is also located in the same area i.e. Kedah. Therefore, it is likely to assume that the original broodstocks may have originated from the same region or at least the nearest location of the wild population.

Samples were collected from TRGN and SRWK mainly because there are no record of hatcheries in both areas and no history of any domestication process. Therefore, it is assumed that the populations were natural and can serve as the uncontaminated populations or control. In addition, the population SRWK was selected because of its geographical factor. Malaysia is divided into two regions i.e. West Malaysia and East Malaysia. Sarawak is the only population in East Malaysia that was analysed in this investigation. Thus, it is isolated from other populations that are being studied.

5.1 Microsatellite amplification and band scoring

The band patterns obtained by electrophoresis of the amplified products can be complicated. Oda *et al.* (1997) reported that the complexity of microsatellite analysis is not due to the heterogeneity of genomic templates but the intrinsic property of *Taq* polymerase, slippage and the terminal deoxynucleotidyl transferase activity of the enzyme. In this study, amplified DNA fragments for all loci were separated by the ABI 3100 Genetic Analyzer and the sizing conducted according to the Applied Biosystems GeneScan^{TM} 500LIZ standard size. Alleles were then visualized and scored by using the GeneMapper version 4.0, which is a fragment analysis software (Applied Biosystems).

According to MacAvoy *et al*. (2008) the sizing of alleles is critical to ensure a reproducible and consistent data in the microsatellite analysis. Hence, capillary electrophoresis system was chosen and used because the size standards and sample fragments were loaded in the same capillary and run under the same electrophoretic conditions providing accurate genotyping. Furthermore, this can reduce the cost and labour thus making the genotyping process simple and fast.

Getting accurate allele sizes during scoring can be problematic due to the appearance of shadow bands or stutter bands. Stutter bands usually occur in the dinucleotide repeat loci. This is due to the high rate mutation of this type of repeat loci which may gain or lose a single repeat unit throughout mutation. Hence, they usually exhibit an extensive allelic variation as well as high levels of heterozygosity. Thus, it is not unexpected if the dinucleotide repeat loci can provide more than 25 alleles and heterozygosities exceeding 90% (Wright and Bentzen, 1994).

Even though the dinucleotide loci provides a higher genetic variation, the interpretation can be complicated especially to samples that are from mixed populations

or when two alleles from a single individual are similar in size (Walsh *et al*., 1996). O'Connel and Wright (1997) suggested that stuttering could be avoided by using tetranucleotide loci or dinucleotide loci with a reduced product size (<120bp). Generally, the tetranucleotide loci are easier to score because of the greater distance between alleles and reduced stutter. On the other hand, dinucleotide loci with smaller sizes usually tend to stutter less and physically easier to separate during electrophoresis.

5.2 Null alleles

In this study, six out of ten loci showed the presence of null alleles. Null allelism is a common feature of microsatellite loci. In some cases, null allele can be amplified by changing the primer binding site. However, considering the large number of microsatellite loci that are needed in a mapping project, it is unnecessary and impractical to redesign the primers.

5.3 Genetic diversity of *M. rosenbergii* within populations

In this study, high levels of genetic diversity of *M. rosenbergii* were detected as evident by the high levels of allelic diversity and heterozygosity. A small number of allele generally is a signature of bottleneck. It may occur in both wild and domesticated populations. Generally, bottleneck may occur in a wild population due to the population isolation or dramatic reduction of effective size while, in the case of hatchery it is due to the small numbers of founders (Ha *et al*., 2009).

According to the present results, the average allele number per locus in wild population of *M. rosenbergii* was 12.1 (SRWK), 12.5 (TRGN), 12.8 (NSBL) and 13.1 (KDH). For hatchery population i.e. THAI, NAPFRE, HatA and HatB the average allele number per locus is 12.5, 12.3, 9.0 and 10.3 respectively. With the presence of high allelic number the populations are not experiencing bottleneck situation. However, the effective number of alleles at each locus is much lower compared to the number of the observed alleles. This could be due to the presence of null alleles present for certain primers.

The loss of alleles in domesticated populations compared to the alleles in wild populations was not obviously seen except for HatA and HatB. HatA (H_o=0.5833; A=9) has both the lowest observed heterozygosity and allele number. HatB (H_o=0.6533; A=10.3) revealed a lower allele number compared to other populations but high level of heterozygosity. Therefore, samples from HatB need to be studied further. Nevertheless, Chareontawee *et al.* (2007) suggested that this condition might occur due to the founder effect and genetic drifts which may eliminate rare alleles that are present in the original population. Founder effect is the loss of genetic variation in new generation due to small number of broodstocks while genetic drift is the changes of allele frequency in a population.

Chareontawee and her colleagues (2007) utilised four microsatellite primers that were also employed in this research to study the genetic diversity of hatchery stocks of *M. rosenbergii* in Thailand. The four primers used were *Mbr*-1, *Mbr*-5, *Mbr*-7 and *Mbr*-8. The allele number obtained from the study for each locus was 10, 27, 7 and 8 respectively. In comparison with Chareontawee *et al*. (2007), there are not many differences in the allele numbers for primers *Mbr*-7 (A=7) and *Mbr*-8 (A=9). However, the allele number for *Mbr*-1 (A=15.5) was higher in this study while for *Mbr*-5 (A=15.6) the allele number was lower when compared to the previous study. Furthermore, using

data from the same primers, it can be seen that wild stocks from Malaysia had higher allele numbers (SRWK and TRGN=12.5; NSBL=11.75; KDH=12.25) compared to wild samples in Thailand (Nat1=9.25; Nat2=10.25).

Heterozygosity is a general index for population diversity at the genetic level and is used as a measurement of the degree of genetic diversity (Yang *et al*., 2008). According to Thai *et al*. (2007), it is important to study the genetic diversity of natural and cultured populations since it provides necessary information of the genotypes for their adaptive response towards the changing conditions. In addition, Beardmore *et al*. (1997) claimed that heterozygous individuals are usually superior in several important characteristics such as growth, fertility and disease resistance compared to the less heterozygous individuals. Hence, it is important to study the genetic diversity of wild and cultured populations of *M. rosenbergii* due to economic reasons.

In domesticated populations i.e. THAI, NAPFRE, HatA, HatB the average of H_o was 0.6367, 0.7067, 0.5833 and 0.6533 respectively while for wild population the H_o was 0.6133 (SRWK), 0.6800 (TRGN), 0.8000 (NSBL) and 0.6333 (KDH). Despite that HatA samples displayed the lowest H_o compared to the other samples screened, no obvious differences in heterozygosity between the domesticated and wild populations as the domesticated stocks also showed relatively high heterozygosity levels.

The levels of heterozygosity from wild samples used in this study were compared with the levels of heterozygosity from Chareontawee *et al.* (2007) wild samples. H_o of wild stocks in Malaysia are 0.6000 (SRWK), 0.7000 (TRGN), 0.8125(NSBL) and 0.6146 (KDH) i.e. higher than the wild samples from Thailand (Nat1=0.5850; Nat2=0.5575). However, the differences is not significant ($P>0.05$). Even though it is not significant, the Malaysian wild stocks are still genetically diverse compared to the Thailand stocks because the genetic diversity of some natural populations in Thailand was contributed by the hatchery populations rather than pure

wild populations. This was a consequence of Thailand's restocking programme history (Chareontawee *et al*., 2007). On the other hand, there is no evidence of restocking programme using hatchery populations was done to Malaysia wild populations. Hence, it can be assumed that Malaysia wild stocks are still conserved.

In wild stocks high expected heterozygosity (*He*) usually indicates that the samples are good representatives in studying the population genetics. On the other hand, high *He* in domesticated stocks directly relates to effective population size. Hatchery populations are often accompanied by low genetic variation compared to the wild population. However, in some cases of hatchery population, the genetic variation retain high variability mainly because of the large number of breeding individuals or sufficient effective population size, *Ne*, short domestication history or the reintroduction of wild stocks into hatchery stocks (Ha *et al*., 2009).

The chances of inbreeding and drift in domesticated samples could be minimised by using large *Ne* in the hactheries, short domestication history and good management practices in the hatchery for example, replacing the poorly performing hatchery stocks with broodstocks obtained from the wild (Li *et al*., 2009a; Aung *et al*., 2010). High levels of heterozygosity in our Malaysian domesticated samples are mainly due to the reintroduction of wild broodstocks into hatcheries after each generation. This had been applied in HatA and HatB based on the information given by the owner of both hatcheries. As for NAPFRE, the high heterozygosity must be due to the selective breeding program that had been practiced at the research centre. The research centre is actively conducting a research on producing freshwater prawn fry through diallel crossing from selected populations. Based on personal communication with the Department of Fisheries, Malaysia, wild broodstocks from Sarawak were used as the founder to produce NAPFRE samples. However for the THAI population, no information on domestication was obtained.

5.4 Hardy-Weinberg equilibrium

Out of 80 tests, 59 significant deviations ($P < 0.05$) from the HWE were observed in this research after the adjustment using FDR procedure. Of that, 29 tests that departed from HWE were from domesticated populations while the remaining were from the wild populations. There are several factors that could have resulted to the departures from the HWE such as the small sample size, mixing of populations within samples, environmental pressures, non-random mating, inbreeding, null alleles and misscoring of microsatellite alleles (Divu *et al*., 2008; Yang *et al*., 2008; Chareontawee *et al*., 2007; Supunggul *et al*., 2000).

The departure from HWE in domesticated populations is usually common as they are exposed to one or a combination of the following forces: selection, genetic drift, migration, inbreeding and non-random mating (Na-Nakorn & Moeikum, 2009). Wild populations may also exhibit a deviation from the HWE. The deviation can be an indicator of possible impact of escaped culture from hatcheries (Xu *et al*., 2001). However, in this study the null alleles of the six primers are responsible for the deviation.

5.5 Genetic differentiation between populations

Genetic differentiations between wild and domesticated stocks have been observed widely in many freshwater and marine species using microsatellite analysis. For example, Li *et al*. (2004) compared the genetic diversity between wild and domesticated Pacific abalone. On the same year, Skaala *et al*. (2004) analysed the

genetic diversity of domesticated and wild Atlantic salmon. In addition, Liu and co-workers (2005) assessed the genetic structure of cultured Japanese flounder.

Pair wise F_{ST} analysis by AMOVA revealed that all populations in this study were significantly different from each other. F_{ST} analysis interpretation guidelines based on Wright (1978) have been summarized by Ong *et al.* (2009). According to Ong *et al.* (2009), there are four qualitative guidelines in interpreting the F_{ST} analysis of Wright (1978) i.e. 0-0.05 for low genetic differentiation, 0.05-0.15 for moderate genetic differentiation, 0.15-0.25 for large genetic differentiation and above 0.25 for extremely large genetic differentiation. In this research, the least genetic differentiation was observed between NSBL and KDH (0.2087). Population NSBL also showed small genetic differentiation with other two populations i.e. HatB and SRWK (0.0382 and 0.0421, respectively). Populations of NAPFRE and HatA showed the highest average pair-wise value among all with the value F_{ST} of 0.1438.

The population genetic distance calculated by Nei's unbiased measures shows that the shortest genetic distance is between the populations NSBL and KDH (0.2087) while the longest is between NAPFRE and HatA (0.9333). These results support the pair-wise F_{ST} analysis data which is discussed earlier. The closest genetic distance of the wild and domesticated samples was recorded between the population NSBL and HatB. This has been expected as the NSBL strains are known to be the broodstocks to the HatB strains.

STRUCTURE that uses a Markov chain Monte Carlo algorithm to cluster individuals into populations on the basis of multilocus genotype data was used to identify the cryptic population structure and detecting migrants or admixed individuals in this study. Pritchard *et al.* (2000) suggested that this method is useful for identifying populations and assigning individuals in situations where there is little information about the population structure. Generally, STRUCTURE assumes that all sampled

individual come from one or more of *K* unobserved populations (Falush *et al*., 2007). In this study, *K*=8 as its Ln P(D) value is nearest to zero i.e. -10092.5. Hence, it indicated that there were eight structured populations.

Figure 4.22 shows the proportional membership (Q) of each individual of *M. rosenbergii* in the eight clusters identified by STRUCTURE. From the figure, it is clearly seen that samples of HatB and NSBL are hybridized together. There are no records from where the hatchery recruits its broodstocks. However, it is believed that the broodstocks of HatB are from NSBL since the result revealed that they were connected to each other. Furthermore, these results supported their geographic location since both came from the same area (Longitude/latitude: NSBL=$2^0$28'29"N/ $102^0$02'05"E; HatB=$2^0$56'49"N/$102^0$05'07"E). It is also shown that minimal mixture occurred in SRWK and KDH population. Populations THAI, NAPFRE, HatA and TRGN seemed to be differentiated.

5.6 Genetic similarity and dissimilarity suggested by phoenetic relationship dendogram

Genetic constructions of a population can be affected mainly by geographic isolation, living environment, genetic bottleneck, gene flow and selection (Li *et al.*, 2007b). Clustering usually reflects the relationships between populations and supports the pair-wise F_{ST} analysis. Phylogenetic tree obtained in this study suggested that the populations HatB and NSBL have the nearest relationship, indicating the two populations that share the highest genetic identity. HatB is a domesticated population whereas NSBL is a wild population. Thus, this may reflect that the origin of HatB

stocks were from NSBL. In addition, both locations are located in the same state and near to each other.

The dendogram also showed that the samples used in this study do not cluster according to their geographical location. This might be due to the natural selection and mutation in order for the organism to adapt to new environment (See *et al.*, 2007). Results of the UPGMA dendogram showed significant population differentiation of HatA samples from the rest. This is not being expected as the HatA is located in the same area with HatB and NSBL i.e. the west side of Peninsular Malaysia (Longitude/latitude: NSBL=$2^0$28'29"N/$102^0$02'05"E; HatA=2°48'11" N/101°59'30" E; HatB=$2^0$56'49"N/ $102^0$05'07"E). However, KDH (northern side of Peninsular Malaysia) was clustered together with HatB and NSBL but into different subclusters.

Another interesting result revealed by the dendogram is that SRWK i.e. a wild sample from the East Malaysia (located at the Borneo Island) was clustered together with NAPFRE. NAPFRE is located in Kg. Pulau Sayak, Kuala Muda, Kedah (northern side of Peninsular Malaysia). It is a research centre that is actively conducting a research on producing freshwater prawn fry through diallel crossing from selected populations. Hence, samples from NAPFRE are not considered as wild. Based on the wide geographical separation between these two populations, it should not cluster together. The most likely explanation for this is that Peninsular Malaysia is separated from East Malaysia by the South China Sea. For this reason, the marine barrier is supposed to block the gene flow between these two sites. By personal communication with the Department of Fisheries, Malaysia, SRWK broodstocks were selected and brought to Kedah in 1980 to produce NAPFRE samples via the selective breeding programme.

The dendogram also indicated that *M. rosenbergii* from a hatchery in Thailand (THAI) clustered with TRGN, forming an independent cluster away from

other populations. Even though TRGN is from Malaysia, it is not clustering with other samples from Malaysia most probably because it is located at the eastern part of Peninsular Malaysia while others are in the western part. The East and West Coast of Peninsular Malaysia are divided by a mountain range called “Banjaran Titiwangsa”. The range covers from the north and ends in the south near Jelebu, Negeri Sembilan. Hence, this might be a barrier to the samples from both sides to be admixed.

A study on the genetic divergence of eastern and western *M. rosenbergii* by de Bruyn *et al*. (2004) revealed that samples from Thailand share the same 16S rRNA haplotypes with the samples from Peninsular Malaysia. This could be an explanation why both samples are clustering together. However, the close relationship between TRGN and THAI could not be discussed further since no information about the THAI samples including the exact location of the samples origin is available.

Nevertheless, Chareontawee *et al*. (2007) reported that several hatchery managers in Thailand have started to introduce non-native stocks from neighbouring countries i.e. Myanmar and India to replace their own stocks. These were done due to the issue of inbreeding and genetic deterioration of the local stocks raised by prawn farmers and scientists. Thus, it might be possible that the domesticated stock of THAI was introduced from TRGN. However, there is no documented record to confirm this assumption.

CHAPTER 6

CONCLUSION

Macrobrachium rosenbergii is one of the most widely distributed freshwater organisms. This study demonstrates the utilization of microsatellite DNA markers in order to meet the objectives i.e. to characterize the genetic diversity of wild and domesticated broodstocks of *Macrobrachium rosenbergii* and also to do a comparison of genetic makeup between the wild and domesticated stocks of the species.

The results presented here suggest that both the domesticated and wild stocks of *M. rosenbergii* are highly diverse in terms of the number of alleles and heterozygosity. Thus, this indicates that the lost of genetic variation did not occur in the studied domesticated stocks. This is due to the reintroduction of wild broodstocks into the hatchery and selective breeding program.

F_{ST} values indicate a highly significant value of differentiation in all populations. STRUCTURE analysis showed that some of the populations demonstrate hybridizations. However, the hybridizations are not obvious or it can be denoted as minimal mixing. Cluster analysis exhibits four independent clusters indicating that there are differentiations among populations which support the F_{ST} data. The distinct genetic differentiation suggested that each population should be managed differently (See *et al*., 2007).

Although all wild and domesticated populations exhibited similar and relatively high levels of genetic variation, it is always important to monitor the changes regularly as the genetic variation are changeable to several factors such as the effective population size and hatchery culture practices. In wild stocks for instance, overfishing and natural habitat changes can reduce the population size. As being reported by Li *et al*. (2009a), smaller population size can lead to inbreeding or bottleneck that may affect the genetic variation.

As for domesticated stocks, regular monitoring is essential because a survey by Chareontawee *et al*., 2007 revealed that most of the small and medium size hatcheries do not practice selective breeding nor maintain their own broodstock. Records on breeding and pedigree information of breeders should be kept to prevent the chances of inbreeding at the later stage of domestication process. It is also important if the hatchery stocks escape to the wild, the information is already available and conservation can be done.

Data obtained can serve as the preliminary information so that it can be used for selective breeding and wild stock management. Wild stock management is vital as it will be a challenge to maintain the high genetic variation of the stocks. For further study, it is suggested to screen on larger samples size so that the data obtained is more robust and the interpretation is more meaningful. Testing larger samples can also confirm on the hypothesis that size of samples can be a factor to cause the reduction of the total number of alleles per locus.