CHAPTER 5: DISCUSSION

In Malaysia, populations of *Channa striata* need to be conserved due to high demand, particularly in the medical field. In addition, people in China, Indochina, Thailand, Singapore, Indonesia, Philippines and India also believe in the biomedical properties of this particular fish. *Channa striata* have white boneless meat with tender taste. Therefore, this fish is both an edible fish and a natural remedy used in traditional medicine, particularly popular among post-operative patients to induce wound healing. Mat Jais *et al.* (2008) reported that these fish have anti-microbial, anti-inflammatory, cell proliferation, induction of platelet aggregation and anti-nociceptive properties of the mucus. Wild populations remain the dominant source of broodstock for some of the most important aquaculture species. Therefore management of the wild resource is vital for the aquaculture industry.

The loss of genetic diversity through inbreeding usually causes weak of adaptations, such as fitness and survival of the population. Therefore, measuring the genetic diversity in wild stocks is essential for the understanding and effective management of *Channa striata*. The aim of this study was to provide useful information on the genetic diversity available in natural populations of *Channa striata* from six locations in Malaysia by using microsatellite marker data to enable the determination of differences and similarities among populations and individuals.

Although microsatellites are continually to increase in popularity for various kinds of genetic studies, a major drawback is the difficulty and expense of de novo isolation of microsatellite for any new species. The conventional method for microsatellite isolation is the screening of genomic short fragment libraries with simple repeat oligonucleotide probes. This enrichment method has a high start-up cost, including material and equipment specific to the protocol but the number of microsatellites isolated was very low. However

this does not seem to be an obstacle for the applications of microsatellite markers because of their wide utility and resolving power in solving biological problems (Chambers and MacAvoy, 2000). In recent years, several alternative methods have been developed to isolate microsatellite markers which reduced the time and effort as well as the cost involved.

The Random Amplified Microsatellites (RAMs) or 5' anchored PCR procedure is an efficient method for isolating microsatellite as it avoids the need for genome. This effective method had been proposed based on some modifications of the Random Amplified Polymorphic DNA (RAPD) (Williams *et al*., 1990) method to get the amplification of unknown microsatellite by using repeat-anchored random primers (Wu *et al.*, 1994). The efficiency and reliability of this technique was confirmed by Usmani *et al.* (2001) who used this technique to isolate microsatellite markers in an aquaculture species *Mystus nemurus*.

The RAM technique offered several advantages. First, this is PCR-based and therefore required only very small amounts of DNA, second, in this technique, the DNA between the distal ends of two closely related microsatellites is amplified thus each PCR fragment could yield at least two microsatellite sequences. Also, by designing RAM with degenerate bases anchored to a certain classes motif, it is possible to produce a microsatellite locus with that motif for the species. If a particular microsatellite motif was abundant in a species of interest for a researcher, such as $(AC)_n$ then primers could be designed with degenerate anchored to the (AC) motif. Then primer pairs flanking the $(AC)_n$ microsatellite region could be isolated. Despite these advantages there is a deterrent with RAM primers. In some cases when the degenerate base sequence was not long enough, designing the primers can be difficult. However, this problem can be overcome by designing degenerate primers with longer degenerate anchored sequences.

The PCR amplification conditions needed to be optimized to amplify *Channa striata*'s microsatellite loci. The most important PCR condition to be optimized was the annealing temperatures. However, the amplification of two loci (SFO-BP8-H2 and SFO-VJ2-H38) out of 12 microsatellite primer pairs used did not produced distinct bands after many trials under different PCR conditions. It was difficult to identify and score them as their bands were not clear. Therefore these loci were ignored in the population study.

To determine the differences in the sizes of microsatellite alleles, high resolution gel electrophoresis is required. In this study, the detection of amplified PCR products from microsatellite markers was done on 4% metaphore agarose gel to separate the alleles of microsatellite markers.

Expressed Sequence Tag or ESTs are tiny portions of an entire gene that can be used to help identify unknown genes to map their positions within a genome. It provides researches with a fast and inexpensive way for discovering new genes since 3' ESTs are unique to a particular species and provide the additional feature of pointing directly to an expressed gene. The Blast results (Appendix A) showed that all the microsatellite loci that were isolated or characterized in this study have significant similarity towards ESTs in zebrafish, *Danio rerio* genome.

Microsatellite development still requires a skilled molecular biologist and a considerable investment of time and cost. Thus, using heterologous markers for cross-species amplification could be used for researchers to isolate new microsatellite loci in other species. The time needed and the cost of isolating microsatellite primer is also saved. However, optimization needed to be carried out to get clear separable bands. In this study a new PCR touchdown protocol was designed to reduce the non specific banding patterns and also to increase the reproducibility and specificity of the bands produced to be used as codominant markers. This process resulted amplified apparent microsatellite containing

fragments in *Channa striata* samples (Figure 4.18 to 4.29). These figures show single and double bands or heterozygous bands. The results demonstrated the conservation of microsatellite flanking regions between two aquatic phyla, *Cyprinus carpio* and *Channa striata*, both vertebrate which is supported the previous empirical finding that cross-species amplification works for closely related taxa.

The efficiency of the heterologous amplification observed between the two species in this study was in agreement to the observation discussed by Schlöttere *et al*. (1991) that microsatellite loci were conserved for aquatic species due to the aquatic environment that were less mutagenic. Although in this study the microsatellite loci could be amplified, there could be problems with their usage since these loci show not very clear bands. Nevertheless, this method is an efficient and low-cost approach to obtain new microsatellite markers for the species of interest.

The numbers and the effective numbers of alleles vary with the loci and populations, and also depended on the sample size analyzed. In this study, the highest mean of the observed and effective number of alleles was 2.10 and 1.66 respectively in the Negeri Sembilan population. Effective number of allele is defines as the number of alleles that can be present in a population. The number of allele tells about the number of allele that would be expected in a locus in each population. It is calculated by inverting the measure of homozygosity in a locus and the calculation may be affected by sample size. The effective number of allele may be informative in measuring diversity. It is for establishing collecting strategies. Collecting strategy needs revising if the figure obtained the second time is less than the first estimated number.

For a locus with several alleles with very different allelic frequencies, the alleles with frequencies far below the average allele frequency contribute very little to genetic diversity and are often of no advantage to the population. These alleles are not considered

in the estimation of effective number of alleles in a population as they are easily lost in due time (Waples, 1989). Therefore the effective number of alleles is a better estimate of genetic variability of a population when compared to the observed number of alleles.

In this study, the mean observed heterozygosity (*Ho*) was more than the expected heterozygosity (*He*) in all populations. Therefore, all populations indicated excess in heterozygosity as shown by the F_{IS} values. Negative values indicate excess in heterozygosity while positive values indicate heterozygosity deficiency (Hedrick, 2005). This value in all populations indicated high variability.

High levels of heterozygosity were observed for all *Channa striata* populations in this study. For excess of heterozygotes, in most cases the probable cause would be small sample sizes. Heterozygosity is one of the major factors that is commonly used to measure genetic diversity and loss (State and Pemberton, 2002). Genetic diversity within and among populations can promote adaptation to environmental changes, thus enabling a species to survive in a wide variety of situations (Carvalho and Hauser, 1998). The value of genetic diversity estimated for *Channa striata* populations provide important information to design future conservation and development programs in Malaysia.

It is important to test if a population is in Hardy-Weinberg equilibrium (HWE) or not at a particular microsatellite locus before undertaking genetic analysis of natural populations. It is performed by using the estimation of exact *P*-value by the Markov chain method. In the present study, most of the microsatellite loci tested was in HWE for most populations. Significant deviations from HWE can be associated with the presence of migration, mutation, non-random mating, selection, and small population size. Deviation from HWE may also indicate inbreeding or even problems in genotyping (Wigginton *et al*., 2005). Homozygote excess caused by the presence of null alleles, which fail to yield an amplification product (Weber and May, 1989), may also account for the deviation from

Hardy-Weinberg expectations. Nonamplifying or null alleles arise due to mutations in one of the priming sites (Callen *et al.*, 1993; Pemberton *et al.*, 1995). Population Negeri Sembilan and Penang show deviation from Hardy Weinberg equilibrium (HWE) because of two null alleles that is occur in loci MFW5 and MFW15.

In this study, locus MFW5 and MFW15 showed the evidence for the presence of a null allele (Table 4.13 and 4.14). Therefore, the possible reason for deviation from HWE for this locus is null allele. Besides, for these microsatellite loci, small sample sizes and sampling errors maybe the main reasons for deviations from HWE. Sampling error is always a factor that could not be avoided as one may have collected samples that were genetically related, thus leading to deficiencies of heterozygotes. MFW5 and MFW15 show homozygote excess in population Negeri Sembilan and Penang respectively. These populations are possibly in Hardy Weinberg equilibrium with loci MFW5 and MFW15 showing signs of a null allele. Other populations are probably in Hardy Weinberg equilibrium and no loci show evidence for a null allele. Loci that gives homozygote excess in some population that is loci SFO-T112-H6, SFO-T112-H6F, SFO-T112-H4, SFO-T112-H6I and MFW7 are maintain because it is not shown null allele in Microchecker analysis.

Linked genes are genes that are located on the same chromosome. If two alleles from different genes on the same chromosome tend to be associated in different individual at a greater frequency than that expected due to random association, linkage disequilibrium is said to exist between those two genes. Any observed deviation from expectation in such a 2 x 2 contingency table can be tested statistically using a chi-squared test or Fisher's exact test. Reduction in the expected frequencies of genotypes is due to linkage disequilibrium.

Linkage disequilibrium (LD) analysis should be performed between pairs of loci since the loci should be unlinked if they are to be used as independent tests of the

relationships among populations (Hedrick, 2005). When alleles at different loci are in LD, it means they tend to be inherited together because crossovers between them during meiosis are relatively rare. Linkage disequilibrium could also be caused by random drift or non-random mating, population subdivisions, interaction between genes, limited migrations and population bottlenecks (Ohta, 1982b; Ohta. 1982a, Kaeuffer *et al*., 2007). In this study, linkage disequilibrium was observed across all populations for several loci. According to the results there was no linkage disequilibrium between the same allele pairs in all the six populations of *Channa striata*. This could be related to the environmental versatility of these populations. Nsengimana and Baret (2004) proposed when there is LD between two or more alleles of a locus with more than one allele of another locus, this may be due to recombination between two linked loci, as in ideal populations, LD should only be influenced by genetic recombination. Thus, the results support the probability of recombination between SFO-T112-H6 and SFO-T112-H6I.

Genetic distances measures based on differences in allele frequency that would be expected to track microevolutionary changes. This genetic distance calculated using Nei's distance. The genetic distance showed that the Penang and Terengganu populations were the most related populations and this is not agree with the geographical origins of these two populations in Malaysia. The Selangor and Negeri Sembilan populations had the largest genetic distances. This could be explained by genetic drift and selection that could cause greater diversity than geographical distances (Kumar, 2003). This is also maybe because of small number of loci. This genetic distance can be realistic with the adding of more loci and samples.

In addition, according to de Bruyn *et al*. (2004), the Pleistocene drainage basins that linked sites on the Sunda Shelf that are today geographically isolated, may have caused gene flow among some populations of *Channa striata* in the past. Therefore, the geological

history would explain the occurrence of shared genotypes in population that are presently hundreds of kilometers apart. Previous work on *Channa striata* by Phon, M.K. (2008) using six microsatellites cross amplification primer in seventh populations of *Channa striata* i.e. Kedah, Terengganu, Pahang, Johor, Negeri Sembilan, Pulau Pinang and Selangor population found that only Pulau Pinang, Kedah and Negeri Sembilan populations were clustered according to their geographical area. Phon, M.K. (2008) also demonstrated that by using LP-RAPD method, almost all the clustering of the wild stock of *Channa striata* were not in accordance with their geographical area.

After removing null alleles, the consensus tree (UPGMA) (Figure 4.36) divided these six populations of the *Channa striata* into 2 major clusters. Populations Kedah, Johor, Negeri Sembilan, Selangor, and Terengganu belonged to one group while the Penang population clustered separately. The first major cluster subdivided into four groups. The populations from Kedah and Selangor belonged to a single group and the Johor, Negeri Sembilan, and Terengganu belonged to another group. Clustering between Negeri Sembilan and Kedah were 54.1%, clustering of Johor with Selangor and Kedah is 54.1%, while the highest clustering is between Negeri Sembilan, Selangor, Kedah, and Johor which is 55.4%.

The analysis of molecular variance (AMOVA) (Table 4.18) which estimated population differentiation showed that 45.8% of the total variation was among the six populations of *Channa striata* and 54.2% of the variation was within populations, related to variations among individuals in the populations. The Fixation index (Wright, 1978) can be applied to determine the level of genetic diversity within a population and is calculated on the basis of heterozygote deficiency or excess. F-statistics indicate the level of genetic differentiation within the entire population, between subpopulations and among individuals the accuracy of F_{ST} interpretation depends on factors like sample size, number of

populations or number of alleles etc. According to Balloux and Moulin (2001), F_{ST} estimates in the range 0-0.05 indicates little genetic differentiation among populations, the range 0.05-0.15 shows moderate genetic differentiation and the value of 0.15-0.25 indicates great genetic differentiation. Based on the population pairwise F_{ST} two populations, Johor and Terengganu (Table 4.20) and Selangor and Terengganu (Table 4.21), showed high differentiation, 0.57 and 0.47 respectively while the other pairwise differences were in the range of 0.33- 0.57, which showed high differentiation.

After removing the loci that show evidence of null allele (MFW5 and MFW15), the pairwise F_{ST} values were change. The pairwise F_{ST} values were lower, and the highest divergence is between Selangor and Terengganu with value 0.47159 (Table 4.21). These result support the geographical distance. The analysis of molecular variance (AMOVA) (Table 4.19) which estimated population differentiation also change and showed that 35.27% of the total variation was among the six populations of *Channa striata* and 64.73% of the variation was within populations. The differentiation within the population was significant.

With Structure (version 2.1; Pritchard *et al.*, 2000; Falush *et al.*, 2003; Pritchard and Wen, 2003) 10,000 replicates of the MCMC performed following a burn-in of 10,000 replicates. The admixture model is used and the allele frequency allowed correlating among subpopulations. This configuration, using the admixture model and correlate allele frequencies, has been consider best in the case of subtle population structure (Falush *et al.*, 2003). Empirical datasets likely will vary in sample size, number of loci, and the variability of loci, all of which may affect the performance of the software programs.

Evanno *et al.* (2005) investigated the impact of sample size (samples of individuals and samples of loci) on the performance of Structure and noticed a decrease in performance with smaller sample sizes. The results are sensitive to the type of genetic marker used, the

number of loci scored, the number of populations sampled, and the number of individuals typed in each sample.

Simulation summaries show that in wild populations low values of F_{ST} (<0.1) are commonly observed and can reflect high levels of gene flow between populations or recent divergence. However, in this study the value of F_{ST} are high although the samples from wild populations. Parameter set for K between one and six, for each of the six simulated datasets at each value of F_{ST}. K is the true number of clusters. Each dataset took approximately 5 minutes to run. Structure search for the most likely number of clusters and perform well at low levels of population differentiation.

When alpha is close to zero, most individuals are essentially from one population or another, while alpha > 1 means that most individuals are admixed (Falush *et al*., 2003). The model choice criterion implemented in structure to detect the true *K* is an estimate of the posterior probability of the data for a given *K*, Pr (*X*|*K*) (Pritchard *et al*. 2000). This value, called 'Ln P(D)' in structure output, is obtained by first computing the log likelihood of the data at each step of the MCMC. Then the average of these values computed and half their variance subtracted to the mean.

Before removing the null allele loci, the result obtain from running microsatellite data is that the cluster number (K) is 3 (Figure 4.37) while after removing the null alleles, the true number of cluster is 6 (Figure 4.38). This suggested that the population is cluster by six genetic differentiations. In addition, result from bar plot (Figure 4.38), there were sign of hybridization occur in all populations. This might be cause of geneflow, migration and the environmental adaptation may cause the differentiation. In order to elucidate genetic structure in detail, we can use more loci and therefore suggest using mitochondrial to infer whether there is unique haplotype (frequencies) for each and between populations. Furthermore, we can say that most of population, population structure within the west coast

of Malaysia comply one haplotype in accordance to Tan Min Pau unpublished result using mitochondrial CO1 gene (Tan M.P. *et al.*, 2009)