CHAPTER 2: LITERATURE REVIEW

2.1 *Channa striata*

2.1.1 Taxonomy

According to Integrated Taxonomic Information System (ITIS), the classification of "*Channa striata*" is as follows;

Kingdom	Animalia
Phylum	Chordata
Subphylum	Vertebrata
Superclass	Osteichthyes
Class	Actinopterygii
Subclass	Neopterygii
Infraclass	Teleostei
Superorder	Acanthopterygii
Order	Perciformes
Suborder	Channoidei
Family	Channidae
Genus	*Channa*
Species	*striata*

Thirty species of genus Channa have been identified; from which eight were reported to be found in Malaysia (Inger and Kong, 1962; Mohsin and Ambak, 1983; Wee, 1982). According to the Integrated Taxonomic Information System (ITIS), there are currently 27 species within the genus *Channa*. *Channa striata*, are known by different names, which are chevron snakehead, striped snakehead, and haruan.

2.1.2 Morphology

Figure 2.1: Morphology of *Channa striata* (Department of Fisheries Malaysia)

Family *Channidae* have some distinguish feature among the species. Their head surfaces have large plate-like scales. The mouths are large with sharp teeth. The fins are hyaline without spines. Dorsal and anal fins are long (Norainy, 2007).

2.1.3 Habitat and distribution

Snakeheads are native to Asia and Africa and have been introduced in the United States, Russia, and the islands of the South Pacific. Most snakeheads live in still or slow-running waters, usually hiding under plants, rocks, and sunken trees. Some large species live in large rivers, swamps, ponds, and reservoirs. Many *Channa striata* can live in the polluted water (Menon, 1999).

2.1.4 Economic and Medical Importance

Channa striata is widely consumed throughout the nation not just as food, but also as a remedy for wound healing. Mid-wives have constantly promoted *C. striata* for wound healing despite without realizing the actual scientific basis to the claimed. Thailand, Indonesia, Indochina, Philippines, China, and India had extensive cultures of this and it is among the most popular table fish even in seafood restaurants. Many believe that *C. striata*

contain all essential elements to bring about good health and helps to recover the lost energy after long illness (Mat Jais *et al.*, 1994, 1997, 1998).

Ng and Lim (1990) and Lee and Ng (1991) indicated that *Channa striata,* along with *C. micropeltes* and *C. lucius*, are utilized for medicinal purposes; particularly in Indonesia and Malaysia. Mention was made of use in a postnatal diet and during recuperation from illnesses or surgery (Lee and Ng, 1991).

Cream extracts of haruan tissues contain high levels of arachidonic acid, a precursor of prostaglandin, essential amino acids (particularly glycine), and polyunsaturated fatty acids necessary to promote prostaglandin synthesis. Treating wounds with these extracts has been demonstrated to promote synthesis of collagen fibers better than standard use of cetrimide, an antimicrobial quaternary ammonium compound, thus increasing tensile strength (Baie and Sheikh, 2000).

2.2 Genetic Diversity of *Channa striata*

2.2.1 Mitochondrial DNA

In a study by Abol-Munafi *et al.*, 2007, the authors analysed partial sequences of the mitochondrial gene for *Cytochrome b* on six channid species; *Channa marulioides, C. melasoma, C. lucius, C. gachua, C. micropeltes* and *C. striata* to determine the phylogenetics relationships amongst these species. From the study, they found that Channids are a taxonomically difficult group and have single common design but considerable morphological diversity. This diversity is linked to ecological factors in such a way that much homoplasy can be expected in morphological characters. They conclude that mitochondrial ribosomal genes of Channa shows high inter- and intraspecific sequence variability.

There are differences between nuclear DNA and mitochondrial DNA (mtDNA). Nuclear DNA has a smaller number of copies per cell than mitochondrial DNA and is inherited from both parents. Mitochondrial DNA is maternally inherited without recombination and, thus, is not unique to an individual. Mitochondrial DNA has a higher rate of substitution than nuclear DNA i.e. mutations when one nucleotide is replaced with another, making it easier to resolve differences between closely related individuals.

The mitochondrial genome is ideal for evolutionary and phylogenetic studies. The mitochondrial genes for 16S ribosomal RNA and *cytochromes b* have proved a powerful tool in phylogenetic studies, providing information on the systematics of terrestrial and marine vertebrates (Allard *et al.*, 1992; Milinkovitch *et al.*, 1993) and on the systematics of marine invertebrates (Cunningham *et al.*, 1992; Geller *et al.*, 1993; Rumbak *et al.*, 1994; Bucklin *et al.*, 1995; Canapa *et al.*, 1996).

Mitochondrial DNA has a number of applications in fisheries biology, management and aquaculture. In the past fifteen years mtDNA has attracted a lot of attention in many species, especially for population and evolutionary studies (Avise, 1994). It has become a very popular marker and dominated genetic studies designed to answer questions of phylogeny and population structure in fish for more than a decade. mtDNA studies particularly can contribute to identification of stocks and analysis of mixed fishery, provide information on hybridization and introgression between fishes, serve as a genetic marker in forensic analysis and provide critical information for use in conservation and rehabilitation programmes (Billington, 2003).

The presence of mitochondrial *pseudogenes* in the nuclear genome of a wide range of organisms is, for population studies, an unwanted reality (Zhang and Hewitt, 2003). The effectiveness of using mtDNA in population genetic studies has been greatly weakened by this fact. In addition mtDNA data on their own have some important limitations. Firstly, mtDNA represents only a single locus. So we can look only through a single window of evolution. This window reflects at best only the maternal lineal history (Skibinski, 1994; Magoulas, 1998), which could well differ from that overall of populations or species.

2.2.2 Microsatellites

2.2.2.1 Definition of microsatellite

Microsatellites are simple sequence tandem repeats (SSTRs). The repeat units are generally di-, tri- tetra- or pentanucleotides. They tend to occur in non-coding regions of the DNA. Microsatellites are known by a number of synonyms such as simple sequence (Tautz, 1989), short tandem repeats (STRs, Craig *et al*., 1988); and simple sequence repeats (SSRs) (Orti *et al*., 1997).

The microsatellite sequences were classified into 3 categories, perfect, imperfect, and compound (Weber, 1990). *Perfect* microsatellites are uninterrupted series of a repeat unit. *Imperfect* microsatellite sequences differ from perfect ones by the presence of one or three bases that arrest an alternating tandem repeat. *Compound* repeats consist of several different repeat types and are separated by less than three bases. A sequence is considered to be a microsatellite when the number of repeats is greater than ten for mononucleotide, six for dinucleotide, four for trinucleotide, and three for tetranucleotide (Stalling *et al.,* 1999; Estroup *et al.,* 1993).

Microsatellite markers are the most important population genetic tools. They are very useful in parent-offspring identification in mixed populations because of their high alleles per locus (>20). Their high polymorphism, and PCR based analysis has made them one of the most popular genetic markers (Wright and Bentzen 1994). Microsatellites are co-dominant, inherited in a Mendelian fashion and among the fastest evolving genetic markers, with 10^{-3}-10^{-4} mutations/generation (Goldstein *et al*., 1995).

2.2.2.2 Advantages of Microsatellites

Microsatellite markers have a number of advantages over other molecular markers and have gradually replaced allozymes and mtDNA. Microsatellite loci are typically short, thus makes it easy to amplify the loci using PCR, and the amplified products can subsequently be analyzed on either "manual" sequencing gels or automated sequencing.

Microsatellites are relatively easy to isolate compared with minisatellites; sample DNA can be isolated quickly because labour-intensive phenol-chloroform steps can generally be eliminated in favour of a simpler form of DNA extraction. Only small amounts of tissue are required for typing microsatellites and these markers can be assayed using non-lethal fin clips and archived scale samples, facilitating retrospective analyses and the study of depleted populations (McConnell *et al*., 1995). Moreover, there is potential for significant increases in the number of samples that can be genotyped in a day using automated fluorescent sequencers. For applications where a large number of loci are required, such as genome mapping or identification of Quantitative Trait Loci (QTL), microsatellites offer a powerful alternative to other marker systems.

2.2.2.3 Disadvantages of Microsatellites

Null alleles

One of the main problems is the presence of so-called "null alleles" (O'Reilly and Wright, 1995; Pemberton *et al*., 1995; Jarne and Lagoda, 1996). Null alleles occur when mutations take place in the primer binding regions of the microsatellite locus, i.e. not in the microsatellite DNA itself. The presence of null alleles at a locus causes severe problems, in particular in individual based analyses such as relatedness estimation and assignment tests, and most researchers prefer to discard loci exhibiting null alleles (Hansen, 2003).

Stutter bands

Another important disadvantage of microsatellite alleles is that amplification of an allele via PCR often generates a ladder of bands (1 or 2 bp apart) when resolved on the standard denaturing polyacrylamide gels. These accessory bands (also known as stutter or shadow bands) are thought to be due to slipped-strands impairing during PCR (Tautz, 1989) or incomplete denaturation of amplification products (O'Reilly and Wright, 1995). The practical outcome of PCR stutter is that it may cause problems scoring alleles. However, trinucleotide and tetranucleotide microsatellite typically exhibit little or no stuttering (O'Reilly *et al*., 1998).

2.2.2.4 Application of Microsatellites Markers in Fisheries Science

Nuclear DNA exhibits the greatest variability of all genetic markers related to fisheries science and will be a highly productive avenue for research and applications in wild and aquaculture stocks. Main applications in fisheries and aquaculture studies are phylogenetics and phylogeography (e.g. Hansen *et al*., 1999; Nielsen *et al*., 1999; Hansen, 2002; Hansen *et al*., 2002), population genetic structure (Scribner *et al*., 1996; O'Reilly *et al*., 1998; Nielsen *et al*., 1997; Shaklee and Bentzen 1998; Nielsen *et al*., 1999), conservation of biodiversity and effective population size (Reilly *et al*., 1999), hybridization and stocking impacts (Hansen *et al*., 2000; Hansen *et al*., 2001; Hansen, 2002; Ruzzante *et al*., 2001), inbreeding (Tessier *et al*., 1997), domestication, quantitative traits (Jackson *et al*., 1998), studies of kinship and behavioural patterns (Bekkevold *et al*., 2002).

2.3 Cross-species amplification in Common carp (*Cyprinus carpio)*

2.3.1 Taxonomy of *Cyprinus carpio*

According to Integrated Taxonomic Information System (ITIS), the classification of *Cyprinus carpio* is as follows:

Kingdom	Animalia
Phylum	Chordata
Subphylum	Vertebrata
Superclass	Osteichthyes
Class	Actinopterygii
Subclass	Neopterygii
Infraclass	Teleostei
Superorder	Ostariophysi
Order	Cypriniformes
Suborder	Cyprinidea
Family	Cyprinidae
Genus	*Cyprinus*
Species	*carpio*

Microsatellite loci of one species can be applied in other closely related species (Schlotter *et al.,* 1991; Zheng *et al*., 1995). Cross species amplification of microsatellite primers from a common carp, *Cyprinus carpio* had successfully demonstrated that microsatellite primer also can be amplified in silver carp, *Hypophthalmichthys molitrix* and bighead carp, *Aristichthys nobilis* (Tong J. *et al.,* 2002). In this study, we attempt to use these primers to amplify target region in *Channa striata* and observed if there is similarity or conserve region between *Channa striata* and *Cyprinus carpio*.

2.4 Population Genetic studies

2.4.1 Genetic variations

The general goals of population genetic studies are to characterize the extent of genetic variation within species and account for this variation (Weir, 1996). A frequency of genes and the forces that affect their frequencies, such as migration, mutation, selection and genetic drift can determine the amount of genetic variation within and between populations. The important factors in genetic variation such as selection, migration and genetic drift are affected by human activities.

For long term preservation of genetic diversity, large population sizes are required so that the loss of genetic variability due to genetic drift and selection and gain of genetic variability by mutation can be maintained in equilibrium (FAO, 1981).

2.4.2 Gene Flow

Gene flow among subpopulations is a characteristic attribute of population genetic studies. The similarity of populations' increases when there are high levels of migration and gene flow between populations (Neigel, 1997).

2.4.3 Heterozygosity (*H*)

A loss of heterozygosity due to inbreeding or isolation may result in a decrease of the potential adaptability of a population (Ferguson *et al*., 1995). Population heterozygosity (*H*), defined as the mean percentage of loci heterozygous per individual.

Estimation of **heterozygosity** for a locus **(H_{obs})** is formulated as the proportion of observed heterozygotes (h) at a given locus:

$$H_{obs} = h/N,$$

where N is the total individuals scored for that locus.

Heterozygosities can also be estimated from observed frequencies of alleles (rather than genotypes), assuming the population is in Hardy-Weinberg equilibrium. Thus;

$$H_{exp} = 1 - \sum_{i=1}^{k} P_i^2,$$

where Pi is the frequency of the i^{th} allele and k the number of alleles at a locus (Nei, 1978).

2.4.4 Allele frequency

Allele frequency or gene frequency is the relative frequency of a particular allele at a genetic locus in a population. It is usually expressed as a proportion or a percentage.

The formula is:

$$f(Ai) = n(Ai\ Ai) + ½\ n\ (Ai\ Aj)] / N$$

where:

f (Ai) = frequency of allele Ai

n (Ai Ai) = number of individuals homozygous for allele

n (Ai Aj) = number of individuals heterozygous for Ai and Aj , $i \neq j$

N = total number of individuals in sample

2.4.5 Genotype frequency

The genotype frequency is the proportion of a particular genotype relative to all genotypes at a specific locus in a population. It is calculated using the formula:

$$f\ (Ai\ Aj) = nij / N$$

where

f (Ai Aj) = frequency of a particular genotype Ai Aj

nij = number of individuals with genotype Ai Aj in the sample

N = total number of individuals in the sample

2.4.6 Percentage of polymorphic loci (*L*)

Other common measures of population variability for genetic marker data are the mean number of alleles per locus and the **percentage of polymorphic loci** (***L***). To avoid an expected positive correlation between *L* and sample size, a locus is usually considered polymorphic only if the frequency of the most common allele falls below an arbitrary cut-off, typically 0.99 or 0.95 (Avise, 1994). Polymorphism corresponds then to the proportion of polymorphic loci: $L = x/l$ x 100**,** where *x* is the number of polymorphic loci and *l* the total number of loci studied and is commonly expressed as a percentage.

2.4.7 Linkage Disequilibrium (*D*)

The linkage (or gametic) disequilibrium, *D*, measures the lack of fit of observed two-locus gametic frequencies to those anticipated based on the product of the single locus allelic frequencies. That is, the frequency of an *A1B2* gamete (loci *A* and *B*) in the population should be equal to frequency of the *A1* allele multiplied by the frequency of the *B2* allele (May and Kruger, 1990).

Furthermore, if the population at issue went through a bottleneck which reduced the effective population size (***Ne***) to a small number of breeding adults, it might be expected to see a significant linkage disequilibrium value for several generations (May and Kruger, 1990).

For a total population that is subdivided into many subpopulations, Wright (1951) defined three **F statistics** (correlation between uniting gametes), to relate the deviation from Hardy-Weinberg in the **total population** ($\boldsymbol{F_{IT}}$), to the genetic divergence **among subdivisions** ($\boldsymbol{F_{ST}}$) and to averaged deviation from Hardy-Weinberg **within subdivisions** ($\boldsymbol{F_{IS}}$) (Yang, 1998). $\boldsymbol{F_{IT}}$ values are seldom used since any type of departure from a single panmictic population will lead to a significant F_{IT} value. $\boldsymbol{F_{IS}}$ values help us to detect departures from Hardy-Weinberg by measuring the amount of heterozygote deficiency or excess observed in the sample. $\boldsymbol{F_{ST}}$ values help us to understand the degree of population differentiation within species.

2.4.8 Null Allele

A null allele can be defined as any allele at a microsatellite locus that is only weakly amplified or not visible after amplification and separation (O'Connell and Wright, 1997) and is recognized, together with population subdivision, as a major factor in depression of observed heterozygosity, as compared with expected on the basis of Hardy-Weinberg equilibrium. Even though the prime reason for null alleles is thought to be deletion and insertion within the priming site of microsatellite DNA, poor DNA preparation and/or mutation within the array may also be responsible (Allen *et al*., 1995; O'Connell and Wright, 1997).