Chapter 1

1.1 The Arowana (*Scleropages formosus*)

1.1.1 Description

The ornamental fish trade is currently a promising business sector in Malaysia. The Asian Arowana (*Scleropages formosus*), more popularly know as Dragon fish is one of the most expensive and sought after ornamental fish in the aquatic world.

In Chinese mythology, dragons have an elongated body, no wings and clawed feet, and look something like a hybrid between a snake and a crocodile. The Dragon fish does not have claw feet but it does have an elongated body which moves gracefully and sinuously as some serpents, and it does have an enormous mouth (as do crocodiles).

In addition the bodies of many dragons are covered in large scales, just as the Dragon fish. They have very large eyes and very importantly , the two long 'tendrils' that emerge from the area between the nostrils and the eyes in dragons which may have their counterparts in the two chin barbels possessed by the Dragon fish (Dawes, 1999).

The Chinese and Japanese believe the fish brings good luck, health and wealth to the owner and drives away evil. Trading has been lucrative as the fish is highly priced because of this myth.

1.1.2 Geographic distribution

The Dragon fish is a primitive freshwater fish from the Jurassic era. It belongs to the Order Osteoglossiformes. Some members of this Order have a characteristic 'bony' (toothed) tongue that distinguishes them as members of the Family Osteoglossidae. Other members of this family include the *Arapaima gigas* (arapaima). As mentioned earlier, the Dragon fish have two distinct mandibular barbels which puts it in the Sub-family, Osteoglossinae. This Sub-family is further divided into two genera, Osteoglossum and Scleropages. Genus Osteoglossum consists of two species: *Osteoglossum bicirrhosum* (silver arowana) and *Osteoglossum ferreirai* (black arowana) . Both of these species are found extensively in South America. The Scleroapages consists of three species: *S. jardini* (pearl arowana) , *S. leichardti* (spotted arowana) and the famous *S. formosus* (Asian arowana). The pearl arowana and spotted arowana are found in Australia and New Guinea, whereas the Asian arowana is distributed in various regions of South-East Asia.

Classification	Geographical Distribution
Osteoglossum bicirrhosum (silver arowana)	South America
Osteoglossum ferreirai (black arowana)	South America
Scleropages jardini (pearl arowana)	Northern Australia and New Guinea
Scleropages leichardti (spotted arowana)	Eastern Australia
Scleropages formosus *(asian arowana)*	South-East Asia

Table 1.1 - Distribution of the sub-family Osteoglossinae.

The present study however is focused on the Asian arowana. The Dragon fish is taxonomically classified as such :

Kingdom : Chordata

Phylum : Chordata

Sub-Phylum : Craniata

Class : Actinopterygii

Order : Osterglossiformes

Sub-order : Osteoglossoidae

Family : Osteoglossidae

Sub-family : Osteoglossinae

Genus : *Scleropages*

Species : *formosus*

Several types of Dragon fish with different colour patterns inhabit different regions of South-east Asia (Goh and Chua, 1999).

Colour variation	Geographic location
Yellow-tailed Malaysian Gold	Taiping and Bukit Merah in Perak.
Red-tailed Malaysian Gold	Pahang.
Indonesian Gold	Pekan Baru, Indonesia.
Red	West Kalimantan (Lake Sentarum and River Kapuas)
Green	Vietnam, Burma, Thailand, Malaysia.

Table 1.2 – Geographic location of the dragon fish in South-east Asia.

Due to their close similarities, it is believed that all arowanas shared a common ancestor at one time during the Pleistocene glacial age where, Borneo, Sumatra and Indochina were probably connected (Goh and Chua, 1999). As these landmasses began to split and drift apart, the different populations of the dragon fish became isolated from each other and thus were unable to interbreed. Factors such as differing environmental and habitat conditions, chance mutation and other influential elements have resulted in speciation (Dawes, 1999).

1.1.3 Breeding

The Dragon fish is found in habitat with still or slow-flowing waters that are often turbid or weedy like rivers and lakes. The Dragon fish is a surface dweller. Their diet consists mainly of insects and smaller fish.

The Dragon fish become mature during the third and the fourth year of their lives and usually begin to spawn from the fourth year onward (Lim *et al.*,1996). At this time they measure between 45-60 cm in length (Dawes, 1999). Dragon fish spawn throughout the year with the peak spawning period between July to December.

Sexes are indistinguishable even after maturity. Actual spawning is preceded by a protracted period of courtship and pair bonding that takes place over a period of two to three months.

Female Dragon fish possess a single ovary which when ripe, contains around 20-30 large ova approximately 1.9 cm in diameter, (Scott and Fuller, 1976). However, later studies and observation (Dawes, 1999) show that the fecundity can be higher. Similar reports have been produced by the Fishery Department, Bt. Berendam, Melaka. For example, in August 1988, a brood of fries harvested from a green dragon consisted of 96 individuals (Dawes, 1989). Lim *et al.* (1996) report that a more common brood size ranges from 4 to 62 for the red dragon.

The male incubates the eggs after fertilization for 5-6 weeks, during which period a very pronounced chin pouch is detected. The male appears not to feed during this period. The large yolk sacs take several weeks to be

fully absorbed by the fries. Feeding, however, starts before the sac is fully absorbed.

1.1.4 Trade of the Dragon fish

Due to its elegance and mythology, the Dragon fish has acquired a special status in Japan and some East Asian countries as a very popular but extremely expensive aquarium fish. A red dragon can easily fetch a price of US$1000. The great demand has led to over exploitation from natural habitat almost causing the species extinction. It has therefore been afforded the highest degree of protection by being listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) Appendix I protected fish (Joseph *et al.* 1986). Extinction of the Dragon fish, with a geological record from the Mezoic Era of Age of the reptiles (Greenwood *et al.* 1996) will mean the end of a unique genome that has been undergoing continuous evolution for at least one hundred million years.

Appendix I is the highest risk category and, consequently, the one with the strictest level of control with regard to trade. Animals and plants to which Appendix I applies " ... are endangered species and commercial import, export and sale is normally prohibited ...".

The 'normally prohibited' clause allows for some flexibility, so that limited trade can be permitted if for example, the specimens concerned:

- are bred in captivity or (in plants) are artificially propagated;
- are required for research purposes;
- are used for non-commercial purposes;

- date from before the Convention came into force (Morgan, 1995).

The basis for CITES to consider lifting the ban on the trading of Dragon fish appears to be whether the fish has been bred in captivity or more precisely according to CITES regulations, whether they are second generation captive breed individuals. The issue, however is not as simple as whether the Dragon fish has actually been bred repeatedly in captivity, as this has been witnessed and confirmed many times over the last ten years or more, and there is no doubt about the ability of some breeders to achieve this. It is more of a problem for the breeders to be able to verify their claims of captive breeding success for two consecutive generations. This means that breeders have to produce verifiable and certifiable evidence to convince CITES that their fish have been bred to the second generation.

Today, breeding of the Dragon fish is done very systematically. Brood stocks are separated from the first and second filial (F1 and F2) generations. Tiny electronic microchips containing a 12 digit code are embedded in the left dorsal muscle of all captive breed Dragon fish. Electronic tagging is a standard requirement by CITES for tagging all captive-bred Dragon fish that are traded. Officers from the Fisheries Department often carry out inspection to CITES registered farms to confirm the identity of the brooders and offspring. Only the F2 generation of the Dragon fish bred in captivity are allowed allow for trade. Each Dragon fish exported has a Certificate of Identity with a unique 27-digit barcode at the bottom of the left hand corner. The bar code is the sole reference for the particular fish described on the Certificate of Identity.

1.1.5 Conservation of the Dragon fish

Conservation in artificial environments is recognized as an important short term measure to protect a limited portion of the gene pool from immediate extinction (Maitland and Evans, 1986). The genetics of small populations must be considered in order to maximize survival and ensure adaptability (Meffe, 1986). It is appropriate then to apply DNA fingerprinting to captive-breed population to ensure that they do not suffer from loss of genetic heterozygosity. Heterozygosity can be evaluated and the survival of various groups compared to determine if heterozygous population are more successful – an approach taken by Dexter National Fish Hatchery (Johnson and Hubbs, 1989). Long term management can prevent the manifestation of growth rate and fecundity declines and an increase in deformed and abnormal fish (Tave, 1995). Thus, to maintain a species in perpetuity (Meffe, 1986), it is imperative to "know it" (Janzen, 1992).

Very little study has been carried out on this ancient fish. The genetic structures of the brood stock used in the commercial trade are unknown due to the lack of availability of genetic markers. Microsatellites are the ideal DNA marker for studying fish populations because they are usually highly polymorphic and easy to analyze (O'Connell and Wright, 1997). Microsatellites also provide undisputed and reproducible evidence of relatedness and stock identification and can be applied for long-term management of domesticated populations through pedigree construction and evaluation of heterozygosity (Fernando *et al.*, 1997).

1.2 Previous molecular studies on the arowana genome

Up till now only two studies have been carried out on the genetics of the Arowana. The first study was by Fernando *et al.* (1997). The aim of the study was to use the data to formulate a broodstock management program for long-term genetic conservation of the Asian arowana.

Different methods were used to obtain DNA fingerprints from the green, gold and red varieties grown in a Singapore fish farm. This was to determine which method was most suitable in providing information on genetic variability. Because DNA fingerprinting is a pattern made up of DNA fragments that are resolved by electrophoresis, so that each individual has its own unique 'fingerprint'. Thus, this study (Fernando *et al.*, 1997) suggests that genetic variability is best studied by developing DNA fingerprints. Several methods were used to develop DNA fingerprints.

Firstly, restriction fragment length polymorphisms (RFLPs) were obtained. Nine restriction endonucleases were used to cleave the Asian Arowana DNA into numerous fragments. Hybridization with four non-radioactively labeled probes produced numerous RFLP fingerprints, indicating numerous binding sites on the fish DNA with the probes used. No discernible differences in the RFLPs between the various varieties were observed. Because RFLPs specific to individual varieties could not be differentiated, RFLP fingerprints were not suitable for detection of genetic variation between the Asian Arowana varieties.

The second method employed was random amplified polymorphic DNA markers (RAPD). Seven primers were used singly in this study. DNA fragments were detected on a polyacrilamide gel as 'band' presence versus 'band' absence. After this, a genetic similarity index (SI) was calculated. The SI calculated for the three varieties of dragon fish indicated high level of genetic similarity within and between varieties. Although more informative DNA fingerprints were obtained by this method, the detection of genetic variability was low.

Lastly, DNA was screened for microsatellite loci. Many highly polymorphic microsatellites were detected simultaneously in the genome by hybridization to probes consisting of tandem repeats. This method provides undisputed and reproducible evidence of relatedness and stock identification and can be applied for long-term management of domesticated populations through pedigree construction and evaluation of heterozygosity.

The second study, was by Yue *et al.* (2000). This paper described the isolation of microsatellites from the green variety of the Asian arowana. A $(CA)_n$ – enriched library was created from arowana genomic DNA by digestion with *Rsa*I. Then, an adaptor produced by hybridizing 5'-phosphorylated 25-mer and 21-mer oligonucleotides (containing *Mlu*I restriction site) was ligated onto the blunt ends. The ligation products were then amplified using PCR. The PCR products were then hybridized to a biotinylated $(CA)_{10}$ probe. The products of this hybridization were captured using streptavidin-coated magnetic beads. Unhybridized genomic DNA was washed away and the captured fragments were used as template for PCR

amplification. The amplification products were digested with *Mlu*I and ligated into vectors. The ligation products were transformed into competent cells. The insert length was determined by colony PCR using M13-20 and M13 reverse primers. Approximately 85% (244/288) of the tested clones contained inserts between 250 and 1000 bp. These were purified and then sequenced on an automated sequencer.

Out of 45 sequenced clones, 33 (77.8%) contained CA repeats proving the efficiency of the approach. Twenty-eight out of 35 (80%) of these CA clones showed sufficient flanking sequences for primer design.

Twenty-eight microsatellite primers were tested in 25 green Asian Arowana individuals obtained from farms in Singapore and Indonesia (pedigrees unknown). Twenty-one out of 28 microsatellites showed specific products and polymorphism .

This study, like the previous one (Fernando *et al.* 1997) showed that microsatellites produce undisputed and reproducible evidence of relatedness.

1.3 Microsatellites

A microsatellite, also known as simple sequence repeats, comprise loci that exhibit variation in tandem repeat number. In microsatellites the tandem arrays are formed from short nucleotide motifs of mono-, di-, tri- and tetranucleotide repeats such as $(AT)_n$ or $(GT)_n$ that frequently extend over less than 300 bp but may be much larger (Tautz, 1989). Microsatellites generally appear to be non-coding and selectively neutral and are frequently

polymorphic as a result of allelic variation in the number of repeats in the tandem array (Queller *et al.*, 1993).

Microsatellites are dispersed at approximately 10 kbp intervals throughout the genome, embedded in unique DNA (Tautz, 1989). Microsatellite loci in the range of 100-300 bp size range, cloned from human, whale , *Drosophila* and mouse DNA, have been amplified by the polymerase chain reaction (PCR) using primers specific to unique flanking domains (Tautz, 1989; Lit and Luty, 1989; Weber and May, 1989). These loci show high levels of polymorphism.

Microsatellites are born from regions of 'cryptic simplicity', i.e. a region which is made up of a few intermixed motifs, which shows little sign of a tandem rearrangement (Tautz *et al.*, 1986). Evidence show that these regions evolve by insertion and deletion of single bases in a manner suggestive of slippage (Volger *et al.*, 1997). Studies have shown that microsatellite arrays tend to evolve faster in longer units than those with shorter units. An early study by Weber and Wong (1993) on human short tandem repeats showed that there was not a simple decrease in microsatellite mutation rate with length of repeating unit and studies later suggested that this was in fact the case (Sia *et al.*, 1997).

To avoid confusion, Chambers and MacAvoy (1999) suggest that microsatellites should have at least four dinucleotide repeats in an array before it can be called a microsatellite. Similarly, a region of cryptic

simplicity must contain less than three trinucleotide repeats, two tetranucleotide or greater nucleotide repeats. Some microsatellites would potentially consist of a mixture within the array, which will further change the level of polymorphism for that particular array. Thus, six classes of microsatellites have been proposed to define these categories. The term pure, compound and complex microsatellites are used to specify one, two or more types of repeat units respectively found within arrays at particular repeats; while interrupted (or imperfect) microsatellites are repeats with an additional hierarchical descriptor to indicate the repeats with one or more single non-repeated units occurring within the array (Chambers and MacAvoy, 1999).

Class	Sequence
Pure / Perfect	--$(CT)_{10}$--
Interrupted pure / perfect	--$TA(CT)_4TA(CT)_6$--
Compound	--$(CT)_{10}(GT)_{16}$--
Interrupted compound	--$(CT)_{10}$- -$ACAA(GT)_{16}$--
Complex	--$(TTTC)_4T_6$- -$(CT)_2$- -$(CYKY)_2$- -CTCC--
Interrupted complex	Some interruptions of nucleotide, which does not match to the repeat pattern of the complex microsatellite above that occurs within the array. -- $(TTTC)_4T_3AT_3$- -$(CT)_2$ - -G - -$(CYKY)_{11}$ - -CTCC--

*K = G or T / Y = T or C

Table 1.3 – Classification of the various types of microsatellites

1.4 Randomly Amplified Microsatellites (RAMS)

Ziestkiewicz *et al.* (1994) suggested a novel method for measuring genetic diversity in plants and animals based on microsatellites. Polymerase chain reaction (PCR) selectively amplifies DNA fragments flanked by sequences complementary to the oligonucleotide primers used (Saiki *et al.*, 1985). Normally two distinct flanking primers are designed on the basis of knowledge of the sequence at the locus of interest. However, if two similar microsatellite sequences are present within an amplifiable distance and in an inverted orientation, a single complementary oligonucleotide will prime amplification of the intervening DNA segment (Ziestkiewicz *et al.*, 1994).

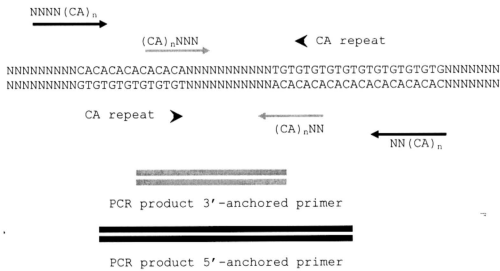

Figure 1.1 - A single primer targeting a $(CA)_n$ repeat, anchored either at the 3' (light arrows) or at the 5' end (dark arrows) of the repeat, is used to amplify genomic sequence flanked by two inversely oriented $(CA)_n$ elements.

Some PCR-based approaches, however, have employed primers containing microsatellite motifs at the 3' end and three or more non repetitive bases at the 5' end (Ziestkiewicz *et al.*, 1994; Hantula *et al.*,1996). The 5' bases are expected to 'anchor' the primer to the 5' end of a microsatellite. In practice this approach has proven difficult to optimize for complex genomes and there has been little evidence that variation in microsatellite repeat length at the primer binding sites has been preserved (Ziestkiewicz *et al.*, 1994). These difficulties arise because the PCR primers fail to anchor at the 5' end of microsatellites, instead slipping to the 3' end of the microsatellites during PCR. Each amplified fragment therefore contains exactly the same number of repeats found in the primers and any repeat length polymorphism is lost.

Fisher *et al.* (1996) described the development of a 5' anchoring procedure that consistently anchors PCR primers at the 5' ends of microsatellites, amplifying two close and inverted simple sequence repeats and the region between them. This technique offers a number of advantages compared with previously published protocols (Reddy *et al.*, 1992 ; Zietkiewicz *et al.*, 1996 ; Hantula *et al.*,1996 ; Wu *et al.*, 1994). Firstly, amplification of genomic DNA yields PCR products containing at least two microsatellites (one on each end of the amplicon) which retain their original repeat length. The multi-locus profiles produced from 5' anchored primers should be more polymorphic than those from non anchored primers. Secondly, this technique can be used to produce libraries which are highly

enriched for single locus microsatellites. Lastly, many of these individual loci can be amplified with only one locus-specific primer which further reduces the cost of microsatellite discovery.

In the study of *Pinus radiata* by Fisher *et al.* (1996), approximately 46 microsatellites, which ranged from 6 to 53 repeats, located at different loci were isolated from only 13 clones that had been sequenced. In addition common ash, sequence analysis of these clones revealed that 86% of them were unique and carried the expected microsatellite patterns at both ends (Brachet *et al.*, 1998). Meanwhile, Fisher *et al.* (1996) compared the number of dinucleotide repeats for the microsatellites discovered from an un-enriched genomic library and RAMS had longer units of repeats (loci sizes ranged from 12 to 63 bp) than the un-enriched genomic library, for which the longest unit was only 18 bp.

1.5 Mutational Model

There are two potential mechanisms that can explain the high mutation rates of microsatellites. The first is recombination between DNA molecules by unequal crossing-over or by gene conversion (Smith 1976, Jeffreys *et al.* 1994). The second mechanism involves slipped-strand mispairing during DNA replication (Levinson and Gutman 1987a). Studies using yeast and *Escherichia coli* as a model organism have shown that replication slippage appears to be the main mechanism that generates length variation mutations in microsatellites (Levinson and Gutman 1987b, Henderson and Petes 1992).

In replication slippage, the nascent DNA strand dissociates from the template strand during the replication of the repeat area and the nascent strand may then reanneal out-of-phase with the template strand. When replication continues, the resulting nascent strand may be longer or shorter than the template, depending on whether the looped-out bases occur in the template strand or the nascent strand. Microsatellites will then lose or gain a single or multiple repeats. These kinds of small length changes are the most common form of mutations at microsatellite loci, and have been detected in *E. coli* and yeast (Levinson and Gutman 1987b, Sia *et al.* 1997, Wierdl *et al.* 1997) as well as in humans (Weber and Wong, 1993).

To estimate population differentiation measures and genetic distance using microsatellite data, a theoretical mutation model for the evolutionary process that changes microsatellite allele size is required. Two theoretical models have been considered for microsatellites (Deka *et al.* 1991). In the infinite allele model (IAM), mutation can involve any number of tandem repeats and always results in a new allele state that had not previously existed in the population (Kimura and Crow, 1964). As mentioned above however, slipped-strand mispairing is currently accepted as the major mechanism considered to produce microsatellite length variation. This mechanism mostly produces small changes in repeat number so that alleles of similar length should be more closely related to each other than will alleles of very different sizes. Alleles may also mutate towards allele states that are already present in the population. The stepwise mutation model (SMM)

(Kimura and Ohta, 1978) developed for allozyme markers provides a better description of these kinds of evolutionary processes. In addition to this model, Di Rienzo *et al.* (1994) described the two phase model (TPM), where a limited proportion of mutations involve several repeats. Although rarely cited in the microsatellite literature, a K-allele model (KAM) could also be considered for microsatellites. Under this model, there are K possible allelic states, and any allele has a constant probability of mutating towards any of the other K-1 allelic states (Crow and Kimura, 1970).

Different kinds of repeat number variance estimators based on the stepwise mutation model (SMM) have been developed for estimating phylogenetic relationships [$(\delta\mu)^2$, Goldstein 1995a,b], genetic distance (D_{SW}, Shriver 1995) and population differentiation (R_{ST}, Slatkin 1995) from microsatellite data. These estimators are based on the following assumptions:

1. mutation results in a change of one repeat unit,
2. the mutation rate is constant and independent of repeat length,
3. there is no asymmetry in the distribution of mutations
4. there are no allele size constraints.

1.6 Objectives of this study

The objectives of this study are :

1. To develop and isolate microsatellite markers for *Scleropages formosus*.

2. To use the markers to study genetic diversity in *Scleropages formosus*.