CHAPTER II

LITERATURE REVIEW

2.1 Arowana

2.1.1 Classification Of Arowana

The arowana is a primitive fish from the Jurassic era (Bonde, 1979), belonging to the order Osteoglossiformes. The osteoglossiformes are considered basal teleosts with their bony tongue being one of their primitive features. Individual members show distinct morphological, physiological and behavorial differences (Nelson, 1994) which obscure their phylogenetic relatioship (Li & Wilson, 1996).

The Osteoglossiforme is devided into two suborders, Osteoglossoidei and Notopteroidei. The former includes the families Osteoglossidae and Pontodontidae whereas the latter comprises of Hiodontidae, Notopteridae, Mormyridae and Gymnarchidae (Nelson, 1994). Osteoglossids or Bonytongues have a world-wide distribution. The splitting of continents caused their ancestors to evlove into different species. This family is further divided into two subfamilies, Osteoglossinae and Heterotidinae. The members of the sumfamily Osteoglossinae have distinct barbels on the mandible while the Heterotidinae have no mandibular barbles (Dawes *et al.*,1999).

The subfamily Osteoglossinae is divided into two genera, *Osteoglossum* and *Scleropages*. There are two species of *Osteoglossum* in South America. *Osteoglossum bicirrhosum* and *Osteoglossum ferreirai* and are refered to as the silver arowana and the black Arowana respectively. There are three species in the genus

Scleropages. Two species, *Scleropages jardinii* , the Gulf Saratoga or Nothern Spotted Barramundi and *Scleropages leichardti*, the Saratoga or Spotted Barramundi are found in Australia while *Scleropages formosus*, the Asian arowana is distributed in the South East Asian region (Dawes *et al.* 1999). The subfamily Heterotidinae consists of *Arapaima gigas* found in South America and *Heterotis niloticus niloticus* distributed in Niger and the upper Nile (Dawes *et al.*, 1999).

2.1.2 Distribution of Asian Arowana

The Asian arowana (*Scleropages formosus*) is also known as *dragonfish*, *Asia Bonytongue*, *kelisa* or *baju rantai*. Records of this fish date back to 1844. This species is distributed in Peninsular Malaysia, Sumatra, Thailand, Cambodia and Kalimantan. In Malaysia, this fish is found in certain rivers like the Krian River, Muda River, Endau River, Rompin River, Pahang River and Terengganu River. This fish can also be found in Kenyir Lake, Bera Lake and Bukit Merah Lake. Although the arowana is widely distributed, the natural population is very low (Suleiman, 1999).

2.1.3 Habitat of Arowana

In nature, the arowana is found in swamps, including herbaceous swamps or marshland, lakes, rivers, mining pools and reservoirs. This fish prefers still or slow-flowing waters which are turbid or weedy. The *dragofish* is a surface dweller and glides along the upper layers of the water column, especially at night and during courtship (Dawes *et al.*, 1999).

2.1.4 Morphology

The arowana shows a knife-shaped compressed body while the abdomen is keeled. The gap of the mouth is very large and the lower jaw sticks out. A pair of barbels is present on the chin while the dorsal fin base is shorter than the anal fin base. The anal fin base is longer than its head length and the caudal fin is truncate. The dorsal, caudal and anal fin bases are covered by fine scales (Suleiman, 1999). The fish can grow up to around 90cm but the average length is shorter (Dawes *et al.*, 1999).

It is difficult to distinguish the sexes when they are young, and even after maturity. Generally, males are larger in size than females and look slimmer while females have more rounded bodies. Males have wider and deeper jaws in which they incubate the eggs (Dawes *et al.*, 1999).

2.1.5 Diet

The arowana is a predator. In nature, it mainly preys on small fish, shrimps, insects, crickets, centipede and frogs. It is a surface dweller and will glide below the water surface waiting to catch low flying insects by surprise. In captivity, the arowana can be fed on live insects, small frogs, fishes, live shrimps and chopped fish meat (Dawes *et al.*, 1999).

2.1.6 Breeding

Male arowana reaches sexual maturity at three to four years of age while the female fish needs between four to five years to mature. In the wild, the fish pair by natural selection. The pair start courting for a period of more than a week (Suleiman, 1999). When a pair is formed, the male trails the female in circular motion and starts to develop its territory. They will then swim side-by-side rubbing each other. When ready, the female releases eggs and the male will fertilise them immediately. Then the male will picks up the fertilised eggs in its mouth and keeps them there until the eggs hatch and the larvae develop into free-swimming fries (Dawes *et al.* 1999).

The diametre of an egg is around 1.5cm. Thirty to sixty eggs will be released at each spawning. The fertilised eggs are orange-red while the unfertilised eggs are light yellow in colour. It takes a week for the eggs to hatch in the father's mouth. The brooding will take fifty to sixty days until the yolk sacs disappear and the fries swim freely. In farm practice, the four-week old larvae, still with their yolk sacs, are removed from the father's mouth (Appendix 1, Fig. 1 to 7). This is to prevent the larvae from being swollowed by the father if it is disturbed (Suleiman, 1999). Sometimes, the fertilized eggs with visible blood vessels can be collected (Ng, H. Y, personal communication). After two months the male can mate with another female again (Suleiman, 1999). The survival rate of the fries is approximately 90% in captivity.

2.1.7 Raising The Larvae

The larvae are raised in aquaria measuring 90 x 45 x 45cm. The aquaria are aerated and the water temperature is maintained at 28 to 30°C by using heater-thermostats. It will take the larvae a month to swim freely. When the larvae start to swim they are fed on live shrimps (*Macrobrachium lanchestri*) or fishes (*Gambusia affinis*). After two to three months, the fries will reach 15 to 20cm. They can then be sold in the market (Suleiman, 1999).

2.1.8 Strains of Arowana

The arowana is famous for its brilliant colours. The colours range from gold, golden red and green. There are four commercial varieties of Asian arowana, namely green arowana, Indonesian red-tail gold arowana, Malaysian gold arowana and Indonesian red arowana.

2.1.8.1 Malaysian Gold Arowana

This variety is very popular and can fetch high price in the market. Malaysian gold arowana is native to Bukit Merah Lake, Perak. It is more attractive than its Indonesian counterparts. When the fish reaches 10cm, it has golden scales to the fourth column and pinkish edge on each scale. When the fish grows up, the whole body turns gold and the pink colour disappears. The scales may have different shade of colours such as gold, silver or blue (Appendix 2, Fig. 1).

2.1.8.2 Indonesian Red Arowana

This is the most well known variety. The red arowana is found in Kalimantan and Sumatra, Indonesia. This variety can be divided into first class red and second class red. It is difficult to differentiate first class red from second class red when the fish is young. Young first class red has orange-yellow scales with some green at the base. The edges of the scales are pink. When the fish is fully grown, the colour of the scales and fins develop into apricot, chilli red, deep red, blood red or brown (Appendix2, Fig.2). On the other hand, the scale colour of the young second class red is not obvious and often whitish silver. The adult second-class red has only pinkish or orange fins and scales. Generally, adult red arowana has longer body than the other varieties. The red arowana has either pear shaped or fan shaped tail fin.

2.1.8.3 Green Arowana

Green arowana is found in Malaysia, Thailand, Vietnam and Myanmar. In Malaysia this variety is distributed in Terengganu, Pahang and Johor. It is generally shorter and smaller than the other varieties. This fish has olive green scales and a prominent lateral line. Young fish has yellow fins while the fins of the adult are dark green in colour (Appendix 2, Fig. 3). Those which have purple-spotted scales are more expensive. This variety is less popular and is the least expensive among the four commercial varieties.

2.1.8.4 Indonesian Red-Tail Gold Arowana

This variety is found in Kalimantan and Sumatra islands of Indonesia. The scales are copper-gold in colour with purplish shine. Scales above the lateral line, dorsal fin and upper half of its tail are dark green. The lower half of its tail fin, dorsal fin and anal fin have purplish-red to brwonish-red colour (Appendix 2, Fig. 4). It is difficult to distinguish between the young of the Malaysian gold arowana and the Indonesian red-tail gold arowana. When the Indonesian red-tail gold arowana reaches 20 cm it does not appear to be as shinning as the Malaysian variety.

2.2 Application of Molecular Genetic Markers

Genetic markers are heritable characters with multiple states of each character. All genetic markers reflect differences in DNA sequences. Information about populations and evolutionary processes can be obtained by examination of genetic markers. The changing rates and the distributions of the different genetic markers depend on the differential actions of fundamental processes, including recombination, mutation and selective constraints (Sunnucks, 2000).

Molecular genetic markers have become a well-established and valuable tool for many applications in population genetics, conservation biology and evolutionary studies as well as for gene mapping projects (Jarne & Lagoda, 1996; Queller *et al*, 1993). These markers have been used to estimate effective population size (Kuhner *et al*., 1998), past bottlenecks (Luikart & Cornuet, 1998), sex-specific gene flow (Latta & Mitton, 1997), founder contributions (Carvajal-Carmona *et al*., 2000) and historical

and geographical relatioships among groups (Cruzan & Templeton, 2000). Different markers are better suited to approach different questions. Selecting the appropriate genetic markers is important for a population genetics survey (Sunnucks, 2000).

Molecular genetic markers are divided into two types, protein and DNA markers. Protein markers include isozyme whereas DNA markers consist of mitochondrial DNA (mtDNA) and nuclear DNA (nDNA). Over the past few years, a number of DNA marker systems, including restriction fragment length polymorphisms (RFLP) and PCR-based methods for the detection of fragment such as random amplification of polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and microsatellites, have been developed.

2.3 Molecular Genetic Markers

2.3.1 Isozymes

Molecular approaches in fish genetics began in the 1950s. Initially serological studies in blood group variants were adopted to demonstrate the existence of genetically controlled variations. However, fishery biologists were more in favour of electrophoretic procedures which revealed genetically determined protein polymorphisms. One of the first of these studies were done by Sick (1961) who described haemoglobin variants in whiting (*Gadus melangus*, Gadidae) and cod (*Gadus morhua*, Gadidae). Extensive work was carried out by Harris (1966) and Lewontin and Hubby (1966) to show the presence of highly polymorphic isozymes. Protein or isozyme electrophoresis was found to be quick, inexpensive and

reproducible but this technique also has several weaknesses. Protein electrophoresis can only provide an indirect assessment of nuclear DNA variability because it can examine only a specific set of genes that code for enzymes. Besides, protein loci evolve more slowly than non-coding DNA sequences. Therefore, there is controversy on whether isozyme variation is selectively adaptive or neutral. This technique also requires multiple fresh or frozen tissues and more material than most methods applying DNA analysis. Fish generally must be sacrificed and this is impossible for endangered species and those of high economic value.

2.3.2 Mitochondrial DNA (mtDNA)

2.3.2.1 Background of MtDNA

The mitochondrial genomes in bony fish are between 17.0 to 18.0 kb (Park & Moran, 1994). MtDNA is a closed circular molecule, made up of about 37 genes coding for 22 tRNAs, 2 rRNAs (12S and 16S), 13 polypeptides and a control region or D-loop of about 1kb. The polypeptides are subunits of enzymes required in electron transport and oxidative phosphorylation, including 3 cytochrome oxidase subunits, one cytochrome reductase unit (cytochrome b), 2 *ATPase* synthetase subunits and 7 NADH dehydrogenase subunits (Nishibori *et. al.*, 2001).

2.3.2.2 Application of MtDNA Analysis

Mitochondrial DNA analysis has become a widely used technique for many applications in population or evolutionary studies (Avise, 1986; 1989). Phylogenetic perspectives on intraspecific population structure were explained based on mtDNA

data since the late 1970s (Avise, 1994). Using mtDNA rather than nuclear DNA (nDNA) has several advantages. Mitochondrial DNA is haploid and maternally inherited. There is no paternal contribution and therefore recombination in the mitochondrial genome is absent (Avise, 1994). Hence, it has an effective population size (N_e) only one quarter of nuclear DNA, and consequently mtDNA variants become a rapid diagnostic marker of taxa (Ward & Grewe, 1995).

MtDNA is relatively easy to be purified and sequenced because the mitochondrial genome is small and contains almost exclusively coding sequence. (Vawter & Brown, 1986). The application of PCR to the analysis of mtDNA solves the problem of sampling logistics faced by allozyme and conventional mtDNA studies which require fresh or frozen tissue. Only small amounts of blood or tissue samples are adequate. The availability of universal primers (Kocher *et al.*, 1989) and the complete sequence data of mitochondrial genome (Gissi *et al.*, 1998; Hurst *et al.*, 1999; Lee *et al.*, 2001; Nishibori *et al.*, 2001; Saitoh *et al.*, 2000) remove the technical difficulties of repeated sequencing and therefore increase the efficiency of detecting nucleotide the diversity of the mtDNA genome. Since mtDNA is maternally inherited, the genetic content is conserved. The sequential order of genes varies only among taxa, making it easy for analyses (Vawter & Brown, 1986).

Different regions of mtDNA evolve at different rates. Cytochrome b and ND genes have been examined (Hansen *et al.*, 1997; Kumazawa & Nishida, 2000) as they are reported to exhibit high variabilites. The fast evolving D-loop is highly variable at

the population level but this is not the case with chum salmon (*Oncoryhnchus keta*) (Park *et. al.*, 1993). The mitochodrial ribosomal genes such as 12S and 16S evolve more slowly and have been used for relationship analyses at a variety of taxonomic levels (Crandall & Fitzpatrick, 1996; Kitaura *et. al.*, 1998; Palumbi & Benzie, 1991). Due to the ability of mtDNA to retain a history of past isolation (Avise, 1989) mtDNA genealogies have been used extensively to detect phylogeographical processes (Avise *et al.*, 1987). Studies have shown that the analyses of mtDNA are effective in demonstrating historical biogeography in shaping intraspecific and interspecific genetic structure (Dodson *et al.*, 1995; Sivasundar *et al.*, 2001). Since mtDNA substitution rates are homogenous across lineages, the date of divergence can be estimated based on genetic distance data (Bermingham & Avise, 1986). This technique has been used to answer questions regarding the phylogenetics of sturgeon (Ludwig & Kirschbaum, 1998) osteoglossoids (Kumazawa & Nishida, 2000) and Asian arowana (Pouyaud *et al.*, 2003).

2.3.2.3 The Drawback of MtDNA

There is controversy over the neutrality of mtDNA haplotypes because the coding regions tend to mutate under selective constraints (MacRae & Anderson, 1988; Nigro & Prout, 1990). For most genetic marker applications, the number of linkage groups represented influences the independence of phylogenetic information. Thus, the number of linkage groups is more important than the number of functional genes assayed. Although mtDNA consists of 37 function genes, all of these genes are transmitted maternally as a non-recombining unit. Hence the entire mtDNA molecule

can be considered as a single non-recombining genealogical unit with multiple alleles (Avise, 1994). Some marine fish populations are often dominated by only one or two mtDNA haplotypes together with many rare variants (Ward & Grewe, 1995). Furthermore, many marine teleosts showed low level of haplotype variations due to genome evolution, recent bottleneck events, family specific mortality or a failure to detect variants (Ovenden, 1990).

2.3.3 Nuclear DNA (nDNA)

The nuclear genomes in bony fish range from 0.3 to 4.0 billion base pairs in size (Ohno, 1974). Nuclear DNA (nDNA) is an important source of genetic information. Many methods such as the analysis of intron, specific genes and repeated sequences (VNTRs) have been used to look at sequence variations in the nuclear genome.

2.4 DNA Markers Systems

2.4.1 Restriction Fragment Length Polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP) markers are very useful because they are co-dominant and DNA fragments of variable lengths can be detected due to gain or loss of restriction sites. RFLP analysis involves development of probe libraries and Southern blotting which are laborious and time consuming. It is impractical for large-scale population studies. The capability of RFLP to detect polymorphisms is low compared with other marker systems. There were no significant

differences in RFLPs between various colour strains of Asian arowana (Fernando *et al.*, 1997).

2.4.2 Random Amplified Polymorphic DNA (RAPD)

Random amplified polymorphic DNA (RAPD) profiles can be generated by single arbitratry oligonucleotide primers and polymerase chain reaction (PCR) (Welsh & McClellan, 1990). This technique is simple, fast and needs no prior sequence information. Polymorphisms can be caused either by failure to prime a site in some individuals due to nucleotide sequence differences at the primer sites, or by insertions or deletions in the fragments between two conserved primer sites (Clark & Lanigan, 1993). RAPD has been successfully used to genetically discriminate species (Dinesh *et al.*, 1993), strains (Bardakci & Skibinski, 1994; Koh *et al.*, 1999) and assessment of Medelian segregation (Chong *et al.*, 1999) in fishes. However RAPD markers have some disadvantages. These markers are usually scored as dominant alleles. Thus the presence of an anonymous amplified fragment might reflect either a homozygous or a heterozygous state (Ward & Grewe, 1995). Furthermore, the reproducibility of RAPD markers is low due to stringent PCR. It has been reported that RAPD markers are not suitable for intraspecific mapping of Asian arowana because the genetic variability between strains was low (Fernando *et al.*, 1997).

2.4.3 Microsatellites

2.4.3.1 Background of Microsatellites

Microsatellites, also known as simple sequence repeat (SSRs) are short segments of DNA in which a specific motif of 1 to 6 bases is repeated up to a maximum of 60 times or so. These repetitive DNAs are dispersed throughout the eukaryotic genome and may constitute as much 66% of the nuclear DNA (Skinner, 1997). They occur approximately once in every 10 kbp in fish species (Wright, 1993). They were first discovered by liquid hybridization (Britten & Kohne, 1968). Tandemly-arrayed repetitive DNAs or satellite DNAs consist of head-to-tail repeats that vary in length from a few to several hundred base pairs (bp) (Brutlag, 1980). These satellite DNAs orbit bulk DNA in buoyant density centrifugation. They have been classified as satellite, minisatellite or microsatellite DNA.

There are different types of microsatellites. A classic microsatellite consists of single repeat only. An interrupted microsatellite is a repeat motif interrupted by base substitutions. A composite microsatellite consists of two adjacent repeats with different sizes or sequences of the repeat unit. Cryptic microsatellites or simple sequences have many interruptions including the addition of a few different motifs (Schlötterer & Zangerl, 1999).

Poly (A) or poly (T) is the most common repeat types in the human genome (Stallings, 1992) but they are not suitable as genetic markers due to their instability during PCR reactions. Dinucleotide repeats are the most widely used loci. The most

common dinucleotide repeats in the animal kingdom are CA (Beckmann & Weber, 1992). In plants, GA and TA repeats are the most common (Stallings, 1992; Lagercrantz *et al.*, 1993). Trinucleotide repeats are frequently used in human disease and cancer studies (Charlesworth *et al.*, 1994; Sutherland & Richards, 1995). They are rarely used in population genetics because they are distributed within exons where they do not disrupt the reading frame (Jurka & Pethiyagoda, 1995). The most common tetranucleotide repeats are GATA. They are often found as compound or interrupted repeats and are highly polymorphic (Weber & Wong, 1993; Ellegren, 1995; Viard *et al.*, 1996).

2.4.3.2 Application of Microsatellite

Due to their exceptional variability and relative ease of scoring, microsatellites are now generally considered the most powerful genetic marker. It is typical to observe loci with more than 10 alleles and heterozygosities above 0.60, even in relatively small samples (Bowcock *et al.*, 1994). Microsatellites are rapidly replacing RFLPs and RAPDs for most applications in population biology, from identifying relatives to inferring demographic parameters (Bowcock *et al.*, 1994).

Microsatellites are suitable as genetic markers for numerous applications in aquaculture and fisheries research as they are very abundant. So sufficient markers can be readily developed for any research objective. Some microsatellites show extremely high levels of allelic variation. Hence they are suitable for species that show low overall levels of variation with conventional markers such as allozymes or

mtDNA, for example Atlantic salmon (*Salmo salar*, Salmonidae) and Atlantic cod (Bentzen *et al.*, 1996; Wright 1993). Microsatellites can be used as genetic markers for populations that are inbred or have experienced severe bottlenecks. Furthermore, microsatellites are attractive in research on recently derived or geographically proximate populations where genetic differentiation may be limited. They are also suitable for pedigree analysis, such as in aquacultural husbandry or studies of variation in reproductive success among individuals. Microsatellite alleles are co-dominant markers inherited in a Mendalian fashion. Co-dominant Mendalion inheritance makes microsatellites more informative in pedigree studies, as well as in population studies.

Since microsatellites are assayed using PCR, only small amounts of tissues are required for analysis. This allows use of logistically convenient tissue sources such as fin-clips and scales preserved by drying or storage in alcohol at ambient temperature, in order to save storage and transport costs. Microsatellites can be assayed rapidly compared with many other DNA markers. Moreover, the highly allelic nature of microsatellites means that they confer more information per unit assay than most other marker systems. Evaluation of microsatellites also becomes even more rapid with the introduction of automation and fluorometric detection methods (Budowle & Morreti, 1999). Lastly, microsatellite primers developed for one species frequently amplify polymorphic loci in related species. For example, Atlantic cod primers amplify microsatellite loci in haddock (*Melanogrammus aeglefinus*, Gadidae) and pollock (*Pollachius virens*, Gadidae) (Brooker *et al.*, 1994); rainbow trout (*Oncorhynchus*

mykiss, Salmonidae) primers amplify polymorphic microsatellite loci in other Pacific and Atlantic salmon (Bentzen *et al.*, 1996).

Identification of microsatellite loci (mostly dinucleotide repeats) is already well documented for several species such as mammals, birds, plants and fish. In fish, microsatellites have been reported for many species such as Atlantic salmon (*Salmo salar* L.) (Slettan *et al.*, 1995a,1995b), brown trout (*Salmo trutta* L.) (Estoup *et al.*, 1993), zebrafish (*Brachydanio rerio*) (Goff *et al.*, 1992), Atlantic cod (*Gadus morhua* L.), rainbow trout (*Oncorhynchus mykiss*) (Brooker et al., 1994), sea bass (*Dicentrarchus labrax* L.) (Garcia de Leon *et al.*, 1995), the Asian river catfish, *Mythua nemurus* (Usmani *et al.*, 2001) and arowana (Yue *et al.*, 1999; Yue *et al.*, 2004)).

2.4.3.3 Microsatellite Evolution

Microsatellites are highly polymorphic. The average mutation rate for microsatellites in *Drosophila melanogaster* is 6.3×10^{-6} per generation (Schug *et al.*, 1997; Schlötterer *et al.*, 1998) while the mutation rate estimates for mammals are 10^{-3} to 10^{-5} (Dallas, 1992; Weber & Wong, 1993; Ellegren, 1995). The most popular explanation for generating length mutations in microsatellites is 'slipped-strand mispairing' during replication. In replication slippage the newly synthesised DNA chain will be longer or shorter than the template, depending on whether the looped-out bases occur in the newly synthesised DNA chain or in the template chain respectively. Then microsatellites will gain or loose of one or more repeat units (Levinson & Gutman, 1987). The length of the microsatellite repeats plays a role in the mutation

rate. Longer repeats (e.g. tetranucleotide repeats) evolve faster than shorter ones (e.g. dinucleotide repeats) (Weber & Wong, 1993; Chakraborty *et al.*, 1997; Primmer *et al.*, 1998; Ellegren, 2000). This is probably due to inefficient repair of longer mismatched segments.

Two leading models, the infinite allele model (IAM) (Kimura & Crow, 1964) and the stepwise mutation model (SMM) (Ohta & Kimura, 1973) have been developed to explain the evolution of microsatellites. The IAM is a classical divergence estimate for neutral loci which assumes that mutation can involve any number of repeats and result in novel alleles (Kimura & Crow, 1964). Under the SMM, mutation at microsatellite loci are stepwise in nature. Each mutation involves gain or loss of a single repeat with equal probablity and the mutation rate is independent of array size (Falush & Iwasa, 1999). However, Valdes *et al.* (1993), Goldstein *et al.* (1995a, b) and Slatkin (1995) suggested that the IAM is not suitable for microsatellites due to their high mutation rates and mutation processes that retain memory of the ancestral allelic states. DiRienzo *et al.* (1994), Estoup *et al.* (1995) and Garza *et al.* (1995) found that the SMM is inadequate to fully explain microsatellite mutational mechanisms. Thus, the two phase model (TPM) which involves a SMM with occasional multi-step mutation was suggested (DiRienzo *et al.*, 1994). Although rarely cited, another classical model, the K-allele model (KAM) (Crow & Kimura, 1970) can also be considered for microsatellite. This model assumes a finite number of K allelic possibilities with a constant probability (μ/K-1) of mutating towards any other K-1 allelic state (Estoup & Cornuet, 1999).

2.4.3.4 Controversy of Microsatellite

2.4.3.4.1 Null Allele

The first problem is pseudo-death of an allele or locus due to nucleotide substitutions, insertions or deletions which occur within the priming site and prevent primer binding (Callan *et al.*, 1993). The reduction or complete loss of amplification, may become fixed in the population. Fixation of null alleles is the major factor of failed cross-species amplification (Chambers & MacAvoy, 2000). The null alleles will not be regconized when there is a product from the other allele homologue. Heterozygosity will be underestimated when compared with the expected on the basis of Hardy-Weinberg equilibrium (Callan *et al.*, 1993; Paetkau & Strobeck, 1995).

Very often the amount of DNA extracted from ancient samples, forensic samples, museum specimens and hair or faeces of free ranging animals can be very low and often in the picogram range. Under these circumstances, one allele of a heterozygous individual may not be detected. This can also lead to serious underestimation of heterozygosity (Taberlet *et al.*, 1996).

2.4.3.4.2 Size homoplasy

Microsatellite variation is usually identified based on conventional length criteria. It is possible that two alleles of identical size are derived from different evolutionary lineages, introducing the possibility of size homoplasy. Microsatellite homoplasy is expected under both the stepwise mutation model (SMM) and the two phase model (TPM) while the infinite allele model (IAM) does not generate

homoplasy (Estoup & Cornuet, 1999). This is because microsatellites may fit either the SMM or TPM (Jarne & Lagoda, 1996). In the stepwise fashion, the degree of homoplasy will increase with the mutation rate on the locus and the time of divergence of the two populations (Estoup & Cornuet, 1999). Grimaldi and Crouau-Roy (1997) showed that the mutational process had involved not only changes in the number of repeats but also perturbations in the non-repeated 5' and 3' flanking sequences. Besides, size homoplasy for example was common at complex microsatellite loci in Lake Malawi cichlids (van Oppen *et al.*, 2000). Size homoplasy should be viewed with caution by population biologists since it may lead to an overestimation of relatedness (Chambers & MacAvoy, 2000) and an alteration of the population structure (Viard *et al.*, 1998). Sequencing the microsatellite loci can reveal hidden genetic variations but this can be very time consuming and costly for any large-scale study. Interrupted microsatellites can be used for investigating population differentiations and evolutionary relationships between relatively distant populations. This is because interrupted microsatellites have higher variance in repeat number and consequently show a lower homoplasy rate than the pure ones (Estoup *et al.*, 1995).

2.4.3.4.3 Shadow Bands

The problems of fidelity of dinucleotide arrays synthesis is associated with the PCR process itself. Various factors including polymerase structure, 3'-5' exonuclease activity, dNTP and divalent cation concentrations and pH affect these problems (Eckert & Kunkel, 1990). The undesirable secondary products are commonly called shadow bands which are shorter than the primary amplification product (length n) by

multiple of 2 nucleotides (n-2, n-4...) (Hauge & Litt, 1993). This can make allele scoring problematic. Shadow bands can be avoided by using thermolabile DNA polymerases, such as Klenow fragment at the extension temperature of 37°C (Hite *et al.*, 1995). Some researchers prefer to work with tri- and tetranucleotide repeat arrays but dinucleotide repeat arrays occur more frequently (Chambers & MacAvoy, 2000). Another method for reducing the potential scoring difficulties is to use dinucleotide loci with a smaller size (<120 bp) as these loci tend to stutter less and their smaller sizes make them easier to separate during electrophoresis (O'Connell & Wright, 1997).