CHAPTER III

MATERIALS AND METHODS

3.1 Sample Collection

Five varieties of arowana were used in this study. Ninety-one arowana samples were obtained from seven sampling sites. The varieties, sampling sites and sample sizes studied are listed in Table 3.1.

Table 3.1. Varieties, sampling sites and sample sizes of arowana

Varieties	Sites	Sample size
Green	Pahang River	6
	Fisheries Department, Malacca	14
Malaysian Red-Tail Gold	Fisheries Department, Malacca	5
	Yong Peng, Johor	16
	Bangsar, Kuala Lumpur	1
Malaysian Yellow-Tail Gold	Bukit Merah Lake	7
	Fisheries Department, Malacca	9
	Yong Peng, Johor	7
Indonesian Gold	Fisheries Department, Malacca	2
	Kluang, Johor	9
Indonesian Red	Fisheries Department, Malacca	7
	Kluang, Johor	7
	King Kong Farm, Malacca	1
	Total	91

The green and Malaysian yellow-tail gold arowana were subdivided into wild and hatchery strains for the population study. The wild strains consisted of six green and seven Malaysian yellow-tail gold arowana collected from the Pahang River and Bukit Merah Lake respectively. Some of the Malaysian yellow-tail gold arowana from the Malacca Fisheries Department were the offsprings of the wild caught fish but the details of their parentage are unclear (Saadun, personal communication). The actual sources of the other fishes from the Malacca Fisheries Department and the commercial farms were unknown.

Blood collection from arowana and fin clippings were not allowed due to its high price. There was degradation of genomic DNA extracted from tissues of dead fishes. This is because some fish have powerful cellular endonucleases, e.g. DNase which prevent DNA extraction (Asahida *et al.*, 1996). Hence scales were collected from live and dead arowana. This sample collection method did not cause major injury to the fish and they would recover within two weeks. Dead fishes were frozen at -20°C for several days to two years. The scale samples collected were stored in absolute ethanol at 4°C until genomic DNA was extracted.

3.2 Genomic DNA Extraction

Approximately 30 mg of each scale was used per extraction. The scales were cut into small pieces. 260µl of TNES-Urea buffer (Asahida *et al.*, 1996) and 40µl of Proteinase (20mg/ml) K were added to the samples. The mixtures were incubated for 15 hours at 50°C. Equal volumes of phenol : chloroform : isoamyl alcohol (25:24:1)

D27	$(CA)_{17}$	F GTGTCAGTATAGTGAATCTGTAG R TGACAATGGCAGCATAATGAGAT	AF219958
D31	$(GATA)_{15}$	F GTTTGTCCCTCCATGCACTGAGAG R GTGATTGCCACATGGTTTTGTTGG	AF219959
D32	$(CA)_{13}$	F AGCACCCTGTTACTGGAAGAGA R AGTGTGATGCTTTTGCTTTGAGAA	AF219960
D33	$(CA)_{12}AAC(CA)_4$	F TATTACCATGCGCCCAGCACAC R TGGGTGAGCCAGAAGCAGGACT	AF219961
D35	$(GT)_{17}$	F CTGGTTTCCTCCCACACAGT R GCCCACACACCTTATTCACC	AF219962
D37	$(GT)_{31}$	F GCCTTACGCCCTGTGTTGC R TGGATATCTGTGAGTGGTGGTGAA	AF219963
D38	$(GT)_{24}$	F TTGGGGTCATGCCACTGG R CAATAAATACCAAACAGGGAACC	AF219964
D42	$(CA)_{19}$	F AGGAACATCACTGACAACACT R TGGACTAACTAGGAGCACAT	AF219965
D72	$(CA)_{14}$	F AGCAGGTTAATTTGGAGACT R CGACCCTGTATGGGACAAG	AF219966
D85	$(CA)_{10}$	F GTTCCACAGGGGCTGAGAAAAT R GAGGACGGAACAAAAGCATTGG	AF219967
D88	$(GT)_{11}$	F TTTCTTTCTGAGACTGAGG R CAACTCTTATCCACCATTT	AF219968
D92	$(GT)_{13}$	F AGTCGCACACCACCACCTCGA R TCAGCGATAACCCCACACCT	AF219969
D94	$(CA)_{16}$	F CAGCAGCAGTGACACGGGTTCG R TCGCAGGCTGATTAAAGGTGTG	AF219970
D95	$(CA)_9$	F CCTGCGGAAGAAGAAAAGACT R CATGGTGTTGGCTGTGAGGAG	AF219971

PCR amplification was performed on a Hybaid thermal cycler in a total volume of 25 µl. The reaction mixture contained 1x PCR buffer (Promega), 1.5 mM $MgCl_2$, 200 µM of each dNTP, 0.2µM of each primer, 1U *Taq* polymerase (Promega) and 20 ng of genomic DNA. Amplification was carried out using 4 minutes of initial denaturation followed by 33 cycles of 30 seconds of denaturation at 94°C, 25 seconds annealing at the temperature detailed in Table 3.3 and 25 seconds extension at 72°C with a final extension period of 5 minutes at 72°C.

Table 3.3. Annealing temperature of each locus

Locus	Annealing temperature
D01, D37	57°C
D04, D11, D13, D14, D15, D16, D27, D31, D32, D38, D85, D92, D95	55°C
D15, D42, D72, D88	50°C
D33	60°C
D94	62°C

The PCR products were run on a 3.5% (w/v) 1x TBE horizontal agarose (MetaPhor) gel containing 0.5 μg/mg ethidium bromide. The gel was electrophoresed at 90 V for 3 to 4 hours. In order to compare the effects of different laboratory procedures, selected PCR products were also run on a 10% (w/v) 1x TBE vertical polyacrylamide gel (20 cm x 20 cm) at 250 for 4 hours. Gel was stained using silver stain (Promega).

3.4 Mitochondrial DNA Analysis

The phylogenetic studies on arowana were based on the nucleotide sequences of the protein coding *ATPase6* and *ATPase8* genes. These mitochondrial DNA regions contain useful information for intraspecific relationship analyses in fishes due to their high evolutionary rates. Two specimens from each of the colour varieties were used in this study.

Two primers ATP8.2L (forward primer: 5'-AAAGCRTYRGCCTTTTAGC-3') and COIII.2H (reverse primer: 5'-GTTAGTGGTCAKGGGCTTGGRTC-3') (Sivasundar *et al.*, 2001) were used to amplify an approximately 950 bp fragment flanking the *ATPase6* and *ATPase8* genes. A fragment roughly 840 bp in length that included the entire *ATPase8* gene of 160 bp and *ATPase6* gene of 680 bp was used for further analysis.

PCR amplification was carried out in a 25 µl reaction mixture containing 2.5 µl of 10x PCR buffer (Roche), 1.5 mM $MgCl_2$, 1U of *Taq* DNA polymerase (Roche), 10 pmol of each primer, 0.4 mM of dNTPs and 1 to 1.5 µl of template DNA. Amplification was performed on an Eppendorf Mastercycler with the cycles as follow: 2 min at 94°C, 30 cycles of 1 min at 94°C, 30s at 50°C and 45s at 72°C and a final extension for 8 min at 72°C. The amplification products were checked for size by loading 5 µl on a 1% agarose gel with 0.5 µg/ml ethedium bromide.

The remaining PCR products were purified using the Qiaquick PCR purification kit (Qiagen). 5 µl of 3 M sodium acetate and 5 volume of PB buffer were added. The mixture was vortexed and transferred to a Qiaquick spin column, followed by centrifugation at 13000 rpm for 1 min. The flow-through was discarded. 0.75 ml of PE buffer was added and the column was centrifuged for 1 min followed by an additional spin for 1 min at maximum speed. The column was placed in a 1.5ml microcentrifuge tube. The PCR product was eluted with 30µl EB buffer (10mM TrisHCl, pH 8.5) for 20 min and the column was centrifuged for 1 min.

The purified PCR products were sequenced directly using the BigDye Terminator cycle sequencing kit (Applied Biosystems Inc.) with an ABI DNA Sequencer 373A (Perkin-Elmer). The sequencing reactions were carried out as suggested by the manufacturer. A total reaction volume of 20 µl contained 30 to 90 ng of purified DNA template, 32pmol of primer and 4 µl big BigDye Terminator Mix. The amplification consisted of 25 cycles of 30 s at 96°C, 15 s at 50°C and 4 min at 60°C.

For each sequencing product, 2 µl of 3 M sodium acetate (pH 4.6) and 50 µl of 95% ethanol were added. The mixture was vortexed and left at room temperature for 15 min to precipitate the product. The mixture was centrifuged at room temperature for 20 min at maximum speed. The supernatant was discarded. The pellet was rinsed with 250µl of 70% ethanol followed by a maximum spin for 5 min and the supernatant was discarded. Pellet was dried by placing the tube with the lid open in a heat block at 90°C for 1 min. Samples were loaded to an ABI DNA Sequencer 373A (Perkin-Elmer).

3.5 Data Analysis

3.5.1 Microsatellite Variation

The microsatellite allele frequencies of each strain at each locus were estimated using the GENEPOP version 3.1c computer software (Raymond & Rousset, 1995). The observed (H_o) and expected heterozygosity (H_e), number of polymorphic

loci and average gene diversity of each strain were estimated using the ARLEQUIN version 2.000 (Schneider *et al.*, 2001).

The microsatellite loci were tested for departures from Hardy-Weinberg equilibria using the exact tests of Guo and Thompson (1992) in GENEPOP version 3.1c. A Markov chain method with 1000 demamorizations, 100 batches and 1000 iterations / batch was used to calculate an unbiased estimate of the P value. The multilocus inbreeding coefficient, F_{IS} (Weir & Cockerham, 1984) was estimated for each strain, to indicate the overall deviation from Hardy-Weinberg expectations. The significanct P values were adjusted following the sequential Bonferroni procedures (Rice, 1989) to eliminate false assignment of significance by chance. Linkage disequlibrium between loci was tested using GENEPOP version 3.1c, again applying the sequential Bonferroni procedures (Rice, 1989).

The genetic distance between strains was measured by calculating F_{ST} (Weir & Cockerham, 1984), using ARLEQUIN version 2.000 (Schneider *et al.*, 2001). The probability associated with the F_{ST} value was evaluated through the random permutation procedure (minimum 1000 permutations). A dendrogram of relationships among the strains was constructed from the F_{ST} matrix using the neighbor joining algorithm (Saitou & Nei, 1967) in MEGA 2 (Kumar *et al.*, 2001).

3.5.2 Bottleneck

When a population experiences a bottleneck the effective population size is significantly reduced. It will show correlatively and progressively reduced numbers of alleles (k) and expected heterozygosities (H_e) at polymorphic loci (Cornuet & Luikart, 1996). Nei *et al.* (1975) suggested that allelic diversity is reduced faster than heterozygosity during a bottleneck. Hence, a recently bottlenecked population generally has less numbers of alleles than expected from the observed heterozygosity under the assumption that the population is at mutation drift equilibrium (Maruyama & Furest, 1985). A test based on the difference between the observed heterozygosity and the expected heterozygosity in the sense of Nei's (1978) gene diversities was applied to detect population bottleneck. A significant heterozygosity excess indicates that a population has experienced a recent bottleneck.

The BOTTLENECK programme of Cornuet and Luikart (1996) was used to test for any excess or deficiency across loci for observed heterozygosity relative to expected heterozygosity, using data from 21 microsatellite loci. Three alternative mutation models, 1) stepwise mutation model (SMM), 2) infinite allele model (IAM) and two phase model (TPM) were proposed in this programme. TPM was used as it is intermediate to the SMM and IAM and most microsatellite loci better fit this model (Cornuet & Luikart, 1996). This model was tested using the Wilcoxon signed test, with default settings and 1000 interations.

3.5.3 Mitochondrial Analysis

Sequence alignment of the *ATPase6* and *ATPase8* sequence data was performed using Clustal W with gap penalty 5 to 10 and gap length penalty 1 to 5 (Gibson *et al.*, 1996). The percentage of each nucleotide of all haplotypes and their substitution were calculated. The amount of genetic variance of the five colour varieties was examined using an analysis of molecular variance (AMOVA; Exoffier *et al.*, 1992). AMOVA analysis was performed using the programme ARLEQUIN version 2.000 (Schneider *et al.*, 2001). The significance of the estimated ϕ_{ST} value was tested for by 1000 times permutation.

The pairwise divergence among the arowana fishes were estimated based on individual mtDNA sequence data using the MEGA2 programme (Kumar *et al.*, 2001). In the analysis, the sequence divergence was estimated based on the Tamura and Nei (1993) model because of the existence of nucleotide bias in the sequences. Since substitution rates varied at different sites, the gamma correction was applied to the distance estimation with 0.5. A phylogenetic tree topology based on the distance measure was constructed according to a neighbor-joining method (Saitou & Nei, 1987) in MEGA2 (Kumar *et al.*, 2001) to display the relationships among the haplotypes of the five strains of arowana. The Tamura and Nei (1993) distance method with gamma correction 0.5 was used to calculate the topology of the phylogenetic tree. 500 bootstrap replicates were used to assess the confidence of the phylogenetic estimates.