

CHAPTER 3

METHODOLOGY

Channa striata is an exotic tropical fish; the interest in the fish arises from its high economical value. The fish was reportedly fetching around S\$10-20 in Singapore, in the late 1980s (Lim & Ng, 1990). This is probably due to its standing status as an oriental delicacy and also as a healing remedy (Baie & Sheikh, 2000; Kumar, et al., 2008). However, in the past decades stocks have declined (Cong, Phuong, & Bayley, 2009) and still not many studies have been conducted on the species on a genetic basis.

Therefore an attempt to study the current status and diversity of the species is the major objective of this study. This would be done by sampling the fish from seven different regions in Peninsula Malaysia and isolating microsatellites, which is the molecular marker that will be used to summon the current status of the fish's population.

3.1. Samples

Tissue samples from adult *Channa striata* were collected from seven different sites. The locations in alphabetical order are: Johore ($N=30$), Kedah ($N=30$), Negeri Sembilan ($N=30$), Pahang ($N=30$), Penang ($N=30$), Selangor ($N=26$) and Terengganu ($N=30$). Figure 10 shows map of locations of samples collected and Table 12 shows longitude and latitude of locations. Tissues were collected between July and August 2008. All samples were stored in -20°c until DNA extraction.

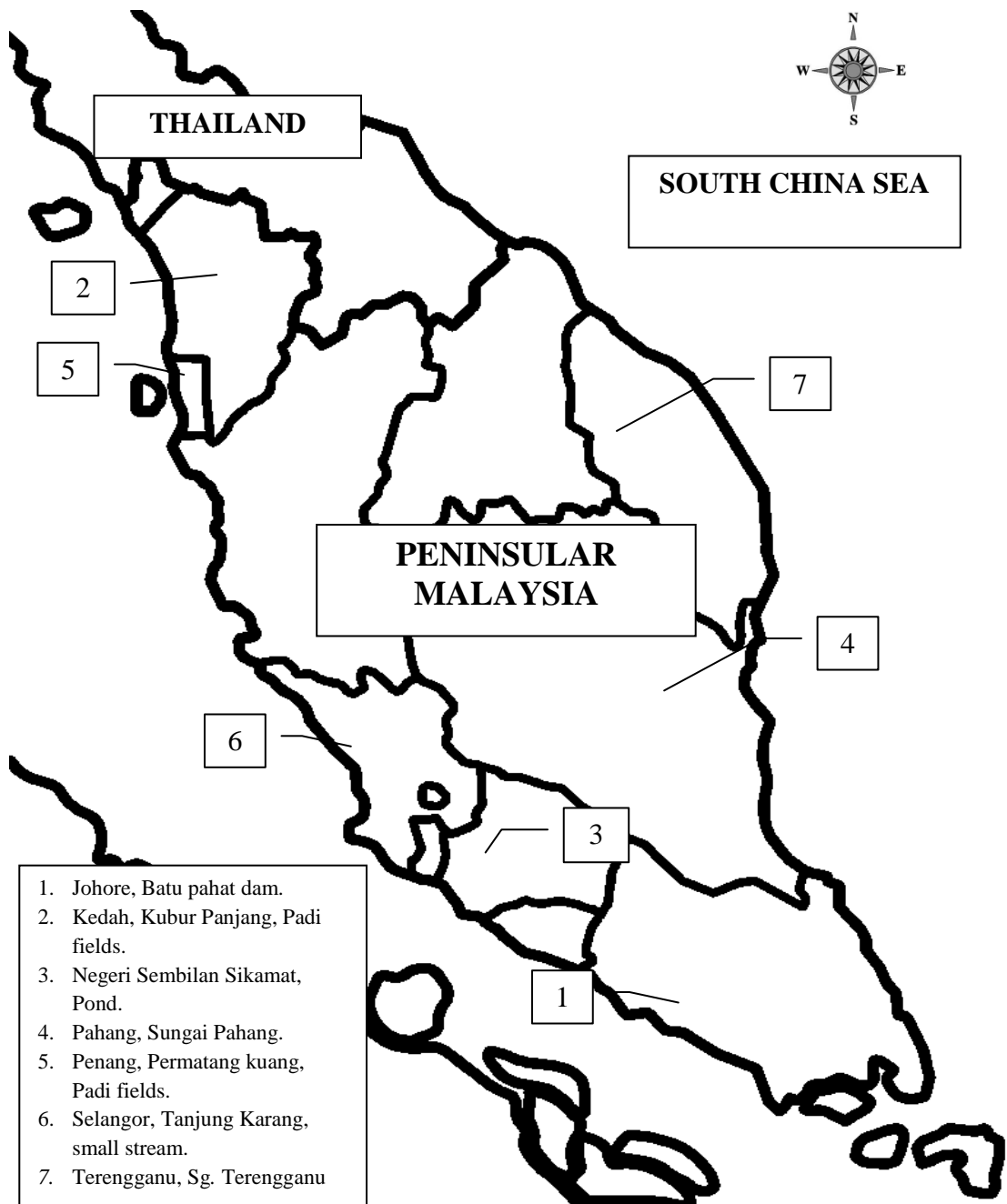


Figure 10 Map of Malaysia showing regions of sample collection. (Courtesy of: Google maps).

Table 2 Samples of *Channa striata* collected from seven locations around peninsular Malaysia.

Latitude and longitude are shown in decimal degrees.

State	Location	Latitude	Longitude
Kedah	Kubur Panjang, Padi fields	6.11427	100.54928
Penang	Permatang kuang, Padi fields	5.40665	100.36783
Selangor	Tanjung Karang, small stream	3.44285	101.18340
Negeri Sembilan	Sikamat, Pond	2.75925	101.94901
Johore	Batu pahat dam	1.85860	102.93777
Terengganu	Sg. Terengganu	5.37950	103.08884
Pahang	Sungai Pahang	3.69172	102.89516

3.2. DNA extraction

Extraction was carried out on all 206 samples using GF-1 Tissue extraction kit (Vivantis, Malaysia). Instructions given by the manufacturer were followed exactly, except for the last elution step, in which the elution buffer was left for at least an hour before centrifuging.

Approximately 25mg of tissue was cut and inserted into a microcentrifuge tube and 250µl of Buffer TL (Tissue lysis buffer) and 20µl Proteinase K (20mg/ml) were added mixed by pulsed vortexing. Next 12µl of Lysis Enhancer was added and mixed immediately. The tissue mixture was incubated in a water bath overnight at 65°C. The following day, 20µl of RNase A (DNase-Free, 20mg/ml) was mixed in and incubated for 10 minutes at 37°C. Then 600µl of Buffer TB was pipetted in and vortexed. The

solution was again incubated in a water bath for 10 minutes at 65°C. Thereafter 200 µl of absolute ethanol was added and pulse-vortexed. Approximately 600µl of the sample was inserted into a new collection tube and centrifuged at 5000 x g for 1 minute and then the flow through was discarded. The step was repeated once more for the remaining sample. Then the column was washed with 750µl of wash buffer and centrifuged at 5000 x g for another minute. Flow through was discarded and the step was repeated again. A centrifugation step followed at 10000 x g for 1 minute in order to eliminate all traces of ethanol. Lastly the column was inserted into a new microcentrifuge tube and approximately 100µl was added of the elution buffer, the tube was left to stand for at least an hour and a final centrifugation step was conducted at 5000 x g for 1 minute to elute the DNA. Samples were stored in -20°C till needed.

The concentrations of the DNA was measured using the NanoPhotometer™,UV/Vis spectrophotometer (Implen, munich, Germany).

3.3. Amplification of microsatellite regions using Random Amplified Polymorphic Primers (RAMs)

3.3.1. Primers

Microsatellites were isolated using random amplified polymorphic (RAM) primers (Fisher, Gardner, & Richardson, 1996). They were used to produce enriched libraries of microsatellite containing sequences. The construct of the primer is rather simple. The 5' end contains random nucleotides, these will form the “anchor” and at the 3' end is the tandem repeat sequence.

The attempt to optimize a number of (RAM) primers resulted in five successfully being optimized; each of which have effectively anchored on to the 5' region of the microsatellite repeats.

3.3.2. Polymerase chain reaction

The polymerase chain reaction (PCR) (C1000™ Thermal cycler; Bio-Rad, USA) was performed in 10µl reaction volumes containing 1x PCR buffer (Promega, USA), 2.5 to 3.75mM of MgCl₂ (Promega, USA), 1.0 unit of Taq polymerase (Promega, USA), 0.25mM of each dNTPs (Promega, USA), 50.0µmole of primer (1st Base), 20ng of genomic DNA and deionised water.

PCR amplification was carried out in a Bio-Rad thermal cycler, starting with a 3 minute pre-denaturation at 96°C, followed by 40 cycles of denaturation at 96°C for 10 seconds. Thereafter an appropriate annealing temperature was applied for 10 seconds and an extension step at 72°C for 30 seconds. A final extension step was performed at 72°C for 7 minutes.

3.3.3. Electrophoresis

The RAMs-PCR products were separated on 1% agarose gels at 70 volts in 1× Tris-borate-EDTA buffer for about an hour. Gels were stained in ethidium bromide and visualized under a 365nm UV light and documented using a gel documentation machine, AlphaImager (Alpha Innotech; CA, USA). PCR products were immediately cloned.

3.4. Cloning

Each RAMs-PCR product was cloned into a yT&A[®] *E. coli* vector (Figure 11), using a (yT&A[®] cloning kit; Yeastern Biotech, Taiwan). Protocol followed was exactly the same procedure on the manufacturers manual: 1µl of ligation buffer A, 1µl of ligation buffer B, 2µl yT&A[®] cloning vector, 5µl RAM-PCR product and 1µl T4 DNA ligase was mixed by pipetting and incubated overnight at 4°C.

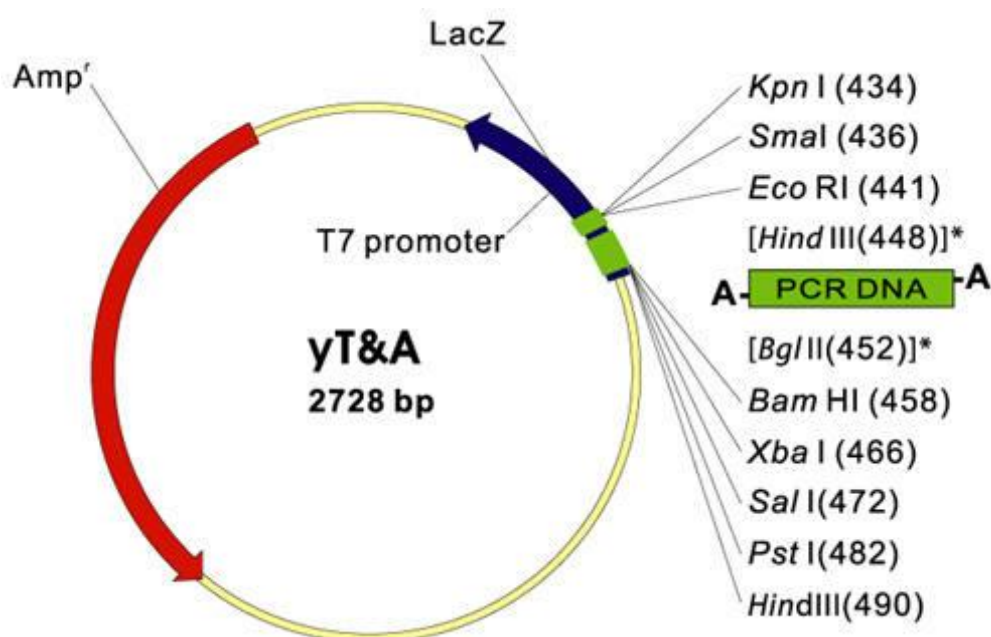


Figure 11 yT&A cloning vector (Wang & Chao, 2005)

Vectors were transformed into *Escherichia coli* competent cells (ECOS[™] 101 competent cells strain: DH5α; Yeastern Biotech, Taiwan) the next day, again following exact manufacturer instructions. Tubes were thawed of competent cells (Each tube typically contained 100µl) for about 30 seconds under tap water or until one half of the contents were thawed. Exactly 5µl of the RAMs-PCR products was added immediately into the competent cells' tube, vortexed for 1 second and placed in a 42°C water bath for 35 seconds. Contents were plated using glass beads on an LB agar (Lennox) (Laboratorios CONDA, Spain) plate containing (0.15mg/ml of LB agar) ampicillin (Bio

Basic Inc., Canada) and spread with (0.012mg/ml of LB agar) X-Gal (5 prime, Hamburg) dissolved in DMSO.

X-Gal tends to give a distinctive blue color when the fragments are not successfully transformed into the vector, however when transformation does occur the cells turn into white; making it easy for us to distinguish cells on the plate. The concept behind this lies in the *LacZ* gene. When interrupted, the gene is no longer transcribes β -galactosidase enzyme and therefore the X-Gal is not cleaved and no blue color is observed. The exact opposite happens when the gene stays intact and a blue color is produced.

The cells were left overnight in a 37°C incubator and the positive inserts were sub-cloned the next day. As a final step each positive white colony was inoculated into 5ml ampicillin (0.05mg/ml) containing LB broth (Conda Laboratories, Spain). The tubes were left in a 37°C shaker at 250rpm for about 12 hours or overnight. The next day, turbidity could be observed in the broth indicating that cells were formed and were ready for plasmid extraction.

3.5. Colony PCR

Colony PCR was conducted on each and every plasmid following the same PCR profile for the RAMs primers. Although the Blue/White screening method is quite effective in determining transformation, this method more or less confirms the success of the previous procedures, eliminating false positives and giving us an idea of the size of the insert and also the orientation in the vector (Sambrook & Russell, 2001; Summers, 2007). M13 forward and reverse primers were used, to give us a product of expected size, which includes the size between the 5' termini of the primers including the

expected size of the inserts. However, if no recombination occurs then the expected size would be the exact length between the M13 primer 5' termini. If in fact recombination does happen, then a definite insert is to be found in the PCR product and the next step of plasmid extraction is bound to give us vectors with inserts, without a doubt.

3.6. Plasmid Extraction

Plasmids were extracted using HighYield Plasmid Mini Kit (Real Biotech. Corporation, Taiwan). Following the protocol on the manufacturer's manual.

Around 1.5ml of the LB broth was poured into a new microcentrifuge tube and centrifuged for 1 minute at 14,000rpm. The supernatant was discarded and 200µl of PD1 buffer (RNase included) was added to resuspend the pellet by vortexing. Next 200µl of PD2 buffer was added and mixed in by inverting the tube 10 times. The tube was left standing for about two minutes and then 300µl of PD3 buffer was added and mixed immediately by inverting 10 times. The tube was vortexed at maximum speed (14,000 rpm) for three minutes.

A new collection tube was prepared in which a PD column was added and the clear lysate was poured into. The tube was centrifuged at full speed for 30 seconds and the flow through was discarded. The column was placed back into the collection tube and 400µl W1 buffer was added and centrifuged at full speed for 30 seconds. Once again the flow through was discarded and the column was returned back into the tube. This time 600µl of wash buffer with ethanol added and centrifuged at full speed for 30 seconds. The flow through was discarded and the PD column was inserted back into the tube, this time the tube was centrifuged with no additions at full speed for three minutes; to dry the column.

Finally the column was added into a new 1.5ml centrifuge tube and 50µl of elution buffer was added right into the middle of the membrane. The tube was left for about an hour and then centrifuged at full speed for two minutes to elute the DNA.

All plasmids were separated on 1.5% agarose gel; in order to assess the plasmid extraction success.

The remaining cells from the falcon tubes were all stored in glycerol for future reference. This was done by adding 150µl of glycerol to 850µl of cells in LB broth in a new 2ml microcentrifuge tube. All tubes were stored in -80°C freezer.

3.7. Sequencing

Extracted plasmids (which have proven to have inserts) were sequenced in an ABI PRISM 377 DNA sequencer using a BigDye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, USA). The universal primer, M13 (-20), 5'-GTA AAA CGA CGG CCA GT-3' was used for sequencing.

3.8. Sequence submission to GenBank

The resulting sequences were purified of restriction enzyme sites and vector sequences using VecScreen system (NCBI). The sequences were submitted to GenBank (NCBI) using a stand-alone software submission tool called Sequin, which was developed by National Center for Biotechnology Information (NCBI) and submitted via e-mail.

3.9. BLAST analysis of microsatellite marker

All sequences that resulted from successful cloning were run through BLASTn (Basic Local Alignment Search Tool) in the NCBI database and SynaTATE, sequence annotation tool (Malaysian Genomics Resource Centre Sdn Bhd); in order to look for annotations that might be of interest to us for further functional studies. The sequences were compared to reference mRNA sequences (refseq_rna) *Danio rario* and other organisms when using BLASTn (somewhat similar sequences); however comparison was only made to *Danio rario* when using the RNA NCBI release of synaTATE. The reason why this species was compared to other model fish; was that not much molecular or genomic work has been done on the indigenous species and the whole genome has not been sequenced yet.

3.10. Designing the microsatellite primers

The sequences were all screened for perfect microsatellites using a software called Perfect Microsatellite Repeat Finder, a web based software that was written by Robert Leach (2000), Los Alamos National Laboratory: <http://sgdp.iop.kcl.ac.uk/nikammar/repeatfinder.html>. A minimum of six repeats were designated for di- repeats, three for tri- repeats and two for both tetra- and penta-repeats.

A pair of primers were designed for each microsatellite, however some of the primers did flank to sets of repeats. This was done via a free online primer designing software, PRIMER 3 (Rozen & Skaletsky, 2000), provided at:

http://www.genome.wi.mit.edu/genome_software/other/primer3.html.

Criteria when designing these primers were as follows:

- The GC content would typically be between 40-60%, with an optimum of 50%.
- Primer size range was to be between 18 and 24 nucleotides, with an optimum of 20 nucleotides.
- An optimal melting temperature of 45°C and 65°C was set.
- Primers that tend to form secondary structures and GC rich 3' ends were avoided.

3.11. Microsatellite primer optimization and amplification

Each of the primers was first optimized, concentrating mainly on getting the annealing temperature optimized and keeping much of the other reaction components constant.

The PCR reactions were carried out in 10µl reaction volumes, which basically contained: 1 x PCR buffer, 1.0 unit of *Taq* polymerase, 0.25mM of each dNTP, 50.0µmole of primer, 20ng of template DNA and MgCl₂ 3.75mM. The PCR amplification was carried out in a Bio-Rad thermal cycler, starting with a three minute pre-denaturation at 96°C, followed by 40 cycles of denaturation at 96°C for 10 seconds. Then an annealing temperature range between 40°C and 65°C was applied for 10 seconds, and an extension step at 72°C for 30 seconds. A final extension step was performed at 72°C for seven minutes.

The PCR products were run on a 4% MetaPhore agarose gel (Cambrex, USA), followed by ethidium bromide staining (0.1 mg/mL) and visualized using AlphaImager (Alpha Innotech; CA, USA).

At first only 16 random individuals are tested; just to see if there are any variations in the banding patterns. If so, then they are suspected to be polymorphic and are promoted to the next step.

3.12. Microsatellite analysis using Genetic Analyzer

When a primer is thought to be polymorphic, it is again tested on a further 16 individuals, making it a total of 32 individuals that confirm a primer's polymorphism. Confirmation is done according to this was according to criteria, derived from experience and others from published sources:

- Amplified DNA should fall within the expected size of the microsatellite primer product.
- The loci should contain at least two alleles, however for those loci with only two alleles some individuals studied should be heterozygote and others homozygote; to be considered polymorphic. All other loci that have more than two alleles were initially considered polymorphic.
- Should amplify successfully, producing clear bright bands on the agarose gel and obvious peaks in the electropherogram of the genetic analyzer (Liu, et al., 2007).
- Should have minimum numbers of null alleles.
- Alleles that appear for each locus should be consistent and reproducible, meaning that it is observed in more than one individual (Liu, et al., 2007).

Once proven to be polymorphic, the forward microsatellite primers is fluorescently labeled using 5' FAM dyes and is used to amplify the same locus in all the individuals from the seven different locations. However instead of viewing them on a gel, electrophoresis is done through a fine capillary using an ABI Prism[®] 3130xl Genetic

Analyzer (Applied Biosystems, USA), in which mobility (electrophoretic movement) is measured instead of base-pair length (Applied Biosystems, 2004).

The exact same PCR profile was used for each of the microsatellite primer; in order to amplify the desired locus, except colorless buffer was used instead of the green buffer. Extra precaution was taken, as the fluorescent primers are sensitive to light.

Once the fragments have been amplified they are diluted 10-20 times using ddH₂O, depending on the height of the peaks shown. If the peaks are off-scale then further dilution is needed, nevertheless if the peaks were not significant less dilution is needed.

Preparation is needed for PCR product, starting off by adding 10µl Hi-Di™ formamide (Applied Biosystems, USA) and 0.2µl GeneScan™ 500 LIZ™ Size Standard (Applied Biosystems, USA) to 1µl of PCR product. The tubes were vortexed for a few seconds and then centrifuged. Next, they were heated for five minutes at 95°C in an ABI GeneAmp® PCR System 2400 (Applied Biosystems, USA). After that the tubes were kept for 5 minutes in an icebox and then loaded into an analysis plate and covered. The plates were tapped; to get rid of bubbles and then loaded onto the ABI Prism® 3130xl Genetic Analyzer (Applied Biosystems, USA). The analysis takes approximately six hours for a plate of 96 wells. Wells that were not loaded were filled with water.

3.13. Data analysis

Results of the Genetic Analyzer were retrieved and scored using a GENEMAPPER Software Version 4.0 (Applied Biosystems, USA). This software translated the results given by the Genetic analyzer into electropherograms that show peaks; indicating the exact size of the amplified fragments, or in other words the amplified loci. The fragments are compared to the loaded ladder, GeneScan™ 500 LIZ™ Size Standard,

which appears as range peaks in the graph. Polymorphic loci had their PIC values calculated manually (Botstein, et al., 1980).

After scoring all alleles of the seven loci under study, they were analyzed using a number of software that apply well recognized principles and algorithms, that brings us closer to understanding the current status of our studied species, as well as a slight insight of its past.

The data sheet was first run through a software named MICRO-CHECKER (van Oosterhout, et. al., 2004) that was used to eliminate stutter bands, null alleles and large allele dropout. The application uses Monte Carlo simulation (Bootstrap), to calculate frequencies of alleles that are either homozygote or heterozygote. Hardy-Weinberg (Levene, 1949) is also used to calculate expected allele frequencies and null allele frequencies. The later were calculated according to both by Chakraborty (Chakraborty, et al, 1992) and Brookfield (Brookfield, 1996).

These calculations were to give us a rough estimate of what to expect throughout the analysis. However this software was not used to compute the official results of the data analysis that were to be documented and published.

Size range of alleles, including number of alleles for each loci and population were calculated using HPRARE v1.0 (Kalinowski, 2004).

Once done the excel file format containing the actual scores of all the microsatellite markers were converted to other formats that are compatible with other software, needed for the analysis such as: GENEPOP, POPGENE, ARLEQUIN, etc. The software used for this purpose is CONVERT: a user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages (Glaubitz, 2004). CONVERT was also used to give a precise calculation of all allele frequencies

encountered. Genotypic frequencies on the other hand were determined using GENEPOP v4.0 (Raymond & Rousset, 1995).

The effective number of alleles (Kimura & Crow, 1964) and allelic frequencies were calculated using POPGENE v1.3 (Yeh, et. al., 1997). The overall observed and expected Heterozygosities (H_{obs} and H_{exp}) were estimated using GDA v1.0 (Lewis & Zaykin, 2001).

Inbreeding coefficient values (F_{is}) were determined using ARLEQUIN software v3.1 (Excoffier, Smouse, & Quattro, 1992). A run was carried out for each loci in every location under study. AMOVA gave an estimate of F_{is} values for each population overall. These assessments are to evaluate the diversities among individuals of a population (Weir & Cockerham, 1984). Inbreeding coefficient is calculated in this general manner: $F_{is} = \frac{H_s - H_i}{H_s}$. Whereby H_s is the average expected heterozygosity of individuals in a population and H_i is the average observed heterozygosity between those same individuals (Hamilton, 2009).

Conformation to Hardy–Weinberg Equilibrium (Hardy, 1908; Weinberg, 1908; Levene, 1949) at each locus, and for each population at each locus separately was conducted using GENEPOP v4.0 software (Raymond & Rousset, 1995). Results were obtained after a dememorization: 10000, batches: 20, iterations per batch: 5000. The principle behind this test is rather basic. It informs us of any unusual occurrence in the population that might lead to the changes in the consistency of allele frequencies across the generations, regardless of what that affect is (Hamilton, 2009). The equation that all software rely on is this: $p^2 + 2pq + q^2 = 1$, whereby p and q are symbols representing the genotype frequencies of any two alleles of a locus. If infarct the frequencies are kept

constant, then the null hypothesis is accepted and mating is random enough to conclude that the population is not subjected to any encounters that are besides the norm.

Linkage disequilibrium was computed using GENEPOP v4.0 [Fisher's method (Sokal & Rohlf, 1994)]. The Markov chain parameters were: Dememorization of 5000, batches: 500 and 5000 Iterations per batch. This would give us an idea of whether any of the loci are not randomly linked.

Genetic distances were used to evaluate the GDA was later on used to calculate Nei's 1978 Genetic distance (Nei, 1978). This software is based on Bruce S. Weir's book "Genetic Data Analysis II: Methods for Discrete Population Genetic Data" (Weir B. S., 1996).

A couple of UPGMA trees were depicted from both genetic distances. The first was a consensus UPGMA tree that results from a summarization of 1000 trees. This was done using PHYLIP (PHYLogeny Inference Package) v3.69 (Felsenstein, 1989). The software uses Nei's 1972 standard genetic distance (Nei, 1972) to construct the tree. The genetic distance is calculated using one the programs in the package named GENDIST that calculates the distances from the gene frequencies. The final output is an unrooted consensus tree with values on the branches that are similar to bootstrapping values.

Another genetic distance tree was constructed using GDA v1.0 (Lewis & Zaykin, 2001). This produced a UPGMA with a genetic distance scale based on Nei's 1978 unbiased genetic distance.

Next, ARLEQUIN was applied to evaluate genetic differentiation among the populations, using Fixation index (F_{st}) (Hartl, 1980). Pairwise F_{st} values were assessed; to compare genetic differentiation among the populations, based on polymorphism data.

The fixation index F_{st} can be used as short term genetic distances between populations. This is done using the formula $F_{st} = \frac{H_T - H_S}{H_T}$, where H_T is the average expected heterozygosity of all the populations and H_S is the average expected heterozygosity of each population (Hudson, Slatkin, & Maddison, 1992)

Lastly AMOVA was results were used to derive the overall molecular variance among and within populations.

A software program called STRUCTURE v.2.2 (Pritchard, Stephens, & Peter, 2000) was used to investigate the genetic structure, this software clusters the populations under study regardless of their geographical locations. Markov Chain Monte Carlo repetition of 500000 was set as a parameter for the project and the length of burning period was also 500000.