# AN ANALYSIS OF POPULATION STRUCTURE USING MICROSATELLITE DNA IN TWELVE SOUTHERN AFRICAN POPULATIONS OF THE MOZAMBIQUE TILAPIA, *Oreochromis mossambicus* (Peters)

**Edward G. Hall**

**Thesis presented in partial fulfillment of the requirements for the degree of Master of Genetics (AB) at the University of Stellenbosch**


**Study leader: Dr Rupert Lewis**

**Co-study leader: Mr Danie Brink**

**December 2001**

# DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

# ABSTRACT

DNA microsatellite loci express extensive allelic variation making them convenient markers for research in many fields employing population genetic tools, including aquaculture and conservation genetics. Twelve *Oreochromis mossambicus* populations from wild, captive and introduced sources in Southern Africa were screened for genetic variation at ten CA repeat microsatellite loci. Three of the loci - UNH104, UNH111, and UNH123 – were sufficiently well resolved to screen extensively and were interpreted according to a model of Mendelian inheritance. Data was analyzed in terms of genetic structure and levels of genetic variation, the effect of management regime in captivity through successive generations on genetic diversity, and the nature of phylogenetic relationships present between populations.

Exact tests, carried out using Monte Carlo type multiple resampling techniques, and F-Statistics were used to detect and quantify genetic structure among the twelve populations. The Exact test $\chi^2$ ($P < 0.001$), a $F_{ST}$ of 0.27 ($P < 0.001$), $\theta_{ST}$ of 0.26, $R_{ST}$ of 0.28, and a $\Phi_{ST}$ of 0.17 all indicated significant structuring among the populations. The evident genetic structuring endorsed the practice of maintaining the populations as separate genetic stocks, in separate tanks, in order to preserve unique genetic material for aquaculture strain development.

Populations also exhibited some significant deviations from Hardy Weinberg equilibrium characterised by an overall reduced heterozygosity across the loci. In microsatellite studies, null alleles are often suggested as major contributors to heterozygote deficits. To test for null alleles, two controlled crosses of *O. mossambicus* were made. The progeny from each cross were examined for expected parental allelic ratios at the UNH104, UNH111 and UNH123 loci. All three loci presented evidence of possible null alleles.

Accelerated inbreeding and genetic drift through successive generations in captivity can reduce heterozygosity and gene diversity. To investigate loss of diversity a sample taken from the Bushmans population in 1999 (N = 25) was compared with a Bushmans 2000 sample (N = 36). The comparison highlighted altered allele frequencies, a significant increase in average observed heterozygosity and a non-significant change in average expected heterozygosity using the UNH104 and UNH123 loci.

Calculation of genetic distances and phylogenetic comparisons between the populations provided insight into the degree of management required in conserving genetic diversity in natural populations of Mozambique tilapia. UPGMA and Neighbour-Joining techniques were used to construct phylogenetic trees using $D_m$ and $(\delta\mu)^2$ distance matrices. Clustering of populations appeared to reflect geographic locality of the source populations, however certain populations were not congruent with geography. Mantel tests were used to expose a possible association between genetic distance matrices generated from each individual locus. An association would support a

geographic background to population genetic structure. The Mantel tests did not provide conclusive evidence. Mantel tests for association between the combined locus $D_m$ and $(\delta\mu)^2$ genetic distance matrices and a geographic distance matrix were similarly non-significant.

Multi-dimensional scaling (MDS) plots of Euclidean distance values for $D_m$ and $(\delta\mu)^2$ matrices presented a two-dimensional view of the genetic distance data. The degree of similarity with the UPGMA and Neighbour-Joining tree-clustering pattern was higher for the $(\delta\mu)^2$ than for the $D_m$ MDS plots. Scatter plots indicated a reliable non-linear correlation between Euclidean distance and genetic distance for the two-dimensional MDS.

The microsatellite markers employed in this research provided molecular information needed for complimenting a co-study on quantitative genetic evaluation of the twelve populations. The quantitative co-study provided measures of average length and weight gain indices for the populations based on progeny growth trials. No significant correlation of average heterozygosity (gene diversity) with either average weight or length gain was found.

The significant genetic diversity and structure present between the twelve populations provided rationale for implementing strategies to conserve natural *O. mossambicus* populations as genetic resources, and manage captive populations for long term maintenance of genetic diversity.

# OPSOMMING

Die verstaffing van groot alleliese variasie deur DNA mikrosateliete maak van hulle gerieflike merkers vir navorsing in 'n verskeidenheid van velde wat gebruik maak van populasie genetiese gereedskap, ingesluit akwakultuur en bewarings genetika. Twaalf *O. mossambicus* populasies wat verkry was vanuit die natuur, in gevangeneskap en ingevoerdes, van Suidelike Afrika was getoets vir genetiese variasie by tien verskillende CA-herhalende mikrosateliet loci. Drie van die loci - UNH104, UNH111 en UNH123 - is op grootskaal getoets en volgens 'n model van Mendeliese oorerwing geinterpreteer. Die data was ontleed volgens genetiese struktuur en vlakke van genetiese variasie, die effek wat bestuur strategie in gevangeneskap op genetiese diversiteit in opeenvolgende generasies uitgeoefen het, so wel as die aard van die filogenetiese verhoudings wat teenwoordig is tussen die populasies.

"Exact" toetse is uitgevoer deur gebruik te maak van Monte Carlo tipe veelvuldige hermonsterinsamelings tegnieke en F-statistieke is gebruik vir die deteksie en kwantifisering van die genetiese struktuur tussen die twaalf populasies. Die Exact toets $\chi^2$ ($P < 0.001$), 'n $F_{ST}$ van 0.27 ($P < 0.001$), $\theta_{ST}$ van 0.26, $R_{ST}$ van 0.28, en 'n $\Phi_{ST}$ van 0.17 gee almal 'n indikasie van betekenisvolle strukturering tussen die populasies. Die genetiese struktuur bevestig die beleid dat die populasies behou moet word as aparte genetiese voorraad, in aparte tenke, om te verseker dat die unieke genetiese materiaal behoue bly om akwakultuur variante te ontwikkel.

Populasies het ook betekenisvolle verskuiwings van die Hardy Weinberg ekwilibrium getoon, wat gekarakteriseer word deur 'n algemene verlaging van heterosigositeit oor die loci. Nul allele word dikwels aanbeveel om in mikrosateliet studies groot bydraes te maak tot hetersigotiese defekte. Om vir nul allele te toets was twee gekontroleerde kruisings van *O. mossambicus* gemaak. Die nageslag van elke kruising was getoets vir verwagte ouer alleliese verhoudings by die UNH104, UNH111 en UNH123 loci. Al drie loci het getoon dat dit moontlike nul allele kan wees.

Versnelde inteling en genetiese drywing deur opeenvolgende generasies in gevangeneskap kan die heterosigositeit en diversiteit verminder. Om die vermindering van diversiteit te toets was 'n monster van die Busmans 1999 (N=25) populasie vergelyk met 'n monster van die Bushmans 2000 (N=36) populasie. Die vergelyking het veranderde alleel frekwensies, 'n betekenisvolle vermeerding in gemiddelde waargeneemde heterosigositeit en 'n onbetekenisvolle verandering in gemiddelde verwagte heterosigositeit getoon deur gebruik te maak van die UNH104 en UNH123 loci.

Berekening van genetiese afstande en filogenetiese vergelykings tussen die populasies het nuwe insig gegee oor die graad van bestuur wat nodig is om genetiese diversiteit in die natuurlike populasies van *O. mossambicus* tilapia te behou.UPGMA en Neighbour-Joining tegnieke was gebruik om filogenetiese bome op te stel deur gebruik te maak van $D_m$ en $(\delta\mu)^2$ afstand matrikse. Populasie bondeling het geblyk om geografiese lokaliteit van die bron populasies te toon, alhoewel

van die populasies nie met die geografiese lokaliteit ooreengestem het nie. Mantel toetse is gebruik om 'n moontlike assosiasie tussen genetiese afstand matrikse wat verkry is van elke loci bloot te stel. 'n Assosiasie sou 'n geografiese agtergrond tot populasie genetiese struktuur steun. Oortuigende bewys is nie deur die Mantel toetse verskaf nie. Mantel toetse vir assosiasie tussen die gekombineerde loci $D_m$ en $(\delta\mu)^2$ genetiese afstand matrikse en 'n geografiese afstand matriks was ook onbetekenisvol.

'n Tweedimensionele beskouing van die genetiese afstand data is voorgestel deur multi-dimensionele skaal (MDS) grafieke van Euclidean afstand waardes van die $D_m$ en $(\delta\mu)^2$ matrikse te teken. Die graad van ooreenstemming met die UPGMA en Neighbour-Joining boom samevoeging patroon was hoër vir die $(\delta\mu)^2$ as vir die $D_m$ MDS grafieke. Verspreiding grafieke het 'n vertroubare nie-liniêre korrelasie tussen Euclidean afstande en genetiese afstande vir die twee-dimensionele MDS grafieke getoon.

Die mikrosateliet merkers wat in die studie gebruik was het molekulêre informasie verskaf wat nodig is vir 'n komplimentêre studie oor die kwantitatiewe genetiese evalueering van dié twaalf populasies. Die kwantitatiewe studie het afmetings van gemiddelde lengte en gewig vermeerdering van die populasies verskaf gebaseer op nageslag proewe. Geen betekenisvolle korrelasie van gemiddelde hetersigositeit (geen diversiteit) was getoon met óf gemiddelde gewig óf lengte vermeerdering.

Die betekenisvolle genetiese diversiteit en struktuur teenwoordig tussen die twaalf populasies het rede gegee om strategieë te implimenteer om natuurlike *O. mossambicus* populasies te konserveer as genetiese bronne en om populasies in gevangeneskap te bestuur vir langtermyn instandhouding van genetiese diversiteit.

# TABLE OF CONTENTS

# LIST OF TABLES

## CHAPTER TWO



## CHAPTER THREE

# LIST OF FIGURES

## CHAPTER TWO



## CHAPTER THREE

## APPENDIX 1

**Figure 1**: Diagrams of *O. mossambicus* PCR amplified products for the UNH104 locus (2% agarose). Lane one contains a 100 base pair DNA ladder

**Figure 2**: Diagram of *O. mossambicus* PCR amplified products for the UNH111 locus (2% agarose). Lane one contains a 100 base pair DNA ladder

**Figure 3**: Diagrams of *O. mossambicus* PCR amplified products for the UNH123 locus (2% agarose). Lane one contains a 100 base pair DNA ladder

# ABBREVIATIONS

| | |
|---|---|
| A | Amatikulu |
| ABI 377 | Automated DNA sequencer 377 |
| AFLP | Amplified fragment length polymorphism |
| AMOVA | Analysis of molecular variance |
| ANOVA | Analysis of variance |
| B | Bushmans |
| bp | Base pairs |
| CM | Centimorgans |
| dCTP | Deoxycytosine triphosphate |
| DFID | Department for International Development |
| DNA | Deoxyribonucleic acid |
| DNTP | Deoxynucleotide triphosphate |
| E | Elands |
| EDTA | Ethylenediaminetetra-acetic acid |
| ESU | Evolutionary significant unit |
| FAO | Food and agriculture organisation |
| FdCTP | Fluorescent deoxycytosine triphosphate |
| GEU | Geminate evolutionary unit |
| H | Heterozygosity |
| HEST | *Haplochromis* ecology survey team |
| HWE | Hardy Weinberg equilibrium |
| *I* | Genetic identity |
| IAM | Infinite alleles model |
| K | Kasinthula |
| Kb | Kilobase |
| Km | Kilometres |
| L | Le Pommier |
| LVFS | Lake Victoria fisheries service |
| M | Makathini |
| MDS | Multi-dimensional scaling |
| $MgCl_2$ | Magnesium chloride |
| MT | Metric tons |
| MtDNA | Mitochondrial DNA |
| MU | Management unit |
| N | Ndumu |
| NJ | Nick James |
| NEJ | Neighbour-joining |
| O | Olifants |
| OTU | Operational taxonomic unit |
| *P* | Probability |
| PAGE | Polyacrylamide gel |
| PCR | Polymerase chain reaction |
| QTL | Quantitative trait loci |
| RAPD | Random amplified polymorphic DNA |
| RFLP | Restriction fragment length polymorphism |
| RPM | Revolutions per minute |
| S | Sucomba |
| SDS | Sodium dodecyl sulphate |
| SMM | Stepwise mutation model |
| SSM | Slipped strand mispairing |
| SSR | Simple sequence repeat |

| | |
|---|---|
| STS | Sequence tagged site |
| STR | Short tandem repeat |
| *Taq* | *Thermus aquaticus* |
| TPM | Two phase model |
| Tris-HCL | Tris[hydroxymethyl]amenomethane hydrochloric acid |
| UNH | University of New Hampshire |
| UPGMA | Unweighted pair group method using arithmetic averages |
| USDA | University of Stellenbosch Division of Aquaculture |
| V | Valley |
| VE | Verloerenvlei |
| VNTR | Variable number of tandem repeats |

**SYMBOLS**

| | |
|---|---|
| $\alpha$ | Mean number of alleles |
| °C | Degrees Celsius |
| *D* | Linkage disequilibrium coefficient |
| $D_m$ | Minimum unbiased genetic distance |
| $D_s$ | Standard genetic distance |
| $H_o$ | Observed heterozygosity |
| $H_e$ | Expected heterozygosity |
| $H_O$ | Observed subpopulation heterozygosity |
| $H_S$ | Expected subpopulation heterozygosity |
| $H_T$ | Total population heterozygosity |
| ng | Nanogram |
| $N_e$ | Effective population size |
| $N_m$ | Number of migrants per generation |
| $R_{sq}$ | Correlation coefficient |
| $\mu$ | Neutral mutation rate |
| µg | Microgram |
| µl | Microlitre |
| $(\delta\mu)^2$ | Delta-mu squared |
| $\chi^2$ | Chi-square |

# ACKNOWLEDGEMENTS

I would like to thank my family and friends for their support and patience throughout the practical and write-up of this thesis. I would like to thank the following people for their contributions towards the completion of this thesis: Dr Rupert Lewis, for his guidance with practical aspects and writing of this thesis, and for his honoured friendship. Mr. Danie Brink and Dr Louw Hoffman, for their financial support, practical guidance, and friendship. Dr Graham Mair and Professor John Beardmore for providing information and coordinating financial support for this study through DFID. Mr. Paul Marais, for his practical guidance and friendship. Mr. Carl Van Heerden and Miss Rene Van Zyl, for their assistance with ABI 377 fragment analysis. The USDA staff, for their assistance in managing the tilapia populations used in this study. Dr Armand Peeters, for the use of his laboratory and its facilities. Miss Bridgette Jacobs, for her assistance with the administration behind this thesis and her friendship.

# CHAPTER ONE

## *INTRODUCTION*

### 1.1 SIGNIFICANCE OF THE MOZAMBIQUE TILAPIA *Oreochromis mossambicus* (Peters - 1852)

The Mozambique tilapia is a member of the Cichlidae family, a family of freshwater and brackish water fishes found in Africa, Central and Southern America, India, Arabia and Madagascar. The African cichlids constitute about 870 species (Skelton, 1993), the majority of which occur in the African Great Lakes (Barel *et al.*, 1985). Members of the family are divided into two main lineages, the tilapiines and the haplochromines (Trewavas, 1983). The haplochromines are mainly predatory fishes, while the tilapiines live predominantly on plant matter or sediment. Cichlids have diverse breeding habits, which in most species involves pair-formation and guarding of offspring by one or both of the parents. *Tilapia* species are classified as substrate spawners as they lay their eggs on available substrate. The *Sarotherodon* and *Oreochromis* species are collectively known as mouth brooders. One of the parents picks up the fertilized eggs and incubates them in the mouth. *Oreochromis* species are entirely maternal mouth brooders while the *Sarotherodon* genus contains paternal and maternal mouth brooders (Skelton, 1993).

Cichlids are important in scientific research, commercial fisheries and aquaculture (Pullin, 1988). Scientific research on the cichlids encompasses the fields of taxonomy, biology and physiology of the African Great Lake cichlids in particular Lakes Victoria, Malawi and Tanganyika which have been extensively explored since the 1950's (Barel *et al.*, 1985; Worthington and Lowe-McConnell, 1994) The rapid diversification of the *Haplochromis* species has been used as a model to study evolution (Greenwood, 1951; Meyer *et al.*, 1990). Cichlids constitute a large proportion of the African annual total fishery of inland freshwater fishes (FAO Fisheries Statistic Database). In aquaculture, a number of cichlid species have been used with differing degrees of success. The *Oreochromis* species generally outperform species from the other genera; however, specific members of the *Sarotherodon* and *Tilapia* genera perform well under favourable conditions (Pullin, 1988).

One of the first *Oreochromis* species to be cultured as a food source, mainly in Asia (1940's and 1950's), was the Mozambique tilapia, *Oreochromis mossambicus* (de Moor and Bruton, 1988; Pullin, 1988; Agustin, 1999). This species has an uncanny ability to adapt to a diverse range of aquatic conditions (de Moor and Bruton, 1988) often resulting in establishment of feral populations where aquaculture is practiced. Indigenous to Southern Africa (Mozambique and South Africa), the species now occurs in 61 countries as either farmed or feral populations (Agustin, 1999). Feral establishment in some South East Asian countries has been so successful that *O. mossambicus* is

regarded as an aquatic pest (Beardmore *et al.*, 1997). In most Asian countries, *O. mossambicus* has subsequently hindered the local fisheries resulting in plans to eradicate unwanted populations, however, in Sri Lanka *O. mossambicus* actually improved the local fisheries (Pullin *et al.*, 1988). A number of aquaculture and conservation issues have arisen following the global introduction and distribution of *O. mossambicus* (Van der Bank and Ferreira, 1987; de Moor and Bruton, 1988; Pullin, 1988; Pullin *et al.*, 1988; Agustin, 1999).

The introduction into Asia was probably based on less than 20 fish. The earliest record of *O. mossambicus* in Asia was of five fish alleged to have escaped from an aquarium store in East Java, 1939 (Agustin, 1999). Four of these fish (three males and a single female) may have formed the founder population for the entire Asian *O. mossambicus* aquaculture industry (Pullin *et al.*, 1988). Agustin (1999) provides a detailed description of the introduction and distribution of *O. mossambicus* in tropical and subtropical countries outside its natural range and how the ancestry of many of these introduced populations can be traced back to the four original aquarium fish from East Java. The small number of originators, inherent founder effects (loss of genetic diversity) and inbreeding (mating amongst relatives), is probably responsible for the relatively poor performance of *O. mossambicus* in Asia. Reduced performance of Asian *O. mossambicus* strains created a negative reputation for *O. mossambicus* in Asia and other nations (Agustin, 1999). Current global tilapia production reflects a preference for the North African tilapia *Oreochromis niloticus* over *O. mossambicus*. Total world production of *O. niloticus* increased from 12000 tons in 1970 to 888000 tons in 1999. Total world production of *O. mossambicus* in 1970 was 1100 tons. Production peaked in 1995 at 58500 tons however subsequently decreased to 46000 tons in 1999 (FAO, 2001).

Fortunately wild populations of *O. mossambicus* tilapia still exist in their native range in Southern Africa (Van der Bank and Ferreira, 1987; Agustin, 1999). The potential of wild *O. mossambicus* from Southern Africa as a source of genetic material for developing and supplementing new aquaculture strains, was considered in 1988 as part of a review on wild tilapia genetic resources. Growth performance and appearance of wild Mozambique tilapia in Southern Africa were quoted as being superior to that of Asian introduced populations (Pullin, 1988). The implementation of a quantitative and population genetic evaluation of *O. mossambicus* at the University of Stellenbosch in 1998 aimed to provide a genetic component to the development of aquaculture strains and the conservation of natural Mozambique tilapia populations.

### 1.1.1 Aquaculture in South Africa – The Role of Tilapia

South African aquaculture production is a relatively underdeveloped industry (Hoffman *et al.*, 2000). Total production of aquaculture products in 1998 was 5301 metric tons (MT). In comparison, Africa as a whole produced 121 905 metric tons. Total world aquaculture production was 36 050 168 metric tons. In South Africa, rainbow trout dominated the annual production of freshwater fish (1650 MT). Ornamental Koi carp production made the second largest contribution (75 MT). Catfish (*Clarias gariepinus*) and tilapia species (e.g. *O. mossambicus*) were the closest rivals. A huge increase in catfish production in the late 1980's was challenged by marketing constraints, which reduced production from 1150 metric tons in 1991 to its current level of 40 metric tons. Total South African tilapia production mainly from small-scale commercial farms, was only 70 metric tons (45 MT *O. mossambicus*, 25 MT other tilapia species) (Hoffman, *et al.*, 2000) contributing little to world tilapia production for 1998 (1 559 967 MT) (FAO Fisheries Statistic Database, 1998 figures). The South African tilapia industry currently consists of commercial farming projects in the Northern and Western Cape Provinces, Mpumalanga, Northern Province and Kwazulu-Natal. Further development of the tilapia commercial sector is expected when knowledge of the production potential of *O. mossambicus* expands (Hoffman *et al*, 2000). Of the tilapias, *Oreochromis mossambicus* has received the most interest in aquaculture in Southern Africa due to its favourable growth compared to other tilapia species in the region (Jubb, 1967; Bruton, 1983; Ferreira and Schoonbee, 1983; Gaigher, 1983; Pike, 1983; McVeigh, 1995; McVeigh, 1998).

Early attempts at farming with tilapia in Southern Africa were hampered by technical and environmental factors. One of the main obstacles was the prolific breeding achieved through precocious maturity that occurred in dams stocked with both sexes leading to the production of small fish of little market value. The technology needed to breed monosex, all male, fry or sterile hybrids was not available or too expensive for the average farmer. Another limiting factor was the susceptibility of most tilapia species, excluding *Tilapia sparmanni*, to cold temperatures experienced in the Western Cape and the Highveld during winter (Jubb, 1967). *Tilapia sparmanni* has little value as an aquaculture species as it rarely grows larger than twenty centimeters (Skelton, 1993). The Department of Inland Fisheries, Stellenbosch, therefore spearheaded a search for cold tolerant species for aquaculture in the Western Cape. A number of tilapia species including *Tilapia aurea* (*Oreochromis aureus*), *Tilapia galilaea* (*Sarotherodon galilaeus*), and *Tilapia zillii* were imported from Israel in 1959 and reared at the Jonkershoek fish hatchery in Stellenbosch. Western Cape farm dams were stocked with tilapia offspring from the Jonkershoek species. A species survey of the dams, in 1964, indicated that *O. aureus* and *S. galilaeus* had survived the cold winters.

*Oreochromis mossambicus* was introduced into Western Cape farm dams and coastal lagoons by the Cape Nature Conservation Department (the Western Cape governmental agency responsible for conservation) in the 1960's (de Moor and Bruton, 1988). The De Hoop Vlei and Zeekoe Vlei brackish coastal lagoons both contained established populations of *O. mossambicus* by 1966. The salinity of the lagoons was often as high as five parts per thousand (Jubb, 1967). In 1998, Stellenbosch University Division of Aquaculture (USDA), in collaboration with University of Wales (Swansea), began developing the technology required to solve the problems surrounding *O. mossambicus* tilapia farming in Southern Africa. The technology developed for aquaculture of *O. niloticus* was adapted for use on *O. mossambicus*. This included research into nutrition, recirculation technology, rearing in tunnel enclosed ground dams, and development of monosex fry. A small-scale tilapia farming co-operative was established by encouraging local farm labourers to rear caged tilapia in existing wine farm dams.

### 1.1.2 Conservation of *O. mossambicus* in South Africa

Artificial introductions of exotic fish, including tilapia, for fisheries or aquaculture have caused ecological disruptions through habitat loss and extinction of endemic species (Pullin, 1988). Lake Victoria offers disturbing examples of biodiversity disruptions possibly caused through introduction of exotic fish species coinciding with or activating other environmental disruptions (Barel *et al.*, 1985; Ogutuohwayo and Hecky, 1991; Lowe-McConnell, 1997). *Lates niloticus*, a voracious predator that can grow to a large size, was suggested as a suitable food species for introduction into Lake Victoria in the 1950's to supplement a declining natural fishery (Barel *et al.*, 1985). Initially, *L. niloticus* introductions occurred through accidental escape from farm ponds neighbouring the lake. The Lake Victoria Fisheries Service (LVFS) (Lowe-McConnell, 1997) deliberately made subsequent introductions. The devastating consequences for indigenous Lake Victoria species have been reviewed by Barel *et al.* (1985), Witte *et al.* (1992) and Lowe-McConnell (1997). Investigations indicated that fisheries landings subsequent to introduction in some parts of the lake consisted almost entirely of the introduced *L. niloticus*. The yearly fishery landings for Lake Victoria increased from in the region of 100 000 tons, prior to *L. niloticus* introduction, to 561 000 tons in 1990 (Nile perch constituting most of the catch) (Lowe-McConnell, 1997). The decline in catches of indigenous species was attributed to heavy predation by *L. niloticus*. Surveys of the lake indicated that more than 260 species, including 200 of the 300 documented cichlids, had disappeared within three decades post stocking of the Nile Perch (Witte *et al.*, 1992; Lowe-McConnell, 1997). This was considered nothing short of disastrous for both the local communities that relied on endemic species as a preferred food source and the scientific community studying the biology and taxonomy of the indigenous fishes (Lowe-McConnell, 1997).

The *Haplochromis* Ecology Survey Team (HEST) implicated *L. niloticus* as a key figure in the decline of Lake Victoria's faunal diversity (Lowe-McConnell, 1997). Extensive eutrofication, algal blooms and zones of reduced dissolved oxygen prominent in various parts of Lake Victoria during the 1950's and 1960's (Ogutuohwayo and Hecky, 1991; Lowe-McConnell, 1997) have been implicated as possible contributing factors to the loss of trophic diversity and the transformation of fish communities (Ogutuohwayo and Hecky, 1991). The influence of pollution and exotic species introduction on the Great Lakes has been extensively researched, however, many of the problems await further investigation (Worthington and Lowe-McConnell, 1994).

The introduction of *L. niloticus* in the 1950's coincided with the introduction of a number of tilapia species. The tilapias, *O. niloticus*, and *Tilapia zillii* were introduced into Lake Victoria in 1953 by the LVFS in an attempt to enhance the existing tilapia fishery (Lowe-McConnell, 1997). Introductions of *O. leucostictus* and *T. melanopleura* were also made in the 1950's and early 1960's (Ogutuohwayo and Hecky, 1991). Prior to introduction the tilapia fishery consisted mainly of the endemic *O. esculentus* and *O. variabilis*. *Oreochromis niloticus* has subsequently become the dominant tilapia in the lake having hybridised with and displaced most of the indigenous tilapias (Pullin, 1988; Lowe-McConnell, 1997).

The adept colonization capabilities of many tilapia species are exemplified by *O. mossambicus*, which has been described as one of the most phenotypically plastic extant species (de Moor and Bruton, 1988). The Asian feral populations of *O. mossambicus* bare sentiment to this (Van der Bank *et al.*, 1989; Agustin, 1999). The remarkable ability to adapt and proliferate in a multitude of environmental conditions would suggest that *O. mossambicus* is not a threatened species. The species is not listed in the South African Red Data book of endangered fishes (Skelton, 1987); however, this does not exclude *O. mossambicus* from requiring conservation and management, particularly at the population level. There are a number of conservation issues surrounding *O. mossambicus* in southern Africa (de Moor and Bruton, 1988).

#### 1.1.2.1 Translocations

The history of *O. mossambicus* introductions in South Africa is documented from the early 1950's to the 1980's (de Moor and Bruton, 1988). There have been few studies on the impact that these introductions have had on local aquatic biodiversity (de Moor and Bruton, 1988) despite the insinuation of *O. mossambicus* in the damaging of many international watercourses (Trewavas, 1983). *O. mossambicus* of Transvaal origin were introduced into farm dams in Malmesbury (Cape Province) in 1936. Some of these fish were transferred to the Jonkershoek Hatchery to act as breeding stock for further distribution. A number of other hatcheries took part in the breeding and distribution of *O. mossambicus* in other parts of South Africa: the Umgeni Warmwater Hatchery

(Kwazulu Natal), the Lydenburg Hatchery (Mpumalanga), the Amalinda Hatchery (East London) and the Lowveld Fisheries Research Station (Mpumalanga). The local translocation of Mozambique tilapia into areas outside its natural range was considered more of a disturbance to natural aquatic ecosystems than an advantage (de Moor and Bruton, 1988).

Translocation in South Africa took place mainly for utlisation in aquaculture, recreational fishing, and biological control of chironomids and macrophytes (de Moor and Bruton, 1988). Evidence transpired that *O. mossambicus* introductions impacted negatively on endemic fish in Kwazulu Natal, Eastern Cape and Western Cape Rivers (Crass, 1969; Gaigher *el al.*, 1980). Translocation of cultured or non-local *O. mossambicus* populations was also described as a threat to the genetic diversity present in local wild *O. mossambicus* populations (Pike, 1983; de Moor and Bruton, 1988). Genetic diversity in wild populations was stated as being worthy of protection due to unique qualities such as cold tolerance and high salt concentration tolerance characteristic of certain populations (de Moor and Bruton, 1988).

### 1.1.2.2 Hybridisation and Displacement

The introduction into South Africa of the popular *O. niloticus* species that performs well in aquaculture in other countries represents a potential threat to natural tilapia populations, since most tilapia species readily hybridise (Wohlfarth, 1994; Van der Waal and Bills, 2000). In a similar manner, *O. mossambicus* represents a threat to tilapia species in areas outside its natural range. Two inhabitants of the Okavango delta, *O. macrochir* and *O. andersonii* can interbreed with *O. mossambicus*. Flooding of neighbouring pans, in which *O. mossambicus* already occurs, constitutes a route for introduction into the Okavango river (de Moor and Bruton, 1988). Hybridisation is more likely between tilapia with close genetic relationships. Starch gel electrophoresis of 24 allozyme loci provided evidence of a close relationship between *O. mossambicus* and *O. niloticus* (Pouyad and Agnese, 1995).

Present South African conservation laws restrict translocation of indigenous fish and introduction of exotic fish into natural water systems. One of the biggest potential threats to the genetic integrity of *O. mossambicus* populations, however, arises from the existence of exotic populations of *O. niloticus* within the natural range of *O. mossambicus* (de Moor and Bruton, 1988; Van der Waal, 2000). Indeed, Moralee and Van der Waal (2000) have already presented morphological and biochemical evidence of wild $F_1$ hybrids between the two species in the Limpopo river system. Ironically, in the 1950's, *O. niloticus* was imported into and distributed in South Africa by local companies and government agencies (reviewed by de Moor and Bruton, 1988). In 1955, the Cape Nature Conservation Department established a population of *O. niloticus* at the Jonkershoek Hatchery for breeding experiments. In 1978, the Natal Fisheries Development Corporation stocked

a small dam in Northern Kwazulu Natal with a population of *O. niloticus* from the Amatikulu Hatchery. The population was later destroyed. The Tongaat Sugar Company introduced *O. niloticus* into the Dudley Pringle dam (Wewe river catchment in Kwazulu Natal) in 1982. The Stellenbosch University Aquaculture Division is, currently investigating DNA markers, to characterise parental and hybrid genotypes in *O. mossambicus* and *O. niloticus*.

#### 1.1.2.3 Habitat Loss

*Oreochromis mossambicus* forms part of the rich fish diversity found in South African aquatic habitats (Skelton, 1993; Van der Waal, 2000). Habitat loss or fragmentation occurs through interventions on natural water systems for power, transport, agriculture and water consumption. The effects of human encroachment, on aquatic habitats are often detrimental, leading to the extinction of aquatic species (Crass, 1969; Avault, 1994). Aquaculture also constitutes a form of encroachment. In the vast majority of cases of habitat loss, the evidence suggest that species diversity has reduced, often dramatically so (Lande, 1998). Shrimp and whitefish culture induced almost fifty percent of mangrove estuary destruction in the Philippines. The direct consequence was a reduction in species diversity in the mangrove forests (Beardmore *et al.*, 1997).

#### 1.1.2.4 Loss of Genetic Diversity through Genetic Interactions

Genetic diversity within populations represents a fundamental level of biological diversity (Biodiversity) from which more commonly utilised estimates of biological diversity (such as numbers of species, families, and phyla) arise (Harper and Hawksworth, 1994, Beardmore *et al.*, 1997). Conservation genetics generally deals directly with the genetic diversity present in populations, how this affects the fitness of the populations and their demographic history (Beaumont, 2001). The conservation of genetic diversity in wild fish populations has been frequently discussed (Adkison, 1995; McConnell *et al.* 1995b; Garcia De Leon *et al.*, 1997; Small *et al.*, 1998; Heist and Gold, 1999).

Genetic factors that can reduce genetic diversity in wild populations include founder events, bottlenecks and artificial supplementation with large numbers of organisms with low overall genetic diversity (Ryman *et al.*, 1995; Ryman, 1997). Agustin (1999) provided evidence of dramatically reduced genetic variation, through founder effects and bottlenecks, in four feral *O. mossambicus* stocks in the Australasian Pacific region when compared to wild populations from the natural range of the species in South Africa. A single mitochondrial DNA (mtDNA) haplotype was found in all the feral stocks. Genetic diversity for mtDNA was reported as being highly diverse between eight wild South African populations that contained 26 different haplotypes. Microsatellite diversity was also considerably reduced in feral stocks (Agustin, 1999).

De Moor and Bruton (1988) cautioned that the translocation of Mozambique tilapia within South Africa could jeopardize the unique genetic variation present in the wild populations of *O. mossambicus*. If this occurred on a large scale, the influx of non-locally adapted genes could reduce the fitness in the wild populations. This phenomenon – known as outbreeding depression - has been demonstrated through simulation studies (Yokota and Watanabe, 1997) and also on wild salmon populations (Kruegar *et al.,* 1981, Thorpe *et al.,* 1981), where the original intention was to supplement fish numbers (see review by Allendorf *et al.,* 1987; Fleming and Gross, 1993; Jones *et al.*, 1996; Tessier *et al.*, 1997).

### 1.1.3 The Natural Distribution of *Oreochromis mossambicus*

Natural populations of *Oreochromis mossambicus* occur from the lower regions of the Shire and Zambezi rivers to the Bushmans river in the Eastern Cape. Populations are predominantly found in East flowing coastal rivers (Jubb, 1967; Trewavas, 1983; de Moor and Bruton, 1988; Skelton, 1993). A temperature cline (12°C) and fewer accessible estuaries prevent the natural distribution of Mozambique tilapia spreading further south or inland. Introduced populations exist in farm dams in parts of the Western Cape and the Northern Cape (Skelton, 1993).

The salt tolerance of *O. mossambicus* enables it to survive periodic exposure to seawater. During heavy flooding, riverine fish are often washed out of estuary mouths into the sea. The salt intolerant fish succumb to the high salt concentration of seawater (Jubb, 1967). The two predominant sea current systems that operate off the East Coast of Southern Africa, the Mozambique and the Benguela (Pearce, 1977), can move the salt tolerant (euryhaline) *O. mossambicus* southward. Migration to more southern rivers is thought to occur when seagoing *O. mossambicus* manage to enter more southern located estuaries whereupon they swim up to the calmer parts of a river and breed with existing *O. mossambicus* or establish completely new populations (Jubb, 1967; Trewavas, 1983).

## 1.2 INTRODUCTION - MOLECULAR GENETIC RESEARCH IN TILAPIA

The comprehensive characterization of genetic variation in the tilapias includes the use of morphometric characters, proteins (in particular allozymes), mitochondrial DNA (mtDNA) and nuclear DNA. These markers have been used to assess population and species level genetic variation for tilapias. Each method has noticeable advantages and disadvantages.

### 1.2.1 Morphometric Markers in Tilapia

Morphometric trait data constitute a vast collection of systematic information. Common morphological features used to distinguish between species are number of gill rakers, number of dorsal or anal fin spines and characteristic colours and markings (Trewavas, 1983). Renowned works on tilapia morphometrics include research done on tilapia species in their natural African ranges (Trewavas, 1983; Teugels and Thys van den Audenaerde, 1992). There are two main limitations in using morphometric characters in taxonomy. Firstly variation in such characters is assumed to reflect only heritable factors, whereas environmental influences are often influential (Van der Bank, *et al.*, 1989; Falk *et al.*, 1996); secondly the highly variable rate of evolution of morphometric characters – particularly those under selection – imposes bias on genetic distance estimates. Morphometric data, however, has necessarily formed the historical backbone to taxonomy, and provides the calibration points for more recent molecular tools (Hillis, 1987).

### 1.2.2 Allozyme Markers in Tilapia

Allozymes probably provide the most rapid method for obtaining data at the molecular level on genetic variation within and between populations or species. Electrophoretic separation detects the charge differences between proteins that arise through some mutations in enzyme encoding loci.

Proteins are products of functional genes, which therefore may be under selection. Interpretation of allozyme variation for the purpose of population genetics, however, relies on the premise that the distribution of alleles among populations is not under selection; i.e. that allozyme variation is selectively neutral. This dichotomy formed the basis of the selectionist/neutralist debate, which has been investigated extensively (Lewontin, 1974; Lewontin, 1991; Avise, 1994). The general consensus of the selectionist/neutralist debate was that in the absence of convincing exclusive evidence, the neutral model as the most parsimonious explanation, was a logical framework for interpretation, and that even under most commonly applied models of selection, the validity of a neutral interpretation was not undermined (see review by Skibinski and Ward, 1998).

Allozymes have been widely used for identifying stocks of tilapia used in aquaculture (McAndrew and Majumdar, 1983), detecting gene introgression between and within aquaculture species (Macaranas *et al.*, 1986; De Silva, 1997) and answering population genetic and systematic questions by providing phylogenetic information for cultured and wild tilapia populations (Van der Bank *et al.*, 1989; Van der Bank, 1994; Pouyaud and Agnese, 1995; Rognon *et al.*, 1996; Vreven *et al.*, 1998). Aside from questions regarding the neutrality of allozymes, their main limitation is the requirement for tissue to remain frozen during transportation and storage. The recent proliferation of DNA techniques whereby tissue can simply be preserved in alcohol (Zhang and Hewitt, 1998) has rendered the logistics of allozyme techniques less attractive.

### 1.2.3 Mitochondrial DNA Markers in Tilapia

Mitochondrial DNA population and species research explores different regions of the mitochondrial genome depending on the level of discrimination required (Kocher and Stepien, 1997). In addition to the D-Loop non-coding region, there are generally around thirteen protein, twenty-two transfer RNA and two ribosomal RNA coding regions in the mtDNA molecule. Variation in these regions includes nucleotide substitution length heteroplasmy (Buroker *et al.*, 1990) and site heteroplasmy (Skibinski *et al.,* 1994) The D-Loop region has higher levels of genetic variation than that of coding regions, a useful attribute in population comparisons (Hillis *et al.*, 1996; Hartl and Clark, 1997). The more conserved coding regions, like the ND5/6 region and Cytochrome B region are useful for species identification, systematics and evolutionary research due to the lower levels of variation (Kocher and Stepien, 1997).

Universal primers have been developed from the mtDNA genome regions of several organisms (Avise, 1994). These primers often amplify analogous regions in related species (Kocher and White, 1989; Bernatchez *et al.*, 1992; Hillis *et al.*, 1996). Tilapia species and population level questions have been addressed using some of these universal mtDNA primers (Rognon and Guyomard, 1997; Agustin, 1999). Agustin (1999) constructed phylogenetic trees, from D–loop mitochondrial haplotype matrices for wild populations of *O. mossambicus* in Southern Africa. Agustin's work represents a preliminary attempt at divulging genetic relatedness and demographic history of wild *O. mossambicus* populations in Southern Africa. The range of *O. mossambicus* distribution covered was stated as being incomplete. Further sampling and analysis of populations from other regions within Southern Africa was suggested (Agustin, 1999).

### 1.2.4. Nuclear DNA Markers in Tilapia

Methods used to investigate genetic questions in tilapia have tended to follow the history of technique development. Early research involved traditional cytogenetic techniques to determine the diploid number (mostly 44) for several *Oreochromis* species including *O. mossambicus* (Majumdar and McAndrew, 1986). With the discovery of large regions of repetitive DNA in the tilapia genome, the first attempts at species comparisons were made using satellite DNA and fingerprinting methods (Wright, 1989; Franck *et al.*, 1992). The Random Amplified Polymorphic DNA (RAPD) technique has also been used for species comparisons (Bardakci and Skibinski, 1994), as have minisatellite (VNTR) techniques (Naish *et al.*, 1995; John *et al.*, 1996).

Recent interest in utilising genetic markers the presence of which is correlated to genes of quantitative performance value has led to Quantitative Trait Loci (QTL) approaches which are most efficiently pursued with a high number of markers such as are typically generated with linkage maps. The Kocher group at the University of New Hampshire have spearheaded this approach in Tilapia with the development of a genetic linkage map for *O. niloticus* using microsatellite and Amplified Fragment Length Polymorphisms (AFLP) markers. The *O. niloticus* linkage map consists of 30 linkage groups encompassing the 22 chromosomes of the haploid genome. The linkage groups contain 62 microsatellite sequence tagged sites (STS) and 112 AFLP loci. Using the distances calculated between the STS and AFLP markers a total map length of 1000 – 1200 cM was calculated for *O. niloticus* (Lee and Kocher, 1996; Kocher *et al.*, 1998).

The groundwork carried out by the Kocher group and previous researchers enabled the initiation of a number of linkage, population and paternity genetic studies on tilapia and related cichlid species (Harris *et al*, 1991; Kellog *et al.*, 1995; Naish *et al*, 1995; Knight *et al.*, 1998; Agustin, 1999; Agresti *et al.*, 2000). Many of the studies made use of repetitive sequences known as satellite (300 – 6000 base pairs), minisatellite (9 – 65 base pairs) or microsatellite (2 – 5 base pairs) DNA, depending on the size of the nucleotide repeats (Wright, 1989; Harris and Wright, 1995; Orti *et al.*, 1997). Of the three repetitive DNA marker types available, microsatellites have the most valuable qualities for use in fish population genetics since their relatively high mutation rates offer a high level of resolution most useful for population studies (O'Reilly and Wright, 1995). Broad conclusions from the research on microsatellite loci describe microsatellites as very useful for intraspecies studies involving populations that have not become too genetically diverged (Takazaki and Nei, 1996). However, microsatellite use relies on the progresses made in the techniques used to analyse other nuclear DNA regions, in particular, the Polymerase Chain Reaction (PCR), sequencing and cloning techniques.

## 1.3 BACKGROUND TO POPULATION GENETIC APPROACHES

### 1.3.1 Mechanisms of Genetic Change

Population genetic research is made possible by mechanisms that create genetic diversity and that lead to changes in allele frequencies between populations (reviewed extensively by Lewontin (1974), Avise (1994), Hillis *et al.* (1996), Hartl and Clark (1997) and Hedrick (2000)). The theories behind these mechanisms in particular the Wright-Fisher theory of random genetic drift formulated by Fisher (1930) and Wright (1931), and Haldane's mathematical models of selection (reviewed by Sarkar, 1992) form the foundation of population genetics.

The two mechanisms that introduce new variation are mutation within and migration between populations. Mutation can create new alleles at a locus while migration allows for movement of the alleles between populations (gene flow) resulting in frequency changes. The two mechanisms that reduce genetic variation are genetic drift (random loss of alleles) and directional selection, which will lead to rapid fixation. Genetic drift leads to a reduction in the number of alleles present at a locus and its magnitude is inversely proportional to the effective population size $N_e$ (size of the breeding population if every individual was unrelated and had an equal probability of passing on offspring to the next generation). Only the individuals that survive and breed pass their alleles on to the next generation. Genetic drift has the potential to reduce the number of alleles at a locus to one (allele fixation). This requires no immigration of new alleles and no mutation. Under the neutral theory then (and in the absence of immigration/emigration), the level of polymorphism in a population generally reflects a balance between allele gain through mutation, and allele loss through genetic drift. Since the mutation rate is generally considered to be constant for a given gene class, and the rate of drift is inversely proportional to $N_e$, levels of genetic variation are expected to be greater in species with larger $N_e$. General reviews support this expectation (e.g. Nevo, 1978).

Early attempts to solve the nature of genetic variation present in populations and the role of natural selection in evolution of a species, led to the formation of two hypotheses (reviewed by Lewontin, 1974). The first, which became known as the Classical hypothesis, stated that individuals from a population were homozygous for wild-type alleles at most loci with very few loci heterozygous for a deleterious mutant gene. The number of loci carrying deleterious mutant genes was deemed the 'mutation load'. Under the Classical theory natural selection was seen as a purifying agent that removes deleterious genes and occasionally fixes advantageous alleles through selective sweeps. The alternate Balance hypothesis stated that individuals from a population were predominantly heterozygous at all loci with no particular wild-type allele at a locus. Homozygous loci exist in different individuals with different alleles, although some loci can be fixed for a single allele among all individuals. Early researchers had no method to test these hypotheses apart from observing

phenotypic ratios in breeding experiments (Lewontin, 1974; Lewontin, 1991). The introduction of the technique for gel electrophoresis of proteins (Harris, 1966; Hubby and Lewontin, 1966; Lewontin and Hubby, 1966) provided, a direct method of assessing the molecular genetic variation between species and populations (Lewontin, 1991).

Aware of the deleterious nature of many visibly detectable mutations in *Drosophila* studies, and the allele fixing nature of directional selection, early population geneticists believed molecular techniques would reveal a lack of allozyme variation within populations. In contrast, early molecular genetic research revealed a great deal of diversity for allozyme markers resulting in controversy over the mechanisms that maintain this variation in populations (Hartl and Clarke, 1997). Initially the traditionalists came up with the concept that the allozyme variation detected was of no functional significance, and so was selectively neutral – the so-called 'neutral' theory (Kimura, 1968; King and Jukes, 1969). This was countered by the idea that since directional selection tended to reduce genetic variation, selection for heterozygosity – so called balancing selection could act to maintain polymorphism (Levene, 1953; Dempster, 1955; Kimura, 1955; Haldane and Jayakar, 1963). Whilst there are many examples of genotype/environment correlations for which selection is a reasonable explanation (see review by Koehn and Hilbish, 1987), this is not in conflict with modified versions of the neutral model, such as the nearly neutral model (Ohta, 1992). The fact that both broad models can be demonstrated to fit the data has left something of a lack of clear resolution to this debate. Despite this, however, most researchers accept the neutral model, for reasons of parsimony (the model makes fewer assumptions), and in most cases for lack of specific evidence to the contrary.

#### 1.3.1.1 Balancing Selection

The concept of balancing selection has many forms based on attempts to try to explain the causes of high levels of polymorphism. One theory is that selection favours heterozygotes leading to an excess of heterozygous individuals, increasing the diversity present in a population through single locus heterosis. The theory is reliant on over-dominance at individual genotypes being a balancing mechanism in populations (Lewontin, 1974). Other forms of balancing selection that have been suggested are frequency-dependent selection (e.g. Cain, 1988) or that selection over time and space is responsible. This is encompassed by Wallace's (1968) theory of marginal over dominance. These and many other forms have been used to try to explain the theory of balancing selection (King, 1967; Milkman, 1967; Sved *et al.*, 1967; Lewontin, 1974; Hedrick, 2000) and support the 'selectionist' view of protein polymorphism.

### 1.3.1.2 Neutrality

The alternative theory put forward was that of neutrality (Kimura, 1968; King and Jukes, 1969). This theory predicts that most observed allelic variation is of no functional significance (i.e. the genetic variation is selectively neutral), and that the level of allelic variation present in a population at equilibrium is a balance between mutation and genetic drift alone. Mutation creates new alleles – those under directional selection are either rapidly fixed or rapidly eliminated, and so don't play a role in maintaining polymorphism. New alleles not under selection change in frequency by chance, eventually becoming fixed or lost by genetic drift. The rate at which new alleles become fixed in a population is equal to the neutral mutation rate ($\mu$) so that at equilibrium, mutation and genetic drift are balanced in a population.

The neutral theory provides the framework for population genetic analyses. Two populations will diverge in gene frequencies by chance alone, the divergence being countered by migration among populations. Differences in gene frequencies among populations then provide the basis for quantifying the rate of exchange and hence allow populations to be classified according to an array of genetic proximity defined by exchange rates.

### 1.3.2 Microsatellites, Effective Markers for Population Genetics

Population geneticists make use of markers that are genetically diverse and assumed to be neutral (Hartl and Clarke, 1997; Hedrick, 2000). Microsatellite DNA is particularly suited to population genetics since sequences are assumed to be non-coding, and therefore presumed to be unlikely to be under direct selection (Goldstein and Schlotterer, 1999). Microsatellite motifs are also very common, widely dispersed throughout the genome, and highly polymorphic (Edwards *et al.*, 1991; Goldstein and Schlotterer, 1999). This latter feature arises since microsatellite or Simple Sequence Repeat (SSR) DNA mutates at a faster rate ($1 \times 10^{-4}$ per site per generation) than that of coding DNA ($1 \times 10^{-9}$) (Levinson and Gutman, 1987; Henderson and Petes, 1992). The faster mutation rate and co-dominant inheritance of microsatellite DNA makes it suitable for population level comparisons. The differences present between populations are the result of genetic drift, mutation and migration (O'Reilly and Wright, 1995; Slatkin, 1995; Hartl and Clark, 1997; Goldstein and Schlotterer, 1999; Hedrick, 2000). The effective use of microsatellites as population genetic markers requires understanding the microsatellite mutation mechanism, the mutation models that operate at microsatellite loci and techniques required for microsatellite application.

### 1.3.2.1 Microsatellite Mutation Mechanism

The mechanism that leads to mutation at microsatellite loci is known as Slipped Strand Miss-pairing (SSM). During replication, slippage of the newly synthesized DNA strand in a 3' to 5' direction, results in an unpaired complimentary repeat unit in the template strand. As replication continues, DNA polymerase adds an extra repeat unit (motif) opposite to the unpaired complimentary template unit, increasing the length of the new microsatellite strand by one motif. If slippage occurs in the newly synthesized strand in a 5' to 3' direction then one repeat unit overshoots the complimentary region in the template strand. As replication continues, DNA polymerase locates the apparent extra unit and removes it resulting in a decreased microsatellite length for the new strand (Levinson and Gutman, 1987; Henderson and Petes, 1992; Schlotterer and Tautz, 1992). Additional mechanisms that have been proposed as mutation mechanisms for microsatellite loci are unequal sister chromatid exchange and genetic recombination (Shriver *et al.*, 1993).

### 1.3.2.2 Microsatellite Mutation Models

Two models of allele mutation have been adapted to explain microsatellite evolution, the Infinite Alleles Model (IAM) (Kimura and Crow, 1964) and the Stepwise Mutation Model (SMM) (Kimura and Weiss, 1964; Ohta and Kimura, 1973). The IAM states that when a mutation occurs at a locus the newly created allele takes on a unique electrophoretic conformation independent of the allele from which it was formed. In this way, an infinite number of alleles can exist for a particular locus. The SMM states that new alleles are created at a locus in a stepwise fashion. A new allele carries information about the allele from which it was formed (Ohta and Kimura, 1973). The SMM, originally developed for allozyme markers, conveniently finds application in the use of microsatellite markers since changes in microsatellite allele repeat size are thought to occur predominantly in a stepwise fashion (Rousset, 1996; Goldstein and Pollock, 1997; Goldstein and Schlotterer, 1999). Repeat size can increase or decrease in single increments of the repeat motif. There appears to be a limit to size of microsatellites and a bias towards increased length as apposed to decreased length however no direct evidence exists to validate these constraints (Bowcock *et al.*, 1994; Garza *et al.*, 1995; Goldstein and Schlotterer, 1999).

There has been some criticism of the single-step, stepwise mutation model proposed for microsatellite DNA. Microsatellite loci have been found that exhibited an allelic mutational nature more like the IAM than SMM (Shriver *et al.*, 1993; Valdes *et al.*, 1993; Di Rienzo *et al.*, 1994). This is particularly true of microsatellites with repeat motifs of two base pairs (di-nucleotide repeats). Simulation studies confirmed this showing that Short Tandem Repeat (STR) DNA with repeat motifs of three to five base pairs followed the SMM model very well (100%) however only

65% of the one to two base pair microsatellite repeats followed a strict SMM mode of evolution. The study also included minisatellite loci (15 to 70 base pair repeats). Seventy three percent of the minisatellite loci did not conform to the SMM. The reduced conformity of microsatellite dinucleotide loci to a one-step SMM was likened to the presence of a low frequency of multi-step mutations resulting in a more IAM affiliated allele distribution (Shriver *et al.*, 1993). SMM is applicable to microsatellite loci but caution must be taken in using it as the sole mechanism behind allelic change (Shriver *et al.*, 1993; Valdes *et al.*, 1993; Di Rienzo *et al.*, 1994). Di Rienzo *et al.* (1994) suggested a model explaining the nature of mutation at di-nucleotide microsatellite loci known as the Two Phase Model (TPM). This model proposes stepwise mutations as the main mechanism of di-nucleotide microsatellite change but also incorporates the possibility of larger mutations in accordance with the IAM model.

### 1.3.3.3 Techniques for Microsatellite Application

The single DNA technique that makes microsatellite analysis so popular for population comparison is the Polymerase Chain Reaction (PCR) (Mullis *et al.*, 1986; Mullis, 1990; Hillis *et al.*, 1996). PCR revolutionized the use of DNA markers in all fields of genetic research (Erlich and Arnheim, 1992). Developments in PCR technology include the amplification of increasingly longer lengths of DNA sequence and the amplification of DNA from miniscule amounts of template DNA. Ultimately it is thought that modified PCR will allow for amplification of entire genomes (Brown, 1997) – indeed this is the objective of genome projects in *E. coli*, humans, mice, yeast, and other organisms (Genome Web – Other Vertebrate Genome Database). The small size of microsatellite sequences enables rapid amplification and analysis, using PCR based methods. Multiplexing (the use of multiple primer sets in one PCR reaction, or the use of simultaneous PCR products in sequencers) and fragment analysis using automated DNA sequencers allows for a high throughput of information (Goldstein and Schlotterer, 1999)

The use of microsatellite DNA as a molecular marker requires preliminary research, outlined by Hammond *et al.* (1998), to develop primers flanking the region containing the microsatellite DNA repeat. There are two approaches to identifying microsatellite repeat regions present within the genome of an animal. The quickest method is to screen GENBANK for microsatellite DNA sequences that have previously been submitted by other researchers working on the organism of concern or a closely related organism. In the absence of existing GENBANK sequence information a cloning technique is employed. Genomic clones are constructed containing short sequences of the genome from the animal for which primers are required. Some of the clones contain DNA with microsatellite repeat regions. These clones are identified using probes homologous to the DNA repeat. The clones are then sequenced and forward and reverse primers for the DNA regions

flanking the microsatellite repeat are designed. Methods have now been developed to render microsatellite cloning techniques more efficient by the use of specific enrichment protocols (Paetkau, 1999).

A problem that often accompanies the use of microsatellites as genetic markers is the presence of stutter bands. So called stutter bands are thought to be generated during PCR by slipped strand mispairing creating additional PCR products that differ in size by one or two repeat units from the true PCR product. They are visible on electrophoresis gels as bands that precede or very rarely trail the true microsatellite band. Dinucleotide microsatellite repeats are more prone to stuttering than trinucleotide repeats. Stutter bands stain less intensely than the true bands however when analysing PCR products of mixed DNA samples identification of the true PCR products can be problematic due to the superimposition of bands (Schlotterer and Tautz, 1992; Goldstein and Schlotterer, 1999). In most cases these can be resolved by close examination of the peak sizes and logical inference. Further problems accompanying microsatellite data include the presence of null alleles (probably due to mutations in a priming site resulting in non-amplification of a specific allele) which can result in an apparent reduced heterozygosity (Callen *et al.*, 1993; Paetkau and Strobeck, 1995; Pemberton *et al.*, 1995) and single base pair shifts in allele size (Ginot *et al.*, 1996; Browstein *et al.*, 1996).

Primers obtained for a particular species are often used for other species due to analogous priming sites (microsatellite flanking regions) (Rico *et al.*, 1996). A comparison done on fish microsatellite flanking regions for 18 microsatellite loci in distantly related taxa revealed evidence of a level of conservation in the region of 470 million years. The finding suggests that there are numerous analogous priming sites between fish species making for a potentially large resource of polymorphic markers for population genetic work. Levels of polymorphism for the microsatellites were higher for species related to the species from which the primers were developed (Rico *et al.*, 1996). The between species conservation of *O. niloticus* microsatellite priming sites enabled construction of a partial linkage map for *O. aureus* using *O. niloticus* primers developed for construction of the *O. niloticus* linkage map (McConnell *et al.*, 2000). In this study, we used primers from the *O. niloticus* linkage map (Kocher *et al.*, 1998) to amplify microsatellite loci in the closely related *O. mossambicus*.

### 1.3.3 Microsatellite Application

Countless research works have applied microsatellites to aquaculture management and fish population conservation. Microsatellites are suitable for use in many different aquaculture applications from strain identification, parentage analysis to mixed stock rearing (O'Reilly and Wright, 1995). Microsatellites have been used to determine basic levels of variation found in hatchery stocks of rainbow trout (Nielsen *et al.*, 1997; Nielsen and Fountain, 1999). The genetic identity of cultured fish in conjunction with growth performance has been monitored in African catfish (*Clarias gariepinus*) (Volckaert and Hellemans, 1999) using microsatellites isolated from the *C. gariepinus* genome (Galbusera *et al.*, 1996).

Aquaculture has also benefited from the application of microsatellites in parentage analysis. Microsatellite tagging allows for rearing of many families (progeny from different parents) in a single communal tank, which can be subsequently identified. This frees available tank space for use in other areas of production and eliminates large environmental effects associated with separate tank family rearing. The technique is potentially very effective depending on the level of polymorphism found in the microsatellite loci used. In a study on the Sea Bass (*Dicentrarchus labrax*) a minimum of two polymorphic microsatellite loci, were needed to unambiguously determine parentage from a three male/three female complete factorial design cross (Garcia De Leon *et al.*, 1998).

Other examples include the use of microsatellites to identify spawning success in female channel catfish in commercial ponds. Females that did not spawn were identified and removed by analysing the genotype of the progeny that came from the ponds. Non-spawning females were removed to maximize the production of fingerlings (Waldbieser and Wolters, 1999).

Ample variability at fish microsatellite loci exists for performing population genetic studies (O'Reilly and Wright, 1995). Results from multiple studies performed on fish indicated a mean number of alleles per locus of 7.5 for freshwater fish, 20.6 for marine fish and 11.3 for anadromous fishes. Average heterozygosities, across loci, were 0.54 (freshwater), 0.68 (anadromous) and 0.77 (marine). (de Woody and Avise, 2000).

The application of microsatellites in conservation of endangered populations can involve hybridisation studies, population phylogeography and demographic history, detection of inbreeding and bottlenecks, and the effects of social structure, dispersal and reproductive behaviour on genetic structure (Goldstein and Schlotterer, 1999). Results from microsatellite studies can be applied in studying the impact that aquaculture and other human activities have on wild populations of exploited fish. Disruption to fish stocks including, natural habitat destruction, overfishing and contamination with exotic species often goes unnoticed due to most fish populations being out of

site to the human eye (Ryman *et al.*, 1995). These 'invisible' populations make up the sea and inland fisheries. The continued existence of our fisheries relies on knowing how to manage potentially exhaustible resources (Ryman *et al.*, 1995; Ryman, 1997; Yokota and Watanabe, 1997; Utter, 1998). Genetic variation, geographic structuring, migration and breeding strategies in fish populations have been monitored using microsatellite markers. Among many species studied are Bigeye tuna (Grewe and Hampton, 1998), European sea bass (Garcia De Leon *et al.*, 1997), Sandbar Shark (Heist and Gold, 1999), catfish (Agnese *et al.*, 1997) and many species of Salmon (Adkison, 1995; McConnell *et al.*, 1995a; McConnell *et al.*, 1995b; Small *et al.*, 1998).

## 1.4 POPULATION GENETIC INDICES AND SIGNIFICANCE TESTS

The application of population genetics in aquaculture and conservation makes use of various significance tests and diversity indices, used independently or combined, to provide a picture of the genetic structure (Hillis *et al.*, 1996). A brief explanation of theories, statistical tests and formulas behind calculating the genetic indices, at both the locus and population level, is presented below.

### 1.4.1 Hardy Weinberg Equilibrium

The Hardy Weinberg equilibrium (HWE - sometimes referred to as the Castle, Hardy, Weinberg equilibrium) is the state of genotypic frequencies arising as a result of random union of gametes that exists in a population with no immigration, no selection, no mutation, and random mating. Knowing observed allele frequencies, HWE expected genotype expectations (such as $H_e$, expected heterozygosity) can be calculated and compared to observed genotype frequencies (such as $H_o$, observed heterozygosity). Departures from expectations are informative regarding population structure. Such departures were extended by Wright (1921) into hierarchical F statistics (based on 'F' expected level of inbreeding (or heterozygote deficit) under various conditions), from which rates of gene flow between populations can be estimated (see section 1.4.10).

There are a number of common tests used to determine the presence or absence of HWE at loci and in individual populations (Weir, 1996). The Exact test for HWE (Haldane, 1954; Louis and Dempster, 1987) is an alternative to the more common $\chi^2$ test for HWE. The Exact test can be applied to multiple alleles with low frequencies whereas the $\chi^2$ test is only effective if the allele frequencies are reasonably high. In order to avoid artificially high $\chi^2$ values generated by very low expected values, Zar (1999) recommends that individual cells in an expected frequency contingency table cannot be less than one, and that no less than 20% of cells can have values less than five. Alleles can be pooled to avoid these situations, however the test becomes less informative.

Due to the increased computational complexity behind calculating the exact probability for large data sets, Guo and Thompson (1992) suggested use of Monte Carlo or Markov Chain, randomisation algorithms to estimate the significance of Exact tests. Such multiple resampling techniques have become accessible in recent years due to increasing computing power. It should be noted that they are often described as $\chi^2$ tests, since they adopt the $\chi^2$ method of calculation, however the distribution from each data set is uniquely calculated, the magnitude of the observed $\chi^2$ like value being compared against expectations generated from multiple random sampling from the original data set. The Markov Chain method is more useful for large sample sizes with few alleles since it does not rely on calculating the Hardy Weinberg proportions of each newly generated random table according to Guo and Thompson (1992). The Monte Carlo method is useful for small

samples with many alleles since allele number does not affect speed of calculation (Guo and Thompson, 1992). Randomization procedures for Exact tests are outlined in Sokal and Rohlf (1995). Absence of HWE due to heterozygote deficiency can indicate possible inbreeding, the presence of null alleles or structuring within the population (Wahlund-Effect) (Hartl and Clark, 1997), or more rarely may possibly be attributable to selection or non-random mating.

### 1.4.2 Allelic Diversity (α)

A high mutation rate (due probably to slipped strand mispairing) can create a large number of alleles at microsatellite loci. Allelic diversity gives an indication of how much mutation has occurred at a locus. A measure of allelic diversity for multiple loci is obtained through averaging the alleles across loci to get a mean number of alleles (α) (Hedrick, 2000).

Alleles that are rare or unique to a population are often useful for assigning unknown fish to particular populations (Grewe and Hampton, 1998). They can also provide evidence of gene flow levels between populations (Slatkin, 1985; Slatkin, 1993).

### 1.4.3 Observed Heterozygosity ($H_o$)

Observed Heterozygosity ($H_o$) is the direct count of heterozygous individuals in a population (sample) calculated as:

$$H_o = \sum_{i<j}^{n} P_{ij}$$ (Hedrick, 2000)

Where: *Pij* is the frequency of the genotype *ij* (Hedrick, 2000).

When populations that were previously isolated join together, the average heterozygosity can increase. This is known as isolate breaking (Hartl and Clark, 1997), although population mixing is commonly cited as giving a heterozygote deficit (compared to expectations).

Reduced $H_o$ can give an indication of the inbreeding that has occurred within a population (Hartl and Clark, 1997). Null alleles can also result in an apparent reduced $H_o$ within populations (Callen *et al.*, 1993; O' Reilly and Wright, 1995; Paetkau and Strobeck, 1995; Pemberton *et al.*, 1995).

### 1.4.4 Expected Heterozygosity ($H_e$)

An informative index of genetic variation is the expected heterozygosity ($H_e$) or gene diversity index, calculated for each population (sample) as:

$$H_e = 1 - \sum_{i=1}^{n} Pi^2$$ (Hedrick, 2000)

Where: $Pi^2$ = the frequency of expected homozygotes (Hedrick, 2000).

Differentiation in $H_e$ between subpopulations occurs in populations that become structured through

isolation leading to different allele frequencies and genotype combinations in the subpopulations (Tessier *et al.*, 1997). Quantitative geneticists can use $H_e$ to compare heterozygosity with selective improvement indices. Population response rate to artificial selection often correlates directly with the amount of genetic variation in the populations undergoing trait selection however this is not always the case (Boulding *et al.*, 1993; Hartl and Clark, 1997).

For unequal population sample size an unbiased estimate of $H_e$ is preferably calculated for each sample (Nei, 1978; Hedrick, 2000).

$$H_e = \frac{2N}{2N-1}\left(1-\sum_{i=1}^{n} Pi^2\right)$$ (Hedrick, 2000)

Where: $N$ = the number of individuals in a population (sample) (Hedrick, 2000).

### 1.4.5 Linkage Equilibrium (Disequilibrium Coefficients (D), Exact tests)

Microsatellite loci can be tested for gametic phase linkage equilibrium to ensure that there is no selection for particular allelic combinations in microsatellite haplotypes however the process is computationally complex. For multiple alleles, disequilibrium coefficients are calculated for each pair of alleles, u and v, between two loci as:

$D_{uv} = p_{uv} - p_u p_v$ (Weir, 1996)

Where: $p_{uv}$ is the frequency of a gamete consisting of two alleles from two different loci A and B. and $p_u p_v$ is the product of allele frequencies for two alleles from the two different loci $A_u$ and $B_v$ (Weir, 1996).

Chi-square tests exist for the significance of individual D coefficients ($\chi^2_{uv}$) and the overall significance that all the coefficients are not greater than zero ($\chi^2_T$). The degrees of freedom for $\chi^2_T$ are calculated as $(k\text{-}1)(l\text{-}1)$ where $k$ and $l$ are the number of alleles at the two loci respectively.

An alternative method for calculating the probability of linkage disequilibrium, between two loci, is with a Fisher's Exact test. The Exact test gives the probability of obtaining n11, n12, n21, n22, counts of gametic combinations A1B1, A1B2, A2B1, A2B2 respectively for the alleles A1, A2 and B1, B2 between two loci (Weir, 1996; Hedrick, 2000).

### 1.4.6 Mutation Rate (μ)

The calculation of mutation rate for a particular DNA marker requires extremely large data sets making the exact calculation for each new organism studied impractical. The mutation rate for most microsatellite loci is generally accepted as $1 \times 10^{-4}$ per site per generation based on work done by Levinson and Gutman (1987) and Henderson and Petes (1992). The mutation rate value forms part of many formulas used in the calculation of equilibrium heterozygosity, the migration rate and

effective population size for populations under both the IAM and SMM mutation models (Hedrick, 2000).

The mutation rate is not constant across different microsatellite loci (Takazaki and Nei, 1996; Weber and Wong, 1993). Estimations for rates of dominant or co-dominant mutation in diploid organisms are calculated according to the formula:

$$\mu = \frac{x}{2N}$$ (Hedrick, 2000)

Where: $x$ = the number of mutant offspring and $N$ = the total number of offspring examined (Hedrick, 2000).

### 1.4.7 Number of Migrants ($N_m$)

The movement of genes from one population to another takes place in one of three validated gene migration models reviewed by Hartl and Clark (1997) and Hedrick (2000). The finite island migration model states that gene flow occurs in a unidirectional fashion from a source population to a recipient population as in the case of a continent to an island. There is no significant reciprocal gene flow from the island to the continent (Maruyama, 1970). The infinite island migration model (Wright, 1931) infers that gene flow can occur in any direction from one subpopulation to any other subpopulation as in the case of a number of islands making up an archipelago. The Stepping Stone migration model, formulated by Kimura and Weiss (1964), proposes a stepwise fashion of gene migration, where gene flow from one population is limited to its most adjacent populations. Freshwater fish often inhabit watersheds that extend down a coastline or along a main drainage basin, as is the case for many salmon populations (Adkison, 1995). Gene migration between these populations may simulate the Stepping Stone model where gene exchange, through migration, occurs from one population to its immediate neighbouring populations (Adkison, 1995).

A widely used formula for estimating the number of migrants per generation between two populations is $N_m = (1-F_{ST})/4F_{ST}$ where $F_{ST}$ is the fixation index between two populations (Hartl and Clark, 1997; Hedrick, 2000).

### 1.4.8 Genetic Drift and Effective Population Size ($N_e$)

From one generation to the next, in the absence of selection, chance determines the fate of an allele in a population according to the Wright-Fisher model of Random Genetic Drift outlined in (Hartl and Clark, 1997; Hedrick, 2000). The frequency, p, of an allele in a population, fluctuates according to random sampling of N number of individuals containing the allele in the previous generation. Ultimately p becomes either zero (lost) or one (fixed). The speed at which a gene is lost or fixed – described by a binomial variance (pq/N) - is inversely proportional to the size of the population

from one generation to the next. This process is known as random genetic drift. The process of drift is influenced by changes in population size, mutation, migration, selection, unequal male to female ratios and overlapping generations. The effective population size ($N_e$) is the proportion of individuals from one generation that extend the existence of certain alleles by breeding and passing their alleles on to the next generation. $N_e$ is effectively the number of individuals that would experience random genetic drift if the population were ideal.

If the exact number of males and females contributing genes to the next generation is known then the $N_e$ can be calculated as:

$$N_e = \frac{4N_f N_m}{N_f + N_m}$$ (Hedrick, 2000)

Where $N_f$ and $N_m$ are the number of females and males respectively that successfully breed.

### 1.4.9 Tests for Significant Population Structuring ($\chi^2$ and Exact tests)

Structuring is brought about by genetic divergence of newly formed populations from a common ancestral population through reproductive isolation, reviewed in Hartl and Clark (1997) and Hedrick (2000). The objective of a population genetic study is to determine the extent of population structuring for a species (one large population, sub-populations or sub-sub-populations). To do this accurately, with microsatellites, a number of genetic structuring tests and indexes have been adapted for use based on allele frequency and mutation at microsatellite loci. The first of these is a simple $\chi^2$ contingency test for inter-population allelic heterogeneity outlined by Workman and Niswander (1970). For small populations with many low frequency alleles, the $\chi^2$ test is biased due to highly inflated $\chi^2$ values arising from low expected values. A more useful method for detecting differentiation in small populations makes use of multiple random resampling algorithms such as the Exact test based on Monte Carlo simulation resampling Monte Carlo Exact test (Raymond and Rousset, 1995).

### 1.4.10 Measures of Population Structuring ($F_{ST}$, $\theta_{ST}$, $R_{ST}$, $\Phi_{ST}$)

Allele frequency based fixation indices, developed by Wright (1921) for use in population structure analysis, can be used to measure population variability between individuals within sub-populations ($F_{IS}$), variability between sub-populations ($F_{ST}$) and total variability across all populations ($F_{IT}$). The three values are interrelated by the equation $1 - F_{IT} = (1 - F_{ST})(1 - F_{IS})$. The F-Statistics originally expressed by Wright in terms of probability of uniting gametes, can be simply expressed in terms of heterozygosity as demonstrated by Nei (1977) for population analysis with multiple loci; $F_{IS} = (H_S - H_O) / H_S$, $F_{ST} = (H_T - H_O) / H_T$ and $F_{ST} = (H_T - H_S) / H_T$. $H_O$ is the average observed sub-population heterozygosity, $H_S$ is the average expected sub-population heterozygosity and $H_T$ is total population

heterozygosity. All three measures are averaged for the number of loci used (Weir and Cockerham, 1984; Hartl and Clark, 1997; Hedrick, 2000).

$F_{ST}$ is the value of importance for population genetic studies as it provides an indication of the amount of genetic variation present between populations, which depends largely on levels of gene flow (migration). Four ranges of $F_{ST}$ relate to different degrees of genetic differentiation between populations. Genetic differentiation has been described by Hartl and Clark (1997) as low for $F_{ST}$ values between zero and 0.05, moderate between 0.05 and 0.15, high between 0.15 and 0.25 and extremely high above 0.25, although these broad categories are rather subjective and don't take sample size or sampling errors into account (Hartl and Clark, 1997). Workman and Niswander (1970), however, described a $\chi^2$ approximation, which accounts for sample size, and facilitates an objective indication of the statistical significance of $F_{ST}$ values. The value, $F_{ST}$, essentially determines the extent of heterozygosity reduction resulting from differential genetic drift among sub-populations. Hartl and Clark (1997) provide a simplified example of the influence of genetic differentiation among subpopulations as follows - a subdivided population has genotype frequencies of $(p_{ave}^2 + (p_{ave} \times q_{ave} \times F_{ST}))$ for AA, $((2 \times p_{ave} \times q_{ave}) - (2 \times p_{ave} \times q_{ave} \times F_{ST}))$ for Aa and $(q_{ave}^2 + (p_{ave} \times q_{ave} \times F_{ST}))$ for aa. $F_{ST}$ values are usually quoted with an accompanying probability value of how significant the $F_{ST}$ values are based on a $\chi^2$ test. The test is dependent on three parameters the total number of individuals sampled (N), the number of alleles at a locus (k) and the number of populations (s).

Degrees of freedom for the test are calculated as (k-1)(s-1) (Workman and Niswander, 1970). An unbiased measure of $F_{ST}$, $\theta_{ST}$, detects correlation of genes between individuals within populations. $\theta_{ST}$ is not affected by population sample size, number of alleles at a locus or the number of populations sampled (Weir and Cockerham, 1984).

Allele frequencies at microsatellite loci can be misleading due to the (SMM) mutation model that the majority of microsatellite loci follow. As a result, $F_{ST}$ values often show too much similarity between populations that diverged long ago or where gene flow is very low. A measure adapted from the $F_{ST}$ and $\theta_{ST}$ measures known as $R_{ST}$ that incorporates allele frequency and the difference in allele size at microsatellite loci, provides an improved indication of population structuring, population divergence time and migration rates for microsatellites under these circumstances. $F_{ST}$ and $R_{ST}$ values can differ from one locus to the next in the same sub-population, therefore $F_{ST}$ and $R_{ST}$ are calculated using all loci to obtain overall $F_{ST}$ and $R_{ST}$ values (Slatkin, 1995).

$\Phi_{ST}$ is another measure of the molecular component of variation within and between populations. Calculations of $\Phi_{ST}$ require the use of analysis of variance (ANOVA), or more aptly for molecular data, analysis of molecular variance (AMOVA) (Excoffier *et al.*, 1992). The AMOVA technique

was originally developed for analysis of haplotype frequencies as in mitochondrial DNA analysis. Michalakis and Excoffier (1996) adapted the AMOVA technique for use with microsatellite diploid data. $\Phi_{ST}$ for diploid organisms makes use of Euclidean distance matrices where the distances can be $F_{ST}$ based, $\theta_{ST}$ based or $R_{ST}$ based. For microsatellite data, $R_{ST}$ - like distance matrices are more suitable.

### 1.4.11 Correction for Multiple Comparison $\chi^2$ Tables (Bonferroni Adjustment)

When using $\chi^2$ to test multiple components making up a comparison table, a common mistake is to reject the Null Hypothesis Ho based on one or more of the components having significant probability (*P*) values. In doing these multiple table tests, chance alone results in *P* values to be less than or equal to 0.05 - a commonly accepted level of statistical significance - in 20% of random tests, (Weir, 1996). A procedure known as sequential Bonferroni adjustment is used to overcome the possibility of a Type I error (rejection of a true null hypothesis (Ho)). The chosen significance level $\alpha$ is sequentially modified for each individual test to prevent the chance occurrence of too many significant *P* values. The *P* values for all the component tests making up the table are ranked from the smallest to the largest. The smallest value *P1* obtained for a multiple table is compared to an adjusted significance level $\alpha/k$ where k is the number of individual test components. If *P1* is smaller than or equal to $\alpha/k$ then the corresponding table component is considered significant. If not then all the table components tests are considered non-significant. The next highest value *P2* is then compared to an adjusted value of $\alpha/(k-1)$. If *P2* is equal to or smaller than $\alpha/(k-1)$ then its corresponding component is also considered significant for the table. The sequential adjustment of $\alpha$ continues until a non-significant *P* is obtained (Hochberg, 1988; Rice, 1989).

### 1.4.12 Genetic Distance ($D_s$, $D_m$ and $(\delta\mu)^2$)

The factors that determine the genetic similarity or distance between two populations include time since the populations separated from a common ancestral population, the amount of gene flow that occurs between them, intensity of genetic drift within each population, and differential selection. If neutral markers are used to assess the genetic distance then the effects of selection will not be detected. Neutral markers are preferred in population genetics exactly for this reason. Many genetic distance measures have been developed for population genetic research. The majority of them rely on the allele frequency differences present between populations under the infinite alleles mutational model (IAM) (Beaumont *et al.*, 1998; Hedrick, 2000).

The most commonly used distance measure is that of Nei's standard genetic distance $D_s$ (Nei, 1972). This measure relies on the genetic identity (I) of the alleles present in two populations. $D_s$ values are overestimated for populations with high heterozygosities. To overcome this, Nei

(1972) introduced an unbiased measure known as the minimum distance ($D_m$). $D_m$ is independent of gene diversity within populations making it more suitable for studying structured populations with different allele frequencies. $D_m$ was further modified to an unbiased minimum distance independent of sample size (Nei, 1978).

The mutation model that the majority of microsatellites are believed to approximate SMM, led researchers to believe that $D_s$ and $D_m$ distance measures would be less effective at calculating the distances between populations for microsatellite marker data. A new distance measure $(\delta\mu)^2$, was developed specifically for microsatellites. It incorporates the SMM by comparing the average allelic states (mean allele size in repeat number) between two populations summed over all loci (Goldstein *et al.*, 1995a; Goldstein *et al.*, 1995b; Ruzzante, 1998). Takezaki and Nei (1996) conducted a simulation study to compare the performance of $D_s$ and $D_m$ (IAM) to that of $(\delta\mu)^2$ (SMM) in determining the genetic distance between populations for microsatellite loci. They concluded that $D_s$ and $D_m$ were better than $(\delta\mu)^2$ however a similar study by Ruzzante (1998) found $(\delta\mu)^2$ to be superior to distances based only on allele frequency. Ruzzante (1998) also noted that $(\delta\mu)^2$ was biased under small sample size conditions, since it relied on the variance in allele state for each population at the loci analysed, but it still performed better than $D_s$ and other allele frequency based measures for microsatellite data. Takezaki and Nei (1996) also pointed out the fact that different microsatellite loci can have vastly different mutation rates. Angers and Bernatchez (1997) and Orti *et al.* (1997) cautioned against using repeat counts as the only measure of historical relationship between populations. The occurrence of substitutions in microsatellite flanking regions leads to the formation of new alleles not in accordance with the SMM (Orti *et al.*, 1997).

## 1.5 PHYLOGENETICS (UPGMA, NEJ AND MDS)

Phylogenetics involves presenting the data present in genetic distance or character state measures in a form that reveals the evolutionary trend of different species or populations. Phylogenetic trees are divided into two types, rooted and unrooted. Rooted trees describe the genetic relation between populations or species (taxa) by forming a topology (branch network) that infers a common ancestor (the earliest point in time). Unrooted trees do not relate taxa based on them having a common ancestor.

The construct of a typical tree starts at the external or terminal nodes representing each taxon. The external nodes are connected via a network of branches. The branches interconnect to form internal nodes. In turn, additional branches radiate from the internal nodes to progressively deeper nodes. Each internal node connects related taxa or groups of taxa (clusters). The extent of divergence between taxa or clusters resides in the length of the branches connecting different nodes (Hillis *et al.*, 1996; Hedrick, 2000).

Bowcock *et al.* (1994) and Di Rienzo et al. (1998) demonstrated the utility of microsatellites for construction of trees depicting human geographic origin. Microsatellites have also been used to make phylogenetic inferences in a number of salmon population genetic studies (Mc Connell *et al.*, 1995a; Mc Connell *et al.*, 1995b; Nielsen *et al.*, 1997; Nielsen and Fountain, 1999).

The phylogenetic tree building methods available are all optimally suited for particular types of DNA markers and distance methods. The chosen method usually incorporates characteristics that will provide the most accurate depiction of the genetic distance data. Two broad categories to tree building methods are those that rely on direct character state differences or similarities (Parsimony methods) and methods that consider the models of evolutionary change that lead to differences between taxa (Distance methods) (Hillis *et al.*, 1996). The negative aspect of parsimony methods is they take longer to perform than distance methods. Time is needed to find the tree with the least number of mutational events between taxa. Parsimony trees often give an incorrect topology due to lack of any background mutation model information whereas distance methods use genetic distance matrices with inherent mutation model abstraction. Two commonly used distance methods for microsatellite data are the Unweighted Pair-Group Method Using Arithmetic Averages (UPGMA) and the Neighbour-Joining method (NEJ) (Hartl and Clark, 1997; Hedrick, 2000; Nei and Kumar, 2000). An alternative way to view distance data is with the use of multi-dimensional scaling (MDS) (e.g. Rao and Majumdar, 1998), which is a commonly applied graphical representation of identity matrix data.

### 1.5.1 Unweighted Pair-Group Method Using Arithmetic Averages

UPGMA is a clustering technique that defines populations in a distance matrix as Operational Taxonomic Units (OTUs). The UPGMA technique sequentially combines the two OTUs with the smallest distance between them into a cluster. A new matrix is then created where the combined cluster of OTUs has a distance value equal to the mean of the constituent OTUs. A new matrix is created comparing the cluster with the remaining OTUs. This process continues until a number of clusters are joined by a final calculation relating the two most distant clusters to one another. UPGMA trees are conventionally rooted since all clusters are eventually connected at a single node representing the root (ancestral population). UPGMA relies on equal mutation rate for all OTUs. The reliability of the tree can be tested using Bootstrapping techniques (Ferguson, 1980; Hillis *et al.*, 1996; Hartl and Clark, 1997; Hedrick, 2000; Nei and Kumar, 2000), however this commonly resamples over loci, resulting in limited application for data sets with small numbers of loci.

### 1.5.2 Neighbour-Joining

Neighbour-Joining is similar to the UPGMA method except that it does not make the assumption of equal mutation rate for all OTUs. The NEJ technique proceeds as follows: A matrix of total branch length distance between all OTUs, assuming each pair of OTUs are potentially neighbours, is constructed from the original distance matrix values. From this total branch length matrix, the two OTUs with the smallest total branch length distance are linked to form a common ancestral node. A new distance matrix is constructed relating the ancestral node to the remaining OTUs. This matrix is then used to construct a second matrix of total branch lengths. Again, the two OTUs with the smallest total branch length are joined to form a new internal node. This procedure continues until there are only two nodes left separated by a single branch (Hillis *et al.*, 1996; Hartl and Clark, 1997; Hedrick, 2000; Nei and Kumar, 2000).

### 1.5.3 Multi-dimensional Scaling

MDS makes use of the Euclidean distance formula to model genetic distance based on stimulus configuration co-ordinates that relate OTUs to one another in the form of a two-dimensional plot. Euclidian distances are obtained using the formula:

$$d_{ij} = \left[ \sum_{a}^{r} \left( x_{ia} - x_{ja} \right)^2 \right]^{1/2}$$ ( Norusis, 1994 )

Where $x_{ia}$ and $x_{ja}$ are the derived stimulus co-ordinates for OTU $i$ in dimension a and OTU $j$ in dimension a respectively (Norusis, 1994).

A linear fit scatter plot of Euclidean distance values generated from a two dimensional plot, versus optimally scaled genetic distance values (disparities) between the OTUs indicates the accuracy to which the Euclidean distances model the actual genetic distances. There are two types of MDS, metric and non-metric. Metric MDS deals with distance data that is interval while non-metric MDS deals with ordinal data. The use of non-metric (ordinal) MDS is more appropriate for genetic distances calculated from qualitative data such as microsatellite alleles (Rao and Majumdar, 1998; Norusis, 1994).

## 1.6 AIMS OF THIS STUDY

Growth of the South African aquaculture industry requires the development and domestication of new fish species for use in aquaculture. *O. mossambicus* represents a promising candidate for such exploitation. The wild populations of Mozambique tilapia represent an important genetic resource from both an aquaculture and a conservation perspective since they are relatively free of the effects of inbreeding and species contamination found in Asian Mozambique tilapia populations. Domesticated strains need to be carefully developed to successfully integrate *O. mossambicus*, free of inbreeding affects, into South African aquaculture while conserving the genetic integrity and diversity in the wild populations. The analysis of population structure in the twelve *O. mossambicus* populations encompasses three objectives; to determine levels of genetic variation found within and between the twelve study populations; to investigate the effect of captivity on the populations; to use F-statistics and genetic distance measures to evaluate genetic structure and phylogenetic relationships between the populations.

The population genetic data obtained through the above objectives will be discussed in the light of quantitative genetic data for the captive populations and the conservation issues surrounding their source populations.

# CHAPTER TWO

## *MATERIALS AND METHODS*

### 2.1 SOURCE MATERIAL

The twelve populations used in this research work were initiated from six wild, four farmed and two introduced sources of *O. mossambicus*, located in the South Eastern area of Southern Africa. Between 50 and 100 individuals were collected from each population between 1997 and 1999, and returned for captive breeding at the University of Stellenbosch.

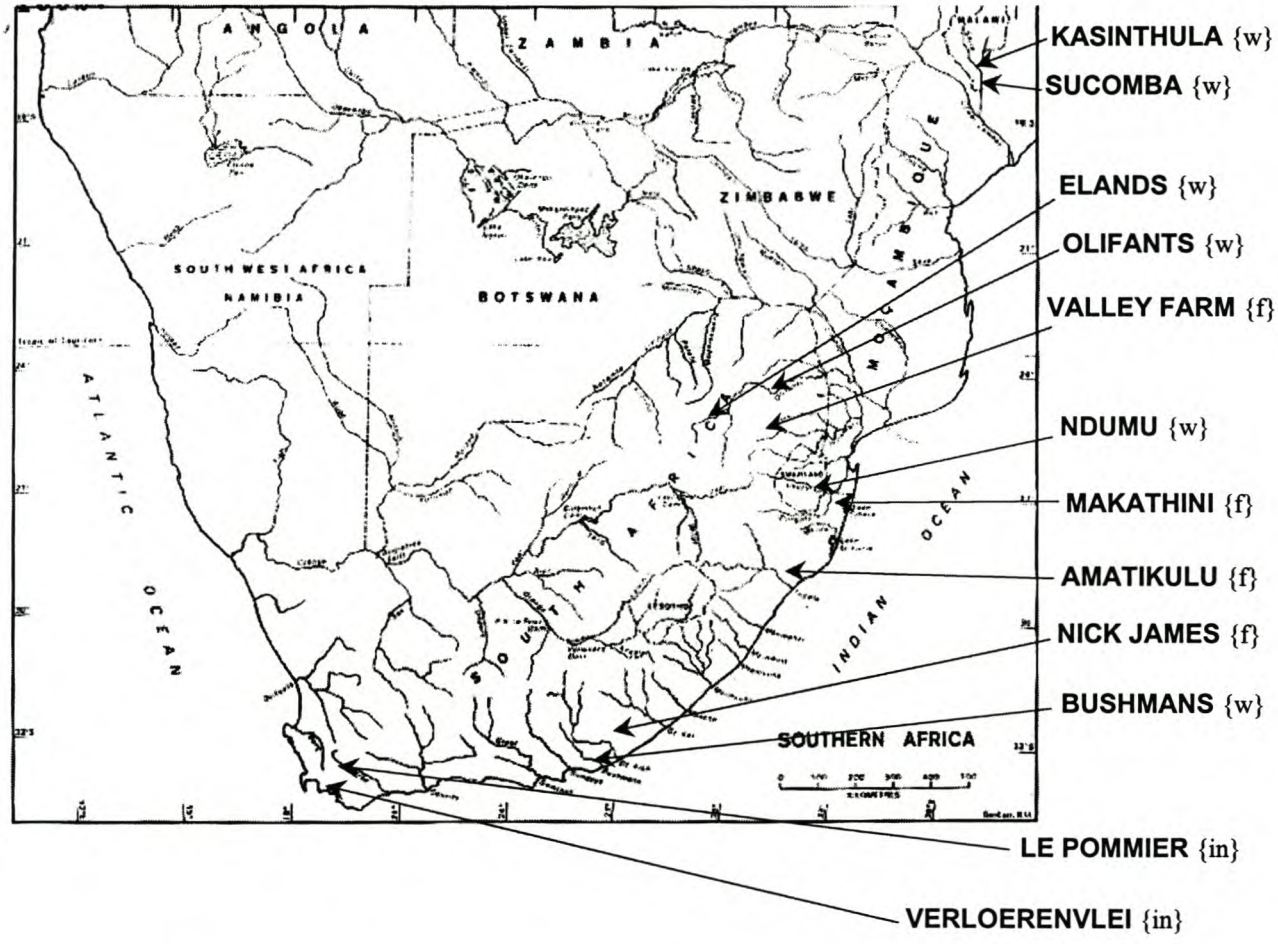


**Figure 2.01**: Diagram of collection locations for the twelve populations in Southern Africa

Key: {w} = a population collected from a wild source

{f} = a population collected from a farm

{in} = an introduced population

**Table 2.01:** Description of collection locations for the twelve populations

| Population | Approximate map grid location | Population description |
|---|---|---|
| Amatikulu | 29° South<br>31° 30' East | A red coloured, ornamental breeding population, collected at the Amatikulu fish farm in Natal |
| Bushmans | 33° 30' South<br>26° 30' East | A wild population from the Bushmans river in the Eastern Cape, South Africa |
| Elands | 25° 15' South<br>29° 15' East | A wild population from the Limpopo river system in Mpumalanga, South Africa |
| Kasinthula | 17° South<br>35° 30' East | A wild population from a research station based on the edge of the Shire river system, Malawi |
| Le Pommier | 34° South<br>18° 45' East | A population introduced in 1940's into the Western Cape, South Africa |
| Makathini | 27° South<br>31° 30' East | A population from the Makathini research station flanking the Pongola river in Northern Kwazulu Natal, South Africa |
| Ndumu | 26° 45' South<br>32° East | A wild population from the Pongola/Usutu river systems in Northern Kwazulu Natal, South Africa |
| Nick James | 33° South<br>27° East | A red coloured, ornamental breeding population collected from an ornamental fish farm near Grahamstown, South Africa |
| Olifants | 24° South<br>31° 45' East | A wild population from the Limpopo river system in Mpumalanga, South Africa |
| Sucomba | 17° South<br>35° 30' East | A wild population from the lower Shire river system, Malawi |
| Valley Fish Farm | 25° South<br>31° East | A red coloured, ornamental breeding population collected from the Valley ornamental fish farm in Mpumalanga, South Africa |
| Verloerenvlei | 34° 10' South<br>18° 30' East | An introduced population from a brackish water lake in the Western Cape, South Africa |

## 2.2 POPULATION SAMPLING SCHEME

A sample consisting of 25 fish was taken from each of the captive populations in March 1999. Larger samples of up to 50 fish (population size dependent) were taken from the captive populations in March of 2000. A fin-clip tissue collection technique was used as detailed below:

Twelve fin-clip collection boxes were assembled each containing 50 screw-cap tubes filled with 2ml of 95% ethanol. Individual fish were removed from their holding tanks and placed between two wet towels. A piece of caudal fin $4mm^2$ to $16mm^2$ in size, depending on the size of the fish, was clipped and placed in an ethanol tube. Subsequent examination of the sampled fish showed that the abscised fin tissue took approximately two to three weeks to fully regenerate after which no evidence of fin clipping was present. Taking fin clips had little or no impact on the long-term survival of the fish.

## 2.3 GENOMIC DNA ISOLATION FROM FIN TISSUE

Genomic DNA extraction from all the sampled individuals made use of a modified commercial DNA kit extraction procedure (Nucleon 1, SCOTLAB, Scotland). Sections of fin clippings ($4mm^2$) were lysed in 1.5ml microcentrifuge tubes (microtubes) containing 40µl of lysis buffer (400mM TrisHCL, 60mM EDTA, 150mM NaCl, 1% SDS). The microtubes were placed on ice while the tissue was macerated inside the tubes using a pair of dissecting scissors. To reduce contamination the scissors were rinsed in ethanol and distilled water, between macerations. A further 340µl of lysis buffer was added to the tubes, which were then placed on ice for ten minutes. 100µl of 5M Sodium Perchlorate was added to each tube to end the lysis stage. The tubes were agitated for 30 seconds after which they were placed in a 37°C water bath for twenty minutes. Vigorous intermittent shaking of the tubes took place every five minutes. The same procedure was performed using a 65°C water bath for twenty minutes. This was followed by addition of 540µl of chloroform (–20°C), five minutes of vigorous shaking, then fifteen minutes of gentle rotation (room temperature). Separation of the aqueous phase from the organic phase was achieved using centrifugation at 14000 RPM for three minutes. The aqueous phases were removed from each tube and placed in new tubes. 880µl of isopropanol (-20°C) was mixed with the aqueous phase. The tubes were gently inverted. Precipitated DNA was then pelleted by centrifugation at 14000 RPM for four minutes. A 70% and a 100% ethanol wash followed. DNA pellets were left to dry for sixty minutes after which they were re-suspended in 50µl of millipore water. DNA samples were then standardized to a concentration of 100ng/µl using a GENEQUANT spectrophotometer followed by storage at 4°C until use.

## 2.4. CHOICE OF MICROSATELLITE LOCI FOR THE ANALYSIS

Primer sets for ten CA repeat loci, chosen on the basis of being present on different linkage groups, were chosen from the tilapia linkage map for *O. niloticus* constructed by Kocher *et al.* (1998). The linkage map was developed using *O. niloticus* DNA, therefore loci were selected based on known amplification in the closely related *O. aureus* and a Haplochromine cichlid from lake Malawi (http: Tilapia Genome Homepage). Forward and reverse primer sequences for the ten loci are listed in GENBANK (Table 2.02).

**Table 2.02**: Sequence of forward and reverse primers for the ten chosen microsatellite loci

| Locus | Forward Primer (5' – 3') | Reverse Primer (5' – 3') |
|---|---|---|
| UNH102 | AAATGATACATGACTGCTTA | TTAGGACTTATCTGTCTACAAGC |
| UNH104 | GCAGUATUGTGGTCACTA | GGTATATGTCTAACTGAAATCC |
| UNH106 | CCTTCAGCATCCGTATAT | GTGTCTUCTCTCTGTCACAAG |
| UNH108 | GGGATCAGCTGUAAGTTT | TGAGTTGATTATTAATTTCTGA |
| UNH111 | TGCTGTTCTTATTTTCGC | ATAAGAGTGTATGCATTACTGG |
| UNH115 | ACCTTCATCTCGGTCAG | TCAAGCAGCTGATTTTT |
| UNH123 | CATCATCACAGACAGATTAGA | GATTGAGATTTCATTCAAG |
| UNHI29 | AGAAGTCGTGCATCTCTC | TGTACATCATCTGTGGG |
| UNH146 | CCACTCTGCCTGCCCTCTAT | AGCTGCGTCAAACTCTCAAAAG |
| UNH160 | CCAUGGCTCTTACATC | GATAGCATTTCTGTAGTTATGG |

### 2.4.1 PCR Protocol for Testing the Microsatellite Loci for Polymorphism

The microsatellite loci were tested for levels of polymorphism using standard PCR techniques with unlabelled deoxynucleotide triphosphates (dNTPs) on a random set of *O. mossambicus* DNA samples from the Bushmans and Ndumu populations. A master-mix of PCR reagents was prepared for each set of PCR reactions (Table 2.03).

**Table 2.03**: Reagents used to make up the PCR master–mix for non-fluorescent PCR microsatellite analysis (Adapted from Kocher *et al.*, 1998)

| Reagent [concentration] | Single tube volume (µl) [conc.] | (n+1) x single tube volume (µl) |
|---|---|---|
| BioTaq buffer [10 X] | 2.5 [1X] | (n+1) x 2.5 |
| $MgCl_2$ [25mM] | 1.5 [1.5mM] | (n+1) x 1.5 |
| dNTPs [10mM of each] | 0.5 [0.2mM of each] | (n+1) x 0.5 |
| Forward primer [20µM] | 0.4 [0.32µM] | (n+1) x 0.4 |
| Reverse primer [20µM] | 0.4 [0.32µM] | (n+1) x 0.4 |
| BioTaq Polymerase [5units/µl] | 0.2 [1 unit] | (n+1) x 0.2 |
| Water | 18.5 | (n+1) x 18.5 |
| Mix Total | 24 | (n+1) x 24 |
| Template [100ng/µl] | 1 | |
| Mix total + template | 25 | |

Master-mix aliquots of 24µl were added to the required number of 200µl, thin-walled, PCR microtubes. In general primer sets were tested for twenty or more individuals, to satisfy the minimum 95% criterion for polymorphism (i.e. a locus is described as polymorphic if the most common allele is present at a frequency of 0.95 or less. One microliter of relevant template [100ng/µl] was added to each tube. Tubes were then transferred to the PCR machine (Hybaid, PCR Express, Thermal Cycler).

The following PCR Thermal Cycle was used to amplify the PCR products (Adapted from Kocher *et al.*, 1998):

94°C for two minutes (denaturation)

30 cycles of:

94°C for 30 seconds (denaturation)

42°C for two minutes (annealing)

72°C for two minutes (extension)

72°C for 10 minutes (final product extension)

4°C indefinitely

PCR products were run on two percent agarose gels. Ethidium Bromide staining (0.5µg/ml in the gel) was used to view the DNA with a Gel Doc system (Bio Rad). Gel pictures were electronically stored as JPEG files using the IPHOTO$^{+4}$ graphics software package (Ulead Systems, Inc., Taiwan) (Appendix 1).

Three of the microsatellite loci, UNH104, UNH111 and UNH123, were polymorphic and sufficiently well resolved (Appendix 1). These loci were subsequently chosen for microsatellite analysis using fluorescently labelled dNTPs and an automated DNA sequencer (ABI 377). The remaining seven loci were excluded from the fluorescent analysis due to problematic PCR optimization and cost of fluorescent fragment analysis.

## 2.5 FLUORESCENT NUCLEOTIDE MICROSATELLITE ANALYSIS

The high cost of performing a microsatellite study with labelled deoxynucleotide triphosphates (FdNTPs) necessitated optimisation of both PCR and labelling techniques prior to the final PCR amplification of DNA samples from all twelve populations (Table 2.04).

**Table 2.04**: Population sample size, source and single letter codes used in the final fluorescent PCR amplification and subsequent genetic analyses.

| Population | Code | Source | Sample Size (N) |
|---|---|---|---|
| Amatikulu | A | F (red) | 36 |
| Bushmans | B | W | 36 |
| Elands | E | W | 9 |
| Kasinthula | K | W | 36 |
| Le Pommier | L | Int | 36 |
| Makathini | M | F | 10 |
| Ndumu | N | W | 36 |
| Nick James | NJ | F (red) | 10 |
| Olifants | O | W | 10 |
| Sucomba | S | W | 36 |
| Valley | V | F (red) | 17 |
| Verloerenvlei | VE | Int | 16 |
| F = Farm population<br>F (red) = Red Strain Farm Population<br>W = Wild natural population<br>Int = Introduced population | | | |

### 2.5.1 Fragment Analysis Optimisation Using the ABI 377 Automated Sequencer

In order to optimise ABI 377 analysis three important parameters were tested:

The amount of template DNA to use in the PCR reactions.

The ratio of unlabelled to labelled dNTP in the reactions.

The degree of PCR product dilution required before loading PCR products on the polyacrylamide sequencer gels.

Template DNA's extracted from four fish (two from the Bushmans population and two from the Ndumu population) were used to test the above parameters, based on clear PCR products obtained for these two populations, using the primer set UNH123 on standard 2% agarose gels (Appendix 1).

#### 2.5.1.1 Amount of Template DNA

The amount of template DNA used in the PCR reactions determined how many copies of the target DNA were present in the first cycle of PCR amplification. Two different dilutions were tested (20ng/µl and 100ng/µl).

### 2.5.1.2 Fluorescent Nucleotide Ratio

Fluorescently labelled deoxycytosine triphosphates (FdCTPs) were used to label the PCR products. The ratio of unlabelled deoxycytosine triphosphates (dCTP) to FdCTP present in the PCR reaction determined the strength of the signal emitted by the PCR products for detection by the ABI laser. Three ratios of unlabelled to labelled dNTP were tested, [1000:1], [500:1], [125:1].

### 2.5.1.3 Dilution of PCR Product

Three dilutions of the PCR products were tested prior to loading on the polyacrylamide gels: [1:10], [1:20] and a [1:30].

### 2.5.2 Final PCR Protocol used for Amplification with Fluorescent Nucleotides:

Three, different coloured, FdCTPs were available for labelling the PCR products. Each locus was assigned a specific FdCTP for analysis on the ABI 377. Tamra (yellow) was used for UNH104 PCR products, R6G (green) for UNH111 products and R100 (blue) for UNH123 products. Each FdCTP, at an original concentration of [100µM], was used at a ratio of 500:1 unlabelled dCTP to labelled FdCTP (see results section 3.1.2.2). A master-mix of PCR reagents was made up for each new set of PCR reactions (Table 2.05).

**Table 2.05**: Reagents used to make up the PCR master–mix for fluorescent PCR microsatellite analysis (Adapted from Kocher *et al.*, 1998 and PE-Biosystems user guide for PCR amplification with FdCTP)

| **Reagent [concentration]** | **Single tube volume (µl) [conc.]** | **(n+1) x single tube volume (µl)** |
|---|---|---|
| BioTaq buffer [10 X] | 2.5 [1X] | (n+1) x 2.5 |
| $MgCl_2$ [25mM] | 1.5 [1.5mM] | (n+1) x 1.5 |
| dNTPs [10mM of each] | 0.5 [200µM of each] | (n+1) x 0.5 |
| FdCTP [100 µM] | 0.1 [0.4µM] | (n+1) x 0.1 |
| Forward primer [20µM] | 0.4 [0.32µM] | (n+1) x 0.4 |
| Reverse primer [20µM] | 0.4 [0.32µM] | (n+1) x 0.4 |
| BioTaq Polymerase [5units/µl] | 0.2 [1 unit] | (n+1) x 0.2 |
| Water | 18.4 | (n+1) x 18.4 |
| Mix Total | 24 | (n+1) x 24 |
| Template [100ng/µl] | 1 | |
| Mix total + template | 25 | |

Master-mix aliquots of 24µl were added to the required number of 200µl, thin-walled, PCR microtubes. One microliter of relevant template [100ng/µl] was added to each tube. Tubes were transferred to the PCR machine (Hybaid, PCR Express, Thermal Cycler).

The following PCR thermal cycle was used for the fluorescent amplification of sample DNA. (Adapted from Kocher *et al.*, 1998):

94°C for two minutes (denaturation)

30 cycles of:

94°C for 30 seconds (denaturation)

42°C for two minutes (annealing)

72°C for two minutes (extension)

72°C for 10 minutes (final product extension)

4°C indefinitely

### 2.5.3 Purification of Fluorescent PCR Products

Seven microlitres of each of the UNH104, UNH111 and UNH123 PCR products, specific to an individual, were mixed in a new 200µl microtube. This mixture was then purified for gel analysis using a Cephadex column.

### 2.5.4. Gel Electrophoresis on the ABI Automated Sequencer

Three microlitres of loading dye (formalin, bromophenol blue, Rox 100 internal fluorescent ladder) was added to each individual's tube of purified PCR products. One microliter of this mix was then loaded into a separate well on a 6%, denaturing polyacrylamide gel (pre-run at 2.97kV for 30 minutes). The gels facilitated loading of combined PCR products for 36 individuals. Gels were then run at 2.97kV, 123 W, 51°C for two hours. Microsatellite locus and population individual information were electronically recorded on a computer coupled to the ABI 377 to ensure the correct assignment of allele peak intensity and size information to each individual.

A laser, operating at 39mW, detected the fluorescence emitted from the PCR products for each individual. The unique wavelength of each FdCTP allowed for unambiguous detection of three different coloured PCR products unique to each fish in a lane. The allele size and peak intensity for each of the three PCR products were recorded.

## 2.6 POPULATION GENETIC ANALYSES

### 2.6.1 Designation of Allele Size

Genescan and Genotyper (proprietary ABI software for the Apple Macintosh platform) were used to calculate the size and intensity of fluorescent peaks obtained for each individual fish. The graphic outputs displayed a set of different coloured peaks with intensity and peak size in base pairs for each individual unique to a population. Each set of peaks represented an individual's PCR products obtained at the three loci. The designation of one (homozygote) or two (heterozygote) peaks for each individual at each locus was complicated by the presence of stutter bands and non-specific PCR products (Figure 2.02). The following criteria were therefore used to make the final allele assignment. A stepwise mutation model (SMM) was assumed to be operating at each locus (for a reasoned justification, see Goldstein and Schlotterer, 1999). The two peaks with the highest peak intensity were chosen as the true peaks. The smaller of the two peaks (in base pair size) was always higher in intensity than the larger peak; if not it was considered carry over from a neighbouring well (observations on neighbouring well peak assignments confirmed this). Peaks that were not whole numbers were rounded down or up to the nearest even whole number. Even numbers were chosen based on the majority of peak sizes being even numbers. The UNH111 allele assignment was further complicated by the presence of one extra peak in some individuals, shadowing the true peak with exact intensity. The cause of this was believed to be 3' single base-pair (A) adenylation of UNH111 PCR products. Adenylated peaks were easy to recognise, as they were always one base pair bigger than the preceding true peaks (Browstein *et al.*, 1996). The UNH104 locus had a non-specific PCR product of 180 base pairs present for the majority of individuals. This peak was ignored, since scoring it would not have altered interpretation of population structure. Microsoft EXCEL (MSEXCEL) spreadsheets were used to store and convert the allele size data from the Genotyper data tables into ASCII data format for subsequent data analysis.

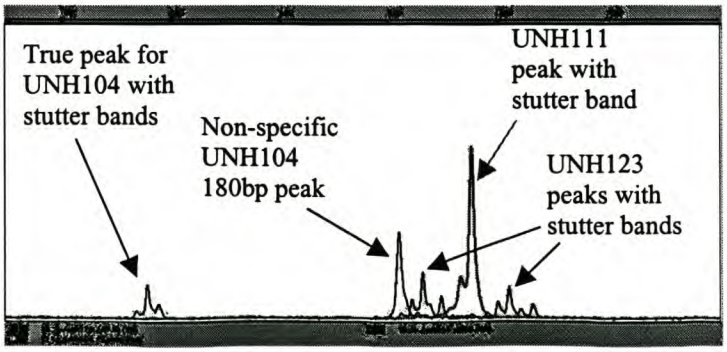


**Figure 2.02**: Diagram of Genotyper peak output showing true peaks, stutter bands and non-specific PCR products for the loci UNH104, UNH111, and UNH123.

### 2.6.2 Conversion of Data

The ASCII allelic data tables were converted to relevant formats necessary for input into the following population genetic software programs: Tools for Population Genetic Analysis (TFPGA) (Miller, 1997), FSTAT (Goudet, 2000); RST-CALC (Goodman, 1997), Genetic Data Analysis (GDA) (Lewis and Zaykin, 2001), ARLEQUIN version 2.0 (Schneider *et al.*, 1997), and PHYLIP (Felsenstein, 1993).

### 2.6.3 Calculating Genetic Variation

MSEXCEL was used to calculate allele frequency, observed heterozygosity and expected heterozygosity (gene diversity) for the three loci across all twelve populations. Additional TFPGA calculations confirmed the allele frequencies and heterozygosities obtained with MSEXCEL. Percentage polymorphism of the populations was calculated as the number of polymorphic loci per individual averaged across all individuals.

Hardy Weinberg Equilibrium for the entire allelic data set and for each population was tested using TFPGA's $\chi^2$ and Exact test options. The low frequency alleles were pooled for the $\chi^2$ test. The Exact test used a Monte Carlo randomization procedure with 10 batches and 1000 permutations per batch. Linkage disequilibrium between the loci in the populations was tested for by using GDA's Exact test option.

The inheritance of microsatellite alleles at the three loci was studied by making two controlled crosses of parents with known genotypes for the three loci. Cross one consisted of a male fish from the Ndumu population bred with an Olifants population female. Cross two consisted of a male fish from the Valley population bred with an Olifants female. The genotypes of the progeny from cross one and two were analysed for expected allele ratios and presence of possible null alleles at the three loci. Putative genotypes including one or two null alleles were constructed for each parent used in the crosses. Microsoft Word (MSWORD) tables of possible allele ratios in hypothetical offspring from the putative parental genotypes were then compared with the actual offspring genotype ratios obtained. In this manner, the presence of one or more null alleles at the three loci was assessed.

### 2.6.4 Assessing the Effect of Captivity on Genetic Diversity

Collection of the populations took place between 1997 and 1999. Upon collection, populations were housed in twelve separate tanks on the tilapia research farm (Welgevallen, Stellenbosch). At the time of initial sampling for this study (March 1999) the Amatikulu, Bushmans, Elands, Olifants, Kasinthula, Sucomba, Nick James and Valley populations had been in captivity for approximately one year. The Ndumu, Makathini, Verloerenvlei and Le Pommier populations were collected at the

end of 1998 and had therefore been in captivity for about six months at initial sampling. When the second larger samples were taken (March 2000) most of the populations had been in captivity for one and a half to two years. During time spent in captivity, the populations underwent periodic population size reductions due to natural deaths and unforeseen infrastructure failures. In order to maintain the population numbers, offspring from commercial breeding groups (group ratio - one male:three females) were used to supplement population numbers. This practice led to the presence of overlapping generations in the tanks and to the potential loss of genetic diversity over time due to the possibility of unbalanced sex ratios (decreasing effective population size and increasing the rate of genetic drift), and possible captive bottlenecks.

In order to ascertain whether gene diversity decreased whilst in captivity, the Bushmans population was used to compare gene diversity between years. The sample of 25 fish taken from the Bushmans population in March 1999 (consisting mainly of original fish collected from the wild and possibly a few offspring) was compared with a sample of 36 fish taken from the population in March of 2000 (consisting of original parent fish and more of their offspring). Sampling was not totally exclusive. Fish from the 1999 sample were possibly resampled as part of the 2000 sample however the proportion of original fish was probably much lower. Samples were compared for significant changes in allele frequencies and heterozygosity. This was done using fluorescent labelling of PCR products for the UNH104 and UNH123 loci on the ABI377 automated sequencer. The UNH111 locus was excluded from the analysis, as the use of two loci was considered sufficient for detecting allele frequency and heterozygosity changes. TFPGA and MSEXCEL were used to perform calculations and construct tables of allele frequencies and heterozygosity. A TFPGA Exact test for allele frequency heterogeneity between years was performed. A *t*-test was used to compare mean observed and expected heterozygosities between years (MSEXCEL).

### 2.6.5 Population Genetic Structure Analysis

To detect the presence of genetic structure between the populations, an Exact test for allele frequency heterogeneity between all the populations was performed using a Monte Carlo algorithm of 1000 dememorization steps, 10 batches and 2000 permutations per batch (TFPGA).

The following measures $F_{ST}$ (FSTAT), $R_{ST}$ (RST-CALC), $\theta_{ST}$ (GDA) and $\Phi_{ST}$ (ARLEQUIN 2.0), were used to quantify genetic structure between the populations. The significance of the overall $F_{ST}$ value was calculated using MSEXCEL and a significance formula (Workman and Niswander, 1970). Individual pair wise differentiation values were obtained between all population pairs using FSTAT. The RST-CALC program facilitated the calculation of a matrix of number of migrants ($N_m$) for the populations of wild origin.

# APPENDIX 1

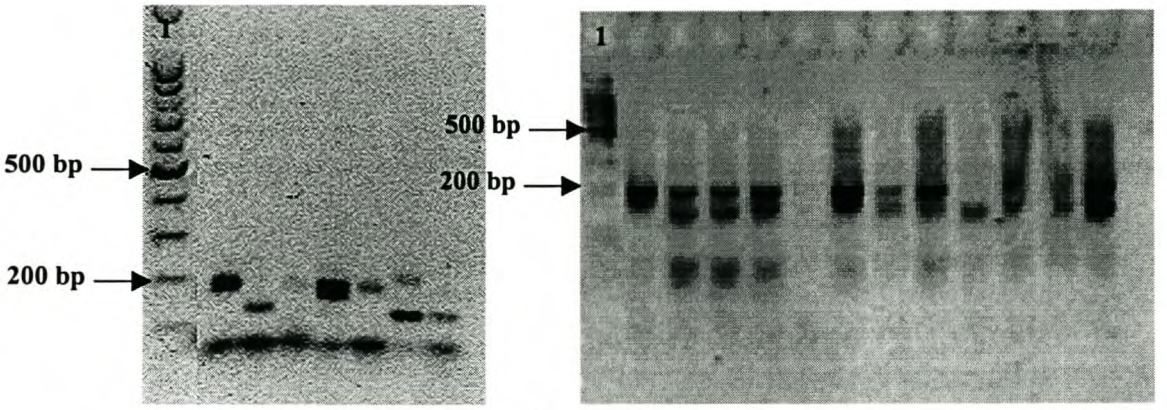


**Figure 1**: Diagrams of *O. mossambicus* PCR amplified products for the UNH 104 locus (2% agarose). Lane one contains a 100 base pair DNA ladder

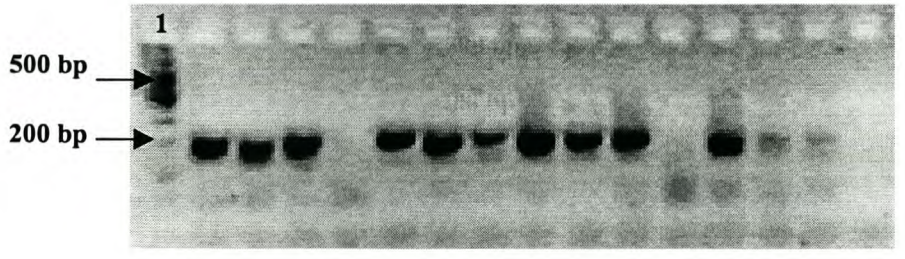


**Figure 2**: Diagram of *O. mossambicus* PCR amplified products for the UNH 111 locus (2% agarose). Lane one contains a 100 base pair DNA ladder

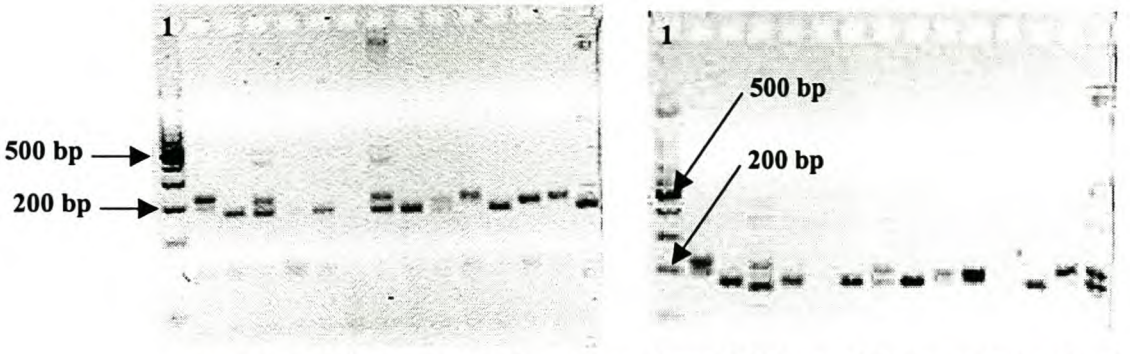


**Figure 3**: Diagrams of *O. mossambicus* PCR amplified products for the UNH 123 locus (2% agarose). Lane one contains a 100 base pair DNA ladder

Several genetic distance matrices were calculated for use in the construction of MDS plots and phylogenetic trees and in determining the affect of geographic locality on genetic structure using Mantel tests. TFPGA was used to generate $D_m$ distance matrices, using the loci separately and combined, for the populations of wild origin only and for all the populations. RST-CALC was used to construct $(\delta\mu)^2$ distance matrices, using the loci separately and combined, for the populations of wild origin only and for all the populations. The RST-CALC and TFPGA programs were sensitive to the absence of alleles at the UNH104 locus for the Nick James and Valley populations therefore these two populations were excluded from certain matrices. The Mantel test option of TFPGA was used to check for association between the locus specific genetic distance matrices of all the populations and of populations of wild origin only. A significant association would indicate a possible geographic background to the genetic structure of the populations. Mantel tests for association between the $D_m$ and $(\delta\mu)^2$ genetic distance matrices and a geographic distance matrix were also performed.

### 2.6.6 Determining Phylogenetic Relationships between the Populations

The PHYLIP Neighbour program was used to generate UPGMA and Neighbour-Joining 'tree files' from the $(\delta\mu)^2$ and $D_m$ matrices. The Drawgram program of PHYLIP was used to construct the final phylogenetic trees. This program allowed for clear construction and storage of the trees as compressed graphics images however, it did not insert a scale indicating the relative measure of genetic distance between the populations. The ratio of genetic distance to actual distance (cm) on the trees was therefore included in the trees' figure captions.

The multi-dimensional scaling option (MDS) of the SPSS (version 10) statistical package was used to calculate derived stimulus coordinates for each population based on the genetic distance values. Two-dimensional plots relating the populations based on best-fit Euclidean distances were constructed from the derived stimulus coordinates for $D_m$ and $(\delta\mu)^2$ distances. The $(\delta\mu)^2$ plot consisted of only ten populations due to the absence of genetic distance data for the Nick James and Valley populations. Non-linear fit plots of Euclidian distances against the appropriately scaled true genetic distances were constructed to indicate the degree of accuracy to which the Euclidean distances represent the genetic distances in two dimensions.

# CHAPTER THREE

## *RESULTS*

## 3.1 LABORATORY PROTOCOL – RESULTS

### 3.1.1 Fragment Analysis Optimisation Using the ABI 377 Automated Sequencer

#### 3.1.1.1 Amount of Template DNA

Addition of 100ng of template to the PCR reactions was more effective than 20ng.

#### 3.1.1.2 Fluorescent Nucleotide Ratio

The 500:1 ratio gave PCR products with ample signal for detection by the ABI 377 laser whereas the 1000:1 and 125:1 ratios gave signals that were too low (not visible as peaks) and too high (peaks were higher than the maximum tolerated by the Genotyper software for accurate size designation) respectively.

#### 3.1.1.3 Dilution of PCR Product

All three dilutions resulted in insufficient signal strength therefore no dilution of PCR product was made for the final analysis.

### 3.1.2 PCR Success Rate

The PCR success rate for UNH104 was 55% calculated as 157 successful PCR products obtained out of 288 attempted PCRs.

The PCR success rate for UNH111 was 68% calculated as 195 successful PCR products obtained out of 288 attempted PCRs.

The PCR success rate for UNH123 was 82% calculated as 237 successful PCR products obtained out of 288 attempted PCRs.

## 3.2 GENETIC VARIATION AT THE LOCI

### 3.2.1 Number of Alleles

The mean number of alleles for the *O. mossambicus* populations was 13.3 (Table 3.01). The number of alleles over the three loci increased with the mean allele size (see Table 3.02 for mean allele sizes calculated as the average repeat count multiplied by two).

**Table 3.01**: The number of alleles and mean allele size present at the loci

| Locus | Sample size | Mean allele size (bp) | Number of alleles |
|---|---|---|---|
| UNH104 | 157 | 138 | 7 |
| UNH111 | 195 | 186 | 16 |
| UNH123 | 237 | 186 | 17 |
| | | | 13.3 (Mean) |

### 3.2.2 Allele Frequencies

UNH104 had the 128 allele as the most frequent, with six alleles larger than 128 base pairs and no smaller alleles. UNH111 had the 180 allele as the most frequent with four alleles smaller than 180 base pairs and eleven alleles larger. The UNH123 locus had the 170 allele as the most common with four alleles smaller than 170 base pairs and twelve larger (Table 3.02).

**Table 3.02**: Allele frequencies present at the loci

| UNH104 (N = 157) | | UNH111 (N = 195) | | UNH123 (N = 237) | |
|---|---|---|---|---|---|
| Allele size | Frequency | Allele size | Frequency | Allele size | Frequency |
| 128 | 0.65 | 150 | 0.08 | 160 | 0.01 |
| 130 | 0.01 | 174 | 0.00 | 162 | 0.08 |
| 132 | 0.10 | 176 | 0.01 | 164 | 0.04 |
| 136 | 0.13 | 178 | 0.23 | 168 | 0.01 |
| 138 | 0.02 | 180 | 0.45 | 170 | 0.35 |
| 148 | 0.08 | 182 | 0.01 | 184 | 0.02 |
| 160 | 0.01 | 184 | 0.03 | 186 | 0.04 |
| Mean (138) | | 186 | 0.01 | 188 | 0.06 |
| | | 188 | 0.02 | 190 | 0.03 |
| | | 192 | 0.02 | 194 | 0.03 |
| | | 194 | 0.03 | 196 | 0.06 |
| | | 196 | 0.03 | 198 | 0.10 |
| | | 198 | 0.01 | 200 | 0.01 |
| | | 202 | 0.04 | 202 | 0.02 |
| | | 204 | 0.02 | 204 | 0.11 |
| | | 208 | 0.01 | 206 | 0.01 |
| | | Mean (186) | | 210 | 0.03 |
| | | | | Mean (186) | |

### 3.2.3 Heterozygosity

Average observed heterozygosity was low (0.35). Average expected heterozygosity was comparably high (0.71). The UNH104 had a substantially reduced observed heterozygosity compared to the other two loci (Table 3.03).

**Table 3.03**: Observed and expected heterozygosity

| Locus | Sample size | Observed Heterozygosity | Expected Heterozygosity |
|---|---|---|---|
| UNH104 | 157 | 0.08 | 0.55 |
| UNH111 | 195 | 0.42 | 0.73 |
| UNH123 | 237 | 0.53 | 0.84 |
| Average | 197 | 0.35 | 0.71 |

### 3.2.4 Test for Hardy Weinberg Equilibrium (HWE)

The $\chi^2$ test, using pooled genotypes (Table 3.04) indicated highly significant deviation from HWE at the loci. The pooled genotype data showed a reduced number of observed heterozygotes compared to the expected number.

**Table 3.04**: Chi–square tests for deviation from HWE using pooled genotypes

| UNH104 | | | UNH111 | | | UNH123 | | |
|---|---|---|---|---|---|---|---|---|
| Genotype | Observed | Expected | Genotype | Observed | Expected | Genotype | Observed | Expected |
| 1/1 | 98 | 66.2675 | 5/5 | 73 | 39.7128 | 5/5 | 61 | 29.0675 |
| 1/* | 8 | 71.465 | 5/* | 30 | 96.5744 | 5/* | 44 | 107.865 |
| */* | 51 | 19.2675 | */* | 92 | 58.7128 | */* | 132 | 100.068 |
| Chi-square | 123.817 | df = 1 | Chi-square | 92.667 | df = 1 | Chi-square | 83.0832 | df = 1 |
| $P < 0.001$ | | | $P < 0.001$ | | | $P < 0.001$ | | |
| * = Pooled Alleles<br>1 = Most common allele type at the UNH104 locus | | | * = Pooled Alleles<br>5 = Most common allele type at the UNH111 locus | | | * = Pooled Alleles<br>5 = Most common allele type at the UNH123 locus | | |

Monte Carlo Exact tests on non-pooled genotypes for deviation from HWE were highly significant for all three loci (Table 3.05).

**Table 3.05**: Monte Carlo Exact tests for deviation from HWE

| Locus | Probability (*P*) | Standard Error (S. E.) |
|---|---|---|
| UNH104 | < 0.001 | < 0.0000001 |
| UNH111 | < 0.001 | < 0.0000001 |
| UNH123 | < 0.001 | < 0.0000001 |
| Randomisation procedure: 10 batches – 1000 Permutations per batch | | |

## 3.3 GENETIC VARIATION WITHIN POPULATIONS

### 3.3.1 Allele Frequencies

For UNH104, all the populations except the Nick James and Valley population had their most common allele present at a frequency greater than or equal to 50%. The 160 allele was unique to the Elands population. The 148 allele was unique to the Bushmans population. The 160 and 148 alleles were both high frequency alleles and may be useful as population specific alleles for identifying members of the Elands and Bushmans populations respectively. Members of the Nick James and Valley populations did not produce any PCR products for the UNH 104 locus (Table 3.06).

**Table 3.06**: Allele frequencies at UNH104 for the twelve populations

| | Population | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Allele** | A{f} (36) [5] | B{w} (36) [27] | E{w} (9) [4] | K{w} (36) [4] | L{in} (36) [32] | M{f} (10) [6] | N{w} (36) [30] | NJ{f} (10) [0] | O{w} (10) [4] | S{w} (36) [33] | V{f} (17) [0] | VE{in} (16) [12] |
| 128 | 0.60 | 0.15 | 0.25 | 0.75 | 0.83 | 0.50 | 0.93 | 0.00 | 0.50 | 0.92 | 0.00 | 0.08 |
| 130 | 0.00 | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 |
| 132 | 0.00 | 0.15 | 0.00 | 0.25 | 0.03 | 0.50 | 0.05 | 0.00 | 0.38 | 0.05 | 0.00 | 0.08 |
| 136 | 0.40 | 0.20 | 0.00 | 0.00 | 0.14 | 0.00 | 0.00 | 0.00 | 0.13 | 0.00 | 0.00 | 0.67 |
| 138 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.17 |
| 148 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 160 | 0.00 | 0.00 | 0.50 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | | | |
| Values in ( ) are sample sizes for which PCR was attempted | | | | | | | | | | | | |
| Values in [ ] are sample sizes for which PCR was successful. | | | | | | | | | | | | |

Ten of the populations had their most common allele present at a frequency greater than or equal to 50% for the UNH111 locus. The 174 allele was unique to the Sucomba population however at a low frequency (0.02) making it unsuitable as a population identifier. The 198 allele was unique to the Bushmans population however also at a low frequency (0.05). The 208 allele was unique to the Nick James population at a high frequency of 67% making this allele a possible identifier for members of the Nick James population. The Amatikulu population was fixed for the 180 allele (Table 3.07).

**Table 3.07**: Allele frequencies at UNH111 for the twelve populations

| | **Population** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Allele** | A{f} (36) [17] | B{w} (36) [32] | E {w} (9) [7] | K{w} (36) [13] | L{in} (36) [33] | M{f} (10) [8] | N{w} (36) [28] | NJ{f} (10) [3] | O{w} (10) [7] | S{w} (36) [27] | V{f} (17) [5] | VE{in} (16) [15] |
| 150 | 0.00 | 0.02 | 0.21 | 0.00 | 0.12 | 0.00 | 0.20 | 0.00 | 0.00 | 0.17 | 0.00 | 0.00 |
| 174 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 |
| 176 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 |
| 178 | 0.00 | 0.50 | 0.57 | 0.50 | 0.20 | 0.25 | 0.21 | 0.00 | 0.21 | 0.04 | 0.00 | 0.03 |
| 180 | 1.00 | 0.06 | 0.00 | 0.08 | 0.41 | 0.75 | 0.48 | 0.33 | 0.64 | 0.61 | 0.80 | 0.60 |
| 182 | 0.00 | 0.06 | 0.00 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 184 | 0.00 | 0.02 | 0.00 | 0.27 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.09 | 0.00 | 0.00 |
| 186 | 0.00 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 188 | 0.00 | 0.00 | 0.07 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.14 | 0.02 | 0.00 | 0.03 |
| 192 | 0.00 | 0.06 | 0.00 | 0.00 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 | 0.20 | 0.00 |
| 194 | 0.00 | 0.08 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 |
| 196 | 0.00 | 0.00 | 0.00 | 0.08 | 0.03 | 0.00 | 0.02 | 0.00 | 0.00 | 0.02 | 0.00 | 0.20 |
| 198 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 202 | 0.00 | 0.02 | 0.14 | 0.00 | 0.18 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 204 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 208 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.67 | 0.00 | 0.00 | 0.00 | 0.00 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | | | |
| Values in ( ) are sample sizes for which PCR was attempted | | | | | | | | | | | | |
| Values in [ ] are sample sizes for which PCR was successful. | | | | | | | | | | | | |

Seven of the populations had their most common allele present at a frequency greater than or equal to 50% for the UNH123 locus. The 168 allele was unique to the Kasinthula population at a low frequency (0.05). No other population unique alleles were identified. The Valley population was fixed for the 204 allele (Table 3.08).

**Table 3.08:** Allele frequencies at UNH123 for the twelve populations

| | Population | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Allele** | A{f} (36) [32] | B{w} (36) [34] | E{w} (9) [7] | K{w} (36) [28] | L{in} (36) [33] | M{f} (10) [9] | N{w} (36) [32] | NJ{f} (10) [8] | O{w} (10) [6] | S{w} (36) [34] | V{f} (17) [7] | VE{in} (16) [7] |
| 160 | 0.00 | 0.00 | 0.00 | 0.02 | 0.00 | 0.00 | 0.06 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 162 | 0.19 | 0.03 | 0.00 | 0.02 | 0.20 | 0.00 | 0.02 | 0.00 | 0.00 | 0.06 | 0.00 | 0.43 |
| 164 | 0.00 | 0.00 | 0.57 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.15 | 0.00 | 0.00 |
| 168 | 0.00 | 0.00 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 170 | 0.61 | 0.43 | 0.36 | 0.25 | 0.03 | 0.44 | 0.77 | 0.56 | 0.92 | 0.00 | 0.00 | 0.00 |
| 184 | 0.00 | 0.07 | 0.00 | 0.07 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 186 | 0.00 | 0.04 | 0.00 | 0.27 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 188 | 0.00 | 0.01 | 0.00 | 0.02 | 0.21 | 0.00 | 0.00 | 0.00 | 0.00 | 0.21 | 0.00 | 0.00 |
| 190 | 0.00 | 0.10 | 0.00 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.00 |
| 194 | 0.00 | 0.12 | 0.00 | 0.00 | 0.02 | 0.00 | 0.05 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 |
| 196 | 0.00 | 0.03 | 0.00 | 0.02 | 0.18 | 0.00 | 0.00 | 0.00 | 0.00 | 0.18 | 0.00 | 0.00 |
| 198 | 0.00 | 0.03 | 0.07 | 0.13 | 0.21 | 0.00 | 0.02 | 0.00 | 0.00 | 0.34 | 0.00 | 0.00 |
| 200 | 0.00 | 0.01 | 0.00 | 0.00 | 0.02 | 0.00 | 0.05 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 202 | 0.00 | 0.09 | 0.00 | 0.02 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| 204 | 0.20 | 0.00 | 0.00 | 0.04 | 0.00 | 0.56 | 0.02 | 0.44 | 0.00 | 0.06 | 1.00 | 0.00 |
| 206 | 0.00 | 0.00 | 0.00 | 0.04 | 0.00 | 0.00 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 |
| 210 | 0.00 | 0.03 | 0.00 | 0.00 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.43 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | | | |
| Values in ( ) are sample sizes for which PCR was attempted | | | | | | | | | | | | |
| Values in [ ] are sample sizes for which PCR was successful. | | | | | | | | | | | | |

### 3.3.2 Population Heterozygosity

Population observed heterozygosities were relatively low. Heterozygosities were particularly low for the red coloured populations, Amatikulu, Nick James and Valley. The Bushmans population had the highest expected heterozygosity (0.73). Overall heterozygosities in populations from wild sources were greater than populations from introduced/farm stocks. The average number of alleles varied greatly between introduced/farm and wild populations. The introduced/farm populations had an average of 8.8 alleles (six populations) while wild populations had an average of 17 alleles (six populations). Ten of the populations were polymorphic (the most common allele present at a frequency of 95% or less) for the three loci UNH104, UNH111 and UNH123. The Nick James and Valley populations did not amplify for the UNH104 locus. The Valley population was fixed for UNH123. Amatikulu was fixed for UNH111. Polymorphism percentage values for the populations were calculated from loci that did amplify for a particular population (Table 3.09).

**Table 3.09**: Heterozygosity, fixation indices ($F_{IS}$), number of alleles and percentage polymorphism averaged across the loci for the twelve populations

| | Heterozygosity | | | Number of alleles | | | |
|---|---|---|---|---|---|---|---|
| **Population** | **$H_o$** | **$H_e$** | **$F_{IS}$** | **UNH104** | **UNH111** | **UNH123** | **% Polymorphism** |
| A{f} (72) | 0.21 | 0.36 | 0.42 | 2 | 1 | 3 | 67 |
| B{w} (72) | 0.37 | 0.73 | 0.49 | 4 | 11 | 12 | 100 |
| E{w} (18) | 0.38 | 0.65 | 0.42 | 3 | 4 | 3 | 100 |
| K{w} (72) | 0.41 | 0.66 | 0.38 | 2 | 6 | 13 | 100 |
| L{in} (72) | 0.58 | 0.63 | 0.08 | 3 | 8 | 11 | 100 |
| M{f} (20) | 0.33 | 0.49 | 0.33 | 2 | 2 | 2 | 100 |
| N{w} (72) | 0.24 | 0.41 | 0.41 | 3 | 7 | 8 | 100 |
| NJ{f} (20) | 0.19 | 0.53 | 0.64 | 0 | 2 | 2 | 67 |
| O{w} (20) | 0.23 | 0.47 | 0.51 | 3 | 3 | 2 | 100 |
| S{w} (72) | 0.37 | 0.51 | 0.27 | 3 | 8 | 7 | 100 |
| V{f} (34) | 0.00 | 0.18 | 1.00 | 0 | 2 | 1 | 33 |
| VE{in} (32) | 0.35 | 0.60 | 0.42 | 4 | 5 | 3 | 100 |
| Average | 0.31 | 0.52 | 0.40 | 2.4 | 4.9 | 5.6 | 88 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | |
| Values in ( ) are 2N sample sizes | | | | | | | |

### 3.3.3 Hardy Weinberg Equilibrium

All twelve populations exhibited significant deviation from Hardy Weinberg Equilibrium after $\chi^2$ tests (Table 3.10). Ten of the twelve populations exhibited significant deviation from Hardy Weinberg Equilibrium after Exact tests (Table 3.11).

**Table 3.10**: Chi–square test *P* values for population HWE using pooled genotypes

| **Population** | **UNH104** | **UNH111** | **UNH123** | **Overall (*P*)** |
|---|---|---|---|---|
| A{f} | 0.7094 | ? | 0.0001 | < 0.001 |
| B{w} | < 0.001 | 0.0047 | 0.0075 | < 0.001 |
| E{w} | 0.0455 | 0.0472 | 0.2703 | 0.001 |
| K{w} | 0.0455 | 0.0769 | 0.0039 | < 0.001 |
| L{in} | < 0.001 | 0.2874 | 0.1220 | < 0.001 |
| M{f} | 0.4142 | 0.0047 | 0.2937 | 0.001 |
| N{w} | < 0.001 | < 0.001 | 0.2211 | < 0.001 |
| NJ{f} | ? | 0.0833 | 0.5007 | 0.042 |
| O{w} | 0.0455 | 0.0684 | 0.8238 | 0.003 |
| S{w} | < 0.001 | 0.0015 | 0.1477 | < 0.001 |
| V{f} | ? | 0.0253 | ? | 0.025 |
| VE{in} | 0.0304 | 0.6670 | 0.0082 | < 0.001 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | |
| ? = calculation not performed due to only one allele or zero alleles present at the locus | | | | |

**Table 3.11**: Exact test *P* values for population HWE using all genotypes

| Population | UNH104 | UNH111 | UNH123 | Overall (*P*) |
|---|---|---|---|---|
| A{f} | 1.0000 | ? | < 0.001 | < 0.001 |
| B{w} | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| E{w} | 0.0300 | 0.7800 | 0.1000 | 0.002 |
| K{w} | 0.1400 | < 0.001 | < 0.001 | < 0.001 |
| L{in} | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| M{f} | 0.4800 | 0.0100 | 0.5500 | 0.003 |
| N{w} | < 0.001 | < 0.001 | 0.0600 | < 0.001 |
| NJ{f} | ? | 0.2000 | 0.5300 | 0.106 |
| O{w} | 0.0900 | 0.1000 | 1.0000 | 0.009 |
| S{w} | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| V{f} | ? | 0.1100 | ? | 0.110 |
| VE{in} | < 0.001 | 0.4200 | 0.0200 | < 0.001 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | |
| ? = calculation not performed due to only one allele or zero alleles present at the locus | | | | |

### 3.3.4 Linkage Disequilibrium

The Bushmans and Le Pommier populations exhibited linkage disequilibrium for all loci combinations. The Kasinthula population indicated linkage disequilibrium for the UNH111/UNH123 combination. The Sucomba population indicated linkage disequilibrium for the UNH104/UNH123 and UNH104/111 combination.

**Table 3.12**: Fisher Exact test probability *P* values of gametic linkage disequilibrium ($\alpha = 0.05$)

| Population | UNH104 / UNH111 | UNH104 / UNH123 | UNH111 / UNH123 |
|---|---|---|---|
| A{f} | 1.0000 | 0.6234 | 1.0000 |
| B{w} | 0.0003 | 0.0000 | 0.0009 |
| E{w} | 0.1703 | 0.1659 | 0.6238 |
| K{w} | 1.0000 | 1.0000 | 0.0059 |
| L{in} | 0.0013 | 0.0000 | 0.0288 |
| M{f} | 1.0000 | 1.0000 | 1.0000 |
| N{w} | 0.1697 | 0.2606 | 0.0509 |
| NJ{f} | 1.0000 | 1.0000 | 1.0000 |
| O{w} | 1.0000 | 1.0000 | 1.0000 |
| S{w} | 0.0338 | 0.0038 | 0.6284 |
| V{f} | 1.0000 | 1.0000 | 1.0000 |
| VE{in} | 0.0800 | 0.3450 | 0.7178 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | |

### 3.3.5 Microsatellite Inheritance

Inheritance patterns were investigated by examining parent/offspring genotypes as follows:

**Table 3.14**: Inheritance of microsatellite alleles at the UNH104 locus for a controlled cross between a Ndumu male and an Olifants female

| Observed alleles | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Parents** | | **Offspring** | | | | | | | | | |
| **Male** | **Female** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | ? | ? | 128 | 128 |
| ? = no alleles present | | | | | | | | | | | |

**Table 3.15**: Inheritance of microsatellite alleles at the UNH104 locus for a controlled cross between a Valley male and an Olifants female

| Observed alleles | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Parents** | | **Offspring** | | | | | | | | | |
| **Male** | **Female** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 136 | 128<br>136 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128 | 128<br>136 |

**Table 3.16**: Inheritance of microsatellite alleles at the UNH111 locus for a controlled cross between a Ndumu male and an Olifants female

| Observed alleles | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Parents** | | **Offspring** | | | | | | | | | |
| **Male** | **Female** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 182 | 150 | 182<br>150 | 182<br>150 | 150 | 150 | 182<br>150 | ? | 182<br>150 | 182<br>150 | ? | 150 |
| ? = no alleles present | | | | | | | | | | | |

**Table 3.17**: Inheritance of microsatellite alleles at the UNH111 locus for a controlled cross between a Valley male and an Olifants female

| Observed alleles | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Parents** | | **Offspring** | | | | | | | | | |
| **Male** | **Female** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 180 | 180<br>150 | 180 | 180<br>150 | 180<br>150 | 180<br>150 | 180 | 180 | 180<br>150 | 180 | 180<br>150 | 180<br>150 |

**Table 3.18**: Inheritance of microsatellite alleles at the UNH123 locus for a controlled cross between a Ndumu male and an Olifants female

| Observed alleles | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Parents** | | **Offspring** | | | | | | | | | |
| **Male** | **Female** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 170 | 170 | 170 | 170 | 170 | 170 | 170 | 170 | 170 | ? | 170 | 170 |
| ? = no alleles present | | | | | | | | | | | |

**Table 3.19**: Inheritance of microsatellite alleles at the UNH123 locus for a controlled cross between a Valley male and an Olifants female

| Observed alleles | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Parents** | | **Offspring** | | | | | | | | | |
| **Male** | **Female** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 170<br>204 | 170 | 170<br>204 | 170<br>204 | 170<br>204 | 170<br>204 | 170<br>204 | 170<br>204 | 170<br>204 | 170<br>204 | 170<br>192* | 162*<br>192* |
| * possible non-specific PCR products | | | | | | | | | | | |

## 3.4 THE EFFECT OF CAPTIVITY ON GENETIC DIVERSITY

### 3.4.1 Change in Allele Frequency

Successive calculation of genetic variation and diversity between the years 1999 and 2000 for the Bushmans population showed significant changes in allele frequency at the UNH104 ($P = 0.0007$) and UNH123 ($P = 0.0002$) loci. One low frequency allele present in the 1999 sample was absent in the 2000 sample at the UNH104 locus. Six low frequency alleles present in the 1999 sample were absent in the 2000 sample at the UNH123 locus (Table 3.20).

**Table 3.20**: Fisher's Exact test for allele frequency heterogeneity between the 1999 and 2000 Bushmans samples at the UNH104 and UNH123 loci

| **Locus UNH104** | | | | | | |
|---|---|---|---|---|---|---|
| | **Allele** | | | | | |
| | 128 | 130 | 132 | 136 | 148 | Total |
| **Sample** | | | | | | |
| 1999 | 15 | 2 | 6 | 0 | 11 | 34 |
| 2000 | 8 | 0 | 8 | 11 | 27 | 54 |
| Total | 23 | 2 | 14 | 11 | 38 | 88 |
| Probability (*P*) of homogeneity at UNH104 ($P = 0.0007$) (S.E. :0.0005) | | | | | | |

| **LocusUNH123** | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Allele** | | | | | | | | | | | | | | | | | | |
| | 162 | 164 | 170 | 178 | 184 | 186 | 188 | 190 | 194 | 196 | 198 | 200 | 202 | 204 | 206 | 210 | 214 | 216 | Total |
| **Sample** | | | | | | | | | | | | | | | | | | | |
| 1999 | 2 | 1 | 21 | 1 | 0 | 1 | 2 | 0 | 1 | 1 | 3 | 0 | 1 | 4 | 1 | 0 | 2 | 1 | 42 |
| 2000 | 2 | 0 | 29 | 0 | 5 | 3 | 1 | 7 | 8 | 2 | 2 | 1 | 6 | 0 | 0 | 2 | 0 | 0 | 68 |
| Total | 4 | 1 | 50 | 1 | 5 | 4 | 3 | 7 | 9 | 3 | 5 | 1 | 7 | 4 | 1 | 2 | 2 | 1 | 110 |
| Probability (*P*) of homogeneity at UNH123 ($P = 0.0002$) (S.E. :0.0002) | | | | | | | | | | | | | | | | | | | |
| Results over loci ($P < 0.001$) | | | | | | | | | | | | | | | | | | | |

### 3.4.2 Change in Heterozygosity

Observed heterozygosity increased significantly between 1999 and 2000 based on a *t*-test for significant change in mean observed heterozygosity ($P$ = 0.003). There were no significant differences in expected heterozygosity between the 1999 sample and the 2000 sample after using a *t*-test ($P$ = 0.631) (Table 3.21).

**Table 3.21**: Table of *t*-test for change in mean expected heterozygosity and mean observed heterozygosity between the 1999 Bushmans and 2000 Bushmans samples

| | **Expected heterozygosity** | | **Observed heterozygosity** | |
|---|---|---|---|---|
| **Locus** | ***1999*** | ***2000*** | ***1999*** | ***2000*** |
| UNH104 | 0.6863 | 0.6771 | 0.0588 | 0.1111 |
| UNH123 | 0.7422 | 0.7862 | 0.4762 | 0.5294 |
| ***t*-Test: Paired two sample test for change in heterozygosity** | | | | |
| | **Expected heterozygosity** | | **Observed heterozygosity** | |
| | ***1999*** | ***2000*** | ***1999*** | ***2000*** |
| Mean | 0.71425 | 0.73165 | 0.2675 | 0.32025 |
| Variance | 0.001562 | 0.005951 | 0.087111 | 0.087487 |
| Df | 1 | | 1 | |
| Probability ($P$) | 0.631221 | | 0.002715 | |

## 3.5 POPULATION GENETIC STRUCTURE

The measures used to detect population structuring all provided highly significant evidence of structuring between the twelve populations.

### 3.5.1 Exact test for Allele Frequency Heterogeneity Between Populations

The Exact tests performed on the allelic data set yielded ($P < 0.001$) indicative of significant allele frequency heterogeneity between the populations at all three loci (Table 3.22 – 3.24).

**Table 3.22**: Fisher's Exact test for allele frequency heterogeneity between populations at UNH104

| | Allele | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Population** | 128 | 130 | 132 | 136 | 138 | 148 | 160 | Total |
| A{f} | 6 | 0 | 0 | 4 | 0 | 0 | 0 | 10 |
| B{w} | 8 | 0 | 8 | 11 | 0 | 27 | 0 | 54 |
| E{w} | 2 | 2 | 0 | 0 | 0 | 0 | 4 | 8 |
| K{w} | 6 | 0 | 2 | 0 | 0 | 0 | 0 | 8 |
| L{in} | 53 | 0 | 2 | 9 | 0 | 0 | 0 | 64 |
| M{f} | 6 | 0 | 6 | 0 | 0 | 0 | 0 | 12 |
| N{w} | 56 | 0 | 3 | 0 | 1 | 0 | 0 | 60 |
| O{w} | 4 | 0 | 3 | 1 | 0 | 0 | 0 | 8 |
| S{w} | 61 | 2 | 3 | 0 | 0 | 0 | 0 | 66 |
| VE{in} | 2 | 0 | 2 | 16 | 4 | 0 | 0 | 24 |
| Total | 204 | 4 | 29 | 41 | 5 | 27 | 4 | 314 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | |
| Probability ($P$) of homogeneity at UNH104 ($P < 0.001$) (S.E.<0.0001) | | | | | | | | |

**Table 3.23**: Fisher's Exact test for allele frequency heterogeneity between populations at UNH111

| | Alleles | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Population** | 150 | 174 | 176 | 178 | 180 | 182 | 184 | 186 | 188 | 192 | 194 | 196 | 198 | 202 | 204 | 208 | Total |
| A{f} | 0 | 0 | 0 | 0 | 34 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 34 |
| B{w} | 1 | 0 | 0 | 32 | 4 | 4 | 1 | 4 | 0 | 4 | 5 | 0 | 3 | 1 | 5 | 0 | 64 |
| E{w} | 3 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 14 |
| K{w} | 0 | 0 | 1 | 13 | 2 | 1 | 7 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 26 |
| L{in} | 8 | 0 | 0 | 13 | 27 | 0 | 0 | 0 | 1 | 1 | 2 | 2 | 0 | 12 | 0 | 0 | 66 |
| M{f} | 0 | 0 | 0 | 4 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 16 |
| N{w} | 11 | 0 | 0 | 12 | 27 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 3 | 0 | 56 |
| NJ{f} | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 6 |
| O{w} | 0 | 0 | 0 | 3 | 9 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 |
| S{w} | 9 | 1 | 2 | 2 | 33 | 0 | 5 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 54 |
| V{f} | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| VE{in} | 0 | 0 | 0 | 1 | 18 | 0 | 0 | 0 | 1 | 0 | 4 | 6 | 0 | 0 | 0 | 0 | 30 |
| Total | 32 | 1 | 3 | 88 | 176 | 5 | 13 | 5 | 6 | 8 | 11 | 12 | 3 | 15 | 8 | 4 | 390 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | | | | | | | | |
| Probability (*P*) of homogeneity at UNH111 ($P < 0.001$) (S.E. < 0.0001) | | | | | | | | | | | | | | | | | |

**Table 3.24**: Fisher's Exact test for allele frequency heterogeneity between populations at UNH123

| | Alleles | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Population** | 160 | 162 | 164 | 168 | 170 | 184 | 186 | 188 | 190 | 194 | 196 | 198 | 200 | 202 | 204 | 206 | 210 | Total |
| A{f} | 0 | 12 | 0 | 0 | 39 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 64 |
| B{w} | 0 | 2 | 0 | 0 | 29 | 5 | 3 | 1 | 7 | 8 | 2 | 2 | 1 | 6 | 0 | 0 | 2 | 68 |
| E{w} | 0 | 0 | 8 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 14 |
| K{w} | 1 | 1 | 0 | 3 | 14 | 4 | 15 | 1 | 4 | 0 | 1 | 7 | 0 | 1 | 2 | 2 | 0 | 56 |
| L{in} | 0 | 13 | 0 | 0 | 2 | 1 | 1 | 14 | 0 | 1 | 12 | 14 | 1 | 2 | 0 | 0 | 5 | 66 |
| M{f} | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 18 |
| N{w} | 4 | 1 | 0 | 0 | 49 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 3 | 0 | 1 | 2 | 0 | 64 |
| NJ{f} | 0 | 0 | 0 | 0 | 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 16 |
| O{w} | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 |
| S{w} | 0 | 4 | 10 | 0 | 0 | 0 | 0 | 14 | 1 | 0 | 12 | 23 | 0 | 0 | 4 | 0 | 0 | 68 |
| V{f} | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 14 | 0 | 0 | 14 |
| VE{in} | 0 | 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 6 | 14 |
| Total | 5 | 39 | 18 | 3 | 166 | 10 | 19 | 30 | 12 | 13 | 27 | 48 | 5 | 9 | 51 | 6 | 13 | 474 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | | | | | | | | | |
| Probability (*P*) of homogeneity at UNH123 ($P < 0.001$) (S.E. < 0.0001) | | | | | | | | | | | | | | | | | | |

### 3.5.2 Quantifying Population Structure (F-Statistics)

An overall $F_{ST}$ value of 0.27 ($P < 0.001$), $\theta_{ST}$ of 0.26, $R_{ST}$ of 0.28 and a $\Phi_{ST}$ of 0.17 were obtained, typical of a high degree of structuring, between the twelve populations.

Pair wise difference *P* values after Bonferroni adjustment appear in Table 3.25.

**Table 3.25**: Population pair wise difference multiple comparison *P* values after Bonferroni correction

| **Population** | A | B | E | K | L | M | N | NJ | O | S | V | VE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A{f} | | 0.0002 | 0.0002 | 0.0002 | 0.0002 | 0.0139 | 0.0002 | 0.0785 | 0.0032 | 0.0002 | 0.0002 | 0.0002 |
| B{w} | * | | 0.0015 | 0.0026 | 0.0002 | 0.0002 | 0.0002 | 0.0517 | 0.0632 | 0.0002 | 0.0003 | 0.0002 |
| E{w} | * | NS | | 0.9941 | 0.0002 | 0.0215 | 0.0002 | 1.0000 | 0.8673 | 0.0002 | 0.9273 | 0.0002 |
| K{w} | * | NS | NS | | 0.0002 | 0.0062 | 0.0002 | 0.7006 | 0.9571 | 0.0002 | 0.2770 | 0.0002 |
| L{in} | * | * | * | * | | 0.0002 | 0.0002 | 0.0014 | 0.0282 | 0.0002 | 0.0002 | 0.0002 |
| M{f} | NS | * | NS | NS | * | | 0.0002 | 1.0000 | 1.0000 | 0.0002 | 1.0000 | 0.0005 |
| N{w} | * | * | * | * | * | * | | 0.9996 | 0.7767 | 0.0002 | 0.6276 | 0.0002 |
| NJ{f} | NS | NS | NS | NS | NS | NS | NS | | 1.0000 | 0.0002 | 1.0000 | 0.0499 |
| O{w} | NS | NS | NS | NS | NS | NS | NS | NS | | 0.0002 | 0.8123 | 0.0002 |
| S{w} | * | * | * | * | * | * | * | * | * | | 0.0599 | 0.0002 |
| V{f} | * | * | NS | NS | * | NS | NS | NS | NS | NS | | 0.1020 |
| VE{in} | * | * | * | * | * | * | * | NS | * | * | NS | |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | | | |
| Upper triangular matrix = *P* values<br>Lower triangular matrix = significance of *P* values<br>* = significant pair wise difference after multiple comparison<br>adjustment of the (5%) nominal value to 0.000758<br>NS = no significant pair wise difference | | | | | | | | | | | | |

### 3.5.3 Genetic Distance Matrices

**Table 3.26**: Matrix of $D_m$ for the twelve populations

| | A | B | E | K | L | M | N | NJ | O | S | V | VE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A{f} | 0.000 | | | | | | | | | | | |
| B{w} | 0.286 | 0.000 | | | | | | | | | | |
| E{w} | 0.371 | 0.149 | 0.000 | | | | | | | | | |
| K{w} | 0.260 | 0.124 | 0.158 | 0.000 | | | | | | | | |
| L{in} | 0.162 | 0.208 | 0.227 | 0.086 | 0.000 | | | | | | | |
| M{f} | 0.095 | 0.220 | 0.289 | 0.156 | 0.173 | 0.000 | | | | | | |
| N{w} | 0.108 | 0.221 | 0.247 | 0.116 | 0.123 | 0.132 | 0.000 | | | | | |
| NJ{f} | 0.213 | 0.226 | 0.327 | 0.260 | 0.259 | 0.133 | 0.161 | 0.000 | | | | |
| O{w} | 0.066 | 0.166 | 0.247 | 0.153 | 0.190 | 0.061 | 0.045 | 0.189 | 0.000 | | | |
| S{w} | 0.165 | 0.303 | 0.285 | 0.132 | 0.026 | 0.171 | 0.136 | 0.273 | 0.215 | 0.000 | | |
| V{f} | 0.269 | 0.501 | 0.603 | 0.478 | 0.347 | 0.103 | 0.433 | 0.294 | 0.472 | 0.289 | 0.000 | |
| VE{in} | 0.186 | 0.252 | 0.363 | 0.315 | 0.203 | 0.276 | 0.364 | 0.321 | 0.282 | 0.285 | 0.356 | 0.000 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | | | |

**Table 3.27**: Matrix of $(\delta\mu)^2$ for ten populations

| | A | B | E | K | L | M | N | O | S | VE |
|---|---|---|---|---|---|---|---|---|---|---|
| A{f} | 0.000 | | | | | | | | | |
| B{w} | 12.521 | 0.000 | | | | | | | | |
| E{w} | 19.874 | 23.249 | 0.000 | | | | | | | |
| K{w} | 6.183 | 11.212 | 41.107 | 0.000 | | | | | | |
| L{in} | 14.035 | 13.220 | 53.543 | 1.722 | 0.000 | | | | | |
| M{f} | 15.289 | 13.612 | 52.860 | 2.626 | 0.324 | 0.000 | | | | |
| N{w} | 2.499 | 24.088 | 24.062 | 10.653 | 19.745 | 20.293 | 0.000 | | | |
| O{w} | 1.050 | 18.275 | 19.054 | 11.506 | 21.969 | 23.913 | 2.876 | 0.000 | | |
| S{w} | 17.170 | 21.036 | 55.845 | 4.939 | 2.815 | 1.606 | 18.332 | 25.964 | 0.000 | |
| VE{in} | 18.812 | 5.309 | 48.232 | 6.966 | 4.512 | 5.142 | 30.583 | 27.354 | 12.171 | 0.000 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | |

**Table 3.28**: Matrix of $D_m$ for the populations collected from wild sources

| | B | E | K | N | O | S |
|---|---|---|---|---|---|---|
| B | 0.000 | | | | | |
| E | 0.149 | 0.000 | | | | |
| K | 0.124 | 0.158 | 0.000 | | | |
| N | 0.221 | 0.247 | 0.116 | 0.000 | | |
| O | 0.166 | 0.247 | 0.153 | 0.045 | 0.000 | |
| S | 0.303 | 0.285 | 0.132 | 0.136 | 0.215 | 0.000 |

**Table 3.29**: Matrix of $(\delta\mu)^2$ for the populations collected from wild sources

| | B | E | K | N | O | S |
|---|---|---|---|---|---|---|
| B | 0.000 | | | | | |
| E | 23.249 | 0.000 | | | | |
| K | 11.212 | 41.107 | 0.000 | | | |
| N | 24.088 | 24.062 | 10.653 | 0.000 | | |
| O | 18.275 | 19.054 | 11.506 | 2.876 | 0.000 | |
| S | 21.036 | 55.845 | 4.939 | 18.332 | 25.964 | 0.000 |

**Table 3.30**: Matrix of approximate geographic straight line distances between the populations (km)

| | A | B | E | K | L | M | N | NJ | O | S | V | VE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A{f} | 0 | | | | | | | | | | | |
| B{w} | 690 | 0 | | | | | | | | | | |
| E{w} | 500 | 980 | 0 | | | | | | | | | |
| K{w} | 1880 | 2420 | 1450 | 0 | | | | | | | | |
| L{in} | 1260 | 700 | 1340 | 1780 | 0 | | | | | | | |
| M{f} | 180 | 840 | 420 | 1690 | 1420 | 0 | | | | | | |
| N{w} | 240 | 900 | 380 | 1620 | 1450 | 70 | 0 | | | | | |
| NJ{f} | 640 | 50 | 920 | 2400 | 700 | 820 | 890 | 0 | | | | |
| O{w} | 580 | 1050 | 90 | 1400 | 1440 | 440 | 380 | 1020 | 0 | | | |
| S{w} | 1890 | 2430 | 1470 | 10 | 2790 | 1700 | 1630 | 2410 | 1410 | 0 | | |
| V{f} | 420 | 1030 | 210 | 1420 | 1500 | 320 | 250 | 1020 | 160 | 1430 | 0 | |
| VE{in} | 1330 | 750 | 1400 | 2820 | 50 | 1480 | 1520 | 760 | 1490 | 2830 | 1540 | 0 |
| Letters in {} indicate population collection site: f = farm, w = wild, in = introduced | | | | | | | | | | | | |

### 3.5.4 Mantel Tests

Mantel tests performed between loci on $D_m$ distance matrices and $(\delta\mu)^2$ distance matrices constructed using all the populations (excluding Nick James and Valley due to absence of PCR products) and populations of wild origin were all non-significant. The Mantel tests of genetic distance matrices and geographic distance also indicated a non-significant association (Table 3.31).

**Table 3.31**: Summary of the Mantel tests

| | $D_m$ | | $(\delta\mu)^2$ | |
|---|---|---|---|---|
| Description of matrices compared for association | | | | |
| Locus combination for all populations excluding NJ and V | Correlation Coefficient (*r*) | Probability (*P*) of significant association | Correlation Coefficient (*r*) | Probability (*P*) of significant association |
| UNH104 / UNH111 | 0.045 | 0.378 | 0.062 | 0.354 |
| UNH104 / UNH123 | 0.092 | 0.323 | 0.234 | 0.084 |
| UNH111 / UNH123 | -0.209 | 0.901 | 0.071 | 0.309 |
| | | | | |
| Locus combination for populations of wild origin | | | | |
| UNH104 / UNH111 | 0.136 | 0.240 | -0.128 | 0.576 |
| UNH104 / UNH123 | -0.319 | 0.854 | 0.105 | 0.279 |
| UNH111 / UNH123 | -0.066 | 0.461 | -0.208 | 0.789 |
| | | | | |
| Combined locus genetic distance / Geographic distance | -0.038 | 0.51 | -0.053 | 0.51 |

### 3.5.5 Number of Migrants between Wild Type Populations

The number of migrants ($N_m$) per generation between the wild populations (Table 3.32) emulated the genetic distances between the wild populations (Tables 3.28 and 3.29). The highest number of migrants per generation was between the Olifants and Ndumu population (4.4).

**Table 3.32**: The number of migrants per generation between the six wild populations

| | B | E | K | N | O | S |
|---|---|---|---|---|---|---|
| B | 0.000 | | | | | |
| E | 1.133 | 0.000 | | | | |
| K | 1.045 | 0.472 | 0.000 | | | |
| N | 0.609 | 1.081 | 1.100 | 0.000 | | |
| O | 0.490 | 1.067 | 0.487 | 4.399 | 0.000 | |
| S | 0.692 | 0.396 | 2.569 | 0.797 | 0.330 | 0.000 |

## 3.6 PHYLOGENETIC RELATIONSHIPS BETWEEN THE POPULATIONS

### 3.6.1 $D_m$ UPGMA and Neighbour Joining trees

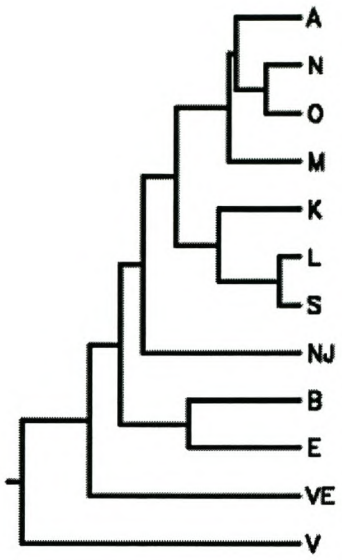


**Figure 3.01**: UPGMA phylogenetic tree constructed using $D_m$ genetic distances for the twelve populations. The tree is scaled to 0.03 genetic distance units per centimeter of branch length

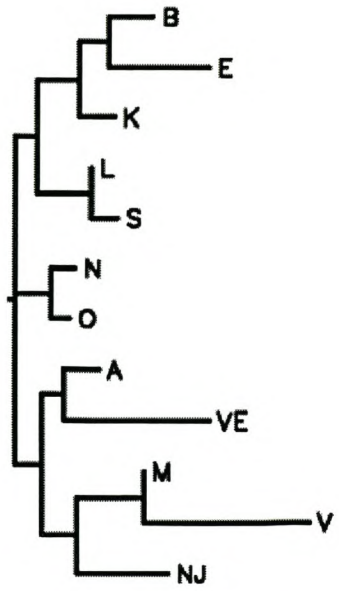


**Figure 3.02**: Neighbour-Joining phylogenetic tree constructed using $D_m$ genetic distances for the twelve populations. The tree is scaled to 0.05 genetic distance units per centimeter of branch length

### 3.6.2 $(\delta\mu)^2$ UPGMA and Neighbour Joining trees

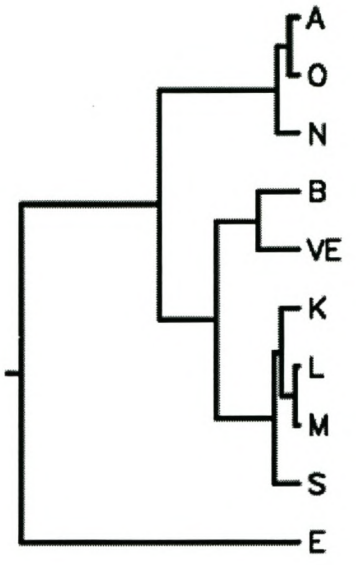


**Figure 3.03**: UPGMA phylogenetic tree constructed using $(\delta\mu)^2$ genetic distances for ten populations. The tree is scaled to 3.3 genetic distance units per centimeter of branch length

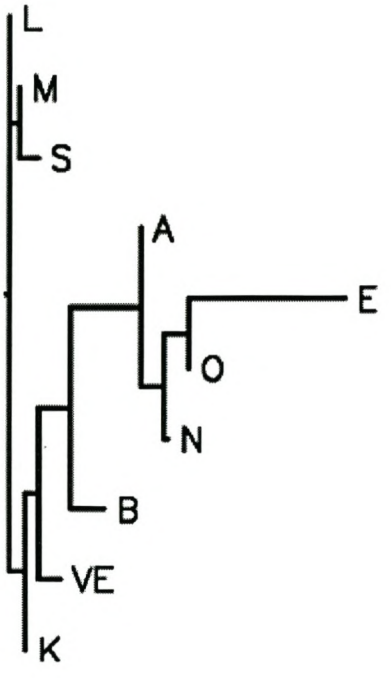


**Figure 3.04**: Neighbour-Joining phylogenetic tree constructed using $(\delta\mu)^2$ genetic distances for ten populations. The tree is scaled to 6.6 genetic distance units per centimeter of branch length

### 3.6.3 $D_m$ MDS plots

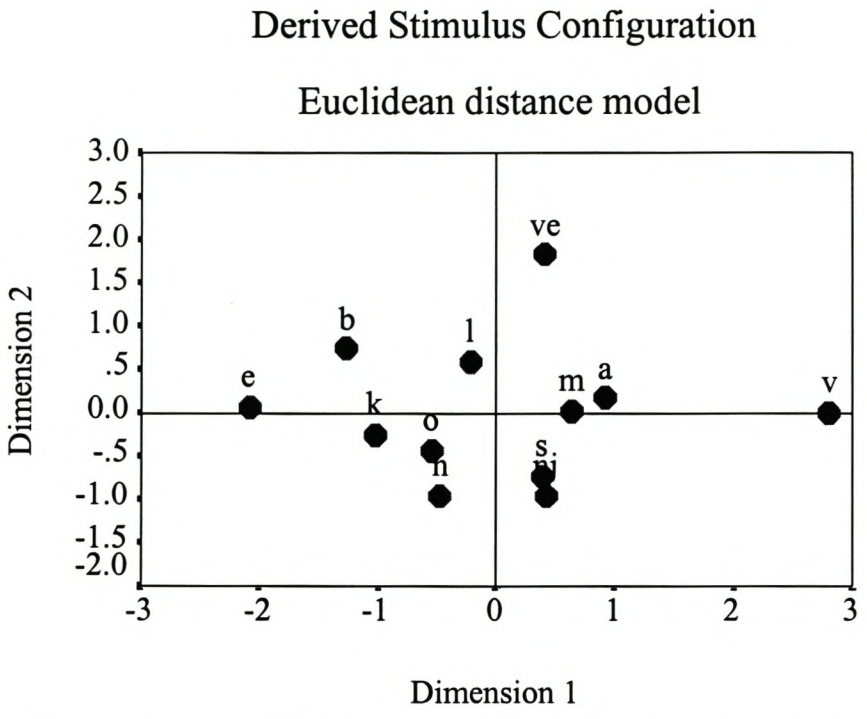


**Figure 3.05**: MDS plot of Euclidean distances between the twelve populations, calculated using the derived co-ordinates of the $D_m$ genetic distances

Key:

| | | |
|---|---|---|
| a = amatikulu (farm) | b = bushmans (wild) | e = elands (wild) |
| k = kasinthula (wild) | l = le pommier (introduced) | m = makathini (farm) |
| n = ndumu (wild) | nj = nick james (farm) | o = olifants (wild) |
| s = sucomba (wild) | v = valley (farm) | ve = verloerenvlei (introduced) |

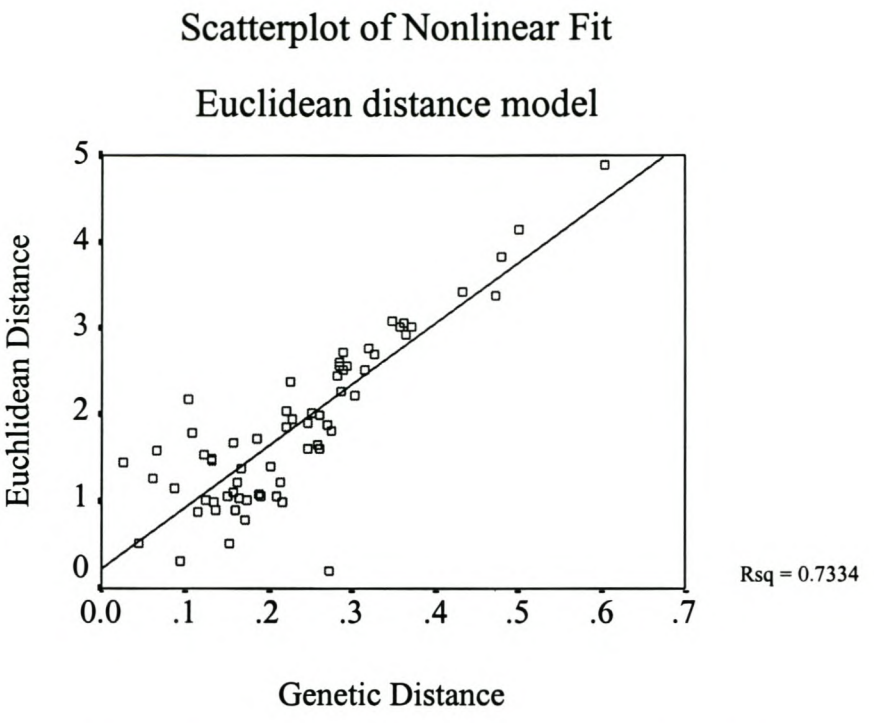


**Figure 3.06**: Scatterplot of Euclidean distances plotted against $D_m$ genetic distances between the twelve populations

### 3.6.4 $(\delta\mu)^2$ MDS plots

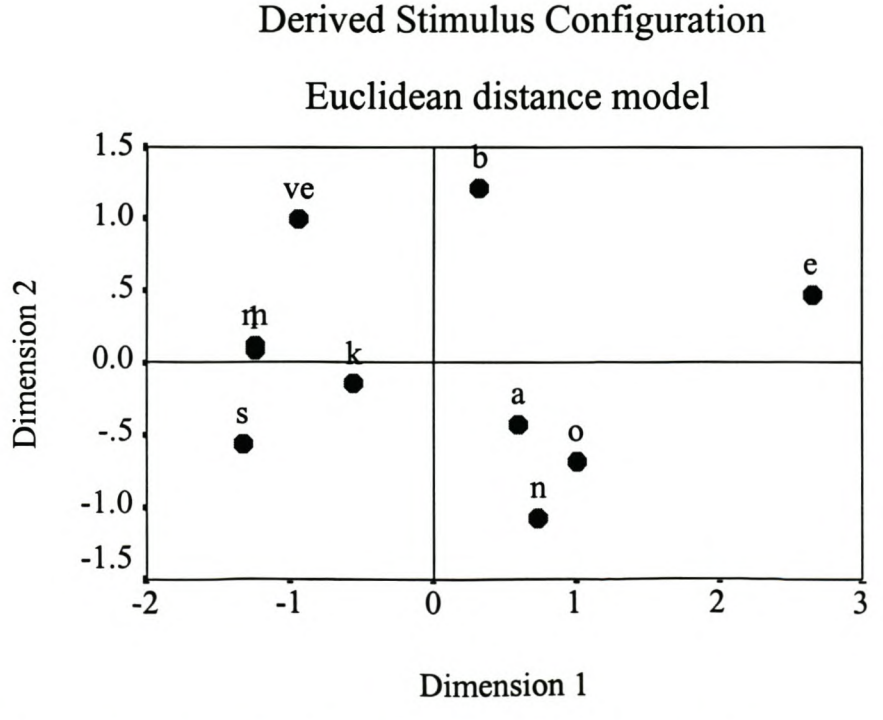


**Figure 3.07**: MDS plot of Euclidean distances between ten populations, calculated using the derived co-ordinates of the $(\delta\mu)^2$ genetic distances

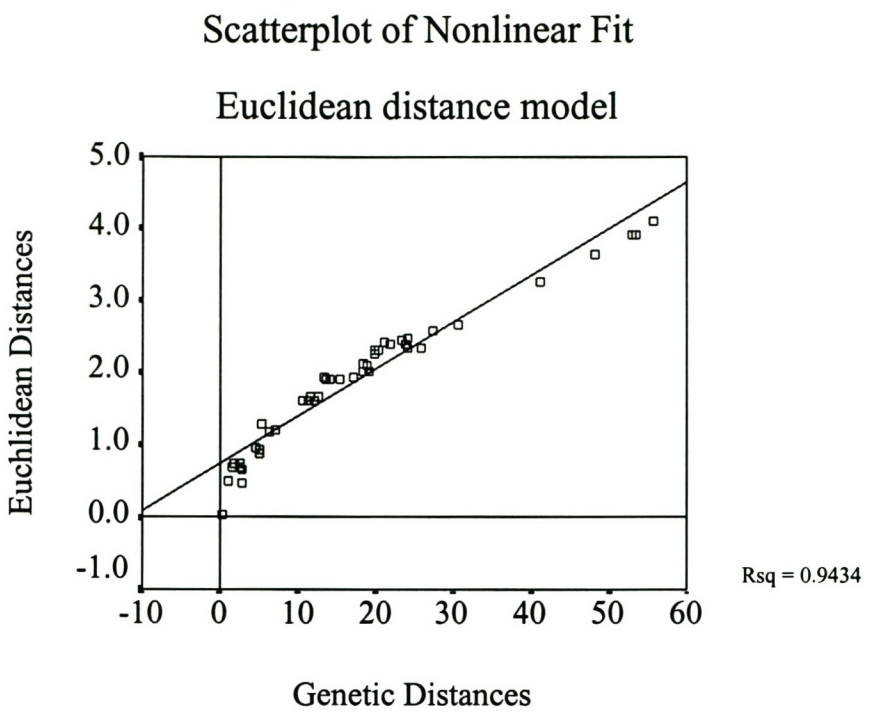


**Figure 3.08**: Scatterplot plot of Euclidean distances against $(\delta\mu)^2$ genetic distances between ten populations

# CHAPTER FOUR

## *DISCUSSION*

Important issues from the laboratory protocols are briefly commented on followed by an in depth discussion of the nature of the genetic variation present in the twelve populations and a general discussion of specific aquaculture and conservation genetic issues.

## 4.1 LABORATORY PROTOCOL – RESULTS

Laboratory procedures were generally satisfactory. The area of most interest was the variability of PCR success for the ten loci. Initial agarose gel electrophoresis clearly indicated this. A more thorough screening of the variables involved in the success of PCR, such as primer annealing temperature and template quality/quantity, may improve the rate of success. This was the first attempt at using *O. niloticus* primers to amplify *O. mossambicus* DNA and perform a population genetic study using FdCTP fragment analysis. The tilapia linkage map primers were developed using *O. niloticus* DNA (Kocher *et al.*, 1998), hence some of them may not behave optimally on *O. mossambicus* DNA. This was supported by the fact that very often completely fresh material was used, and that individuals, which amplified well, always amplified well, in contrast to those that did not amplify well. Thus the phenomenon is clearly a systematic problem to do with specific individuals, rather than a general variability with technique. Future research on *O. mossambicus* could involve development of primers specific to *O. mossambicus* DNA. However the cloning and sequencing protocols required would add to the costs and time of such a project.

### 4.1.1 Success of PCR Amplification

The success of average PCR amplification indicated locus specificity (see Tables 3.06 to 3.08) possibly due to different levels of priming success with non-specific primers. UNH104, in particular, had a reduced PCR success rate. Absence of PCR products at the UNH104 locus for the Nick James and Valley populations (Table 3.06) indicate that the *O. niloticus* primers may be specific to a population, which in turn leads one to question the specific status of the Nick James and Valley fish. Alternatively, the Nick James and Valley populations may both be fixed for a null allele. The origin of the red coloured Nick James and Valley populations is unknown. They may be hybrids with divergent priming site and hence low specificity for UNH104. The fact that this is a systematic error – that this PCR failure was population specific, that the populations do amplify with the other primers, and that the material was prepared in an identical way, and often at the same time as other successful DNA samples, indicates the systematic diversion of the red tilapia as the causative factor, rather than any procedural problems.

## 4.2 GENETIC VARIATION AT THE LOCI

### 4.2.1 Number of Alleles

The number of alleles present at the loci varied from seven at UNH104, to seventeen at UNH123 (Table 3.01). The UNH104 locus may have had more null alleles than the other loci resulting in fewer observed alleles. The greater heterozygote deficit at UNH104 – which may also be indicative of inherently high numbers of null alleles – is supportive of this suggestion. Additional supportive evidence comes from the lack of PCR product for specific red populations (Valley and Nick James). This systematic PCR failure is indicative of excessive primer mismatch for UNH104. This is also likely to have manifest itself as partial PCR failures – where only one allele of a heterozygote is amplified – i.e. a high rate of null alleles inherent in a mismatched primer. The lower sample size for UNH104 may also have contributed to lower estimates of allele numbers due to sampling errors. Although the number of loci sampled is too small to be conclusive, there is a suggestion that greater mean allele lengths result in greater genetic numbers of alleles. There is some evidence of this in other organisms. Human and *Drosophila* microsatellite DNA provide evidence for this (Weber, 1990; Goldstein and Clark, 1995; review in Goldstein and Schlotterer, 1999). Some simulation studies, however, have found no correlation between repeat size and microsatellite variation (Valdes *et al.*, 1993). UNH104 had an average repeat size of 138 base pairs and the lowest number of alleles (seven). UNH111 had an average repeat size of 186 base pairs and sixteen different alleles. UNH123 had an average repeat size of 186 base pairs and seventeen different alleles (see Table 3.02).

Freshwater fish are thought to have lower levels of allelic diversity than anadromous and marine fish due to smaller effective population sizes and a less stable environment (de Woody and Avise, 2000). The average number of alleles (13.3) obtained for *O. mossambicus* (Table 2.01) is higher than the mean per locus for freshwater fishes (7.5) reported by de Woody and Avise (2000), but similar to Agustin's average of 15.4 alleles found at five microsatellite loci for five populations of wild and feral origin (Agustin, 1999). Wild *O. mossambicus* may have a larger effective population size than most freshwater fishes due to the ability to enter seawater and migrate to neighbouring river systems (personal observation). In a sense *O. mossambicus* may be more similar to anadromous fishes, which have an average of 11.3 alleles per locus as reported by de Woody and Avise (2000). Considering the fact that null alleles may be present at the loci (see section 4.3.5) the average may be even higher.

### 4.2.2 Allele Frequencies

A single allele at relatively high frequency was present at each locus accompanied by a number of smaller and larger low frequency alleles (Table 3.02) as expected under the SMM mode of microsatellite evolution (review in Goldstein and Schlotterer, 1999). There were more alleles with a repeat size larger than the most common allele. Tentative evidence exists supporting this as a form of mutation bias towards increased repeat size (more gains than losses of repeat units) at microsatellite loci (Weber and Wong, 1993; review in Goldstein and Schlotterer, 1999).

### 4.2.3 Locus Heterozygosity

Average observed heterozygosity at the loci was low (0.35), compared to an average of (0.46) obtained in a review of genetic variation at microsatellite loci in freshwater fish (de Woody and Avise, 2000) (Table 3.03). Average expected heterozygosity, on the other hand, was higher (0.71) than documented in the literature (de Woody and Avise, 2000). In fact the average expected heterozygosity was comparable to that found at microsatellite loci in marine fish (0.79), which are naturally more diverse largely due to larger and more stable long-term effective population sizes (de Woody and Avise, 2000). The large reduction in observed heterozygosity compared to expected heterozygosity may be indicative of null alleles at the loci leading to an apparent excess of homozygotes, or may result from inbreeding after generations of captivity that did not attempt to rigorously control inbreeding levels.

### 4.2.4 Test for Hardy Weinberg Equilibrium (HWE)

Both the $\chi^2$ test and the Exact test detected significant deviation from HWE at the loci (Tables 3.04 and 3.05). The expected values for the $\chi^2$ test were higher than five and therefore did not violate the assumptions of the $\chi^2$ test. The Exact test is more powerful than the $\chi^2$ test for many low frequency alleles, since no information is lost. In this case the tests produced comparable results. The principal reason for deviation from HWE was a reduction in heterozygosity at all loci. This is typical of microsatellite studies which are known to suffer from relatively high null allele frequencies, however in this case it may also be a result of inbreeding due to lack of prevention of inbreeding in captivity. In this regard it is interesting to note that the long term captive red populations of Nick James, and Valley sites, - where one might expect the highest level of inbreeding – do indeed have the highest $F_{IS}$ values (Table 3.09).

## 4.3 GENETIC VARIATION WITHIN POPULATIONS

The amount of genetic variation within a population can be influenced by selection, migration, mutation and genetic drift. In this study, the effects of selection can be only speculative, since the microsatellite loci are necessarily presumed to be neutral, since this is the most parsimonious explanation of microsatellite genetic variation. Of the other three factors, genetic drift across generations is responsible for losses in genetic variation while immigration and mutation increase genetic variation. The extent of genetic drift is dependent on effective population size changes for past and present generations. If a population experiences a bottleneck with subsequent population number expansion the level of genetic variation present in the resultant population will be determined by the narrowest part of the bottleneck. This phenomenon is demonstrated by the calculation of the effective population size ($N_e$) over many generations as the harmonic mean over those generations (Maynard-Smith, 1989). The harmonic mean will be closest to the $N_e$ of the smallest population over the time period. The level of variation reflects the genetic drift that occurred during the bottleneck. Table 3.09 contains a summary of the general levels of genetic variation present in the twelve *O. mossambicus* populations. The farm populations (e.g. Amatikulu, Nick James and Valley) may have suffered reduced effective population size in the past, evident in their reduced observed heterozygosities compared to the wild and introduced populations.

### 4.3.1 Allele Frequencies

Populations that have a few high frequency alleles have low levels of genetic variation. Many of the *O. mossambicus* populations had a common high frequency allele at the loci however there were many low frequency alleles (see Tables 3.06 to 3.08). Genetic drift acts on the low frequency alleles first, resulting in their loss before the high frequency alleles. The high frequency alleles may become low frequency alleles over time. However, if there is migration occurring between the populations the common alleles will remain common since migrating individuals are more likely to contain common alleles. Immigration therefore has a stabilizing effect on allele frequency.

Supplementation of the populations using progeny from unequal family sizes may have increased the frequency of some alleles at the expense of others. There appears to be a tendency for more of the wild *O. mossambicus* populations to share a common allele than farmed or introduced populations (Tables 3.06 to 3.08). This would suggest that the common allele is present in the source populations, which have probably undergone more migrations and less genetic drift than captive isolated populations. It is also reasonable to suppose that captive populations, which are not rigorously managed to maintain genetic variation, will have high rates of genetic drift compared to wild populations.

### 4.3.2 Population Heterozygosity

Observed heterozygosities were lower than expected heterozygosities for all twelve populations (Table 3.09). Possible reasons for an increased number of homozygotes include the presence of null alleles, inbreeding or population substructure (i.e. the Wahlund effect, - see Hartl and Clark (1997)). The controlled crosses summarized in Tables 3.14 to 3.19 produced some evidence of null alleles however this data is not conclusive (see section 4.3.5). The presence of inbreeding is difficult to determine without the knowledge of effective population sizes for the populations. A suitable means of calculating effective population size was not found due to the nature of *O. mossambicus* breeding leading to the presence of overlapping generations within the populations. Overlapping generations may have also resulted due to the supplementation of the populations using progeny from family crosses. This process in itself may have reduced the effective population sizes of the populations, if unequal family contributions were made, leading to inbreeding (increased homozygosity). If substructure exists within the source populations its detection will have been obscured by time spent in captivity after one generation of random mating.

Populations with excess heterozygotes indicate the occurrence of outbreeding. Analysis of the loci individually detected excess heterozygosity for the Elands and Le Pommier populations at the UNH111 locus and the Le Pommier and Sucomba populations at the UNH123 locus (results not included). The Le Pommier population is an introduced population on a wine farm in the Western Cape. If a number of fish from different sources were introduced into the dam, this may have resulted in the formation of an excess of novel heterozygous genotypes. Possible reasons for outbreeding in the Elands and Sucomba population have not been determined, however a single generation of outbreeding, either in the source population, or by contamination during captivity, is a likely explanation. An alternative is the possibility of linkage with a gene for which over dominance is present – however one might expect this in all populations.

### 4.3.3 Hardy Weinberg Equilibrium

The $\chi^2$ tests detected significant deviation from HWE in all twelve *O. mossambicus* populations however the Exact test detected deviation in only ten (Tables 3.10 and 3.11). The two populations that were in HWE according to Exact test are the Nick James and the Valley populations. Note that the UNH104 locus did not amplify for these two populations. UNH104 possibly had a large influence on overall HWE in the other populations or the Exact test provides more information on the extent of HWE within the populations.

### 4.3.4 Linkage Disequilibrium

Apart from the Bushmans, Le Pommier, Sucomba and Kasinthula populations the Fisher Exact test for linkage disequilibrium gave non-significant results (Table 3.12). This supports the loci being on different linkage groups and the occurrence of independent allele assortment at the loci in most of the populations. The few significant tests found are in line with predictions of Type I errors from carrying out multiple significance tests, and are no longer significant if Bonferroni *P* value adjustments are made. Heterogeneous linkage disequilibrium among populations can be a result of population mixing – however the lack of corroborative evidence from raised heterozygote deficits in these populations argues for a Type I error explanation.

### 4.3.5 Microsatellite Inheritance

For the Ndumu male and Olifants female cross, the observed offspring alleles at the UNH104 locus (Table 3.14) were best explained by the putative parental genotypes of Case Two (see Table 4.01). PCR failure is unlikely since the primers were successful in both parents and eight out of ten offspring.

**Table 4.01**: Putative genotypes for the UNH104 Ndumu and Olifants cross

| Case One<br>Expected genotype ratios<br>(if no null alleles are present) | | |
|---|---|---|
| **Parents** | | **Offspring** |
| **Male** | **Female** | All 128/128 |
| 128<br>128 | 128<br>128 | |
| **Case Two**<br>**Expected genotype ratios**<br>**(if null alleles are present in both parents)** | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ¼ 128/128<br>½ 128/null<br>¼ null/null |
| 128<br>null | 128<br>null | |
| **Case Three**<br>**Expected genotype ratios**<br>**(if a null allele is present in one parent)** | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ½ 128/128<br>½ 128/null |
| 128<br>128 | 128<br>null | |

For the Valley male and Olifants female cross (Table 3.15); the observed offspring alleles at the UNH104 locus were best explained by the putative parental genotypes of Case Three (see Table 4.02). The 136 allele obtained for the Valley male was unexpected since no other members of the Valley population amplified for any alleles at UNH104. This raises doubts as to either the authenticity of the Valley male or the presence of the 136 allele. It is possible that a male from the red Amatikulu population (which does contain the 136 allele) accidentally contaminated the red Valley population. There are no obvious visible differences between a red Amatikulu male and a red Valley male. The 136 allele was present in the Olifants population at a relatively high frequency (0.13) and is therefore probably a true allele of the Olifants female. If however the 136 allele was merely an artifact of both the Valley male and the Olifants female, the genotype ratios obtained would be expected. In either case, genotype ratios do not fit the expectations of any scenario very well.

**Table 4.02**: Putative genotypes for the UNH104 Valley and Olifants cross

| Case One<br>Expected genotype ratios<br>(if no null alleles are present) | | |
|---|---|---|
| **Parents** | | **Offspring** |
| **Male** | **Female** | ½ 136/136<br>½ 136/128 |
| 136<br>136 | 128<br>136 | |
| Case Two<br>Expected genotype ratios<br>(if null alleles are present in both parents)<br>(Not possible) | | |
| Case Three<br>Expected genotype ratios<br>(if a null allele is present in one parent) | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ¼ 136/128<br>¼ 128/null<br>¼ 136/136<br>¼ 136/null |
| 136<br>null | 128<br>136 | |

For the Ndumu male and Olifants female cross (Table 3.16); the observed offspring alleles at the UNH111 locus can best be explained by the putative parental genotypes of Case Four (see Table 4.03).

**Table 4.03**: Putative genotypes for the UNH111 Ndumu and Olifants cross

| Case One<br>Expected genotype ratios<br>(if no null alleles are present) | | |
|---|---|---|
| **Parents** | | **Offspring** |
| **Male** | **Female** | All 182/150 |
| 182<br>182 | 150<br>150 | |
| **Case Two**<br>**Expected genotype ratios**<br>**(if null alleles are present in both parents)** | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ¼ 182/150<br>¼ 182/null<br>¼ 150/null<br>¼ null/null |
| 182<br>null | 150<br>null | |
| Case Three<br>Expected genotype ratios<br>(if a null allele is present in the female parent) | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ½ 182/150<br>½ 182/null |
| 182<br>182 | 150<br>null | |
| Case Four<br>Expected genotype ratios<br>(if a null allele is present in the male parent) | | |
| Parents | | Offspring |
| Male | Female | ½ 182/150<br>½ 150/null |
| 182<br>null | 150<br>150 | |

For the Valley male and Olifants female cross (Table 3.17); the observed offspring alleles at the UNH111 locus can best be explained by the putative parental genotypes of Case One (see Table 4.04).

**Table 4.04**: Putative genotypes for the UNH111 Valley and Olifants cross

| Case One<br>Expected genotype ratios<br>(if no null alleles are present) | | |
|---|---|---|
| **Parents** | | **Offspring** |
| **Male** | **Female** | ½ 180/180<br>½ 180/150 |
| 180<br>180 | 180<br>150 | |
| Case Two<br>Expected genotype ratios<br>(if null alleles are present in both parents)<br>(Not possible) | | |
| Case Three<br>Expected genotype ratios<br>(if a null allele is present in the male parent) | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ¼ 180/180<br>¼ 180/null<br>¼ 150/null<br>¼ 180/150 |
| 180<br>null | 180<br>150 | |

For the Ndumu male and Olifants female cross (Table 3.18); the observed offspring alleles at the UNH123 locus can best be explained by the putative parental genotypes of Case Two (see Table 4.05).

**Table 4.05**: Putative genotypes for the UNH123 Ndumu and Olifants cross

| Case One<br>Expected genotype ratios<br>(if no null alleles are present) | | |
|---|---|---|
| **Parents** | | **Offspring** |
| **Male** | **Female** | All 170/170 |
| 170<br>170 | 170<br>170 | |
| **Case Two**<br>**Expected genotype ratios**<br>**(if null alleles are present in both parents)** | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ¼ 170/170<br>½ 170/null<br>¼ null/null |
| 170<br>null | 170<br>null | |
| **Case Three**<br>**Expected genotype ratios**<br>**(if a null allele is present in the female parent)** | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ½ 170/170<br>½ 170/null |
| 170<br>170 | 170<br>null | |

For the Valley male and Olifants female cross (Table 3.19); the observed offspring alleles at the UNH123 locus can best be explained by the putative parental genotypes of Case One (see Table 4.06), however it is curious that the expected 170/170 homozygote is not present in offspring. This appears to be a non-random distribution – i.e. it does not conform to Mendelian expectations. Non-specific PCR products were present in the offspring individuals nine and ten. The fact that these alleles occur in adjacent individuals, suggest that they could be artifacts caused during screening on the polyacrylamide gel as opposed to PCR artifacts.

**Table 4.06**: Putative genotypes for the UNH123 Valley and Olifants cross

| Case One<br>Expected genotype ratios<br>(if no null alleles are present) | | |
|---|---|---|
| **Parents** | | **Offspring** |
| **Male** | **Female** | ½ 170/170<br>½ 170/204 |
| 170<br>204 | 170<br>170 | |
| Case Two<br>Expected genotype ratios<br>(if null alleles are present in both parents)<br>(Not possible) | | |
| Case Three<br>Expected genotype ratios<br>(if a null allele is present in the female parent) | | |
| **Parents** | | **Offspring** |
| **Male** | **Female** | ¼ 170/170<br>¼ 170/null<br>¼ 204/170<br>¼ 204/null |
| 170<br>204 | 170<br>null | |

Most of the above inheritance patterns indicate the possible presence of null alleles at the loci however the likelihood of other factors like low PCR success cannot be ruled out. A future investigation using *O. mossambicus* pedigrees could aid the proper identification of null alleles. The breeding of two heterozygotes would throw better light on the possibility of non-Mendelian inheritance.

## 4.4 THE EFFECT OF CAPTIVITY ON GENETIC DIVERSITY

The extent of change in genetic variation for aquaculture populations over generations holds inherent information on the effectiveness of current and past management designs (Tave, 1993). Consecutive generations can be analysed for inbreeding, genetic drift or signs of pedigreed mating. The random mating of *O. mossambicus*, which has the potential to produce a new generation every three to four weeks in a breeding season (Skelton, 1993), and the supplementation of populations using unequal family contributions from mating groups, presented a challenge in detecting genetic changes.

### 4.4.1 Change in Allele Frequency

The frequency of the alleles in the Bushmans population fluctuated significantly between 1999 and 2000 at both the UNH104 and UNH123 loci (Table 3.20). Both random genetic drift and supplementation, resulting in overlapping generations within the populations, possibly caused the shift in allele frequencies present between years. There were seven alleles in the 1999 sample that were absent in the 2000 sample having potentially been reduced in number through genetic drift (Table 3.20). A new strategy to maintain population numbers and prevent loss of genetic variation within the aquaculture populations may be worth considering.

### 4.4.2 Change in Heterozygosity

A common concern for aquaculture populations over generations is the reduction in heterozygosity through inbreeding. Inbreeding would be evident as a reduction in observed heterozygosity as opposed to a shift caused by overlapping generations, which would not necessarily lead to a reduction in heterozygosity unless a large number of homozygous individuals were used to supplement a population. This was implied from the non-significant change in $H_e$ between the 1999 population sample and the 2000 Bushmans population sample (Table 3.21). In fact, there was a significant increase in $H_o$ between the 1999 sample and the 2000 Bushmans population sample. It may be that there was inbreeding already present which decreased over the course of one year.

## 4.5 POPULATION GENETIC STRUCTURE

The molecular genetic data can verify whether genetic structure exists between populations and provides a reason for maintaining them separately (Tave, 1993). Overall, a significant genetic structure was present between all twelve populations of *O. mossambicus* evident in the various structure indices that were calculated (see section 3.5). In aquaculture, genetic structure information serves as a guide to which populations should be kept separately for evaluation. Space permitting, however, all populations should be kept separately until their biological potential is assessed since some populations can have phenotypic advantages (Tave, 1993).

The conservation of fish populations has often been neglected based on an 'out of sight out of mind' premise. Consequently, data concerning vulnerability to extinction of aquatic species and populations is sparse compared to that of terrestrial species (Ryman *et al.*, 1995). To obtain a true reflection of a species or population's conservation status requires data on genetic structure, genetic variation and associated demographic history (Avise, 1994; Zhang and Hewitt, 1998; Goldstein and Schlotterer, 1999). From this information, identification and nominations of populations with evolutionary significance can be made (Avise, 1994; Bowen, 1998). Populations are referred to as management units (MUs) if they contain sufficient genetic differentiation from other populations and evolutionary significant units (ESUs) if they are believed to be important to the present and future survival of a species (Bowen, 1998; Parker *et al.*, 1999). The significant structure found among the twelve *O. mossambicus* populations presents evidence indicating that all of the populations represent MUs and that some of the populations (e.g. Elands) constitute possible ESUs. The Elands population exists in the upper reaches of the Limpopo river system where it may be protected from the invading *O. niloticus*. Without the concomitant demographic history of the Southern African *O. mossambicus* populations, the assignment of ESU status can only be tentative. The current level of evaluation for the twelve *O. mossambicus* populations is therefore not sufficient to accurately classify ESUs however the fact that structure exists encourages the protection of habitats and river systems in which the wild source populations of *O. mossambicus* occur.

### 4.5.1 Exact test for Allele Frequency Heterogeneity Between Populations

The Exact tests for allele frequency heterogeneity between the populations (see Tables 3.22 to 3.24) were quick to perform providing basic evidence to the presence of highly significant structuring between the populations and reason for performing more involved (F-statistic) type analyses.

### 4.5.2 Quantifying Population Structure (F-Statistics)

The overall $F_{ST}$ (0.27) and other structure indices were all highly significant (see section 3.5.2). Congruency across the different indices provided assurance of accurate calculation of the extent of genetic structure. According to the AMOVA ($\Phi_{ST}$), 17% of genetic variation was due to differences between populations of *O. mossambicus*.

The pair wise $F_{ST}$ matrix produced 36 significant differences out of the 66 population comparisons, after Bonferroni *P* value correction (Table 3.25). Considering that by chance alone only four significant differences (5%) were expected - without Bonferroni correction – great structuring exists between the populations.

### 4.5.3 Genetic Distance Matrices

Both IAM based ($D_m$) and SMM based $(\delta\mu)^2$ genetic distances were calculated to compare the utility of each genetic distance measure in divulging the extent of differentiation and demographic relationship between the populations (Tables 3.26 to 3.29). A matrix of geographic distance was included as a reference to the straight-line distances between the populations (Table 3.30). No direct correlation was found between genetic distance and geographic distance (Table 3.31), however since movements between populations are extremely unlikely to occur overland in straight-line fashion (even allowing for anthropogenic introductions) – such a correlation is not an expectation.

The level of genetic differentiation between *O. mossambicus* populations based on $D_m$ distances (Table 3.26) was similar to that reported between five populations of Atlantic salmon (McConnell *et al.*, 1995b). de Silva (1997) provided estimates of Nei's 1972 genetic distances between *O. mossambicus* and *O. niloticus* populations in Sri Lanka ranging from 0.1 to 0.25. Estimates of Nei's 1972 genetic distance between the *O. mossambicus* populations ranged from zero to 0.0006. These *O. mossambicus* populations were undoubtedly highly related - being offspring of a single small, introduced population (Agustin, 1999) and therefore not a true reflection of the level of genetic differentiation expected between wild Southern African *O. mossambicus* populations. In addition, $D_m$ is an unbiased version of Nei's 1972 distance – $D_m$ and Nei's 1972 distance may not be comparable. Alternatively, if the genetic distances are comparable then the populations in this study may be representative of subspecies of *O. mossambicus*, although this is unlikely due to the migration habits of the species. The inferiority of allele frequency based genetic distance measures found for some microsatellite analysis techniques (Ruzzante, 1998) warrants the use of genetic distance measures like $(\delta\mu)^2$ which perform more realistically for microsatellite data. The $(\delta\mu)^2$ values obtained were similar to those in human population comparisons (reviewed by Nei and Kumar, 2000) and salmon population comparisons (Nielsen *et al.*, 1997).

### 4.5.4 Mantel Tests

Evidence for an association between geographic locality and genetic distance based on gene flow was not conclusively found after using Mantel tests performed between loci, for all the *O. mossambicus* populations, and between loci for the wild populations only. Mantel tests performed between the combined locus genetic distance matrices and the geographic distance matrix similarly indicated a non-significant association (Table 3.31).

Gene flow between the populations may take place between populations that are distantly located due to the ability of *O. mossambicus* to enter the sea and move in the sea currents. This ability possibly reduces the extent of isolation by distance. Whilst presence of Mantel test inter-locus correlations would have supported a geographic isolation by distance model of population structure, absence of such correlations does not preclude an identical by decent IBD model, since random genetic drift among populations would not be expected to operate similarly for different genes.

### 4.5.5 Number of Migrants between Wild Populations

The high number of migrants (4.4) (Table 3.32) between the Ndumu and Olifants populations has indications of a geographic nature. The Ndumu source population occurs in the Usutu river, a tributary of the Maputo river, which drains into the Maputo Bay in Mozambique. The Olifants source population occurs in the Olifants river, a tributary of the Limpopo river, which drains into the sea north of Maputo bay near Inhambaan (Mozambique). The predominant sea currents on the coast of Mozambique move southward. They possibly provide the Olifants population with a direct migration route from the Limpopo estuary to the Maputo estuary and ultimately the Ndumu population in the Usutu river. The geographically close Sucomba and Kasinthula populations from Malawi had the next highest number of migrants (2.6). This was expected as the Kasinthula population at the Kasinthula research station was initiated from the Shire river, in which the Sucomba population naturally occurs.

## 4.6 PHYLOGENETIC RELATIONSHIPS BETWEEN THE POPULATIONS

The use of only three loci to obtain genetic distance measures prevented the testing of tree topology using bootstrapping or jack knife methods, since sampling with replacement strategies would have resulted in over-representation of most site combinations.

### 4.6.1 $D_m$ UPGMA and Neighbour-Joining Trees

The UPGMA phylogenetic tree constructed using the $D_m$ genetic distance matrix for all twelve populations displayed clusters that resembled the geographic location of the populations' sources. The Ndumu, Olifants, Amatikulu and Makathini populations, which are all geographically close, (Northern Kwazulu Natal) formed a cluster. The Kasinthula, Sucomba and Le Pommier populations formed a second cluster. The clustering of Le Pommier, the second most southern situated population, with the two most northern situated populations suggests that the Le Pommier population possibly had founder individuals from more northern latitudes. A possible connection lies in the 1936 distribution of *O. mossambicus* from the Transvaal (Northern latitudes) to the Jonkershoek Hatchery in the Western Cape. The Bushmans and Elands populations formed a third cluster. The Limpopo and Bushmans river may have shared a common tributary for previous drainage topologies (Figure 3.01).

The Neighbour-Joining phylogenetic tree of $D_m$ genetic distances indicated similar clusters linked to geographic location of the populations sources. Three broad clusters formed. The Bushmans, Elands, Kasinthula, Sucomba and Le Pommier populations formed a cluster. The Olifants and Ndumu populations formed a second cluster. A third cluster encompassed the Amatikulu, Verloerenvlei, Nick James and Valley populations (Figure 3.02). Interestingly, the three red coloured populations clustered relatively closely suggesting a common origin.

### 4.6.2 $(\delta\mu)^2$ UPGMA and Neighbour Joining Trees

For the UPGMA tree, the Ndumu, Olifants and Amatikulu populations formed a cluster as expected from the $D_m$ UPGMA tree; however, Makathini was absent from this cluster. A second cluster linked the Kasinthula, Sucomba, Le Pommier and Makathini populations. The Verloerenvlei population resided with Bushmans in a third cluster (Figure 3.03).

The Neighbour-Joining tree of $(\delta\mu)^2$ genetic distances produced two main clusters. The first cluster encompassed the Ndumu, Olifants, Elands and Amatikulu populations. The Olifants and Elands populations have source populations connected to the Limpopo drainage systems. The Ndumu population may be linked to the Limpopo drainage system via the sea as described in section 4.5.5. The second cluster consisted of Le Pommier, Sucomba and Makathini populations (Figure 3.04).

### 4.6.3 $D_m$ MDS Plots

The graphic representation of the $D_m$ genetic distances between the populations, in two dimensions, did not conform well with the geographic locality of the source populations despite the very high correlation between Euclidean and Genetic distances in two dimensions. No obvious clusters were present in the plot (Figure 3.05). A non-linear relationship present between the $D_m$ genetic distances was stipulated evident in the scatter plot of Euclidean distances against equivalent $D_m$ distances ($R^2$ = 0.733) (Figure 3.06). A limitation of MDS is that a few populations can be misplaced in two dimensions whilst the weight of the majority of populations is sufficient to maintain a high Euclidean distance/genetic distance correlation. An alternative explanation for the lack of utility of this scatter plot, is the possible limitation of $D_m$ as a genetic distance measure in microsatellites.

### 4.6.4 $(\delta\mu)^2$ MDS Plots

The scatter plot of Euclidean distances for the $(\delta\mu)^2$ genetic distances (Figure 3.07) formed clusters similar to the clusters in the UPGMA tree of $(\delta\mu)^2$ (see Figure 3.03). $(\delta\mu)^2$ was found to be more affective than allele frequency based methods at depicting the true demographic structure between populations for microsatellite loci (Ruzzante, 1998). This may be the case between $(\delta\mu)^2$ and $D_m$ in this study. A non-linear relationship present between the $(\delta\mu)^2$ genetic distances was stipulated evident in the scatter plot of Euclidean distances against equivalent $(\delta\mu)^2$ distances ($R^2$ = 0.943) (Figure 3.08). The high correlation between Euclidean and genetic distances suggests that the two-dimensional plot is a useful interpretive tool, particularly since it confirms the UPGMA interpretation.

### 4.6.5 The Utility of Phylogenetics and MDS

A useful application of phylogenetics and MDS is in divulging the possible origins of the introduced or farmed population of *O. mossambicus* based on wild populations. For example, the origins of the red coloured *O. mossambicus* populations in South Africa are unclear. A possible source includes the original introduction of red (*O. mossambicus* x *O. niloticus*) hybrids from Taiwan in 1981 to the Umgeni Hatchery in Kwazulu Natal (reviewed by de Moor and Bruton, 1988). One can make inferences of population origin from the phylogenetic trees and MDS plots obtained for the *O. mossambicus* populations in this study (Figures 3.01 to 3.08). The Amatikulu population generally clusters with the Ndumu and Olifants populations, all three of which occur in the Northern Kwazulu Natal region. The Amatikulu red coloured fish may have originated from a mutation that occurred in a wild grey population in the Northern Kwazulu Natal region or through breeding with the red hybrids from Taiwan. The other two red populations, Nick James and Valley

did not amplify for the UNH104 locus limiting their phylogenetic placement. These strains may not have originated naturally in South Africa. They may have possible roots with the introduced Taiwanese hybrids. It may be useful in this respect to record their lack of PCR product as a monomorphic product different from other alleles. This would undoubtedly have placed them as an outgroup. The systematic PCR failure for the Nick James and Valley groups almost certainly represented phylogenetic divergence rather than methodological problems.

## 4.7 GENERAL DISCUSSION

Genetic management has become an essential part of both exploitation and conservation of aquatic resources. Effective genetic management can minimise the accumulation of deleterious mutations in wild and captive fish; protect wild populations that may harbour potential evolutionary essential genes; and preserve the unique combinations of genetic traits in aquaculture strains. The benefits procured may in the long term justify the initial capital outlay, particularly from the point of view of sustainable utilisation of genetic resources - however the molecular genetic management and evaluation of fish populations is often a luxury only afforded by research organizations or large commercial aquaculture companies (Tave, 1993).

The acquisition of genetically diverse fish populations is essentially the most important stage in developing strains for aquaculture. Reduced genetic variation transpires into fewer traits for exploitation (Tave, 1993). For this reason the acquisition of the twelve *O. mossambicus* populations took place across a broad spectrum of wild and artificial environments (Figure 2.01 and Table 2.01). The maintenance of diversity within the populations over generations constitutes the next most important consideration however this is not always possible when dealing with large numbers of populations. A degree of inbreeding per generation is acceptable. It is the rapid accumulation of inbreeding that is undesirable. Unfortunately, the average farmer may not have the financial power or the knowledge required to implement a genetic program as part of a management scheme. In this situation over zealous production objectives can lead to loss of genetic potential and ultimately loss of biological performance in aquaculture stocks (Tave, 1993). To prevent the rapid loss of diversity, the effective sizes ($N_e$) of the *O. mossambicus* populations must be controlled and maximised to allow for the lowest rate of accumulated inbreeding and minimise gene losses through genetic drift.

Reduced genetic variation is often an outcome of artificial manipulation geared to produce true breeding monotypes suited to aquaculture conditions. Alternatively, inbreeding and genetic drift in captivity may cause reduced variation (Tave, 1993). Genetic variation should be used as a guide to the biological potential of a population and not merely taken as a definitive measure of fitness (Avise, 1994). A population with lowered genetic variation can still perform well particularly if it is a population from a farm that has been conditioned to an aquaculture environment (Tave, 1993), however the intrinsic improvement potential is limited, since this depends on selection of the most useful genes from an array of variation. Alternatively, some genetically heterogeneous and homogeneous wild tilapia populations can outperform farm populations as demonstrated by performance trials with eight populations of *O. niloticus* (Eknath *et al.*, 1993). Quantitative growth and weight trials performed on eight of the *O. mossambicus* populations, which included wild and farm populations, yielded similar results. The Bushmans, Kasinthula and Ndumu populations (all

from wild sources) were identified as having the most potential out of the eight analysed populations. The Bushmans and Kasinthula populations had relatively high expected heterozygosities (0.73 and 0.66 respectively) however the Ndumu population had a low expected heterozygosity (0.41) (Hoffman *et al.*, 2001). The Ndumu population was collected from a wild source therefore the low diversity of the Ndumu population may be attributed to demographical and historical factors present in the wild. Therefore, a population's genetic variation can provide an indefinite measure of performance potential.

Ignorance towards genetic mechanisms operating in wild and captive populations can result in some annoying consequences. If the collection of populations for aquaculture strain development requires concurrent conservation of the source populations as genetic resources, then it is important to take blood or tissue samples for genetic analysis at the onset.

The delay in taking samples from the *O. mossambicus* populations may have allowed for changes in the genetic relationships compared to those present in the wild source populations. Even if populations are sampled early, an aquaculturist is usually unwilling to delay production based on pending completion of a genetic characterization for purely conservation purposes. Careful decisions on production, breeding and maintenance of populations have to be made initially in the absence of molecular genetic indicators. This is a challenge faced by modern aquaculturists and conservationists.

Early conservation genetics correlated the presence of reduced genetic variation with high levels of inbreeding and reduced fitness. Conservation strategies were often poorly planned due to this misconception (Avise, 1994). Some species and populations were given unwarranted protection under the Endangered Species Act (1973) while others that should have qualified did not receive protection (reviewed in Avise, 1994). Recent debate over MU and ESU concepts (reviewed by Bowen, 1998) has lead to the development of the concept of a geminate evolutionary unit (GEU). A GEU constitutes a population that has recently become isolated from other populations and may develop unique genetic combinations that may have important evolutionary relevance in future environments (Bowen, 1998). Some of the *O. mossambicus* wild populations may qualify as GEUs.

Conservation genetics makes use of genetic techniques to study the many forms of biodiversity (Pimm, 1995). One measure of biodiversity, often used in taxonomic studies, is the number of higher order taxa present in a community. Intra-specific genetic variation, however, is considered to be the ultimate source of all biodiversity and the unit of conservation in endangered species studies (Bowen, 1998). This study covered many elements of conservation genetic issues including the threat of loss of trademark genotypes through hybridisation, loss of locally adapted genomes through translocation, unintentional changes in genotype during captivity and finally the concept of balanced exploitation of a species through sound management of captive populations and

conservation of their source populations. The parties involved in exploitation need to be made aware of the long-term benefits of conservation in terms of sustainability. This study has explored and highlighted many of these issues and represents the groundwork for further research into a species that is under threat from a variety of sources, and has demonstrated great aquaculture potential. This study has provided the tools to monitor the rate of genetic change of genotype in captive populations to prevent the rapid loss of diversity, which is a common goal of conservation and sustainable exploitation.

# CHAPTER FIVE

## *CONCLUSION*

The analysis of genetic structure provided evidence for high levels of variation within and between the twelve *O. mossambicus* populations. The populations therefore represent genetic resources from a conservation perspective and a sustainable aquaculture standpoint. Genetic diversity within the populations varied according to origin. Farm populations had reduced levels of variation compared to populations from wild sources. Introduced populations had higher levels of variation than farm populations. The reduced heterozygosity observed for most of the populations may merely be a consequence of a high prevalence of null alleles although it is recommended that the populations be monitored occasionally to detect possible reduction in heterozygosity through inbreeding. The loss of some low frequency alleles between years in the Bushmans population warns of similar genetic drift occurring in the other populations. A more applied approach to preventing effective population size reductions caused through infrastructure failure or through unequal family contributions when supplementing is advised.

Genetic structure indices and phylogenetic methods divulged interesting facts on the genetic relatedness between the populations and the possible origins of some of the farm and introduced populations. All the populations were sufficiently distinct to warrant designation of management unit conservation status although insight into the evolutionary significance of the populations will require further research on the demographic history of *O. mossambicus* populations.

Comparison of heterozygosity with quantitative weight and length data for eight of the twelve populations displayed no apparent correlation, and the possibility of phenotypic influence was evident in some of the populations' performances (e.g. Ndumu).

The use of microsatellites will undoubtedly be applied in future research of wild and captive *O. mossambicus* populations. Anticipated areas of microsatellite application include; parentage analysis with loci that have a sufficiently lowered prevalence of null alleles; the continued monitoring of genetic variation over time by analysing random samples from the populations at subsequent time intervals and genetic tagging growth trials for identification of Quantitative Trait Loci.

The development of aquaculture strains from the twelve *O. mossambicus* populations for aquaculture in South Africa is actively underway. The data presented in this study provides a basis for future molecular analysis of conservation and exploitation issues pertaining to wild and captive populations of *Oreochromis mossambicus*.

## REFERENCES

Adkison, M. D, 1995. Population Differentiation in Pacific Salmon: Local Adaptation, Genetic Drift, or the Environment? *Canadian Journal of Fisheries and Aquaculture Science* 52: 2762 - 2777

Agnese, J. F., Teugels, G. G., Galbusera, P., Guyomard, R. and Volckaert, F., 1997. Morphometric and Genetic Characterization of Sympatric Populations of *Clarias gariepinus* and *C. anguillaris* from Senegal. *Journal of Fish Biology* 50: 1143 – 1157

Agresti, J. J., Seki, S., Cnaani, A., Poompuang, S., Hallerman, E. M., Umiel, N., Hulata, G., Gall, G. A. E. and May, B., 2000. Breeding New Strains of Tilapia: Development of an Artificial Centre of Origin and Linkage Map Based on AFLP and Microsatellite Loci. *Aquaculture* 185: 43 - 56

Agustin, L. Q., 1999. Effects of population bottlenecks on the levels of genetic diversity and patterns of differentiation in feral populations of *Oreochromis mossambicus*. PhD Thesis, Queensland University of Technology, Australia.

Allendorf, F. W., 1977. Electromorphs or Alleles. *Genetics* 87: 821 - 822

Allendorf, F. W., Ryman, N. and Utter, F. M., 1987. Genetics and Fishery Management: Past Present and Future. In: Population Genetics and Fishery Management. University of Washington Press, Washington, SA, pp 1-19

Angers, B. and Bernatchez, L., 1997. Complex Evolution of a Salmonid Microsatellite Locus and Its Consequences in Inferring Allelic Divergence from Size Information. *Molecular Biology and Evolution* 14: 230 - 238

Avault, J. M., 1994. What is Happening to Our Ocean Fisheries. What is the Implication for Aquaculture. *Aquaculture Magazine* 20: 80 - 84

Avise, J. C., 1994. Molecular Markers, Natural History and Evolution. Chapman and Hall Publishers. New York

Barel, C. D. N., Dorit, R., Greenwood, P. H., Fryer, G., Hughes, N., Jackson, P. B. N., Kawanabe, H., Lowe-McConnell, R. H., Nagoshi, M., Ribbink, A. J., Trewavas, E., Witte, F. and Yamaoka, K., 1985. Destruction of Fisheries in Africa's Lakes. *Nature* 315: 19 - 20

Beardmore, J. A., Mair, G. C. and Lewis, R. I., 1997. Biodiversity in Aquatic Systems in Relation to Aquaculture. *Aquaculture Research* 28: 829 - 839

Beaumont, M. A., Ibrahim, K. M., Boursot, P. B. and Bruford, M. W., 1998. Measuring Genetic Distance. Molecular Tools for Screening Biodiversity. Edited by Karp, A., Isaac, P. G. and Ingram, D. S. Chapman and Hall, London. pp 316 - 325

Beaumont, M. A., 2001. Conservation Genetics: Handbook of Statistical Genetics. Edited by Balding, D. J., Bishop, M. and Cannings, C. John Wiley and Sons, Ltd

Bernatchez, L., Guyomard, R. and Bonhomme, F., 1992. DNA Sequence Variation of the Mitochondrial Control Region Among Geographically and Morphologically Remote European Brown Trout *Salmo trutta* Populations. *Molecular Ecology* 1: 161- 173

Boulding, E. G., Boom, J. D. G. and Beckenbach, T., 1993. Genetic Variation in One Bottlenecked and Two Wild Populations of the Japanese Scallop (*Patinopecten yessoensis*): Empirical Parameter Estimates from Coding Regions of Mitochondrial DNA. *Canadian Journal of Fisheries and Aquaculture Science* 50: 1147 - 1157

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R. and Cavalli-Sforza, L. L., 1994. High Resolution of Human Evolutionary Trees with Polymorphic Microsatellites. *Nature* 368: 455 - 457

Bowen, B. W., 1998. What is Wrong with ESUs?: The Gap between Evolutionary Theory and Conservation Principles. *Journal of Shellfish Research* 17: 1355 - 1358

Browstein, M. J., Carpten, J. D. and Smith, J. R., 1996. Modulation of Non – Templated Nucleotide Addition by *Taq* DNA Polymerase: Primer Modifications that Facilitate Genotyping. *Biotechniques* 20: 1004 - 1010

Bruton, M. N., 1983. Breeding, Juvenile Biology and Growth of *Oreochromis mossambicus* in its natural range in South East Africa. International Symposium on Tilapia in Aquaculture, Tel Aviv.

Buroker, N. E., Brown, J. R., Gilbert, T. A., O'Hara, P. J., Beckenbach, A. T., Thomas, W. K and Smith, M. J., 1990. Length Heteroplasmy of Sturgeon Mitochondrial DNA: An Illegitimate Elongation Model. *Genetics* 124: 157 - 163

Cain, A. J., 1988. The Scoring of Polymorphic Colour and Pattern Variation and its Genetic Basis in Molluscan Shells. *Malacol.* 28: 1-15

Callen, D. F., Thompson, A. D., Shen, Y., Phillips, H. A., Richards, R. I., Mulley, J. C. and Sutherland, G. R., 1993. Incidence and Origin of Null Alleles in the (AC)n Microsatellite Markers. *American Journal of Human Genetics* 52: 922 - 927

Crass, R. S., 1969. The Effects of Land Use on the Freshwater Fish in South Africa, with Particular Reference to Natal. *Hydrobiologia* 34: 38 - 56

Dempster, E., 1955. Maintenance of Genetic Heterogeneity. Cold Spring Harbor Symposium. *Quantitative Biology* 20: 25 -32

de Moor, J. I. and Bruton, M. N., 1988. Atlas of Alien and Translocated Indigenous Aquatic Animals in Southern Africa. *South African National Scientific Programmes Report* 144. South Africa

de Silva, C. D., 1997. Genetic Variation in Tilapia Populations in Man-made Reservoirs in Sri Lanka. *Aquaculture International* 5: 339 - 349

de Woody, J. A. and Avise, J. C., 2000. Microsatellite Variation in Marine, Freshwater and Anadromous Fishes Compared with other Animals. *Journal* of *Fish Biology* 56: 461 - 473

Di Rienzo, A., Peterson, A. C., Garza, J. C., Valdes, A. M. and Slatkin, M. *et al*, 1994. Mutational Processes of Simple-Sequence Repeat Loci in Human Populations. *Proceedings of the National Academy of Science of USA* 91: 3166 - 3170

Edwards, A., Civitello, A., Hammond, H. A. and Caskey, C. T., 1991. DNA Typing and Genetic Mapping with Trimeric and Tetrameric Tandem Repeats. *American Journal of Human Genetics* 49: 746 - 756

Eknath, A. E., Tayamen, M. M., Palada-de Vera, M. S., Danting, J. C., Reyes, R. A., Dionsio, E. E, Capili, J. B., Bolivar, H. L., Circa, A. V., Bentsen, H. B., Gjerde, B., Gjedrem, T. and Pullin, R. S. V., 1993. Genetic Improvement of Farmed Tilapias: The Growth Performance of Eight Strains of *Oreochromis niloticus* Tested in Different Farm Environments. *Aquaculture* 111: 171 - 188

Erlich, H. A. and Arnheim, N., 1992. Genetic Analysis using the Polymerase Chain Reaction *Annual Review of Genetics* 26: 479 - 506

Excoffier, L., Smouse, P. E. and Quattro, J. M., 1992. Analysis of Molecular Variance Inferred from Metric Distance Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* 136: 343 - 359

FAO Fisheries Department, Fishery information, Data and Statistics Unit. 2001. Fishstat Plus: Universal Software for Fishery Statistical Time Series. Version 2.3. 2001.

Falk, T. M., Abban, E. K., Oberst, S., Villwock, W., Pullin, R. S. V. and Renwrantz, L., 1996. A Biochemical Laboratory Manual for Species Characterization of Some Tilapiine Fishes. ICLARM Educ. Ser. 17: 93p

Felsenstein, J., 1993. (Phylogeny Inference Package), Version 3.5c. Department of Genetics. University of Washington. Seattle

Ferguson, A., 1980. Biochemical Systematics and Evolution. Blackie. Glasgow

Ferreira, J. T. and Schoonbee, H. J., 1983. Use of Heated Sewage Effluent from a Power Station for the Culture of *Oreochromis mossambicus* and *Hypophthalmichthys molitrix*. *International Symposium on Tilapia in Aquaculture*, Tel Aviv.

Fisher, R. A., 1930. The Genetical Theory of Natural Selection. Oxford: Clarendon Press

Fleming, I. A and Gross, M. R., 1993. Breeding Success of Hatchery and Wild Coho Salmon (*Oncorhynchus-kisutch*) in Competition. *Ecological Applications* 3: 230 - 245

Franck, J. P. C., Wright, J. M. and McAndrew, B. J., 1992. Genetic Variability in a Family of Satellite DNA's from tilapia (Pisces: Cichlidae). *Genome* 35: 719 – 725

Gaigher, I. G., Hamman, K. C. D. and Thorne, S. C., 1980. The Distribution, Conservation Status and Factors Affecting the Survival of Indigenous Freshwater Fishes in the Cape Province. *Koedoe* 23: 57 - 88

Gaigher, I. G., 1983. Cage Culture of Mozambique Tilapia, *Oreochromis mossambicus*, without Artificial Feeding in Wastewater in South Africa. *International Symposium on Tilapia in Aquaculture*, Tel Aviv.

Galbusera, P., Volckaert, F. A., Hellemans, B. and Ollevier, F., 1996. Isolation and Characterization of Microsatellite Markers in the African Catfish *Clarias gariepinus* (Burchell, 1822). *Molecular Ecology* 5: 703 – 705

Garcia De Leon, F. J., Chikhi, L. and Bonhomme, F., 1997. Microsatellite Polymorphism and Population Subdivision in Natural Populations of European Sea Bass *Dicentrarchus labrax* (Linnaeus, 1758). *Molecular Ecology* 6: 51 – 62

Garcia De Leon, F. J., Canonne, M., Quillet, E., Bonhomme, F. and Chatain, B., 1998. The Application of Microsatellite Markers to Breeding Programmes in the Sea Bass, *Dicentrarchus labrax*. *Aquaculture* 159: 303 – 316

Garza, J. C., Slatkin, M. and Freimer, N. B., 1995. Microsatellite Allele Frequencies in Humans and Chimpanzees, with Implications for Constraints on Allele Size. *Molecular Biology and Evolution* 12: 594 - 603

Ginot, F., Bordelais, I., Nguyen, S. and Gyapay, G., 1996. Correction of some Genotyping Errors in Automated Fluorescent Microsatellite Analysis by Enzymatic Removal of One Base Overhangs. *Nucleic Acids Research* 24: 540 - 541

Goldstein, D. B. and Clark, A. G., 1995. Microsatellite Variation in North American Populations of *Drosophila melanogaster*. *Nucleic Acids Research* 23: 3882 - 3886

Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L. and Feldman, M. W., 1995a.Genetic Absolute Dating Based on Microsatellites and the Origin of Modern Humans. *Proceeding of the National Academy of Science USA* 92: 6723 - 6727

Goldstein, D. B., Ruiz Linares, A., Cavalli-Sforza, L. L. and Feldman, M. W., 1995b. An Evaluation of Genetic Distances for use With Microsatellite Loci. *Genetics* 139: 463 -471

Goldstein, D. B. and Pollock, D. D., 1997. Launching Microsatellites: A Review of Mutation Processes and Methods of Phylogenetic Inference. *Journal of Heredity* 88: 335 -342

Goldstein, D. B. and Schlotterer, C. (Editors), 1999. Microsatellites. Evolution and Applications. Oxford University Press, Great Britain

Goodman, S. J., 1997. RST CALC: A collection of computer programs for calculating unbiased estimates of genetic differentiation and determining their significance for microsatellite data. *Molecular Ecology* 6:881-885.

Goudet, J., 2000. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.1). Available from http://www.unil.ch/izea/softwares/fstat.html. Updated from Goudet (1995)

Greenwood, P. H., 1951. Evolution of the African Cichlid Fishes; the *Haplochromis* Species Flock in Lake Victoria. *Nature* 167: 19

Grewe, P. and Hampton, J., 1998. An Assessment of Bigeye (*Thunnus obesus*) Population Structure in the Pacific Ocean, Based on Mitochondrial DNA and DNA Microsatellite Analysis. CSIRO REPORT. Joint Institute of Marine and Atmospheric Research, University of Hawaii

Haldane, J. B. S., 1954. An Exact test for Randomness of Mating. *Journal of Genetics* 52: 631 - 635

Haldane, J. B. S. and Jayakar, S. D., 1963. Polymorphism Due to Selection of Varying Direction. *Journal of Genetics* 58: 237 - 242

Hammond, R. L., Saccheri, I. J., Ciofi, C., Coote, T., Funk, S. M., McMillan, W. O., Bayes, M. K., Taylor, E. and Bruford, M. W., 1998. Isolation of Microsatellite Markers in Animals. Molecular Tools for Screening Biodiversity. Edited by Karp, A., Isaac, P. G. and Ingram, D. S. Chapman and Hall, London. pp 279 - 285

Harper, J. L. and Hawksworth, D. L., 1994. Biodiversity - Measurement and Estimation - Preface. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 345: 5-12

Harris, H., 1966. Enzyme Polymorphism in Man. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 164: 298 - 310

Harris, A. S. and Wright, J. M., 1995. Nucleotide Sequence and Genomic Organization of Cichlid Fish Minisatellites. *Genome* 38: 177 - 185

Harris, A. S., Bieger, S., Doyle, R. W. and Wright, J. M., 1991. DNA Fingerprinting of tilapia, *Oreochromis niloticus*, and its Application to Aquaculture Genetics. *Aquaculture* 92: 157 - 163

Hartl, D. L. and Clark, A. G., 1997. Principles of Population Genetics. Sinauer Associates, Inc. Publishers. Massachusetts.

Hedrick, P. W., 2000. Genetics of Populations. Second Edition. Jones and Bartlett Publishers. London

Heist, E. J. and Gold, J. R., 1999. Microsatellite DNA Variation in Sandbar Sharks (*Carcharhinus plumbeus*) from the Gulf of Mexico and Mid Atlantic Bight. *Copeia* 1: 182 – 186

Henderson, S. T. and Petes, T. D., 1992. Instability of Simple Sequence DNA in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* 12: 2749 - 2757

Hillis, D. M., 1987. Molecular Versus Morphological Approaches to Systematics. *Annual Review of Ecology and Systematics* 18: 23 - 42

Hillis, D. M., Moritz, C., Mable, B. K. Editors., 1996. Molecular Systematics Second Edition. Sinauer Associates, Inc. Publishers. Massachusetts. USA

Hochberg, Y., 1988. A Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika* 75: 800 - 802

Hoffman, L. C., Swart, J.J., Brink, D., 2000. The 1998 Production and Status of Aquaculture in South Africa. *Water SA* 26: 133 - 135

Hoffman, L. C, Brink, D., Mair, G. C. and Beardmore, J. A., 2001. Annual Report. Genetic Improvement and Utilisation of Indigenous Tilapia in Southern Africa. DFID Fish Genetics Programme (R 7284)

Hubby, J. L. and Lewontin, R. C., 1966. A Molecular Approach to the Study of Genetic Heterozygosity in Natural Populations. I. The Number of Alleles at Different Loci in *Drosophila pseudoobscura*. *Genetics* 54: 577 - 594

Jones, M. W., Clay, D. and Danzmann, R. G., 1996. Conservation Genetics of Brook Trout (*Salvelinus fontinalis*): Population Structuring in the Fundy National Park, New Brunswick and Eastern Canada. *Canadian Journal of Fisheries and Aquaculture Science* 53: 2776 - 2791

Jubb, R. A., 1967. Freshwater Fishes of Southern Africa. Gothic Printing Company. Cape Town

Kellog, K. A., Markert, J. A., Stauffer Jr, J. R. and Kocher, T. D., 1995. Microsatellite Variation Demonstrates Multiple Paternity in Lekking Cichlid Fishes from Lake Malawi, Africa. *Proceedings of the Royal Society of London B* 260: 79 – 84

Kimura, M., 1955. Stochastic Processes and Distribution of Gene Frequencies Under Natural Selection. Cold Spring Harbor Symposium. *Quantitative Biology* 20: 33 - 53

Kimura, M., 1968. Evolutionary Rate at the Molecular Level. *Nature* 217: 624 -626

Kimura, M. and Crow, J. F., 1964. The Number of Alleles that can be Maintained in a Finite Population. *Genetics* 49: 725 - 738

Kimura, M. and Weiss, G. H., 1964. The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics* 49: 561 -576

King, J. L., 1967. Continuously Distributed Factors Affecting Fitness. *Genetics* 55: 483 - 492

King, J. L. and Jukes, T. H., 1969. Non-Darwinian Evolution: Random Fixation of Selectively Neutral Mutations. *Science* 164: 788 - 798

Knight, M. E., Turner, G. F., Rico, C., van Oppen, M. J. H. and Hewitt, G. M., 1998. Microsatellite Paternity Analysis on Captive Lake Malawi Cichlids Supports Reproductive Isolation by Direct Mate Choice. *Molecular Ecology* 7: 1605 - 1610

Koehn, R. K. and Hilbish, T. J., 1987. The Adaptive Importance of Genetic Variation. *American Scientist* 75: 134 - 141

Kocher, T. D. and White, T. J., 1989. Evolutionary Analysis Via PCR. In H. A. Erlich (ed.), PCR Technology: Principles and Applications for DNA Amplification. Stockton Press, New York

Kocher, T. D. and Stepien, C. A. eds., 1997. Molecular Systematics of Fishes. Academic Press

Kocher, D. K., Lee, W., Sobolewska, H., Penman, D., McAndrew, B., 1998. A Genetic Linkage Map of a Cichlid Fish, the Tilapia (*Oreochromis niloticus*). *Genetics* 148: 1225 - 1232

Kruegar, C. C., Gharret, T. R., Dehring, T. R. and Allendorf, F. W., 1981. Genetic Aspects of Fisheries Rehabilitation Programs. *Canadian Journal of Fisheries and Aquaculture Science* 38: 1877-1881

Lande, R., 1998. Anthropogenic, Ecological and Genetic Factors in Extinction and Conservation. *Researches On Population Ecology* 40: 259 - 269

Lee, W., Kocher, D., 1996. Microsatellite DNA Markers for Genetic Mapping in *Oreochromis niloticus*. *Journal of Fish Biology* 49: 169 - 171

Levene, H., 1953. Genetic Equilibrium when More than One Ecological Niche is Available. *American Naturalist* 87: 131 - 133

Levinson, G., and Gutman, G. A., 1987. Slipped-strand Mispairing: A Major Mechanism for DNA Sequence Evolution. *Molecular Biology and Evolution* 4: 203 - 221

Lewis, P. O. and Zaykin, D., 2001. Genetic Data Analysis: Computer Program for the Analysis of Allelic Data. Version 1.0 (d16c). Free Program Distributed by the Authors over the Internet from http://lewis.eeb.uconn.edu/lewishome/software.html

Lewontin, R. C. and Hubby, J. L., 1966. A Molecular Approach to the Study of Genetic Heterozygosity in Natural Populations. II. Amount of Variation and Degree of Heterozygosity in Natural Populations of *Drosophila pseudoobscura*. *Genetics* 54: 595 - 609

Lewontin, R. C., 1974. The Genetic Basis of Evolutionary Change. Columbia University Press. New York

Lewontin, R. C., 1991. Twenty-Five Years Ago in Genetics: Electrophoresis in the Development of Evolutionary Genetics: Milestone or Millstone? *Genetics* 128: 657 - 662

Louis, E. J. and Dempster, E. R., 1987. An Exact test for Hardy-Weinberg and Multiple Alleles. *Biometrics* 43: 805 - 811

Lowe-McConnell, R., 1997. EAFRO and After: A Guide to Key Events Affecting Fish Communities in Lake Victoria (East Africa). *South African Journal of Science* 93: 570 - 574

Majumdar, K. C., McAndrew, B. J., 1986. Relative DNA Content of Somatic Nuclei and Chromosomal Studies in Three Genera, *Tilapia, Sarotherodon, and Oreochromis* of the Tribe Tilapiini (Pisces, Cichlidae). *Genetica* 68: 175 -188

Maruyama, T., 1970. Effective Number of Alleles in a Subdivided population. *Theoretical Population Biology* 1: 273 - 306

McAndrew, B. J. and Majumdar, K. C.,1983. Tilapia Stock Identification Using Electrophoretic Markers. *Aquaculture* 30: 249 - 261

Maynard-Smith, J., 1989. Evolutionary Genetics. Oxford University Press, Oxford

McConnell, S. K., Hamilton, L., Morris, D., Cook, D., Paquet, D., Bentzen, P. and Wright, J., 1995a. Isolation of Salmonid Microsatellite Loci and their Application to the Population Genetics of Canadian East Coast Stocks of Atlantic Salmon. *Aquaculture* 137: 19 – 30

McConnell, S. K., O'Reilly, P., Hamilton, L., Wright, J. M. and Bentzen, P., 1995b. Polymorphic Microsatellite Loci from Atlantic Salmon (*Salmo salar*): Genetic Differentiation of North American and European Populations. *Canadian Journal of Fisheries and Aquaculture Science* 52: 1863 – 1872

McVeigh, S., 1995. Ganskraal – The First Small Scale Fish. *Farmers Weekly* May 12: 8 – 11

McVeigh, S., 1998. Fish: Small Scale Farming with a Big Future. *Farmers Weekly* July 31: 10 - 13

Meyer, A., Kocher, T. D., Basasibwaki, P. and Wilson, A. C., 1990. Monophyletic Origin of Lake Victoria Cichlid Fishes Suggested by Mitochondrial DNA Sequences. *Nature* 347: 550 - 553

Michalakis, Y. and Excoffier, L. , 1996. A generic estimation of population subdivision using distances between alleles with special reference to microsatellite loci. *Genetics* 142: 1061 - 1064

Milkman, R. D., 1967. Heterosis as a Major Cause of Heterozygosity in Nature. *Genetics* 55: 493 - 495

Miller, M. P., 1997. Tools for Population Genetic Analysis (TFPGA) 1.3: a Windows Program for t he Analysis of Allozyme and Molecular Population Genetic Data. Computer Software Distributed by Author.

Moralee, R. D. and Van der Waal, B. C. W., 2000. Biochemical Genetic Markers to Identify Hybrids Between the Endemic *O. mossambicus* and the Alien Species, *O. niloticus* (Pisces: Cichlidae). *Water SA* 26: 263 - 268

Mullis, K., Falcoma, F., Scharf, S., Snikl, R., Horn, G. and Erlich, H., 1986. Specific Amplification of DNA *in vitro*: The Polymerase Chain Reaction. *Cold Spring Harbor Symposium on Quantitative Biology* 51: 260

Mullis, K., 1990. The Unusual Origin of the Polymerase Chain Reaction. *Scientific America* 262: 56

Nei, M., 1977. F – Statistics and Analysis of Gene Diversity in Subdivided Populations. *Annals of Human Genetics* 41: 225 - 223

Nei, M., 1978. Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics* 89: 583 - 590

Nei, M. and Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press. New York.

Nevo, E., 1978. Genetic Variation in Natural Populations: Patterns and Theory. Theoretical Population Biology 13: 121 - 177

Nielsen, J. L., Carpanzano, C., Fountain, M. C. and Gan, C. A., 1997. Mitochondrial DNA and Nuclear Microsatellite Diversity in Hatchery and Wild *Oncorhynchus mykiss* from Freshwater Habitats in Southern California. *Transactions of the American Fisheries Society* 126: 397 – 417

Nielsen, J. L. and Fountain, M. C., 1999. Microsatellite Diversity in Sympatric Reproductive Ecotypes of Pacific Steelhead (*Oncorhynchus mykiss*) from the Middle Fork Eel River, California. *Ecology of Freshwater Fish* 8: 159 – 168

Norusis, M. J., 1994. SPSS for Windows Professional Statistics Release 6.1. SPSS Inc., Michigan.

Ogutuohwayo, R. and Hecky, R. E., 1991. Fish Introduction in Africa and Some of their Implications. *Canadian Journal of Fisheries and Aquaculture Science* 48: 8 - 12

Ohta, T. and Kimura, M., 1973. A Model of Mutation Appropriate to Estimate the Number of Electrophoretically Detectable Alleles in a Finite Population. *Genetic Research* 22: 201 - 204

Ohta, T., 1992. The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics* 23: 263 - 286

O'Reilly, P. and Wright, J. M., 1995. The Evolving Technology of DNA Fingerprinting and its Application to Fisheries and Aquaculture. *Journal of Fish Biology* 47: 29 - 55

Orti, G., Pearse, D. E. and Avise, J. C., 1997. Phylogenetic Assessment of Length Variation at a Microsatellite Locus. *Proceedings of the National Academy of Science* 94: 10745 - 10749

Paetkau, D. and Strobeck, C., 1995. The Molecular Basis and Evolutionary History of a Microsatellite Null Allele in Bears. *Molecular Ecology* 4: 519 – 520

Paetkau, D., 1999. Microsatellites Obtained Using Strand Extension: An Enrichment Protocol. *Biotechniques* 26: 690 - 696

Parker, K. M., Sheffer, R. J. and Hedrick, P. W., 1999. Molecular Variation and Evolutionary Significant Units in the Endangered Gila Topminnow. *Conservation Biology* 13: 108 - 116

Pemberton, J. M., Slate, J., Bancroft, D. R. and Barrett, J. A., 1995. Non-amplifying Alleles at Microsatellite Loci – A Caution for Parentage and Population Genetic Studies. *Molecular Ecology* 4: 249 – 252

Pike, T., 1983. *Oreochromis mossambicus* in Natal: Natural Distribution and Growth Studies. *International Symposium on Tilapia in Aquaculture*, Tel Aviv.

Pearce, A. F., 1977. The Shelf Circulation off the East Coast of South Africa. M. Sc. Thesis. Department of Physics, University of Natal

Pimm, S. L., Rusell, G. J., Gittleman, J. L. and Brooks, T. M., 1995. The Future of Biodiversity. *Science* 269: 347 - 350

Pullin, R. S. V., 1988. Tilapia. Genetic Resources for Aquaculture. ICLARM, Philippines

Pullin, R. S. V., Bhukaswan, T., Tonguthai, K., Maclean, J. L., 1988. *Oreochromis mossambicus* is not Universally a Nuisance Species: the Sri Lankan Experience. *The Second International Symposium on Tilapia in Aquaculture*. DOF, Thailand and ICLARM, Philippines

Pouyaud, L. and Agnese, J., 1995. Phylogenetic Relationships Between 21 Species of Three Tilapiine Genera *Tilapia, Sarotherodon and Oreochromis* Using Allozyme Data. *Journal of Fish Biology* 47: 26 - 38

Rao, C. B. and Majumdar, K. C., 1998. Multivariate Map Representation of Phylogenetic Relationships: Application to Tilapiine Fish. *Journal of Fish Biology* 52: 1199 - 1217

Raymond, M. and Rousset, F., 1995. An Exact test for Population Differentiation. *Evolution* 49: 1280 - 1283

Rice, W. R., 1989. Analyzing Tables of Statistical Tests. *Evolution* 43: 223 - 225

Rico, C., Rico, I. and Hewitt, G., 1996. 470 Million Years of Conservation of Microsatellite Loci Among Fish Species. *Proceedings of the Royal Society of London B* 263: 549 - 557

Rognon, X., Andriamanga, M., McAndrew, B. and Guyomard, R., 1996. Allozyme Variation in Natural and Cultured Populations in Two tilapia Species: *Oreochromis niloticus* and *Tilapia zillii*. *Heredity* 76: 640 - 650

Rognon, X. and Guyomard, R., 1997. Mitochondrial DNA differentiation among East and West African Nile tilapia Populations. *Journal of Fish Biology* 51: 204 - 207

Rousset, F., 1996. Equilibrium Values of Measures of Population Subdivision for Stepwise Mutation Processes. *Genetics* 142: 1357 - 1362

Ruzzante, D. E., 1998. A Comparison of Several Measures of Genetic Distance and Population Structure with Microsatellite Data: Bias and Sampling Variance. *Canadian Journal of Fisheries and Aquaculture Science* 55: 1 - 14

Ryman, N., Utter, F. and Laikre, L., 1995. Protection of Intraspecific Biodiversity of Exploited Fishes. *Reviews in Fish Biology and Fisheries* 5: 417 – 446

Ryman, N., 1997. Minimizing Adverse Effects of Fish Culture: Understanding the Genetics of Populations with Overlapping Generations. *Journal of Marine Science* 54: 1149 - 1159

Sarkar, S., 1992. Haldane and the Emergence of Theoretical Population Genetics, 1924 – 1932. *Journal of Genetics* 71: 73 - 79

Schneider, S., Kueffer, J., Roessli, D. and Excoffier, L., 1997. ARLEQUIN (version 1.1). Geneva.

Schlotterer, C. and Tautz, D., 1992. Slippage Synthesis of Simple Sequence DNA. *Nucleic Acids Research* 20: 211 - 215

Shriver, M. D., Jin, L., Chakraborty, R. and Boerwinkle, E., 1993. VNTR Allele Frequency Distributions Under the Stepwise Mutation Model: A Computer Simulation Approach. *Genetics* 134: 983 - 993

Skelton, P. H., 1987. South African Red Data Book – Fishes. *South African National Scientific Programmes Report* 137. South Africa

Skelton, P., 1993. A Complete Guide to the Freshwater Fishes of Southern Africa. Southern Book Publishers, Halfway House

Skibinski, D. O. F., Gallagher, C. and Beynon, C. M., 1994. Mitochondrial-DNA Inheritance. *Nature* 368: 817 - 818

Skibinski, D. O. F. and Ward, R. D., 1998. Are Polymorphism and Evolutionary Rate of Allozyme Proteins Limited by Mutation or Selection? *Hereditas* 81: 692 - 702

Slatkin, M., 1985. Rare Alleles as Indicators of Gene Flow. *Evolution* 39: 53 - 65

Slatkin, M., 1993. Isolation by Distance in Equilibrium and Non-equilibrium populations. *Evolution* 47: 264 - 279

Slatkin, M., 1995. A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. *Genetics* 139: 457 - 462

Small, M. P., Withler, R. E. and Beacham, T. D., 1998. Population Structure and Stock Identification of British Columbia Coho Salmon, *Onchorhynchus kisutch*, Based on Microsatellite DNA Variation. *Fisheries Bulletin* 96: 843 – 858

Sokal, R. R. and Rohlf, F. J., 1995. Biometry. Third Edition. W.H. Freeman and Company. USA

Sved, J. A., Reed, T. E. and Bodmer, W. F., 1967. The Number of Balanced Polymorphisms that can be Maintained in a Natural Population. *Genetics* 55: 469 - 481

Takazaki, N. and Nei, M., 1996. Genetic Distances and Reconstruction of Phylogenetic Trees from Microsatellite DNA. *Genetics* 144: 389 - 399

Tave, D., 1993. Genetics for Fish Hatchery Managers. Second Edition. Van Nostrand Reinhold. New York

Tessier, N., Bernatchez, L. and Wright, J. M., 1997. Population Structure and Impact of Supportive Breeding Inferred from Mitochondrial and Microsatellite DNA Analysis in Land-locked Atlantic Salmon *Salmo salar* L. *Molecular Ecology* 6: 735 - 750

Thorpe, J. E.; Koonce, J. F., Borgeson, D., Henderson, B., Lamsa, A., Maitland, P. S., Ross, M. A., Simon, R. C. and Walters, C., 1981. Assessing and Managing Man's Impact on Fish Genetic Resources. *Canadian Journal of Fisheries and Aquaculture Science* 38: 1899 - 1907

Trewavas, E., 1983. Tilapiine Fishes of the Genera *Sarotherodon*, *Oreochromis* and *Danakilia*. British Museum (Natural History). London

Utter, F., 1998. Genetic Problems of Hatchery-Reared progeny Released into the Wild, and how to Deal with them. *Bulletin of Marine Science* 62: 623 - 640

Valdes, A. M., Slatkin, M. and Freimer, N. B., 1993. Allele Frequencies at Microsatellite Loci: the Stepwise Mutation Model Revisited. *Genetics* 133: 737 -749

Van der Bank, F. H. and Ferreira, J. T., 1987. Morphometric and Electrophoretic Analyses of Three Wild Populations of *O. mossambicus*. *Proceedings of the World Symposium on Selection, Hybridisation and Genetic Engineering in Aquaculture*, Bordeaux, 27 – 30 May, 1986

Van der Bank, F. H., Grant, W. S. and Ferreira, J. T., 1989. Electrophoretically Detectable Genetic Data for Fifteen Southern African Cichlids. *Journal of Fish Biology* 34: 465 - 483

Van der Bank, F. H., 1994. Molecular Systematics of Fifteen Old World Cichlid Species Using Allozyme Data. *Biochemical Systematics and Ecology* 22: 791 - 801

Van der Waal, B. C. W., 2000. Mozambique Tilapia Threatened by the Exotic Nile Tilapia and Hybridisation. Department of Biological Sciences. University of Venda for Science and Technology

Van der Waal, B. C. W. and Bills, R., 2000. *Oreochromis niloticus* (Teleostei: Cichlidae) now in the Limpopo River System. *South African Journal of Science* 96: 47 - 48

Volckaert, F. A. M. and Hellemans, B., 1999. Survival, Growth and Selection in a Communally Reared Multifactorial Cross of African Catfish (*Clarias gariepinus*). *Aquaculture* 171: 49 - 64

Waldbieser, G. C. and Wolters, W. R., 1999. Application of Polymorphic Microsatellite Loci in a Channel Catfish *Ictalurus punctatus* Breeding Program. *Journal of the World Aquaculture Society* 30: 256 – 262

Wallace, B., 1968. Topics in Population Genetics. Norton, New York

Weber, J. L., 1990. The Informativeness of Human $(dC\text{-}dA)_n.(dG\text{-}dT)_n$ Polymorphisms. *Genomics* 7: 524 - 530

Weber, J. L. and Wong, C., 1993. Mutation of human short tandem repeats. *Human Molecular Genetics* 2: 1123 - 1128

Weir, B. S. and Cockerham, C. C., 1984. Estimating F-Statistics for the Analysis of Population Structure. *Evolution* 38: 1358 - 1370

Weir, B. S., 1996. Genetic Data Analysis 2nd edn. Sinauer, Sunderland, MA

Witte, F., Goldschmidt, T., Wanink, J. H., van Oijen, M. J. P., Goudswaard, P. C., Witte-Maas, E. L. M. and Bouton, N., 1992. The Destruction of an Endemic Species Flock: Quantitative Data on the Decline of the Haplochromine Cichlids of Lake Victoria. *Environmental Biology of Fishes* 34: 1 - 28

Wohlfarth, G. W., 1994. The Unexploited Potential of Tilapia Hybrids in Aquaculture. *Aquaculture and Fisheries Management* 25: 781 - 788

Workman, P. L. and Niswander, J. D., 1970. Population Studies on Southwestern Indian Tribes. II. Local Genetic Differentiation in the Papago. *American Journal of Human Genetics* 22: 24 - 29

Worthington, E. B. and Lowe-McConnell, R., 1994. African Lakes Reviewed – Creation and Destruction of Biodiversity. *Environmental Conservation* 21: 199 - 213

Wright, S., 1921. Systems of Mating. *Genetics* 6: 111 - 178

Wright, S., 1931. Evolution in Mendelian Populations. *Genetics* 16: 97 - 159

Wright, J. M., 1989. Nucleotide Sequence, Genomic Organization and Evolution of a Major Repetitive DNA Family in tilapia (*Oreochromis mossambicus/hornorum*). *Nucleic Acids Research* 17: 5071 - 5079

Yokota, M. and Watanabe, S., 1997. One-way Gene Flow by Stocking and its Effects on a Fish Population. *Fisheries Science* 63: 539 - 542

Zhang, D. and Hewitt, G. M., 1998. Field Collection: Animals. Molecular Tools for Screening Biodiversity. Edited by Karp, A., Isaac, P. G. and Ingram, D. S. Chapman and Hall, London. pp 46 – 48

Zar, J. H., 1984. Biostatistical Analysis. 2nd ed. Prentice-Hall International, Inc., Englewood Cliffs, New Jersey.

## Internet Resources

FAO Fisheries Statistic Database: ***www.fao.org/fi/statist/statist.html***

Genbank Entrez Viewer: ***www.ncbi.nlm.nih.gov:80/Database/index.html***

Genome Web – Other Vertebrate Genome Database: ***www.Hgmp.mrc.ac.uk/Genome.Web/vert-gen-db.html***

Lake Malawi Cichlid Research: ***http://tilapia.unh.edu/WWWPages/TGP/CA-Melanochromis.html***

Linkage maps for Aquaculture Species: ***http://tilapia.unh.edu/WWWPages/TGP/RegProject.html***

Tilapia DNA Reference Families: ***http://tilapia.unh.edu/WWWPages/TGP/Families.html***

Tilapia Home Page: ***http://tilapia.unh.edu/WWWPages/TGP/CA-tilapia.html***

Tom Kocher Home Page: ***http://tilapia.unh.edu/default.html***