Microsatellite markers to identify two species of Tilapiine fish, *Oreochromis mossambicus* (Peters) and *O. niloticus* (Linnaeus)

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

March 2002

Candidate: M.M. Esterhuyse

Supervisors: dr. M.E. D'Amato

Prof F.A.M. Volckaert

DECLARATION

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

Abstract

Forming part of a conservation programme, this study was concerned with two species of Cichlid fish (*Oreochromis mossambicus* and *O. niloticus*), which were brought into contact with each other by unnatural ways. They are now hybridizing to some extent and there is also evidence that the foreign *O. niloticus* may out compete the native *O. mossambicus*. To cast light on what the current distribution is of both these species and the hybrids in Southern Africa, it is important to identify specimens very accurately.

In attempting to find genetic markers to distinguish between two species of Cichlids we tested 20 microsatellite dinucleotide (CA_n) repeats during a preliminary study and found five of these promising to exhibit little intra-specific genetic diversity but large genetic variation between species. We amplified these five loci in 145 individuals from 10 populations, which included the two species and their hybrids. Exact sizes of the fragments were determined using an automated DNA sequencer.

Between the two species, allele sizes were overlapping, but when data were analyzed by statistical models, the differences could be seen for populations, however on individual level there was overlap between the species. The hybrids were found to be intermediate positioned between the two pure species. Our attempt to assign individuals to populations provided doubtful results. Thus, using this set of markers, populations can be ascribed to one of these species, but not individuals by themselves.

Opsomming

As deel van 'n natuurbewarings program, word daar in hierdie studie twee spesies van vis ondersoek was in kontak met mekaar gekom het op onnatuurlike wyse. Hierdie twee visspesies vanuit die CICHLIDAE familie (*Oreochromis mossambicus* en *O. niloticus*) kan hibridiseer wanneer hul saam voorkom, maar dit is ook bekend dat die uitheemse *O. niloticus* die inheemse *O. mossambicus* kan bedreig in terme van leefruimte, kos en broeispasie. Om die voorkoms van hibriede tussen die twee spesies te ondersoek in Suider Afrika se varswater opvangsgebiede, is dit baie belangrik om individue baie akkuraat te identifiseer.

In hierdie poging om genetiese merkers te vind wat die twee spesies van mekaar onderskei, het ons 20 mikrosateliet di-nulkleotied (CA_n) herhalende volgordes op verskillende loci ondersoek. Vyf daarvan het belowend voorgekom om as spesie spesifieke merkers te dien. Die fragmente op die vyf loci is ge-amplifiseer in 145 individue vanuit 10 populasies. Presiese groottes van die fragmente is bepaal met behulp van 'n ge-outomatiseerde DNA volgorde bepaler waarna genotiepes vir elke individu toegeken is.

Tussen die twee spesies het alleel groottes oorvleuel, maar wanneer data ge-analiseer word met behulp van statistiese metodes, was verskille tussen die spesies duidelik op populasie vlak. Die hibriede het intemediêr tussen die twee spesies voorgekom. Dus met behulp van hierdie stel merkers kan onderskei word tussen die twee spesies op populasie vlak, hoewel individue nie op sig self identifiseer kan word nie.

Table of contents

List of tables

List of figures

Acknowledgements

I would like to thank my family and friends for their support and patience throughout the stages 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 the first practical aspects of this study; Mr Danie Brink for the financial support; Prof. Ben van der Waal, Edward Hall and dr Graham Mair for sample material which I obtained from them for this study; Carl Van Heerden and Rene Van Zyl, for their assistance with ABI fragment analysis; dr Armand Peeters, for the use of his laboratory and its facilities; dr Christophe Pampoulie for guidance in assignment techniques; Karl Cottenie and Joost Raeymaekers for assistance in other statistical matters; University of Stellenbosch (Department of Genetics) and Catholic University of Leuven, (Laboratory for Aquatic sciences) for providing staff and equipment to obtain training as scientist; Bridgitte Jacobs and Etienne Hurter for their assistance with the administration behind this thesis. Finally, thanks goes to dr Eugenia D'Amato and Prof. Fillip Volckaert for all their trouble in trying to train me as scientist and helping enormously with their faithful guidance at all times.

Chapter 1 Introduction

1 Overview of project

Ever since ecology started, hybridization of species was evident and raised various problems for conservationists, ecologists and aquaculturists. Research done in this study is concerned with the hybridization of two Tilapiine fish species in Southern Africa, both of which belong to the genus *Oreochromis*.

Cichlid fishes from the genus *Oreochromis* are known to have occurred naturally in Northern and Central Africa and was first named in the nineteenth century (Trewavas, 1983). Some animals were able to distribute to southern parts of the continent and because of natural barriers (such as rivers that change morphology after floods) it could not return an evolved as a separate species. This species is now called *Oreochromis mossambicus* (Peters 1852).

Oreochromis niloticus (Linnaeus 1758), the other species of interest in this study, has its natural distribution ranging North, Northwest, and Central Africa. This globally successful aquaculture Tilapiine species has been introduced in many countries all over the world, including Southern Africa (FAO 2001).

From studies done by Iversen (1968) and others (see Trewavas, 1983) it is clear that above-mentioned two tilapia species (which have similar patterns of breeding, feeding, growth and behaviour), hybridize successfully in captivity. These hybrids can often be identified using morphological characteristics, but in most cases, individuals that are expected to be hybrids in the wild, especially F3, F4 etc. resemble either species (*pers. comm.* Ben van der Waal[1]). Some of these invading *O. niloticus* were found in some of the Northern freshwater river systems of South Africa, including the Limpopo River drainage (see Table 1.2).

Management strategies for the conservation of species and enhancement for breeding/commercial purposes require the use of markers which can be used to

[1] University of Venda, RSA

identify the representatives of individual stocks. To distinguish the populations of fish to explore their relationships, diagnostic features must be identified. Any discriminatory feature, may it be morphological or genetic, can be used in this context (Waldman & Wirgin, 1994).

Several techniques could be used to assist in identifying the two species apart from each other and their hybrids. These techniques are mentioned shortly in section 1.5. But in this study we used nuclear molecular genetic markers, particularly, microsatellite markers (shortly: microsatellites).

Microsatellites are tracts of repetitive nucleotides, whether di-, tri-, tetra- or penta-repeats (Tautz, 1989). They have become a very popular marker system which is used in many studies particularly because of their high mutation rate, high allelic number, and practically, minimal amounts of DNA which are required to conduct a study (for detailed discussion, see section 1.5).

There are several positions in the genome where these microsatellite repeats occur, and to study these, one can easily amplify them using the Polymerase Chain Reaction (PCR). When these fragments are obtained, their length in number of base pairs (bp) is measured. For a study such as this, when primers for microsatellite loci are unavailable for the focal species, one would use primers designed for homologous loci in related species. It is possible that different levels of diversity are observed when nonspecific primers are used, and that different species display an overlapping range of allele sizes, although a few of them might be species-specific. In such cases these loci can still be used because of the development of statistical methods.

Microsatellites are thus used to "mark" certain differences between species (or populations, or individuals, depending on the aim of the study). If "good" differentiating markers can be found, these can be applied to identify hybrids, for if an individual carrying alleles occurring in the two species from which it is hybridised, it can be identified as a hybrid.

The aim of this study is to identify species specific markers which could be used in following studies to identify hybrids.

2 Objectives of the study

It is a well-known fact that the Nile tilapia competes successfully with the Mozambique tilapia in terms of natural resources such as food and shelter. These two species have similar patterns of feeding, breeding, growth and behavior, and it is documented that these two species hybridise successfully in captivity producing all male offspring (Iversen, 1968 and Trewavas, 1983).

Because of the fact that the species under investigation are both used in aquaculture, which implies an unnatural movement of genetic material, it is necessary to manage the genetic resources of the fish properly in order to prevent problems mentioned below. Kocher & Carleton (1997) states some three "genetic problems" as a result of poorly managed genetic resources:

1. Loss of pure species through mismanagement of interspecific hybridization;
2. High level of inbreeding depression and
3. Contamination of genetically improved strains by introgression from feral species.

If the genetic resources are not managed properly, the following scenarios are relevant in this specific case:

1. *Competition:* Nile tilapia could represent a selective pressure on the Mozambique tilapia and possibly on other fish species as well. The species may co-exist for some time until one becomes locally rare or dominant. Total displacement of Mozambique tilapia by the Nile tilapia as a result of direct competition for food, shelter, and breeding sites as has happened in Lake Victoria, could lead to the extinction of the Mozambique tilapia (Trewavas, 1983).
2. *Habitat modification*: Nile tilapia can change the environment to such an extent that it is not anymore favorable for local fish (e.g. depleting natural resources can result in destruction of other animals' habitat) or some of the other organisms in the same habitat.

3. *Loss of local genetic resources through hybridization*: Nile tilapia hybridizes with local Mozambique Tilapia producing a cross with unknown characteristics, many of which are unexpected and might be negative (e.g. lower growth rate, lower maximum size attained, lower resistance to cold conditions, higher susceptibility to local parasites). Depending on the success of hybridization, one or both species can disappear completely (*pers comm.* Ben van der Waal).

But what if *O. niloticus* is a "better" aquaculture species than *O. mossambicus*, and what if the hybrids are even better than both? A few studies were done already dealing with this question (see section 3.5 of this chapter), but no *molecular* markers were used in this respect yet. Our research is thus aiming to provide baseline knowledge for the identification of *O. mossambicus, O. niloticus* and its hybrids by the use of molecular nuclear markers.

Various factors play a role in the aquaculture of tilapia in South Africa. First, South African export of fish is regulated by the Department of Trade and Industry of the South African government, stating that no export or import of *O. niloticus* is permitted. Secondly, farming with *O. niloticus* in South Africa is not permitted if conditions on farms are not completely preventing the escape of fish into natural ecological systems. In such cases no fish may leave the farm alive. Regulations are therefore so strict, which make it almost not profitable to farm these fish in Southern Africa. Thus we find it rather appropriate to develop/cultivate indigenous *O. mossambicus,* which occur in most of the region's rivers already. With selection programmes it might be possible to develop an indigenous aquaculture species with characteristics that suit the needs of aquaculturists, but perform as aquaculture species better or equal to the Nile tilapia, thus providing a reason for us to conserve the indigenous *O. mossambicus*.

The Mozambique tilapia has however a different genetic make-up than that of the Nile tilapia, and therefore enlarge the "gene pool" for the "aquaculture tilapias". This provides another reason for us to conserve the Mozambique tilapia.

This study forms part of the bigger picture sketched above in the way that it provides a "tool" to investigate the current state of genetic resouces. The main questions in this bigger picture are:

1. Is *O. niloticus* endangering the local species and the ecosystem?
2. What is the current conservational status in the Southern African ecology regarding these Tilapias?
3. What management should be installed if necessary?

All these have to be answered in the light of a very controversial question: What is a species? This topic was studied in literature to cast light on whether this research project is necessary at all. As we allied us with the phylogenetic species concept[2] (for this is an ongoing debate), and found the two groups of fish at hand are in fact different species, question 1 above has been answered.

In answering the question of the conservational status in South Africa, groups, which are concerned with tilapia ecology in the country, were approached. Several cases are reported (see section 3.3; Table 1.2) of *O. niloticus* in South Africa, which indicates a danger of this exotic species to pose a threat to the local species and ecology.

To be able to study the conservational status it is necessary to distinguish the two species from each other, as well as from their hybrids. In doing this, we had to be sure that the individuals we use were (a) pure species, (b) include populations from different sites for each species, (c) use methods not relying on morphology because it is not always possible to identify these fish, especially the hybrids, and (d) if using molecular markers, identifying the correct number of markers for use in statistical analysis.

Allozyme markers to identify the two species of interest were reported in McAndrew & Majumdar (1983), Pouyaud & Agnèse (1995) and Moralee et al. (2000). Hybrids were included only in last study.

---

[2] See section 1.4 for a brief discussion on species concepts and its applications in terms of this project.

We made use of dinucleotide nuclear genetic markers (CA-microsatellites) to distinguish the two species and also included some fish that appeared as (are expected to be) hybrids to provide a tool available to study the current conservational status of the two species in South Africa (question 2 above).

3 Background on Tilapia

3.1 Taxonomy

The taxonomic position of both tilapias studied is as follows:

Order: PERCIFORMES
Family: CICHLIDAE
Genus: *Oreochromis*
Species: *O. mossambicus*
O. niloticus

The family Cichlidae constitutes a monophyletic group in the order Perciformes. Monophyly of the cichlid family is indicated by the presence of at least nine synapomorphic morphological characters (Trewavas, 1983). Since the distribution of cichlids ranges from South and Central America and Mexico to tropical Africa, Madagascar, southern India, and Sri Lanka, the cichlid family must have arisen before the separation of Africa, South America, and India by continental drift more than 100 million years (MY) ago. The scientific names of Tilapia species have been actively revised a lot in the last 30 years (Popma & Masser, 1999), causing terrible confusion. "Tilapia" is the common name for about 70 species of fish in the family CICHLIDAE (perch-like fish), and according to the literature, native to the fresh waters of tropical Africa (Trewavas, 1983).

Until the late 1970's the tilapias were all classified into a single genus: *Tilapia*. However, most taxonomists now classify them into three genera according to their breeding behavior: *Tilapia* (substrate spawners), *Sarotherodon* (parental or biparental mouthbreeders), and *Oreochromis* (maternal mouthbreeders). (Trewavas, 1983; Popma & Masser, 1999).

Not all taxonomists accept this classification and it is still common to see the Nile tilapia, *Oreochromis niloticus* referred to as *Tilapia nilotica*, particularly in American literature, or *Sarotherodon niloticus* although the latter of less frequent use.

3.2 Morphology

Tilapiine fish can easily be identified by an interrupted lateral line characteristic of the Cichlid family of fishes. The body shapes are laterally compressed and deep-bodied with long dorsal fins. The forward portion of the dorsal fin is heavily spined. Spines are also found in the pelvis and anal fins. There are usually wide vertical bars down the sides of fry, fingerlings, and sometimes adults. Nile tilapia has strong vertical bands, while Mozambique tilapia have weak or no bands on the caudal fin (see Fig. 1.1 and 1.2 and Table 1.1). Male Mozambique tilapias also have upturned snouts. Mature male Nile tilapia has grayer pink pigmentation in the throat region, while Mozambique tilapia have a more yellow coloration. However, coloration is often an unreliable method of distinguishing tilapia species because environment, state of sexual maturity, and food source greatly influence color intensity.

Figure 1.1 *Oreochromis mossambicus* picture with courtesy of Windsor Aguirre, Gulf Coast Research Laboratory Museum Institute of Marine Sciences. Picture taken at J.L. Scott Marine Education Center. Biloxi, MS.

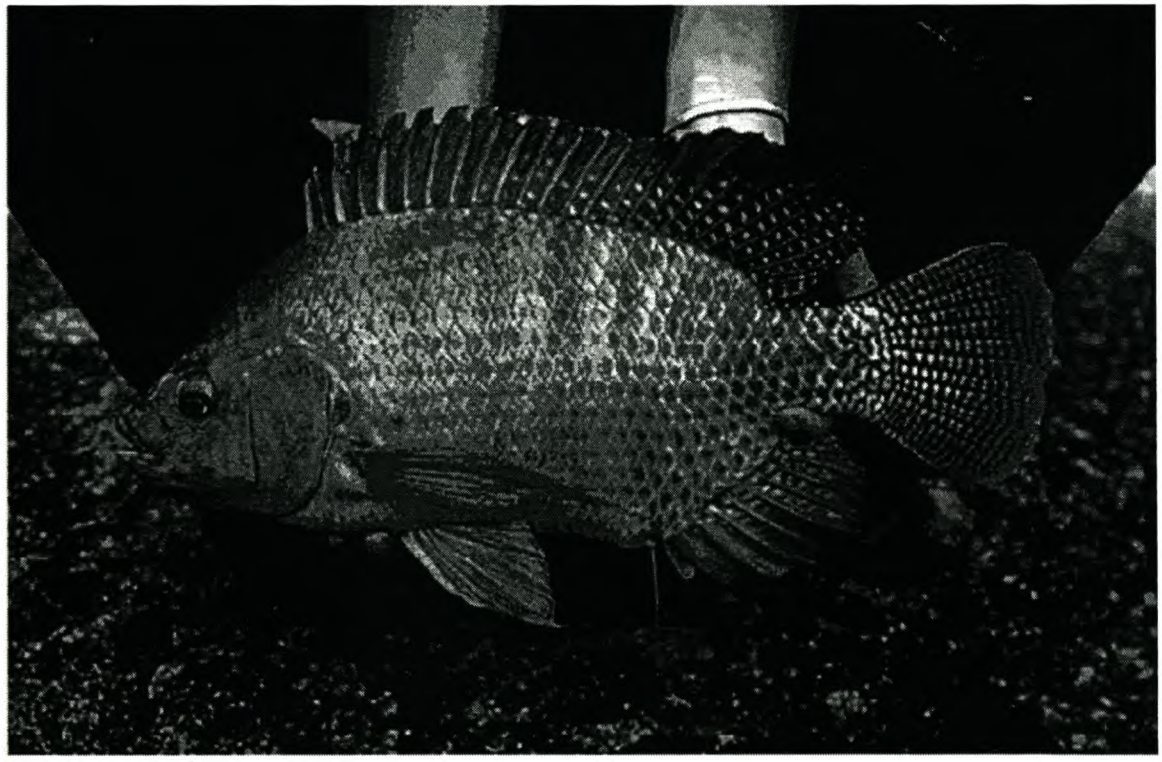

Figure 1.2 *Oreochromis niloticus*

Table 1.1 Morphological differences between *O. niloticus* and *O. mossambicus* (*pers comm.* Ben van der Waal; Skelton, 1993)

Nile tilapia	Mozambique tilapia
Prominent vertical bars on tail fin	Tail fin with spots but never in vertical lines
Dorsal fin never red edged, tail end can be red	Tail and especially dorsal fin edged with red
Head and body dark green to silvery grey	Head often darker as body, light silver brown to yellowish to blue grey and even black
Eight clear vertical bands on body and tail base	Three unclear spots in a horizontal row on flanks, six or seven unclear vertical band on body
Spines of dorsal fins very thick and prominent	Anal and dorsal fin spines well developed but thickened
Dorsal and anal fins relatively high compared to depth	Dorsal and anal fins not as high as in the Nile tilapia
Eye typically red coloured	Eye from yellow to dark brown, never red
Males in breeding dress with a pinkish head and Body	Males in spawning dress very dark to black with a white chin
Males never have a prominent enlarged mouth and concave profile of the snout	Males often with enlarged mouth and concave head profile
Lower half of first gill arch with 20 to 26 rakers	Lower half of first gill arch with 16 to 20 rakers

3.3 Distribution

Tilapias are native only to the African continent (Popman & Masser, 1999). As mentioned before, *O. mossambicus* is indigenous to the Southern African subregion, whilst *O. niloticus* original distribution is Northern, Western and Central Africa.

Oreochromis mossambicus was originally found in the area from the lower Zambezi up to Tete and the coastal part of Mozambique from Quelimane, just north of the Zambezi delta, southwards nearly to the Limpopo (Trewavas, 1983). In 1976 the Pongola River (about 27°S 31°N) was given the southern limit of the distribution in fresh water, but because of its wide salinity tolerance it was able to extend in estuaries farther south, as is confirmed by many records and examples (Trewavas, 1983).

The latest documentation on Southern African Tilapia distribution (Skelton, 1993) states that *O. mossambicus* is distributed: (a) along the east coastal rivers from

the lower Zambezi system south to the Bushmans system, eastern Cape Province; (b) South of the Phongolo system, naturally confined to closed estuaries and coastal reaches of rivers; (c) widely dispersed beyond this range to inland regions and to the south-west and west coastal rivers including the lower Orange and rivers of Namibia; (d) introduced to tropical and warm temperate localities throughout the world.

Dispersal of the Nile tilapia started in 1924 (Trewavas, 1983) and it was introduced into Southern Africa before 1955 from Israel for aquacultural use as a fodder fish (Skelton, 1993). *O. niloticus* occupy a range that extends from 8° S to 32° N from 1830 m to sea level, which includes the Nile basin, Rift Valley lakes and certain West African rivers. The temperature, seasonal range of temperatures, and the chemistry of waters cover wide ranges (Trewavas, 1983) as the distribution/occupation changes. In Southern Africa it is distributed to the Cape Flats area, southwest Cape, Natal and Kariba basin in Zimbabwe.

Figure 1.3 Natural distribution of *Oreochromis niloticus* (Trewavas, 1983)

Nile tilapia was first released in Lake Victoria where they displaced two valuable large indigenous and endemic tilapias, *O. esculentus* and *O. variablilis*. Today these species are threatened or extinct. As a result of the proven success of *O. niloticus* in fish farming, it was imported into Zambia in 1982 to two fish farms.

From there it escaped into the Kafue River where it was collected in 1992. Then it was transferred further south to fish farms near Lake Kariba and its presence has been reported from the lake. It is just a question of time before it will hybridize with or outcompete the local species. In the nineties it was distributed to fish farms and angling ponds south of Lake Kariba. Anglers and fish farmers distributed it to dams around Bulawayo, Zimbabwe, including the Umzingwane River (Van der Mheen, 1997). Specimens from the Zhovhe Dam, about 60 km from the Limpopo River, were positively identified by the JLB Smith Institute in 1999. Additionally, it is known from no less than twelve different dams in Zimbabwean tributaries of the Limpopo River (Ben van der Waal, *pers comm.*).

Table 1.2 Presence of *O. niloticus* in the Limpopo river system (*pers comm* Ben van der Waal.)

Place	Date
Manxeba Pan, Phauri, Kruger National Park	20 November 1996
Dam in Mzingwane River, Bubiana Conservancy, West Nicolson, Zimbabwe	1996*
Fish pond on Farm Den Staat, Owner G. Hodgson	6 April 1998**, 23 July 1988
Pool in Limpopo River opposite Den Staat	23 July 1998
Pool at Confluence of Limpopo and Shashe River	24 July 1998
Zhovhe Dam, Umzingwane River, Southern Zimbabwe	1998
Limpopo River, Farm Teuniskloof, Tolwe	1999***
Pool in Limpopo River, Malips Drift, Farm	16 November 1999
Confluence of Luvuvhu and Limpopo Rivers	19 October 1999**
Mangala, lower Luvuvhu River, KNP	19 October 1999**

(Van der Waal, 1997, Van der Waal & Bills 1997)

* Brian Marshall, University of Zimbabwe, in Litt.

** Mick Angliss, report 1998, personal communications, 2000

***Tinus du Plessis, personal communications, 2000

3.4 Biology

O. niloticus directly competes with Mozambique tilapia (*O. mossambicus*) for food and breeding place, and hybridization has been reported (Moralee, et al., 2000). Thus the biology of these two Tilapiine species under investigation is important in the light of this study for several obvious reasons:

1. It was the salt- and drought tolerance as well as the temperature range of the Mozambique tilapia, which made it able to distribute south of its range.
2. The similarities between the species regarding their feed, habitat and breeding requirements, made them able to either outcompete one the other, or successfully hybridize (Moralee, et al., 2000).
3. Certain aspects of its biology make either one of them best suited to farm with as aquacultural species.

Positive aquacultural characteristics of both species *O. mossambicus* and *O. niloticus* are their tolerance to poor water quality and the fact that they eat a wide range of natural food organisms. Constraints to the development of commercial tilapia farming mentioned in the literature, are their inability to withstand sustained water temperatures below 10 to 10.5 °C and early sexual maturity that results in spawning before fish reach market size.

(a) Feed

Of all cichlids, the genus *Oreochromis* contains all the best species for aquaculture regarding feeding behavior and diet. They have similar feeding habits to *Sarotherodon* spp.

When feeding, tilapias do not disturb the pond as aggressively as common carp. However, they effectively browse on live benthic invertebrates and bacteria-laden detritus. Tilapias also feed on mid-water invertebrates. They are not generally considered piscivorous, but juveniles do consume larval fish. In general, tilapia use natural food so efficiently that crops of more than 3 000 kg/ha can be sustained in well-fertilized ponds without supplemental feed (Popma & Masser,

1999). The nutritional value of the natural food supply in ponds is important, even for commercial operations that feed fish intensively.

Two mechanisms help tilapia digest filamentous and planktonic algae and succulent higher plants: (1) Physical grinding of plant tissues between two pharyngeal plates of fine teeth and (2) a stomach pH below 2, which ruptures the cell walls of algae and bacteria. The commonly cultured tilapias digest 30 to 60 percent of the protein in algae; blue-green algae are digested more efficiently than green algae (Popma & Masser, 1999).

As a whole, tilapias ingest a variety of natural food organisms, including plankton, some aquatic macrophytes, planktonic and benthic aquatic invertebrates, larval fish, detritus, and decomposing organic matter. Tilapias are often considered filter feeders because they can efficiently harvest plankton from the water. However, tilapias do not physically filter the water through gill rakers as efficiently as true filter feeders such as gizzard shad and silver carp (*Hypopthalmichthys molitrix*). The gills of tilapia secrete a mucous that traps plankton. The plankton-rich mucous, or bolus, is then swallowed. Digestion and assimilation of plant material occurs along the length of the intestine (usually at least six times the total length of the fish). The Mozambique tilapia is less efficient than the Nile or Blue tilapia at harvesting planktonic algae (Popma & Masser, 1999).

The Mozambique tilapia mainly consumes a detritivorus diet although it is able to take advantage of particulate food in plankton when available. It feeds on algae, especially diatoms, and detritus, but large individuals may take insects and other invertebrates (Skelton,1993). In several studies the fry was found to consume zooplankton, with cyclopoid copepods, rotifers, ostracods (Green et al. 1974), Entomostraca (Le Roux, 1956) and omnivory up to about 8 cm (Bruton & Boltt, 1975).

The Nile tilapia is almost entirely herbivorous from a length of about 5 cm, feeding on various types of plant material as seasons change. The relative coarseness or fineness of the pharyngeal dentition, and to some extent also of the teeth of the jaws reflects the different emphasis on possible food items. In

areas where phytoplankton is abundant, it forms the bulk of the fishes' diet. (Trewavas, 1983). The fry are omnivorous, consuming copepods, hydracarines and various insects, and peck at aufwuchs and detritus.

Some aspects of the diet or mechanisms in the two tilapia's feeding behaviour may be a reason for one of them to outcompete the other in natural environment, and/or make either one of them more popular as an aquaculture species. It is thus clear in the literature that these two species compete for similar feed in cases where they appear in physically the same habitat and that in some aspects one is preferred above the other as aquaculture species.

(b) Breeding

The three genera in this family (CICHLIDAE) are distinguished from each other mainly by the differences in their way of breeding: all *Tilapia* are nest builders – fertilized eggs are guarded in the nest by a brood parent. The other two genera, *Sarotherodon* and *Oreochromis* are mouth brooders – eggs are fertilized in the nest but parents immediately pick up the eggs in their mouths and hold them through incubation and for several days after hatching. In *Sarotherodon* species, either the male or both male and female are mouth brooders, where, in *Oreochromis* species, only females practice mouth brooding.

For most species of *Oreochromis*, water temperatures need to be above 22°C for spawning to take place. If temperatures are above 22°C all year, spawning will be continuous, if it falls below 22°C, spawning will be seasonal (Pullin, 1997; Popma & Masser, 1999).

Naturally, breeding by *O. mossambicus* takes place in summer, and females raise multiple broods every 3-4 weeks during a season (Skelton, 1993). After hatching, the male plays no part in parental care and can mate with many females at a time, therefore sex ratios in breeding ponds can be as high as 7 females: 1 male. The mouth size of the female determines how many fry are in a brood, bigger females have bigger broods, however usual brood sizes would be 100-500 fry. The eggs are relatively large, producing large fry, which do not

need live feed at first feeding. Removing the eggs or fry prematurely from a brooding female, will increase the frequency of spawning. In poor conditions, species such as *Oreochromis mossambicus* can become sexually mature at a small size (from 10 g) within 4-6 months of hatching. This can lead to ponds becoming overpopulated with small, unmarketable fish. In actively breeding populations of tilapia, much of the resources of the females are tied up with reproduction, either producing eggs, or being unable to feed during mouth brooding. This means that the growth rates of males are much higher than females.

In general, tilapia populations in large lakes mature at a later age and larger size than the same species raised in small farm ponds (Popma & Masser, 1999). The Mozambique tilapia reaches sexual maturity at a smaller size and younger age than the Nile tilapia. When growth is slow, sexual maturity in Nile tilapia is delayed a month or two, but stunted fish may spawn at a weight of less than 20 grams. In poorly fertilized ponds sexually mature Mozambique tilapia may be as small as 15 grams. Under good growing conditions in ponds, the Mozambique tilapia may reach sexual maturity in as little as 3 months of age, when they seldom weigh more than 60 to 100 grams. Males grow about twice as fast as females (Popma & Masser, 1999).

(c) Salt/drought tolerance

Tilapias are more tolerant than most commonly farmed freshwater fish to high salinity, high water temperature, low dissolved oxygen, and high ammonia concentrations.

All *Oreochromis* spp. are tolerant to brackish water. The Nile tilapia is the least saline tolerant of the commercially important species, but grows well at salinities up to 15 ppt, although when hybridized with *S. aureus*, the offspring can tolerate very saline ponds (Trewavas, 1983). Mozambique tilapia grows well at salinities near or at full strength seawater. Therefore, the Mozambique tilapia and some *mossambicus*-derived "red" tilapia are preferred for saltwater culture.

Some strain of the Mozambique tilapia reportedly have spawned in full strength seawater, but its reproductive performance begins to decline at salinities above 10 to 15 ppt, however it performs better at salinities below 5 ppt. Fry numbers decline substantionally at 10 ppt salinity (Popma & Masser, 1999).

O. mossambicus tolerates fresh, brackish or marine waters. It survives lower temperatures (below about 15°C) in brackish or marine waters (Skelton, 1993). Various studies were done on the ability of *O. mossambicus* to tolerate wide ranges of temperature and salinity. These studies were mostly encouraged by the coastal distribution of the species and its presence in the lower reaches only of the southern rivers. In a report by [Lt Col L. Basil] Gardiner in 1947, the species was stated to be found not more than a mile from the ebb and flow at the Bushmans River near Algoa Bay (Trewavas, 1983). This fact suggests that it may only be its tolerance of salt water that permits its migration to a lower temperature area than any of its congeners.

A report cited in Trewavas (1983), stated that from the Umtavina River, on the Natal-Cape Province border, southward to Algoa Bay the species is present in the lower reaches of the rivers, but not above the first formidable obstacle. This means that fish could not come from upstream, but had to swim from the river mouth.

In one of these toleration studies, the biology of these fish was studied to find any association with the ability for salt toleration (Perez & Maclean, 1976). Interestingly, it appeared that the toleration of this species to wide ranges of temperature and salinity is associated with the presence in adults of two types of haemoglobin. One of these, developing at the age of 47 days, is less adversely affected in its affinity for oxygen by high temperatures and salinity than the other, which is present throughout life. Perez & Maclean (1976) suggest that 'the presence of a second haemoglobin allows the adult to exploit environments which are both warmer and more saline than can be tolerated by the larvae'. Additionally, *O. mossambicus* not only survives as an adult in such conditions,

but can also breed in them, and young *Oreochromis* of 2 to 5 cm commonly live in warmer water than adults (Trewavas, 1983).

O. mossambicus appear to be extremely drought tolerant too. They can survive extreme reduction of the body of water in which they find themselves during the dry season by precocious breeding. Several authors documented the survival of this species in a layer of moist sand covered by up to 3 m of sand, the upper two meters of which was quite dry (Trewavas, 1983). This, along with its salt tolerance abilities make *O. mossambicus* a very favourable aquaculture species.

(d) Water temperature

The intolerance of tilapia to low temperatures is a serious constraint for commercial culture in temperate regions. The lower lethal temperature for most species is 10 to 11° C for a few days, but the Blue tilapia (*O. aureus*) tolerates temperatures to about 8.9° C. *Oreochromis* generally stops feeding when water temperature falls below 17.2° C. Disease-induced mortality after handling seriously constrains sampling, harvest and transport below 18.3° C. Reproduction is best at water temperatures higher than 26.7°C and does not occur below 20° C. In subtropical regions with a cool season, the number of fry produced will decrease when daily water temperature averages less than 23.9°C. After 16-20 day spawning cycles with 250-gram Nile tilapia, fry recovery was about 600 fry per female brooder at a water temperature of 27.8°C, but only 250 fry per female at 23.9°C. Optimal water temperature for tilapia growth is about 29.4 to 31.1°C. Growth at this optimal temperature is typically three times greater than at 22.2°C (Popma & Masser, 1999).

(e) Habitat

In a survey study of the fishes of the Limpopo system (Gaigher, 1973), *O. mossambicus* was part of the fish described as "pool-living species confined to the warmer Middle- and Low-Veld streams".

Apparently it is only in closed estuaries and coastal lakes that this southern distribution is represented (Trewavas, 1983). Several factors affect it. The fish avoid strong currents and in such conditions will remain in vegetation near the bands. Suitable breeding sites and marginal vegetation for the shelter of the young are required, and although considerable salinity is tolerated, rapid changes are not. In such favourable coastal waters if marine piscivorous fishes are absent *O. mossambicus* may be common.

3.5 Use of animal

Today, all commercially important tilapia outside of Africa belong to the genus *Oreochromis*, and more than 90 percent of all commercially farmed tilapia outside of Africa are Nile tilapia. Less commonly farmed species are Blue tilapia (*O. aureus*), Mozambique tilapia (*O. mossambicus*) and the Zanzibar tilapia (*O. urolepis hornorum*).

Worldwide harvest of aquaculture tilapia[3] has now surpassed 800 000 metric tons. It is only carp as the most widely farmed freshwater fish in the world (Popma & Masser, 1999) that has higher harvesting. It is documented that the Nile tilapia was one of the first cultured species – more than 3 000 years ago (Popma & Masser, 1999). Tilapias have been called "Saint Peter's fish" in reference to biblical passages about the fish fed to the multitudes. Still today, Nile tilapia is the most cultured species of tilapia in Africa.

O. mossambicus is mainly used in aquaculture as well as commercial and subsistence fisheries (Skelton, 1993). It is also considered a valued angling species. In many biological, physiological and behavioral research studies these fish are being used.

The following factors all contribute to the ease of cultivation of tilapias; a) Resistance to poor water quality and disease, b) tolerance to a wide range of environmental conditions, c) ability to convert efficiently organic, domestic and

[3] The family as a whole

agricultural wastes, into high quality protein, d) good growth rates and e) easy to grow in intensive culture.

Commercial use of hybridization is not widespread because little is known about the characteristics of these hybrids other than their sex ratio.

3.6 Hybrids

Early studies on the hybrids of these fish showed that their temperature tolerance was intermediate between those of the parent species. Growth-rate and survival rate were said to be better, and food conversion more efficient than in the parent samples (Trewavas, 1983). Thus the indigenous Mozambique tilapia may lose its genetic purity and be replaced by hybrid wild populations throughout most of its natural range in time (Moralee et al., 2000). The morphology however was not described.

Tilapia genetics is complex. Hybridization between species sometimes produces offspring with a skewed sex ratio (more males than females or more females than males). Over 25 different hybrid combinations of tilapias have been shown to produce 80 % males (this is desirable for commercial use of the fish because males grow faster than females).

The end result of hybridization between any two *Oreochromis* species is unpredictable. In most cases, the hybrids in fact have a lower adaptive value than either of the two pure species and these hybridizations usually lead to the disappearance of one of the two pure species and eventually the hybrids as well (Agnèse et al., 1998; Moralee et al., 2000). In Lake Victoria (Welcomme, 1967) the hybrids were all males and soon after the introduction of the second species (*O. niloticus*), both species[4] disappeared from the lake. *O. niloticus* is suspected of being the cause of these disappearences. The double experiment of the introduction of *O. niloticus* and *O. mossambicus* in Lakes Italy in Madagascar and Ihema in Rwanda also shows that one cannot predict which species will win

[4] *Oreochromis niloticus* and *O. variablis*

the competition. The consequences of the elimination of a species after hybridization with another species are often also not known. In particular, the vanished species (both *O. mossambicus* and *O. niloticus*) may have left some of its genes in the established species (Agnèse et al., 1998; Moralee et al., 2000).

Several studies report hybridization between *O. mossambicus* with *O. niloticus* (Iversen, 1968; Mires, 1977). The morphology of the hybrids has not been described, but their temperature tolerance was said to be intermediate between those of the parent species; growth-rate and survival rate were said to be better, and food conversion more efficient than in the parent samples (Trewavas, 1983).

In a study by Moralee et al. (2000) morphological data of samples from *O. niloticus* and *O. mossambicus* and their hybrids were studied in an attempt to determine if hybrid identification would be possible. This included counting the dorsal and anal fin spines and rays, lateral scales and gill rakers on the lower part of the first gill arch. The length and width of the last four dorsal fin spines were measured, and X-rays of the fin spines were taken to determine if they are different for the taxa studied. These parameters were compared with results from allozyme data. The results showed that the identification was not always possible.

Mires (1977) reports an approximately 1:1 sex ratio in crosses (both ways) between *O. mossambicus* (stock from South Africa) and *O. niloticus* (stock from Lake Albert). *O. niloticus* female and *O. aureus* male, produces 80-90 % males, with the growth vigour of *O. niloticus* and the cold tolerance of *O. aureus*. *O. hornorum* is the only known species, which consistently produces all male fry when crossed with *O. niloticus* or *O. mossambicus*.

Hybridization between the two species of interest can also produce what is known as Red Tilapia. In one study similar to the present one (Moralee, et al., 2000) red tilapias from aquaria at the Rands Afrikaanse Universiteit (RAU) were analysed to determine whether they were either mutants or *O. mosssambicus*

hybrids. The Red Tilapia in this case were found to be of the species *O. mosssambicus*.

Red tilapia is a hybrid produced by the inter-breeding between *Oreochromis niloticus* and *O. mossambicus* (Fitzgerald, 1979), or it could be a mutant of *O. mossambicus*. The biology and behavior of red tilapia is extremely similar to that of the common mouth breeding tilapia. Red tilapia is also omnivorous, reproductive and euryhaline, and is highly resistant to diseases. Having a glorious reddish coloration and lacking black coloration on the peritoneum, red tilapia look very similar to sea bream (*Chrysophrys major*), and are highly preferred by the consumers (Liao and Chang, 1983). It is for these reasons that red tilapias are cultured, and similarly why biologists dread that they escape into natural waters. There is, however, a large misconception whereby any tilapia that appears red, orange, gold or pink in color is termed a "red tilapia", and that they are always hybrids. Red tilapia can be obtained when two individuals from the same species, which have mutations for color, are crossed, and then backcrossing F_1 to parental individuals, or by crossing $F_1 \times F_1$.

This problem was studied by Moralee et al. (2000). The population they used was from fish living in captivity at RAU. Their conclusion for these samples was that all were in fact *O. mossambicus* (based on allozyme data), and that their morphological data corresponded to that of *O. mossambicus* too as defined by Skelton (1993). It was however mentioned that natural red tilapias usually do not survive as they are soon eradicated through natural selection by predatory birds and larger fish.

3.7 The tilapia genome

It is known that karyotypes of all the *Oreochromis* species are highly similar. It consists of 22 pairs with no morphologically distinct sex chromosomes. In fact, only 2 pairs are recognisable, with the remaining 20 being similar in size and morphology (Majumdar & McAndrew, 1986).

The first steps towards constructing a genetic map for cichlids, was taken by Lee & Kocher (1996) when they isolated 133 $(CA)_n$ and 7 $(AAC)_n$ microsatellite loci from *Oreochromis niloticus*.

The genome of several species of tilapias has been measured at around 1000 Mb (1 pg) – about one third the size of many mammalian genomes. The variation, which does occur (from 840 to 1210 Mb), is probably because of the evolution of repetitive element families. The SATA family consists of repeats ~230bp long, and represents 7% of the *O. mossambicus* genome (Wright, 1989).

A major study was undertaken to map the tilapia genome by Universities of New Hampshire and Stirling (Kocher et al., 1998). The well-known cultured *O. niloticus* was used to construct this map, which forms the basis or starting point for many molecular studies of the tilapias, such as the mapping of single loci and quantitative trait loci in cichlid fishes.

Forty one haploid embryos derived from a single female were used to investigate the segregation of 62 microsatellites and 112 amplified fragment length polymorphisms (AFLPs). A total of 162 (93,1%) linkages were found between the markers in the final map. The map spans 704 kosambi cM in 30 linkage groups covering the 22 chromosomes of this species. Twenty four of these linkage groups contain at least one microsatellite polymorphism.

4 Conservation aspects

It is reported that freshwater fishes are the most threatened of all vertebrate groups exploited by humans (Moralee et al., 2000). About 160 species are endangered and about one species per year becomes extinct. The threats include water extraction, pollution, overfishing and the impacts of exotic species (Pullin, 1997).

In order to discuss/study interspecies hybridization, it is necessary to clearly define the concept of a "species".

"Species are normally the units of biodiversity and conservation" (Wilson, 1992).

In literature various approaches can be found on this topic, based on various schools of thought, as well as the development of the concept throughout history. In a review by Mayden (1997), 22 concepts were identified at that time. The author attributes this turmoil embodied in the species problem to "packaging of inappropriate criteria for species into a single concept".

A brief overview of the most popular concepts if species is given here, but for more detailed discussions, see Mayr (1942 and 1969), Claridge *et al.* (1997) and Lèvêque (1997).

The fist approach is the ***biological*** approach, from which a species is defined as: "Reproductively isolated species are separate evolutionary entities characterized by unique specific mate recognition systems" (Claridge *et al.*, 1997), or "groups of actually or potentially interbreeding natural populations which are reproductively isolated from other such groups" (Barton & Hewitt. 1985).
Some remarks on this concept are:

1. Reproductively isolated sibling species show no clear morphological differentiation BUT are in fact reproductively isolated;
2. Morphological, cytological, behavioral, or molecular markers usually diagnose the presence of high levels of reproductive isolation;
3. This concept can only be applied to biological systems that are biparental and reproduce sexually – or at least regularly exchange genetic material;
4. It is difficult and usually subjective in dealing with populations that are isolated in space (allopatry).

These difficulties gave rise to another more broadly approach: ***phylogenetic species***, by which species are considered as "*diagnosable distinct clades*" (Claridge, *et al.,* 1997). Remarks about this concept are:

1. It is a broader concept than the biological species.
2. Allopatric and asexual populations do not challenge this concept.
3. Difficulties exist on how to decide objectively on what is a diagnosable distinct clade.

In both of these groups of concepts the possibility of ignoring sibling species exits. In our case, F_3 and F_4 are almost undistinguishable from parental species. This problem is overcome in this project by taking samples from the wild, from a geographical range as wide as possible, and also by the use of non coding nuclear genetic material markers.

It is clear that *O. mossambicus* and *O. niloticus* do hybridize successfully in the natural environment, which indicates that the reproductive isolation of what is considered as two species, is incomplete. Therefore it is decided to follow the approach of the phylogenetic concept mainly because we are dealing with populations, which were allopatric before human interference, but now occurring together in the physical environment (see point 2 above).

"There is clearly common ground between these two general concepts for describing biological diversity and together they form a unitary *taxonomic* or *evolutionary species*" (Claridge, *et al.,* 1997).

Often taxonomists are forced back to describe dead specimens, without knowing much of their habitats or habits. Thus identification of species have to be based mainly on morphological characters to identify species, posing a *theoretical* problem with hybrids.

In cases of interspecies hybridization the progeny sometimes display characters that resemble only one of the parental species or intermediate characters between parental species. These problems usually hamper the identification of hybrids and are thus an *operational* problem when dealing with hybrids.

Biodiversity

Considerable concern is expressed about the possible effects of direct competition between cultured exotic species and wild endemic species but this may be, in part, founded upon the commonly held view that biodiversity involves mainly or only a species diversity element (Beardmore et al., 1997). The consequences of aquaculture on biodiversity in aquatic systems are still imperfectly analyzed.

Various definitions of biodiversity have been formulated as yet, but one stands out: "*biodiversity forms a biospatial, hierarchically distributed structure of variability among living organisms with five levels of complexity and which, unusually, includes agricultural systems*" (Beardmore et al., 1997). These five levels refer to:

1. whole system such as ecosystems or landscapes (greater order of complexity)
2. assemblages with their habitats such as associations or communities of species
3. species
4. populations within species
5. genes within populations (lesser order of complexity)

Importance of biodiversity

High levels of genetic variability for animals in natural populations have been shown to be closely associated with characters related to fitness. The characters referred to here are amongst others reproductive rates, growth and developmental stability.

There are three basic reasons for the long-term conservation of our genetic resources. First, it is generally agreed by ecologists and evolutionary biologists that species diversity and genetic variability are necessary for the long-term maintenance of stable, complex ecosystems and species. In this regard, conservation of genebanks also have applications concerned with breeding and selection as well as pharmaceuticals, pesticides (biological-), etc..

Second, there is often local pride in populations or species that are characteristic of an area. People often become disturbed when some local form of animal is threatened by extinction, and this concern is an important reason for conservation of at least some species.

Finally, diversity or variability seems aesthetically pleasing in most environments. This is not only true in general but often applies to specific species frequently encountered by man, such as certain common and/or colourful birds.

Genetic diversity among species within ecosystems is the basis of different functional forms or niches. Niche simplification frequently leads to an increased need for energy supplements from man to maintain the system. Farmlands are classic examples of systems with reduced diversity planned by man for specific biological reasons that require considerable attention and energy for their maintenance. Most highly diverse natural systems normally require little or no energy subsidy from man for their maintenance. Genetic variability can also directly affect the quality of individuals within species through single-and multi-locus effects (Falconer, 1960; Dobzhansky, 1970). These effects are well known in applied genetics and their applications of great economic significance. Maintaining genetic variability can be important because of its potential use under a variety of environmental conditions that currently exist in the future (Beardmore et al., 1997).

Knowledge of aquatic biodiversity

Compared with terrestrial ecosystems, marine ecosystems are poorly understood, although they occupy close to three-quarters of the surface of the earth and contain the greater proportion of biological diversity, particularly in relation to the number of species. One consequence of this is that the species of marine macrofauna so far described may represent only a small proportion of extant species (Gaston & May, 1992). The oceans have the highest diversity of animals and plants with 28 phyla, freshwater ecosystems contain 14 phyla whilst terrestrial systems contain only 11 phyla (Beardmore *et al.*, 1997).

Loss of Biodiversity

It is clear that, if the definition of biodiversity illustrated as above is accepted, reductions in total biodiversity in a given area may come about through change in any one of a number of the components. There is, however, a natural tendency to focus upon species loss as the most potent index of reduced biodiversity. It should be remembered, nevertheless, that the normal fate of a given species, in the long run, is to become extinct (or evolve into another species). The species now existing on earth constitute a very small fraction (possibly as low as 10^{-3}) of all the species that have ever existed. Concern should therefore be focused upon rate of extinction rather than extinction *per se* (Beardmore, *et al.*, 1997).

Impact of aquaculture on biodiversity (Beardmore *et al.*, 1997)

Possible or actual impacts of the development of aquaculture on biodiversity may result from a variety of causes. It has been argued that these impacts are occasionally positive, sometimes neutral and usually negative. Negative impacts may be direct, e.g. by genetic introgression, or indirect (and probably more importantly) by e.g. loss of habitat (Beveridge *et al.*, 1994). Clearly therefore, on the whole, positive measures have to be taken to ensure that biodiversity is protected, so as far as possible, from such impacts which include:

1. habitat destruction to create ponds;
2. pollution of local water, by intensive production;
3. effects of antibiotics and other chemical treatments on local microfauna and macrofauna;
4. intensive collection of wild seed;
5. competition with endemic fauna by escaped exotics;
6. introduction of pathogens and parasites;
7. genetic introgression with local fauna by introduction of:
   - 7.1 populations,
   - 7.2 species,
   - 7.3 transgenics.

These impacts are more thoroughly discussed by Beardmore *et al.* (1997), but in this study, emphasis is on impact number 7 with which this study is concerned – in particular, the species impact.

Species

One example of a potential major conflict between the interests of aquaculture and those of biodiversity is posed by catfish culture in Thailand. The walking catfish, *Clarias macrocephalus* (Gunter), is widely cultured in Thailand and has desirable flesh characteristics but is slow growing. The African walking catfish, *C. gariepinus* (Burchell), has a faster growth rate but less desirable flesh characteristics. The hybrid has acceptable flesh, a growth rate intermediate between the two parent species and is now widely cultured. It is not clear what the level of risk posed in local populations of *C. macrocephalus* may be, but F1 hybrids may cross readily with the resident fish. Laboratory experiments suggest that the ability of future specific F1 hybrids to produce reliable backcross progeny is low, but nevertheless there is the possibility that the viability of populations of *C. macrocephalus* may be threatened. Similar problems may arise in Bangladesh through the use of hybrid *C. batrachus* X *C. gariepinus* in aquaculture (Rahman, 1995).

In each of these cases, a possible outcome is the development of local populations that consist of individuals of hybrid origin, with long-term threats to survival of one or more species. The hazard may be greater in cases such as these where geographically distant populations of two species are intercrossed than in cases like the brown trout, *Salmo trutta* L., and Atlantic salmon (*S. salar*). With these species, natural hybrids are found at low frequency, but there is no evidence of introgression probably because of considerable disturbances to meiosis and generation of triploids in backcross progeny. However, the same view of the desirability of use of sterile strains applies here, especially as species such as Nile tilapia are being spread geographically at a rapid rate.

Genetic variability

There are two components that builds the concept of genetic variability: allelic diversity (i.e. number of alleles at a given locus) and genetic heterozygosity (i.e. expected proportion of genes that are heterozygous in the average individual) (Bruford & Wayne, 1994). In the typical individual about 3 to 10 % of all genes are heterozygous, whereas in outbred populations, typically 10 to 30 % of genetically variable structural genes detected by protein electrophoresis have two or more alleles. Analysis of noncoding variable DNA regions, such as mini- and microsatellites reveal higher values, with greater than 50 % of loci being polymorphic, and typical heterozygosity values varying from 50 to 94 %. In small and isolated populations, genetic variation may be drastically reduced, initially by a loss of allelic variability, followed by a decline in heterozygosity. Low levels of heterozygosity, especially when associated with inbreeding, may correspond with decreases in viability and increases in juvenile mortality. Genetic variability may be important for the long-term persistence of and adaptive change in populations, and management of captive and wild populations of endangered species are designed to minimize the loss of genetic variation.

5 Methods available

"*Molecular biology was born after the discovery of the structure (and rapidly thereafter the function) of DNA in the 1950's. It was after the development in 1970's of techniques that allowed for the manipulation of genes, and the ability to study the genetic process at a depth not dreamt of only a decade earlier, that the field of molecular biology exploded with major impact on research on a wide variety of biological sciences.*" (Garte, 1994).

Identification of systematic units and the determination of taxonomic uniqueness are crucial information for conservation programmes. Given the difficulty of assigning priority to the numerous species that are rare or endangered, molecular techniques can be used to better define the distinctiveness of species and the taxonomic units they contain. Informed decisions can then be made regarding the relative significance of the species and the amount of effort which

should be devoted to its conservation. Molecular data can be analysed using phylogenetic techniques to provide information of the evolutionary heritage of a species and to determine their phylogenetic distinctiveness by ranking them according to the number of close relatives and phylogenetic position. Species which represent monotypic genera may, for example, be regarded as more significant than ones in more specious genera (Bruford & Wayne, 1994).

Genetic studies on fish started by focusing on protein level: electrophoretic techniques were used to identify variants (allozymes) at enzyme coding loci (isozymes), and their frequency were quantified in conspecific and higher taxonomic comparisons. Unfortunately, in many cases this approach suffered from a lack of sufficient levels of detectable protein variability to discriminate taxa due to the inherent sensitivity of the technique and its focus on gene products which evolve slowly. The conventional electrophoretic techniques only detect about one third of all amino acid substitutions in protein molecules, and the products of isozyme loci evolve slowly due to their critical functions in cellular activities (Waldman & Wirgin, 1994).

As techniques to isolate purified DNA were developed in the 1970's, it became possible to visualize selected DNA fragments (radio labeling / fluorescent techniques), and reproducibly cleave DNA molecules at selected, short, 4 to 8 base pair sequences by the use of restriction endonucleases. Restriction endonuclease digestion of DNA provided a rapid and highly reproducible method to screen any DNAs for differences in their base sequences at randomly distributed recognition sites. The number and size of resulting DNA fragments could be determined by electrophoretic analysis. Probes were developed that allowed for the visualization of distinct DNA segments by their hybridization to complementary sequences in the target DNA after blotting the target DNA to a solid support (Southern, 1975). These probes could be used to visualize DNA sequences at coding genes of any region of interest in the nuclear or mitochondrial genomes.

The advent of the polymerase chain reaction (PCR - Mullis, *et al.* 1986) allowed for the direct and rapid replication of sequences of target DNA producing an exponential growth in all the scientific fields related to molecular genetics. With these tools, the stage was set to characterize DNA sequences in individuals and compare the frequencies of variant genotypes among individuals, populations, species, and higher taxonomic comparisons. (Waldman & Wirgin, 1994).

Amongst others, DNA-based markers provide two major advantages to phenotypic features as tools to use in discriminating methods. These are: (a) that DNA sequences are heritable and, in the short term, free from environmental influences thus offering long term stability without the need for frequent "recalibration" of the marker; and (b) different DNA sequences evolve across a wide range of rates. Therefore DNA sequence analysis offers the potential to quantify genetic relationships at differing levels of taxonomic divergence, from the individual to interspecific level.

Microsatellites

Variation in microsatellite repeat number is found to have pronounced effects on their regulatory and coding functions, and is associated with phenotypic variation at both the biochemical and organismic levels. The combination of high mutation rate and regulatory function raises the possibility that microsatellites are a major source of eukaryotic genetic variation (Kashi & Soller, 1999). But these functional roles as coding and regulatory elements are only explored in late years (Künzler *et al.*, 1995) - micro- and minisatellites are most generally considered in terms of their roles as genetic markers for studies in population genetics, evolutionary relationships, and gene mapping.

Microsatellites (or short tandem repeats – STR's, or simple sequence repeats – SSR's) have repeat units of only 2 to 5 bp. STR and VNTR (variable number tandem repeats) allelic numbers are given by the number of repeats. These markers can therefore be used for identification as well as associations with traits

of interest. For minisatellites[5] (Jeffreys, *et al.* 1985) the repeat unit is of the order of 10 to 60 bp and the number of units can be in the thousands (Weir, 1996).

Advantages of microsatellites are (a) high allelic number, (b) high mutation rate – 10^3-10^5 mutations per gamete, which is three times higher than for mitochondrial DNA (Bowcock, *et al.*, 1994) and (c) minimal amounts of DNA is required.

Using this marker system, firstly, one excludes the main disadvantages of coding regions, eliminating the chances of environmental (external) effects on the part of the genotype/individual being studied. Secondly, the resolution/sensitivity of these markers is suitable for this study in seeking differences between two species. The dinucleotide repeat $[CA]_n$ is one of the most abundant microsatellite families in vertebrate genomes occuring on average every 15 to 30 kb (Stallings, *et al.*,1991 and Estoup *et al.*, 1993).

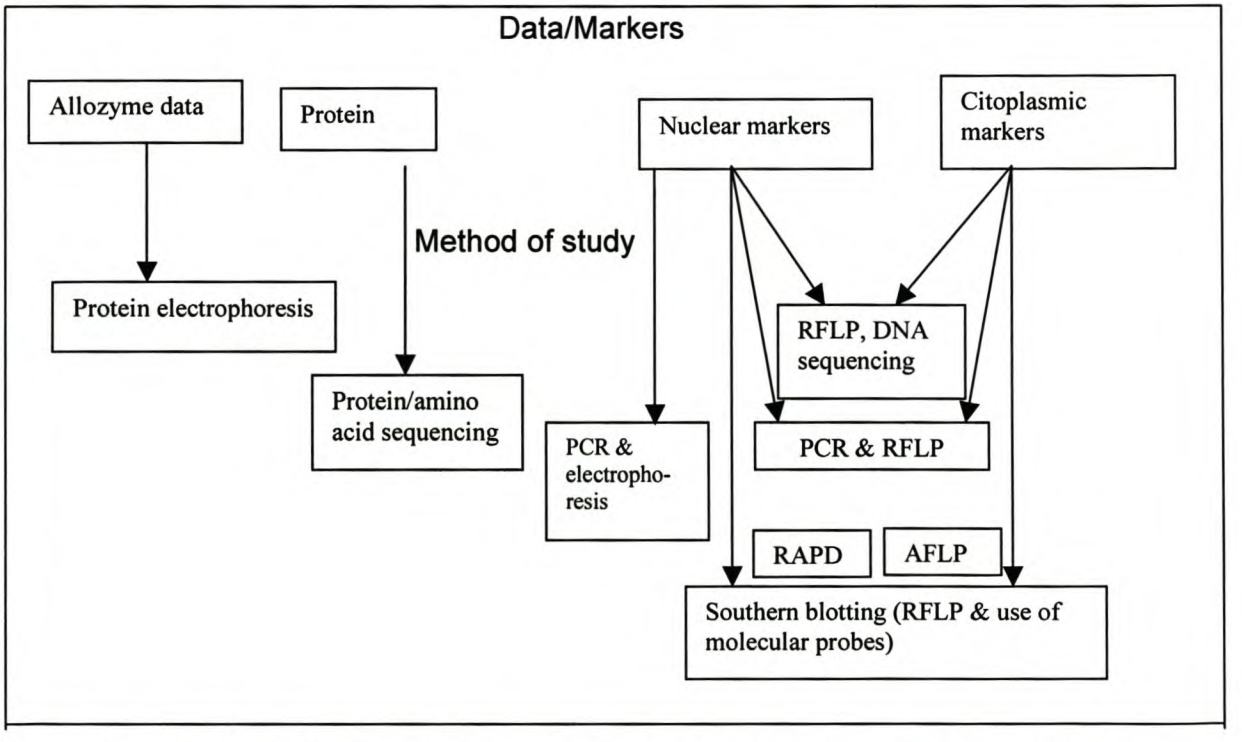

Figure 1.4 Schematic view of methods of study available for different types of data/markers

[5] or Variable number of tandem repeats – VNTR's

When designing and carrying out experiments, certain factors should be kept in mind, which compose a sound experimental design. These are: (a) adequate number of independent replicates, (b) simultaneous and appropriate controls and (c) the number of observations should be large enough so that the likelihood of statistical anomalies is low (Hedrick, 2000)

Statistical analysis of microsatellite data

(1) Genetic Variation

It was only when Mendelian genetics and Darwin's theory of natural selection were synthesized into the neo-Darwinian theory of evolution in the 1930's, that a substantial effort was initiated to document the amount of genetic variation within and between populations (Hedrick, 2000).

The amount and kind of genetic variation in a population are potentially affected by a number of factors, such as selection, inbreeding, genetic drift, gene flow and mutation. They may have general effects; one may increase genetic variation whilst another decreases it. Several factors in combination may though generate any amount or pattern of genetic variation.

Measures of Genetic Variation

To measure genetic variation, there are two ways necessary: (a) an unbiased manner, and (b) some way to measure selected effects of different genotypes, such as homozygotes and heterozygotes (Hedrick, 2000). The measure of genetic variation of most extended use are heterozygosity, number of alleles per locus and proportion of polymorphic loci.

(a) Proportion of polymorphic loci

A practical approach to defining polymorphism is to decide arbitrarily on a limit for the frequency of the most common allele; that is, polymorphic loci are those for which the frequency of the most common allele is smaller than 0.99, or smaller than 0.95. Both of these arbitrary cutoff points have been used, but 0.99 is used most frequently if the sample size is adequate (approximately 100 individuals or

more) (Hedrick, 2000). To estimate the proportion of polymorphic loci (P) for a population where a number of loci have been examined, one must first count the polymorphic loci and then calculate the proportion that these loci represent of all the loci examined. In other words, the proportion of polymorphic loci is :

$$P=x/m$$

where x is the number of polymorphic loci in a sample of m loci.

(b) Heterozygosity

The amount of heterozygosity[6] is the most widespread measure of genetic variation in a population and the second popular means to measure genetic variation within and between populations. In general, heterozygosity indicates how many individuals in a population are heterozygous at a certain locus. Therefore heterozygosity can be mathematically indicated as:

$$H = y/n$$

where H is the proportion of heterozygotes, y the number of individuals which are heterozygous, and n the number of individuals tested at the particular locus.

Heterozygosity should though be further discussed, because if a population is in Hardy-Weinberg equilibrium, certain heterozygosity values are expected to be evident for that population. But populations are not always in Hardy-Weinberg equilibrium (in fact with so many different factors affecting genetic variation in different ways, most often populations are *not* in Hardy-Weinberg equilibrium). One can thus calculate two different heterozygosity values – one for observed heterozygosity (H_O) and one for expected heterozygosity (H_E). If these two values differ for a certain population at a certain locus, the phenomena is called *heterozygote deficiency*, but this will be discussed later.

When data from various loci is available, heterozygosity over all loci can be calculated as an average over individual heterozygosity values at each locus individually.

[6] Proportion of individuals that are heterozygous in the sample

When more than one population is considered, heterozygosity over all populations can also be measured. This value will thus indicate the proportion of heterozygotes in the whole group of individuals examined irrespective of the nature of the alleles at the locus. By this I mean that whether in one population (if we consider variable numbers of repeat units at a certain locus) the alleles are of size 102, 106, 112 and 120 bp and in another population alleles are of size 124, 128, 134 and 140, this difference will not be indicated. In this case only genetic variation with regards to what proportion of individuals has two different alleles will be evident. Therefore it depends on what the question of the study is, whether heterozygosity values can be calculated over all populations or perhaps only over few groups of populations.

Another remark on heterozygosity is that sampling bias may play a role in calculating observed heterozygosity values. This may not be the case in many studies, and may be eliminated in particular by planning of sampling (e.g. large enough sample size). In cases where an endangered species in studied and the individual have to be killed, large sample sizes are just not possible – that's if conserving the species is a priority For these cases Nei (1978) developed a mathematical correction for this value of heterozygosity.

First it is necessary to clearly define observed and expected heterozygosity. When a population is in HWE the allele frequencies in the population are expected to be in the following proportions:

$$p^2 + 2pq + q^2 = 1$$

where p is the frequency of one allele and q the frequency of the second allele (in the case of only two alleles present). The expected heterozygosity for such a population at such a locus will then be defined as

$$H = 1 - \Sigma p_i^2$$

where H is the heterozygosity (expected) and p_i is the frequency of the ith allele at a locus. The correction Nei (1978) suggested is the following:

$$h = 2n(1-\Sigma x_i^2)/(2n-1)$$

(for a single locus)

where *n* is the number of individuals per locus. The correction, when *r* loci are concerned, can still be made but here the number of individuals per locus is considered independently because it may differ from locus to locus. Thus the corrected expected heterozygosity for *r* loci would be:

$$H = \Sigma h_k / r \quad \text{(the summation is for k=1 to } r\text{)}$$

And h_k is the value for *h* for the *k*th locus (Nei, 1978).

The observed heterozygosity should be irrespective of the HWE, and can be simply described as

$$H_O = y/n$$

where *y* is the number of individuals which are heterozygous, and n the number of individual tested at the particular locus. Because individuals in diploid species are either heterozygous or homozygous at a given locus, this measure represents a biologically useful quantity. However, the theoretical properties of the distribution of heterozygosity are complicated, and measures of heterozygosity are not very sensitive to additional variation because the upper limit, unity, is the same for any number of alleles. This limit makes it difficult to differentiate between populations for highly variable loci, such as microsatellites, where the heterozygosity may be 0.8 or higher.

For allozymes, microsatellites, or other diploid loci detected using molecular techniques, one can obtain simultaneously information concerning the heterozygosity of a number of loci in many individuals in a population. For a given locus in a particular individual, there is either a heterozygous or a homozygous state.

The sampling variance has two components: that due to variation in heterozygosity among individuals and that due to variation in heterozygosity among loci. These values may be quite different.

Nei (1987) has suggested that, given the choice of examining more loci or more individuals, it is best to examine more loci because variation among loci is generally high. In fact, if only a few individuals are available, as in a rare species, then one can obtain a fairly good estimate of heterozygosity by

examining a few individuals for a number of loci. However, it is doubtful whether the interlocus variation will ever be reduced very much because of the real heterozygosity differences among loci (not just the result of sampling), resulting in extreme polymorphism for some loci and monomorphism for others.

Similar but less extreme effects may also be seen for highly variable diploid loci, such as microsatellites. One should be aware that when utilizing microsatellite loci, researchers often do not include invariant or low-variation loci, and they may even select for loci with high heterozygosity (Paetkau, et al., 1997) or a specific range of heterozygosity (Dib, et al., 1996) to obtain loci that are most useful for the purpose of their study.

In gene mapping or in application to diagnosis, the presence of a segregating null allele will not corrupt the linkage data but may cause loss of information (Callen, *et al.* 1993; Paetkau & Strobeck, 1995; Pemberton, *et al.*, 1995). In data sets such as in this study, null alleles may cause heterozygote deficiencies, which in turn may be interpreted wrongly (may be ascribed to be due to selection, inbreeding or population subdivision) prompting wrong conclusions.

Heterozygote deficiency

When the observed and expected heterozygosity (whether considering either biased or unbiased expected heterozygosity) differ, the phenomena is called heterozygote deficiency indicating that the population(s) is(are) not in HWE.

This can be due to various different factors. These factors can be put into two different groups: methodological and biological factors.

With methodological factors I mean factors that are due to human activities concerning the study such as sampling bias. The researcher could take too small amounts of individuals from the population (or populations from a species) which do not represent the population's alleles properly, or even lacking some alleles totally. This is called small sampling bias. Another factor is the presence of null alleles. In this case during electrophoresis, a certain band do not get

scored because the band could perhaps run off the gel into the buffer because the focus is on bands of a particular size, and no such a small band was expected.

Biological factors comprise of factors that are based on the biological aspects of the populations. These deficiencies can be due to: null alleles, Walhund effect arising from population subdivision; inbreeding due non random mating and selection (Foltz, 1986).

(i) Null alleles

In gene mapping or in application to diagnosis, the presence of a segregating null allele will not corrupt the linkage data but may cause loss of information. In microsatellite studies null alleles may cause heterozygote deficiencies, which in turn may be interpreted wrongly (may be ascribed to be due to selection, inbreeding or population subdivision), prompting wrong conclusions. In microsatellite studies, null alleles may be evident because of abnormally high levels of homozygosity in only one or a very few loci with respect to all scored loci. It is simply logic that null alleles will give a lower indication of genetic variation than what it should be because one is missing alleles which should be part of the study. There is a statistical correction to handle this problem. The frequency of null alleles can be estimated as:

$$p_n=(H_E-H_O)/(1+H_E)$$

where the recessive homozygote (individual that do not display any PCR product and are detected as blank gels) is not present (Brookfield, 1996).

(ii) Wahlund effect

When two populations which are in Hardy-Weinberg genotypic proportions, but with different allele frequencies, are pooled together, the pooled population will have an allele frequency equal to the average of those in the two populations. The effects of population pooling must be considered as potential complications to genetic analysis, because unrecognized subdivision in samples may be evident in samples from virtually all natural populations. The first generation of random mating in the pooled population will result in Hardy-Weinberg proportions

of genotypes, but the frequency of homozygotes in the two unmixed subpopulations. Reduction in homozygosity in this way, is called the Walhund principle (Hartl & Clark, 1989).

Hedrick (2000) mentions that it can sometimes be difficult to determine whether a deficiency of heteroygotes is the result of inbreeding or of the Walhund effect. The heterozygote frequency at all loci should be affected by inbreeding, whereas only the heterozygote frequency at those loci with allelic frequency variation over subpopulations should be reduced by the Wahlund effect. When there are multiple alleles, all heterozygotes should be reduced in frequency by inbreeding, whereas some may be decreased and others may remain unaffected or be increased by the Wahlund effect (Hedrick, 2000).

(iii) Inbreeding

Inbreeding refers to matings between related individuals, and the primary effect of inbreeding is departure from Hardy-Weinberg genotype frequencies toward an excess of homozygotes. The amount of inbreeding can be quantified in a population. "In the absence of any other evolutionary factors, inbreeding alone does not change allele frequencies – it changes only the assembly of genes into genotypes" (Hartl & Clark, 1989)

However, if natural selection is also operating, then inbreeding can have a profound effect on the course of evolution. Similarly, inbreeding and random genetic drift results in an increased likelihood of bringing together genes from a common ancestor due to loss of genetic variation.

(iv) Selection

Selection can be described as "the tendency of the bearers of particular genotypes to reproduce more of less than others in the population, thereby systematically altering allele frequencies (Goldstein & Schlötterer, 1999). Selection on microsatellite loci has been assumed to be absent, except when the locus occurs in a region affecting gene regulation (Carrington, *et al.*, 1999).

(c) Allele number

Another measure that is sometimes used is the number of actual alleles, n, a count of the alleles observed at a locus in a population. However, this measure is often strongly influenced by the sample size, so comparison across population with different sample sizes should be made cautiously. It is a measure of higher importance in microsatellite markers when nonspecific primers are used to amplify homologous locus to the focal species.

F-statistics

Heterozygote deficiency can be partitioned into a within and among population component, and F-statistics provide a set of tools for this partitioning. At first it was devised by Wright (1921 & 1969) but during the last 30 years estimation of F-statistics has been debated in literature since the early work of Cockerham (1969, 1973) and Nei (1973 & 1975) and two "families" of estimators are derived – those of Weir & Cockerham (1984) and that of Nei (1987).

F-statistics is used to assess levels of structuring in samples of natural populations and are defined as:

F_{IS} = (mean $H_S - H_I$)/mean H_S

= inbreeding coefficient

= measures reduction in heterozygosity of an individual due to nonrandom mating within its subpopulation

and H_S is the expected Hardy-Weinberg heterozygosity in subpopulations.

Inbreeding refers to matings between related individuals, and the primary effect of inbreeding is departure from Hardy-Weinberg genotype frequencies toward an excess of homozygotes. The amount of inbreeding can be quantified in a population. "In the absence of any other evolutionary factors, inbreeding alone does not change allele frequencies – it changes only the assembly of genes into genotypes" (Hartl & Clark, 1989).

The interpretation of F in terms of probability is this: the inbreeding coefficient F is the probability that the two alleles of a gene in an individual are identical by descent (autozygous) (Hartl & Clark, 1989)

$0 < F < 1$ with 0 = outbred, 1 = inbred

Even thought the inbreeding coefficient is defined as the probability of an individual having a pair of alleles that are identical by descent, there are two uses of the inbreeding coefficient that are subtly distinct. When the inbreeding coefficient is used in models, such as models that predict genotype frequencies, then it is a parameter, and is assumed to be known without error. On the other hand, when F is estimated from population heterozygosity or from a pedigree, then it is a statistical estimator, and as such represents a description of an observation.

F_{ST} = $(H_T - \text{mean } H_S)/H_T$
- = fixation index
- = measures the effects of population subdivision
- = reduction in heterozygosity of a subpopulation due to random genetic drift
- ≥ 0, because the Walhund effect assures that $H_T >$ mean H_S. Thus if all subpopulations are in HWE with the same allele frequencies, then $F_{ST}=0$

F_{IT} = $(H_T-H_I)/H_T$
- = overall inbreeding coefficient for an individual
- = contributors: nonrandom mating within subpopulations (F_{IS}) and subdivision itself (F_{ST})

Cockerham uses some other symbols to describe his F-statistics' values: inbreeding coefficient (F_{IS})=f; total inbreeding coefficient (F_{IT})=F and coancestry coefficient (F_{ST})=θ.

F-statistics become more complicated with multiple alleles and multiple genes. For solving this problem, Weir & Cockerham (1984) introduced bootstrapping.

This principle involves resampling the data set over alleles and genes to obtain an empirical distribution of the F-statistics using subsamples of the complete data set.

Population structure

Differences in genetic variation among constituent parts of species is considered as structure in the population, and is the result of several different evolutionary factors (Hedrick, 2000). Two models are used to elucidate population structure: the *continent-island* and *general* models. These general models may not precisely fit a particular biological example, but they give close approximations to many situations and enable us to evaluate the effect of limited gene flow. For thorough discussions on this topic see Hartl & Clark (1997) and Hedrick (2000). Here it is only important that population structure is taken into account when processing data.

Principal components analysis

Principal component analysis (PCA) is a type of multivariate analysis dealing with data of which the measurements are often correlated and thus making conclusions much more complex. With such data it is frequently useful to attempt to find uncorrelated composite measures by the method of Principal components before attempting any analyses such as clustering, etc., since it is often found that the data may be expressed in terms of far fewer than that of the composite measures without any significant loss of information.

PCA thus leads to a set of *p* composite characters that are uncorrelated and are arranged in order of decreasing variance. If it is found that the first few principal components account for most of the variation, it might be possible to use only these in subsequent analyses and thus achieve a considerable simplification. Mathematically, the first principal component of the observations is that linear combination y_1, of the original variables,

$$y_1=a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p$$

where a and x are the original variables.

The usefulness of this artificial variate constructed from the observed characters is under question, and the answer lies in the proportion of the total variance attributable to it. If 87% of the variation in an investigation in six characters could be accounted for by a simple weighted average of the character values, it would appear that almost all of the variation could be expressed along a single continuum rather than in six-dimensional space (Dunn & Everitt, 1982).

It should be mentioned that the method of principal components is not independent of the scales(s) of the original measurements. Multiplying one of the variables by a constant (for example, by altering the scale from meters to centimeters) will change the convariance matrix and produce a different set of principal components. It should also be remembered that where the original characters are measured in widely different units, linear combinations of them will have no sensible physical dimensions.

Consequently, the analysis is often carried out on standardized measurements and the components extracted from the correlation rather than the covariance matrix. Examples of situation where no standardization is needed, however, include those where all measurements are proportions (e.g. gene frequencies) of where they are all logarithms or lengths. The effect of the logarithmic transformation in the latter case is to give measurements with the same proportional variability the same variance, so that measurements that are relatively more variable will have higher variance and will be given more weight in the subsequent analysis.

Genetic distance

Different methods to measure genetic distance between populations were developed by various researchers. Basically there are two groups of these distances: those which use stepwise mutation models and those using the infinite allele model.

It is very important to keep in mind when considering genetic distances, that it only is a tool to consolidate the data into manageable proportions and aid one in visualising general relationships among the groups. Although some information is lost when arrays of frequency data are reduced to a single value, patterns among populations obscured by the mass of numbers may be made apparent by utilising genetic distance values. Distance measures are generally analogous to geometric distance; i.e. zero distance is equivalent to no difference between the groups.

Distance measures are based on measures of genetic variation (like H or allele frequencies) and therefore care should be taken to use the correct groups of populations in cases where different species or different types (wild vs. farm) animal populations are considered.

In this study, we use two groups of populations belonging to two different species, which evolved one to two million years ago (Trewavas, 1983). When considering microsatellite markers, one needs to keep in mind that alleles differ

because of differences in numbers of repeat units. Thus one cannot follow the infinite allele approach, because the genome cannot have extended infinitely. (the amount of bases which can be transcribed cannot be infinitely long for then a stop and start codon for transcription is needed which will break up the microsatellite region of repeats) It is though documented that the number of alleles do not increase at microsatellite loci after about 50 000 years – which is much less than the amount of years passed from speciation of the two species under investigation in this study. Thus both species obtained their maximum amount of alleles already.

The method of Cavalli-Sforza & Edwards (1967) holds true for SMM, and will thus be used in this study. Other methods are those of Nei (1972 and 1978) and Weir & Cockerham (1984) but it follows the infinite allele model and it states that no differential selection is evident. Last mentioned methods will not be discussed here, but I refer to Hartl & Clark (1989) and Hedrick (2000) for discussions of these.

The type, amount and pattern of genetic variation between populations can be the result of various factors which cause either similarities or differences between populations. These factors are: (a) the time from separation until now – recently separated populations are normally genetically very similar (b) gene flow may of may not occur – populations between which gene flow occurs are genetically similar (c) size of population – in large populations the amount of genetic drift is little and therefore are more similar to each other, (d) selection pressures – if the same selection pressures affects the two populations, they would be very similar. (stochastic effects). Unfortunately, if these factors affected the populations long ago in historical times, it is very difficult to detect their relative roles.

The first step to measure genetic distances, is to see whether the allelic frequencies are significantly different. This is done by using the χ^2 test for heterogeneity over populations where for two alleles:

$$\chi^2=2NV(p)/\text{mean } pq$$

where N is the total sample size, V is the weighted variance, p is and estimate of the average allelic frequency of allele A_1 and $q = 1 - p$.

A number of genetic similarity and distance measures have been proposed and used to evaluate the amount of variation shared among groups. These measures help to consolidate data into manageable proportions and aid one in visualizing general relationships among the groups. Although some information is lost when arrays of frequency data are reduced to a single value, patterns among populations occurred by the mass of numbers may be made apparent by utilizing genetic distance values. Distance measures are generally analogous to geometric distance; that is, zero distance is equivalent to no difference between the groups.

Similarities or differences in the type, amount, and pattern of genetic variation between populations can be the result of many factors. Genetically similar populations may perhaps appear as such because they (1) recently separated, or (2) geneflow occurs between them, or (3) they were large populations and genetic drift has not been long enough to produce differential accumulation of genetic variants between locations (with little genetic drift), or similar selection pressures affected loci similarly in both populations. In the same way, if two populations are very different from each other, it could be because (1) they have been isolated for a long time and there has been no gene flow between them and genetic drift has generated large differences, or (2) there are different selective pressures in the two populations, or (3) stochastic historical factors like population crashes or founder effects. Often all of these factors stated above may be important in a particular situation. Furthermore, if these forces have been important in the past, it is often (not always) difficult to reconstruct historically their relative roles. For a proper discussion on these factors, see Hedrick (2000).

A number of measures of genetic distance have been suggested over the past several decades. In practice, many of these measures are highly correlated, particularly when the differences between the populations are small, even though

the measures are often based on different biological or mathematical assumptions. However, when the differences become larger, then there are often substantial differences between genetic distance measures on the same data set. This is particularly true when comparing genetic distance measures developed for microsatellite loci that assume particular modes of mutation and generally weight genetic distance by the square of the difference in the number of repeats for different alleles (Hedrick, 2000).

Cluster analysis

Various methods of grouping the variables according to the magnitudes and interrelationships among their correlations have been developed. These methods are known generally as *cluster analysis*. An introduction to cluster analysis is given by Everitt (1980).

Ways of performing cluster analysis are to construct distance methods (e.g. Neigbor–Joining (Saitou & Nei, 1987) and UPGMA (unweighted-pair-group method) (Sokal & Sneath, 1963), parsimony methods and maximum likelihood methods. The first methods construct trees from a distance matrix containing measures of genetic distance between all the pairs of taxa under examination. Only distance methods are used in this study.

Whereas cluster analysis groups variables that are highly correlated with each other and excludes from a cluster those that are not, *factor analysis,* this is another type of multivariate analysis, aims to express covariation in terms of *k* underlying factors that explain a large part of the variance and covariance of the original variables. In this model, the variables are considered linear combinations of the underlying factors. The number of factors considered is usually much lower than the number of variables in the study.

Multivariate analysis

In multivariate analysis the technique of principle axes is important, where instead of ellipses (see Sokal & Rohlf, 1995) we encounter clouds of observations describing hyper ellipsoids in a multidimensional space. To simplify the description of these clouds of points, we calculate principle axes through the hyper ellipsoids. An important property that we have not yet emphasized is that the eigen values, which represent the variance along the principle axes, are such that as we determine successive principle axes representing the major axis, the second major axis, and so forth of the hyper ellipsoid, we are successively finding the orthogonal axis that accounts for the greatest, second greatest, and successively smaller amounts of variation. This technique is called principal component analysis, a branch of multivariate analysis (Dunn & Everitt, 1982).

Assignment of individuals to populations

With this method, a model is assumed in which there are K populations, each of which is characterized by a set of allele frequencies at each locus. Individuals in the sample are assigned to populations, or jointly to two or more populations of their genotypes indicate that they are admixed. It is assumed that within populations, the loci are at Hardy-Weinberg equilibrium, and linkage equilibrium. This is very important to keep in mind when interpreting data.

Chapter 2 Materials and Methods

1. Introduction

The screening of the genotype of tilapia populations requires a number of consecutive procedures. First, non-destructive sampling of taking finclips allowed to return fish to the environment and to keep broodstock (section 2.2). The samples were preserved in 99% ethanol and taken to the laboratory for further work (section 2.3). DNA was then extracted from the samples for use in amplification experiments with twenty microsatellite primer pairs to amplify loci of CA repeats. Initially amplified products were separated on vertical polyacrylamide gels to check which of the twenty primer sets would be suitable in further experiments. This represents the "preliminary study" (section 2.4). Five microsatellite primer sets were then chosen for genotyping all populations (section 2.5). One oligonucleotide of each of these primer sets was fluorescently labelled and PCR products were run on sequencing gels, using the automated ABI 3100 DNA analyser. The results were statistically evaluated with appropriate software programmes to calculate measures of genetic variation and differentiation (section 2.5).

2. Collection of samples

O. mossambicus samples were taken by several collectors at different locations and dates. Five samples were taken in its natural distribution range; one sample was taken by Ben van der Waal[1] in 2000, and the samples Kasinthula, Le Pommier, Ndumu and the Bushmans were taken by Edward Hall[2] in 1999 and 2000.

Samples from the Limpopo river were caught with a 50 m long by 3 m seine net in large (2-10 ha) irrigation ponds on the farm Samaria, bordering the Limpopo River, (owner H. Heyns) on 31 August 2000. Some fish were also collected with a 17 m long by 2 m deep seine net in a drainage channel outside these ponds. Both ponds and the drainage channel had had a surface connection with the

[1] Department of Biological Sciences, University of Venda, South Africa
[2] Aquaculture division, University of Stellenbosch, South Africa

Limpopo River during the extremely high flood of February 2000. Additional *O. mossambicus* fish were collected with the smaller seine net in a small farm pond of Ben van der Waal on the farm Rondebosch near Louis Trichardt. These fish originated from the Albasini Dam, Luvuvhu River, also Limpopo System.

The Limpopo (*O. mossambicus*) sample (*n*=21) was taken from a wild population in the Limpopo river, which forms the Northern boarder of Verre Noord province in South Africa. The Boesmans sample (*n*=10) was taken from the Boesmans river in the Southeast of South Africa. Ndumu is a site at the Pongola/Usuthu river systems in Northern KwaZulu Natal, South Africa (*n*=9). The Le Pommier sample (*n*=7) was taken from the Le Pommier farm in the Western Cape, South Africa, and the population was introduced in the 1940's. The Kasinthula sample (*n*=14) is taken at a site in the Shire river (Malawi) nearest to the village called Kasinthula.

For the group of hybrids we had only one sample from a population, which we could say for sure to be hybrids between the two species under investigation. This sample of 7 individuals was taken from a wild population in the Limpopo river by Ben van der Waal in 2000.

O. niloticus was sampled at 4 sites; 20 individuals from a farm population in Thailand were collected by Graham Mair[3] in 1999. Another sample of 20 individuals was taken by Henk Stander[4] from a research farm in the Phillipines in 2001 where experiments are conducted to obtain YY males. The 10 males from this sample were YY males. A sample of 21 wild fish was taken by M.A. Hussein[5] in Egypt in 2001 and another wild sample of 16 by Mahmoud A. Rezk[6] at another site in Egypt, which forms part of its natural distribution range.

[3] DIFID, University of Stirling, UK
[4] Division of Aquaculture, University of Stellenbosch, RSA
[5] Central Laboratory for Aquaculture Research, Agricultural Research Center, Egypt
[6] ICLARM, Egypt

Table 2.1 Samples collected for this study. N refers to the number of individuals in that sample; date refers to the year in which the sample was taken. * indicates information not available.

Species		Sample					
Scientific name	Abbr.	Location	Drainage.	Co-ordinates	Wild/farm population	N	Date
Oreochromis mossambicus	MWLI	Samaria/ Levuvhu	Limpopo river	22° 16' S 29° 16' E	Wild	21	2000
	MWBO	*	Boesmans river	33° 30' S 26° 30' E	Wild	10	1999
	MWND	Ndumu	Pongola/Usuthu river	26° 45' S 32° E	Wild	9	1999
	MILE	Le Pommier	*	34° S 18° 45' E	Introduced	7	1999
	MWKA	Kasinthula	Shire river	17° S 35° 30' E	Wild	14	1999
Hybrids	HWLI	Samaria/ Levuvhu	Limpopo river	22° 16' S 29° 16' E	Wild	7	2000
Oreochromis niloticus	NFTH	Thailand	*	*	Farm	20	1999
	NFPH	Philippines	*	*	Farm	20	2001
	NWE1	Egypt	Nile river	*	Wild	21	2001
	NWE2	Egypt	Nile river	*	Wild	16	2001

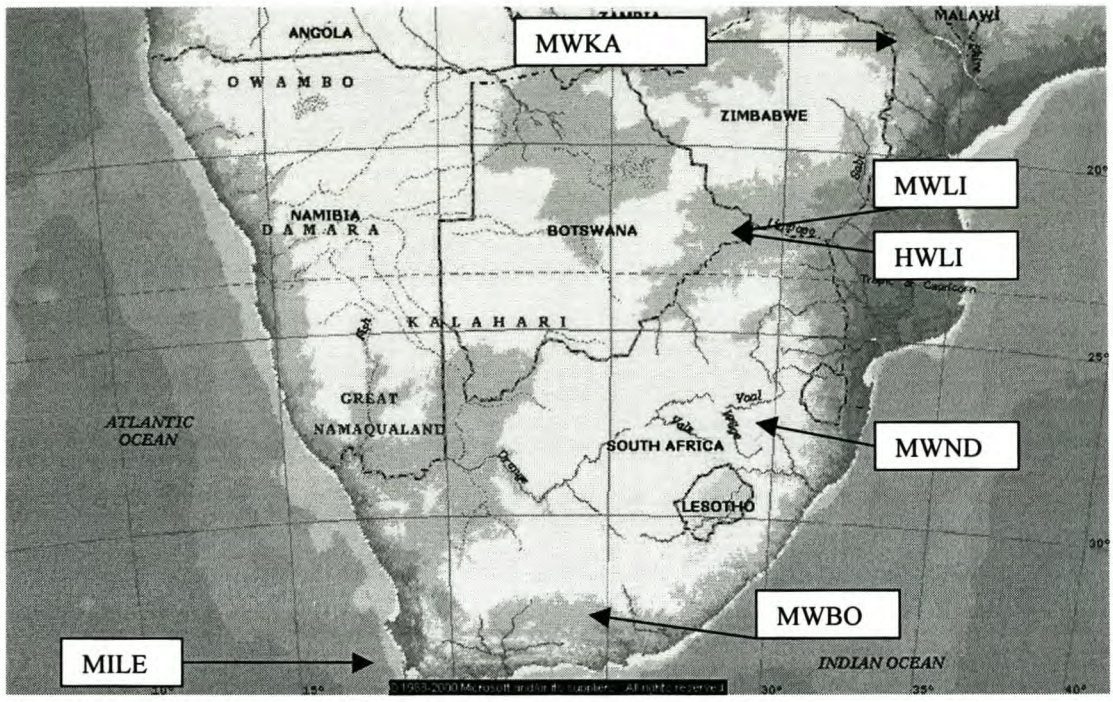

Figure 2.1 Map of collection locations for 5 populations of *O. mossambicus* and 1 population of hybrids in Southern Africa. Map is constructed with Encarta®©.

3. Sample processing

Tissue collection

Sample tubes were prepared for fieldwork in the lab by putting 1 ml 95% ethanol into 1,5 ml numbered microcentrifuge tubes.

The fish collected by the seine nets were provisionally identified as either *O. mossambicus, O. niloticus* or a hybrid. An additional distinction was made in the case of hypothetical hybrids in that they were grouped as hybrids showing more *O. mossambicus or O. niloticus* characteristics. Representative fishes of these groups were photographed and the locality recorded[7]. A piece of about 1 cm² was removed from the anal fin and placed in a prepared sample tube.

As for the other wild samples, detailed methods were not available. These fish were caught with nets in the drainage systems and finclips of about 1 cm² was also removed from the anal fin and placed in prepared sample tubes.

[7] Information at Ben van der Waal

Samples from farm locations were taken when fish were slaughtered for marketing purposes by taking about 1 cm^2 of tissue from the anal fin and placing it in a prepared sample tube

All samples were stored at room temperature and protected from UV light until laboratory analysis.

DNA extractions[8]

Grinded tissue was digested in 340 µl of lysis buffer (400 mM Tris-HCl pH8.0, 60 mM EDTA, 150 mM NaCl, 1% SDS) for 10 min. One hundred microlitres of 5M Sodium Perchlorate was added to each sample and shaken vigorously. Incubation occurred at 37°C for 20 minutes while being shaken vigorously every 3 to 5 minutes, and then again for 20 minutes at 65°C. After chloroform extraction the DNA was precipitated with isoporopanol, and washed with cold 70% ethanol and again with 99% ethanol. Pellets were resuspended in 55 µl dH_2O and stored at 4°C.

4. Microsatellite screening

Twenty microsatellite loci isolated on *O. niloticus* by Kocher *et al.* (1998) were screened for amplificatability, the number of alleles, the absence of null alleles and distinctness between *O. niloticus* and *O. mossambicus*.

PCR reactions for the preliminary study contained 1X (total concentration) reaction buffer, 0.2 mM of each nucleotide, 0.025 U/µl Taq DNA polymerase, 1.5 or 2.0 mM $MgCl_2$ (see table 2.2), 1 µl of both primers (stock 100 µM) to which is added ~25 ng/µl of genomic DNA.

Cycling conditions: Reactions were run for 5 minutes at 94°C, and 45 cycles for 20 s at 94°C, 30 s at the optimal annealing temperature and 45 s at 72°C. A final 7 min was spent at 72°C. Primers tested are listed in Table 2.2.

[8] Nucleon® extraction booklet

Table 2.2 PCR primer sequences, optimised PCR temperatures and $MgCl_2$ concentrations, product lengths of five loci used in final study for *Oreochromis niloticus* and *O. mossambicus*. For the primers, F = forward; R = reverse. T_A refers to the optimal annealing temperature; [$MgCl_2$] is the concentration of $MgCl_2$.

Primer	Primer sequence (5'-3')	T_A (°C)	[$MgCl_2$]	Product length[9]
UNH102F	AAA TGA TAC ATG ACT GCT TA	53	1.5	149-165
UNH102R	TTA GGA CTT ATC TGT CTA CAA GC			
UNH104F	GCA GTT ATT TGT GGT CAC TA	55	1.5	
UNH104R	GGT ATA TGT CTA ACT GAA ATC C			
UNH106F	CCT TCA GCA TCC GTA TAT	55	1.5	
UNH106R	GTC TCT TTC TCT CTG TCA CAA G			
UNH108F	GGG ATC AGC TGT TAA GTT T	58-62 touchd	1.5	
UNH108R	TGA GTT GAT TAT TAA TTT CTG A			
UNH111F	TGC TGT TCT TAT TTT CGC	56	1.5	
UNH111R	ATA AGA GTG TAT GCA TTA CTG G			
UNH115F	ACC TTC ATC TCG GTC AG	58	1.5	
UNH115R	TCA AGC AGC TGA TTT TT			
UNH123F	CAT CAT CAC AGA CAG ATT AGA	58	1.5	
UNH123R	GAT TGA GAT TTC ATT CAA G			
UNH129F	AGA AGT CGT GCA TCT CTC	50	2.0	172-214
UNH129R	TGT ACA TCA TCT GTG GG			
UNH146F	CCA CTC TGCCTG CCC TCT AT	53	1.5	118-132
UNH146R	AGC TGC GTC AAA CTC TCA AAA G			
UNH160F	CCA TTG GCT CTT ACA TC	50	2.0	
UNH160R	GAT AGC ATT TCT GTA GTT ATG G			
UNH124F	AATTTGGCAGCTTCTTTT	50	2.0	265-315
UNH124R	CCCACAAGCATAGTAAACT			
UNH135F	TATGTGTGTGAAGGCTTTT	58	1.5	
UNH135R	CTCTGACTATATGTCTATAGCTGG			
UNH142F	CTTTACGTTGACGCAGT	34	1.5	
UNH142R	GTGACATGCAGCAGATA			
UNH173F	CGTGAGAAAACAATGGT	50	2.0	
UNH173R	TATTGATTTTATAGCTGTCTGG			
UNH188F	ATTTAGACAGGGGTGGAGTTTCAA	50	2.0	

[9] Exact product length only available for primer sets used in final study

UNH188R	AAGTGCTGGAGAGTTCTGCTGGAC			
UNH191F	ACACACTCCAGACTGTG	50	2.0	
UNH191R	TTGAGATGAGCTGAGTG			
UNH192F	GGAAATCCATAAGATCAGTTA	50	2.0	128-170
UNH192R	CTTTTTCAGGATTTACTGCTAAG			
UNH204F	ACAAGACTGTTGAACTGTGAT	50	2.0	
UNH204R	TCTCCACTTCATAGCGTTTAT			
UNH205F	TATATTTAGACTGGGTGACTGA	50	2.0	
UNH205R	CTTACAATGAGATCATCCC			
UNH216F	GGGAAACTAAAGCTGAAATA	50	2.0	
UNH216R	TGCAAGGAATATCAGCA			

Electrophoresis

(a) Polyacrylamide gels

To test PCR reactions, products were loaded into polyacrylamide gels to separate the amplified fragments. Five percent polyacrylamide (19:1 acrylamide:bis acrylamide) gels were used – either on small-sized (82 x 100 mm) or medium-sized (200 x 220 mm) rigs. 3,6 µl DNA is loaded per well and 3 µl of a 100 bp ladder (Promega®) was used. Each gel was run for 45 min (82 x 100 mm) or 200 min (200 x 220 mm) at 200V or 300V at 4°C.

(b) Silver staining

Manual silver staining was used to stain amplified DNA fragments. Immediately after electrophoresis, gels were taken from the plates, and fixed with 100 ml of 10% ethanol and 0.5% Acetic Acid for 15 min using a nutator to spread the solution over gels continuously. Then gels were rinsed twice for 1 min each time in de-ionised distilled water (Millipore®). Silver staining was done with 50 ml (per gel) of 0.1% silver nitrate solution for 12 min at room temperature. This solution is washed quickly (5 s) off the gels by using de-ionised distilled water (Millipore®).

The gels were then developed using 1.5% NaOH and 0.15% formaldehyde (chilled to 4°C) for 20 min. The developing solution was rinsed using de-ionised

distilled water (Millipore®) and sealed in clear plastic[10] (normal transparency film was used).

(c) Ethidiumbromide staining

Alternatively ethidiumbromide was used. After this initial test of heterologous primers, the primers for further analysis were selected on the basis of consistency of amplification, sharpness of bands and absence of spurious (extra) bands other than expected. Thus Ethidiumbromide was used (6 µl ethidiumbromide in 200 ml of TBE buffer) for staining (20 min) after the gels were run. Images were viewed on the Geldoc (Biorad®) system and pictures taken were saved and printed.

5. Microsatellite genotyping

Forward primers priming for loci UNH102, UNH124, UNH129, UNH146 and UNH192 selected for genotyping of *O. mossambicus* and *O. niloticus*. The fluorecently labelled forward primers were used in the main study along with unlabelled reverse primers of the loci to amplify fragments at these loci by PCR. The high cost of performing a microsatellite study with labelled primers was taken into account in deciding which label to use for which oligonucleotide. The forward primers of those two loci whose fragment lengths were expected to differ by more than 150 bp, were labelled with the same dye. Forward primers of locus UNH102 and UNH129 were labelled with R100 (blue), those of loci UNH124 and UNH192 were labelled with Tamra (yellow) and that for locus UNH146 was labelled with R6G (green). The same PCR conditions were applied as during the preliminary study (see Table 2.2).

PCR products of primers labelled with fluorescent dyes were electrophoresed on an automated ABI 3100 sequencer. Samples were multiplexed into the same lane in the sequence gels (no samples were multiplexed in the PCR reactions). The locus and individual of each sample was electronically recorded

[10] I made enlarged photocopies of the gels when sealed to do measurements and calculations properly. These copies were filed after use for future reference.

on a computer connected with the ABI 3100 when a laser lens (operating at 39 mW) detected the fluorescence emitted from the PCR products.

Data was analysed using Genescan® analysis 3.7 and Genotyper®2.1. This software displays the outputs as a set of different coloured peaks with intensity and peak size in base pairs for each individual separately.

Allele sizes were then documented into spreadsheets using Microsoft Excel® for each individual at each locus, from which input files were created for statistical analysis programmes as needed. The designation of peaks for each individual was complicated by the presence of stutter bands and non-specific PCR products. The criteria to assign alleles to samples are discussed in Chapter 4.

6. Statistical data analysis

Statistical analyses to evaluate genetic diversity, structure and assignment were conducted using GENEPOP version 3.3 (Raymond & Rousset, 1995), GENETIX version 4.01 (Belkhir *et al.*, 1999), FSTAT version 2.9.1 (Goudet, 2000), PHYLIP version 3.57c (Felsenstein, 1989), PCAgen version 1.2 (Goudet, 1999), RSTcalc (Goodman, 1997) and STRUCTURE (Pritchard *et al.*, 2000).

6.1 Genetic diversity

(a) Allele frequencies

For each of the five loci, allele frequencies were calculated with GENEPOP. Sigmaplot 2000® for Windows (version 6.0) was used to present the results graphically with bubble plots.

(b) Polymorphism

The level of polymorphism was calculated for each population by species at the 95% and 99% confidence level using GENEPOP.

(c) Heterozygosity

Observed and expected heterozygosities were calculated by GENEPOP. Some sample sizes were very small (as little as 7 individual per sample) thus we used

the non-biased expected heterozygosity of Nei (1978) (see Chapter 1, section 1.5). One-tailed probabilities of departure form Hardy-Weinberg equilibrium (HWE) were estimated within each population for each locus. Because testing for HWE in each population involves a large number of tests, the critical probability for each test was adjusted using the sequential Bonferroni procedure (Rice, 1989) to maintain an overall significance level of 0.05.

Heterozygosity values were calculated within each species separately as well as the total heterozygosity over all loci. Significance in pairwise comparisons was evaluated after a sequential Bonferroni adjustment of critical probabilities (Rice, 1989).

6.2 Population structure

(a) Linkage disequilibrium

Genotype linkage equilibrium among the entire set of pairwise population samples were analysed by Fisher's exact test, using GENEPOP V. 3.1a (Raymond & Rousset, 1995).

(b) F-statistics

(i) Inbreeding coefficient

The inbreeding coefficient was calculated per locus and population as well as over all loci (multilocus) and over populations per species (multipopulation) according to Weir & Cockerham (1984) using Fstat and Genetix. Probability values were calculated by GENEPOP to indicate the probability of error when rejecting the null hypothesis (HWE) as well as standard error values for the estimates. To obtain an unbiased estimation of exact Hardy-Weinberg probability (statistical significance of average and pairwise F_{ST}) and because more than four alleles were observed, calculations were performed using the Monte Carlo method as described by Guo & Tompson (1992) after 1000 dememorization steps and 20 batches of 1000 iterations per batch.

In cases were null alleles were suspected, frequency of null alleles were estimated as:

$$p_n=(H_E-H_O)/(1+H_E)$$

where the recessive homozygote is not present (Hedrick, 2000).

(ii) Fixation index

Various fixation indices were calculated by locus using methods of Weir & Cockerham (1984), Rousset (1996), Robertson & Hill (1984) and Raufaste & Banhomme (2000). Multilocus estimates of F_{ST} were obtained by averaging over variance components (Weir & Cockerham, 1984) and the significance of these estimates was determined by Fisher's method (Manly, 1985 and Raymond & Rousset, 1995). In addition to testing for differences in allele frequency, we tested for differentiation among populations in the distribution of allele lengths using Goodman's estimate of R_{ST} (Goodman, 1997). We estimated R_{ST} over all populations and for each pair of populations using Rstcalc (Goodman, 1997) and evaluated statistical significance with a permutation test with 1000 permutations. Multilocus estimates were obtained in the same way as for F_{ST}. For pairwise comparisons between populations over all loci, the method of Weir & Cockerham (1984) was used.

(iii) Overall inbreeding coefficient for an individual

The method of Weir & Cockerham (1984) was used to calculate multiallelic F_{IT} values per locus for all populations as well as multilocus and multipopulation values. Multipopulation values were calculated for species separately.

(c) Genetic distance

Nei's method of calculating genetic distance was used for a single locus with *n* alleles (Nei, 1972). The distance between two populations is in this way defined as

$$D = -\ln(I)$$

Where the definition of $I = J_{xy} / (J_xJ_y)^{1/2}$ with J indicating the sum of all allele frequencies of the x and/or y populations. These distances were then used to construct distance matrices in order to perform cluster analysis. Pairwise

genetic distance values were calculated between all populations to investigate relationships in allele frequencies according to Cavalli-Sforza & Edwards (1967).

The Neighbor-Joining method of clustering (Nei and Saitou, 1987) was implemented. The programme NEIGHBOR (Felsenstein, 1995) constructs a tree by successive clustering of lineages, setting branch lengths as the lineages join. The tree is not rearranged thereafter. The tree does not assume an evolutionary clock, so that it is in effect an unrooted tree. NEIGHBOR is well suited to bootstrapping studies and to analysis of very large trees. Trees were made and bootstrapped (1000 times) for cluster analysis using the programme PHYLIP (Felsenstein, 1989).

Multidimentional scaling was performed with S-PLUS (Mathsoft, Inc.).

(d) PCA

Because it is not known whether the data are correlated or not, it was important to attempt to find uncorrelated composite measures by the method of principal components before attempting any analyses such as clustering. Since it is often found that the data may be expressed in terms of far fewer than the expected amount of composite measures without any significant loss of information (Dunn & Everitt, 1982).

No standardization of measurements was needed since allele frequencies are proportional values.

This analysis was done using the programme PCAgen (Goudet, 1999) that is developed specifically for principal component analysis of microsatellites.

6.2 Assignment of individuals to populations

A model-based clustering method for inferring population structure using genotype data consisting of unlinked markers is implemented by the programme Structure (Pritchard, 2000). We used this programme to assign individuals to populations and identify migrants and admixed individuals.

Chapter 3 Results

Preliminary study

During the preliminary study, twenty microsatellite markers were tested in one population of *O. mossambicus* and *O. niloticus* each. These markers were the microsatellite loci UNH102, UNH104, UNH105, UNH106, UNH108, UNH111, UNH115, UNH123, UNH124, UNH129, UNH135, UNH142, UNH146, UNH160, UNH173, UNH188, UNH191, UNH192 and UNH196.

It was found that loci UNH104, UNH105, UNH106, UNH108, UNH111, UNH115, UNH123, UNH160, UNH135, UNH142, UNH173, UNH188, UNH191, and UNH196 were either monomorphic or non-specifically amplifying.

Five loci were found to be informative and were suspected to be useful as species specific markers. They were selected on the basis of their apparent polymorphism when run on 8% polyacrylamide gels. These loci are: UNH102, UNH124, UNH129, UNH146 and UNH192. These markers showed least differences within species and most constant differences between the populations of the two species used. They were also consistent during PCR amplification.

Main study

1. Microsatellite genotyping

Genotypes collected from the automated ABI Genotyper® were displayed as a set of coloured peaks with intensity and peak size in base pairs for each individual. These results were interpreted and studied to compile individual genotypes. Figures 3.1a, b and c show some of these peaks. 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. 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 (see Goldstein & Schlötterer, 1999). The two peaks with the highest intensity were chosen as the true peaks. The smaller to the two peaks (in bp size) was always higher in intensity than the larger peak; if not it was considered as 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 or uneven[1] whole number.

Whether even or uneven numbers were to be used, was decided based on the majority of peak sizes being even (or uneven) numbers. The assumption was made that a locus fragment may have started with an adenine (A) and after this follows a certain number of CA repeats. For peak intensity of less than 80 intensity units, peaks were again considered as carry over from next door wells because photo-reactions are used to detect fragments. The UNH146 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 adenylation of PCR products. Adenylated peaks were easy to recognise, as they were always one base pair bigger than the preceding true peaks (Brownstein *et al.* 1996).

The UNH102 locus had a non-specific PCR product of 141 base pairs present in the majority of individuals. This peak was ignored, since scoring it would not have altered interpretation of species specific data analysis.

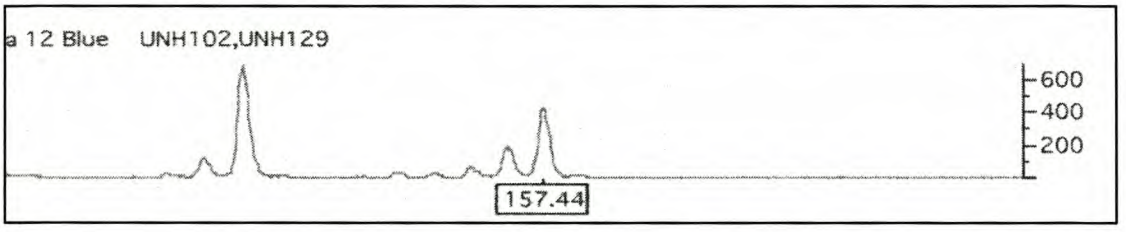

Figure 3.1a Picture of results from ABI Genotyper® for individual 142 (population NWE2 from Egypt) at locus UNH102. The picture shows the non-

[1] Depending on which locus, even or uneven numbers were used. For loci UNH102, UNH124 and UNH129 uneven numbers were used; for loci UNH 146 and UNH192 even numbers were used.

specific amplification of allele 141 bp at the left and the true scored allele 157 bp (157.44 which is rounded down to 157 bp).

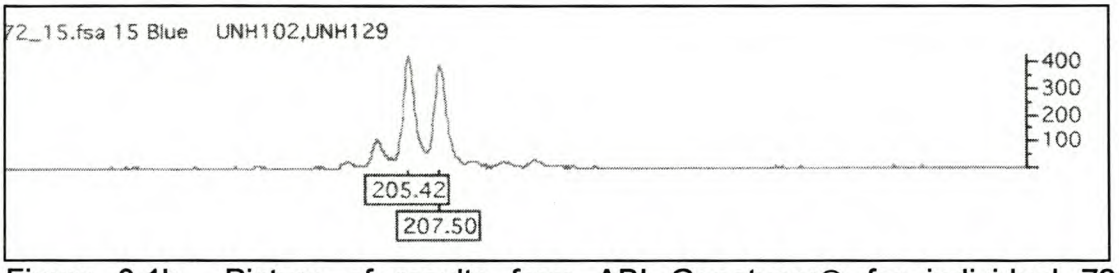

Figure 3.1b Picture of results from ABI Genotyper® for individual 72 (population NFTH from Thailand) at locus UNH129. This result was interpreted to be a case where an adenine is added to the 3' end of the PCR product.

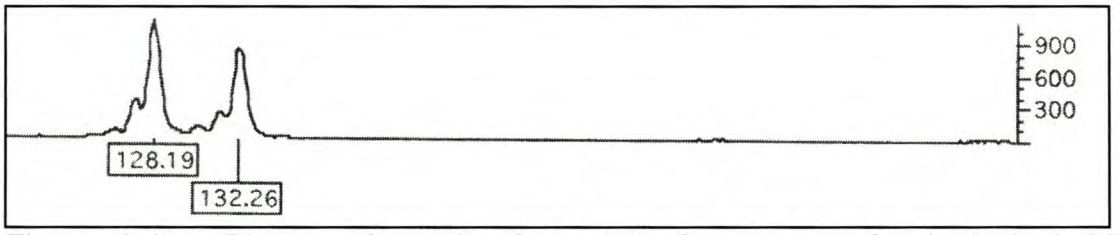

Figure 3.1c Picture of results from ABI Genotyper® for individual 26 (population MWBO from the Boesmans drainage in South Africa) at locus UNH192. This result was interpreted to be a heterozygote with alleles 128 and 132.

Genotypes for every individual at each locus are available in Appendix A. The graphic output of data is available for insight from Mr D. Brink, Division of Aquaculture, Department of Genetics, University of Stellenbosch, South Africa.

All five of the loci used in the final study were composed of repeat regions of pure CA_n. These loci varied considerably in their allele numbers (Table 3.1).

2 Statistical data analysis

Genotypes for five microsatellite loci were determined for 145 adult *O. mossambicus*, *O. niloticus* and their hybrids sampled from nine locations covering most of the *O. mossambicus* range with no distinction among year

classes. A total of 40 alleles were observed across the five loci ranging from 6 alleles in UNH129 to 11 in UNH124.

2.1 Genetic diversity

2.1.1 Number of alleles and allele frequencies

The total number of alleles over all populations ranged from 6 to 11 per locus. For the locus with the fewest alleles (UNH129) there was only one population (i.e. Nil A) that contained almost all alleles (5 out of the possible 6). The majority of alleles at the loci tested, occurred at a frequency of <5%. *Within* the *O. niloticus* group, the same alleles were common in all populations, but *between* taxa (*O. niloticus* vs. *O mossambicus*) common alleles varied.

Five populations were approaching fixation[2] in at least one locus. These were three populations from the *O. niloticus* group, the single hybrid population, and one population from the *O. mossambicus* group. The number of loci ranged from 1 in MWND, HWLI, NFPH and NWE2 to 2 in the NFTH population. At loci UNH124 and UNH146 no alleles approached fixation, whereas at locus UNH102, the allele approaching fixation was the same allele, and occurred only in the three populations from the *O. niloticus* group in high frequencies.

Allele frequencies of all populations at all loci are illustrated by figures 3.2a to e, and values for allele frequencies are shown in Appendix B.

[2] Here defined as major allele with allele frequency of ≥ 0.9

Table 3.1 Number of alleles. O.moss refers to *Oreochromis mossambicus*, O.nil refers to *O. niloticus*, and Hyb refers to a population of hybrids between *O. mossambicus* and *O. niloticus*. Total N is the number of individuals in the populations that were included in the reactions – all of which did not necessarily amplify. Values in brackets indicate the number of individuals per sample amplified. Common alleles refer to the number of alleles with a total frequency of at least 5%. Numbers in brackets after these values indicate the total number of alleles over all populations.

	Population	Total N	Number of alleles					Mean number of alleles
			UNH102	UNH124	UNH129	UNH146	UNH192	
O.moss	1	21	4 (21)	4 (18)	3 (21)	3 (21)	5 (20)	3.8
	2	10	4 (10)	4 (10)	2 (8)	2 (10)	3 (10)	3.0
	3	9	2 (7)	3 (9)	2 (8)	3 (9)	1 (9)	2.2
	4	7	4 (4)	3 (5)	2 (6)	3 (7)	4 (7)	3.2
	5	14	6 (13)	4 (12)	3 (12)	3 (12)	3 (13)	3.8
	Total		**6 (55)**	**6 (54)**	**3 (55)**	**3 (59)**	**8 (59)**	**5.2**
Hyb	6	7	2 (4)	5 (4)	2 (5)	2 (5)	4 (7)	3.0
O.nil	7	20	2 (17)	5 (18)	2 (17)	4 (17)	5 (20)	3.6
	8	20	1 (10)	2 (11)	2 (19)	2 (20)	3 (20)	2.0
	9	21	4 (13)	7 (12)	4 (16)	5 (15)	5 (17)	5.0
	10	16	1 (15)	7 (16)	5 (16)	4 (16)	5 (16)	4.4
	Total		**4 (55)**	**11 (57)**	**6 (68)**	**7 (68)**	**8 (73)**	**10**
Common alleles			4 (7)	8 (11)	4 (6)	4 (7)	7 (9)	

2.1.2 Allele frequencies

The predominance of alleles at various loci differed between the two groups of populations tested (fig. 3.2a to j). In UNH102 the *O. niloticus* group of populations was dominated by one allele (allele 157 bp, see Appendix A), which occurred only in three of the five *O. mossambicus* populations, in frequencies of 0.36 or less. The *O. mossambicus* group had no allele of 159 bp.

At locus UNH124, no particular allele dominated any of the two groups of populations. All other frequencies were less than 0.48. Alleles 203, 295, 313 and 317 did not occur at all in the *O. mossambicus* populations.

At locus UNH129 all populations of both groups were dominated by one allele – allele 207 bp; alleles 195, 209 and 215 only occurred in *O. niloticus*. As for locus

UNH146, alleles 118bp and 120bp were completely absent for *O. mossambicus* as well as the hybrids.

For locus UNH192 allele in the *O. mossambicus* group, two alleles (alleles 128 and 130) dominated clearly, although frequencies are spread over all alleles but 128 in the *O. niloticus* group.

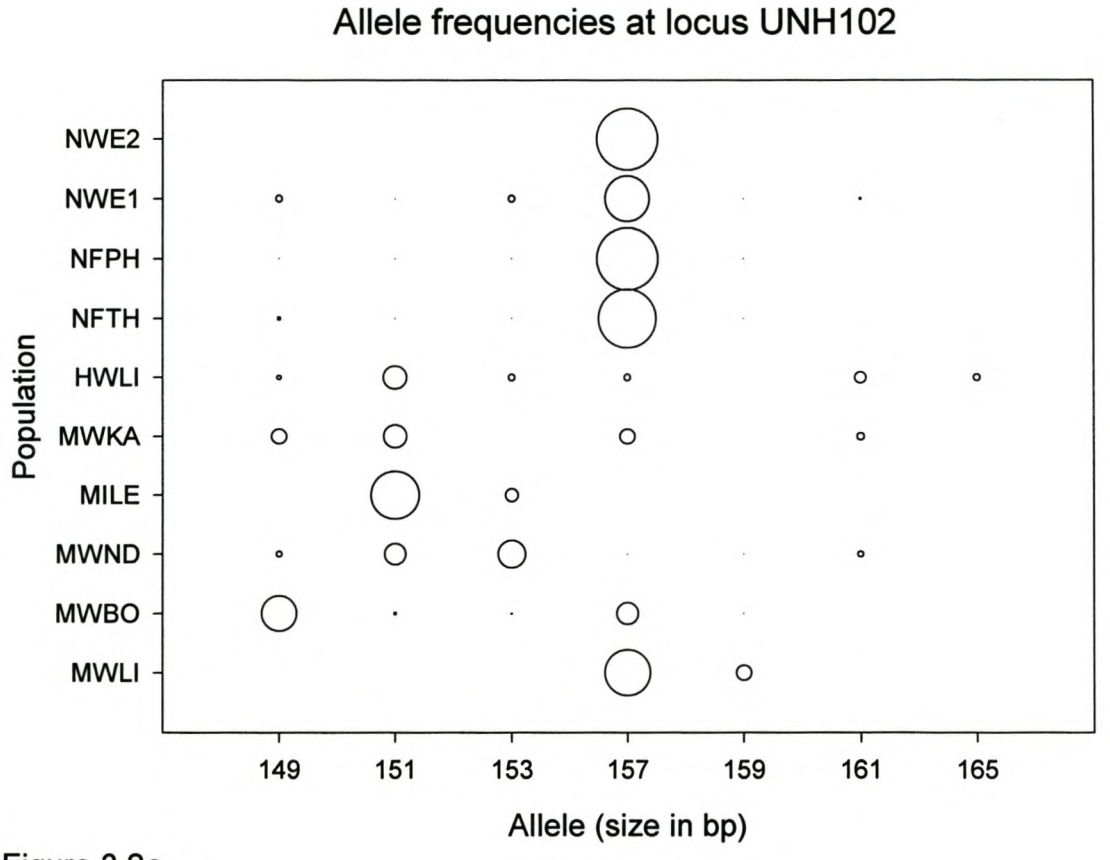

Figure 3.2a

Figure 3.2 Allele frequencies of five microsatellite loci in 10 populations of *Oreochromis mossambicus*, *O. niloticus* and their hybrids. Area size of bubbles is proportional to the respective allele frequency (ten times). See Appendix B for detail on values of frequencies of all alleles. Graphs were produced by Sigmaplot® 2000.

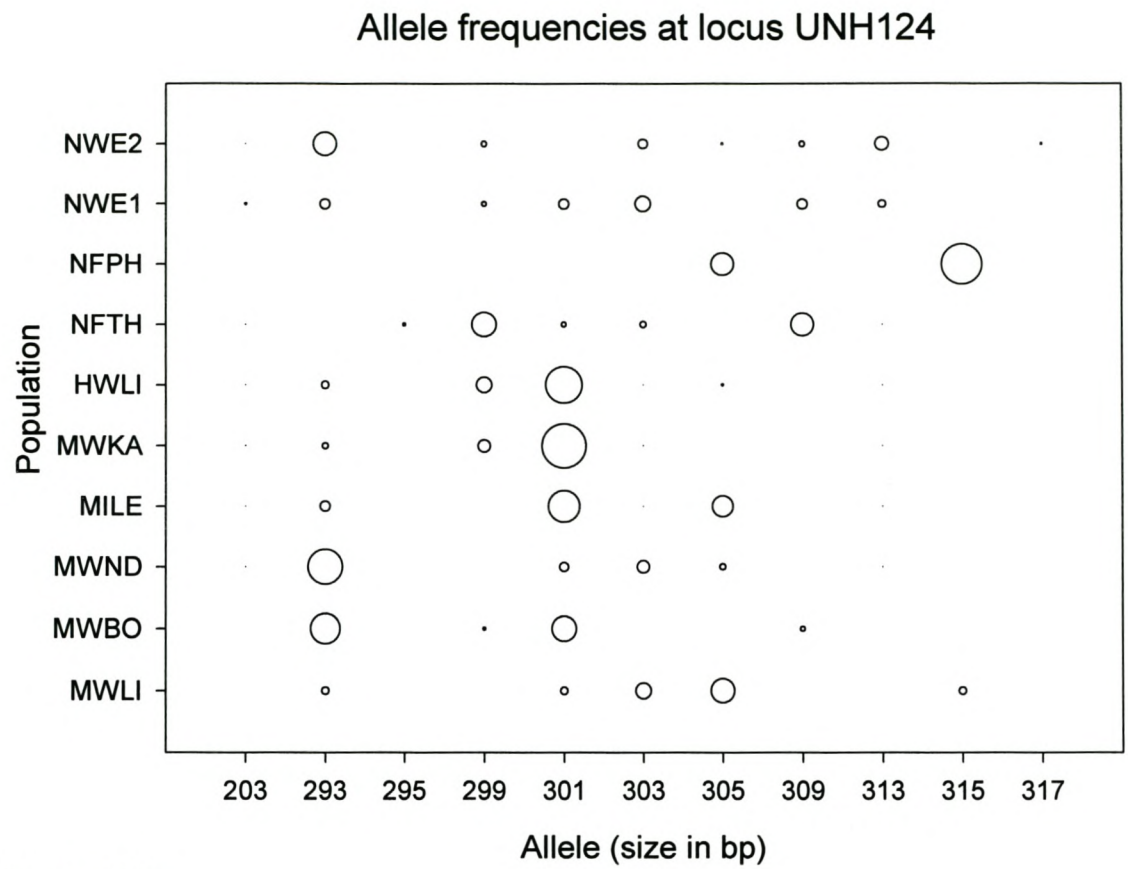

Figure 3.2b

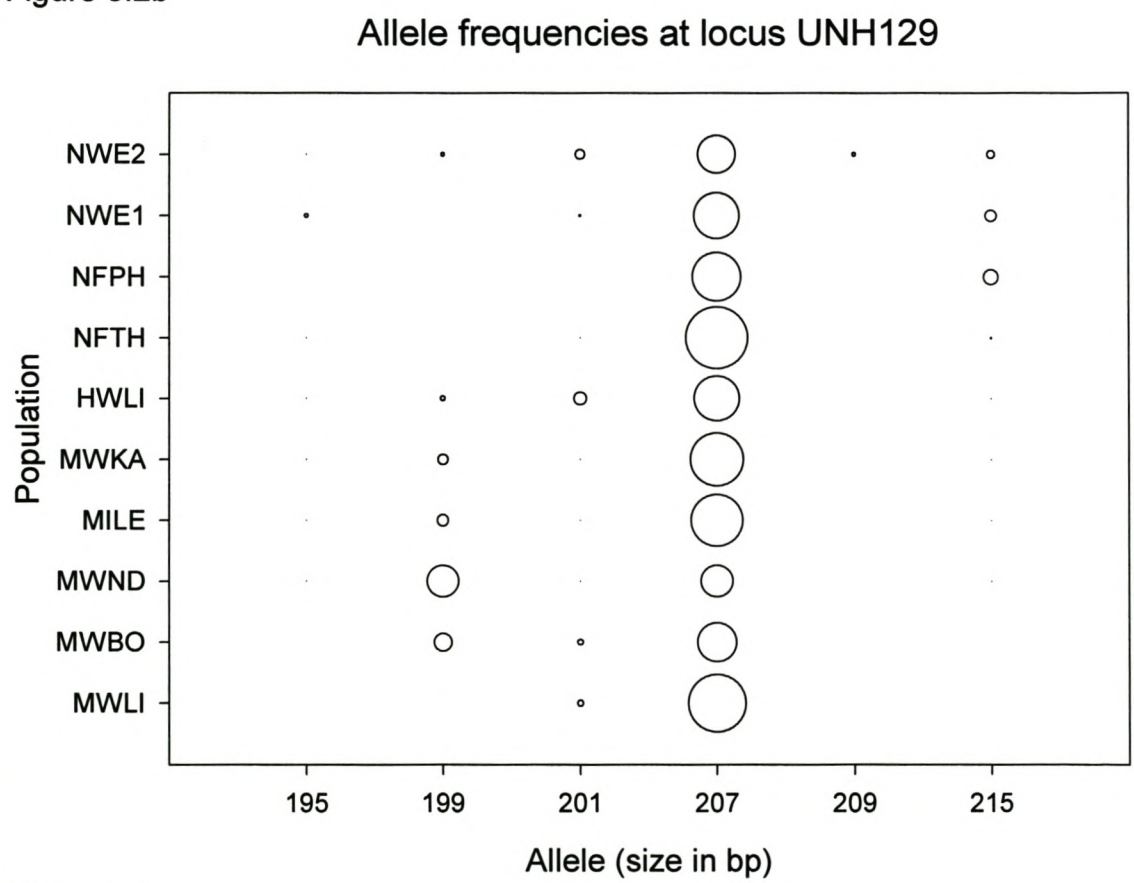

Figure 3.2c

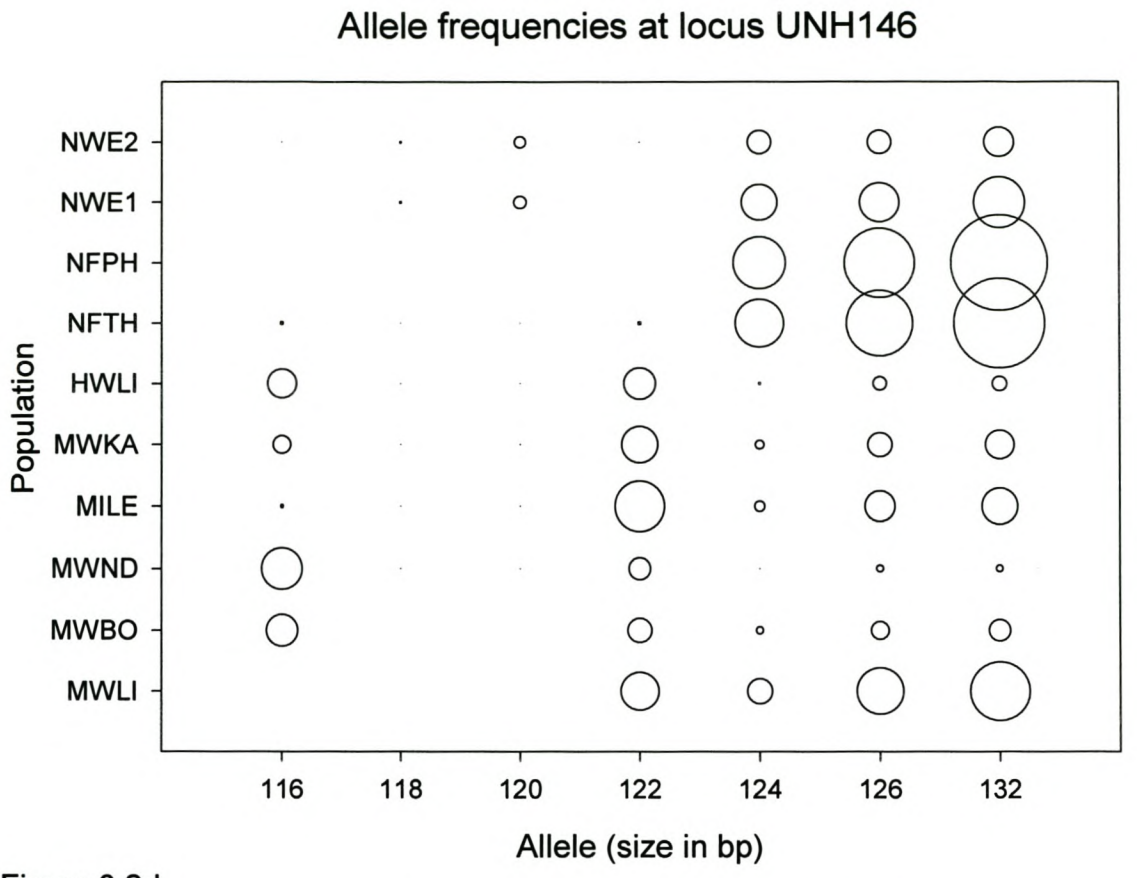

Figure 3.2d

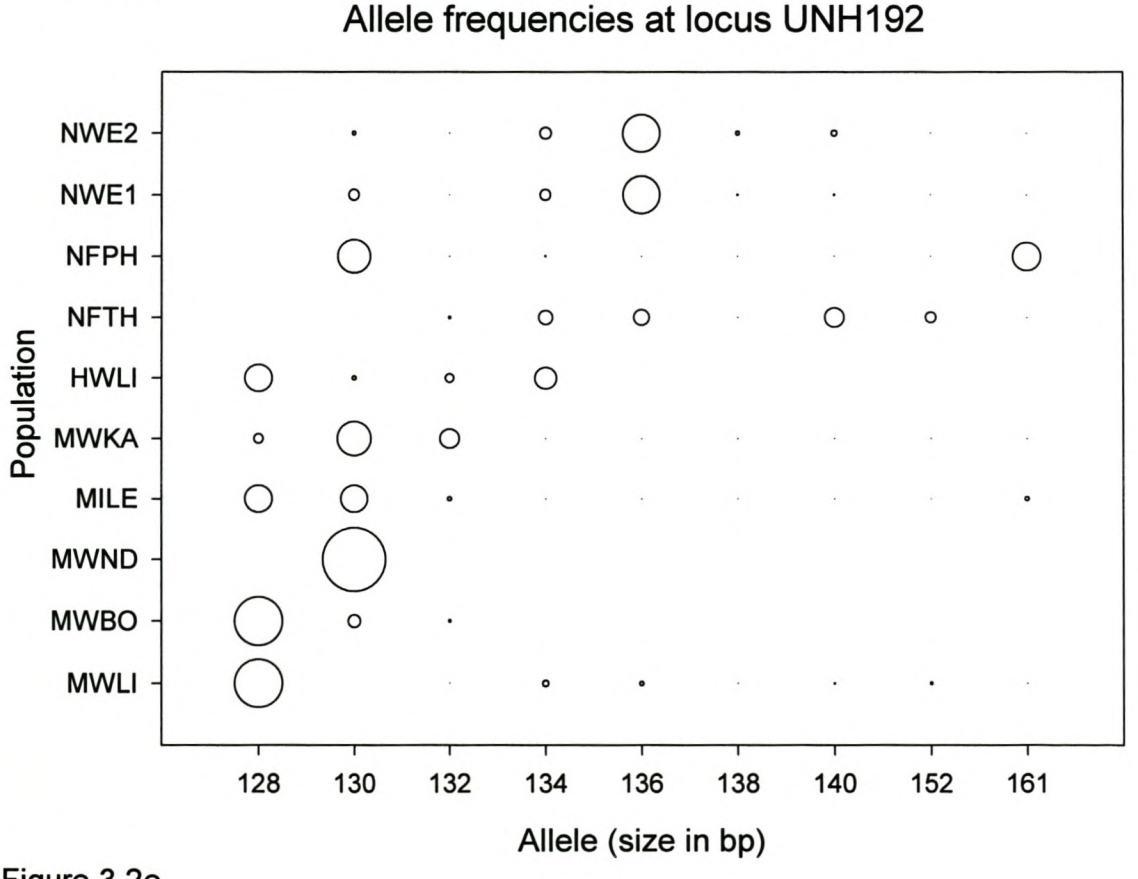

Figure 3.2e

2.1.3 Gene diversity

Estimates of gene diversity are given in Table 3.2 per locus and sample using an unbiased estimator as described by Nei (1987). It takes into account the sample size, allele frequencies and observed heterozygosities.

The mean gene diversity values for four of the five *O. mossambicus* populations and two of the four populations of the *O. niloticus* were high (>0.5). The population of hybrids shows an intermediate value for the mean gene diversity over all loci.

Table 3.2 Gene diversity per locus and population.

Locus	MWLI	MWBO	MWND	MILE	MWKA	HWLI	NFTH	NFPH	NWE1	NWE2
UNH102	0.558	0.689	0.357	0.917	0.808	0.500	0.114	0.000	0.458	0.000
UNH129	0.545	0.518	0.339	0.300	0.470	0.200	0.059	0.368	0.454	0.629
UNH146	0.601	0.467	0.382	0.595	0.549	0.550	0.404	0.295	0.629	0.690
UNH124	0.639	0.683	0.625	0.550	0.610	0.875	0.717	0.500	0.867	0.800
UNH192	0.436	0.433	0.000	0.714	0.625	0.726	0.782	0.524	0.612	0.615
Mean	0.556	0.558	0.341	0.615	0.612	0.570	0.415	0.337	0.604	0.547

2.1.4. The proportion of polymorphic loci

The proportion of polymorphic loci in a population is a measure of genetic variation which is often used. Results from polymorphism for all populations show high values (≥ 0.8) at both the 95 and 99% confidence intervals (Fig. 3.3).

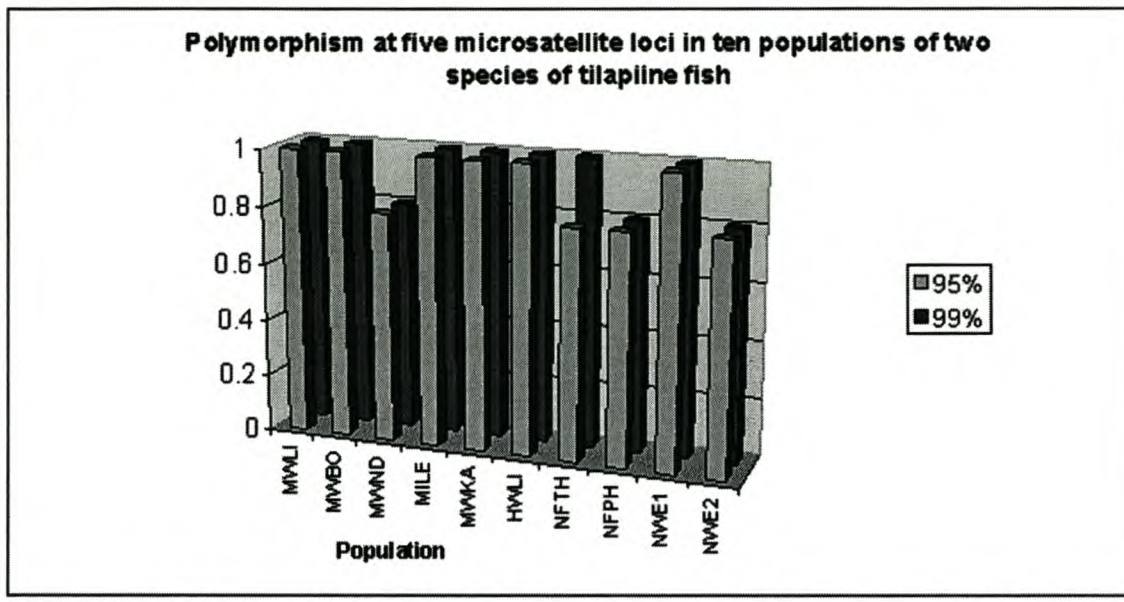

Figure 3.3 Polymorphism. Proportions of polymorphism are indicated where 95% indicates polymorphism at the 95% frequency level and 99% at the 99% frequency level. Polymorphism was calculated for groups of populations within a species but for species separately

2.1.5. Heterozygosity

Randomization tests showed that most of the samples were consistent with Hardy-Weinberg expectations when each locus from each population was analysed separately. Sixteen loci (32%) were found to have P-values significant at the 0.05 level after Bonferroni correction. When results are combined across loci for each location 6 (60%) of the populations showed significant departure from the Hardy-Weinberg expectations after adjustment for multiple tests.

Expected heterozygosities within each population and locus range from 0 to 0.86 in the *O. mossambicus* group, 0 to 0.87 in the *O. niloticus* group, and 0.47 to 0.80 for the population of hybrids. Averaged multi-locus heterozygosity ranged from 0.34 to 0.58 in the *O. mossambicus* group and from 0.34 to 0.80 for *O. niloticus*. For the population of hybrids, expected heterozygosity over all loci was 0.61 (when the difference in H_E and H_O is measured over all loci). Values were very similar for the expected and observed heterozygosity in each population.

Large differences were found between expected and observed heterozygosities in locus UNH102 for the introduced sample MILE (H_E = 0.82; H_O = 0.25) and the hybrid population HWLI (H_E = 0.43; H_O = 0.00). For locus UNH124 such big difference occurred only in the wild *O. mossambicus* sample MWBO (H_E = 0.66; H_O = 0.20). At loci UNH29 and UNH146 no such big differences occurred between the expected and observed heterozygosities. But again for the locus UNH192, the introduced *O. mossambicus* sample MILE showed big difference (H_E = 0.67; H_O = 0.14) as well as for the one *O. niloticus* farm population NFPH (H_E = 0.53; H_O = 0.95). This was the only case of heterozygote excess.

Table 3.3 Heterozygosity values. H_E. refers to the expected heterozygosity without bias using the method of Nei (1978) and H_O is the observed heterozygosity.

Species		*O.mossabicus*					Hybrids	*O.niloticus*			
	(N)	20	10	9	7	13	7	20	20	17	16
Locus		MWLI	MWBO	MWND	MILE	MWKA	HWLI	NFTH	NFPH	NWE1	NWE2
UNH102	H_E	0.5563	0.6895	0.3626	0.8214	0.8	0.4286	0.1141	0	0.4554	0
	H_O	0.4762	0.7	0.4286	0.25	0.6154	0	0.1176	0	0.3846	0
UNH124	H_E	0.6333	0.6579	0.6471	0.5111	0.6051	0.8571	0.7159	0.4848	0.8659	0.7923
	H_O	0.4444	0.2	1	0.2	0.5	0.75	0.6667	0.1818	0.8333	0.5625
UNH129	H_E	0.5389	0.5333	0.325	0.303	0.4674	0.2	0.0588	0.3713	0.4577	0.619
	H_O	0.2857	0.75	0.125	0.3333	0.4167	0.2	0.0588	0.4737	0.5625	0.3125
UNH146	H_E	0.6051	0.4789	0.3856	0.6154	0.5616	0.5333	0.4064	0.2962	0.6299	0.6794
	H_O	0.7619	0.7	0.4444	0.8571	0.8333	0.4	0.4706	0.35	0.6667	0.375
UNH192	H_E	0.4295	0.4158	0	0.6703	0.6154	0.7143	0.7833	0.5346	0.6078	0.6149
	H_O	0.2	0.1	0	0.1429	0.3846	0.5714	0.85	0.95	0.4706	0.625
Total	H_E	0.5526	0.5551	0.3441	0.5843	0.6099	0.5467	0.4157	0.3374	0.6033	0.5411
	H_O	0.4337	0.4900	0.3996	0.3567	0.5500	0.3843	0.4327	0.3911	0.5835	0.3750

2.1.6 Linkage disequilibrium

Evidence for linkage disequilibrium was found to be significant between loci UNH102 and UNH129 when all populations (10 combinations) are considered (Table3.4). When only the *O. mossambicus* populations were considered, significant linkage occurred again between UNH102 and UNH129 but also between UNH146 and UNH192. When only *O. niloticus* were considered, no significant values were observed. After sequential Bonferroni correction, these values are considered not anymore significant.

Table 3.4 P-value for each locus pair across all populations (Fisher's method). Values significant P-values at the 0.05 level are indicated by bold numbers.

	All populations			*O. mossambicus*			*O. niloticus*		
Locus pair	Chi2	df	P-value	Chi2	df	P-value	Chi2	df	P-value
UNH102 & UNH129	25.109	14	**0.03350**	20.063	8	**0.01010**	0.988	4	0.91162
UNH102 & UNH146	6.536	14	0.95111	5.317	8	0.72325	1.333	4	0.85572
UNH129 & UNH146	5.905	20	0.99902	2.033	10	0.99608	3.947	8	0.86189
UNH102 & UNH124	5.020	12	0.95731	4.936	8	0.76442	0.000	4	1.00000
UNH129 & UNH124	13.440	18	0.76482	7.683	10	0.65977	5.529	8	0.69985
UNH146 & UNH124	22.852	18	0.19633	8.734	10	0.55751	14.045	8	0.08060
UNH102 & UNH192	12.429	12	0.41186	9.150	6	0.16529	2.229	4	0.69380
UNH129 & UNH192	8.918	18	0.96162	2.470	8	0.96313	5.793	8	0.67039
UNH146 & UNH192	23.694	18	0.16531	15.924	8	**0.04348**	7.310	8	0.50354
UNH124 & UNH192	23.515	16	0.10064	14.615	8	0.06707	8.973	8	0.34461

2.2 Population structure

2.2.1. F-statistics

2.2.1(a) Inbreeding coefficient

Five out of the ten populations showed an overall significant ($P<0.05$) departure from Hardy-Weinberg genotypic proportions, showing an excess of heterozygotes in all five populations, with a mean inbreeding coefficient ranging

from 0.102 (MWKA) to 0.420 (MILE). Three out of five loci show a significant departure from Hardy-Weinberg equilibrium.

Table 3.5 F_{is} value per locus for all populations according to Weir & Cockerham (1984). Multilocus and multipopulation values are calculated using the method of Weir & Cockerham (1984) (FSTAT & GENETIX). Values were calculated by species. Values in brackets indicate P-values as calculated within each species group of populations. These P-values are unbiased estimates of Hardy-Weinberg exact P-values by Markov chain method after 1000 dememorization steps and 20 batches of 1000 iterations per batch. The alternative hypothesis of interest is the heterozygote state. Significant values at P<0.05 are indicated in bold.

	O.mossambicus						Hybrid			*O.niloticus*		
Pop	MWLI	MWBO	MWND	MILE	MWKA	Multipop	HWLI	NFTH	NFPH	NWE1	NWE2	Multipop
UNH102	0.147	-0.016	-0.200	0.727	**0.238**	**0.173**	1.000	-0.032	NA (-)	0.161	NA (-)	0.113
	(0.0084)	(0.0225)	(1.0000)	(0.0000)	**(0.0168)**	**(0.0000)**	(0.1429)	(1.0000)		(0.2245)		(0.2360)
UNH124	0.304	**0.707**	-0.600	0.636	0.180	**0.236**	0.143	**0.071**	0.636	0.039	**0.297**	**0.207**
	(0.1434)	**(0.0043)**	(1.0000)	(0.1111)	(0.0802)	**(0.0075)**	(0.1272)	**(0.0404)**	(0.0588)	(0.2472)	**(0.0214)**	**(0.0004)**
UNH129	**0.476**	-0.448	0.632	-0.111	**0.113**	**0.230**	0.000 (-)	0.000 (-)	-0.286	-0.239	**0.503**	**0.052**
	(0.0004)	(0.9801)	(0.2000)	(1.0000)	**(0.0467)**	**(0.0001)**			(1.0000)	(1.0000)	**(0.0006)**	**(0.0046)**
UNH146	-0.267	-0.500	-0.164	-0.440	-0.517	-0.363	0.273	-0.164	-0.188	-0.061	**0.456**	**0.066**
	(0.6799)	(1.0000)	(1.0000)	(1.0000)	(0.9902)	(0.9941)	(0.6190)	(1.0000)	(1.0000)	(0.6009)	**(0.0163)**	**(0.2646)**
UNH192	0.541	**0.769**	NA	**0.800**	**0.385**	**0.579**	0.213	-0.088	-0.814	0.231	-0.017	-0.166
	(0.0530)	**(0.0093)**	(1.0000)	**(0.0017)**	**(0.0180)**	**(0.0003)**	(0.3313)	(0.1030)	(1.0000)	(0.1214)	(0.4971)	(0.3964)
Multilocus	**0.220**	**0.122**	-0.173	**0.420**	**0.102**	**0.158**	**0.326**	-0.042	-0.159	0.034	**0.314**	**0.408**
	(0.0002)	**(0.0005)**	(0.8282)	**(0.0042)**	**(0.0009)**	**(0.0000)**	**(0.0390)**	(0.0503)	(0.9913)	(0.1028)	**(0.0000)**	**(0.0002)**

Null alleles

Frequencies of null alleles were low for most samples at most loci. In population MILE, more than 20 % null alleles were registered at 3 of the 5 loci. For populations MWBO and HWLI one locus shows expected null alleles at more than 20 %

Table 3.6 Expected frequency of null alleles according to Brookfield (1996)

Locus	MWLI	MWBO	MWND	MILE	MWKA	HWLI	NFTH	NFPH	NWE1	NWE2
102	0.051	-0.006	-0.048	0.314	0.103	0.300	-0.003	0.000	0.049	0.000
124	0.116	0.276	-0.214	0.206	0.065	0.058	0.029	0.204	0.017	0.128
129	0.165	-0.141	0.151	-0.023	0.035	0.000	0.000	-0.075	-0.072	0.189
146	-0.098	-0.150	-0.042	-0.150	-0.174	0.087	-0.046	-0.042	-0.023	0.181
192	0.161	0.223	0.000	0.316	0.143	0.083	-0.037	-0.271	0.085	-0.006
Total	0.077	0.042	-0.041	0.144	0.037	0.105	-0.012	-0.040	0.012	0.108

2.2.1(b) Fixation index (F_{ST})

A comparison was made among various global fixation indices (Table 3.7). The method of Weir & Cockerham (1984) measured the fixation index at locus UNH192 the highest of the five loci, but according to the of Rousset (1996), UNH192 was the lowest. When the indices of Roberston & Hill (1984) and that of Raufaste & Bonhomme (2000) were used, F_{ST} of UNH192 was second highest and in the middle (3rd) respectively.

Table 3.7 Fixation indices by locus as calculated using methods of Weir & Cockerham (1984) (Theta); Rousset (1996) (R_{ST}); Robertson & Hill (1984) (RH) and Raufaste & Banhomme (2000) (RH').

Locus	*O. mossambicus*				*O. niloticus*			
	Theta	R_{ST}	RH	RH'	Theta	R_{ST}	RH	RH'
UNH102	0.234	0.113	0.177	0.177	0.101	0.034	0.036	0.086
UNH124	0.118	0.030	0.112	0.294	0.216	0.284	0.147	0.147
UNH129	0.036	0.034	0.046	0.049	0.072	0.023	0.037	0.109
UNH146	0.102	0.147	0.067	0.073	0.106	0.003	0.045	0.217
UNH192	0.381	-0.024	0.114	0.114	0.257	0.061	0.172	0.172
Multilocus	0.188	0.066	0.118	0.162	0.181	0.170	0.104	0.154

Pairwise comparisons among populations (Table 3.8) also revealed substantial genetic differentiation between populations of different species. As expected, they were lower for populations of the same species. Particular strong divergence seems to occur when NFTH and NFPH are compared to MWBO and MWND.

Table 3.8 Pairwaise comparisons of Weir & Cockerham's (1984) estimation of the fixation index in the upper triangle, and the probability values in the lower triangle. Bold values indicate significance at the 95% level after sequential Bonferroni correction.

Theta	MWLI	MWBO	MWND	MILE	MWKA	HWLI	NFTH	NFPH	NWE1	NWE2
MWLI	-	**0.106**	**0.378**	0.089	**0.184**	**0.183**	**0.328**	**0.436**	**0.281**	**0.324**
MWBO	0.005	-	**0.336**	0.143	**0.167**	**0.266**	**0.455**	**0.511**	**0.344**	**0.401**
MWND	0.000	0.002	-	0.148	**0.133**	**0.368**	**0.493**	**0.538**	**0.387**	**0.462**
MILE	0.029	0.019	0.010	-	-0.022	0.114	**0.310**	**0.417**	**0.245**	**0.311**
MWKA	0.000	0.004	0.004	0.710	-	**0.199**	**0.340**	**0.411**	**0.274**	**0.327**
HWLI	0.001	0.000	0.000	0.063	0.000	-	**0.184**	**0.319**	0.129	**0.187**
NFTH	0.000	0.000	0.000	0.000	0.000	0.000	-	**0.286**	**0.099**	**0.153**
NFPH	0.000	0.000	0.000	0.000	0.000	0.002	0.000	-	**0.236**	**0.286**
NWE1	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	-	0.023
NWE2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.054	-

2.2.1(c) F_{IT}

The total inbreeding coefficient for the *O. mossambicus* ranges from –0.224 to 0.739 with a value over all loci of 0.317. For *O. niloticus*, inbreeding coefficients range from 0.120 to 0.379 and multilocus F_{IT} of 0.220.

Table 3.9 Multiallelic F_{IT} per locus for all populations according to Weir & Cockerham (1984, Fstat). Multilocus and multipopulation values are calculated using the method of Weir & Cockerham (1984) (Genetix). All values were calculated of each group of populations within a species separately.

Species	*O. mossambicus*	*O. niloticus*
UNH102	0.366	0.203
UNH129	0.258	0.120
UNH146	-0.224	0.165
UNH124	0.326	0.378
UNH192	0.739	0.134
Multilocus	0.317	0.220

2.2.2 Genetic distance

2.2.2(a) Matrixes

Pairwise genetic distance according to Cavalli-Sforza & Edwards's (1967) values were calculated between all populations to investigate relationships among allele frequencies (Table 3.10). The greatest genetic difference was observed between MWLI and NFPH (0.77), MWND and NWE2 (0.86), NFPH vs MWBO (1.05) and MWND (1.16).

When populations from *O. mossambicus* are compared among each other, distances ranged from 0.02 to 0.61; in *O. niloticus* distances ranged from 0.22 to 0.40. When populations of different species are compared, the values of genetic distance ranged from 0.32 to 1.16. The distances between hybrids and *O. mossambicus* populations ranged from 0.13 to 0.58 and for hybrids and *O. niloticus* populations, distances ranged from 0.15 to 0.47.

Table 3.10 Distance matrix according to Cavalli-Sforza and Edwards (1967). P-values are shown in the lower triangle and bold values indicate significance at the 95% level after sequential Bonferroni correction.

Fst/1-Fst	MWLI	MWBO	MWND	MILE	MWKA	HWLI	NFTH	NFPH	NWE1	NWE2
MWLI	-	**0.132**	**0.254**	**0.099**	**0.146**	**0.197**	**0.200**	**0.383**	**0.273**	**0.303**
MWBO	0.000	-	**0.149**	**0.123**	**0.114**	**0.261**	**0.384**	**0.464**	**0.367**	**0.448**
MWND	0.000	0.000	-	**0.130**	**0.115**	**0.250**	**0.326**	**0.356**	**0.317**	**0.372**
MILE	0.000	0.000	0.000	-	**0.050**	**0.202**	**0.247**	**0.309**	**0.308**	**0.348**
MWKA	0.000	0.000	0.000	0.006	-	**0.218**	**0.283**	**0.367**	**0.316**	**0.358**
HWLI	0.000	0.000	0.000	0.000	0.000	-	**0.177**	**0.247**	**0.214**	**0.253**
NFTH	0.000	0.000	0.000	0.000	0.000	0.000	-	**0.237**	**0.114**	**0.136**
NFPH	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-	**0.225**	**0.226**
NWE1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	-	**0.060**
NWE2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	-

2.2.2(b) Cluster analysis of genetic distances

Genetic distances were sampled in a phenogram with the UPGMA method. One thousand trees were bootstrapped to produce the final tree. There is a clear differentiation between the group of populations from *O. mossambicus* and *O. niloticus*. At this first separation point (node), the hybrid population groups with the *O. niloticus*. A high percentage of trees showed the differentiation of MILE and MWKA in the *O. mossambicus* group, but the other separations were weak and may not be trusted. The NWE1 and NWE2, and NFTH groups of *O. niloticus* separated clearly.

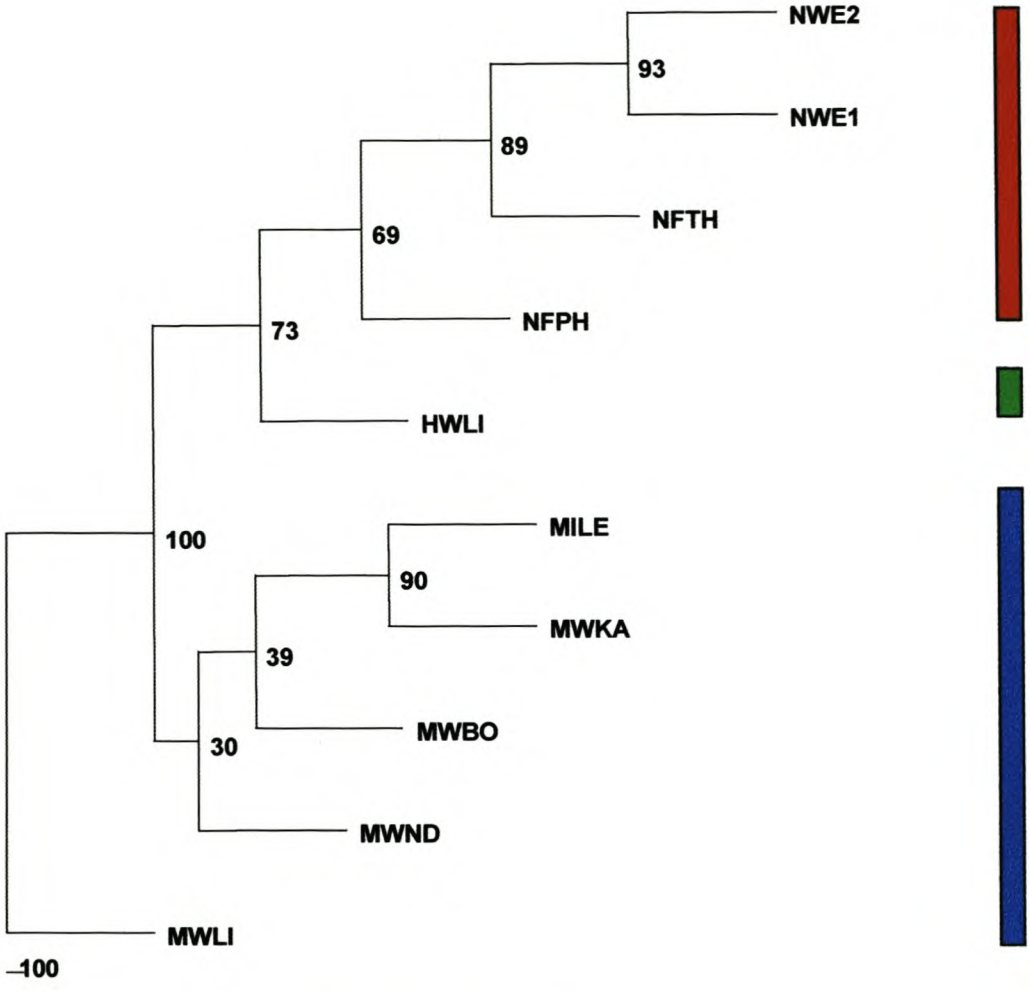

Figure 3.4 An UPGMA consensus tree (unrooted) after bootstrapping 1000 sets of data. The values at the nodes indicate the number of times (as percentage) the group consisting of the populations, which are to the right of the fork occurred among the trees (bootstrap values). The red line on the right of the picture indicates the populations of *O. niloticus*, the green that of the hybrids and the blue line indicates the populations of *O. mossambicus*.

2.2.2(c) Multivariate analysis by multi-dimensional scaling of genetic distances

With Figure 3.5, a "map" of the relationship among populations using multidimensional scaling, based on genetic distances of Cavalli-Sforza (1967) is constructed. Two dimensions were sufficient to represent the relationships among all populations, while preserving most of the information contained in the estimated pairwise distances.

Four individuals from the MWLI populations (individuals # 6,10,11 and 13) were found to relate closely to the group of populations from *O. niloticus*. One individual from *O. niloticus* (NWE1, individual #122) is plotted onto an individual from population MWND.

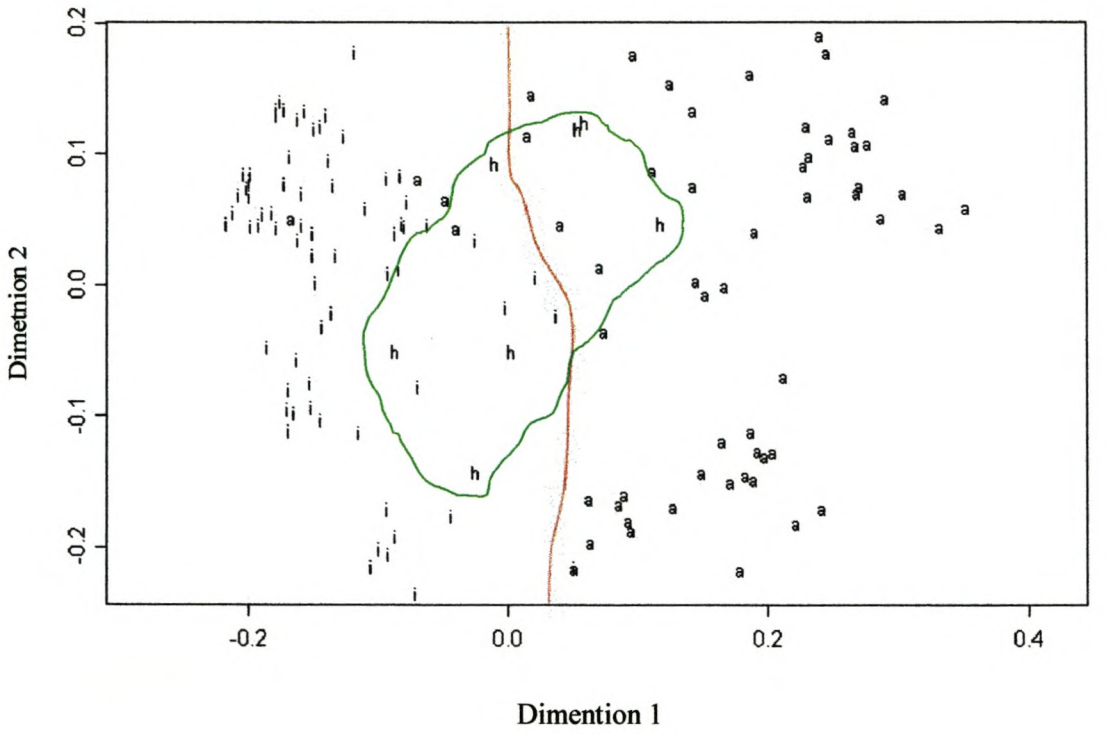

Figure 3.5 Multidimentional scaling plot based on the genetic distance as proposed by Cavalli-Sforza (1967). "a" Refers to individuals belonging to *O. mossambicus*, "h" refers to hybrids and "i" refers to individuals of *O. niloticus*. The green line crops the hybrid populations.

2.2.2(d) Correspondence factor analysis

Correspondence factor analysis plots individuals belonging to the populations for both species, except for some overlap in the middle. The hybrids are plotted throughout the ranges of both *O. mossambicus* and *O. niloticus*.

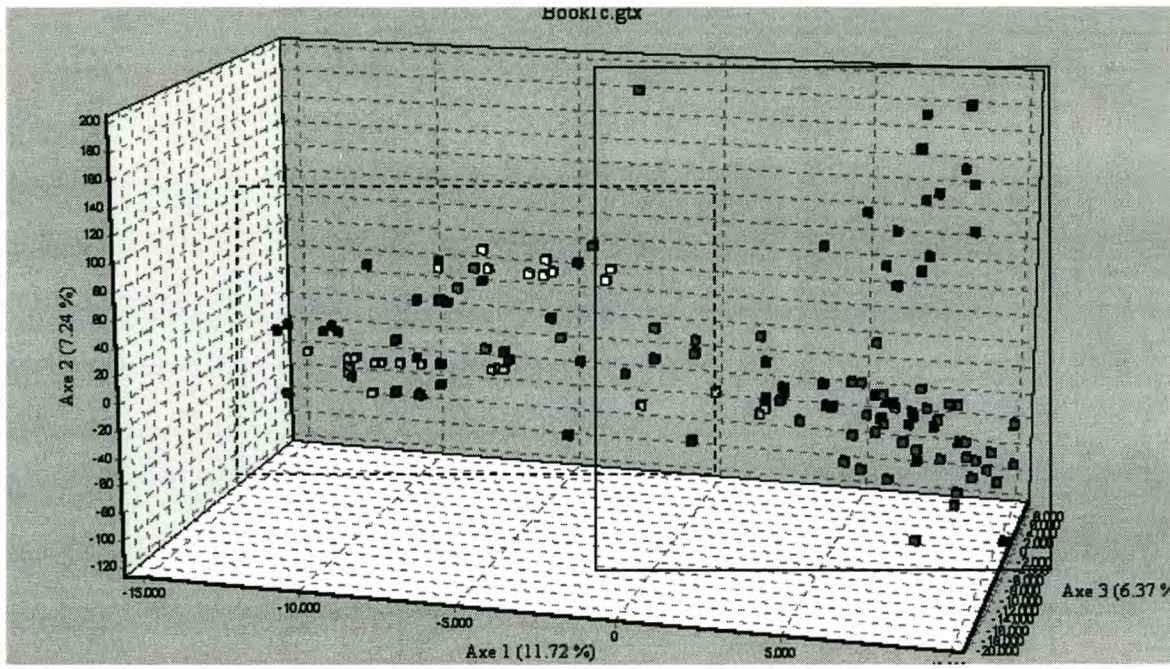

Figure 3.6 Three dimensional plot of correspondence factor analysis. The dotted line groups all individuals from *O. mossambicus*, and the solid line those of *O. niloticus*. Axe 1 refers to dimension 1, Axe 2 to dimension 2 and Axe 3 tot dimension 3.

2.2.3 Principal components analysis

All loci and populations were used in a principal components analysis (Figures 3.6 a to c).

Total heterozygosity was 0.697 whereas total inertia (distance X number of loci) was 1.058. The percentage of inertia was 50.34, 18.76 and 12.35, for the first 3 factors respectively. Cumulatively, the 80% inertia was reached after the third factor (81.45%).

Planes between factors 1 and 2, and 1 and 3 show clear differences between the populations of the two species. The plane between factors 2 and 3 was

less informative. One drawback of this programme though is that it is not possible to tell what the factors are accounted for. This is the same for some other analyses such as correspondence factor analysis and multi-dimensional scaling.

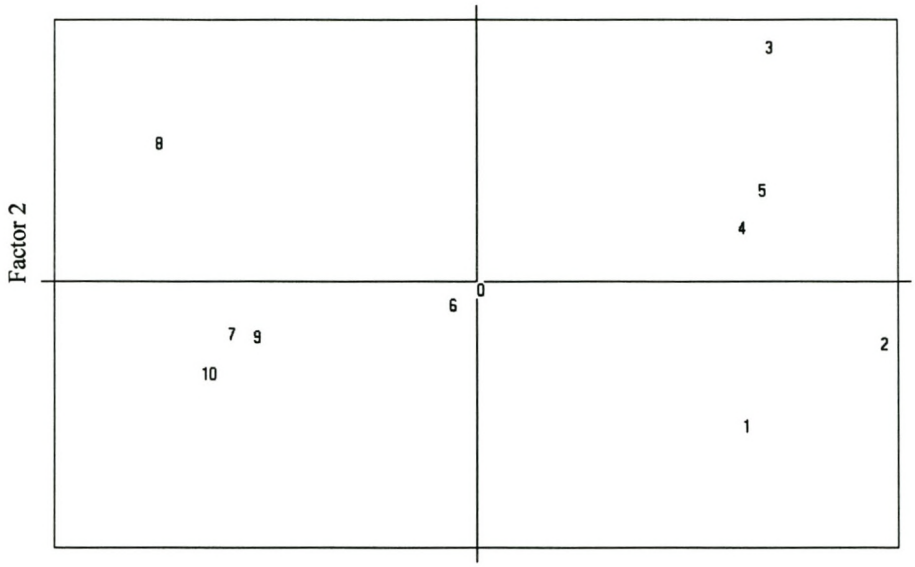

Figure 3.6a

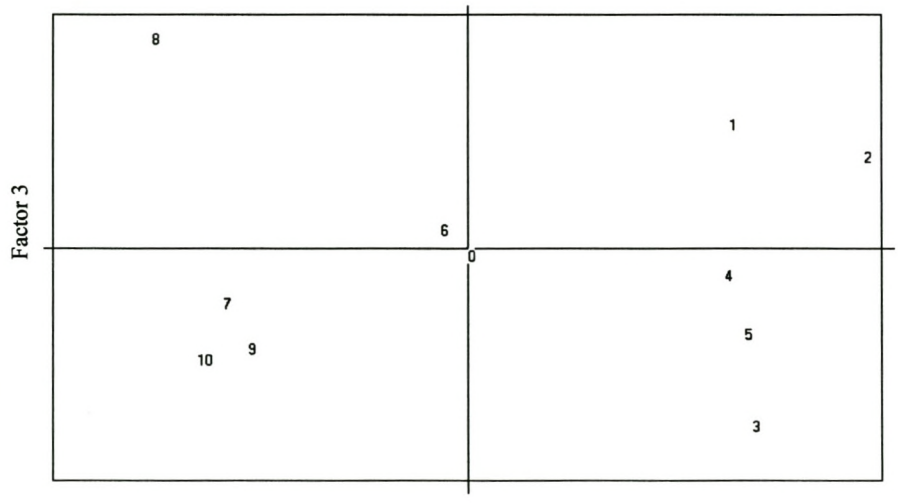

Figure 3.6b

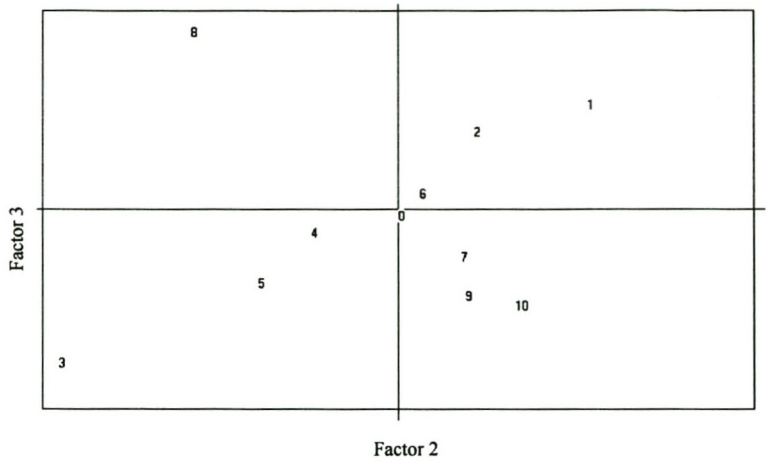

Figure 3.6c

2.3 Assignment of individuals to populations

All individuals from all samples were clustered according to Pritchard (2000) (Table 3.11 and Appendix C). The hybrids maintained their distinctiveness (100%). Most of the individuals from the samples of *O. mossambicus* had highest probability to be grouped into the hybrids, but *O. niloticus* had three of the four populations which belongs to this taxa assigned to *O. niloticus*, and one population's individuals (NFPH) assigned to *O. mossambicus*.

Table 3.11 Assignment of individuals within populations to certain taxa. Total refers to the total amount of individuals in the populations, *O. mossambicus* in the first row refers to the number of individuals within each population which are assigned to the species *O. mossambicus.* The same goes for Hybrid and *O. niloticus* in the fist row. Values in bold indicate percentages higher than 50 % of the individual of a population assigned to the species indicated in the fist row.

Species	Population	Total	*O. mossambicus*	Hybrid	*O. niloticus*
O. mossambicus	MWLI	21	0 (0)	**15 (71)**	6 (29)
	MWBO	10	0 (0)	**10 (100)**	0 (0)
	MWND	9	0 (0)	**9 (100)**	0 (0)
	MILE	7	0 (0)	**7 (100)**	0 (0)
	MWKA	14	1 (7)	**13 (93)**	0 (0)
Hybrid	HWLI	7	2 (29)	**4 (57)**	1 (14)
O. niloticus	NFTH	20	1 (5)	0 (0)	**19 (95)**
	NFPH	20	**20 (100)**	0 (0)	0 (0)
	NWE1	21	3 (14)	0 (0)	**18 (86)**
	NWE2	16	0 (0)	0 (0)	**16 (100)**

Chapter 4 Discussion

This study attempts to utilize five microsatellite DNA markers to genetically discriminate between two species of *Oreochromis* fish.

1. Microsatellite genotyping

1.1 Artifacts associated with microsatellite PCR

Although amplification of microsatellites by PCR is generally simple and reliable, microsatellites are known to show an artifact called "stutter bands". These are additional products produced during PCR and is a particular problem in analysis of mixed DNA samples (Armour *et al.*, 1999). In this study, during designation of peaks for each individual at each locus, mistakes could have been made because of this reason. Double-checking by independent readers should have limited such errors. Dinucleotide microsatellites were used which are more prone to produce stutter bands, as opposed to tetranucleotide microsatellites, although much less work has been done on tetranucleotides than on dinucleotides. Developing tetranucleotide microsatellite primers for this study would have been cost and time consuming (as apposed to using markers already identified for at least one of the species) and could perhaps still produce some (but probably less) stutter bands. For this reason we made an attempt to find markers which were already available to serve the purpose of this study.

Appearance of null alleles is another artifact, which is observed during this study. Non-amplification of certain alleles due to substitutions, insertions, or deletions within the priming sites can lead to null alleles appearing in population studies (e.g. Pemberton *et al.* 1995). For the introduced *O. mossambicus* population, more that 20 % null alleles were registered at 3 of the 5 loci and two other populations had at one locus each more than 20 % null alleles. This high appearance of null alleles could be ascribed to the fact that this study is conducted interspecifically, which could lead to serious ascertainment bias. Because these microsatellite markers were developed for *O. niloticus* (the focal

species), and applied to amplify fragments in *O. mossambicus*, some alleles could have been lost. This phenomenon was found in several other studies where microsatellites were developed for one species and applied in another one (e.g. Scott *et al.*, 2001; Zenger & Johnston, 2001). In these there were always much less alleles in the second or third species, than found in the focal species. Thus this high percentage of null alleles found in *O. mossambicus* may either be an over-estimation, or it may be that as in the other studies mentioned, they found less alleles because of null alleles due to non-specificity of the primers.

Products generated by *Taq* polymerase show "slippage" by this enzyme, which could cause adding of the wrong nucleotide during amplification. It is seen especially in mono- and dinucleotide microsatellite loci (e.g. Schlötterer & Tautz, 1992), which can sometimes make allele scoring problematic.

Taq polymerase also has a tendency to add an additional dATP to PCR products, which can cause single-base shifts and additional sizing problems (Ginot *et al.*, 1996). This artifact may explain the extra peaks at locus UNH 146.

An artifact of this marker system as opposed to mitochondrial markers, is that the material used has only one copy of a locus per cell, which can lead to stochastic amplification problems, especially where alterations of stringency conditions have been made. Several artifacts have been observed in these situations, including allelic "drop-out", non-specific amplification fragments obscuring analysis of the locus being studied, and generation of incorrect genotypes (Taberlet *et al.*, 1996 and 1997; Beaumont & Bruford, 1999).

Lastly, the major problem associated with microsatellite genotyping, is that no well-sustained evolutionary model is available which can be applied universally, or even to a subset of markers used in most studies. Without such a model, inference is difficult using allele frequency distribution, accurately quantifying genetic differentiation in problematic, and it is hard to explain the differences we see across species in terms of allele length and variability. There are mainly four models, which are used in conjunction with microsatellites in the recent literature

(Deka *et al* 1991; Estoup & Cornuet, 1999). These are (a) stepwise mutation model (SMM, Kimura & Ohta, 1978), (b) infinite allele model (IAM, Kimura & Crow, 1964), (c) two phase model (TPM, DiRienzo *et al.*, 1994) and (d) K-allele model (KAM, Crow & Kimura, 1970). Because microsatellites are such popular markers and have been used in so many studies, many conclusions have been drawn from studies which shape the form of future research (to a certain extend since several studies are conducted based on conclusions derived from microsatellite data). Thus it is extremely important to know as much as possible about mutation in microsatellite regions. Modeling becomes even more important to understand certain aspects. The four models developed to explain the mutation of microsatellites provide in most studies a reasonable explanation of results and a platform for interpretation. In answering which of the existing models should be used, care must be taken since the estimation of numerous population parameters depends upon the mutation model assumed for the markers. Even more, the sensitivity to the mutation model increases normally with mutation rate, and since this is normally very high for microsatellites, it makes the choice of a model even more critical.

One major flaw found in all these models, is that they assume that allelic differences are entirely due to changes in the number of the basic repeat unit. Since it is known that allelic size distributions indicate that microsatellite allele sizes do not always represent an increment of an entire number of repeats, they suggest that other forms of mutational change occur. Sequencing of microsatellite alleles has shown that insertions and deletions also occur in the flanking sequences, especially when different species are compared (e.g. Van Treuren *et al.* 1997). It would thus be of special interest to do some sequencing experiments for this study and also developing a more appropriate model.

1.2 Instability of microsatellites

Their inherent instability is the characteristic that makes loci containing microsatellite repeats particularly useful for evolutionary and genetic studies. Eisen (1999) summarizes what is known about the mechanism underlying

microsatellite instability and discuss some of the factors that cause variation in stability within and between species. The mechanism and role of slip-strand mispairing (SSM) cause (amongst others) microsatellite instability, but is regarded to be conserved between species. Thus it is likely that the specific rates and patterns of SSM differ greatly between species. Many of the factors that influence SSM errors within a species differ greatly between species (e.g. GC content, temperature and methylation).

During the course of this study, we were looking for markers as a tool to screen for two species of Tilapiine fish, and in both species microsatellite fragments were treated under the same amplification conditions. In the light of above-mentioned discussion, it would perhaps pay to investigate different amplification conditions for the two species in order to make the results more reliable and unambiguous.

1.3 Dinucleotides versus tri- and tetranucleotides

We used dinucleotide repeats in this study, mainly because these were available in literature. But how does it compare to other microsatellite repeat motives that are larger?

Studies done on humans, using di-, tri- and tetranucleotides, cast light on the informativeness of these different microsatellite motives (Deka *et al*, 1996; Jorde *et al*, 1995; de Knijff *et al*, 1997; Perez-Lezaun *et al*, 1997). Quantitatively, the results obtained with dinucleotide repeats and STRs are similar. However, data produced with the longer repeats do not seem to reflect population structure to the same extent as dinucleotide repeats. The longer repeats produce mean F_{st} values somewhat lower that CA repeats such as used in the present study, and result in trees less informative than the CA repeats.

Linares (1999) and others stated that the lower informativeness of tri- and tetranucleotides could relate to an allele size constraint acting on microsatellites, particularly considering that tetranucleotide repeats seem to have mutation rates

of the order of 10^{-3} while dinucleotide repeats mutate perhaps an order of magnitude more slowly. In two other studies (Weber & Wong, 1993; Heyer *et al*, 1997), mutation rates were *directly* estimated as four times higher for tetranucleotides than for dinucleotides. On the other hand, *indirectly* estimated relative mutation rates of di-, tri- and tetranucleotides showed different results (Chakraborty *et al*, 1997; Zhivotovsky & Feldman, 1995). Dinucleotides had mutation rates 1.5 to 2 times higher than tetranucleotides, with non-disease-causing trinucleotides intermediate between the di- and tetranucleotides, which is contradictory to studies where mutation rates were directly estimated. One explanation could be the choice of the mutation models which either of these studies accepted.

It is thus not yet clear whether dinucleotides would be more appropriate as species specific markers as opposed to tri-or tetranucleotides with regard to mutation rates. This study does not cast more light on the subject because no comparison was made with tri- or tetranucleotides. It could be useful to develop some longer microsatellite motifs to add to the knowledge of this topic.

1.4 Species differentiation

During the preliminary study, constant differences were evident on 8% polyacrylamide gels between the two species. The microsatellites in one species were constantly longer than in the other, similar to several other studies (Ellegren *et al*, 1995; Rubinsztein *et al*, 1995; van Treuren *et al*, 1997; Crawford *et al*, 1998). By definition, the lineage in which the microsatellites are longer must have experienced a greater average number of expansion mutations since the most recent common ancestor. This could come about either by a shift in the equilibrium length distribution in one or both genomes, or if a biased mutation process were accompanied by a difference in the average mutation rate between the lineages (Amos, 1999).

In our study *Oreochromis niloticus* had the longer fragments, and in the light of the discussion in the previous paragraph, this is not consistent with the theory of

its evolutionary developmental history. From studies by Nagl *et al.* (2001) and Pouyaud & Agnèse (1995) it is evident that *O. mossambicus* evolved more recently relative to *O. niloticus*.

Heterozygote instability provides one possible mechanism by which the genome-wide microsatellite mutation rate could change rapidly over short periods of evolutionary time. Assuming an expansion-prone mutation process, the prediction would be that expanded populations would carry longer microsatellites than their homologues in smaller populations (Rubinsztein *et al.*, 1995). Several relevant data sets allow this prediction to be examined as discussed by Amos (1999). Amongst others, there is the example of human microsatellites, which are longer than their homologues in chimpanzees (*Pan troglodytes*). Similar cases were reported for barn swallows (Ellegren *et al.*, 1995), rats (Beckmann & Weber, 1992) and sheep (Crawford *et al.*, 1998).

Ellegren *et al.* (1995) suggests that it is the cloning process, not population size differences, which cause the observed length differences. It must be kept in mind that most microsatellites are cloned from abundant species. The proposition is that, since long microsatellites are selected as markers, marker loci are likely to be longer than their homologues in other lineages, an effect known as ascertainment bias. In this study, microsatellites were developed for the more abundant species (*O. niloticus*), but it is also this species, which is the less recently evolved of the two. There is a critical need to determine the size of this ascertainment bias, since locus length comparisons between populations and species need to allow for this artefactual component. Any effect over and above that ascribable to ascertainment bias would imply either a change in average mutation rate or a shift in the equilibrium state.

Because these microsatellite markers are selected in a focal species (*O. niloticus*), it may differ systematically from its orthologues in related species (such as *O. mossambicus*) due to the criteria used to isolate it in the focal species (Ascertainment bias, Goldstein & Schlötterer, 1999). For microsatellites, most work on mutation has naturally come from these loci with the highest rates of

mutation. The vast majority of loci, including some of considerable interest (Morral *et al.*, 1994), have rates too low to measure by direct observation. In this particular study, we are not able to confirm whether the loci used are subject to locus ascertainment or not because we could not infer any mutation rates. We attempt to find an answer in the light of work done by Goldstein & Pollock (1997).

Goldstein & Pollock (1997) state that an estimation of ascertainment bias can be obtained as half of the minimum length acceptable as a marker in the focal species (C/2 where C is the minimum length acceptable as a marker in the focal species). Of course assumptions have to be made: unbiased mutation, all correlation between the lengths of homologous loci due to shared ancestry has been lost, uniform mutation rate over all lengths, and reflecting upper and lower length boundaries. The validity of most of these assumptions can be questioned, but the formula provides a useful starting point for thinking about how to estimate ascertainment bias (see Amos, 1999 for a detailed discussion and examples). The minimum length acceptable as a marker in the focal species (*O. niloticus*) is estimated at 149, 293, 195, 116, and 128 bp in loci UNH102, UNH124, UNH129, UNH146 and UNH192 respectively. Thus the estimation of the amount of ascertainment bias is 74.5, 146.5, 97.5, 58 and 64 for the respective alleles. No other studies were found to relate these results to. These values can still not be trusted completely since the assumptions necessary in order to use the formula of Goldstein & Pollock (1997) cannot be made in all cases to its full extend.

2. Genetic Variation

2.1 Per locus and species

2.1.1 Number of alleles

DNA sequences evolve through mutations, creating novelty, followed by the actions of neutral genetic drift and natural selection, which determine whether and how fast new forms spread within and between populations. Of these three fundamental processes, the *nature of mutation* is least understood mainly because individual events occur too frequently to be studied directly. Lately, several aspects related to mutations and how it happen became clear: (a) many

short tandem repeat markers have mutation rates high enough that significant numbers of germline mutations can be identified in pedigrees; (b) PCR techniques made it possible to study *de novo* mutant molecules in DNA isolated from sperm, and (c) the combination of high throughput genotyping systems and the ability of many microsatellite markers to amplify across species has resulted in large amounts of data, both in terms of hundreds or even thousands of loci within key species (see review by Amos, 1999).

The number of alleles present at the loci varied from 6 at locus UNH 129 to 11 in UNH 124. The expected number of null alleles for locus UNH 129 was close to zero and thus the low allelic variation at UNH 129 can only be accounted to the choice of loci and not due to missed alleles because of null alleles.

Common alleles (alleles with total frequency of at least 5 % in a population) were low (4) for three loci (UNH 102, UNH 129 and UNH 146). The other two loci (UNH 124 and UNH 192) exhibited intermediate numbers of common alleles (8 and 7 respectively), which is common in fish (and even higher numbers had been found, e.g. van Treuren *et al.*, 1997). No reason for this phenomenon could have been found in literature apart from the fact that they are selected as such. In a study conducted by Newman & Squire (2001) on population structure in wood frogs, the common alleles were ranging from 2 to 5, which was even lower.

When the number of alleles is compared between species, we found a total mean number of alleles over all loci of 5.2 in *O. mossambicus* and 10.0 in *O. niloticus*, which indicates more diversity among the *O. niloticus* group of populations. This would be unexpected, since two of the four populations in this group were expected to be inbred farm populations. Alternatively, outcrossing between inbred lines could have let to this.

2.1.2 Allele frequencies

No clear differentiation between the two species could be made on account of allele frequencies because there was an overlap between alleles at all loci tested. It could well be seen in Figure 3.2a that at locus UNH102, the majority of alleles

in the *O. niloticus* group belonged to allele 157 bp. In the *O. mossambicus* group, the only population, which exhibited higher frequency of that allele, is MWLI and the majority of alleles were smaller than 157 bp.

At locus UNH147, higher frequencies of smaller alleles were estimated in *O. mossambicus* than in *O. niloticus*, where high frequencies of larger alleles were evident.

These results are too limited to conclude anything, and I suggest increasing the amount of loci in order to make meaningful conclusions from these allele frequencies. The present frequencies/numbers of alleles may be used in some statistical models, in order to explain data meaningfully.

2.1.3 Polymorphism

Even though the microsatellite markers used in this study have been available for the focus species for 5 years (Lee & Kocher, 1996), information on their polymorphism was never published. Even more, no species differences studies were available for microsatellites in these two species under investigation for comparison or to improve on.

All the Tilapiine populations sampled for this study were polymorphic at all five loci, except for the Ndumu population at locus UNH192 and the Philippines population (*O. niloticus*) at locus UNH102. The Philippines population is expected to be very inbred leading to these results, and the Ndumu population may be again subject to severe selection.

2.1.4 Heterozygosity

Average heterozygosity at the loci was intermediate (0.51) when compared with values estimated for freshwater fish (0.46 – Table 4.1) by de Woody and Avise (2000). Macaranas *et al.* (1995) found lower heterozygosity values with allozymes in five cultured African strains of *O. niloticus.* They describe these low values to be due to limited sampling of the wild populations. Alternatively, occurrence of bottlenecking due to bouts of drought in the wild habitats may have

resulted in decreased heterozygoity since McAndrew and Majumdar (1983) reported similarly low levels of heterozygosity for African *O. niloticus* and several other Tilapiine species. Macaranas *et al.* (1995) also found higher heterozygosities in their Philippine strains and ascribed that phenomena to introgressive hybridizations with feral *O. mossambicus* (Macaranas *et al.*, 1986). We found rather low values for our sample of Philippine *O. niloticus*, but this may be due to the selection of markers used (selected to be homogeneous within species) and also because we compared two different species in our study (again, selection of markers in preliminary study provided least heterozygosity within species) (refer to next two paragraphs).

The interest in a deficiency of heterozygotes is supported by three reasons: (a) unexpected subdivision within a sample is expected to lead to a deficiency of heterozygotes ("Wahlund-effect"); (b) presence of "null" or nonamplifying alleles within a population will lead to an excess of "apparent" homozygotes (Brookfield, 1996), and (c) non random mating (inbreeding). Null alleles (Pemberton, *et al.*, 1995) pose a challenge to investigators of microsatellite variation because they may vary in identity of frequency among populations, leading to unknown biases in estimates of allele frequencies and population differentiation. Strategies for dealing with null alleles include dropping loci suspected of having null alleles from the study, redesigning primers, or statistical correction (Chakraborty *et al.*, 1992; Brookfield, 1996).

A global test for heterozygote deficiency is not significant for both species (separately or pooled – Table 3.1). When individual populations are considered at individual loci, there are some populations that show very low heterozygosity values: the Limpopo (MWLI) and Ndumu (MWND) populations have very low heterozygosity values at locus UNH192. Both these populations are wild populations in which genetic drift (Walhund effect - Walhund, 1928) may have occurred. The population from the Philippines (NFPH) is fixed for allele UNH102. This sample was taken from a farm population and therefore these values can be attributed to inbreeding. A good proof for the presence of null alleles designates individuals that do not amplify at a certain locus. Appendix a summarizes such

individuals, the populations in which they occur and the loci at which they do not amplify. Null alleles are thus suspected to occur at loci UNH102 and UNH124, and further analyses were done with both data sets, either including or excluding these loci.

The level of microsatellite diversity of the loci used in this study in comparison with microsatellite diversity in other fish, and other loci, is demonstrated in Table 4.1. The lowest heterozygosity observed was in five populations of *O. mossambicus*, although *O. niloticus* also exhibited low heterozygosity. This compares well with earlier studies of microsatellites with fish in that it is similar to the mean heterozygosity value of freshwater fish. The low allelic diversity of *O. niloticus* may be explained by the fact that the species is so highly cultivated. There are the inbred groups of farm animals, and few natural populations left, because this species serves as a main source of aquacultured Tilapia. This implies that the loci used for analysis in this study are subject to inbreeding, and serve therefore to a lesser degree as species specific marker, than hoped for. This is only a suggestion, and can be tested if more populations of *O. niloticus* from more areas are included to improve the analysis.

Table 4.1 Table comparing the genetic diversity of various groups of fish. Values for freshwater fish, anadromous fish and marine fish were taken from a review by de Woody and Avise (2000).

	Mean heterozygosity	Mean number of alleles	Allozyme heterozygosity
Freshwater fish	0.46	9.10	0.046
Anadromous fish	0.68	10.80	0.052
Marine fish	0.77	19.90	0.059
O. mossambicus	0.49	3.56	
O. niloticus	0.53	5.00	0.046*

*from Macaranas *et al.* (1995)

2.1.5 Linkage disequilibrium

All populations exhibited significant linkage disequilibrium at just one locus combination (UNH102 & UNH129). When only *O. mossambicus* is considered, significant linkage disequilibrium was found between loci UNH102 & UNH129 as well as UNH146 & UNH192. For the *O. niloticus* group, no significant linkage disequilibrium was found (Table 3.4).

This indicates that the loci of both *O. mossambicus* and *O. niloticus* are on different linkage groups. Independent allele assortment of loci occurs in most of the populations. Thus the loci used are all appropriate fro a study such as this, because they will all evolve independently from each other.

2.2 Between all populations

The total number of alleles over all populations ranged from 6 to 11 per locus, which is low compared to a review on marine, freshwater and anadromous fish (de Woody & Avise, 2000). This can be explained by the fact that the two species of interest in this study, is more closely related than the groups studied by de Woody & Avise (2000). In other studies where microsatellites were used, the number of alleles observed is higher (e.g. Brunner *et al.*, 1998). Some microsatellite studies concerned with several populations of a single species show values with the same range of allele numbers we have found (e.g. Daemen *et al.*, 2001).

The allele frequencies exhibited different patterns for most of the populations. At locus UNH129, high allele frequencies occurred for all populations at one allele (207 bp). In this respect, this allele could be a non-specific allele, or data from the automated sequencer might have been wrongly interpreted.

Gene diversity per population over all loci ranged from 0.337 in NFPH to 0.615 in MILE, which are intermediate values, compared to other studies (e.g. Hall, 2001).

2.3 Farm populations

We tested several populations to avoid, if there exists, a source of error in allele number/frequency due to population structure or differential genetic drift between populations within species. Similar problems are faced when species are of farm origin, where a certain level of inbreeding is expected. To overcome these problems, a large number of individuals per sample and at least four samples (populations) per species were included. Three farm populations were used: one *O. mossambicus* from Le Pommier (MILE), and two *O. niloticus* populations, one from Thailand (NFTH) and one from the Philippines (NFPH).

The unbiased estimator of gene diversity (Nei, 1987) for these was calculated as 0.165 (MILE), 0.415 (NFTH) and 0.337 (NFPH). This is consistent with expectations for these populations since MILE has been introduced in the 1940's. The other two populations were established long before and cultured for many years. These two populations are expected to even have the same source population. Still, the NFPH population, which has the lowest gene diversity, comes from a farm where experiments are conducted to produce YY males, and half of the sample (10 out of 20) included YY males.

With regards to expected and observed heterozygosity in farm populations, the two *O. niloticus* populations exhibited heterozygote excess ($H_O > H_E$). These results contradict what is expected for these populations, because heterozygote excess results from outbreeding and/or gathering of populations (opposite of population subdivision).

On the other hand, the other farm population, MILE, had the largest heterozygote deficiency exhibited in the study. This could be because it is a farm population cultured for almost 60 years. Thus it is especially prone to inbreeding (since the genetic management of genetic stock became evident in South Africa only in recent years), which could result in a heterozygote deficiency. The sample size of the population was also small (7 individuals that amplified at these loci), and thus small sample bias may result in some alleles to be missed from this population. The expected frequency of null alleles for

this population over all loci is 14 %; hence it is reasonable to expect null alleles to be responsible for the heterozygote deficiency in this population. Finally, this population is subject to selection (selection for better aquaculture performance), which can contribute to heterozygosity.

2.4 Wild populations

Apart from population MWND, the wild populations exhibited higher values of non-biased genetic diversity (Nei, 1987) than found in the farm populations. Still, the factor that lowers the gene diversity value for MWND, is the zero diversity at locus UNH192. If this locus is excluded for MWND, the gene diversity value is in the range of other wild populations.

Polymorphism at all five loci was again higher in most wild populations than in farm populations.

All wild populations but one (MWND) exhibited heterozygote deficiency, although *O. mossambicus* populations and one *O. niloticus* population had a low deficit. The other wild *O. niloticus* population and the hybrid population showed larger heterozygote deficiencies. For the hybrid population, small sample bias or null alleles (10.5 %) can be responsible for the heterozygote deficiencies (hybrids are usually expected to have a heterozygote surplus).

Heterozygote deficiencies in other populations can be accounted for by selection and the Walhund effect. For population NWE2, 10.8 % null alleles are calculated to be expected, providing additional explanation for heterozygote deficiency.

3. Population structure

3.1 F-statistics

In this section, first, factors influencing deviations from expected heterozygosity (and thus Hardy-Weinberg equilibrium) are discussed. F_{IS} measures the reduction in heterozygosity of an individual due to non-random mating within its

F-statistics would reveal inbreeding to a degree that it could be quantified in a population. Genetic differentiation between populations can be measured by the F_{ST} values in pairwise comparisons between populations (Table 3.5).

3.1.1 Inbreeding coefficient

Almost all F_{IS} values calculated for the populations were not significant (at $P < 0.05$ level). Only one population of *O. niloticus* had a significant F_{IS} value of 0.314, which is, compared to a mean value of 0.037 found with allozymes for this species (Macaranas *et al.*, 1995), a rather high coefficient of inbreeding. Since microsatellites are considered much more sensitive to inbreeding, this is not surprising. In comparison to similar studies where microsatellites were used (e.g. Daemen *et al.*, 2001; Goodisman *et al.*, 2001), we found a high inbreeding coefficient in our study, which was expected.

Another, even higher significant value, was exhibited by the hybrid population. This is not what would be expected from this population, since it is a wild population in which random mating can occur. The high inbreeding coefficient in this population may be explained by two factors in particular: (a) small sample bias, by which some alleles in the population may be missed, or (b) selection by one species for certain traits or gender in individuals of the other species.

The inbreeding coefficient over all loci and populations is significant for both *O. mossambicus* and *O. niloticus* (0.158 and 0.408 respectively), but is much higher in *O. niloticus*. It can be explained by the fact that this group contains two populations (out of four), which are suspected to be highly inbred.

Within the *O. mossambicus* group, the population from Le Pommier (MILE), which is the only farm population included, exhibits also a significantly high inbreeding coefficient (0.420). Three of the remaining four populations in this group showed also significant F_{IS} values but are of much lower magnitude.

3.1.2 Fixation index

It is evident from Table 3.5 that genetic variation between populations of the same species is lower than between populations of different species. In this study, F_{ST} values (Table 3.5) are not discriminatory enough to ascribe unknown populations to a certain species. Here I would like to refer to the overlap in allele frequency distribution when pooled data are considered (Fig. 3.2a-e). Thus I would suggest if microsatellites were to be used in future to ascribe a certain individual to a certain species, it would be necessary to search for loci where there is no overlap in allele frequency distribution (private alleles).

Most values in pairwise comparisons (33 out of 45) lead to rejection of homogeneity. Such differentiation strongly rejects a high level of gene flow among samples of populations, which is meaningful because none of the populations used in this study is able to mix naturally because of their geographical range. Very low (~0.00) values are noted between the population from Le Pommier (MILE) and from Kasinthula (MWKA), and also between the two populations from Egypt (NWE1 and NWE2). MILE is a farm population and is strongly suspected to have its stock taken from Malawi (where MWKA occurs). NWE1 and NWE2 are two wild populations form Egypt. Therefore, gene flow or a common source population may be the reason in both cases (for such a degree of homogeneity in genetic composition at the polymorphic loci). If time since separation has been insufficient for divergence due to drift or accumulation of novel mutations, populations may remain similar for some time, depending on the rates of drift and mutation, even in the absence of ongoing gene flow. At the spatial scale separating populations NWE1 and NWE2, this last argument (in addition to the first – that of gene flow), may be very relevant. For populations MILE and MWKA, a high level of gene flow is not evident because of the total spatial separation of the two populations.

Different methods were used to calculate the fixation index. It was found that the methods used differed greatly in estimations of the fixation index, although the methods of Robertson & Hill (1984) and Raufaste & Bonhomme (2000) were very similar. We decided to continue further calculations using the method of Weir &

Cockerham (1984) because it is the most often used method in literature, and it will enable us to compare our data more meaningfully.

According to Weir & Cockerham's (1984) estimate of the fixation index, *O. mossambicus* and *O. niloticus* were very similar and both intermediate, although somewhat lower compared to the measure of population subdivision found in *O. niloticus* when four wild strains were tested with allozymes (Macaranas *et al.*, 1995, Table 4.2).

Table 4.2 Comparison of F-statistics in *O. mossambicus* and *O. niloticus*.

	F_{IS}		F_{ST}		F_{IT}	
	O. mossam-bicus	*O. niloticus*	*O. mossam-bicus*	*O. niloticus*	*O. mossam-bicus*	*O. niloticus*
Present study	0.158	0.408	0.188	0.181	0.317	0.220
Macaranas *et al.*,1995	-	0.229	-	0.037	-	0.257
Hall, 2001	-	-	0.260	-	-	-

On account of these results, using this set of primers, we found that neither of the two species tested were in Hardy-Weinberg equilibrium, which is consistent to what was expected. The amount of population subdivision exhibited can be explained largely by inbreeding, especially in the case of *O. niloticus*, or selection, null alleles and non-random mating.

In comparison with Hall (2001), a significant F_{ST} value of 0.26 (P<0.001) was calculated for *O. mossambicus*, using three microsatellite primers. This value is slightly higher than what is found in our study with five other microsatellite markers.

3.1.3 Total inbreeding coefficient

According to this measure, individuals from *O. mossambicus* exhibited higher overall inbreeding than *O. niloticus*, which is inconsistent with the F_{IS} and F_{ST} measures.

3.2 Genetic Distance

3.2.1 Cluster analysis

Phylogenetic analysis revealed that the populations of *O. mossambicus* (*n=5*) and *O. niloticus* (*n=4*) formed distinct clades (Cavalli-Sforza's method, 100 % bootstrap support). Within these two species, populations separated to some degree with the set of five microsatellite markers we used in the main study. This is consistent with the expectation.

Within the *O. niloticus* group, the two wild populations (NWE1 and NEW2) as well as the farm populations from Thailand separated clearly. The NFTH population is also the farm population that grouped closest to the wild populations.

For *O. mossambicus*, the introduced population (MILE) separated clearly (90% bootstrap support) from the wild populations, but the wild populations did not separate clearly among each other. Four of the five populations included in the present study, were used by Hall (2001) where he was able to discriminate between the populations using three microsatellite markers (CA dinucleotide repeats). The levels of discrimination (e.g. bootstrap values) were unfortunately not available, but his separations correspond well with what we found using five different microsatellite markers.

3.2.2 Multi-dimensional scaling

With this technique we intended to construct a "map" of the relationships among populations using multidimentional scaling, based on Cavalli-Sforza's (1967) measures of genetic distance. Two dimensions were sufficient to represent the relationships among all populations, while preserving most of the information contained in the estimated pairwise distances.

It appears that this method can be used to diffentiate between the two species, although an overlap is evident between the individuals. The hybrid population fits into the "map" as was expected in the middle of the two species. However it will be necessary to find more loci to separate the two species more clearly.

The three loci used by Hall (2001) is now known to amplify successfully in *O. mossambicus* and since they were isolated using *O. niloticus* it may be used to find more clear separations between the species.

3.2.3 Principal component analysis

Although an earlier study (Moralee *et al.* 2000) showed that hybrids between the exotic and endemic tilapias have occurred, our study provides evidence that nuclear differences between the two species exist and with the five markers identified, it is possible to differentiate between the species. It is still necessary to identify more loci in order to make the groups of scattering populations in a principle component analysis (see figures 3.5 a to c) more concise and stringent. "One can obtain a fairly good estimate of heterozygosity by examining a few individuals for a number of loci." (Nei, 1987).

A similar study was conducted by Agnèse *et al.* (1997) on two species of African Catfish (*Clarias spp.*) using various marker systems. They obtained meaningful results showing two clear clouds of specimens from the two different species tested. They also identified a hybrid taking intermediate position in between the two clouds, which indicates the possible meaningful use of this method of analysis. We had less clear results, but including more markers and different kinds of marker systems (e.g. mitochondrial markers) this analysis seems to be promising.

4 Assignment techniques

Assignment of individuals to certain populations appeared to be a promising method to be incorporated in this study, especially since fragment sizes of the set of markers used, were overlapping between the two species.

When results from the assignment is considered statistically (Table 3.11), individuals in only 4 out of the 10 populations are assigned to the taxa they are morphologically sampled for. This asks for serious consideration of the applicability of this technique in this study.

Firstly, mistakes could have been made during designation of genotypes (peaks) to individuals, but to come to such contradictory results in this analysis, these mistakes should have been seen to a larger extent in the other analyses.

Secondly, during sampling, mistakes may have been made in the morphological identification of individuals. We see in other analyses that some individuals in each population may have been wrongly identified as a certain species, but if fact appear to be a hybrid. However with the assignment method, most individuals of not even one of the populations of *O. mossambicus* are assigned as to be in fact *O. mossambicus*. This contradicts with other methods of analysis used in this study. Again, individuals from the farm population from the Philippines (NFPH) were all (100%) assigned to *O. mossambicus*. Macaranas *et al.* (1986) reported the introgression of genes from feral populations of *O. mossambicus* into *O. niloticus* in the Philippines. The sample for our study came from a farm which has (as considered) highly inbred populations of *O. niloticus*. Even if we accept that some hybrids were present in this population, 100% of individuals assigned to *O. mossambicus* (not even the hybrids), is not acceptable. In a case where all individuals in this population would be wrongly identified morphologically, it should at least be hybrids (and not *O. mossambicus*) since they were all identified as pure *O. niloticus*.

Thirdly, two assumptions are made when assigning individuals to populations (or taxa, in this case): within populations, the loci are in Hardy-Weinberg- and linkage equilibrium. (The method can be applied to microsatellites provided that the loci are unlinked. When considering Table 3.4, for most locus combinations, no significant linkage disequilibrium has been found, but significant linkage disequilibrium as found between loci UNH102 and UNH129. Since the programme we used (STRUCTURE) is relatively recently set up, and results from using the programme is not very often found in literature, we could not find to what extent the linkage disequilibrium of one locus combination may influence the results. On account of the results shown in this study, it appears that this factor may alter results largely.

5 Suggestions for further research

First, to improve on this study, sampling has to be extended. More individuals per sample should be taken, and more populations should be included. The ideal would be to sample 50 individuals from as many wild populations for the pure species as possible. The problem exists though that there is hardly any population of *O. mossambicus* left of which one can be sure that it is not intruded with *O. niloticus*. On the other hand, the hybrid populations studied should ideally be cultured populations and individuals should be used which are surely F_1 hybrids. More of these populations should also be included.

Secondly, the type of markers used in this study is not completely suitable at these early stages. Microsatellites are much too sensitive and reveal data which cannot be interpreted fully yet. The first study to answer the same question, has been done using allozymes (Moralee *et al.*, 2000) and another, which also casts some light on the topic, by Pouyaund *et al.* (1995). Allozymes are often considered in literature to be of little help, but in studies such as this, it would provide a good platform and background knowledge on species-specific differences. Thus I would suggest to first find a good battery of allozyme makers from trusted pure populations [such as done by Rognon *et al.*, 1998 on African catfishes (*Clarias gariepinus* and *C. anguillaris*)]. Secondly as a marker system following allozymes, maternally inherited mitochondrial markers would be appropriate (especially on Cytochrome b and ND 5 and 6 regions) (D'Amato, unpublished). Thirdly, a set of AFLP markers would provide for each species a fingerprint and a good starting point could be at nucleons (introns) such as for Aldolase. With this battery of markers (allozymes, mtDNA and AFLP's), some individuals may appear as hybrids in the populations and thus can be excluded in order to make analyses more secure. At that stage a set of microsatellite markers should be incorporated and only then should populations of hybrids be included. At this stage a linkage map should be helpful to quantify the degree of hybridization and introgression.

Chapter 5 Conclusions

In conclusion, the five microsatellite loci, UNH102, UNH124, UNH129, UNH146 and UNH192 are enabling us to discriminate between populations from the species *Oreochromis mossambicus* and *O. niloticus*. However, it is not possible to use these markers to identify populations of hybrids since there is still uncertainty of the position of hybrid populations in the various analyses which is possible to perform.

Secondly, it is not possible to use this set of dinucleotide repeat markers to identify individuals as either pure species or hybrids. The method of assigning individuals to certain populations or taxa, do not cast more light on this issue as applied in this study.

Thirdly, if the results of this study are considered, we suspect the natural drainage systems of Southern Africa to be invaded to a larger extend by *O. niloticus* than what was expected. What is morphologically expected to be *O. mossambicus,* appear to be later generations (F_3 or F_4) of backcrossed hybrids.

Finally, the set of markers used in this study need to be expanded, in both the type of marker system and the number of markers used, as well as numbers and locations of samples.

This study thus provides preliminary work from which hypotheses may be made and tested in providing a tool to study the current situation of invasion by *O. niloticus* in the Southern African freshwater systems.

REFERENCES

Agnèse J-F; Adepo-Gourene B & Pouyand L (1998) Natural hybridization in tilapias. In: *Genetics and Aquaculture in Africa*. Editios I' Institut Francais de recherche scientifique pour le development en cooperation. Collectia colloques et seminares, Paris, 1998.

Agnèse JF; Teugels GG; Galbusera, P; Guyomard, R & Volckaert, F (1997) Morphometric and genetic characterization of sympatric populations of *Clarias gariepinus* and *C. anguillaris* from Senegal. *J. Fish Biol.* 50:1143-1157.

Allendorf, FW; Ryman, N; Utter, FM (1987) Genetics and Fishery management: past present and future. In: *Population genetics and fishery management*. University of Washington Press, Washington, USA, 1-19.

Amos, W (1999) A comparative approach to the study of microsatellite evolution. In: *Microsatellites: Evolution and applications*. Eds. D.B.Goldstein & C.Schlötterer. Oxford University Press, New York.

Armour JAL; Alegre, SA; Miles, S; Williams, LJ & Badge, RM (1999) Minisatellites and mutation processes in tandemly repetitive DNA. In: *Microsatellites: Evolution and applications*. Eds. D.B.Goldstein & C.Schlötterer. Oxford University Press, New York.

Barton, NH & Hewitt, GM (1985) Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.* 16:113-148.

Beardmore, JA; Mair, GC; Lewis, RI (1997) Biodiversity in aquatic systems in relation to aquaculture. *Aquaculture Research* 28:829-839.

Beaumont, MA & Bruford, MW (1999) Microsatellites in conservation genetics. In: *Microsatellites: Evolution and applications*. Eds.

D.B.Goldstein & C.Schlötterer. Oxford University Press, New York.

Beckman, JS & Weber, JL (1992) Survey of human and rat microsatellites. *Genomics* 12:627-631.

Belkhir, K (1999) GENETIX, logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome et Populations, CNRS UPR 9060, Université de Montpellier II, Montpellier (France).

Beveridge, MCM; Ross, LG; Kelly, LA (1994) Aquaculture and biodiversity. *Ambio* 23:497-502.

Bowcock, AM; Ruiz-Linares, A; Fomfohrde, J; *et al.* (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature.* 368:455-457.

Brookfield JFY (1996) A simple new method for estimating null allele frequency from heterozygote defeciency. *Molecular Ecolology.* 5:453-455.

Brownstein, MJ; Carpten, JD & Smith, JR (1996) Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: Primer modifications that facilitate genotyping. *Biotechniques.* 20:1004-1010.

Bruford, MW & Wayne, RK (1993) The use of molecular genetic techniques to address conservation questions. In: *Molecular techniques in environmental in environmental biology* Ed. S.J. Garte. CSC Press, USA.

Brunner, PC; Douglas, MR & Bernatchez, L (1998) Microsatellite and mitochondrial DNA assessment of population structure and stocking effects in Arctic charr *Salvelinus alpinus* (Teleostei: Salmonidae) from central Alpine lakes. *Molecular Ecology.* 7:207-223.

Bruton, MN & Boltt, RE (1975) Aspects of the biology of *Tilapia mossmabica* Peters (Pisces: Cichlidae) in a natural freshwater

lake (Lake Sibaya, South Africa). *Journal of Fish Biology.* 7:423-445.

Callen, DF; Thompson, AD; Shen, Y *et al.* (1993) Incidence and origin of "Null" alleles in the $(AC)_n$ microsatellite markers. *American Journal of Human Genetics.* 52:922-927.

Carrington, M; Marti, D; Wade, J; *et al.* (1999) Microsatellite markers in complex disease: mapping disease-associated regions within the human MHC. In: *Microsatellites: Evolution and applications.* Eds. D.B.Goldstein & C.Schlötterer. Oxford University Press, New York.

Cavalli-Sforza, LL & Edwards, AWF (1967) Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics.* 19:233-257.

Chakraborty R; De Andrade M; Daiger SP; *et al.* (1992) Apparent heterozygous deficiencies observed in DNA typing data and their application in forensic applications. *Annals of Human Genetcs.* 46 45-57.

Chakrabory, R; Kimmel, M; Stivers, DN; *et al.* (1997) Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. *Proc. Natl. Acad. Sci. USA.* 94:1041-1046.

Claridge, MF; Dawah, HA and Wilson, MR (1997) Practical approaches to species concepts for living organisms. In: *Species: The units of Biodiversity.* Eds.: M.F. Claridge, H.A. Dawah and M.R. Wilson. Chapman & Hall, London.

Cockerham, CC & Weir, BS (1993) Estimation of genefolow from F-statistics. *Evolution.* 74:855-863.

Cockerham, CC (1969) Variance of gene frequencies. *Evolution* 23:72-84.

Cockerham, CC (1973) Analysis of gene frequencies. *Genetics* 74:679-700.

Crawford, A; Knappes, SM; Paterson, KA; *et al.* (1998) Microsatellite evolutions: testing the ascertainment bias hypothesis. *Journal of Molecular Evolution*. 46:256-260.

Crow, JF & Kimura, M (1970) *An introduction to population genetics theory.* Harper and Row, New York, Evanston and London.

Daemen, E; Cross, T; Ollevier, F; Volckaert, FAM (2001) Analysis of the genetic structure of European eel (*Anguilla anguilla*) using microsatellite DNA and mtDNA markers. *Marine Biology* 139:755-764.

Deka, R; Chakraborty, R & Ferrel, RE (1991) A population genetic study of six VNTR loci in three ethnically defined populations. *Genomics*. 11:83-92.

De Knijff, P; Dayser, M; Corach, D; *et al.* (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. *International Journal of Legal Medicine*. 110:134-140.

De Woody, JA & Avise, JC (2000) Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of fish biology* 56:461-473.

Dib, C; Faure, S; Fizames, C; *et al.* (1996). A comprehensive map of the human genome based on 5364 microsatellites. *Nature* 380:152-154.

DiRienzo, A; Peterson, AC; Garza, JC; *et al.* (1994) Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences of the USA*. 91:3166-3170.

Dobzhansky, T (1970) *Genetics and the Evolutionary Process.* New York. Columbia Univ. Press.

Dowling, TE; Moritz, C; Plamer, JD; *et al.* (1996) Nucleic acids III: Analysis of fragments and restriction sites. In*:* *Molecular systematics* 2nd ed. Eds. D.M. Hills, C.Moritz, B.K. Mable. Sinauer, USA.

Dunn, G & Everitt, BS (1982) *An introduction to mathematical taxonomy.* Cambridge University Press, Cambridge.

Eisen, JA (1999) Mechanistic basis for microsatellite instability. In: *Microsatellites: Evolution and applications*. Eds. D.B.Goldstein & C.Schlötterer. Oxford University Press, New York.

Eldredge, N (1995) Species, selection, and Paterson's concept of the specific-mate recognition system, in *Speciation and the Recognition Concept* (eds) D.M. Lambert & H.G. Spencer. The John Hopkins University Press, Baltimore, London.

Ellegren, H; Primmer, CR & Sheldon, BC (1995) Microsatellite evolution: directionality or bias in locus selection. *Nature Genetics*. 11:360-362.

Estoup, A & Cornuet, JM (1999) Microsatellite evolution: inferences from population data. In: *Microsatellites: Evolution and applications*. Eds. D.B. Goldstein & C. Schlötterer. Oxford University Press, New York.

Estoup, A; Rousset, F; Michalakis, Y; Cornuet, M; Adriamamga, M & Guyomard, R (1993) Comparative analysis of microsatellite an dallozyme markers: a case study investigating microgeographic differentiation in brown trout (*Salmo trutta*). *Molecular Ecology.* 7:339-353.

Estoup, A; Presa, P; Krieg, F; Vaiman, D & Guyomard, R (1993) $(CT)_n$ and $(GT)_n$ microsatellites: a new class of genetic markers for *Salmo trutta* L. (Brown trout). *Heredity* 71:488-496.

Everitt, BS (1980) *Cluster analysis* 2nd ed. Halsted Press, New York.

FAO Fisheries department, Fishery information, Data and Statistics Unit. (2001). Fishstat plus: Universal software for Fishery Statistical Time series V.2.3 2001.

Falconer, DS (1960) The genetics of litter size in mice. *J. Cell. Comp. Physiol.* 56:153-167.

Felsenstein, J (1988) Phylogenies from molecular sequences: Inference and stability. *Ann.Rev.Genet.* 22:521-565.

Fitzgerald WJ (1979) Red-orange tilapia – A hybrid that could become a world favorite. *Fish Farming Inst.* 6 26-27.

Fleming, IA & Gross, MR (1993). Breeding success of hatchery and wild coho salmon (*Oncorhyncus kisutch*) in competition. *Eclological Applications* 3:230-245.

Foltz, DW (1986) Null alleles as a possible cause of heterozygote deficiencies in the oyster, *Crassostrea virginica* and other bivalves. *Evolution.* 40:869-870.

Gaigher, IG (1973) The habitat preferences of fishes from the Limpopo Rever system, Transvaal and Mozambique. *Koedoe.* 16:103-116.

Garte, SJ (1994) *Molecular techniques in environmental in environmental biology.* CSC Press, USA.

Gaston, KJ & May, RM (1992) The taxonomy of taxonomists. *Nature.* 536:281-281.

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 DB, Linares AR, Cavalli-Sforza LL *et al.* (1995) Genetic absolute dating on microsattelites and the origen of modern humans. *Proc. Nat. Ac. Sci. USA.* 92:6723-6727.

Goldstein, DB & Pollock, DD (1997). Launching Microsatellites: A Review of Mutation Processes and Methods of Phylogenetic Inference. *Journal of Heredity* 88: 335 –342.

Goldstein, DB &.Schlötterer, C (1999) *Microsatellites: Evolution and applications.* Oxford University Press, New York.

Goodman, SJ (1997) Dinucleotide repeat polymorphism at seven anonymous microsattelite loci cloned from the European harbour seal (*Phoca vitulina vitulina*). *Animal Genetcs.* 28 310-311.

Goodisman, SJ; Hidetoshi, BT; Wilson, R (2001) Bottleneck, drift and differentiation: the population structure and demographic history of sika deer (*Cervus nippon*) in the Japanese archipelago. *Molecular Ecology* 10:1357-1370.

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.

Green, J; Corbet, SA & Betney, E (1974) Ecological studies on crater lakes in West Cameroon. The blood of endemic cichlids in Barombi Mbo in relation to stratification and their feeding habits. *J.Zool.Lond.* 170:299-308.

Guo, SW & Thompson, EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 48:361-372.

Hall, EG (2001) An analysis of population structure using microsatellite DNA in twelve Southern African populations of the Mozambique tilapia, *Oreochromis mossambicus* (Peters). M.Sc. thesis, University of Stellenbosch, South Africa.

Hartl, DL & Clark, AG (1989) *Pricipals of population genetics* 2nd ed. Sinnauer. Sunderland, USA.

Hedrick, PW (2000) *Genetics of Populations* Second Edition. Jones and Bartlett Publishers. London. UK.

Heyer, E; Puymirat, F; Dieltjes, P *et al.*(1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Human Molecular Genetics*. 6:799-803.

Hindar, K; Jonsson, B; Rayman, N et al. (1991). Genetic-ralationships among landlocked, resident, and anadromous brown trout, salmo-trutta L. *Hereditas*. 1989 66(Pt1), 83-91.

Iversen, ES (1968) Farming the edge of the sea In: FR Irvine *et al. The fishes and fisheries of the Gold Coast*. Crown Agents, London.

Jeffreys, AJ; Wilson, V & Thein, SL (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature, Lond.* 316:67-73.

Jorde, LB; Bamshad, MJ; Watkins, WS; *et al.* (1995) Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. *American Journal of human Genetics.* 57:523-538.

Kashi, Y & Soller, M (1999) Functional roles of microsatellites and minisatellites. In: *Microsatellites: Evolution and applications.* Eds. D.B.Goldstein & C.Schlötterer. Oxford University Press, New York.

Künzler, P; Matsuo, K & Schaffner, W (1995) Pathological, physiological, and evolutionary aspects of short unstable DNA repeats in the human genome. *Biological Chemistry Hoppe Seyler.* 376:201-211.

Kocher, TD & Carleton, KL (1997) Base substitution in fish mitochondrial DNA: Patterns and Rates. In: *Molecular systematics in fishes*, Eds: Thomas D.K & Carol A.S. Academic Press, London.

Kocher, TD; Lee, WJ; Sobolewska, H; *et al.* (1998). A Genetic Linkage Map of a Cichlid Fish, the Tilapia (*Oreochromis nilocus*). *Genetics.* 148:1225-1232.

Kimura, M & Crow, JF (1964) The Number of Alleles that can be Maintained in a Finite Population. *Genetics* 49: 725 – 738

Kimura, M & Otha, T (1978) Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Science of the USA.* 75:2868-2872.

Lee, W-J & Kocher, TD (1996) Microsatellite DNA markers for genetic mapping in the tilapia, *Oreochromis niloticus. Journal of Fish Biology* 49:169-171.

Le Roux, PJ (1956) Feeding habits of the young of four species of *Tilapia. South African Journal of Science.* 53:33-37.

Lèvêque, C (1997) *Biodiversity dynamics and conservation - the freshwater fish of tropical Africa*. Cambridge University Press, Cambridge.

Liao, I-C. & Chang, S-L. (1983) Studies on the feasibility of red tilapia culture in saline water. In: Fischelson L and Yqron Z. (eds.) *Int. Symp. on Tilapia in aquaculture*. Tel Aviv Univerity Publishers, Tel Aviv, Israel. 524-533.

Linares, AR (1999) Microsatellites and the reconstruction of the history of human populations. In: *Microsatellites: Evolution and applications*. Eds. DB Goldstein & C Schlötterer. Oxford University Press, New York.

Macaranas, JM; Agustin, LQ; Ablan, MCA; *et al.* (1995) Genetic improvement of farmed tilapias: biochemical characterization of strain differences in Nile tilapia. *Aquaculture International* 3:43-54.

Majumdar, KC & McAndrew, BJ (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.

Manly, BFJ (1985) *The statistics of Natural Selection.* Chapman & Hall. London.

Manly, BFJ (1997) *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Chapman & Hall, London.

Mayden, RL (1997) A hierarchy of species concepts: the denouement in the saga of the species problem. In: *Species: The units of Biodiversity*. Eds.: MF Claridge, HA Dawah and MR Wilson. Chapman & Hall, London.

Mayr, E (1942) *Systematics and the origin of species*. Columbia University Press, New York.

Mayr, E (1969) *Principles of systematic zoology*. McGraw-Hill, New York.

McAndrew BJ and Majumdar KC (1983) Tilapia stock identification using electrophoretic markers. *Aquaculture* 30:249-261.

Minch, E; Ruiz-Linares, A; Goldstein, D; *et al.* (1995) MICROSAT (Ver. 1.4d): A computer program for calculating various statistics on microsattelite allele data. www:http://lotka.stanford.edu/distance.html.

Mires D (1977) Theoretical and practical aspects of the production of all male *Tilapia* hybrids. *Bahmidgeh* 29:94-101.

Moralee, RD; Van der Bank, FH & Van der Waal BCW (2000) Biochemical genetic markers to identify hybrids between the endemic *Oreochromis massambicus* and the alien species, *O. niloticus* (Pisces: Cichlidae). *Water SA* 26:263-268.

Morin,PA & Woodruff, DS (1996) Non-invasive genotyping for vertebrate conservation. In: *Molecular genetic approaches in conservation.* Eds. TB Smith & RK Wayne. Oxfort University Press.

Morral, N; Bertranpetit, J; Estivill, X; *et al.* (1994) The origin of the major cystic fibrosis mutation (?F508) in European populations. *Nature Genetics*. 7:169-175.

Mulaik, SA (1972) *The foundations of factor analysis*. McGraw-Hill, New York.

Mullis, K; Faloon, F; Scharf, S; *et al.* (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. *Cold Spring Harbor Sym.quant. Biol.* 51:263-273.

Nagl, S; Herbert, T; Mayer, WE; *et al.* (2001) Classification and phylogenetic relationships of African Tilapiine fishes inferred from mitochondrial DNA sequences. *Molecular phylogenetics and evolution.* 20:361-374.

Nei, M (1972) Genetis distance between populations. *Am. Nat.* 106:283-292.

Nei, M (1973) Ananlysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70:3321-3323.

Nei, M (1975) *Molecular population genetics and evolution*. Amsterdam. North Holland.

Nei, M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583-590.

Nei, M (1987) *Molecular evolutionary Genetics*. Columbia University Press. New York.

Nei, M & Saitou, N (1987) On the maximum likelihood method for molecular phylogeny. *Japanese journal of genetics*. 62:547-548.

Newman, RA & Squire, T (2001) Microsatellite variation and fine-scale population structure in the wood frog (*Rana sylvatica*). *Molecular Ecology* 10:1087-1100.

Paetkau, D & Storbeck, C (1995) The molecular basis and evolutionary history of a microsatellite null allele in bears. *Molecular Ecology* 4:519-520.

Paetkau, D; Waits, LP; Clarkson, PL; *et al.* (1997) An imperial evaluation of genetic distance statistics using microsatellite data from bear (*Ursidae*) populations. *Genetcs.* 147:1943-1957.

Pemberton, JM; Slate, J; Bancroft, JR; *et al.* (1995) Non-amplifying alleles at microsatellite loci – a caution for parentage and population studies. *Molecular Ecology.* 4:249-252.

Perez JE and Maclean N (1976) The haemoglobins of the fish *Sarotherodon mossambicus* (Peters): functional significance and ontogenetic changes. *Journal of Fish Biology.* 9:447-455.

Perez-Lezaun , A; Calafell, F; Mateu, E; *et al.* (1997) Microsatellite variation and the differentiation of modern humans. *Human Genetics.* 99:1-7.

Popma, T & Masser, M (1999) Tilapia: Life history and biology. *SRAC Publication* 283.

Pouyaud L and Agnèse J-F (1995) Phylogenetic relationships between 21 species of three genera *Tilapia, Sarotherodon* and *Oreochromis* using allozyme data. *Journal of Fish Biology.* 47:26-38.

Pritchard, JK; Stephens, M; and Donnelly, P (2000) Inference of population structure using multilocus genotype data. *Genetics.* 155:945-959.

Pullin, RSV (1997) International concerns on diversity and genetic resources management. In: Eds.: Pullin RSV, Casl CMV, Abban EK and Fal TM *Characterisation of Ghanaian Tilapia Genetic Resources for Use in Fisheries and Aquaculture.* ICLRAM *Conf. Porc.* 52 58

Rahman, MH (1995) Experimental pathology of feeding leucaena-leucocephala in rats. *Indian Journal of animal sciences.* 65:1216-1218.

Raymond M and Rousset F (1995) GENEPOP (Ver. 1.2):a Population genetics software for exact tests ecumenicism. *Journal of Hereditary,* 86 248-249.

Rice WR (1989) Analyzing tables of statistical tests. *Evol.* 43(1) 223 225.

Raufaste, N & Banhomme, F (2000) Properties of bias and variance of two multiallelic estimators of F_{ST}. *Theoretical population biogogy.* 57:285-296.

Roberston, A & Hill, WG (1984) Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics.* 107:703-718.

Rognon, X., Teugels, G.G., Guyomard, R., Galbusera, P., Andriamanga, M., Volckaert, F. & Agnèse, J.F. (1998) Morphometric and allozyme variation in the African catfishes *Clarias gariepinus* and *C. anguillaris. Journal of Fish Biology* 53:192-207.

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

Rubinsztein, DC; Amos, W; Leggo, J; *et al* (1995) Microsatellite evolution – evidence for directionality and variation in rate between species. *Nature Genetics*10:337-343.

Saitou, N & Nei, M (1987) The neighbor-joining method: A method for reconstructing phylogenetic trees. *Mol.Biol.Evol.* 4:406-425.

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

Scott, IAW; Hayes, CM; Keogh, JS & Webb, JK (2001) Isolation and characterization of novel microsatellite markers from the Australian tiger snakes (Elapidae: *Notechis*) and amplification in the closely related genus *Hoplecephalus*. *Molecular Ecology notes*. 1:117-119.

Simpson, GG (1961) *Principles of animal taxonomy*. Columbia University Press, New York.

Skelton PH (1993) *A Complete Guide to Freshwater Fishes of Southern Africa*. Southern Book Publishers. 325-330.

Slatk, M & Barton, NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349-1368.

Sokal, RR & Rohlf, FJ (1995) *Biometry: the principles and practice of statistics in biological research* 3rd ed. Freeman & Company, New York.

Sokal, RR & Sneath, PHA (1963) *Principles of Numerical Taxonomy.* Freeman Publications. San Francisco.

Southern, EM (1975) Detection of specific sequences among DNA fragments separtated by gel electrophoresis. *Journal of molecular biology.* 89:503-517.

Stallings, RL; Ford, AF; Nelson, D; *et al.* (1991) Evolution and distribution of $(GT)_n$ repetitive sequences in mammalian genomes. *Genomics*. 10:807-815.

Taberlet, T; Griffin, S; Goossens, B; *et al.* (1996) Reliable genoyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*. 24:3189-3194.

Taberlet, P; Camarra, JJ; Griffin, S; *et al.* (1997) Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Molecular Ecology*. 6:869-876.

Tautz, D (1989) Hypervariablility of simple sequences as a source of ploymorphic DNA markers. *Nucl. Acids. Res.* 17:6463-6471.

Thorpe, JE; Koonce, JF; Borgeson, D; *et al.* (1981) Assessing and managing man's impact on fish genetic resources. *Can. J. Fish. Aquat. Sci*. 38:1899-1907.

Trewavas, E (1983) Tilapiine fishes of the genera *Serathorodon, Oreochromis* and *Danakilia*. British Museum. London.

Van der Mheen, H (1997) Review of introduction and translocation of aquatic species in the Limpopo River system and regional co-operation for policy development. Aquaculture for local development Programme. ALCOM Report No 25, Harare.

Van der Waal BCW (1997) Vreemde kurpers bedreig die bloukurper in die Limpopo. *Stywe Lyne/Tight Lines* (December) 50-51.

Van der Waal BCW and Bills R (1997) *Oreochromis niloticus* in the Limpopo System. *Ichthos* 52:14-16.

Van Treuren, R; Kuittinen, H; Kärkkäinen, K; *et al.* (1997) Evolution of microsatellites in *Arabis petra* and *Arabis lyrata*, outcrossing relatives of *Arabidopsis thaliana. Molecular Biology and Evolution*. 14:220-229.

Vazquez-dominguez, E; Paetkau, D; Tucker, N *et al.* (2001) Resolution of natural groups using iterative assignment tests: and example

from two species of Australian native rats (*Rattus*). *Molecular Ecology*. 10:2069-2078.

Waldman, JR & Wirgin, II (1994) What DNA can do for you. *Fisheries* 19:16-27.

Walhund, S (1928) Compositions of populations from the perspective of the theory of heredity. *Hereditas* 11:65-105.

Weber, JL & Wong, C (1993) Mutation of human short tandem repeats. *Human Molecular Genetics.* 2:1123 – 1128.

Weir BS (1996) *Genetic data analyses II.* Sinauar. Canada.

Weir BS and Cockerham, CC (1984) Estimating F-statistics for the analyses of population structure. *Evolution.* 38:1358-1370.

Welcomme, RL (1967) Observations on the biology of the introduced species of *Tilapia* in Lake Victoria. *Revue Zool. Bot. Afr.* 76: 249-279.

Williams, AGK; Kubelik, AR; Livak, KJ; *et al.* (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18:6531-6535.

Wilson, DS (1992) Complex interactions in metacommunities, with implications for biodiversity and higher levels of selection. *Ecology* 73:1984-2000.

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

Wright, S (1921) Systems of mating. *Genetics* 6:111-178.

Wright, S (1969) *Evolution and the Genetics of Populations. Volume 2. The theory of Gene frequencies.* University of Chicago Press, Chicago.

Zenger,KR & Johnston, PG (2001) Isolation and characterization of microsatellite loci in the southern brown bandicoot (*Isoodon*

obesulus), and their applicability to other marsupial species. *Molecular Ecology Notes.* 1:149-151.

Zhivotovsky, LA & Feldman, MW (1995) Microsatellite variability and genetic distances. *Proceedings of the National Academy of Sciences of the USA*. 92:11549-11552.

Appendix A Genotypes

Values are shown for each individual at each locus. The values indicate the size of the fragment amplified in three digits for the first allele and another three to follow indicating the size of the second allele. Zero indicates no amplification.

Population	Individual	Locus				
		UNH102	UNH129	UNH146	UNH124	UNH192
MWLI	1	149149	199207	116122	293293	128128
MWLI	2	149149	199207	116122	293293	128128
MWLI	3	149149	199207	116122	293293	128128
MWLI	4	149157	207207	116116	0	134134
MWLI	5	149157	199207	116122	293299	128128
MWLI	6	149157	207207	122124	0	0
MWLI	7	149149	199207	116122	293293	128128
MWLI	8	149157	207207	116122	0	128128
MWLI	9	149157	207207	116122	301301	128128
MWLI	10	149157	207207	122124	301309	134136
MWLI	11	149157	207207	122124	301301	134136
MWLI	12	157157	207207	116116	293299	128128
MWLI	13	157157	207207	124124	301309	136152
MWLI	14	157157	207207	116122	301309	140152
MWLI	15	149157	199199	116122	293293	128128
MWLI	16	149157	199199	116122	293301	128128
MWLI	17	149149	199207	116122	293301	128128
MWLI	18	149153	201201	116116	301301	128128
MWLI	19	149149	199199	116116	301301	128128
MWLI	20	149149	207207	116122	293301	128128
MWLI	21	151151	201201	116122	293293	128128
MWBO	22	151153	199199	116122	293293	128128
MWBO	23	153153	199207	116122	293293	128128
MWBO	24	151153	199207	116122	293293	128128
MWBO	25	151153	0	116122	293305	130130
MWBO	26	161161	207207	116122	301301	128132
MWBO	27	149149	199207	116122	303303	128128
MWBO	28	151153	199207	116116	303303	128128
MWBO	29	151153	199207	116122	301305	130130
MWBO	30	151153	0	116116	293293	128128
MWBO	31	151153	199207	116116	293293	128128
MWND	32	151153	207207	122122	293301	130130
MWND	33	151153	207207	122122	301305	130130

MWND	34	0	207207	122122	301305	130130
MWND	35	151151	0	122122	301305	130130
MWND	36	0	207207	122124	293301	130130
MWND	37	151153	199207	122124	301305	130130
MWND	38	151151	207207	116122	301305	130130
MWND	39	151151	199199	122122	301305	130130
MWND	40	151151	207207	122124	293301	130130
MILE	41	151151	207207	116122	293301	130130
MILE	42	157157	207207	116122	301301	132162
MILE	43	0	207207	122124	0	128128
MILE	44	0	207207	122122	0	130130
MILE	45	151161	199207	116122	301301	130130
MILE	46	149149	199207	116122	299299	128128
MILE	47	0	0	122124	301301	128128
MWKA	48	157161	207207	122122	301301	128128
MWKA	49	151157	207207	116122	301301	128132
MWKA	50	149149	0	116116	293301	0
MWKA	51	0	0	0	0	130130
MWKA	52	151151	207207	122124	301301	130130
MWKA	53	151151	207207	116122	299301	128130
MWKA	54	151151	207207	116122	299299	130132
MWKA	55	161165	201207	116122	301301	132132
MWKA	56	153157	207207	116122	0	130132
MWKA	57	151161	201207	116122	299301	130130
MWKA	58	151161	201207	116122	299301	130130
MWKA	59	151165	201207	116122	299301	130132
MWKA	60	153153	199199	0	293293	130130
MWKA	61	161165	201207	116122	301305	132132
HWLI	62	159159	201207	122124	301305	134134
HWLI	63	157157	207207	122122	0	128128
HWLI	64	157157	207207	122124	0	128128
HWLI	65	0	207207	0	305315	130134
HWLI	66	0	0	124124	303303	132134
HWLI	67	0	0	0	0	128132
HWLI	68	157157	207207	122122	293305	128134
NFTH	69	157157	207207	122124	301301	136152
NFTH	70	149157	207207	124124	301309	134136
NFTH	71	157157	207207	116124	301301	134136
NFTH	72	157157	207207	124126	305309	136140
NFTH	73	157157	207207	124124	303309	136152
NFTH	74	157157	207215	116124	303309	136140
NFTH	75	157157	207207	124124	303309	140152
NFTH	76	0	0	0	299299	134134
NFTH	77	157157	207207	124124	301309	134140
NFTH	78	157157	0	124126	305309	134140
NFTH	79	157157	207207	124124	301305	136140

NFTH	80	149157	207207	124124	301301	140152
NFTH	81	157157	207207	124124	301301	134134
NFTH	82	157157	207207	124126	305309	136140
NFTH	83	0	0	0	0	136152
NFTH	84	157157	207207	124124	301309	140152
NFTH	85	157157	207207	124126	301309	136140
NFTH	86	157157	207207	122124	301309	140152
NFTH	87	0	207207	0	309309	134140
NFTH	88	157157	207207	124124	0	132132
NFPH	89	157157	207207	124126	307315	130130
NFPH	90	157157	207215	124124	315315	130162
NFPH	91	157157	207215	124124	0	130162
NFPH	92	0	207207	124126	307307	130162
NFPH	93	157157	207215	124124	0	130162
NFPH	94	0	207215	124124	315315	130162
NFPH	95	0	207215	124124	0	130162
NFPH	96	0	207207	124124	315315	130162
NFPH	97	0	0	124124	0	130162
NFPH	98	0	207207	124124	315315	130162
NFPH	99	0	207207	124124	315315	130162
NFPH	100	0	207207	124124	0	130162
NFPH	101	157157	207207	124126	307315	130134
NFPH	102	157157	207207	124124	0	130162
NFPH	103	157157	207215	124126	307307	130162
NFPH	104	0	207207	124124	0	130162
NFPH	105	0	207215	124124	0	130162
NFPH	106	157157	207215	124126	315315	130162
NFPH	107	157157	207207	124126	0	130162
NFPH	108	157157	207215	124126	307307	130162
NWE1	109	157157	195207	124124	305309	130136
NWE1	110	153157	207207	120120	309313	134136
NWE1	111	157157	207207	124126	303303	134136
NWE1	112	157157	207215	124124	295301	134136
NWE1	113	0	207207	0	0	0
NWE1	114	157157	207215	120126	295309	134140
NWE1	115	0	0	0	0	0
NWE1	116	157161	207215	120124	303305	136136
NWE1	117	0	0	124126	295305	136136
NWE1	118	0	0	124124	301305	134138
NWE1	119	0	207207	124126	0	130136
NWE1	120	0	0	124124	0	136136
NWE1	121	0	207207	120124	293305	134136
NWE1	122	0	0	0	0	130130
NWE1	123	153153	201207	124126	295303	130130
NWE1	124	149157	207207	124132	0	136136
NWE1	125	149157	207215	118124	0	0

NWE1	126	157157	195207	0	0	136136
NWE1	127	157157	207207	0	0	0
NWE1	128	149157	207215	120124	305309	136136
NWE1	129	157157	207215	0	313313	136136
NWE2	130	157157	201207	120120	295301	130136
NWE2	131	157157	207215	126126	295309	134140
NWE2	132	157157	207215	124126	305317	134134
NWE2	133	157157	201201	126126	305305	136136
NWE2	134	157157	207207	126126	309309	136140
NWE2	135	157157	207207	120124	301313	136136
NWE2	136	157157	199199	126126	295307	136140
NWE2	137	157157	207207	118120	295295	134136
NWE2	138	157157	207207	124124	305313	136136
NWE2	139	157157	209209	120124	301313	130136
NWE2	140	157157	207207	124124	295313	136136
NWE2	141	157157	207215	124124	305313	134136
NWE2	142	157157	207215	124124	313313	136138
NWE2	143	157157	207207	120126	295295	134136
NWE2	144	157157	207207	124126	295295	136138
NWE2	145	0	201201	126126	295295	136136

Appendix B Allele frequencies

Locus: UNH102

Population	Allele						
	1	2	3	4	5	6	7
MWLI	0.571	0.048	0.024	0.357	0.000	0.000	0.000
MWBO	0.100	0.350	0.450	0.000	0.000	0.100	0.000
MWND	0.000	0.786	0.214	0.000	0.000	0.000	0.000
MILE	0.250	0.375	0.000	0.250	0.000	0.125	0.000
MWKA	0.077	0.385	0.115	0.115	0.000	0.192	0.115
HWLI	0.000	0.000	0.000	0.750	0.250	0.000	0.000
NFTH	0.059	0.000	0.000	0.941	0.000	0.000	0.000
NFPH	0.000	0.000	0.000	1.000	0.000	0.000	0.000
NWE1	0.115	0.000	0.115	0.731	0.000	0.038	0.000
NWE2	0.000	0.000	0.000	1.000	0.000	0.000	0.000

Locus: UNH129

Population	Alllele					
	1	2	3	4	5	6
MWLI	0.000	0.286	0.095	0.619	0.000	0.000
MWBO	0.000	0.500	0.000	0.500	0.000	0.000
MWND	0.000	0.188	0.000	0.813	0.000	0.000
MILE	0.000	0.167	0.000	0.833	0.000	0.000
MWKA	0.000	0.083	0.208	0.708	0.000	0.000
HWLI	0.000	0.000	0.100	0.900	0.000	0.000
NFTH	0.000	0.000	0.000	0.971	0.000	0.029
NFPH	0.000	0.000	0.000	0.763	0.000	0.237
NWE1	0.063	0.000	0.031	0.719	0.000	0.188
NWE2	0.000	0.063	0.156	0.594	0.063	0.125

Locus: UNH146

Population	Allele						
	1	2	3	4	5	6	7
MWLI	0.500	0.000	0.000	0.381	0.119	0.000	0.000
MWBO	0.650	0.000	0.000	0.350	0.000	0.000	0.000
MWND	0.056	0.000	0.000	0.778	0.167	0.000	0.000
MILE	0.286	0.000	0.000	0.571	0.143	0.000	0.000
MWKA	0.458	0.000	0.000	0.500	0.042	0.000	0.000
HWLI	0.000	0.000	0.000	0.600	0.400	0.000	0.000
NFTH	0.059	0.000	0.000	0.059	0.765	0.118	0.000
NFPH	0.000	0.000	0.000	0.000	0.825	0.175	0.000
NWE1	0.000	0.033	0.200	0.000	0.567	0.167	0.033
NWE2	0.000	0.031	0.188	0.000	0.375	0.406	0.000

Locus: UNH124

Popula -tion	Allele										
	1	2	3	4	5	6	7	8	9	10	11
MWLI	0.472	0.000	0.056	0.389	0.000	0.000	0.000	0.083	0.000	0.000	0.000
MWBO	0.550	0.000	0.000	0.150	0.200	0.100	0.000	0.000	0.000	0.000	0.000
MWND	0.167	0.000	0.000	0.500	0.000	0.333	0.000	0.000	0.000	0.000	0.000
MILE	0.100	0.000	0.200	0.700	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MWKA	0.125	0.000	0.250	0.583	0.000	0.042	0.000	0.000	0.000	0.000	0.000
HWLI	0.125	0.000	0.000	0.125	0.250	0.375	0.000	0.000	0.000	0.125	0.000
NFTH	0.000	0.000	0.056	0.389	0.083	0.111	0.000	0.361	0.000	0.000	0.000
NFPH	0.000	0.000	0.000	0.000	0.000	0.000	0.364	0.000	0.000	0.636	0.000
NWE1	0.042	0.167	0.000	0.083	0.167	0.250	0.000	0.167	0.125	0.000	0.000
NWE2	0.000	0.375	0.000	0.094	0.000	0.156	0.031	0.094	0.219	0.000	0.031

Locus: UNH192

Population	Allele								
	1	2	3	4	5	6	7	8	9
MWLI	0.750	0.000	0.000	0.100	0.075	0.000	0.025	0.050	0.000
MWBO	0.750	0.200	0.050	0.000	0.000	0.000	0.000	0.000	0.000
MWND	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
MILE	0.429	0.429	0.071	0.000	0.000	0.000	0.000	0.000	0.071
MWKA	0.154	0.538	0.308	0.000	0.000	0.000	0.000	0.000	0.000
HWLI	0.429	0.071	0.143	0.357	0.000	0.000	0.000	0.000	0.000
NFTH	0.000	0.000	0.050	0.225	0.250	0.000	0.300	0.175	0.000
NFPH	0.000	0.525	0.000	0.025	0.000	0.000	0.000	0.000	0.450
NWE1	0.000	0.176	0.000	0.176	0.588	0.029	0.029	0.000	0.000
NWE2	0.000	0.063	0.000	0.188	0.594	0.063	0.094	0.000	0.000

Appendix C Assignment of individuals to species

Individual	Species	% assignment to *O. mossambiucs*	% assignment to Hybrids	% assignment to *O. niloticus*
1	1	0.0049	0.9898	0.0053
2	1	0.0057	0.9880	0.0064
3	1	0.0042	0.9900	0.0058
4	1	0.0223	0.2549	0.7229
5	1	0.0086	0.9778	0.0136
6	1	0.1251	0.2942	0.5806
7	1	0.0054	0.9907	0.0039
8	1	0.0156	0.9650	0.0194
9	1	0.0082	0.9784	0.0133
10	1	0.0111	0.0260	0.9630
11	1	0.0154	0.0603	0.9243
12	1	0.0325	0.9402	0.0273
13	1	0.0233	0.0078	0.9689
14	1	0.0057	0.0792	0.9152
15	1	0.0120	0.9803	0.0077
16	1	0.0083	0.9811	0.0106
17	1	0.0052	0.9915	0.0033
18	1	0.0041	0.9897	0.0062
19	1	0.0039	0.9918	0.0043
20	1	0.0067	0.9887	0.0046
21	1	0.0049	0.9904	0.0046
22	1	0.0048	0.9903	0.0049
23	1	0.0035	0.9921	0.0044
24	1	0.0050	0.9885	0.0065
25	1	0.0092	0.9841	0.0067
26	1	0.0070	0.9889	0.0040
27	1	0.0116	0.9760	0.0124
28	1	0.0093	0.9804	0.0103
29	1	0.0073	0.9878	0.0049
30	1	0.0068	0.9885	0.0048
31	1	0.0044	0.9899	0.0057
32	1	0.0079	0.9845	0.0076
33	1	0.0131	0.9777	0.0092
34	1	0.0281	0.9562	0.0157
35	1	0.0122	0.9820	0.0059
36	1	0.1123	0.8524	0.0353
37	1	0.0381	0.9437	0.0182
38	1	0.0066	0.9859	0.0075
39	1	0.0062	0.9892	0.0046
40	1	0.0395	0.9514	0.0091
41	1	0.0057	0.9894	0.0049
42	1	0.2131	0.7107	0.0761
43	1	0.0592	0.8980	0.0429
44	1	0.0284	0.9611	0.0106
45	1	0.0080	0.9860	0.0060

46	1	0.0046	0.9909	0.0045
47	1	0.0183	0.9561	0.0256
48	1	0.0089	0.9765	0.0146
49	1	0.0115	0.9759	0.0126
50	1	0.0082	0.9835	0.0083
51	1	0.5649	0.4197	0.0154
52	1	0.0442	0.9218	0.0340
53	1	0.0079	0.9857	0.0065
54	1	0.0081	0.9852	0.0067
55	1	0.0061	0.9882	0.0057
56	1	0.0479	0.9395	0.0126
57	1	0.0064	0.9878	0.0057
58	1	0.0045	0.9922	0.0034
59	1	0.0068	0.9876	0.0056
60	1	0.0100	0.9846	0.0055
61	1	0.0075	0.9837	0.0088
62	2	0.0315	0.0244	0.9440
63	2	0.0521	0.9121	0.0359
64	2	0.1662	0.6243	0.2096
65	2	0.8927	0.0355	0.0718
66	2	0.6379	0.0135	0.3486
67	2	0.0577	0.9221	0.0202
68	2	0.0522	0.7207	0.2271
69	3	0.0062	0.0407	0.9532
70	3	0.0171	0.0103	0.9727
71	3	0.0138	0.0435	0.9427
72	3	0.0090	0.0078	0.9832
73	3	0.0343	0.0068	0.9589
74	3	0.0141	0.0102	0.9757
75	3	0.0581	0.0044	0.9375
76	3	0.1618	0.2544	0.5837
77	3	0.0086	0.0056	0.9859
78	3	0.0251	0.0047	0.9702
79	3	0.0151	0.0078	0.9770
80	3	0.0126	0.0194	0.9680
81	3	0.0138	0.0064	0.9798
82	3	0.0085	0.0056	0.9860
83	3	0.0520	0.0251	0.9230
84	3	0.0134	0.0078	0.9788
85	3	0.0172	0.0077	0.9750
86	3	0.0101	0.0129	0.9770
87	3	0.0082	0.0080	0.9838
88	3	0.7562	0.0348	0.2090
89	3	0.9798	0.0041	0.0160
90	3	0.9832	0.0063	0.0106
91	3	0.9731	0.0075	0.0193
92	3	0.9842	0.0068	0.0089
93	3	0.9743	0.0081	0.0176
94	3	0.9874	0.0059	0.0066
95	3	0.9783	0.0103	0.0114
96	3	0.9805	0.0100	0.0095
97	3	0.9689	0.0119	0.0192
98	3	0.9826	0.0085	0.0088
99	3	0.9844	0.0076	0.0080
100	3	0.9608	0.0092	0.0301
101	3	0.9789	0.0066	0.0145

102	3	0.9666	0.0151	0.0183
103	3	0.9878	0.0035	0.0087
104	3	0.9741	0.0101	0.0158
105	3	0.9816	0.0094	0.0090
106	3	0.9893	0.0037	0.0070
107	3	0.9805	0.0058	0.0138
108	3	0.9811	0.0075	0.0114
109	3	0.1292	0.0111	0.8597
110	3	0.0126	0.0118	0.9757
111	3	0.0570	0.0085	0.9345
112	3	0.0165	0.0055	0.9780
113	3	0.3491	0.2452	0.4057
114	3	0.0112	0.0050	0.9838
115	3	0.3191	0.3401	0.3408
116	3	0.0446	0.0086	0.9468
117	3	0.0182	0.0064	0.9753
118	3	0.0421	0.0149	0.9430
119	3	0.4419	0.0228	0.5353
120	3	0.0449	0.0077	0.9474
121	3	0.0185	0.0469	0.9346
122	3	0.5950	0.3797	0.0254
123	3	0.7168	0.1754	0.1079
124	3	0.0506	0.0120	0.9374
125	3	0.3768	0.0265	0.5967
126	3	0.0194	0.0096	0.9710
127	3	0.4992	0.0315	0.4693
128	3	0.0068	0.0062	0.9870
129	3	0.0117	0.0057	0.9827
130	3	0.0289	0.0283	0.9428
131	3	0.0266	0.0048	0.9686
132	3	0.2052	0.0045	0.7903
133	3	0.0071	0.0043	0.9886
134	3	0.0078	0.0038	0.9884
135	3	0.0113	0.0055	0.9832
136	3	0.2114	0.0685	0.7201
137	3	0.0093	0.0070	0.9838
138	3	0.0087	0.0047	0.9866
139	3	0.0508	0.0186	0.9307
140	3	0.0070	0.0064	0.9866
141	3	0.0250	0.0044	0.9706
142	3	0.0295	0.0057	0.9648
143	3	0.0112	0.0046	0.9842
144	3	0.0156	0.0041	0.9803
145	3	0.0105	0.0054	0.9840