Genetic connectivity of fish in the Western Indian Ocean

Oskar Henriksson

Department of Zoology
Stockholm University
“If you really think the environment is less important than the economy try holding your breath while you count your money.”

—Dr. Guy McPherson
ABSTRACT

An almost unbroken fringing reef runs along the east coast of Africa, the lagoon inside the reef is the foundation of almost all artisanal fisheries. It is a low-tech fishery conducted by many people. Some areas can have up to 19 fishermen per square kilometer. High fishing pressures, coupled with declining fish stocks has led to changes in mean size and reproductive age of many exploited species. There is clearly a vital and urgent need for scientifically based management systems, including the utilization of genetic information to guide management practices.

This thesis aims to investigate the presence of genetic structures in fish in the Western Indian Ocean. In order to do that we first investigated the historical patterns of connectivity throughout the region (paper I). In papers II and III we focused on local scale connectivity in Kenya and Tanzania and finally in paper IV we investigate the large-scale contemporary gene flow throughout the Western Indian Ocean. In paper III we also investigate the temporal genetic variation at one site and compare it to the small-scale genetic variation along a stretch of the Kenyan coastline. Some overall conclusions that can be drawn from my body of work are: there are genetic structures present in the Western Indian Ocean even though the apparent lack of physical barriers. Major oceanic currents aid evolutionary dispersal patterns. A single geographic site need not be genetically homogenous or temporally stable. Island sites are genetically more homogenous than mainland sites.

In conclusion, there are clear and distinct genetic structures present especially in *Siganus sutor*, the most targeted fish for the artisanal fishery in East Africa.

**Keywords:** population genetics, Indian Ocean, *Siganus sutor*, *Valamugil buchanani*, *Scarus ghobban*, connectivity, aflp, mtDNA, d-loop, CO1
## CONTENTS

Abstract........................................................................................................................................... 5
List of papers ...................................................................................................................................... 9
Introduction ...................................................................................................................................... 11
  State of affairs ................................................................................................................................. 11
  Genetic population structure ........................................................................................................... 12
  Genetics and management ............................................................................................................... 13
Scope and aim of this thesis .............................................................................................................. 15
  Study region ................................................................................................................................... 15
  Sampling methods ........................................................................................................................... 16
  Study organisms ............................................................................................................................... 16
    Scarus ghobban ............................................................................................................................... 16
    Mullets (Mugilidae) ....................................................................................................................... 17
    Siganus sutor ................................................................................................................................. 17
  Methods used in the thesis ............................................................................................................... 18
Statistics ........................................................................................................................................... 20
  AFLP statistics ................................................................................................................................. 21
  DNA sequence statistics .................................................................................................................. 24
Combining mtDNA-sequencing and AFLP to speed up and minimize costs in genetic species identification ............................................................................................................................... 25
Summary of the thesis research papers and major findings ............................................................. 28
  paper I Genetic connectivity and historical demography of the Blue barred parrotfish (Scarus ghobban) in the western Indian Ocean (2011) ........................................................................ 28
  paper II Genetic identification and population structure of juvenile mullet (Mugilidae) collected for aquaculture in East Africa. (2012) ........................................................................ 29
  paper III Temporal genetic variability of landed Siganus sutor reveals a mixed stock fishery in coastal Kenya (Manuscript) .................................................................................................................. 29
  paper IV Contrasting population genetic structure of Siganus sutor between mainland coastal and oceanic island populations (Manuscript) ................................................................. 30
Discussion ......................................................................................................................................... 31
References ......................................................................................................................................... 36
Populärvetenskaplig sammanfattning ............................................................................................... 43
Acknowledgements ............................................................................................................................ 47
LIST OF PAPERS

This thesis is based on the following papers, which are referred to by their roman numerals


II. Henriksson, Mwandya, Gullström, Thorberg, Grahn (2012) Genetic identification and population structure of juvenile mullet (Mugilidae) collected for aquaculture in East Africa. Western Indian Ocean J. Mar. Sci. 11:41-54 +

III. Henriksson, Larsson, Grahn (Manuscript) Temporal genetic variability of landed Siganus sutor reveals a mixed stock fishery in coastal Kenya

IV. Henriksson, Grahn (Manuscript) Contrasting population genetic structure of Siganus sutor between mainland coastal and oceanic island populations

Papers not included in this thesis


* Reprinted with the kind permission of Springer. License number 3207690487355

+ Reprinted with kind permission of the publisher under creative commons license
INTRODUCTION

STATE OF AFFAIRS

In developing countries coral reef fisheries provide up to 25% of the total harvest of finned fish (Souter and Lindén 2000) as well as a large proportion of other resources such as clams, octopus, sea urchins and income from tourism. Intense harvesting of marine resources has led to local and in some cases, global extinction of several species of marine organisms and it is predicted that the majority of marine fish stocks will be close to complete depletion by 2050 (Worm et al. 2006). High fishing pressures, coupled with declining fish stocks has led to changes in mean size and reproductive age of many exploited species (Kamukuru et al. 2005; Yemane et al. 2004). Coral Reefs worldwide are under threat from habitat destruction and degradation from wide-ranging causes, including climate-change-associated coral bleaching (Glynn 1983; Brown 1997; Douglas 2003; Hoegh-Guldberg 1999) ocean acidification (Veron 2008), over-fishing (Valentine and Heck Jr. 2005), pollution (Dubinsky and Stambler 1996) and coastal development (Rogers 1990). The dramatic decline of coral reef-associated fisheries has been linked to the deterioration in the general health and condition of coral reefs (Graham et al. 2005; Jacobsen et al. 2001). These problems are particularly pronounced in the Western Indian Ocean, where a large share of the world’s coral reefs are located, and where traditional fisheries are often concentrated around coral reefs. There is a vital and urgent need for scientifically based management systems, including the utilization of genetic information to guide management practices.

Previously marine populations where often considered to be panmictic (Hauser and Carvalho 2008) i.e. everybody has the same chance to mate with each other, the boundlessness of the open ocean was thought to hold no real barriers to restrict gene-flow. Since most fish move very little in their adult stage larval dispersal is thought to be the greatest contributor to gene-flow between populations. Larvae was thought to passively drift with oceans currents and settle when their energy resources where depleted or they have gone through metamorphosis. Larval duration period (LDP) was thought to correspond to migration distances (Hauser and Carvalho 2008). There are some imminent issues to this hypothesis that recent research has shed light on; pacificity of larvae, dilution factor of freshly spawned
eggs, suitable habitat, major ocean currents are typically offshore, spawning site preferences of adult fish as well as marine barriers. Panmictic marine populations exist but are mostly limited to large solitary pelagic fish such as swordfish (Muths et al. 2009). There is a growing body of scientific literature that disputes the passive nature of larvae dispersal for example, larvae have been found to adhere to chemical cues in the water i.e. smell their home reef (Gerlasch et al. 2006). And self-recruitment of larvae for Pomacentrus amboinensis has been estimated to range between 15-60% when chemically tagging embryos (Jones et al. 1999).

**GENETIC POPULATION STRUCTURE**

A genetic population is "a community of individuals of a sexually reproducing species within which mating takes place" (Dobzhansky 1970). As opposed to a traditional population that can be any number of individuals one chooses to group together based on a certain criteria e.g. all animals inhabiting a certain area or all fish caught in an area. Genetic structure are caused by, among other things, differences in mating success. When a number of individuals have a higher than average chance of mating with each other than with the general populous, genetic structures emerge. This is called non-random mating, and is one of the corner-stones in the field of population genetics. Random mating and infinite population sizes are the fundamental assumptions and the base of the Hardy Weinberg (HW) equilibrium. A population is in HW equilibrium when there is no selection acting upon it, no genetic drift present, random mating exists, mutations happen somewhere else and the populations are infinitely large. When comparing two hypothetical populations the assumption is made that both the compared populations are in HW equilibrium within each other, but not between each other. By making this assumption it is possible to compare the allele frequencies between the populations and estimate how much they differ from each other, this estimate is called fixation index or Fst (Holderegger et al. 2006; Wright 1950).

Reproductive isolation is an important factor in creating genetic structure, reproductive isolation can occur on different time scale such as reproductive isolation due to lowering of sea level and forcing the animals into glacial refugias. If populations are kept isolated long enough genetic mutations and genetic drift will start to differentiate the populations, when the populations then reconnect differences attributed to drift quickly disappear while the different mutations accumulated during the separation are spread through both of the
populations. Even if the population is in complete panmixia for the next millennia modern genetic tools can detect these mutations (Horn and van Herwerden 2013).

**GENETICS AND MANAGEMENT**

“Recent history shows that natural renewable resources -- such as fishery stocks -- are depleted in the absence of effective governance as soon as the demand outstrips the biological capacity of sustaining the particular fish stock.” (FAO 2001-2013)

The overall aim of fisheries management is to ensure the sustainability of the fisheries. To ensure sustainability of the fisheries methods such as maximum sustainable yield, population growth, fishing effort, and stock assessment have been suggested and extensively used. Some common tools available for managers are gear restrictions, creation of no take zones (MPAs most often), yield quotas, and limiting of the fishing fleet (McClanahan and Mangi 2004).

Managing fisheries without prior knowledge of the genetic fish stock composition can put the fishery at a risk (Arnason et al. 2009), hence knowledge about the genetic stock composition of a species is an essential tool for proper sustainable management. An example is the nearly extinct common skate *Dipturus batus* that has always been considered a single species and has been managed as such, molecular markers has identified that the skate is actually two different species that should be managed separately (Griffiths, Sims et al. 2010). By using high resolution genetic markers it is possible to determine the number and extent of genetic populations in a certain area (Bay et al. 2008; Lin et al. 2009) with this information conclusions about gene flow can be made. This is especially important, since knowing the location from where a harvested, degraded or depleted population recruits, allows for management or protection of the source population, with benefits downstream. This type of reasoning has stimulated a number of studies on genetic connectivity and its implications for the design, i.e. number, size and placement, of marine protected areas (MPAs) (Palumbi 2003). Genetic data can also be used to determine stock composition in harvested species as, in species with strong genetic structure, intense harvest of a local population may lead to the collapse of genetically distinct and locally adapted stocks (Hutchinson 2008). Genetic structures have been found in several commercially important fish species on spatial scales ranging from tens of kilometers to ocean-wide homogenous
populations with important consequences for management of harvested stocks (Laikre et al. 2005).

One issue in marine management is poorly defined management units which stems from an inherent lack of baseline data including information regarding dispersal and connectivity of key species and populations. One way to obtain this information is through genetic studies, and therefore they should be an essential tool for the conservation of coral reef ecosystems (Baums 2008; Jones, Almany et al. 2009). Information from such studies provides estimates of genetic diversity, a measure of an organism’s adaptive potential and ability to survive in the face of environmental change. Moreover, since marine propagules are almost impossible to physically monitor, genetic markers can track dispersal of larvae throughout a vast geographic area, providing an indication of how distant reefs are ‘connected’. This enables a measurement of genetic structure as well as contemporary gene flow. A common assumption is that marine species with pelagic larval stages show genetic uniformity over a wide geographic range. However an increasing number of studies determine strong genetic differentiation among populations from different geographic regions, indicating that larval dispersal can be more limited than we believe (Benzie et al. 1994; Barber, Palumbi et al. 2000) and it seems that several ecological and behavioral factors influence the genetic structure of fish populations. The most important factors are; homing behavior (Gerlach et al. 2007), timing of reproduction (Selkoe et al. 2006), habitat specialization (Knudsen et al. 2006), larval dispersal capabilities and effective versus census size of the populations. Marine organisms can have an open, closed or continuous population structure (Knutsen et al. 2003; Laikre et al. 2005; Dorenbosch et al. 2006; Sonstebo et al. 2007), with spawning mode, kinship schooling as well as the ability to return to the same spawning site as contributing factors. On a local scale the term “chaotic genetic patchiness” (Johnson and Black 1982) describes the stochastic effects of all of the above mentioned factors. For example, it has been shown that juvenile fish are recruited back to their birth site, despite the ability to disperse, making the local populations genetically distinct from each other (Planes et al. 2002).

In East Africa management has traditionally been focused on ecosystem protection rather than sustainable fishing. A number of MPAs have been set up and more are being established (Chircop et al. 2010). Population genetic information can be a valuable tool
when it comes to the design and management of a well functioning MPA, since the aid of molecular methods can answer questions that are difficult or even impossible to investigate using ecological methodology. Molecular ecology has been introduced as a promising tool for the effective design of MPAs worldwide (Gaines et al. 2010). By estimating the correlation between genetics and geography within a region it is possible to assess the number and extent of different stocks present within a species. Knowing the extent and size of a single stock is of great importance as the area inhabited by a single stock can be much larger than the area protected, or the spawning grounds of a stock can be located outside the protected area and thus be subject to intense fishing pressure. The Masai-Mara/Serengeti national parks (Kenya and Tanzania) are a good parallel; if only one country had chosen to protect the ecosystem the great migration would not properly have been protected and a steady decline in for example wildebeest population size would have been expected due to hunting or change in land use.

**SCOPE AND AIM OF THIS THESIS**

The aims of this thesis was to investigate the genetic connectivity of fish in the Western Indian Ocean (WIO). We have also tried to illustrate how our findings can aid the local management of the artisanal fisheries. We addressed the overall aim by examining the evolutionary phylogeography of *Scarus ghobban*, the small scale genetic variation was examined using *Valamugil Buchanan* and temporal genetic variation within a site as well as large scale genetic variation was examined in *Siganus sutor*. Two different genetic marker systems were used, mitochondrial DNA (mtDNA) and Amplified fragment length polymorphism (AFLP).

**STUDY REGION**

The status of artisanal marine fisheries in Kenya and Tanzania is varied; however most show signs of overexploitation (Jiddawi and Öhman 2002), the catch per unit effort has steadily declined since 1985 (Mkenda and Folmer 2001). In Seychelles many species are considered to be overfished (Seychelles fishing authorities, 2006). In Mauritius, fish catch from the artisanal fishery has been in constant decline since the 1980s and the artisanal fishermen
have been given incentives such as duty concessions on outboard motors and soft loans, to encourage them to enter the off lagoon fishery (Annual report 1999, Ministry of Fisheries and Co-operatives, Mauritius). The driving forces of overfishing in artisanal fisheries are; the lack of compliance and weak enforcement of management regulations (Cinner et al. 2006), as well as poorly defined management units (Cinner et al. 2009). For example, in East Africa the artisanal fisheries has so far been managed with gear restriction and fishing exclusion zones (MPAs) (McClanahan and Mangi 2004). Gear restrictions are typically made to limit the efficiency of the individual fisherman’s catch. The gear restriction includes a minimum mesh size for nets and a minimum size of the individual fishes caught so that all fish caught are mature and have spawned at least once. However, the enforcement of the mesh size is in reality very weak in East Africa and especially in Kenya (Mangi and Roberts 2006).

**Sampling Methods**

Samples collected for papers I, III and IV were all collected at local fish landing sites when fishermen were bringing in their catch of the day. Local fisherman seldom fish far from where the fish is landed i.e. fish collected at local landing sites are caught within a 1-5km radius (Monywoki et al. 2007). For paper II schooling fish were caught during low tide using a seine net. Each haul swept an area of approximately 170 m2. All fish collected were juveniles.

**Study Organisms**

*Scarus ghobban*

*S. ghobban* is a mass-spawning reef fish with a pelagic larval phase (Leis and Carson-Ewart 2000). It has a complex social structure, and is a sequential hermaphrodite (Allsop and West 2003). The fish is widespread and abundant in the Indo-Pacific Ocean region with moderate-to-high economic value to the artisanal fishery in countries from the WIO region (Sousa and Dias 1981) in which the study was carried out. Tailfin clips of *S. ghobban* were obtained from artisanal fishermen at local fish landing sites.
**MULLETS (MUGILIDAE)**

Grey mullets are distributed worldwide from approximately 42°S to almost 51°N where they inhabit estuarine, intertidal, freshwater and coastal marine habitats (Odum, 1970; Ross, 2001). Reproductive patterns in grey mullet involve migration from shallow coastal habitats to offshore waters where spawning takes place in large schools. Thereafter, larvae and juveniles migrate to inshore environments where they inhabit shallow intertidal habitats such as mangrove creeks (Odum, 1970; Saleh, 2008). Grey mullet is considered to be isochronal spawners, characterized by synchronous gamete development and spawning of all eggs at once or in batches within successive nights (Render et al. 1995). The mullets (Family Mugilidae) are important in commercial and subsistence fisheries in many parts of the world (FAO, 2000; Ross, 2001) and, because of their high tolerance to environmental change, they have a great potential for aquaculture in many countries (Oren, 1981; Lee and Menu, 1981; Pillay and Kutty, 2005). They constitute priority species for marine aquaculture development in East Africa (Mmochi and Mwandya, 2003).

**SIGANUS SUTOR**

Shoemaker spinefoot (*Siganus sutor,* Valenciennes 1835) is a fairly well researched species however there are no genetic studies prior to the two presented in this thesis. It is endemic to the Western Indian ocean. It’s resilient with a minimum population doubling time of less than 15 months (Froese and Pauly 2010). *S. sutor* spawn year round with two peaks occurring one to two months after each monsoon starts (Ntiba and Jaccarini 1990). The spawning aggregations are both temporally and spatially stable, in the Seychelles some spawning aggregations have been targeted for over 70 years (Robinson et al. 2004). Spawning occur in large aggregations and takes place in open water over hard substrate, the eggs are negatively buoyant and sink onto the substrate below the spawning aggregations (Woodland 1990). Acoustic tagging of *S. sutor* shows a high affinity for a single spawning aggregations (Bijoux et al. 2011). Little is known of what happens to the larvae, but juveniles thrive in seagrass beds and large adults prefer coral reefs (Kimerei et al. 2011). There is some debate over the phylogeny of Siganids Borsa et al. (2007) place *Siganus canaliculatus* and *S. fuscescens* as distinct species while Hsu et al. (2011) see the two species as synonyms. Nobody contends the status of *S. sutor* as a distinct species. Population genetic studies done on other Siganids confirm population structures on similar scales as in paper IV (Ravaga-
Gotanco 2010; Magsino et al. 2008). A limited number of studies on Siganid population genetic structure have been reported. Genetic structure of S. spinus and S. guttatuus in Japanese waters performed by Iwamoto et al. (2009) reports gene-flow to be higher in S. spinus than S. guttatuus they attribute this to the “non dispersal strategy” i.e. small sized juveniles and an inner bay habitat preference, employed by S. guttatuus as opposed to S. spinus. S. guttatuus in the northwest pacific show a high degree genetic population structure with almost all pair-wise comparisons between population pairs significant, also the demographic history of S. guttatuus reports stable populations for the last 55,000-188,000 years (Iwamoto et al. 2011). The gene-flow associated with the Lessepsian invasion was investigated with two siganids, S. rivilatus and S. luridus (Hassan et al. 2003) There was no genetic difference between the Mediterranean and the Red Sea populations illustrating that the invasion occurred in great numbers and not just by a few immigrants. The influence of egg type (demersal vs pelagic) explain contrasting population genetic structures for S. fuscescens and S. argentus along the eastern Philippine coast (Magisino and Juinio-Menez 2008).

**Methods Used in the Thesis**

Molecular markers have different resolution, think of slow evolving conserved markers such as mitochondrial DNA (mtDNA) as looking at a forest with binoculars while high resolution rapidly evolving markers such as amplified fragment length polymorphism (AFLP) and microsatellites as a viewing the same forest with a magnifying glass. Depending on whether one views the forest with binoculars or a magnifying glass different patterns will emerge, with the binoculars one will be able to see the difference between different tree species, while the magnifying glass shows differences between individuals within a single species of trees. It's possible to use the magnifying glass to see differences between trees as it is possible to look at within species differences using a binoculars, it is however not optimal. Choosing the proper molecular marker is essential. In this thesis I have deployed two different molecular markers AFLP and mtDNA. Each with its own strength and weakness as listed below.
AFLP

Amplified fragment length polymorphism, AFLP, randomly amplifies the genome through a three step progress 1. Cutting of the genomic DNA with restriction enzymes 2. amplifying the restricted DNA with labeled primers 3. Analysis of the amplified DNA with an automated sequencer, but see Bensch and Åkesson (2005) for a full review. AFLP is a binary marker, the data is represented by presence absence of a loci. This poses some problem, the presence of a loci can either be a heterozygote or a homozygote, an absence of a loci is however always homozygote for an absence. When using dominant markers such as AFLP roughly 4 to 10 times as many loci have to be used when compared to co-dominant markers (Mariette et al. 2002). According to Sonstebo et al. (2007), as a rule of thumb, ten AFLP loci translates to 1 variable microsatellite. Also by deploying a number of different statistical analysis which all are based on different metrics (F-statistics, dissimilarity matrix and Bayesian assignment test) one can further ensured that the patterns presented in these paper are reliable. AFLP has been used extensively in plant ecology and is growing in popularity for marine population genetics as well. AFLPs ability to scan the genome and generate hundreds of informative loci makes it suitable when dealing with a previously uninvestigated species.

Mitochondrial DNA (mtDNA)

The mtDNA is the DNA contained within the mitochondria and is, in most organisms, inherited maternally. The common way to utilize mtDNA is PCR based sequencing of specific regions of the mtDNA, were the control region also called the D-loop and Cytochrome oxidase flanking the D-loop (CO1) being two of the most common. The mutation rate, hence resolution, of the two regions differs. The control region has a higher mutation rate making it more suitable for within species phylogeography and evolutionary demography, while the coding genes, in particular CO1 is used for between species phylogenetics, it is also useful for species identification and is the primary marker of the barcode of life project (Hebert et al. 2003). When analyzing mtDNA sequence data, the raw data, is in the form of DNA nucleotides (ACTG) in a sequence of a couple of hundred nucleotides stringed together.
Differences between individuals sequences arise through mutations, these mutations are very uncommon, about 3.6% per million years is average for teleosts (Donaldson and Wilson 1999).

**STATISTICS**

When allele frequencies have been calculated it is possible to calculate a number of different measurements, such as Wright’s F-statistics, including Fst, Fis, Nei’s average genetic diversity per locus, etc. The F stands for fixation index; fixation is measured as the increase in homozygosity as a result of inbreeding. Fst is the proportion of differences between groups in the data set compared to the data set as a whole, i.e. how much of the differences in the data set is explained by our grouping. Wright developed three fixation indexes Fis (individual), Fst (subpopulation), Fit (total population). These measurements all provide estimates of population diversity and can be interpreted as proportion of difference between groups.

\[ F_{st} = 1 - \frac{H_s}{H_t} \]  

Where \( H_t \) is the average gene diversity calculated for the whole data set the same way as \( H_e \) (2) is calculated. \( H_e \) is the most commonly used measurement for population genetic diversity; it represents the probability of sampling two different alleles from a population (Holderegger et al. 2006). \( H_s \) is the mean of each population within the total dataset. A high Fst is indicative of strong population subdivision, i.e. that the alleles have a non-uniform distribution among the populations.

\[ H_e = 1 - \sum_{i=1}^{n} p_i^2 \]  

In summary Fst is a measurement of population differentiation, the value given can be interpreted as the percentage difference between groups.

Gene flow
If we deplete stock A will it be replenished by another stock? This question is of fundamental importance when deciding on management actions especially for marine species. Rate of effective migration is usually calculated by the formula

\[ F_{ST} = 1/\left(1 + 4mN_e\right) \] (3)

\( mN_e \) is the product of the migration rate and the effective population size, which is interpreted as the number of effective migrants per generation between each population (Waples 1998). Around 10 migrants per generation imply an Fst of around 0.02 according to formula 3. But are 10 migrants per generation enough to replenish a depleted stock? When calculating rate of migration a number of assumptions are made; (1) the number of subpopulations is infinite; (2) \( N_e \) is the same and constant over time in every subpopulations; (3) breeding is random within subpopulations; (4) generations are discrete; (5) migration is constant over time and the same for each of the subpopulations; (6) \( m \) is small; (7) alleles are selectively neutral; and (8) there is no mutation (Waples 1998). The important thing to remember is that \( mN_e \) illustrates gene flow on an evolutionary and not ecological timescale making it hard to apply it to management strategies. Therefore assignment of individuals to genetic clusters (or stocks) as described below is of more practical use.

**AFLP STATISTICS**

There are three fundamentally different approaches to working with AFLP statistics, frequency based statistics, band-based statistics and Bayesian statistics. Band based and frequency based statistics are the traditional choices while Bayesian statistics has only recently been widely used (Bonin et al. 2007). Band and frequency based methods both predefine what a population is and then compare genetic differences between these. Bayesian clustering, most commonly used as assignment tests, does not use a priori defined populations but rather groups the dataset into genetic clusters that are in Hardy-Weinberg equilibrium. Band based metrics are built on dissimilarities between individuals, that is presence or absence of loci. The dissimilarities can be measured by a number of different indices, see Bonin et al. (2007 for a full review). In paper II and IV we used the Jaccard coefficient (Zhivotovsky 1999) which only takes into account bands present in at least one individual, and is therefore unaffected by homoplastic absent bands. The band based
approach is used when visualizing the data on a principal coordinate analysis (PCO) as described in paper II and IV. Frequency based statistics calculate allele frequencies, the dominance of AFLP poses a problem here since there is no way of knowing if a presence band is a homozygote or heterozygote, an absence of a band is always a homozygote. Dominance implies that it is not possible to determine if a presence band is present in one (homozygote) or two copies (heterozygote). There are a number of ways to estimate the allele frequencies, the Bayesian approach or the inbreeding coefficient and the square root of the null allele frequency. The dominance of AFLP creates problems when calculating hetrozygosity, there are however a number of different ways to calculate dominance by assuming Hardy-Weinberg equilibrium and calculating hetrozygosity from Fst. The Bayesian approach implemented in AFLPsurv (Zhivotovsky 1999) seems to be the most robust (Bensch and Åkesson 2005).

An insertion/deletion mutation in the amplified sequence would alter the length of the fragment and cause it to be scored as another presence band, a mutation in the primer site which inhibits the amplification of that site represents is the most common type of mutation (Caballero et al. 2008). Homoplasic absent bands describe the fact that there are a number of mutations that can alter the primer site resulting in an absence of a band.

Homoplasy is the appearance of a band of a certain length in different individuals that are different in sequence but have the same length. This can lead to an overestimation of the allele frequency for the present band, an underestimation of degree of differentiation between subpopulations and an overestimation/underestimation of the heterozygosity, depending on the frequency of the marker (Caballero et al. 2008). The impact of homoplasy becomes more prominent when many loci are scored from a single primer combination, hence no more than 100 fragments per primer combination are recommended (Caballero et al. 2008). Homoplasy leads to underestimation of genetic differences between populations i.e. it increases the chance of doing a type II statistical error.

Bayesian assignment tests are growing in popularity and have now been adopted to work with dominant data, the software STRUCTURE being the most commonly used. The benefit of Bayesian assignment tests are that no prior definition of populations is needed. The previous methods described can only be used when a grouping of individuals have been
made prior to the test, which is fine if there is a correlation between genetic and geographic groups (Pritchard et al. 2000, Falush et al. 2003, 2007). The fact that no a priori population grouping is necessary is particularly useful in highly mobile organism such as marine fish species where individuals of different populations may mix during non-reproductive seasons. A good example of when genetics and geography coincide is Sea turtles sampled while laying eggs (Dethmers et al. 2006). By grouping turtles sampled on the same nesting beach there will be a true reflection of population structure when using traditional statistics, band or frequency based. However if the turtles are sampled randomly at sea during non spawning, traditional statistics may fail to detect a genetic structure due to the fact that the turtles have to be grouped prior to the statistical testing. Since no such a priori grouping is needed with Bayesian statistics it is possible to detect the genetic structure independent of geography. It’s worth noting that these results can be difficult to interpret without sound ecological knowledge of the studied species. Bayesian statistics is not hypothesis testing and therefore doesn’t provide any p values, instead probabilities are produced. When STRUCTURE is used you set the program to test a range of different groups (K), for example if there are 5 sampling sites you set the program to explore K 1-8, each K is then iterated 3-5 times depending on the size of the data-set. The absolute value P(X| K) for each K is given and the K with the highest value is the most probable K (Pritchard et al. 2000). An alternative method to choosing the optimal number of K is proposed by Evanno et al. (2005), instead of plotting P(X| K) you plot the parameter ΔK. ΔK is based on the rate of change in the log probability of data between successive K values and better reflects the “true” number of clusters.

Constrained Ordinations

A Constrained Analysis of Principal Coordinates (PCoA/CAPscale) can be used to visualize AFLP data. A CAPscale is a multidimensional scaling that allows for non-Euclidian dissimilarity indices and is a very visual way to illustrate the data since the data is plotted based on dissimilarities. A priori definition of groups such as populations or genetic clusters are necessary, it is also possible to test if the centrioles (genetic averages of predefined groups) are significantly separated from each other through an anova (Anderson and Willis 2003).
Isolation by distance

Isolation by distance (IBD) is the inverted relationship of genetic diversity and geographic distance (Wright 1943). The further apart two sites are from each other the more genetically different they should be. IBD is based on the assumption of Wright's island model of genetic migration which simply is the assumption that migration occurs in a stepping stone fashion from one population to the next. Populations close to each other are more similar than populations further apart. A mantel test of correlation between Fst and geographic distance is the simplest way to calculate IBD. IBD is best used when comparing sites along a straight line such as a coastline or island separated along a geographic gradient.

DNA SEQUENCE STATISTICS

When modeling the evolution of DNA there are three main parametric approaches, (1) base frequency parameters, (2) base exchangeability parameters and (3) rate heterogeneity parameters (Whelan et al. 2001). The base frequency parameters average the frequencies of all the bases at all the sites. Base exchangeability parameters describe the substitution tendencies of the bases. The most widespread approach to modeling rate heterogeneity amongst sequence sites is to describe each site’s rate as a random draw from a gamma distribution. A gamma distribution can be used in combination with base frequency and exchangeability parameters (Whelan et al. 2001). Pair-wise mismatch distribution plots the probability that two randomly sampled sequences from the same neutral loci will differ as the total number of differences increases. If a population has remained constant over many generations (100-1000) the probability that two random neutral gene sequences will differ at exactly the same nucleotide will decrease as the number of pair-wise differences increase, if however the population has increased suddenly in size, the shape of the curve will resemble a wave (Rodger and Harpending 1992).

Bayesian skyline plot – illustrates past female effective populations size through time, using a standard MCMC sampling procedure directly from sequence data (Drummond, Rambaut et al. 2005). The method takes into account both phylogenetic reconstruction errors as well as
the stochastic error intrinsic to the coalescent process. The method allows for discovery of novel demographic signatures undetectable to previous skyline plots (Drummond et al. 2005).

**COMBINING mtDNA-SEQUENCING AND AFLP TO SPEED UP AND MINIMIZE COSTS IN GENETIC SPECIES IDENTIFICATION**

When using non species-specific genetic markers such as AFLP, correct species identification is important. However some marine species are difficult to properly identify in the field using morphological characteristics alone. An alternative is genetic identification using DNA sequencing, mainly sequences of mitochondrial DNA (mtDNA). DNA sequencing of specific slow evolving marker genes such as D-loop or CO1 has proven to be a robust technique for genetic species identification that works well with marine fish (Herbert et al. 2003). However mtDNA data analysis seldom require as large sample sizes as AFLP. Five to ten samples per site are usually sufficient for mtDNA analysis; as opposed to AFLP that require sample sizes of about 30 samples per site, thus sequencing hundreds of fish can be time consuming and costly but by combining a large AFLP dataset with mtDNA sequencing of a few individuals we were able to genetically identify individuals that have only been identified using morphological characteristics, thus we were able to discard misidentified samples.

The analyses by CAPscale and STRUCTURE group individuals based on genetic similarities (Anderson and Willis 2003; Pritchard et al. 2000). Individuals that are genetically similar get grouped together, by cross referencing individuals that were analyzed both with AFLP and mtDNA it was possible to cross reference the clades identified with the mtDNA phylogenetic tree with the groupings identified by CAPscale and STRUCTURE on the AFLP data.

Initially a subset of individuals were sequenced using mtDNA, and next a phylogenetic tree was constructed showing the relationship between the sequenced samples. The entire dataset was then analyzed based on the AFLP data using different assignment tests that group the data by genetic similarities and then cross-referenced with the mtDNA phylogenetic tree to see if other non identified samples grouped with mtDNA clades Dodgy1 and Dodgy2 (fig 1a-b). All individuals grouping with highly divergent mtDNA samples can be
deleted based on the assumption that genetically similar AFLP samples group together and thus should belong to the same mtDNA clade.

To ensure that only *Siganus sutor* was sampled for study III and IV, mtDNA genotyping of the D-loop was carried out in 120 individuals. mtDNA samples were primarily collected and analyzed by a coauthor. Because of this the samples do not cover the entire geographic range of papers III and IV, we do however have expertly identified *Siganus sutor* from Kenya which represents the other extreme of *S. sutors* habitat. 120 individuals of 656 (roughly 20%) were genotyped for the D-loop region. Also ten samples of *S. sutor* and ten samples of the closely related *Siganus canaliculatus* were morphologically identified by a *S. sutor* specialist Dr. Agembe and later analyzed only using AFLP. When analyzing the 120 individuals using D-loop sequences, three clades were discovered (fig 1a), the majority of samples (107) grouped into one clade, the other two clades were composed of five and eight individuals each. *Siganus fuscescens* is the only *Siganus* species with an available mtDNA sequence in GenBank and was used as an out group. *S. fuscence* doesn’t inhabit the Western Indian Ocean. There was a huge spread of AFLP data on the CAP scale (fig 1b) and the mtDNA-identified samples were concentrated on the bottom half of the graph with the exception of the samples identified as "dodgy", which were located high up on the y-axis. "Dodgy2" is quite distinct as it occupies the upper left quadrant of the plot with a distinct “swarm” of AFLP samples around it, "dodgy1" was not as clear cut but had its centriole (genetic average) equally high up on the y – axis. In STRUCTURE we predefined five groups using the popflag option, the five groups were; morphological *S. sutor* and *S. canaliculatus*, mtDNA sutor, "dodgy1" and "dodgy2". In essence we told the software that these five groups exist and sought to place all the other samples into one of these groups. As seen in fig 1b-c there is no genetic difference between *Siganus sutor* and *Siganus canaliculatus* regarding AFLP data, although no conclusive results can be drawn from our small sample sizes so it is worth investigating further. Borsa et al. (2007) did however find genetic differences between *S. sutor* and *S. canaliculatus* when investigating the lineage diversification in Rabbitfish using the mtDNA markers CytB and 16S. In the STRUCTURE analysis there is a clear distinction between the mtDNA samples and the samples identified morphologically by a specialist (fig 1c); the samples are however collected in different years and at different sampling localities (morphologically identified sampled in Kenya and mtDNA sampled from the Islands of the
Western Indian Ocean). Structure assigned the unknown samples to a mixed lineage rather than distinct populations. Which is to be expected since the reference samples are from the extremes of the distribution continuum. Dodgy2 was entirely composed of a single genetic cluster (cluster 3) and all unknown individuals that assigned to that cluster were deleted, in total ten individuals. Dodgy1 was only partially defined by structure, most samples were assigned to cluster 4 in some extent and thus we used a cautionary approach and sorted out individuals that had more than 50% of their genotype from cluster 4 a total of 46 individuals. This is probably an over cautionary approach since a weak correspondence between nuclear and mtDNA markers in phylogenetic analyses is not uncommon (Toews and Brelsford, 2012). In total we deleted 56 individuals that probably are of the wrong species, a feat we would not have been able to do without the aid of both mtDNA and AFLP.

Fig 1 a. Neighbor-joining tree of 120 samples based on D-loop sequences of morphologically identified S. sutor. S. fuscescens was used as an out-group. b. CAPscale analysis of 656 samples of morphologically identified S. sutor based on AFLP data, individuals with mtDNA data available are identified as well as the AFLP samples identified by Dr Agembe. c. STRUCTURE analysis using popflag option to assign unknown samples to known populations. The predefined populations were; morphologically identified S. sutor and S. canalicularus, mtDNA clade S. sutor, mtDNA clade dodgy1 and dodgy2.
SUMMARY OF THE THESIS RESEARCH PAPERS AND MAJOR FINDINGS


In this study we aimed to investigate the historical demography and genetic connectivity of Scarus ghobban in the Western Indian Ocean. A total of 83 samples were collected throughout the region, DNA was extracted and the D-loop was sequenced. The phylogeny of S. ghobban D-loop exhibits three clades without geographic relationship. “Clade 3” formed a polytomy cluster at the base of the tree, which were all sister to clades 1 and 2. This suggests that clades 1 and 2 are more recent than the other clades, which presumably acted as the source populations giving rise to the clade 1 and 2 populations, all of which have since dispersed extensively in the region leading to the observed lack of geographic structure. It is likely that all Scarus ghobban in the WIO shares a common demographic history due to their high capacity of dispersal. Analysis of molecular variance (AMOVA) showed genetic differentiation between two groups consisting of samples from Mauritius and Tanzania in one group, and samples from Kenya and Seychelles in another group.

Asymmetric or directional gene flow, favoring migrants from Mauritius to Tanzania, and Tanzania to Kenya, migrants proceed downstream with the South equatorial current and the East African counter current, respectively. Population size fluctuations match sea surface level variations, historical decline in population size is due to reduction in suitable habitat rather than elevated sea surface temperatures. The distinction between historical and contemporary gene flow is relevant to the issue of management of fish stocks as when interpreting patterns of gene flow, it is important to consider the geological and oceanographic history of the region (Barber et al. 2000; Fauvelot et al. 2003). If there has been a relatively recent divergence between populations of S. ghobban in the region then the rate of evolution of the mitochondrial marker used for this study may not be sufficiently high to detect that divergence.
Field identification of juvenile fish using morphological characteristics can be difficult. In this study we sampled individuals what was initially morphologically identified to be *Mugil cephalus*. However, two distinct genetic clusters emerged when analyzing the AFLP data using the Bayesian assignment test in STRUCTURE 2.2, thus we employed genetic barcoding of the CO1 gene to ensure that we had sampled a single species and not two species as indicated by the Bayesian assignment test. The samples were identified as *Valamugil buchanani*, *Moolgara seheli* and *Moolgarda cunnesius*. Individuals identified by DNA-barcoding as *V. buchanani* were found in both the genetic clusters identified by the Bayesian analysis of the AFLP markers. Our two markers, the CO1 gene (mtDNA) and AFLP (nuclear DNA), revealed two different scenarios with no correspondence between them. A weak correspondence between nuclear and mtDNA markers in phylogeographic analyses is not uncommon (Toews and Brelsford, 2012) and our combined information on AFLP markers and CO1 sequences did not clearly separate out different species. The pairwise comparisons between sampling sites revealed that there were some fine-scale genetic differences. This study has further shown that the species currently regarded as *Mugil cephalus*, in fact, is *Valamugil buchanani* and, as there are differences in growth rate and maximum size between the two species (Froese and Pauly, 2010), it is possible that aquaculture systems may be deemed a failure due to the fact that the fish under culture are slower-growing than *M. cephalus*.

The aim of this study was to investigate both the spatial and temporal genetic variation in *S. sutor* along the coast of Kenya. A total of 322 samples were collected from six sites along...
the Kenyan coast, one site was continuously sampled for five weeks. We then compared the temporal genetic variation with the spatial variation. This is the first study to reveal both spatial and temporal genetic variation on this scale within this species and we found the temporal genetic variation to be about one fifth of the spatial genetic variation. The temporal genetic variation indicates that the genetic stocks differ in their distribution over time, i.e. a change in the proportions of the major genetic clusters present. It seems that *S. sutor* is not stationary and that the different populations have overlapping home ranges and several populations can be found at a single site depending on sampling time.

**PAPER IV CONTRASTING POPULATION GENETIC STRUCTURE OF Siganus sutor BETWEEN MAINLAND COASTAL AND OCEANIC ISLAND POPULATIONS (MANUSCRIPT)**

In this study we aimed to investigate the genetic population structure of *Siganus sutor* in the Western Indian Ocean. We did this by sampling 506 samples from 20 different sites in 6 different countries. The genetic marker used for this study was AFLP. We found a clear genetic population structure for *S. sutor* in the Western Indian Ocean. The genetic variation follows an isolation by distance models when testing for the entire region, along the mainland there is no isolation by distance. The genetic data is partitioned into three genetic clusters that are further sub dived into to clusters each. The sub clusters are defined by geography but there seems to be little genetic structuring within island sites. Our main result is that there were extensive mixing of genetic clusters along the African coastline, including Zanzibar Island, and that the oceanic island sites were dominated by single clusters with very little evidence of other genetic clusters at each site. This contrasting pattern of genetic structure is probably explained mainly by the difference in habitat continuity where sites on oceanic islands, showing lower levels of gene-flow than the sites along the coastal mainland. The effect of habitat discontinuity on genetic structure has previously been described on smaller geographic scales for other marine organisms (Riginos and Nachman 2001, Alberto et al. 2010). Our main conclusion with regards to management of the *S. sutor* fisheries is that different strategies have to be applied for oceanic islands and the main coast of East Africa. The East African coast line should be managed as a mixed stock fishery according to the
portfolio model. By aiming at extracting fish from as many genetically distinct units and by ensuring that as many genetic units are maintained the overall resilience of the fisheries is maintained (Schindler et al. 2010). The oceanic islands should be managed with more caution since they are to a greater extent composed of only one genetic cluster. This makes the fishery more susceptible to collapse caused by genetic depletion (Biro and Post 2008) and their clear genetic separation indicates that rebuilding after a collapse would take very long time (Hutchings and Reynold 2004).

**DISCUSSION**

This thesis aims to investigate the presence of genetic structures in three species of fish in the Western Indian Ocean and I have also tried to illustrate how these findings can be useful in the local management of the artisanal fisheries. In order to do that we first investigated the historical patterns of connectivity throughout the region (paper I). In papers II and III we focused on local scale connectivity in Kenya and Tanzania and finally in paper IV we investigate the large-scale contemporary gene flow throughout the Western Indian Ocean. In paper III we also investigate the temporal genetic variation at one site and compare it to the small-scale genetic variation along a stretch of the Kenyan coastline. Some overall conclusions that can be drawn from my body of work is: there are genetic structures present in the Western Indian Ocean even though the apparent lack of physical barriers. Major oceanic currents aid evolutionary dispersal patterns (paper I). A single geographic site need not be genetically homogenous or temporally stable (paper III). Island sites are genetically more homogenous than mainland sites (paper IV). Recently, climate change has become a serious cause of concern for coral reef organisms, partly due to rising sea level and elevated temperatures. With respect to this, our demographic studies (paper I) serve to illustrate that reef fish with a greater capacity for dispersal could have higher resilience. The relatively high level of gene flow that we report for *S. ghobban* in the WIO is consistent with the general pattern of high mtDNA gene flow for reef fish with a pelagic larval phase (Chen et al. 2004; Craig et al. 2007). This high capacity for dispersal is principally brought about by ocean currents (Roberts 1997). Convincing support for the role of regional currents in the dispersal of *S. ghobban* also comes from our estimates on the direction and number of migrants per generation between locations. As illustrated by Mora et al. (2011) even the most geographically distant reefs of the Western Indian Ocean are not further apart than 30 days.
of passive transportation of oceanic currents. This suggests that the limited connectivity is a behavioral adaptation rather than an accessibility hindering (Iwamoto et al. 2008, Magisino and Junio-Menez 2008) where the fish can disperse throughout the region, but choose not to.

A species need not always be a species is a lesson learned from paper II, we morphologically identified our samples in the field using relevant literature, worth noting is that we sampled the juvenile fish more or less in the same way as it is sampled fish for aquacultures. Two distinct genetic clusters were found in the AFLP data, and an equal amount of samples from both clusters were sequenced for the CO1 gene. A DNA blast search (GenBank, BLAST) revealed potentially five different species with no correlation to the AFLP clusters (see fig 2 paper II). Our two markers, the CO1 gene (mtDNA) and AFLP (nuclear DNA), revealed two different scenarios with little correspondence between them. The two AFLP clusters each contained several species identified by the CO1 neighbour-joining tree. The AFLP variation in cluster 1 was greater within *V. buchanani* than between *V. buchanani* and *M. seheli*. In addition, *V. buchanani* mtDNA haplotypes were found in both the AFLP-based genetic clusters, with no clear delineation in the AFLP phenotypes in the continuum between *V. buchanani* and *M. seheli*. It is also worth noting that Durand et al. (2012) questioned the validity of the two genera, *Valamugil* and *Moolgarda*.

One of the key findings of paper III was that populations of *S. sutor* were not stable in time, a well documented phenomena for cod (Lindegren et al. 2013), salmon (Beacham et al. 2004), European eel (Dannewitz et al. 2005) and other temperate and arctic species. From a management perspective this can be an issue since monitoring can be difficult, near impossible without genetic tools. The spatial genetic structure found in this study, over this geographic scale, could be influenced by difference in reproductive output and larval settlement as previously described for damselfish (Hepburn et al. 2009; Hogan et al. 2010). The pattern found is in line with previous studies on Siganides using mtDNA markers (control region) reporting significant genetic structure, however on a much larger spatial scale (Ravago-Gotanco et al. 2010). The artisanal fishery along the coast of Kenya is managed by the semi-autonomous beach management units (BMUs) that operate on a small scale i.e. 2 km, the currently used management strategy is temporary closure of fishing grounds (Ocheiwo 2004). Our results imply that this strategy may not be successful due to the
temporal fluctuations. As the BMUs operate on such a small scale and the temporary closure might protect a number of stocks in various proportions but when the closure opens those stocks might not be present, reducing the efficiency of the closure. Furthermore, our results implies that the fishery is a mixed stock fishery and when fishing in the vicinity of a spawning aggregation the fishermen are extracting fish from several genetic stocks, which is a better alternative than extracting fish from the spawning aggregation itself as it probably is composed of a single genetic stock. Using this strategy the fishery can and should be managed as a mixed stock fishery, similar to the North American salmon (e.g. Beacham et al. 2011) and North Atlantic cod (e.g. Pampoulie et al. 2011). We believe that the economical importance of S. sutor warrants a species-specific management plan. However, there is a need to determine the number and extent of stocks present throughout the WIO.

Contemporary gene-flow in the WIO is illustrated in paper IV. Three genetic clusters that correlates with geography dominate the population genetic structure of Siganus sutor. Our main result is that we found extensive mixing of genetic clusters along the African coastline, including Zanzibar Island, and that the oceanic island sites were dominated by single clusters with very little evidence of other genetic clusters at each site. This contrasting pattern of genetic structure is probably explained mainly by the difference in habitat continuity where sites on oceanic islands, showing lower levels of gene-flow than the sites along the coastal mainland. The effect of habitat discontinuity on genetic structure has previously been described on smaller geographic scales for other marine organisms (Riginos and Nachman 2001, Alberto et al. 2010). Our results are in accordance with previous studies on Siganids. Iwamoto et al. (2008) investigated genetic structure of S. spinus and S. guttatus in Japanese waters and found that gene-flow is higher in S. spinus than S. guttatus. They attribute these differences to the “non dispersal strategy” of S. guttatus, were small sized juveniles of have an inner bay habitat preference. S. guttatus in the northwest pacific had a high degree of genetic population structure, with populations significantly separated on a geographic scale of about 100 - 2,000 km. The demographic history of S. guttatus also indicates stable populations for the last 55,000-188,000 years (Iwamoto et al. 2012). The contrasting population genetic structures for S. fuscescens and S. argentus along the eastern Philippine coast (Magisino and Juinio-Menez 2008) can be explained by egg type (demersal vs. pelagic) were species with demersal eggs have a more pronounced genetic structure. The demersal
nature of *S. sutors* eggs and the clear populations structure found in our study support the notion of egg type being important for the genetic population structure of Siganids.

To put my research into a larger perspective I have created a hypothetical seascape system that is highly probable for *S. ghobban*, *V. buchanani* and *S. sutor* (Fig. 1). Currently managers have no genetic information available, and thus manage the fisheries by gear restriction and no take zones as previously described. I have so far in this thesis illustrated the use of two molecular markers that yield complementary results. mtDNA provides information about the genetic evolutionary demography and history and AFLP describes the contemporary genetic structure. By combining the information generated by these two markers to the hypothetical seascape in fig. 2 the evolutionary history, evolutionary demography (paper II), the contemporary genetic stock structure (papers II, III and IV) and the extent of the stocks can be added to the manager’s toolkit. The hypothetical seascape also illustrates some important issues that need to be addressed in fisheries management, issues that would not be raised without the additional knowledge generated by the aid of molecular techniques. These are among other, that the spatial distribution of stocks is not temporally stable (paper III and IV), the effectiveness of the MPA might be less than optimal, and each village is extracting resources from the same genetic stocks. It’s important to realize that the spatial distribution of fish stock along a continuous coastline is not temporally stable (paper III), what is stable is the occurrence of spawning aggregations (Robins, Marguerite el al. 2007). After spawning the larvae passively or actively drift with ocean currents and settle at a suitable habitat, the adults disperse and, at the next spawning event the fish would then return to their natal spawning aggregation using cues previously described.

The genetic composition of spawning aggregations has to my knowledge not yet been investigated for tropical coral reef associated fish. Regarding the effectiveness of the no take zone one can argue that since no spawning aggregation is situated within the no take zone, the area is not fully protecting the species. Still, protecting an area ensures that at least some of the fish from all genetic clusters have the ability to reach maturity under the condition that the species investigated is relative stationary as an adult.
Fig 2. A hypothetical east African coast line, an MPA is set up to protect a highly diverse coral reef. Along the coast are three fishing villages with adjacent fishing grounds, the fishing grounds are mostly located with the continuous lagoon. The genetic composition of a theoretical fish species is marked by different colored unique genetic clusters. The fish is an annual mass-spawner, and the spawning takes place outside the lagoon in deeper water one of the spawning aggregations is targeted by a fisheries. The three different times represent a continuous year.

In this scenario all fishing villages are fishing from all genetic stocks which reduce the fishing pressure on individual stocks, however management decisions made on a very local scale (single fishing ground) has less impact than decisions made on a regional scale. As can be seen by this hypothetical seascape, population genetics is a valuable tool in artisanal marine fisheries management.

Our main conclusion with regards to management of the S. sutor fisheries is that different strategies have to be utilized for oceanic islands and the main coast of East Africa. The East African coastline should be managed as a mixed stock fishery according to the portfolio model (Schindler et al. 2010). The oceanic islands should be managed with more caution since they are to a greater extent composed of only one genetic cluster.
REFERENCES


Ross ST (2001) *The inland fishes of Mississippi*. University Press of Mississippi, Jackson, 624 pp


Sousa MI, Dias M (1981) *Catalogo de peixes de Mocambique—Zona Sul*. Instituto de Desenvolvimento Pesqueiro Maputo 121 pp


Woodland DJ, Ichthyology BPBMD of (1990) Revision of the fish family Siganidae with descriptions of two new species and comments on distribution and biology. *Bernice Pauahi Bishop Museum*


Längs med Östafrikas kust samt runt många öar i Indiska oceanen löper ett nästan oavbrutet korallrev, lagunen innanför korallreven utgör basen för nästan allt småskalig fiske längs med kusten. Detta ska inte förväxlas med det kommersiella fisket med avancerade redskap och stora båtar, detta fiske bedrivs med de enklaste redskapen såsom mjärde, krok, nät, och lina. Det förekommer även harpunfiske, dynamitfiske och notvarp trots att de är olagliga. Vad fiskare saknar i teknologi tar de ikapp i antal, vissa fiskeområden kan ha upp till 19 fiskare per kvadratkilometer. Sedan 1980 talet har mängden fisk i lagunerna minskat och det finns tydliga tecken på överfiske. Förvaltningen av de marina områdena har hittills fokuserat på att främst skydda korallreven, detta genom att ett antal marina naturreservat har etablerats. Fisket har reglerats genom vilka redskap som får användas och var fisket får ske. För att kunna förbättra förvaltningen av fisket behövs mer information om bland annat fiskbeståndens storlek och spridning. En viktig aspekt som saknas är populationsgenetiken, dvs. hur många populationer finns det i det förvaltade området och hur är de släkt med varandra. Ett av målen med denna avhandling har varit att bidra till förvaltningen av det småskalig kustnära fisket i Östafrika, genom att studera hur fiskbestånd är besläktade och hur fiskar sprids i västra indiska oceanen.

förklaringar till IBD, begränsad spridningsförmåga på äggen, aktivt val av yngel etc. det finns dock populationer som inte uppvisar IBD. Dolda genetiska strukturer är ett problem, ta lax som ett exempel, om man samlar in lax när de leker i sina älvor kommer en tydlig genetisk struktur synas, laxar från samma älv är mer släkt med varandra än laxar från en annan älv. Hur ser populationsstrukturen ut om laxen samlas in ute till havs? Beroende på vilken metod man använder för att analysera data kommer svaret bli olika, F-statistik analyserar den genetiska variationen baserat på förbestämda grupper, i detta fall skulle det vara fångstlokal. I älv exemplet skulle F-statistiken visa på genetisk struktur medan i havsfallet skulle ingen struktur finnas. Alternativet är att använda ett Bayesianskt assignment test, detta test skapar grupper som befinner sig i genetisk jämvikt oberoende av fångstlokal. I teorin kommer det Bayesianska testet visa på samma populationsstruktur i både älv- och havsexempen. I denna avhandling har jag använt mig av både F-statistik och Bayesianska test för att undersöka populationsstrukturerna. Om en lax simmar fel och leker i en annan älv kallas detta för genflöde.

Strömförhållanden är till viss del viktiga för spridningen av yngel och ägg, västra indiska oceanen domineras av sydekvatoriella strömmen som flyter mellan Seychellerna och Mauritius mot Östafrika, när strömmen når Afrika precis norr om Madagaskar delar den på sig, den nordliga strömmen blir den Östafrikanska kustströmmen. Denna ström stannar av under den nordliga monsunen och då domineras kusten av den Somaliska motströmmen.

I artikel I undersöker vi det historiska genetiska släktskapet hos papegojfisken *Scarus ghobban*. Hur har fisken spritt sig och hur har populationsstorleken varierat över de senaste 300 000 åren? Havsströmmar har troligen haft en betydande roll för spridningen av *Scarus ghobban*. Främst har spridning skett från de östra öarna Mauritius och Seychellerna till Östafrikas kust. Genetisk kunde vi identifiera tre olika klader (grupper) som inte hade någon koppling med geografi. Vi kan också se att det har skett en populations expansion för ca 160 000 år sedan. Att förstå migrationsmönster är viktigt när det till exempel gäller nyetableringen av marina reservat, speciellt om en av reservatets funktioner är att förse andra områden med föryngring genom nya individer.

I artikel II använder vi mtDNA för att bekräfta identiteten på en multefisk. Vi gjorde detta för att AFLP data visar två tydliga genetiska grupper, dessa grupper är så distinkt separerade att vi misstänker att det kan röra sig om två olika arter och därför sekvenserade vi deras mtDNA också. Sekvensering av dessa individer visar på att det inte ens är den art vi har identifierat i fält utan en helt annan art, vi misstänker också det kan förekomma flera arter. Trots problematiken med olika arter så finns det en tydlig populationsstruktur då de genetiska grupper vi såg på AFLP inte har något att göra med de olika arter vi uppräckte via sekvenseringen. Den multeart vi trodde vi samplade har framhävts som en lovande art för fiskodlingar, den arten vi identifierade blir betydligt mindre och växer långsammare. Detta kan ha lett till att försök med fiskodlingar har dömts som misslyckade eftersom de odlat en mycket mindre art. I denna studie samlade vi in prover på samma sätt som lokalbefolkningen samlar in yngel till just fiskodlingarna, vi var oroliga för att den genetiska mångfalden hos de vida populationerna hotades att urholkas men det visade sig att så var inte fallet.

fisketrycket, detta kan ha begränsad effekt om olika bestånd hela tiden rör sig igenom det skyddade området.

I del IV tittar vi på hur den vitfläckiga kaninfisken *Siganus sutors* populationsstruktur ser ut i en stor del av Västra indiska oceanen. 506 fiskar har samlats in från 20 olika lokaler i 6 olika länder. Vi använder oss av AFLP för att titta på den genetiska variationen. Tre genetiska grupper med tydlig geografisk koppling. Önationerna (Comorerna, Seychellerna, Mauritius och Rodriguez), samt Zanzibar domineras av en genetisk grupp, de andra två grupperna dominerar på fastlandet, dock finns det en viss geografisk uppdelning mellan dem. Samtliga av de tre genetiska grupperna har tre undergrupper som också har kopplingar till geografi. I och med att vi vet att den temporala variationen för kaninfisk är ca 1/5 av den totala kan vi jämföra olika lokaler och se om dess genetiska skillnader beror på tidsberoende variation eller faktiska skillnader mellan populationer. Insikten att önationerna genetiskt mer enhetliga är viktig för förvaltningen av fisket. Fisket runt öarna är mer känsligt för överfiske eftersom väldigt lite av föryngringen kommer utifrån.
ACKNOWLEDGEMENTS

There are so many people to thank, doing a Phd is a journey you start alone but quite quickly you find yourself surrounded by all these wonderful people. Words cannot express the gratitude I feel towards all of you who have honored me with your presence. Petra thank you for guiding me down this path, Emma and Jossan thank you for all the laughs, the hard work, and especially thanks for all the help in completing this thesis. To the proprietors of hotel “Ängel och Korpen” Asim and Linn, how can I thank you enough for all your support. A big thanks to all the wonderful co-workers I’ve had the pleasure to laugh with over the years, Lars, Mange, Anna-Karin, Stefan, Petter, Natasha, Paulina, Andrea. My x-jobbare, Jossan (again), Marika, Erik you made my work so much better. Dr. Dark and Dr. Dawa may the saga continue.

My supervisor Mats has given all the freedom and independence, but when I needed support all I had to do was ask. For this I thank you. A big thanks to all the institutions that have supported me, SH, SU, SIDA, CORDIO, IMS, MOI, WIOMSA.

I couldn’t have done this without the support of my family. If for no other reason I undertook this journey to get some leverage for our dinner discussions about the Baltic cod. Dad you have taught me to treat everybody as an equal and to be the hardest worker of them all. Mom from you I learned that home is where your loved ones are. Mom and dad thank you for instilling in me that I could be anything when I grow up, even a biologist. My brother has taught me that to succeed all it takes is a lot of hard work, my sister has taught me that the sky is the limit. You are both my inspiration. My mother and father in-law thank you for all the help with taking care of Hampus so I could finish this thesis.

I used my research as a pickup line when I met my wife, thank you Sara for putting up with me and supporting me through all this. Little did you know that you didn’t just marry me, but also my thesis. Well that’s over now, it’s just me from now on.

To the fisher folks who so graciously let me sample their livelihood I say

“So long, and thanks for all the fish.”