ROYAL HOLLOWAY UNIVERSITY OF LONDON

School of Biological Sciences

POPULATION STRUCTURE AND PHYLOGEOGRAPHY OF *OCTOPUS CYANEA* AND *LETHRINUS* SPECIES IN THE SOUTHWESTERN INDIAN OCEAN

By

Amy Louise Taylor

Thesis submitted as part of the requirements for the awarding of the degree Doctor of Philosophy

September 2014

Declaration of Work

I, Amy Louise Taylor, hereby declare that all the work presented in this thesis is from my authorship, with the exception of the laboratory techniques used to isolate microsatellite markers for *Octopus cyanea* that were performed by my colleague Dr Niall McKeown and the control region sequences of *Lethrinus nebulosus* from the South African Institute of Aquatic Biodiversity (Chapter 6) which were sequenced by my colleague Dr Gavin Gouws.

Signed:

Name:

Date:

Acknowledgements

I would like to thank everyone who has helped me to complete this thesis. In particular I would like to thank Niall for spending many hours discussing the various principles and analyses used in population genetics with me when I was confused (frequently!) or needed to clarify a particular point. I would also like to thank Martin Baxter for proofreading my thesis, which was invaluable. Thanks also to my supervisor Prof. Paul Shaw.

In Madagascar I would like to thank Sophie Benbow and the staff at Blue Ventures for collecting octopus and fish samples.

In South Africa I would like to thank Dr Gavin Gouws, for collecting and making available to me a number of *Lethrinus* samples and *L. nebulosus* CR sequences.

In the Seychelles I would like to thank; The Seychelles Fisheries Authority staff, Daryl and Stella Green on Praslin who invited us into their home for dinner and drinks and to collect octopus samples, Dan the octopus fisherman, and Andre and Marcelle the fishermen from Mahe without whom I would not have collected nearly as many samples of *Lethrinus* sp.

In Mozambique I would like to thank Dr Adriano Junior and the staff at CEPAM – the Centre for Marine Research in Pemba, who helped us patrol the reef flat and collect samples of octopus from the various fishers on the beach.

I cannot write the acknowledgements without mentioning my boyfriend Tom, who very patiently travelled round helping me to collect samples from the Seychelles and Mozambique and didn't complain about chasing fishermen on a reef flat in the baking sun for hours at a time, even when he stood on a sea urchin!

I am also extremely grateful to my family and friends for their understanding and support throughout the 4 years that this thesis took to complete.

Abstract

Tropical reefs such as those in the Southwestern Indian Ocean (SWIO) are the most biologically diverse of shallow water marine ecosystems. Despite this, biodiversity of the SWIO is understudied and with fishing pressure on reefs increasing there is a need to document genetic diversity of species and communities. This thesis set out to investigate intra-specific genetic diversity in four species important to subsistence and commercial fisheries: *Octopus cyanea*, and the emperor fish *Lethrinus mahsena*, *L. nebulosus* and *L. harak*. The aim was to examine factors underlying the partitioning of this genetic diversity, including historical and contemporary drivers. Population structuring and phylogeography were assessed using nuclear microsatellite genotyping and mtDNA sequencing.

Mitochondrial DNA revealed two previously unrecorded cryptic species, *L. nebulosus* sp. off the coast of South Africa and southern Mozambique and *Lethrinus* sp. A around the Seychelles, within what were previously assumed to be single stocks. The detection of cryptic species and accurate species identification is crucial for future maintenance of genetic biodiversity and management of sustainable fisheries.

High levels of genetic connectivity were observed across the majority of the SWIO in all species. Positive selection detected in a microsatellite locus of *O. cyanea* and mtDNA data for *O. cyanea*, *L. harak* and *L. nebulosus* indicated differentiation and possible emerging isolation of the Mauritian populations of these species from populations across the rest of the SWIO which may reflect local oceanographic processes. Future fisheries management of Mauritius should therefore be carefully considered as Mauritian stocks may be more vulnerable to overexploitation and environmental changes than other SWIO stocks. Additionally, patterns of non-equilibrium conditions were reported across the studied taxa and emphasise how current neutral genetic patterns may underestimate contemporary population differentiation. Future genetic based studies may therefore benefit from adopting genomic approaches.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

1.1 Population genetics and the marine environment

In the marine environment a particular challenge is to understand the processes by which populations become genetically distinct. Historically, the potential for dispersal over large distances, high dispersal potential of adults and/or larvae and the absence of apparent barriers in the marine environment led to the theory of demographically open marine populations over hundreds and even thousands of kilometres (Féral 2002). However, in the last 20 years, research has revealed highly differentiated marine populations (Ayre and Hughes 2004; Hilbish 1996; Reeb and Avise 1990) and the structure of marine populations is now known to range along a cline from completely open to completely closed populations (Cowen and Sponaugle 2009).

Genetic structure in marine populations reflects the historical and contemporary interplay among a complex set of ecological, demographic, behavioural, genetic, oceanographic, climatic and tectonic processes (Coscia and Mariani 2011; Doubleday, et al. 2009; Gerlach, et al. 2007; Janko, et al. 2007; Koizumi, et al. 2012; Lessios, et al. 1999; Olivares-Banuelos, et al. 2008; Palero, et al. 2008; Perez-Losada, et al. 2002; Purcell, et al. 2009; Purcell, et al. 2006; Selkoe, et al. 2006; Selkoe, et al. 2010; von der Heyden, et al. 2008). In particular, physical oceanographic features such as eddies have recently been identified as a factor in the sub-structuring of marine populations, as they can disrupt gene flow between adjacent populations and induce local larval retention (Galarza, et al. 2009; Hoffman, et al. 2012; Jackson, et al. 2014; Mokhtar-Jamai, et al. 2011; Selkoe, et al. 2010). Past climatic fluctuations, particularly during the Pleistocene epoch, greatly affected the distribution of biomes and sea levels, acting as drivers of population sub-structuring and species diversification (Gaither, et al. 2011; Han, et al. 2012; Janko, et al. 2007; Kenchington, et al. 2009; Marko, et al. 2010).

1.2 Molecular markers

Genetic markers are heritable characters, with multiple states for each character, reflecting modifications in the DNA sequence. In a diploid organism each individual may have one or more different states (alleles/nucleotides) for each character (locus). Therefore, the use of separate loci may provide an independent test of hypotheses and the combination of several loci may yield high statistical power (Sunnucks 2000).

Marker selection for population genetics has progressed considerably since its initial use in the 1960s from allozymes, randomly amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP) and amplified fragment length polymorphism (AFLP), to DNA sequencing, microsatellite markers and single

nucleotide polymorphisms (SNPs). A brief description of some of these markers is included below.

1.2.1 **Mitochondrial DNA**

Mitochondrial DNA (mtDNA) is a small extra-nuclear part of the genome found in the mitochondria, organelles that occur in the cytoplasm of most eukaryotic cells. In most metazoa, the mtDNA genome consists of a circular loop of DNA which contains a non-coding region (Control Region or D-Loop), thirty six to thirty seven genes, two ribosomal RNAs, twenty two transfer RNAs and twelve to thirteen subunits of multimeric proteins of the inner membrane of mitochondria (Beebee and Rowe 2001) In vertebrates, mtDNA is a haploid, (in most cases) maternally inherited, closed circular molecule approximately 16,000 base pairs long (Beebee and Rowe 2001).

Mitochondrial DNA possesses a number of unique properties which make it popular as a marker to investigate phylogeography:

1. The higher substitution rate of mtDNA compared with the coding nuclear genome usually results in genetic variation in all but the most inbred populations;
2. Intra-specific nucleotide polymorphism is generally neutral so haplotype distribution is influenced more by demographic events in the population history than by selection;
3. The effective population size of mtDNA is one quarter of diploid nuclear genes so haplotype frequencies can drift rapidly, creating genetic differences among populations on relatively short timescales;
4. mtDNA does not undergo recombination, allowing the genetic relationships among haplotypes to be reconstructed; and
5. Mitochondrial DNA is relatively easy to extract.

Due to these unique properties, mtDNA has traditionally been the molecule of choice for phylogeographic and phylogenetics studies and has been used in over 80% of studies to date (Avise and Walker 1998). In fish species, the control region (CR) an area important in replication and transcription, is commonly used in phylogeographic and phylogenetics studies because of its relatively high mutation rate and polymorphism (Tang, et al. 2006).

Studies of the CR may allow the examination of finer scale demographic features than less polymorphic mtDNA regions over tens of thousands of years. On longer timescales, however, homoplasy may compromise phylogenetic accuracy (Tamura and Nei 1993). Therefore, in addition to the CR, other regions of mtDNA with lower mutation rates have been utilised in phylogenetics and phylogeographic studies such

as cytochrome b, cytochrome oxidase I (COI) and the ribosomal RNA 12S and 16S subunits. The use of the COI gene as a molecular marker is particularly widespread because of the Barcode of Life Initiative (Hebert, et al. 2003), a project which aims to use the COI to differentiate between species and to define new species (DeSalle, et al. 2005).

In recent years there has been an increasing awareness that inference based on demographic history and phylogeography using mtDNA alone may not always be able to accurately reflect demographic, phylogeographic and phylogenetic history of a species (Ballard and Whitlock 2004). Additionally as mtDNA is generally maternally inherited, this may also cause limitations in studies where males and females of the same species differ in behaviour or movement (Beebee and Rowe 2001)

Additionally, as mtDNA is assumed to not undergo recombination, the entire molecule displays the same history, making independent sampling of loci impossible. The value of obtaining additional data from independent markers is illustrated by the number of studies which have displayed conflicts between mitochondrial and nuclear data (Ballard and Whitlock 2004; Rognon and Guyomard 2003; Shaw 2002; Sota 2002). The utilisation of several different single locus markers, including both nuclear and mtDNA markers is therefore preferable in studies of population biology (Ballard and Whitlock 2004; Sunnucks 2000).

1.2.2 **Microsatellites**

Microsatellite markers are the most common nuclear markers used in population studies and are currently one of the most powerful tools for investigating population structure (Selkoe and Toonen 2006; Sunnucks 2000). Microsatellites are tandem repetitions of between one and six nucleotides found at high frequencies (10^4 to 10^5 copies) throughout the nuclear genome and are organised as perfect e.g. $(GC)_n$, interrupted e.g. $(GC)_nAT(GC)_n$ or compound loci e.g. $(GC)_n(AT)_n(GC)_n$ (Estoup and Cornuet 1999). Microsatellites with small repeat units such as mono or di-nucleotides are more common than larger repeat units, which may be due to slippage of bases during replication. Slippage is thought to occur within the protein complex that mediates DNA replication, as a consequence of mis-pairing between the original and newly synthesised DNA strand (Bennett 2000). The resulting region of unpaired DNA is then forced to “loop out” and if this loop is on the new strand, the overall effect is an addition of a repeat unit. Alternatively if the loop is on the template strand it is removed by enzymes and the overall effect is the loss of a repeat unit (Estoup and Cornuet 1999). Larger repeat units require one strand to slip a considerable number of bases before the bases could correctly pair again and this may possibly explain why larger repeat units are less common and are often more stable than microsatellites containing

smaller repeat units (Bennett 2000). The rate at which slippage occurs is not the same as the mutation rate, since most “loops” are repaired by the DNA mismatch repair system (Bennett 2000).

Microsatellites are co-dominant markers which allow the identification of heterozygous and homozygous individuals for the same locus (Selkoe and Toonen 2006). Consequently, it is possible to analyse genetic relationships and structure at both population level, through allelic frequencies, and at the individual level using genotypes (Selkoe and Toonen 2006). Due to the high mutation rate of microsatellites (up to five times higher than other parts of the genome) it is possible to assess fine scale evolutionary processes, making microsatellites invaluable for population studies (Heddrick 2005). Furthermore, microsatellites exhibit highly conserved adjacent sequence regions, known as flanking regions, which can be used to site DNA primers to allow amplification of the targeted microsatellite (Selkoe and Toonen 2006). These characteristics of microsatellites and the small amounts of tissue required for microsatellite analysis have made microsatellites a popular and widely used marker in population biology (Selkoe and Toonen 2006; Sunnucks 2000).

1.3 The Southwestern Indian Ocean (SWIO)

For the purposes of this thesis, the SWIO is defined as the coastline of East Africa and surrounding islands, including; Madagascar, Mozambique, Kenya, South Africa, Tanzania (including Zanzibar), Comoros Islands, the Seychelles, Rodrigues and Mauritius.

1.4 Oceanography of the SWIO

The oceanography of the Indian Ocean was first studied in 1964-66 during the International Indian Ocean Expedition, a basin-wide survey which produced a comprehensive hydrographic atlas (Wyrtki 1973). More recently a number of regional studies in the Indian Ocean have shed light on the oceanography of various areas (de Ruijter, et al. 2002; Schott and McCreary 2001; Schott, et al. 2009).

The oceanography of the SWIO is complex and subject to seasonal variation (Figure 1.1). The main oceanographic features of the SWIO are;

- The South Equatorial Current: flows west across the Indian Ocean towards Madagascar. This current splits off the east coast of Madagascar to form the East Madagascar current and the northern flowing north equatorial current (NEC) (Figure 1.1).

- On reaching the east coast of Africa, the NEC splits into two, forming the northward flowing East African Coastal Current (EACC) and a current which enters the Mozambique Channel.
- Circulation in the Mozambique Channel is complex, forming a series of eddies which may or may not traverse the channel from north to south.
- The East African Coastal Current (EACC) travels north up the coast of Kenya towards Somalia. During the NE monsoon, changing winds slow this current and eventually reverse it forming the Somali Counter Current.
- South of the Mozambique Channel is the start of the Agulhas Current, one of the strongest western boundary currents in the world.
- Major downwelling areas and associated low nutrient waters are present along Tanzania and South Kenya.
- The southern SWIO is characterised by coral reefs and benthic productivity associated with low nutrient, warm waters.

A more detailed description of the most important oceanographic features of the SWIO is provided below.

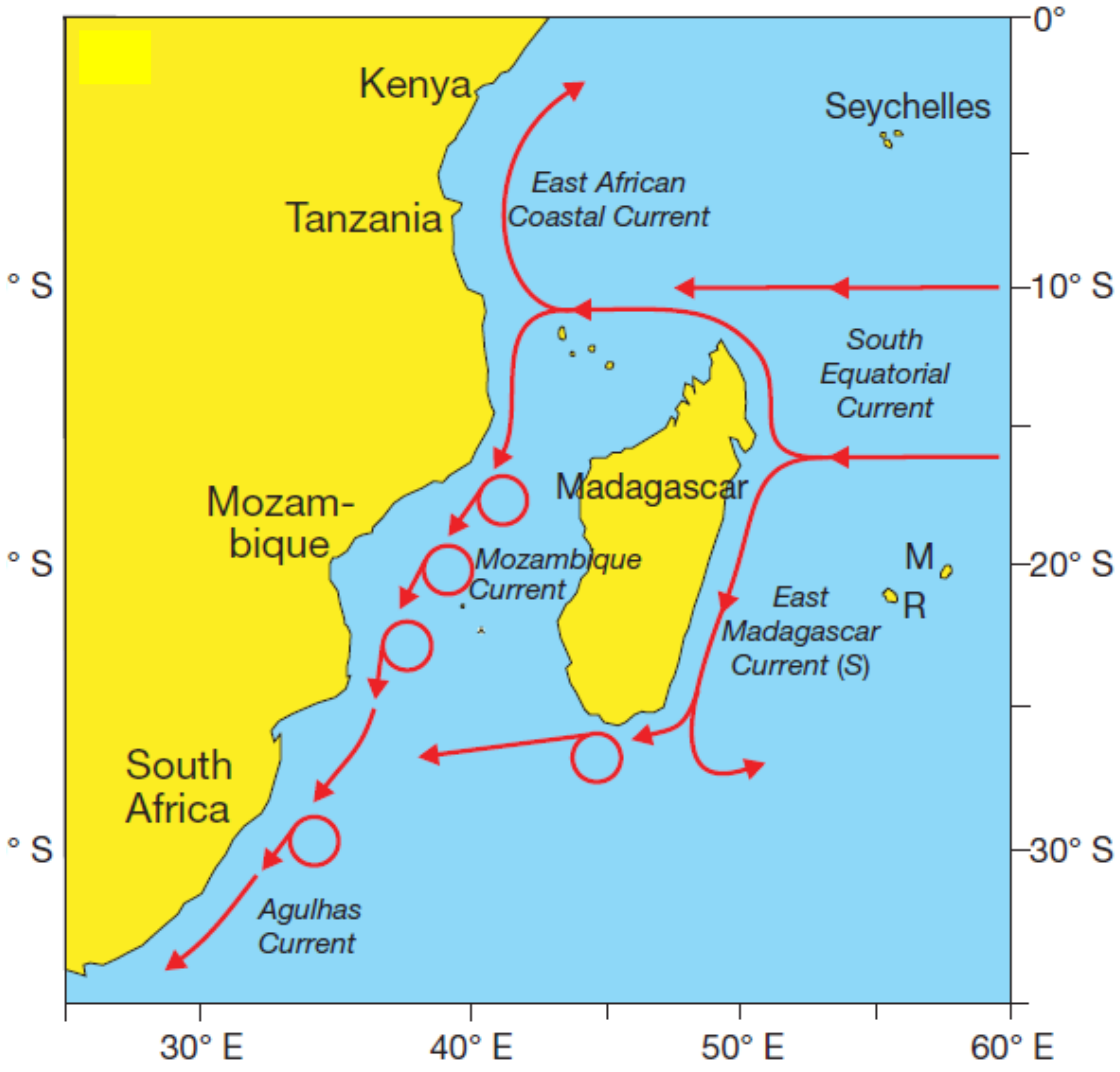

Figure 1.1. Surface water ocean circulation in the Southwestern Indian Ocean. M – Mauritius, R- Rodrigues. Reproduced from Ali and Huber (2010).

The South Equatorial Current

The South Equatorial Current (SEC) flows west across the southern Indian Ocean from the Indo-Pacific (Schott and McCreary 2001) crossing the Mascarene Plateau at 60°E. The Mascarene Plateau is a series of plateaus, banks, shoals and islands, separated by deep ridges or channels extending over 2000 km from the Seychelles to Mauritius. The shoals are generally 20 – 200 m deep, coral topped and may occasionally break the surface. Around the shoals, water depth steeply increases to 2000 – 3000 m, eventually reaching depths of over 4000 m (New, et al. 2007; Smith and Sandwell 1997). The Plateau is composed of four major elevations and banks: the Seychelles Bank, Correira Bank Rise, Saya de Malha Bank and Nazareth Bank (Figure 1.2). Between the Seychelles and the Saya de Malha Bank there is a 400 km ridge, the crest of which is 1000 – 1500 m (Smith and Sandwell 1997). Between the Saya de Malha Bank and the Nazareth Bank there is a deep narrow channel 1100 m deep and 11 km

wide combined with a ridge at 400 – 450 m running to the south-southwest for at least 45 km.

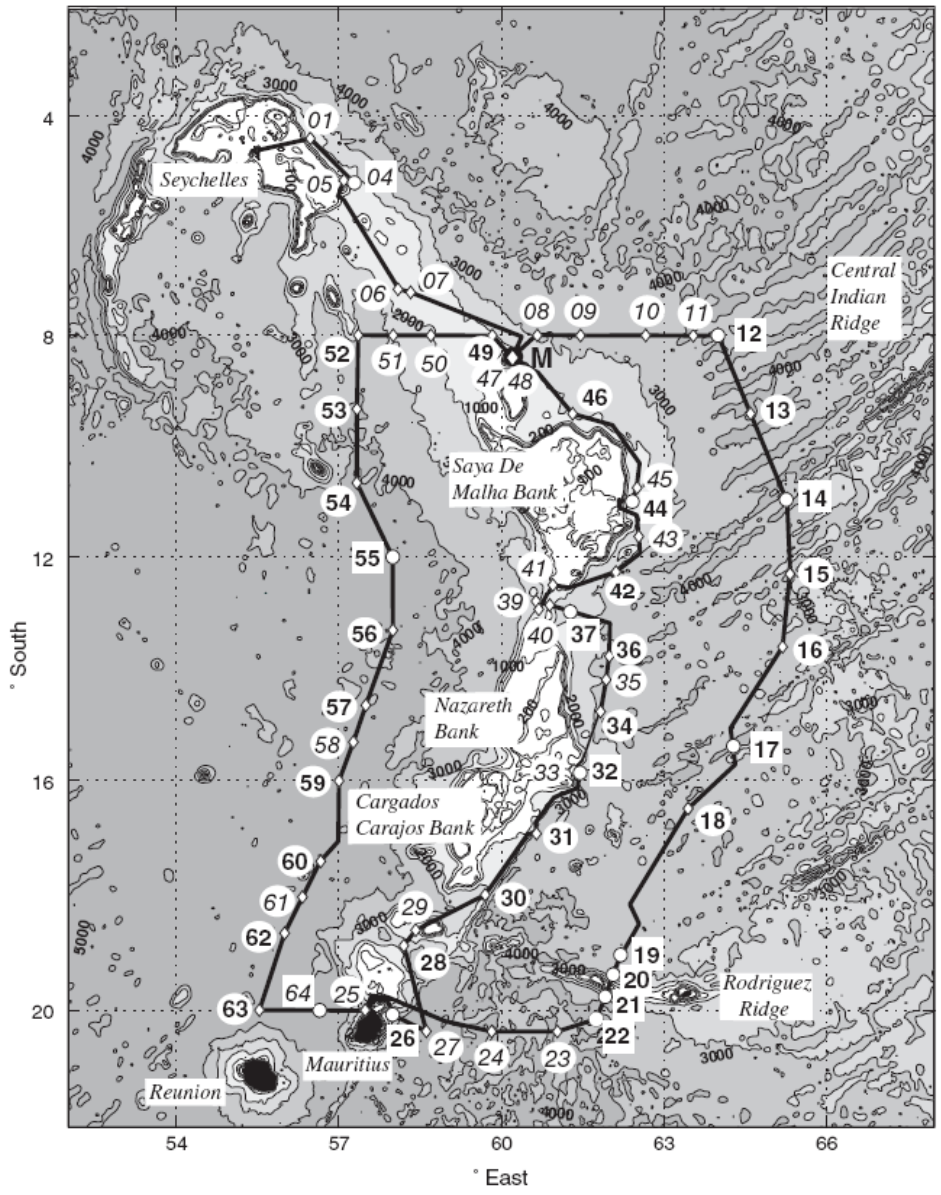

Figure 1.2. Seafloor topography of the Mascarene Plateau and the track of the R.R.S. Charles Darwin cruise June-July 2002 (New, et al. 2007; Smith and Sandwell 1997).

As the SEC approaches the Mascarene Plateau between 10 °S and 16 °S it carries 56 Sverdrups (Sverdrup = 10^6 m^3 s^{-1}: Sv) westwards. Upon reaching the Plateau, 23 Sv is constricted to pass over the sill between the Saya de Malha and Nazareth Banks at 12 °S – 13 °S (Figure 1.3) (New, et al. 2007). This forms the northern core of the SEC between 10 °S and 14 °S downstream of the Plateau. The remainder of the flow passes around the northern edge of the Saya de Malha Bank or between Mauritius and the Cargados Carajos Bank. It is thought that the northern core may retroflect and flow eastwards, joining the South Equatorial Counter-Current (SECC) whilst the remainder of the flow forms a southern core to the SEC.

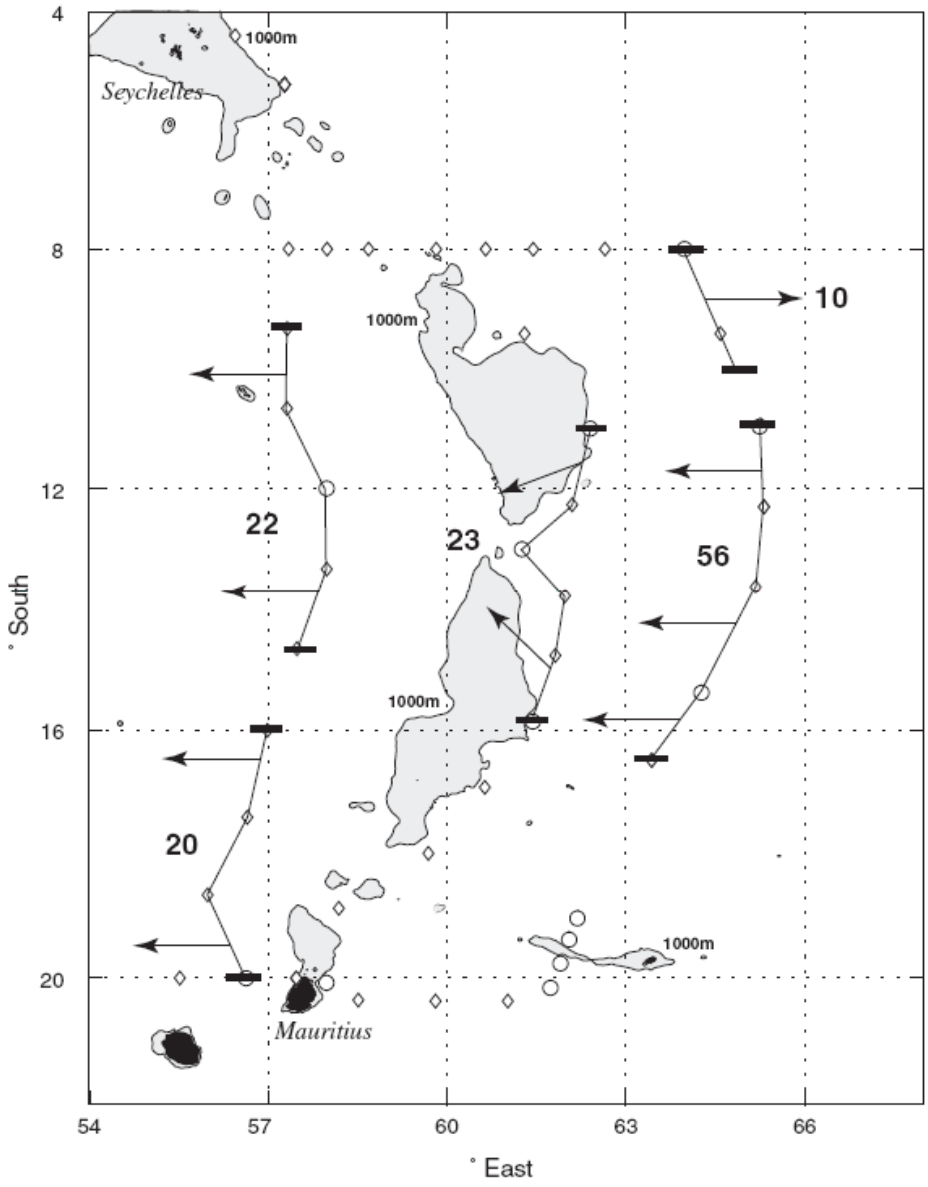

Figure 1.3. Lowered acoustic Doppler Current Profiler transports (Sv) above 2000 m from a cruise in the Mascarene Plateau in 2002. Totals are the transport crossing those cruise sections indicated, arrows show indicative direction of flow reproduced from New, et al. (2007).

The Plateau therefore effectively splits the SEC into two cores and it is thought that upon reaching the coast of Madagascar, these cores may form the northerly flowing Northeast Madagascar Current (NEMC) and the southerly flowing Southeast Madagascar Current (SEMC) (New, et al. 2007). The NEMC transports about 29.6 ±8 Sv (Swallow et al., 1988) and flows up past the north of Madagascar supplying water for the Mozambique Channel and the East African Coastal Current (EACC). The SEMC transports about 20 Sv to the southern tip of Madagascar where it is thought to dissipate in a series of eddies and dipoles which migrate to the African coast (de Ruijter, et al. 2004). Observational data from the Southern Indian Ocean show little seasonal variability and so circulation is thought to be identical in both the South East and North West Monsoon (Schott and McCreary 2001).

The Mozambique Channel

Water flows into the Mozambique Channel from the NEMC. Upon reaching the northern tip of Madagascar, the NEMC splits into two currents, the northward EACC and a southward flow which enters the Mozambique Channel. The flow in the Mozambique Channel was classically thought of as a strong western boundary current which then extended to form the Agulhas Current (Quartly and Srokosz 2004; Schouten, et al. 2003). However, research has now discovered that there is no western boundary current in the Mozambique Channel and the circulation actually takes the form of anticyclonic eddies (Quartly and Srokosz 2004; Schouten, et al. 2003).

The circulation in the Mozambique Channel is very different on the western and eastern side of the Channel. The western side of the channel is dominated by large, anticyclonic eddies. Schouten, et al. (2003) found that on average four anticyclonic eddies travelled southward through the Mozambique Channel per year. The number of eddies in the channel showed a southward decrease from 7 per year in the extension of the South Equatorial Current to 4 per year south of Madagascar. All followed a similar path along the bathymetry of the African coast on the western side of the channel. The speed of southward travel varied along the channel. From 12 °S - 27 °S southward transport was 6 km/day with the exception of 18 °S - 21 °S where it was 3 – 4 km each day. Towards the southern end of the channel (27 °S - 35 °S) the southward transport increased to 8 -10 km/day. Eddy diameter ranged from 300 - 350 km and eddies were unaffected by seasonality. Mean transport through the Mozambique Channel has been measured at 17 Sv (de Ruijter, et al. 2002).

Little research has been conducted into the flow on the eastern side of the channel, but surface drift measurements have indicated that the currents here are weak and variable in direction (Lutjeharms 2006).

1.5 Changes in the SWIO oceanography through time

The Miocene (23-6MYA) was a time of profound change in the world atmospheric and oceanic circulation patterns. In particular, during mid to late Miocene (12.7-7.7 MYA) major climate changes occurred due to the beginning of the uplift of the Isthmus of Panama and expansion of Antarctic ice sheets, leading to a generalised cooling of surface and deep waters (Gourlan, et al. 2008). In the SWIO, this led to a complete reorganisation of the palaeoceanic circulation. A strong westward oceanic current, the Miocene Indian Ocean Equatorial Jet (MIOJet) was established at ~14 MYA, as a result of the progressive closure of the Indonesian Strait and which dominated Indian Ocean circulation until ~ 3.5 MYA (Figure 1.4).

Around the same time as the MIOJet was established, the convergence of the African and Eurasian plates closed the connection of the Indian Ocean to the Tethys Sea (Mediterranean) causing rifting and uplift in East Africa. The closure of the Tethys connection and the formation of the westward flow of the equatorial current in the form of the MIOJet into the Mozambique Channel led to the formation of the Mozambique Current and the accompanying Delagoa Bight Lee eddy during the early/mid Miocene (Preu, et al. 2011). At the end of the Miocene 5.3 MYA, near the Pliocene boundary, the Agulhas current was established off the east coast of Africa. Since then, main path flows in the SWIO have been fairly stable (Martin, et al. 1982).

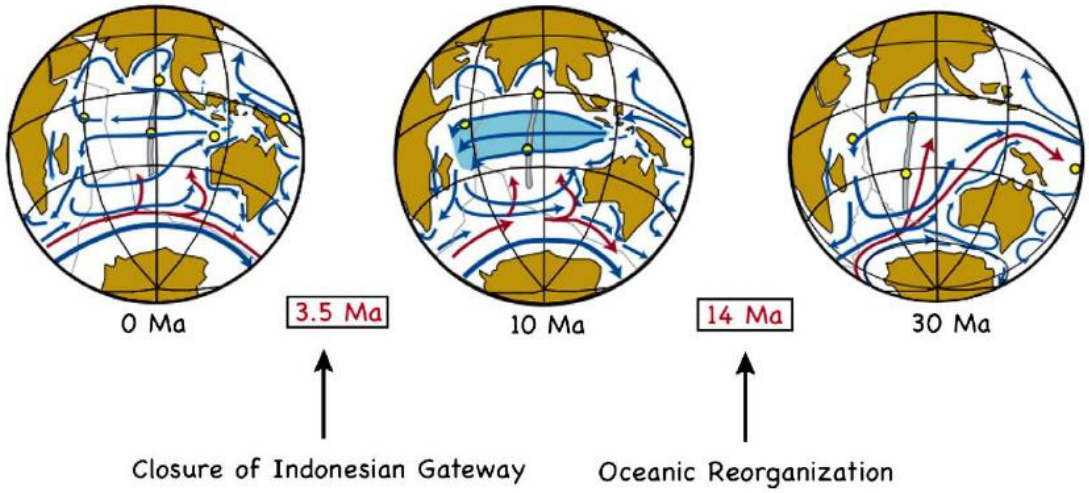

Figure 1.4. Paleogeography and Paleocurrent Reconstruction of the Indian Ocean over the past 30 million years reproduced from Gourlan, et al. (2008)

The Pleistocene epoch (2.5MYA – 11,700 years ago) was a period of considerable global climate change, marked by repeated glacial cycles in which continental glaciers advanced and retreated. At maximum glacial extent, 30% of Earth's surface was covered by ice. This dramatically affected the distribution and diversity of both terrestrial and aquatic organisms (Hewitt 1999; Provan and Bennett 2008). Over 11 major glacial events have been identified as well as many minor glacial events (Pillans, et al. 1998). During the late Pleistocene, sea level in the SWIO was closely linked to the climate fluctuations, with drops in sea level of up to 130 m during glacial events (Lisiecki and Raymo 2005). Sea surface temperature (SST) variations are not available for the SWIO for the majority of the Pleistocene, although it is likely that SST varied considerably between glacial and interglacial periods.

The last ice age occurred between 120,000 and 10, 000 years ago in a period known as the Late Quaternary (Siegert 2002). The end of the Last Glacial Maximum (LGM) marks the end of the Pleistocene epoch. Although global ice sheets varied in size throughout this period, they reached their maximum at around 21,000 years ago. During the LGM, ice sheets had captured so much water that sea levels are predicted

to have been on average 120 -135 m lower than present day, although sea levels in the Indian Ocean range from 109 – 143 m lower than present day (Fleming, et al. 1998). The equatorial region of the Indian Ocean was thermally stable during the LGM, with SST remaining within 2 °C of current day temperatures (Prell, et al. 1980). Within the subtropical areas of the Indian Ocean, cooling of up to 4 °C occurred. This is particularly apparent south of Madagascar, where winter water temperatures are estimated to have been 4°C cooler than present day. During the LGM, there was a northern migration of the 23 °C isotherm to 17.5 °S during August, compared to 30 °S today.

Initial rates of sea level rise from around 21 – 17 thousand years ago (KA) were relatively slow, on average 6 m per thousand years, meaning that these sea levels were maintained for approximately 3000 years. This was followed by a fast period of rise of 10 m per KA for the next 10 KA. Departure from these average rates may have occurred due to the Younger Dryas but most ice sheets were completely decayed by 7 KA, with a small rise of 3-5 m in sea level to present day levels (Fleming, et al. 1998).

The Quaternary in the SWIO was therefore characterised by abrupt changes in sea level and decreased sea surface temperatures in subtropical regions of the SWIO.

1.6 Summary of current knowledge of population structuring in the SWIO

Tropical reefs such as those in the Southwestern Indian Ocean (SWIO) are the most biologically diverse of shallow water marine ecosystems (Roberts, et al. 2002). The SWIO has been identified as a biodiversity hotspot (van der Elst, et al. 2005) with high ichthyofauna richness; over 2200 species of fish have been recorded to date, totalling 15% of the global number of marine fish species (Smith and Heemstra 1986). Despite the SWIO's importance as a biodiversity hotspot it remains relatively understudied with regards to population genetics and phylogeography.

To date most research on marine population connectivity and phylogeography involving the Indian Ocean has focused on population connectivity on an inter-ocean scale (Duda Jr and Palumbi 1999; Gaither, et al. 2011; Gaither, et al. 2010; Hoareau, et al. 2012; Lessios, et al. 1999; McMillan and Palumbi 1995; Read, et al. 2006). These studies typically include only one or two sampling locations within the Indian Ocean and so their ability to assess genetic structuring within the SWIO is limited.

A review entitled "Population genetic status of the Western Indian Ocean (WIO): What do we know?" (Ridgway and Sampayo 2005) found only 20 studies relevant to the Western Indian Ocean, all of which had appeared in the literature within the last six years and were limited to genetic analyses examining variation in allozymes,

mitochondrial DNA or intron variability in nuclear DNA on starfish, coral, crustaceans, holothurians and some fish species. The majority of these studies included in this review showed little genetic exchange between the WIO and the rest of the Indo-Pacific whilst within the WIO, the authors conclude that the limited information available suggests widespread genetic structuring in the reefs off tropical Africa and the Indian Ocean islands (Ridgway and Sampayo 2005).

Since the publication of this review, a number of studies have been carried out on the population structure of marine species in the SWIO which, conversely, appear to show high gene flow between localities within the SWIO. Sequencing of COI mtDNA of the mud crab *Scylla serrata* revealed a lack of genetic population structure throughout the SWIO and the existence of a unique metapopulation spreading from Kenya to South Africa and including the Indian Ocean islands (Fratini, et al. 2010). Sequencing of COI mtDNA and morphometric examination in the fiddler crab *Ucas annulipes* revealed panmictic populations over 3000 km of African coastline (Silva, et al. 2010b). A single panmictic population of bigeye tuna (*Thunnus obesus*), was found throughout the Indian Ocean, revealed by sequencing the hypervariable region 1 (HVR1) of the CR mtDNA (Chiang, et al. 2008) and the Dory snapper, *Lutjanus fulviflamma* which was investigated using AFLP techniques, has low genetic differentiation among populations in the Western Indian Ocean, (Dorenbosch, et al. 2006).

In contrast, a number of studies show some limited population structuring; two species show high gene flow across the majority of the SWIO but display isolation of the Seychelles from the mainland; the blue-barred parrotfish *Scarus ghobban* (Visram, et al. 2010) and the mangrove crab *Neosarmantium meinerti* (Ragionieri, et al. 2010). Muths, et al. (2012) found little structuring across the SWIO but mtDNA sequencing indicated some local structuring in Mauritius of *Lutjanus kasmira*. Interestingly, two studies show structuring along the African coast with a genetic break located somewhere along the Mozambique Channel. In the spiny lobster *Palinurus delagoae* CR mtDNA sequencing revealed highly differentiated northern and southern populations along the African coast, which were split around the interface of the southern end of the Mozambique Channel and the start of the Agulhas Current (Gopal, et al. 2006). Silva, et al. (2010a) detected significant population structuring, using COI mtDNA, in the mangrove crab *Perisesarma guttatum* which displays two highly supported phylogeographic clades; a southern Mozambique clade and a northern clade comprising North Mozambique, Tanzania and Kenya.

From these limited studies it appears that for some benthic invertebrates in the SWIO, the current regime in the Mozambique Channel may act as a barrier to gene flow thereby causing significant structuring in populations north and south of the

Mozambique Channel. The Mozambique Channel circulation consists of southward moving anticyclonic eddies on the western side of the channel and poorly defined weak currents on the east side of the channel (Quartly and Srokosz 2004; Schouten, et al. 2003). Eddies, such as those present in the Mozambique Channel have recently been shown to play a considerable role in the genetic sub-structuring of populations as they may disrupt gene flow between adjacent populations and encourage local larval retention (Galarza, et al. 2009; Hoffman, et al. 2012; Jackson, et al. 2014; Mokhtar-Jamai, et al. 2011; Selkoe, et al. 2010). Circulation in the Mozambique Channel has been relatively constant for the past 15 million years (Obura 2012) and it was therefore predicted that it may have acted as a long term barrier to gene flow, ultimately generating separate phylogeographic lineages north and south of the channel.

1.7 The SWIO ecosystem

Tropical reefs such as those in the SWIO are the most biologically diverse of shallow water marine ecosystems (Roberts, et al. 2002). The SWIO has been identified as a biodiversity hotspot (van der Elst, et al. 2005) with high ichthyofauna richness; over 2200 species of fish have been recorded to date, totalling 15% of the global number of marine fish species (Smith and Heemstra 1986). Whilst this high icthyodiversity is scientifically interesting, it creates problems for resource management as only a small number of the 1051 species listed as relevant to the region's fishers (Fischer and Bianchi 1984) have been assessed or researched (van der Elst, et al. 2005). Therefore, developing sustainable harvesting strategies for many of the fisheries are compromised through lack of information.

Despite accounting for approximately 8% of the world's ocean, the Western Indian Ocean, which encompasses the SWIO, is only moderately productive, contributing 4% to the industrial global catch (FAO 2013). In addition, coastal productivity is small for such a large area. Coastal fishery yield along the entire western boundary of the Indian Ocean (South Africa to Arabia) including various island states represents less than 1% of global catches (FAO 2013) and most coastal fish stocks are considered to be fully exploited or even overexploited (Guard 2009).

The reason for this moderate productivity is not entirely clear, although the SWIO lacks the large nutrient upwelling systems of other regions and narrowness of the continental shelf along the coast of Africa which are often suggested as contributing factors (FAO 2013; Fischer and Bianchi 1984; van der Elst, et al. 2005). In addition, the reporting of landings in artisanal and subsistence fisheries is often not established in developing countries, such as those surrounding the SWIO, because of lack of infrastructure and detailed species-specific catch data. Ardill and Sanders (1991) estimated that non-reporting to the FAO of artisanal landings prior to 1988 was around 69.7% for 8

countries within the SWIO. Since 1988, more countries in the SWIO have reported artisanal catches in their fisheries statistics. Annual artisanal catch values are often very inconsistent, however, suggesting inaccurate recording of catches (van der Elst, et al. 2005). Until submissions of fishery statistics are accurately reported across the SWIO, the absolute quantities landed will remain uncertain. It is possible that FAO data may only reflect half of actual catches in SWIO (van der Elst, et al. 2005).

1.8 Study species

1.8.1 Study species selection

Four study species, *Octopus cyanea*, *Lethrinus harak*, *L. nebulosus* and *L. mahsena* were selected for this research for the following reasons:

1. **The four species have a similar geographic range throughout the SWIO** To investigate how oceanographic conditions and past climate may have shaped population structure in the SWIO it is useful to compare taxa with similar geographical distributions. Geologic processes such as glaciations, sea level changes and tectonic activity have altered ecological conditions through time and these past events may have promoted biological diversification and patterns of genetic variation within species (Riginos and Harrison 2005). Modern intraspecific divisions could have resulted from two non-mutually exclusive processes. First, a changing landscape could have imposed temporary physical barriers to gene exchange. The genetic structures of this past vicariance may be evident within species whose distribution is currently continuous (Avise and Walker 1998). However, distinct environmental conditions in neighbouring regions could also result in adaptation of populations to local conditions and thus restrict gene flow, similarly resulting in genetic discontinuities within species (Riginos and Harrison 2005). Phylogeographic comparisons across multiple taxa may therefore clarify the histories of biogeographic regions. In particular, historic barriers to movement should have affected multiple species and this should result in a pattern of concordant intraspecific genetic divisions among species (Kuo and Avise 2005; Riginos and Harrison 2005). It is therefore useful to compare multiple taxa with similar geographical distribution.
2. **The four species all possess a planktonic larval phase** - For benthic marine species, which have adult stages that do not generally migrate large distances, dispersal primarily occurs during the pelagic larval phase or the earliest life history phase (Cowen and Sponaugle 2009). The dispersal of this pelagic larval phase, combined with the settlement, survival and reproduction of these individuals creates gene flow between marine populations, providing the link

between marine connectivity and population genetics. Oceanographic features such as currents, eddies and fronts are known to disrupt gene flow between populations by limiting dispersal of the pelagic larval phase. The use of species with a pelagic larval phase should therefore allow the investigation into whether the anticyclonic eddy system in the Mozambique Channel has led to genetically divergent sub-populations occupying the northern and southern coasts of the Mozambique Channel.

3. **The four species have varying pelagic larval duration (PLD)** - Although the species chosen have similar life histories which include pelagic larvae, the planktonic larval duration (PLD) varies with species. This should allow the investigation as to whether any differences between population structure across species are due to variations in life history characteristics rather than oceanographic conditions or past climate fluctuations.

4. **All study species are valuable fisheries resources in the SWIO** – There is currently very little data on the genetic diversity of species in the SWIO (Ridgway and Sampayo 2005). In recent years major signs of environmental degradation and declines in natural resources and biodiversity have been observed in the SWIO (Berg, et al. 2002; Ridgway and Sampayo 2005). Effective conservation and management informed by empirical population genetic studies is therefore required to prevent any further decline. Therefore any data generated will contribute to future fisheries management in the region.

1.8.2 *Octopus cyanea* (Gray, 1849)

Commonly known as the big blue octopus or day octopus, *O. cyanea* is a benthic species and is widespread on coral reefs throughout the Indo Pacific, from East Africa to the Hawaiian Islands and in tropical and warm waters including the Red Sea, India and Australia (Figure 1.5).

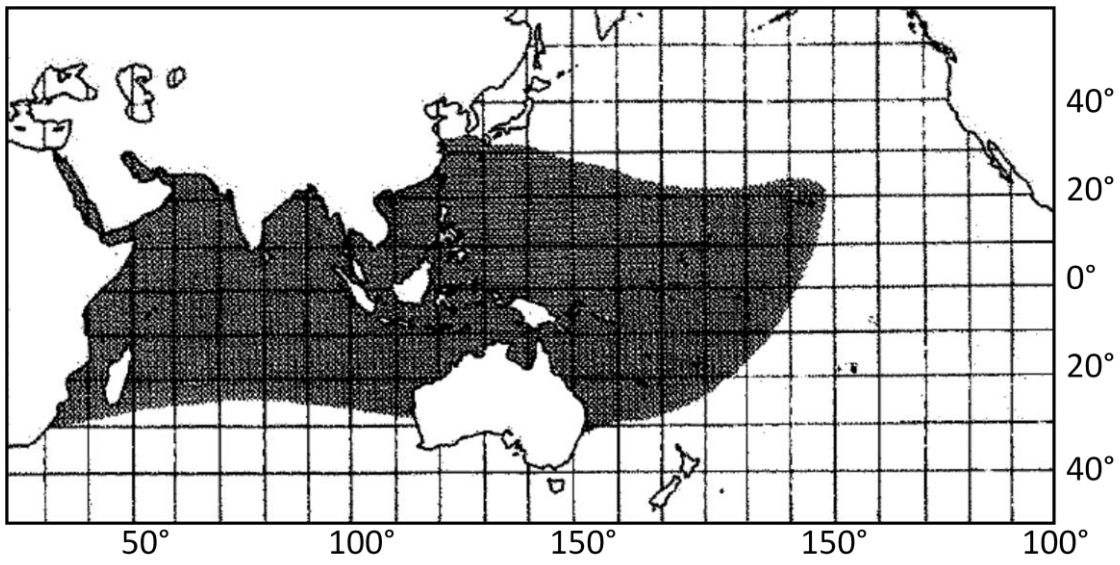

Figure 1.5. Current range of *O. cyanea* reproduced from FAO species catalogue (Roper, et al. 1984).

It is a large octopus on average weighing about 6kg but can reach up to 11.7 kg (Guard and Mgaya 2002; van Heukelem 1973) and unusually for octopuses, it hunts during daylight hours (Figure 1.6). Its mantle is round to oblong, smooth-skinned with a few large tubercles. It has one large cirrus over each eye, usually with 2 secondary smaller tubercles and a dark purple ringed ocellus or “eye spot” located on the web on each side between the eye and the base of arms 3 and 4. The mantle is mottled and the arms have purple-brown blotches (Roper, et al. 1984).

Figure 1.6. Photograph of *O. cyanea* on a reef in the SWIO (reproduced with the permission of Andy Murch/Elasmodiver)

O. cyanea has a maximum lifespan of 400 days (van Heukelem 1973) for males and females. Females may copulate many times but reproduce only once, barricading themselves in their den, laying eggs and then brooding the eggs for 20-30 days until they hatch (van Heukelem 1976). Females die shortly after the eggs hatch. Males die at approximately the same age but may mate with many females (van Heukelem 1976).

There is some debate in the literature as to whether *O. cyanea* exhibits seasonal reproduction. According to van Heukelem (1973) *O. cyanea* does not exhibit seasonal reproduction. However, van Heukelem observes that local fishermen know that large octopuses are more abundant in the autumn, which would suggest some form of seasonality. Research into the artisanal fishery for *O. cyanea* in Tanzania found clear evidence of seasonal reproduction at one of three sites sampled (Guard and Mgaya 2002). *O. cyanea* were found to reproduce throughout the year but peak brooding periods occurred in June, and to a lesser extent in December. This was supported by both an increase in mature individuals in the catch and a peak in the gonadosomatic index. The other two sites studied showed peaks in average weight which were less pronounced and did not follow the same basic pattern as the other site, perhaps suggesting some other control on reproduction (Guard and Mgaya 2002).

Benthic octopuses of the family Octopodidae adopt two different major life-history patterns (Villanueva and Norman 2008). The first is the production of a few, large eggs resulting in well-developed hatchlings that closely resemble adult form and quickly settle into the benthic habitat of their parents. The second strategy is the production of numerous small eggs that hatch into planktonic free-swimming larvae known as paralarvae. In octopuses, egg size relative to mantle length is used as a predictor of life history strategy (Villanueva and Norman 2008). Octopus species with eggs <10% mantle length are likely to produce planktonic hatchlings. *O. cyanea* is recorded as having a very small egg size to mantle length (ELI_{max}). *O. cyanea* has an ELI_{max} of 1.7 making it highly likely to possess a planktonic larvae. Observations of *O. cyanea* larvae in the laboratory also confirm that *O. cyanea* has a planktonic life history phase (van Heukelem 1973).

Immediately after hatching, larvae are approximately 2.5 mm long and 1 mm wide, strongly phototropic and negatively geotropic (Villanueva and Norman 2008). Larvae swim actively using their siphons (Dew 1959). Paralarvae differ from conspecific adults in their morphology, physiology and behaviour (Villanueva and Norman 2008). The duration of the pelagic larval phase ranges from 3 weeks to 6 months in Octopodidae. Observations of *O. cyanea* larvae in the laboratory confirm that *O. cyanea* has a life history with a considerable PLD although larvae have not been raised to settlement in

the laboratory so exact larval duration times are unknown (van Heukelem 1973) but are likely to be between 50 days and 2 months, similar to *O. vulgaris*, which employs a similar life history strategy (Villanueva 1995). After time in plankton, larvae metamorphose to adult form and settle to live a benthic life.

Fisheries of *O. cyanea*

O. cyanea is fished along the East African coastline and the surrounding islands (Guard and Mgaya 2002). Octopus fishing is a valuable source of food and money to coastal communities in the SWIO. Traditionally octopus are collected from intertidal reef flats and sub-tidal inner reefs either by walking on the reef or using snorkel equipment. When a hole containing an octopus is found, a metal or wooden stick is inserted into the hole. The octopus seizes the stick and is then pulled out of the hole. The stick is used to kill the animal by pushing it through the beak and into the brain (Guard and Mgaya 2002). In recent years fishing activity for octopus has increased in the SWIO due to the rising demand of exports, with local catches of octopus often sold as *O. vulgaris* to appeal to the European market (A. Taylor personal observation). Outside buyers, who process and export octopus, now operate along the coastal regions of Tanzania, Mozambique and Madagascar and may use specially commissioned boats to take fishers to fishing sites (Guard 2009). In areas such as the Seychelles, scuba equipment is now being used to increase the available reef area that can be fished for octopus (A. Taylor, pers. obs.).

Exact catch statistics for *O. cyanea* are difficult to ascertain. The FAO fisheries yearbook for 2011 does not identify *O. cyanea* by species, despite its importance in SWIO fisheries but classifies it under the general octopus category (FAO 2013). Across the Western Indian Ocean, the dominant octopus species in the artisanal catch is *O. cyanea* (Guard 2002, A Taylor pers. comm.). Capture production for the entire Western Indian Ocean, (FAO fishing zone 51), of octopus including *O. cyanea* from 2002 – 2011 ranged from 1020 – 2535 tons annually. However, this value is likely to be considerably larger as a number of countries in this fishing area submit catch data under the more general term of cephalopods instead of specifically by group (India, Somalia, Madagascar and Mozambique).

Of the SWIO countries which do submit specific catch data for octopus, Tanzania has the highest catch production with between 700 – 1700 tons harvested annually (Figure 1.7). Mauritian catch statistics show a sharp decrease in catch from 335 tons in 2002 to 84 tons in 2006 with catches remaining low from 2006 up to the latest FAO catch data in 2011 (FAO 2013), possibly suggesting overexploited populations in this area.

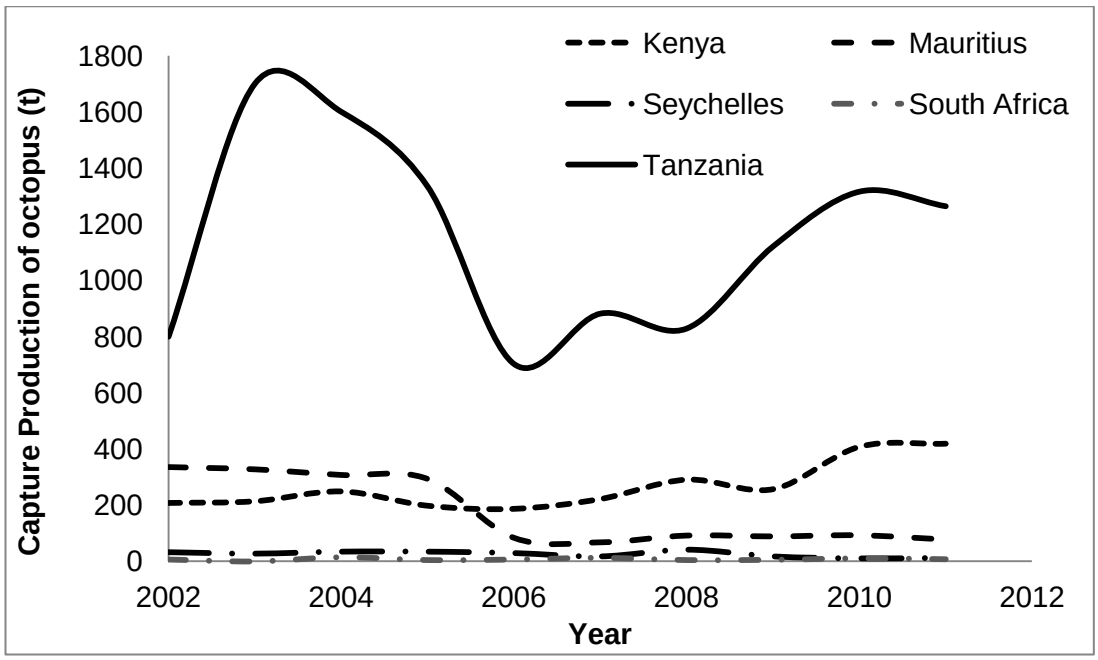

Figure 1.7. Annual catch data for octopus fisheries by country in the Southwestern Indian Ocean.

Catch data specifically on octopus has not been submitted for Madagascar and Mozambique (FAO 2013) with these countries instead reporting total cephalopod catch annually (Figure 1.8). Of this catch, a significant proportion is expected to be *O. cyanea* as significant artisanal and export fisheries for *O. cyanea* operate in both countries (A Taylor pers. obs). Catches for both countries show a steady increase in from 2002 – 2011 (Figure 1.8).

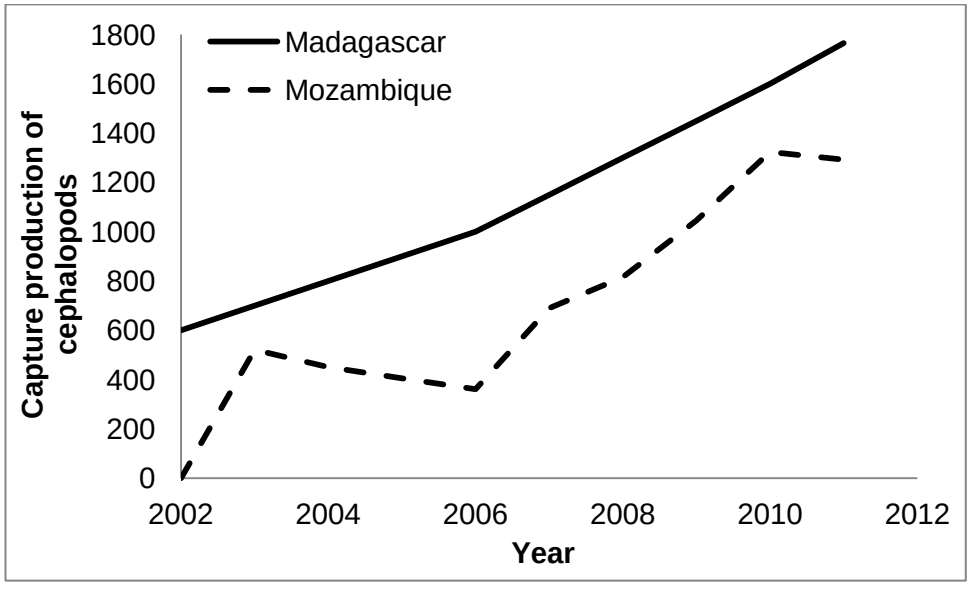

Figure 1.8. Annual catch data for cephalopod fisheries in Madagascar and Mozambique in the Southwestern Indian Ocean.

Exports of *O. cyanea* have increased markedly across the SWIO in recent years (Guard 2009; Guard and Mgaya 2002). Guard reports that in Madagascar a 35% increase in exports of octopus to France was reported between 2002 and 2003 (FAO

2013). However, since 2006, the FAO has not provided totals of octopus exports from individual countries making estimates of the extent of exports difficult to quantify accurately.

Despite the socio-economic importance of the octopus fishing industry and the increasing fishing intensity, little research has been conducted into the stocks of *O. cyanea* across the SWIO. The most extensive research to date has been conducted in Tanzania (Guard and Mgaya 2002). Research into the artisanal octopus fishery in Tanzania analysed specimens from three sites from April 2000 until June 2001. The total economic value of the octopus fishery at these sites for the local fishermen during this time period was $245,000, a considerable income for this area. At two of the sites average weight and catch per unit effort (CPUE) were significantly lower than at the third site. Higher abundance of smaller individuals and a lower overall biomass was also observed at the first two sites. The results suggest overfishing of octopus may be occurring at two of the sites, significantly reducing both reproductive output and success. Interestingly, capture records suggest that females may be more prone to capture during brooding periods, possibly due to the obviousness of their dens, an important consideration in management of stocks (Guard and Mgaya 2002).

With fishing pressure on *O. cyanea* in the SWIO increasing and the lack of stock assessment and management of this species, research to accurately identify population sub-structuring patterns and assess current biodiversity levels are a key step towards understanding the impact of fishing and providing a base for future sustainable harvesting plans.

1.8.3 ***Lethrinus* species**

Emperor fish, *Lethrinus* sp. (Family Lethrinidae) are indigenous to the tropical and subtropical Indo-Pacific region, encompassing the Indian Ocean and Western Pacific (Carpenter and Allen 1989). Their life history is composed of a planktonic larval phase, juvenile phase and demersal adult phase. Larval duration for *Lethrinus* sp. ranges from 28 – 37 days (Nakamura, et al. 2010). A size related discrepancy in male to female sex ratio has been observed in *Lethrinus* with males generally predominating in smaller sizes and females in the larger sizes (Kulmiye, et al. 2002), in keeping with the fact that most *Lethrinus* sp. display protogynous hermaphroditism (Bertrand 1986; Ebisawa 2006; Ebisawa and Ozawa 2009; Kulmiye, et al. 2002).

1.8.4 ***Lethrinus mahsena* (Forsskål, 1775)**

L. mahsena (Figure 1.9) is a tropical reef-associated fish, found on coral reefs and adjacent sand and seagrass patches up to depths of 100 m throughout the Western Indian Ocean including the Red Sea and the coast of East Africa to Sri Lanka

(Carpenter and Allen 1989). *L. mahsena* can reach lengths of 65 cm and feeds on echinoderms, crustaceans, fish, molluscs, tunicates, sponges and polychaetes (Carpenter and Allen 1989).

L. mahsena has a maximum lifespan of 27 years and is a protogynous hermaphrodite (Bertrand 1986; Grandcourt 2002). Individuals become sexually mature at 2 – 3 years and sex reversal takes place between the ages of 5 - 7 years (Bertrand 1986). Research by the Mauritian government on Saya de Malha Bank suggests spawning occurs throughout the year with two peaks of spawning activity in December-January and May-June (Bautil and Samboo 1988). Prior to this research, a single spawning season was thought to occur each year between October and February (Bertrand 1986). *L. mahsena* possesses planktonic larvae, although the length of the larval duration is at present unknown.

The appearance of *L. mahsena* varies across its range (Figure 1.9), with some individuals possessing a bright red oblique streak from above to below the pectoral fin base, a morphotype previously identified as *L. sanguineus* (Carpenter and Allen 1989). Morphological examination of both colour types has shown no other discernible difference between the two colour variants (Carpenter and Allen 1989). Additionally, the red marking is highly variable with many intermediates between the red streak (*sanguineus* patterning) and the red pectoral fin base commonly seen in *L. mahsena*. As a result, these two morphotypes are considered a single species. Genetic analyses of these two morphotypes has not yet been undertaken.

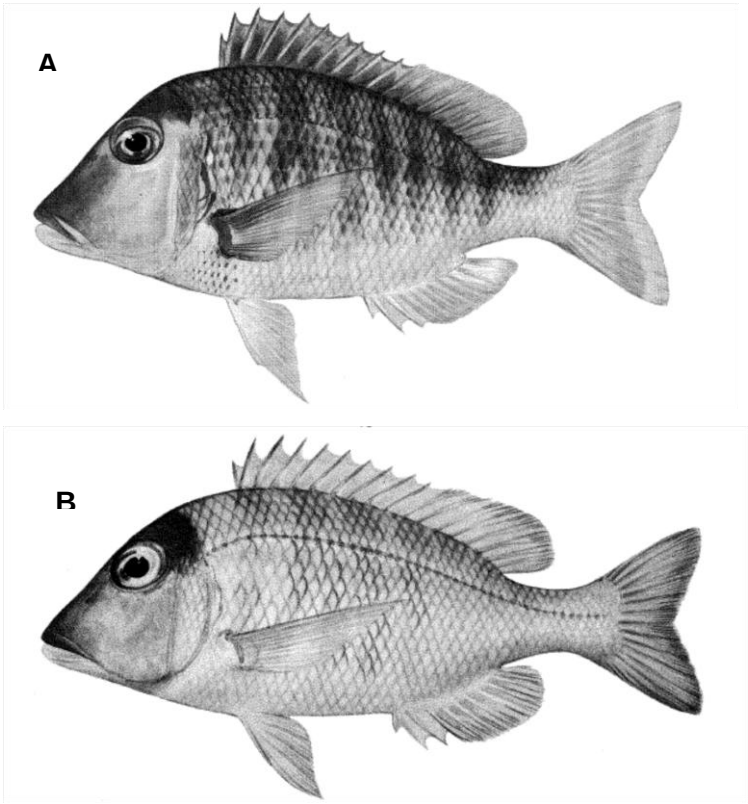

Figure 1.9. *L. mahsena* morphotypes **A**: *L. sanguineus* pattern, **B**: *L. mahsena* pattern (reproduced from Carpenter and Allen 1989).

Fisheries of *L. mahsena*

L. mahsena is fished throughout the SWIO and is a particularly important component of the Banks fisheries, where it contributes up to 90% of the catch. The Banks fisheries comprise the Nazareth and Saya de Malha Banks, large submerged banks along the Mascarene Ridge. The fishing season extends from September to July (Bautil and Samboo 1988). Small dories using handlines are worked in conjunction with mother ships to fish depths from 15 to 50 m. The fish are gutted shortly after capture and landed in Mauritius as frozen product (Bautil and Samboo 1988).

Information on the Banks fisheries after 1988 and more general statistics on fisheries of *L. mahsena* throughout the rest of the SWIO are limited as countries do not report species catch data on *L. mahsena* (FAO 2013). An additional concern is that the Banks fisheries are remote and beyond any national Economic Exclusion Zone (EEZ) and as such are not governed by fishing regulations or subject to patrols. High seas' fishing is therefore likely to increase in the future, further increasing the pressure on *L. mahsena* in the SWIO.

1.8.5 ***Lethrinus nebulosus* (Forsskål, 1775)**

Lethrinus nebulosus is a tropical reef-associated fish, widespread in the Indo-West Pacific including the Red Sea and Arabian Gulf, East Africa to southern Japan and Samoa (Figure 1.10, Carpenter and Allen 1989). It inhabits nearshore and offshore coral reefs, coralline lagoons, seagrass beds, mangroves, coastal sand and rock areas up to depths of 75 m. *Lethrinus nebulosus* feeds on echinoderms, molluscs, crustaceans and occasionally polychaetes and fish (Carpenter and Allen 1989). It can reach lengths of up to 80 cm, or more commonly 20 – 50 cm.

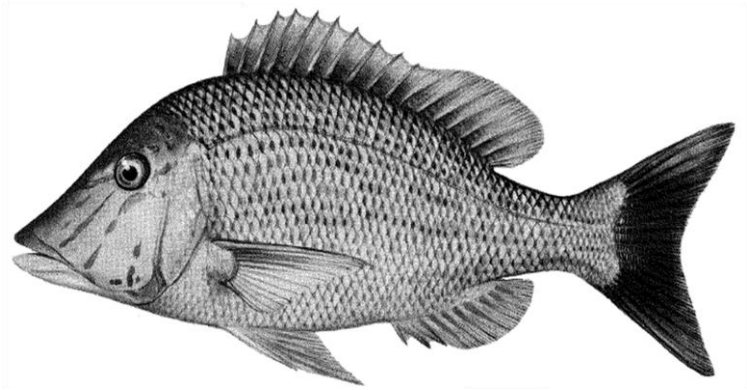

Figure 1.10. *L. nebulosus* (reproduced from Carpenter and Allen 1989).

The life history, population dynamics, demography and reproductive biology of *L. nebulosus* have been studied extensively in the Gulf of Aden and Red Sea (Aldonov and Druzhinin 1979; Edwards, et al. 1985). Like most emperor fish*, L. nebulosus* is a protogynous hermaphrodite. Sex change occurs at 2-3 years and at a mean length of 277.5 mm (Marriott, et al. 2010). *L. nebulosus* spawns once a year and is a multiple spawner (Gouws 2012). Spawning activity peaks between April and July in the northern Red Sea, in March to May and September in the Gulf of Aden (Edwards, et al. 1985) and between February and June in the Persian Gulf and Oman Sea (Grandcourt, et al. 2006). Sexual maturity is reached between 4.5 - 6 years. The length of larval duration was 37 days at the Great Barrier Reef in Australia (Brothers, et al. 1983) but only 28.4 ± 2.1 days off the Southern Ryukyu Islands of Japan (Ebisawa and Ozawa 2009; Nakamura, et al. 2010). Size at settlement is recorded as 18.6 ± 4.2 mm (Nakamura, et al. 2010).

Fisheries of *L. nebulosus*

L. nebulosus is a major component of artisanal, subsistence, recreational and commercial fisheries throughout its distribution (Gouws 2012) and is considered to be one of the best food fishes in many countries. In the SWIO *L. nebulosus* is caught from coral and rocky reefs and coastal waters using bottom trawls, seine nets, traps, hand-lines and gill nets. The species is commercially important in Mauritius, India and Sri

Lanka (Fischer and Bianchi 1984; van der Elst, et al. 2005). Despite this, published information on the importance of this species to artisanal, subsistence, recreational and commercial fisheries in the SWIO is scarce and species catch data are not available (FAO 2013; Gouws 2012). In South Africa, *L. nebulosus* is a minor component of both recreational and commercial fishing. However, its importance is increasing as the abundance and catches of other preferred species decline (Mann 2000).

1.8.6 ***Lethrinus harak* (Forsskål, 1775)**

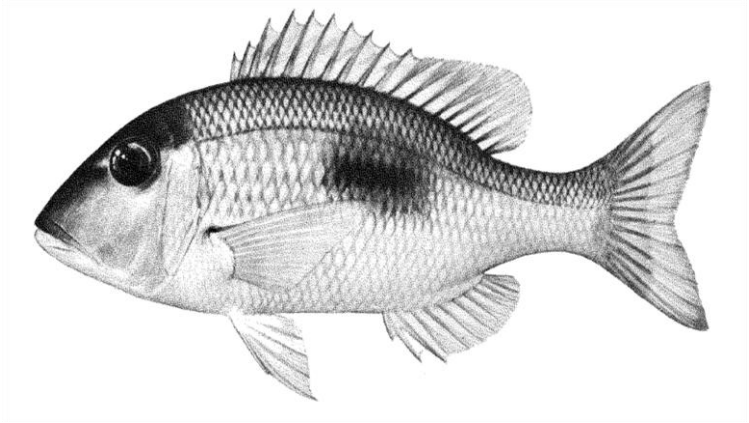

Figure 1.11. *L. harak* (reproduced from Carpenter and Allen 1989)

Lethrinus harak (Figure 1.11) is found throughout the Indian Ocean and western Pacific, including the Red Sea, East Africa, Seychelles, Maldives and Sri Lanka. It inhabits shallow, sandy coral rubble, mangroves, lagoons, inshore seagrass areas and areas adjacent to coral reefs (Carpenter and Allen 1989). *L. harak* has an extended spawning season extending from October to April with peaks in October and February in Kenya (Kulmiye et al. 2002). In the Ryukyu Islands, spawning period was from April to November (Nakamura et al. 2010). Average larval duration is 29.2 days (Nakamura et al. 2010).

Fisheries of *L. harak*

L. harak is an important component of reef-based fisheries and is caught by shore seines, gill nets, traps and handlines (Carpenter and Allen, 1989). *L. harak* is one of the most abundant and commercially fished species in Kenya (Kulmiye et al. 2002). The Fishery Statistics Year Book of 1996 lists Lethrinids together with Siganids as the most important marine fish, constituting about 31% of total reef fish landings over the last 5 years in Kenya (FAO 2013). Despite this, published information on the importance of this species to artisanal, subsistence, recreational and commercial fisheries in the SWIO is scarce and species catch data are not available after 1996 (FAO 2013).

The life history characteristics of Lethrinidae; slow growth, longevity and protogynous hermaphroditism, make *Lethrinus* sp. particularly vulnerable to overexploitation.

Species which are protogynous hermaphrodites are particularly vulnerable to the effects of fishing (Young and Martin 1982) as size-selective fishing may lead to the removal of larger individuals, which are predominantly one sex, reducing reproductive success and subsequent recruitment (Marriott, et al. 2010).

1.9 Aims and objectives

Despite the SWIO's importance as a biodiversity hotspot it remains understudied with regards to genetic characterisation of this biodiversity and its evolutionary drivers. In recent years major signs of environmental degradation and declines in natural resources and biodiversity have been observed in the SWIO (Berg, et al. 2002; Ridgway and Sampayo 2005). Effective conservation and management, informed by empirical population genetic studies, is therefore required to prevent any further decline. Ridgway and Sampayo (2005) outline three important research directions for the SWIO:

1. Measurement of population genetic structure;
2. Study of the influence of historical biogeographical events on present day populations; and
3. Use of more realistic models to understand the genetics of natural populations.

This thesis set out to address numbers one and two of these research themes: specifically this thesis investigates intra-specific genetic diversity across the SWIO in a number of species and examines the underlying factors associated with the partitioning of this genetic diversity. Additionally, where sufficient geographic samples could be obtained, the role of the Mozambique Channel oceanography in population structuring of populations north and south of the Mozambique Channel was investigated. The empirical data may be incorporated into the generation of more accurate models of natural populations in the future.

Chapter 2: Methodology

2.1 Sampling

Samples were collected from a number of locations across the Southwestern Indian Ocean (SWIO) with emphasis on sample ranges encompassing areas where known physical oceanographic features may influence genetic structuring (Table 2.1). Samples were collected from fish markets and local subsistence fishermen and were preserved in 100% ethanol. Samples were collected by the PhD student from Mozambique and the Seychelles. The remaining samples were collected by various collaborators.

2.2 Molecular techniques

2.2.1 DNA extraction

Total genomic deoxyribonucleic acid (DNA) was extracted using a phenol/chloroform/isoamyl alcohol (IAA) method adapted from (Sambrook, et al. 1989) and (Winnepenninckx, et al. 1993). Precipitated DNA was suspended in 100µl (microlitres) of Sigma water and stored at -20°C.

The quality of extracted DNA (i.e. high molecular weight) was assessed using gel electrophoresis with ethidium bromide staining and size comparison with Bioline size marker Hyperladder II. The concentration and purity of DNA was assessed using a spectrophotometer (Nanodrop).

2.2.2 PCR amplification and sequencing of mitochondrial DNA

Octopus cyanea

For *O. cyanea*, three mitochondrial regions, cytochrome oxidase I (COI), 16S ribosomal DNA (16S rDNA) and octopus non-coding region (NCR) were amplified using polymerase chain reaction (PCR) and subsequently sequenced (Table 2.2).

COI mtDNA was amplified as it is the universal bar-coding gene and is frequently used to assess phylogeny and phylogeography in marine taxa (Silva, et al. 2010a; Strugnell and Lindgren 2007; Takumiya, et al. 2005; Ward and Holmes 2007; Ward, et al. 2005). Species-specific primers (Table 2.2) were designed using Primer 3.0 (Untergasser, et al. 2012) with partial COI sequences of *O. cyanea* from GenBank (AB430534.1, AB430535.1, GQ900742.1, GQ900741.1, GQ900740.1) and used to amplify a 628bp segment of COI.

Table 2.1. Sampling locations in the Southwestern Indian Ocean for the study species used in this thesis.* indicate details of specific sampling locations were not provided for these localities.

Species	Country	Location	Sample Code	Sample Size
Octopus cyanea	Madagascar	Andavadoaka	AN	80
		Ankiembe	AK	80
		Beheloke	BE	80
	Zanzibar, Tanzania	Kizim Kazi	KK	34
		Stone Town fish market	ST	80
	Tanzania	Kunduchi	ZZ	60
	Kenya	Mombassa fish market	MO	58
	Rodrigues	N.Rodrigues	NR	51
		S.Rodrigues	SR	52
	Mauritius	Mahebourg	MA	80
	Seychelles	North Island	MS	49
		Grand Anse, Praslin	PS	62
	Madagascar	Ivovona	I	80
	Mozambique	Murrumbue Beach, Pemba	PE	52
		Vamizi Island	VI	13
Lethrinus atkinsoni sp.	Seychelles	North Island	MS	10
		Grand Anse, Praslin	PS	9
Lethrinus harak	Madagascar	Andavadoaka	AN	23
		Belo Sur Mer	B	51
	Zanzibar Tanzania	Stone Town fish markets	ZZ	78
	Kenya	Mida Creek, Dabaso	DA	24
	Mauritius	Mahebourg	MA	43
	Mozambique	Pemba	PE	50
Lethrinus mahsena	Madagascar	Belo sur Mer	B	5
	Zanzibar Tanzania	Stone Town fish market	ZZ	18
	Seychelles	North Island, Mahe	MS	3
	Mauritius	Mahebourg	MA	20
	Mozambique	Maputo fish market	MM	10
		Pemba	PM	7

Table 2.1. (continued) Sampling locations in the Southwestern Indian Ocean for the study species used in this thesis.* indicate details of specific sampling locations were not provided for these localities.

Lethrinus nebulosus	Madagascar	Andavadoaka	AN	8
		Belo sur Mer	B	44
		Nosy Bé	NB	6
		St Marie	SM	9
		Toamasina	TO	3
	Zanzibar Tanzania	Stone Town fish market	ZZ	10
	Tanzania	Tanzania*	T	30
	Mauritius	Mahebourg	MA	9
		Mauritius*	MA	30
	Seychelles	Mahe	MS	13
		Mahe plateau	MP	7
		Farquhar atoll	F	30
	Kenya	Kenya	MB	29
	Comoros Islands	Comoros Islands*	C	5
	Mozambique	Maputo fish market	MM	21
		Pemba	PM	1
		Quissico	Q	24
	South Africa	Banga Nek	BN	3BN
		Richard's Bay	RB	3
Lethrinus nebulosus sp.	South Africa	Richard's Bay	RB	55
	Mozambique	Zavora	Z	4
		Quissico	Q	16
		Maputo fish market	MM	3

In cephalopods 16s rDNA is considered to be more variable than COI (Lindgren, et al. 2005) and is frequently used in phylogenetic and phylogeographic studies (Lindgren, et al. 2005; Takumiya, et al. 2005; Teske, et al. 2007). Universal primers 16Sar-L and 16Sbr-H (Palumbi 1996) have previously been used successfully on both *O. vulgaris* and (inadvertently) *O. cyanea* (Teske, et al. 2007) and were used in this study to amplify a 508bp segment of 16s rDNA.

The octopus non-coding region was also employed as a phylogeographic marker. MtDNA in coleoid cephalopods has undergone significant rearrangement, both between cephalopod taxa (Tomita, et al. 2002) and in relation to other metazoans (Takumiya, et al. 2005). The first cephalopod mtDNA genome sequenced was *Loligo bleekeri* (Tomita et al 2002). The mitochondrial genome of *L. bleekeri* had undergone significant rearrangement and contained several long non-coding regions (NCRs), areas assumed to be involved in transcription and replication (Tomita, et al. 2002) and thought to be similar to the control region in metazoans. In Octopodidae, a single mtDNA non-coding region is present in *O. vulgaris* and *O. ocellatus* between the

MCYWQGE tRNA group and COIII (Akasaki, et al. 2006; Takumiya, et al. 2005; Tomita, et al. 2002). Due to the non-coding nature of this section of mtDNA, it was hypothesised that the NCR would have a higher mutation rate than coding regions of mtDNA such as COI or 16s rDNA, thus providing increased sequence information which would be extremely useful in the investigation of population structure and phylogeography of *O. cyanea*.

To test this hypothesis, primers were designed in the flanking regions of COIII and tRNA glu using the *O. vulgaris* complete mtDNA genome (NC_006353.1), COIII sequence of *O. cyanea* (AJ628220.1) and other closely related octopus species; *O. aculeatus* (AJ628213.1), *Ameloctopus litoralis* (AJ628207.1) and *Cistopus indicus* (AJ628208.1).

Alignment of COIII revealed a highly conserved region across the related species in which the reverse primer OcNCRR CCTTGAAATGTGCTTTCTCG was sited. The forward primer OcNCRF TCGTGCAAATTACACCACAAA was sited in tRNA glutamine, designed solely from *O. vulgaris sequence* (AC53793887) as no other closely related species' tRNA sequences were available (Table 2.2).

Five samples were amplified by PCR using these primers and sequenced producing five 825bp sequences. An internal sequencing primer OcNCRi TTGTTCAATAATAGTGCTGTTGAAAG was then designed to confirm the accuracy of the sequences obtained. The resulting sequences were then used to design primers NCR F2 and NCR R2 to reliably amplify a 500bp portion of NCR region.

All PCRs were carried out in a PTC-200 thermocycler (MJ Research) using the following conditions: 180 s at 95°C, 45 cycles of 30 s at 95°C, 45 s at the specified annealing temperature (Table 2.2) followed by 30 s at 72°C with a final extension step of 120 s at 72°C. Each reaction mix contained 4µL template DNA, 2 mM $MgCl_2$, 0.5 µM forward primer and 0.5 µM of reverse primer, 0.2 mM dNTP mix (20 µM each dATP, dCTP, dGTP, dTTP), 1x reaction buffer [75 mM Tris-HCl, 20 mM $(NH_4)_2SO_4$] and *Taq* polymerase (BIOTAQ, 5 U/µl) in a final reaction volume of 20 µL.

To check that unique haplotypes were not a product of DNA degradation, 5 samples with unique haplotypes were independently re-amplified using PCR and then sequenced. The two sequences were then checked to ensure that both sequences were identical.

Table 2.2. Mitochondrial DNA primers and annealing temperatures for amplification of COI, 16s rDNA and NCR in *O. cyanea*.

Amplified region	**Primer name (primer source)**	**Primer sequence 5'-3'**	**Annealing temperature (°C)**
COI	OC COI F OC COI R (this thesis)	CATTTTCGGAATTTGATCAGG GATTGGGTCTCCTCCACCTC	55
16S rDNA	16Sar-L 16Sbr-H (Palumbi 1996)	CGCCTGTTTATCAAAAACAT CCGGTCTGAACTCAGATCACGT	55
NCR	OC NCR F2 OC NCR R2 (this thesis)	TCCTGTTAATGGTCAGGGTCTAA TTCAACAGCACTATTATTGAACA	54

<u>*Lethrinus* fish species</u>

To investigate population structure and phylogeography of *Lethrinus harak*, *L. mahsena* and *L. nebulosus*, two regions of mitochondrial DNA, cytochrome oxidase I (COI) and the hypervariable region I (HVR1) of the control region (CR) were sequenced. Both regions are commonly used in studies of phylogeography and population structure of fish populations (Janko, et al. 2007; Kon, et al. 2007; Ward, et al. 2008; Ward, et al. 1995; Ward and Holmes 2007; Ward, et al. 2005; Zemlak, et al. 2009).

PCR of COI was carried out using universal fish primers FishF1, FishR1 (Ward, et al. 2005), (Table 2.3). However, PCRs using universal fish primers were unsuccessful in all *L. mahsena* samples and *L. harak* samples from Kenya, Mauritius and Zanzibar. Species-specific primers (Table 2.3) were therefore designed in Primer 3.0 (Untergasser, et al. 2012). To amplify the *L. mahsena* samples, primer pairs Lm COIF and Lm COIR were used initially. In samples which failed to amplify using this primer pair, Lh COIF and Lm COIR were subsequently used which successfully amplified all remaining samples. Samples of *L. harak* from Kenya, Mauritius and Zanzibar were amplified using primer pair Lh COIF, Lh COIR.

No published primers for amplification of the control region reliably produced PCR products in *Lethrinus* species. Primers were therefore designed to amplify the hypervariable region 1 (HVR1) in *Lethrinus* species in this study. Primers were designed in both flanking regions of the CR using the complete mtDNA sequence of *L. obsoletus* from GenBank (AP009165.1). Six samples (2 samples per species) were

sequenced for both forward and reverse sequences, allowing the entire CR in *L. mahsena*, *L. harak* and *L. nebulosus* to be sequenced. Sequences were aligned and examined for conserved regions across the 3 species. Primers were then designed in conserved regions either side of HVR1 (Table 2.3) using Primer 3 (Untergasser, et al. 2012).

PCRs were carried out in a PTC-200 thermocycler (MJ Research) using the following conditions: 180 s at 95 °C, 45 cycles of 30 s at 95 °C, 45 s at the specified annealing temperature (Table 2.3) followed by 60 s at 72 °C with a final extension step of 600 s at 72 °C. Each reaction mix contained 4 µL template DNA, 2 mM $MgCl_2$, 0.5 µM forward primer and 0.5 µM of reverse primer, 0.2 mM dNTP mix (20 µM each dATP, dCTP, dGTP, dTTP), 1x reaction buffer [75 mM Tris-HCl, 20 mM $(NH_4)_2SO_4$] and *Taq* polymerase (BIOTAQ, 5 U/µl) in a final reaction volume of 20 µL.

2.2.3 **Preparation of PCR amplicons for sequencing**

Amplified products were resolved on a 2% agarose gel stained with ethidium bromide. PCR products were then purified through enzymatic removal of excess nucleotides and primers using exonuclease I and thermosensitive alkaline phosphatase (EXOSAP) in the thermocycler under the following conditions; 30 minutes at 37°C, 15 minutes at 80°C. DNA sequencing was then performed on an Applied Biosystems 3500 sequencing platform

Table 2.3. Mitochondrial DNA primers and annealing temperatures for amplification of COI and HVR1 in *L. mahsena*, *L. nebulosus* and *L. harak*.

Species	**Amplified Region**	**Primer name (primer source)**	**Primer sequence**	**Annealing Temp** (°C)
L. harak	COI	FishF1 FishR1 (Ward, et al. 2005)	TCAACCAACCACAAAGACATTGGCAC TAGACTTCTGGGTGGCCAAAGAATCA	55
L. harak	COI	Lh COIF Lh COIR (this study)	CGAACTTAGTCAGCCCGGA TGCTGATAGAGGATTGGGTC	55
L. nebulosus	COI	Fish F1 Fish R1 Fish R2 (Ward, et al. 2005)	TCAACCAACCACAAAGACATTGGCAC TAGACTTCTGGGTGGCCAAAGAATCA ACTTCAGGGTGACCGAAGAATCAGAA	55
L. mahsena	COI	Lm COIF Lm COI R (this study)	TGGTAGGAACAGCCCTAAGC AGAATTGGGTCCCCTCCTC	55
L. harak, *L.nebulosus*	CR	LCR1F LCR1R (this study)	CGGTCTTGTAAACCGGATGT GTCATGGCCCTGAAATAGGA	57

2.2.4 **Isolation of microsatellite loci for *Octopus cyanea***

The most powerful studies into population genetics and phylogeography combine the use of nuclear and mitochondrial markers (Sunnucks 2000). Microsatellite markers were therefore chosen as suitable nuclear markers to study the population connectivity and structure of *O. cyanea* in conjunction with the mtDNA regions being sequenced.

Eleven microsatellite loci were developed for *O. cyanea* by Dr Niall McKeown of our research group (Table 2.4). Microsatellites were isolated from an enriched partial genomic library created by methods outlined by (Glenn and Schable 2005) and (McKeown and Shaw 2008). Genomic DNA extracted from *O. cyanea* tissue was

digested with restriction enzymes (New England Biolabs) producing fragments suitable for the ligation of oligonucleotide linkers. Linker-ligated DNA was amplified by PCR using the linkers as PCR primers. Enrichment was then performed by selective hybridisation of biotin-labelled repeat motif oligonucleotide probes to the PCR products. Hybridised complexes were captured using streptavidin-coated magnetic beads (DYNAL). Microsatellite enriched eluates were thereafter PCR amplified and cloned into bacterial vectors using the TOPO-TA cloning kit (Invitrogen). Recombinant colonies were identified by inactivation of the beta-galactosidase gene. Recombinant colonies were individually transferred into 50 ml of 10 mM Tris-HCl (pH 8.5) and incubated at 95^{o}C for 10 min to promote plasmid DNA release. One microlitre of each plasmid extract was submitted to PCR involving the standard M13 forward and M13 reverse primers. The amplification reaction contained 1X buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 U of *Taq* DNA polymerase (Bioline, UK), 10 pmol of each primer and was performed through 30 cycles of 30 s at 95^{o}C, 30 s at 52^{o}C and 30 s at 72^{o}C. PCR products were cleaned using ExoSap and sequenced using the internal T7 vector primer. Sequences were analysed using the TANDEM REPEATS FINDER (Benson 1999) and, where appropriate, primer pairs designed using Primer 3 (Untergasser, et al. 2012).

Table 2.4. Primer sequences and characteristics of 11 microsatellite loci developed for *O. cyanea*.

Locus Name	Primer sequences (5'-3')	Microsatellite Sequence	Product Size (bp)	Annealing Temp. (°C)
ROC32	F:TTTCATCCACCCATTGATATCTC, R:GCCATATCAGACACCATGATG	$(GTAT)_{13}$	180	55
ROC6	F:TAGAATGTGCGTCGCAAAAG, R:GCGTAAGCATGCATGTATGC	$(CATA)_{16}$	180	55
OC22	F:TTCAACACTCAAATCGTTAGGG, R:TGAACAGAGCTCTTAATTACTATGA	$(TATA)_{17}$	200	55
ROC28	F:TCCCGCATCAACCTTTAATG, R:TTAGCCGGACCGTTAGTTTG	$(GATA)_{19}$	200	56
OC31	F:CGTCAGGTCCAGGACTTATTG, R:AAGACACGTTTTGGATATACTTC	$(GTAT)_{21}$	200	58
OC19	F:TGCCCCTTTAAAATAGAGTTGC, R:TAAGTCCTGGGATCGAATGG	$(CATA)_{13}$	220	58
ROC17	F:TAAGTCCTGGGATCGAATGG, R:TGCCCCTTTAAAATAGAGTTGC	$(GTAT)_{13}$	220	58
OV4	F:CAGCCAGCACCGTAATACATC, R:ACCAGGCCTTGTGCCTTTAG	$(AAT)_{22}$	220	59
OC15	F:TTCAACACTCAAATCGTTAGGG, R:ATGCTTTTTGCTTTCACAGG	CATACTCATACAGATA$(CATA)_{15}$CATGAGTGTA$(CACA)_2$TA$(CA)_3$CT$(CACA)_{2(}$TA$)_3$	274	55
ROC1	F:AAATGCACACGCATAAAAACA, R:GCGTCAGGTCCAGGACTTAT	$(CATA)_{20}$	207	55
OC18	F:ACATCGACCCCAGGACTTAG, R:TGCAATAACCTTCGGTTGTG	$(CACA)_{12}$TAC A$(CACA)_{15}$	348	55

2.2.5 **Development of a more economical method of microsatellite screening**

Traditionally, microsatellite loci are developed and the forward primers are then labelled with a fluorescent dye to allow the detection of the product during screening such as Applied Biosystems' (AB) 5-fluorescent dye system (6-FAM, NED, VIC and PET). However, ordering fluorescently labelled primers is expensive, typically £150 per labelled primer. Therefore during this study a more economical method was trialled to label the microsatellite loci based on Schuelke (2000). Schuelke carried out a nested

PCR performed with three primers: a sequence-specific forward primer with M13(-21) tail at its 5' end, a sequence-specific reverse primer and the universal fluorescently-labelled M13(-21) primer. Using optimised PCR conditions, Schuelke was able to both amplify and label the product during this single PCR reaction. This method was trialled using the *O. cyanea* microsatellite loci developed for this study following conditions suggested by Schuelke (2000). However, despite optimisation procedures, this nested approach did not produce consistent results using the *O. cyanea* microsatellites. A slightly modified method was therefore developed involving two separate PCR reactions; the first PCR to amplify the product followed by a second PCR to label the product. The first PCR was performed with two primers; a sequence-specific forward primer with M-13 tail at its 5' end and a sequence-specific reverse primer. The product of this PCR was then used in a further PCR which contained the universal fluorescent labelled M13 primer. Six microsatellite loci OC31, OC22, OV4, ROC32, OC19 and ROC6, were screened using this method.

Traditional fluorescently labelled sequence-specific forward primers using Applied Biosystems' (AB) 5-fluorescent dye system (6-FAM, NED, VIC and PET) were used for OC18, ROC1, ROC17, OC15 and ROC28 as despite various optimisation procedures, the M13 labelling technique was not consistently successful for these microsatellites.

2.2.6 **PCR amplification and screening of microsatellite data**

PCRs were carried out in a PTC-200 thermocycler (MJ Research) using the following conditions: 180 s at 95°C, 45 cycles of 30 s at 95 °C, 45 s at the specified temperature for the microsatellite (Table 2.4) followed by 30 s at 72 °C with a final extension step of 120 s at 72 °C. Each reaction mix contained 4 µL template DNA, 2 mM $MgCl_2$, 0.5 µM forward primer with M13 tail at 5' end and 0.5 µM of reverse primer, 0.08 mM dNTP mix (20 µM each dATP, dCTP, dGTP, dTTP), 1x reaction buffer [75 mM Tris-HCl, 20 mM $(NH_4)_2SO_4$] and 0.04 µL *Taq* polymerase (BIOTAQ, 5 U/µl) in a final reaction volume of 10 µL. Products were resolved on a 2% agarose gel to check that a product had been produced which was a similar size to the amplification product.

M13 Labelling

PCR products were then labelled with universal M13 primers PET, NED, VIC, 6FAM (6-carboxy-fluorescine) (Table 2) using a PCT-200 thermocycler under the following conditions: 180 s at 95 °C, then 35 cycles of 30 s at 95 °C, 60 s at 50 °C followed by 30 s at 72 °C. Reaction mixes contained 7 µL template DNA, 2mM $MgCl_2$,1.2ul of M13 forward primer, 5 mM dNTPs, 1x reaction buffer [75 mM Tris-HCl, 20 mM $(NH_4)_2SO_4$] and 0.04 µL *Taq* polymerase (BIOTAQ) in a final reaction volume of 12 µL.

Genotyping

PCR products were genotyped on an ABI 3500 Genetic Analyzer using Applied Biosystems LIZ 500(-250) size marker. Alleles were scored manually using GENEMAPPER version 4.1 (Applied Biosystems).

2.3 Statistical analyses

2.3.1 Mitochondrial analyses

Population diversity

Genetic variation was assessed by the calculation of number of haploptypes (H), haplotype diversity (h) and nucleotide diversity (π) in Arlequin 3.5.1.2 (Excoffier and Lischer 2010).

Population Structure

Statistical power of the mtDNA dataset to detect genetic structuring was assessed using POWSIM (Ryman and Palm 2006), adjusted for mtDNA after Larsson, et al. (2009) using a population size of 10,000. The statistical power of the dataset to detect a range of F_{ST} values from 0.01 to 0.05 was tested using these settings by varying the number of generations (*t*). Type 1 error was also estimated from the dataset using F_{ST} = 0.

Population structure was assessed using F_{ST}, estimated by θ (Weir and Cockerham 1984) both globally and pairwise between samples. Exact tests of population structure were also calculated in Arlequin 3.5.1.2 (Excoffier and Lischer 2010). For *L. nebulosus* and *L. harak*, Φ_{ST} was additionally calculated in Arlequin 3.5.1.2 (Wegmann, et al. 2010) to account for divergence between haplotypes in the calculation of population structuring. Where multiple tests were conducted, significance levels were adjusted according to a Bonferroni correction (Rice 1989).

Evolutionary History

Departures from neutrality in mitochondrial DNA indicate non-neutral evolution of mtDNA sequences. Departures from neutrality can be caused by selection or changes in past population size. To test for departures from neutrality Fu's Fs (Fu 1997), Tajima's D (Tajima 1989) and Ewens-Watterson test (F) (Slatkin 1994a) were calculated using Arlequin 3.5.1.2 (Excoffier and Lischer 2010). Statistical significance was assessed after 1000 permutations.

To test the theory of a past population expansion, mismatch analysis was applied to the dataset. Mismatch distribution is based on the theory of the coalescence. Mitochondrial DNA is the signature of an ancestral population which can be traced back through mtDNA lineages to the most recent common ancestor. Changes in past population size leave specific signatures in mtDNA of organisms, in particular the distribution of pairwise differences between pairs of sequences. In populations which have been stable for considerable periods of time, this distribution is multimodal. However in populations which have undergone a significant population expansion, a unimodal peak in the number of nucleotide differences is often observed (Rogers and Harpending 1992). Mismatch analysis assumes a single stepwise expansion model, an instantaneous population expansion with populations remaining stable since expansion. Mismatch distribution was calculated for both a pure demographic expansion and spatial expansion in Arlequin v. 3.5.1.2. (Excoffier and Lischer 2010) assuming a finite sites model and heterogeneity of mutation rates (Schneider and Excoffier 1999). Confidence intervals for mismatch analysis were also calculated using a parametric bootstrap approach based on 1000 replicates. The sum of squares deviation (SSD), test statistic was used to compare how well the expected data fit the observed data. The p value of the SSD statistic was used to test whether the null hypothesis of an instantaneous population expansion was accepted or rejected. If the p value was > 0.05, the hypothesis of a population expansion was accepted, if the p value was < 0.05, the hypothesis of a population expansion was rejected.

Mismatch analysis cannot be relied upon to calculate accurately the magnitude of the population expansion and can only estimate the expansion time (Tau, τ) with any confidence (Schneider and Excoffier 1999). The mismatch equation is $\tau = 2\mu t$, where μ is the mutation rate, and t is the generation time. Using this equation, an estimate of time since expansion can be calculated. However, there are high rates of calculation errors in mismatch distribution analysis (Schenekar and Weiss 2011), resulting in numerous false inferences of biological importance. In the articles surveyed by Schenekar and Weiss (2011), 49% of mismatch calculations were found to be incorrect. Calculation of time since expansion using mismatch analyses were therefore

checked using the online spreadsheet http://www.uni-graz.at/zoowww/mismatchcalc/index.php (Schenekar and Weiss 2011).

Bayesian skyline plots, implemented in BEAST v.1.7.4 (Drummond, et al. 2012) were performed to explore demographic changes through time within each population. Skyline plots employ coalescent theory to reconstruct fluctuation through time in effective population size (N_e) using DNA sequence variation, when the nucleotide substitution model and rate are known (Ho and Shapiro 2011). The appropriate model was estimated for each population in jModelTest2 (Darriba, et al. 2012). Two independent runs were performed, using the piece-wise constant method for population expansion, for 40 million MCMC generations, sampling every 2,000 generations. Convergence and visualization of median and 95% highest posterior probability density intervals (HPD) were assessed in Tracer v.1.5 (Rambaut and Drummond 2009), using the effective sample size (ESS>500) as an indicator. To incorporate variation among related lineages, a local molecular clock was enforced, using an appropriate sequence divergence rate.

Phylogenetic Tree Reconstruction

Haplotype networks were constructed using maximum likelihood methods in PHYLIP DNAml (Felsenstein 1989) and subsequently constructed in Hapview (Ewing 2010).

A phylogenetic tree depicts the lines of evolutionary descent of different species, organisms, or genes from a common ancestor (Hall 2011). Phylogenies are useful for organizing knowledge of biological diversity, for structuring classifications and for providing insight into events that occurred during evolution.

The most suitable model of nucleotide substitution to account for the observed sequence polymorphism in samples was calculated using jModelTest version 2.1.3 (Darriba, et al. 2012; Guindon and Gascuel 2003). Models of nucleotide substitution are crucial to molecular phylogenetics (Darriba, et al. 2012). Model selection may affect estimates of phylogeny, substitution rates, bootstrap values and posterior probabilities (Posada and Buckley 2004). In general, phylogenetic methods may be less accurate (recover an incorrect tree more often) or may be inconsistent (converge to an incorrect tree with increased amounts of data) when the wrong model of evolution is assumed (Posada and Buckley 2004; Posada and Crandall 2001). No model of evolution will completely reflect the true model but it is important to select a model which is a suitable approximation to reality on which to base phylogenetic analysis.

JModelTest v 2.1.3. (Darriba, et al. 2012) was run using the following likelihood settings; 11 substitution schemes, equal or unequal rate variation, proportion of invariable sites and rate variation among sites with a number of rate categories using

an optimised maximum likelihood tree. The tree search selected was “Best”, which uses both subtree pruning and regrafting (SPR) and nearest neighbour interchange (NNI) techniques. The most suitable substitution model was then chosen based on the Bayesian Information Criterion (BIC) (Posada 2008).

Reconstruction of phylogenetic relationships between individuals was carried out using Bayesian and maximum likelihood methods using the most appropriate substitution model indicated by JModelTest v 2.1.3 (Darriba, et al. 2012). Bayesian Markov Chain Monte Carlo (MCMC) analyses were carried out using the appropriate model and default settings in MrBayes v 3.2 (Ronquist, et al. 2012). Two independent sets of four chains were run for 5 - 9 million MCMC generations and the parameters recorded every 500 generations in MrBayes. All branches with posterior probabilities above 0.5 were retained. The convergence and stability of runs were estimated by examining the log likelihood values and the standard deviation of the split frequencies.

Maximum likelihood analyses were conducted using a neighbour-joining starting tree and implemented using the online version of PhyML 3.0 (Guindon, et al. 2010). The most suitable substitution model was selected and tree searching was carried out using nearest neighbour interchange (NNI) and subtree pruning and regrafting (SPR) methods. Confidence in tree topology was estimated using the approximate likelihood-ratio test (aLRT) method (Anisimova and Gascuel 2006).

Trees were constructed in Figtree v.1.4 (Rambaut 2012).

Calculation of sequence divergence

Sequence divergence was calculated using the p-distance (the proportion (p) of nucleotide sites at which two sequences being compared are different). Minimum rather than mean p-distances were reported as where species identification is being considered, mean distances artificially inflate the barcoding gap and may lead to misclassification (Meier, et al. 2008).

Time since divergence

Time since divergence was estimated by calculating the percentage divergence between clades or species and then applying an appropriate species mtDNA divergence rate. For *Lethrinus* fish species, a divergence rate of 3.6% per MYR for CR mtDNA was utilised, calculated from the closure of the Isthmus of Panama (Donaldson and Wilson 1999). For Lethrinus COI mtDNA, a divergence rate of 1.2% per MYR was utilised, which was also calculated from the closure of the Isthmus of Panama (Bermingham, et al. 1997)

Additionally, a coalescent approach was employed to estimate time since most recent common ancestor (tmrca) from the mtDNA dataset in *L. nebulosus*. A total of 45

individuals of *L. nebulosus* (five from each sampling location) and 1 *L. obsoletus* individual, a closely related sister species to *L. nebulosus*, were used to estimate time since most recent common ancestor. Two independent runs (10 million MCMC generations, sampling every 1,000 generations), using the HKY+G nucleotide substitution model and establishing the tree prior as Yule Process, were conducted in BEAST v.1.7.4 (Drummond, et al. 2012). Convergence of run parameters (ESS>500), tmrca and 95% HPD intervals were reviewed in Tracer v.1.5 (Rambaut and Drummond 2009) to ensure convergence (Drummond, et al. 2012). Calibration of the molecular clock was conducted by fixing the lineage mutation rate at 1.8×10^{-5} per thousand years for the CR (calculated from 3.6% divergence rate per MYR) (Donaldson and Wilson 1999). Estimates of tmrca were relative only as no fossil or biogeographical data were available to use as node calibrators.

2.3.2 **Microsatellite analyses**

Data were screened for amplification errors such as null alleles, large allele drop-out and stuttering using MICRO-CHECKER (Van Oosterhout, et al. 2004) and FreeNA (Chapuis and Estoup 2007).

All mtDNA sequences were inspected and aligned in BIOEDIT (Hall 1999) and multiple alignments were performed using CLUSTAL X (Thompson, et al. 1997)

Genotype frequencies were tested for departure from Hardy Weinberg Equilibrium expectations and linkage disequilibrium (non-random association of alleles between loci) using Genepop v4.1 (Raymond and Rousset 1995b). Statistical significance was assessed after 1000 permutations and adjusted using Bonferroni correction (Rice 1989) for multiple tests.

Allele number (N_a), allelic richness (AR), observed and expected heterozygosity (H_o and H_e) and the inbreeding coefficient (F_{IS}) were calculated using FSTAT v 1.2 (Goudet 1995). Statistical significance of F_{IS} was calculated in FSTAT v 1.2 using 1000 permutations (Goudet 1995). Significant departures from expected heterozygosity at a particular locus across all populations may indicate the presence of null alleles. Wrights' inbreeding coefficient (FIS) estimates the partitioning of genetic diversity within subpopulations and ranges from +1, an excess of homozygotes, to -1, an excess of heterozygotes.

Microsatellites are generally non-coding and so are assumed to be selectively neutral markers (Nielsen, et al. 2006). Explanations of patterns of microsatellite genetic differentiation among populations therefore commonly invoke genetic drift and migration as the drivers of population structuring. However, microsatellites are often utilised in linkage mapping of important quantitative trait loci under selection in crops

and livestock (Vasemägi, et al. 2005), making an assumption of neutrality not always accurate for microsatellites. Genetic hitch-hiking, when the frequency of a microsatellite locus changes because of linkage with a positively selected allele at another locus, has been observed in a number of studies of population structuring (Beaumont 2005; Nielsen, et al. 2006; Rhode, et al. 2013). Directional selection among populations at any locus linked to a hitch-hiking neutral genetic marker would inflate measures of population differentiation, whilst loci under overall stabilizing selection through genetic hitch-hiking would decrease measures of population differentiation.

Tests for detection of microsatellites under selection were performed using the F_{ST}-outlier detection methods implemented in LOSITAN (Antao, et al. 2008). Deviation of allele frequencies between samples from those expected under neutrality were tested using the F_{ST}-outlier method of Beaumont and Nichols (1996). Runs were performed initially using the entire microsatellite dataset. LOSITAN was then run using pairs of loci and finally for pairs of locations using all loci. Two independent runs were performed (100000 MCMC generations) using the "neutral" mean F_{ST}-estimated under the infinite alleles mutation model (IAM) as recommended by Antao, et al. (2008).

The statistical power of the genetic data to detect genetic homogeneity and estimates of the statistical α (type 1) error were assessed using POWSIM (Ryman and Palm 2006). The concept of power and resolving power in genetic differentiation studies is often discussed, with justifications that the dataset has enough power to detect structuring due to large sample sizes, particular markers used, number of loci etc. However, these descriptions are qualitative and do not provide any quantitative estimate of power. POWSIM (Ryman and Palm 2006) is a simulation-based program that estimates power (and α error) for chi-square and Fisher's exact tests when evaluating the hypothesis of genetic homogeneity and degree of differentiation (quantified as F_{ST}).

POWSIM (Ryman and Palm 2006) was run to estimate the power of the microsatellite dataset using chi-square and Fisher's exact tests under the following settings; Markov chain parameters - number of dememorisations, batches and iterations per batch were set to 1000, 100 and 1000 respectively, 7 polymorphic loci, 13 subpopulations and an effective population size 10,000. The statistical power of the dataset to detect a range of F_{ST} values from 0.01 to 0.05 was tested using these settings by varying the number of generations (*t*). Type 1 error was also estimated from the dataset using $F_{ST} = 0$.

Population structure

Genetic differentiation among samples was analysed using exact tests of differences in allele frequencies (using GENEPOP) and F_{ST} (Wright 1965). Levels of F_{ST}, estimated by θ (Weir and Cockerham 1984) both globally and pairwise between samples, were

calculated and tested for significant departure from zero (no differentiation) within FSTAT v2.9.3.2 (Goudet 1995).

Molecular ecologists have used a number of different population methods to test population structure including F_{ST} (Wright 1949), G_{ST} (Nei 1973) and R_{ST} (Slatkin 1995). Despite the large number of methods available, F_{ST} is still the most common statistic used to test for population structuring and often performs better than its R_{ST} counterparts (Balloux and Lugon-Moulin 2002).

F_{ST} was designed by (Wright 1949) to test population structuring using low diversity markers such as allozymes. It is based upon the infinite alleles model and originally ranged from 0 – 1, with values close to 0 indicating little to no differentiation and values close to 1 indicating nearly complete differentiation (Wright 1949). It was designed to test low diversity systems (Wright 1949). However, as molecular techniques have advanced, highly polymorphic markers have been developed which, even in the presence of high differentiation, often result in drastically reduced values of F_{ST} or even nonsensical values (Charlesworth 1998; Gerlach, et al. 2010; Jost 2008). A good example of this is highlighted by Jost (2008). A study of two races of the common shrew where analysis of the Y chromosome microsatellites resulted in an F_{ST} of 0.19, suggesting little differentiation, despite the authors showing that no alleles were shared between the races and gene flow was essentially zero (Balloux and Lugon-Moulin 2002). The reason for this is the way that these estimates of differentiation are calculated. Additive partitioning of heterozygosity into mean within subpopulation heterozygosity and between subpopulation diversity (used to calculate F_{ST}) means these two are not independent, therefore when subpopulation heterozygosity is large, subpopulation diversity and hence F_{ST} has to be small regardless of the differentiation between subpopulations (Jost 2008).

Researchers therefore need to be extremely cautious when interpreting F_{ST} and its relatives. In such situations it is therefore useful to have an independent method to test for population structuring, such as non-parametric tests based on a permutation procedure (Balloux and Lugon-Moulin 2002). Due to the high polymorphism of microsatellites, exact tests are very powerful at detecting population substructuring (Balloux and Lugon-Moulin 2002). Therefore in addition to calculating F_{ST}, this study also includes an exact test of population structure. An exact test is based on a permutation procedure in which alleles or genotypes are shuffled among subpopulations repeatedly (Raymond and Rousset 1995a). From these datasets differentiation can be estimated and the proportion of values larger than or equal to the one estimated from the real data will provide the un-biased probability (p) value of the test (Raymond and Rousset 1995a). Allelic exact tests take into account whether two

genotypes share one allele or not, which is not considered for genotype exact tests and are more powerful than genotype statistics when sampling is balanced (Goudet, et al. 1996). When sampling is unbalanced (when sample size varies between subpopulations) allelic exact tests have a considerably higher power than F_{ST} tests (Goudet, et al. 1996). Allelic exact tests with significance determined by a Markov chain method were calculated using GENEPOP v4.1 (Raymond and Rousset 1995b).

Controlling the Rate of Type 1 Error

The rate of false positives (type 1 error) where the null hypothesis is falsely rejected, is a common challenge to studies on population genetics, which frequently involve multiple testing of hypotheses or data. Where multiple hypothesis tests are carried out, type 1 errors quickly increase at the rate of $1 – (1-\alpha)^k$ where α is the P value and *k* is the number of tests simultaneously carried out (Rice 1989). Rice (1989) therefore argued that correction of false type 1 error rate be employed wherever; 1. a group of two or more tests are scanned and the p values of component tests are used to determine where significant differences occur; and 2. where two or more tests address a common null hypothesis and the rejection of the null hypothesis is only possible when some of the tests are found to be individually significant (Cabin and Mitchell 2000; Rice 1989). To correct this, Rice (1989) suggested the use of Bonferroni correction. Bonferroni modifies the significance criteria to α/k to reduce the frequency of type 1 errors. Since Rice's (1989) popular publication, Bonferroni correction has used frequently in studies where multiple comparison tests have been used (Cabin and Mitchell 2000). However, debate on the use of Bonferroni correction has been increasing in recent years as the correction is very conservative and typically results in a large increase in type 2 error - the proportion of the false null hypotheses that are correctly rejected (Cabin and Mitchell 2000; Nakagawa 2004; Narum 2006; Ryman and Jorde 2001). A number of less conservative alternatives such as false discovery rate (FDR) have been developed which exert a less stringent control over type 1 errors compared to familywise error rates such as Bonferroni correction. False discovery aims to reduce the expected proportion of type 1 errors as opposed to Bonferroni correction which seeks to eliminate all type 1 errors. To investigate the likelihood of type 1 error, analyses were run in POWSIM (Ryman and Palm 2006) using the microsatellite dataset. In addition, the significance of results was considered using no correction and Bonferroni correction.

Population structuring was assessed using STRUCTURE (Pritchard, et al. 2000), a model based Bayesian clustering analysis program. STRUCTURE (Pritchard, et al. 2000) uses a model-based clustering method to infer population structure using genotype data and may be used to demonstrate evidence of population structuring, to

identify distinct genetic populations, to assign individuals to populations, and to identify migrants and admixed individuals.

STRUCTURE assumes a model where there are a number (K) of populations (K may be unknown) each of which are defined by a specific set of allele frequencies at each locus. Individuals in the samples may then be assigned to a population, or their genomic proportions to multiple populations (Hubisz, et al. 2009). STUCTURE may be run using a number of models; the simplest model is the "no-admixture model", which assumes that each individual belongs to a single cluster or population. The less simple "admixture" model assumes that an individual may cluster with more than one population (Hubisz, et al. 2009). STRUCTURE may also be run with or without *a priori* sample partitioning. Using *a priori* sample partitioning allows STRUCTURE to place significantly more weight on clustering outcomes that are correlated with the sampling locations (Hubisz, et al. 2009), when the data suggest that this would be useful. Alternatively STRUCTURE may be run with no *a priori* sample partitioning, if the data is considered not to be partitioned into clusters according to locality. Three independent runs were performed for each value of K between 1 and 12 using 1,000,000 Iterations and a burn in of at least 100,000 iterations. Multiple runs were performed varying the parameter sets to thoroughly explore the dataset. The parameters which were varied were; admixture, no admixture, locprior (sample partitioning)/no locprior (no sample partitioning) and population information/no population information.

Results were interpreted using the Evanno Method (Evanno, et al. 2005) in STRUCTURE HARVESTER (Earl and vonHoldt 2012). STRUCTURE HARVESTER is a Python program which can quickly parse and summarise output data from STRUCTURE (Earl and vonHoldt 2012). The Evanno method uses the model value of ΔK, an ad hoc quantity based on the second order rate of change of the likelihood function with respect for K to determine the most likely number of populations present in the sample data (Evanno, et al. 2005). Measures of LnP(K) were also examined to determine the most likely value of K (Pritchard 2010) as the Evanno method is not able to determine if K is one (Evanno, et al. 2005).

Demographic history

Changes in effective population size due to environmental events can have a profound effect on the genetic diversity of a species. The identification of population bottlenecks is extremely important in conservation because populations which experience significant reductions in population size are at risk from a variety of demographic and genetic processes that can lead to extinction (Frankham 2005). Traditionally, methods to detect recent bottlenecks in populations have been based upon the assumption that during a bottleneck allele diversity will be reduced resulting in the observed

heterozygosity being higher than the heterozygosity expected from the observed allele number were the locus at mutation-drift equilibrium (Frankham 2005).

Two methods have been commonly used to detect bottlenecks in populations, BOTTLENECK and the M ratio (Garza and Williamson 2001; Piry, et al. 1999). BOTTLENECK (Piry, et al. 1999) compares for each population sample and for each locus the distribution of the heterozygosity expected from the observed number of alleles (*k*), given the sample size (*n*) under the assumption of mutation-drift equilibrium. This distribution is obtained through simulating the coalescent process of *n* genes under three possible mutation models; the infinite alleles model (IAM), the two phase mutation model (TPM) and the stepwise mutation model (SMM). This enables the computation of the average heterozygosity which is compared with the observed heterozygosity to establish whether there is an heterozygosity excess or deficit at this locus. In addition, the standard deviation (SD) of the mutation-drift equilibrium distribution of the heterozygosity is used to compute the standardized difference for each locus.

An alternative method which is often used in conjunction with BOTTLENECK to test for a historical population bottleneck is the M ratio (Garza and Williamson 2001). Allele frequency distributions contain information including the total number of alleles (k) and the overall range in allele size (r). When a population decreases in size, the effects of genetic drift are increased, causing the loss of rare alleles. Any loss of alleles will contribute to a reduction in k but only the loss of the largest or smallest allele will result in a reduction of r. The frequency distribution of microsatellite alleles is rarely a bell shaped distribution and so the loss of the rarest alleles is generally not the loss of the smallest or largest alleles. Therefore k is expected to reduce more quickly than r during a bottleneck. The ratio M = k/r is therefore expected to be smaller in populations which have recently experienced a bottleneck than in equilibrium populations (Garza and Williamson 2001).

To test for a historical bottleneck, both the M ratio (Garza and Williamson 2001) and the variation between the expected and observed heterozygosity relative to the number of alleles was calculated using BOTTLENECK (Piry, et al. 1999). The hypothesis of recent bottleneck events was tested in BOTTLENECK (Piry, et al. 1999) under the three commonly accepted potential mutation models for microsatellite evolution. The TPM was run with an assumption of 70% of stepwise mutations. Significant heterozygote excess was tested for using the Wilcoxon signed rank test, after 10000 iterations.

To test for a recent bottleneck the M-ratio, the mean ratio of the number of alleles (k) to the range in allele size (r), was calculated using the software M_P_VAL (Garza and

Williamson 2001) as described in Chapter 2.3.2. Effective population size (N_e) was estimated to be 5000 pre-bottleneck in *O. cyanea*. Since the absolute rate of microsatellite mutation in our study species is unknown, θ was calculated using a range of mutation rates 1×10^{-4}, 3×10^{-4} and 5×10^{-4}, which are typical of mutation rates of microsatellites in other species. M_P_VAL was therefore run with θ of 2, 6 and 10. To determine the significance of our M ratio we compared our observed value to a distribution of values obtained from a simulated population given our sample size and mutation model. Critical values (Mc) set at the lower 5% tail of the distribution were determined using the program CRITICAL_M, below which it can be assumed that an observed ratio is from a population that has experienced a significant reduction in size (Garza and Williamson 2001)

A review of published literature (Peery, et al. 2012) has indicated that microsatellite based bottleneck tests often fail to detect bottlenecks in vertebrate populations which are known to have experienced declines. Simulations have revealed that bottleneck tests have limited statistical power to detect bottlenecks largely as a result of the limited sample sizes typically used in studies. In addition, commonly assumed mutation model parameters do not incorporate the variation in microsatellite evolution often observed in vertebrates and multi-step mutations may be underestimated. This may lead to the bottleneck test having a higher probability of detecting bottlenecks in stable populations than expected (Peery, et al. 2012). Studies have not yet addressed whether this pattern is also true of invertebrates such as octopus. However the recent development of Bayesian methods which estimate posterior distributions of demographic parameters and evaluate competing models of demographic history are rapidly evolving and are increasingly being utilised in studies to estimate demographic change in populations. For this reason, this thesis incorporates both the traditional tests and Bayesian methods which are described in further detail in the mitochondrial data analysis section.

Chapter 3. Molecular identification of three co-occurring and easily misidentified octopus species in the Southwestern Indian Ocean using PCR-RFLP techniques

This work has been published;

Taylor, A. L., McKeown, N. J., Shaw, P. W. (2012) Molecular identification of three co-occurring and easily misidentified octopus species in the Southwestern Indian Ocean using PCR-RFLP techniques. Molecular Genetics Resources 4: 885-887

3.1 Introduction

In phylogenetics, autapomorphic characters are those shared only with a terminal group in a phylogenetics tree. As such they are considered to be diagnostic of phylogenetic species (Baum 1992). When these characters are genetically detectable it provides a useful method for species identification.

3.1.1 Applications of molecular methods to fisheries

Identification of species using molecular methods has been used in a wide range of fields including regulation of consumer products such as meat (Hunt, et al. 1997) investigation and prosecution of illegal wildlife hunting/smuggling (Cronin, et al. 1991; Dawnay, et al. 2007) and criminal forensic work (Saigusa, et al. 2009). In the marine environment, molecular species identification techniques are increasingly being utilised for a range of purposes such as accurate species identification for fisheries management, consumer protection/quality control issues, fish fraud detection, compliance with conservation/endangered species treaties and identification of life-history stages.

Fisheries management

Accurate fish identification is a prerequisite for effective fisheries management (Ward 2000). Catch data are typically used to estimate biomass sizes and for species managed under quotas, to monitor the quota use throughout the year. If species are morphologically similar, or are cryptic species, misidentification of these species may be fairly common. This misidentification of fish greatly reduces the accuracy and therefore value of the recorded catch data and may therefore contribute to overfishing of a particular species (Ward 2000). Molecular techniques have therefore been developed to differentiate a number of cryptic species, including northern and southern

bluefin tuna (Ward, et al. 1995), skate species (Griffiths, et al. 2010) and hammerhead shark species (Abercrombie, et al. 2005).

Species substitution in the consumer market/seafood fraud

Molecular techniques are also highly valuable in quality regulation and enforcement in the consumer market (Rasmussen and Morrissey 2008) where rising international trade, increasing global seafood consumption and varying levels of supply and demand have caused an increase in species substitution and economic fraud in consumer markets. Many commercially valuable species have closely related, less valuable, congeners. Alternatively, where processes which remove morphological identification characters are carried out, less valuable alternatives may be used in place of the original species. Examples of this include red snapper being replaced by various other white fish species (Cawthorn, et al. 2012), cod and haddock replaced by cheaper white fish species (Miller and Mariani 2010), replacement of Greenland halibut for common sole (Céspedes, et al. 2000), grouper and wreck fish being replaced by a number of cheaper substitutes in restaurants (Asensio and Samaniego 2009) and even cases where toxic puffer fish were substituted for mullet roe, causing tetrodotoxin poisoning (Hsieh, et al. 2003). Mislabelling of fish is a global problem, with significant substitutions occurring in South Africa, Europe, the USA and the Far East (Cawthorn, et al. 2012; Garcia-Vazquez, et al. 2010; Hsieh, et al. 2003; Jacquet and Pauly 2008; Miller and Mariani 2010). The USA imports 80% of its seafood, a third of which may be mislabelled (Jacquet and Pauly 2008). Mislabelling of seafood products becomes even more problematic where canning or extensive processing is involved as there is greater opportunity for species substitution. A recent study into 69 fish products available in Italy found 22 were incorrectly identified (32%) and of these 26% were serious frauds from both economic and nutritional points of view (Filonzi, et al. 2010).

To combat mislabelling, regulatory bodies such as the European Union (EC No. 2065/2001) have established labelling laws for fish and aquaculture. These laws require products to be traceable, accurately identified to species, and the production method labelled (Moretti, et al. 2003). However, mislabelling of seafood globally continues to remain a significant concern, making reliable molecular methods of species identification crucial in trying to combat this black market and maintaining consumer confidence and safety.

Conservation of vulnerable marine species

The majority of commercially valuable marine species are subject to annual quotas and international agreements. Due to their high demand, these marine species often have

substantial retail prices, such as tuna ($60,000 per kg), shark fins ($700 per kg), caviar ($2,400 per kg) and invertebrates such as abalone (Abercrombie, et al. 2005; Courchamp, et al. 2006; Ward, et al. 2005). In these cases only a limited amount of product is available and the cost of the product is inversely proportional to availability (Courchamp, et al. 2006; Sweijd, et al. 2000). This often leads to the highest value products being those of endangered species.

Due to high demand for these valuable items there is the potential for significant monetary reward from the illegal exploitation of these species. Illegal, unreported and unregulated (IUU) fishing is estimated to be worth between $10bn and $23bn per year and poses a serious threat to the sustainable management of regional fisheries through direct depletion of stocks and by undermining legal fishing efforts (Ogden 2008). For prosecution of illegal exploitation of species to take place, it must be proved that the product being sold is the particular species which is protected under treaties and/or trade agreements. Molecular techniques are therefore invaluable in this area such as in the case of the South African abalone (*Haliotis midae*). Abalone stocks in South Africa have been steadily decreasing in recent years, predominantly due to poaching. In 2008 all abalone fishing was banned in South Africa to try to prevent extinction of the populations (Raemaekers, et al. 2011). Prior to this ban, abalone was being illegally sold deliberately mislabelled as a product of Australia. A DNA based molecular technique allowed the product to be identified as the South African abalone *Haliotis midae* as opposed to the Australian abalone species (Hauck and Sweijd 1999; Raemaekers, et al. 2011), effectively stopping the illegal poaching of this species in South Africa.

Life history research

The application of molecular identification techniques is not confined to commercial fisheries and is also widely used in the research of life history characteristics. Early life history phases of marine organisms are often morphologically cryptic in relation to other species making ecological studies and conservation of these species difficult. To this end, molecular techniques have been used to identify larvae and juvenile fish such as sea bass, cardinal fish, Atlantic salmon and brown trout, rock cod and scombrids (Perrier, et al. 2010; Van de Putte, et al. 2009) and invertebrate larvae (Garland 2002; Goffredi, et al. 2006; Jones, et al. 2008; Webb, et al. 2006) to species level.

3.1.2 **Octopus *cyanea* (Gray, 1849) fisheries**

Artisanal fishing for the day octopus, *Octopus cyanea* is an important economic and subsistence activity along the eastern coast of Africa and islands of the Southwestern

Indian Ocean (SWIO)(Guard and Mgaya 2002). *O. cyanea* is a protein-rich food source for local people, a national source of foreign currency earnings through export to Europe and the Far East and an important component of reef communities (Guard and Mgaya 2002). Exportation of *O. cyanea* has increased in recent years due to an expanding market in Europe and the Far East. Despite this increase in fishing pressure, little is known on the ecology and distribution of *O. cyanea* across the SWIO. Recent research has confirmed the existence of *O. vulgaris* (Lamarck 1798) populations in the SWIO at Amsterdam and St Paul Islands (Guerra, et al. 2010), suggesting that the two species have overlapping ranges. The similarity of appearance, habits and general behaviour of *O. cyanea* and *O. vulgaris* make it difficult to differentiate the two species (Norman 1992; van Heukelem 1973; Wells and Wells 1970; Yarnell 1969). Research has also revealed the co-occurrence of a third species, *Callistoctopus ornatus* (Gould 1852), within the same habitat and among catches of *O. cyanea* in the SWIO (A. Taylor unpublished).

Samples of octopus collected during this study by collaborators were not generally identified to species level during sampling. A simple and effective technique was therefore required to screen octopus samples collected in the SWIO to identify the samples to species level before any analysis was carried out. In addition, accurate species identification is a fundamental prerequisite for stock conservation and management and may also be useful in preventing the mis-selling of *O. cyanea* to foreign markets.

The purpose of this study was therefore to develop a low cost identification method to differentiate between *O. cyanea*, *O. vulgaris* and *C. ornatus*.

3.2 **Materials and methods**

All cytochrome oxidase I (COI) sequences available for the three species on GenBank (*O. cyanea* n = 3, *O. vulgaris* n = 47, *C. ornatus* n = 2), together with additional sequences collected as part of ongoing octopus population genetic research in our group, were aligned using BioEdit (Hall 1999). The alignment was used to identify suitable conserved regions for primer sites for Polymerase Chain Reaction (PCR) amplification of a section of the COI region. Primers OctCOI-F1 (5'-GTTATAATCGGAGGATTTGG-3') and OctCOI-R1 (5'-ATATAGAATTGGATCTCCTC-3') were designed by eye using this BioEdit alignment.

Restriction enzyme cleavage sites within this section were also identified using NEBcutter v.2.0 (Vincze, et al. 2003). Patterns of cleavage site presence were

compared across species to identify species-diagnostic enzyme combinations for Restriction Fragment Length Polymorphism (RFLP) screening. Restriction enzyme assays were performed following manufacturer (New England Biolabs) recommendations with products separated on 2% agarose gels and visualised by ethidium bromide staining. To validate the PCR-RFLP assay restrictions were performed on known specimens of each species; 50 specimens of *O. cyanea* (10 each from Madagascar, Tanzania, Zanzibar, Rodrigues in the Republic of Mauritius and Kenya); 9 specimens of *O. vulgaris* (5 from South Africa (Eastern Cape) and 4 from Vigo, Spain); and 3 specimens of *C. ornatus* (1 from Madagascar and 2 from Mauritius). In all cases samples consisted of arm tips which were preserved in absolute ethanol. DNA was extracted following the method described by (Winnepenninckx, et al. 1993).

Primers OctCOI-F1 (5'-GTTATAATCGGAGGATTTGG-3') and OctCOI-R1 (5'-ATATAGAATTGGATCTCCTC-3') were used to amplify an homologous stretch of 477bp in the three species. The optimised PCR thermoprofile was: 180 s at 95 °C followed by 45 cycles of 30 s at 95 °C, 45 s at 54 °C, 30 s at 72 °C with a final extension step of 120 s at 72°C. Each reaction mix contained 5 µL template DNA (~100 ng), 2 mM $MgCl_2$, 1.0 µM of each primer, 0.08 mM dNTP, 1x reaction buffer (Bioline UK) and 0.08 µL *Taq* polymerase (BIOTAQ, 5 U/µl) in a final reaction volume of 20 µL. A suite of three restriction enzymes was identified that allow unambiguous discrimination of the three species.

Each enzyme recognises a single cleavage site in one of the species that is absent in the other two, producing clear restriction patterns with fragment sizes that correspond with predictions from GenBank sequences (Figure. 3.1). BsdrI restricts solely in *O. cyanea* producing 2 fragments (111bp and 366bp), AlwNI restricts only in *C. ornatus* producing 2 fragments (127bp, 350bp), and HinfI restricts only in *O. vulgaris* producing 2 fragments (91bp and 369bp) (Table 3.1).

Once the above method had been tested, samples of unknown octopus species for the study of *O. cyanea* population structure and phylogeography (Chapter 4) were screened using the above method.

Table 3.1. Table indicating the restriction characteristics of *O. cyanea*, *O. vulgaris* and *C. ornatus*.

Restriction Enzyme	**Species**		
	O. cyanea	***O. vulgaris***	***C. ornatus***
BsdrI	2 fragments	X	X
AlwNI	X	X	2 fragments
HinfI	X	2 fragments	X

3.3 **Results**

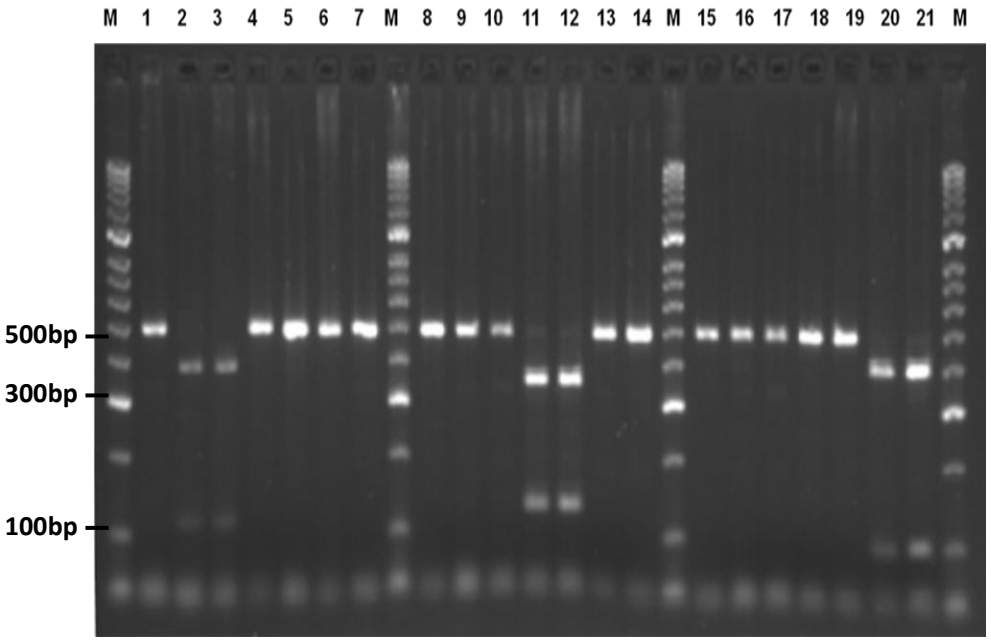

Figure 3.1. Mitochondrial DNA COI gene digest of the three species and their associated DNA fragment patterns separated by agarose gel electrophoresis. 1: unrestricted PCR product *O. cyanea*, 2-3: gene digest BsdrI *O. cyanea*, 4-5: gene digest BSdrI *C. ornatus*, 6-7: gene digest BsdrI *O. vulgaris*, 8: unrestricted PCR product *C. ornatus*, 9-10: gene digest HinfI *O. cyanea*, 11-12: gene digest HinfI *C. ornatus*, 13-14: gene digest HinfI *O. vulgaris*, 15: unrestricted PCR product *O. vulgaris*, 16-17: gene digest AlwNI *O. cyanea*, 18-19: gene digest AlwNI *C. ornatus*, 20-21: gene digest AlwNI *O. vulgaris*, M molecular weight marker Bioline Hyperladder II.

The suite of restriction enzyme assays unambiguously identified the known specimens of each species (Figure 3.1). Following this, all unidentified samples to be used for population genetic studies of *O. cyanea* (Chapter 4) were screened using the above method. Five specimens of *O. vulgaris* and 1 specimen of *C. ornatus* were subsequently identified (Table 3.2).

Table 3.2. Location and number of specimens of *O. vulgaris* and *C. ornatus* collected across the Southwestern Indian Ocean.

Sample Location	Country	*Octopus vulgaris*	*Callistoctopus ornatus*
Ankiembe	Madagascar	0	1
Mauritius	Mauritius	0	1
Mahe	Seychelles	2	0
Praslin	Seychelles	1	0

3.4 Discussion

This report demonstrates a simple, effective technique to differentiate between the three most commonly caught octopus in the SWIO. Although the method is based upon a limited number of polymorphisms, no evidence of noise due to intra-specific mutation has been reported in subsequent work.

In total, over 500 samples were tested during this thesis, only 5 of which were found not to be *O. cyanea*. All samples screened in this study were collected from small-scale fishermen either for local subsistence, sale to local restaurants and in one case export. Small-scale fisheries for *O. cyanea* in the SWIO appear to effectively target *O. cyanea*, with very low catch rates of non-target species. This catch specificity could be due to a number of factors; all fishing for octopus was carried out during daytime hours. *O. cyanea* is the only octopus that actively hunts during daytime hours, making it more likely to be caught by subsistence fishermen. In addition, the range and population sizes of *O. vulgaris* and *C. ornatus* in the SWIO are not well documented. It may be that *O. cyanea* is the dominant octopus species in this region, with very low population numbers of *O. vulgaris* and *C. ornatus* occurring in some parts of the SWIO, resulting in very low levels of other octopus species in the catch.

A fourth species of octopus, the Marbled Octopus, *O. aegina* is recorded as occurring in the SWIO. *O. aegina* is a small octopus reaching a total length of 30cm. It lives in sand and mud (Roper, et al. 1984) and is found across the Indian and Pacific Ocean. It is one of the commonest species in Indo-Malayan markets as well as contributing to subsistence fisheries in East Africa (Roper, et al. 1984). No samples of *O. aegina* were collected during our sampling and it was not therefore possible to include *O. aegina* in our RFLP method. However, the primers used in this study, OctCOI-F1 (5'-GTTATAATCGGAGGATTTGG-3') and OctCOI-R1 (5'-ATATAGAATTGGATCTCCTC-3') were designed to amplify the COI region of *O. aegina* should any samples be collected. In addition, restriction enzyme cleavage work on sequences of *O. aegina* from GenBank using NEBcutter v 2.0 have shown that the restriction enzyme TfiI would

restrict solely in *O. aegina*, thereby differentiating this species from the other octopus species in this study. This RFLP method can therefore easily be adjusted to accommodate the identification of a fourth species of octopus found in the SWIO should any more samples be collected.

Authentication of seafood species is an increasingly important issue within the seafood industry. As international trade in seafood increases, driven by increasing worldwide consumption, cases of economic fraud are becoming increasingly common (Civera 2003). Seafood mis-marketing may put consumers at risk and reduce the effectiveness of marine conservation and management programs designed to protect at risk species (Civera 2003). DNA-based identification techniques are therefore essential to the management of fisheries. With catches of *O. cyanea* for export increasing (Guard and Mgaya 2002) and the similarity in appearance of *O. cyanea* to *O. vulgaris*, the development of a simple species identification technique represents an important genetic tool that can monitor exports of *O. cyanea* to foreign markets, facilitate the collection of species-specific octopus catch data and contribute to the conservation and management of octopus stocks across the SWIO.

Chapter 4: Population Structuring and Phylogeography of *Octopus cyanea* across the South-west Indian Ocean

4.1 Introduction

Many marine species, including some of the most economically important taxa, exhibit features such as high dispersal potential, high fecundity, large population sizes and wide geographic distributions. These traits, in conjunction with the lack of conspicuous barriers to dispersal in the marine environment, traditionally led to expectations of high gene flow, sparse local adaptation and low levels of genetic drift culminating in limited opportunities for population divergence (Féral 2002; Kenchington, et al. 2009; Waples 1998; Ward, et al. 1994). However, population genetic studies of marine taxa have provided evidence of significant intra-specific neutral and adaptive structuring (Pérez-Losada, et al. 2002; Perez-Losada, et al. 2007; Santos, et al. 2006; Zardoya, et al. 2004). Structuring may be driven by mechanisms that are extrinsic (Doubleday, et al. 2009; Fraser, et al. 2009; García-Rodríguez and Perez-Enriquez 2006; Janko, et al. 2007; Koizumi, et al. 2012; Palero, et al. 2008; Perez-Losada, et al. 2007; Selkoe, et al. 2006; von der Heyden, et al. 2008) or intrinsic to a species (Gerlach, et al. 2007; Purcell, et al. 2009; Purcell, et al. 2006), or an interplay of both and be significant to varying degrees on ecological and evolutionary time frames. Resolution of the extent of such structuring and its determinants are central to our ability to manage and conserve genetic biodiversity of species and are fundamental to predicting responses to harvesting and environmental changes.

In benthic species where adults have limited dispersal abilities, long-distance dispersal and consequently gene flow is facilitated by the pelagic larval phase. Traditionally, duration of the pelagic larval phase has been used to predict dispersal capability of larvae and hence connectivity of populations (Bohonak 1999; Scheltema 1968). However, it is now apparent that long pelagic larval duration may not always confer high levels of genetic connectivity among populations, with life history traits such as larval behaviour and extrinsic factors such as oceanographic features (Mokhtar-Jamai, et al. 2011; Selkoe, et al. 2010; Thomas and Bell 2013) considerably influencing dispersal of larvae and hence gene flow between populations.

The big blue octopus or day octopus, *Octopus cyanea* is a benthic species and is widespread on coral reefs throughout the Indo-Pacific from East Africa to the Hawaiian Islands in tropical and warm waters, including the Red Sea, India and Australia. It is a large octopus on average weighing about 6kg (Guard and Mgaya 2002; van Heukelem 1973) and hunts during daylight hours, an unusual feature among Octopodidae. The duration of the pelagic larval phase ranges from 3 weeks to 6 months in Octopodidae

(Villanueva and Norman 2008). Observations of *O. cyanea* larvae in the laboratory confirm that *O. cyanea* has a life history with a considerable planktonic larval phase and produces paralarvae which are specialised for a pelagic existence (van Heukelem 1973; Villanueva and Norman 2008). Larvae have not been raised to settlement in the laboratory but pelagic larval duration is likely to be approximately 50 days, similar in length to that of *O. vulgaris*, which possesses similar life history characteristics (van Heukelem 1973; Villanueva 1995; Villanueva and Norman 2008).

The range of *O. cyanea* includes the Southwestern Indian Ocean comprising East Africa, Tanzania, Madagascar, Mozambique, Kenya, Comoros Islands, the islands of the Mascarene Plateau including the Seychelles and Mauritius and South Africa. The large-scale oceanographic currents in this region may have a profound effect on population structuring of *O. cyanea*. Prevailing currents in the SWIO flow from east to west in the form of the South Equatorial Current (SEC) (Schott, et al. 2009). The SEC flows west across the Indian Ocean towards Madagascar transporting 5055 Sverdrups (Sverdrup = 10^6 m^3 s^{-1}, Sv) westwards and reaching speeds of 11cm s^{-1}. This current divides off the east coast of Madagascar to form the East Madagascar current and the northern flowing North Equatorial Current (NEC) (Schott, et al. 2009). On reaching the east coast of Africa, the NEC splits in two, forming the East African Coastal Current which travels northwards along the coast of East Africa and a current which enters the Mozambique Channel (see Chapter 1.4 for further information) .

It is predicted that due to *O. cyanea*'s considerable pelagic larval duration and the numerous large-scale currents present across the SWIO, which may provide *O. cyanea* with significant dispersal potential, intra-specific population structure within the region will be limited.

Despite the socio-economic importance of octopus in the SWIO, little research has been conducted into the stocks of *O. cyanea*. The most extensive research to date has been conducted in Tanzania (Guard and Mgaya 2002). Research into the artisanal octopus fishery in Tanzania analysed specimens from three sites from April 2000 to June 2001. The total economic value of the octopus fishery at these sites for the local fishermen during this time period was $245 000, a considerable income for this area. At two of the sites average weight and catch per unit effort (CPUE) were significantly lower than at the third site. Higher abundance of smaller individuals and a lower overall biomass was also observed at the first two sites. The results suggest overfishing of octopus may be occurring at two of the sites, significantly reducing both reproductive output and success. Interestingly, capture records suggest that females may be more prone to capture during brooding periods, possibly due to the obviousness of their dens, an important consideration in the management of stocks.

Information on population structuring is vital for sustainable management of fishery resources and conservation of marine biodiversity. The need for population studies of *O. cyanea* is emphasised by increasing fishing pressures. Exact catch statistics for *O. cyanea* are difficult to obtain as catch is not grouped by species and some countries only submit catch under the category of cephalopods rather than by specific cephalopod groups (FAO 2013). However, a conservative estimate would be more than 3000 tons of *O. cyanea* are harvested annually across the SWIO (FAO 2013). Across the SWIO in 2011, 1781 tons of octopus were harvested and a further 3057 tons of unspecified cephalopods were also caught, a significant proportion of which are likely to be *O. cyanea*. Of the countries submitting specific catch data for octopus, Tanzania has the highest catch production of any of the SWIO countries, catching between 700 – 1700 tons annually (FAO 2013). Of this catch, 99% is expected to be *O. cyanea* as it is the dominant species in artisanal octopus fisheries (Guard 2009). Madagascar and Mozambique catches for cephalopods in 2011 were 1765 and 1292 tons respectively. Mauritian catch statistics show a sharp decrease in catch from 335 tons in 2002 to 84 tons in 2006 with catches remaining low from 2006 up to the latest FAO catch data in 2011, possibly suggesting overexploited populations in this area (FAO 2013).

Exports of *O. cyanea* have increased markedly across the SWIO in recent years (Guard 2009; Guard and Mgaya 2002). Guard and Mgaya (2002) report that in Madagascar a 35% increase in exports of octopus to France occurred between 2002 and 2003. However, since 2006, the FAO has not provided totals of octopus exports from individual countries making estimates of the extent of exports difficult to quantify accurately (FAO 2013).

This study employed a combination of microsatellite and mitochondrial DNA markers to investigate the genetic diversity and population structuring of *O. cyanea* throughout the SWIO. An additional objective was to partition historical and contemporary drivers of genetic structuring to inform fishery management and increase our understanding of biodiversity within the region.

4.2 Methods

4.2.1 Sampling and DNA extraction

Sampling was carried out at 15 locations across the SWIO between 2010 and 2012 (Figure 4.1, Table 4.1). Arm clips were collected from samples obtained from local fishermen and fish markets and subsequently preserved in 95% ethanol.

Total genomic DNA was extracted using a modified phenol/chloroform method (Sambrook et al. 1989).

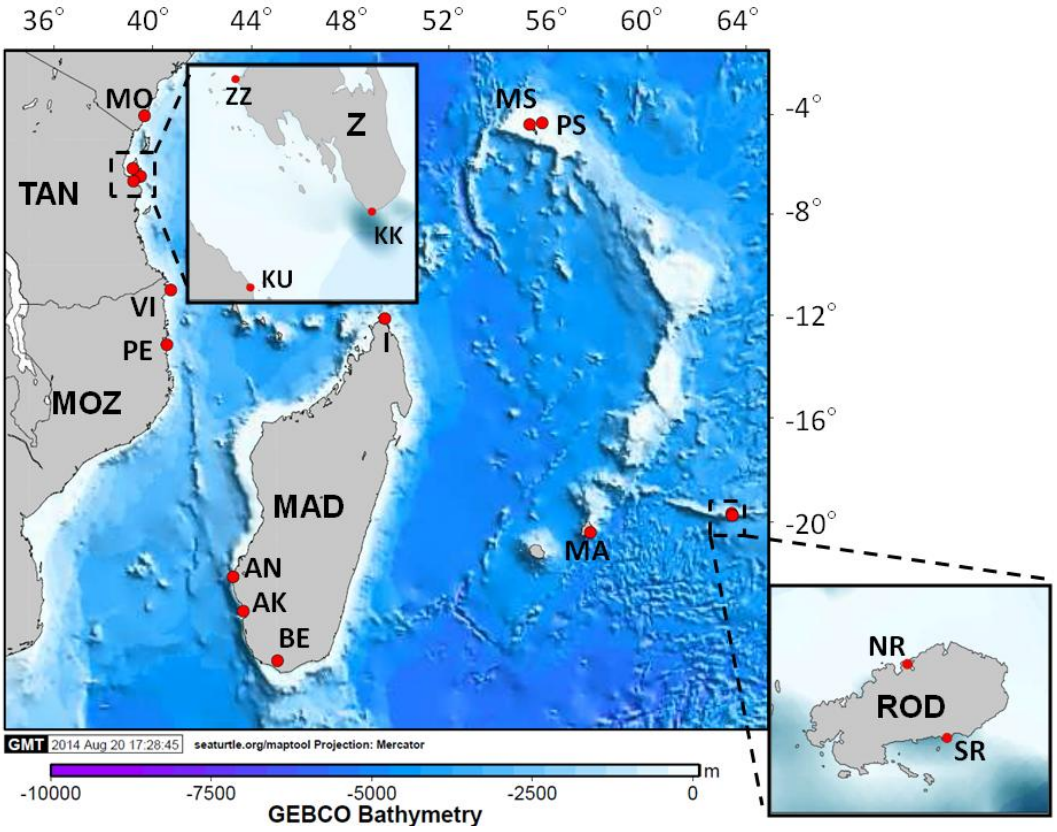

Figure 4.1. Sampling locations of *O. cyanea* across the SWIO. TAN-Tanzania, MOZ-Mozambique, ROD-Rodrigues, MAD-Madagascar, Z-Zanzibar Island, MA-Mauritius.

4.2.2 **Genetic screening**

Individuals were genotyped at 10 species-specific nuclear microsatellite loci (ROC32, ROC6, ROC28, OC22, OC31, OC19, ROC17, OC15, ROC1 and OC18 (see Chapter 2.2.3. for microsatellite development) using protocols described in Chapter 2.2.5.

Individuals were also sequenced at three mitochondrial DNA (mtDNA) regions cytochrome oxidase I (COI), 16S rDNA and the octopus non-coding region (NCR) following protocols outlined in Chapter 2.2.2.

Table 4.1. Sampling information for *O. cyanea* across the SWIO.

Country	Location	Sample Code	Sample Size
Kenya	Mombassa fish market	MO	58
Zanzibar, Tanzania	Kizim Kazi	KK	34
	Stone Town fish market	ZZ	80
Tanzania	Kunduchi	KU	60
Mozambique	Murrumbue Beach, Pemba	PE	52
	Vamizi Island	VI	13
Madagascar	Andavadoaka	AN	80
	Ankiembe	AK	80
	Beheloke	BE	80
	Ivovona	I	80
Rodrigues, Mauritius	Northern Rodrigues	NR	51
	Southern Rodrigues	SR	52
Mauritius	Mahebourg	MA	80
Seychelles	North Island	MS	49
	Grand Anse, Praslin	PS	62

It was hypothesised that, as in the case of the control region in metazoans, this non-coding may have a higher mutation rate than coding regions of the mitochondrial genome and may therefore be a more informative marker than COI and 16S rDNA in *O. cyanea* (for more information see Chapter 2..2.2). As an initial sample, the same 124 individuals which were sequenced for both COI and 16S rDNA were sequenced for NCR. The NCR sequences displayed higher nucleotide and haplotype diversity than coding regions of the mtDNA and sample size was subsequently increased to 206 individuals.

Table 4.2. Primer pairs, primer sequence and PCR conditions for amplification of COI, 16S rDNA and NCR of *O. cyanea* used in this study.

Amplified region	Primer name	Primer sequence 5'-3'	Sample Size	Annealing temperature °C
COI	OC COI F OC COI R (this thesis)	CATTTTCGGAATTTGATCAGG GATTGGGTCTCCTCCACCTC	124	55
16S rDNA	16Sar-L 16Sbr-H (Palumbi 1996)	CGCCTGTTTATCAAAAACAT CCGGTCTGAACTCAGATCACGT	124	55
NCR	NCR F2 NCR R2 (this thesis)	TCCTGTTAATGGTCAGGGTCTAA TTCAACAGCACTATTATTGAACA	206	54

4.2.3 **Microsatellite analyses**

Genotypes at all pairs of loci were tested for genotypic linkage disequilibrium and deviation from Hardy Weinberg expectations using an Exact Test with significance determined by a Markov chain method in GENEPOP v 4.1, (Raymond and Rousset 1995b) (as described in Chapter 2.3.2). Data were screened for amplification errors such as null alleles, large allele drop-out and stuttering using MICRO-CHECKER (Van Oosterhout, et al. 2004) and FreeNA (Chapuis and Estoup 2007).

Population Diversity

The levels of genetic variation within samples were estimated as the proportion of heterozygotes observed (H_o), allelic richness (AR, number of alleles observed corrected across all samples for the minimum m sample size screened), observed and expected heterozygosity (H_O, H_E) and Wright's inbreeding coefficient (FIS) were calculated in FSTAT (Goudet 1995) (see Chapter 2.3.2). Statistical significance between populations was assessed using Mann Whitney U tests.

Population structure

POWSIM (Ryman and Palm 2006) was used to assess the statistical power of the microsatellite dataset to detect population structuring using the settings as described in Chapter 2.3.2.

The genetic differentiation among samples was analysed using exact tests of differences in allele frequencies using GENEPOP (Raymond and Rousset 1995b) and F_{ST} (Wright 1949) as described in Chapter 2.3.2. Exact tests of genic differentiation were calculated for all seven loci, six loci and also for each locus separately in

GENEPOP (Raymond and Rousset 1995b). Correction for Type 1 error was carried out using Bonferroni correction (Rice 1989). Tests for detection of microsatellites under selection were performed using the F_{ST}-outlier method of Beaumont and Nichols (1996) implemented in LOSITAN (Antao, et al. 2008) (see Chapter 2.3.2).

In addition, population structuring was assessed using STRUCTURE (Pritchard 2010; Pritchard, et al. 2003) as described in Chapter 2.3.2. Results were then analysed by STRUCTURE HARVESTER (Earl and vonHoldt 2012) as described in Chapter 2.3.2. Isolation-by-distance was assessed using Mantel's test (Mantel 1967) from the comparison of all pairwise F_{ST} values with pairwise geographic distances and pairwise oceanographic distances using the Isolation-by-distance web service (Jensen, et al. 2005).

Demographic history

To test for a recent bottleneck the M-ratio, the mean ratio of the number of alleles (k) to the range in allele size (r), was calculated using the software M_P_VAL (Garza and Williamson 2001) as described in Chapter 2.3.2. Effective population size (N_e) was estimated to be 5000 pre-bottleneck. Since the absolute rate of microsatellite mutation in our study species is unknown, θ was calculated using a range of mutation rates 1 x 10^{-4}, 3 x 10^{-4} and 5 x 10^{-4}, which are typical of mutation rates of microsatellites in other species. M_P_VAL was therefore run with θ of 2, 6 and 10. Significance of our M ratio was then calculated using the program CRITICAL_M (see Chapter 2.3.2). To further test for a population bottleneck, variation between the expected and observed heterozygosity relative to the number of alleles was calculated by the program BOTTLENECK (Piry, et al. 1999) using the settings described in Chapter 2.3.2.

4.2.4 **Mitochondrial DNA analyses**

Population diversity

Arlequin 3.5.1.2 (Excoffier and Lischer 2010) was used to calculate the number of haploptypes (H), haplotype diversity (*h*) and nucleotide diversity (π) for COI, 16S rDNA and NCR as described in Chapter 2.3.1. Samples were rarefied to the smallest sample size to compare haplotype richness across sampling locations using Microsoft Excel. Haplotype networks were constructed using maximum likelihood methods in PHYLIP DNAml (Felsenstein 1989) as described in Chapter 2.3.1 and subsequently constructed in Hapview (Ewing 2010).

Population structure and phylogeography

Statistical power to detect population structuring was assessed using POWSIM v 4.0 (Ryman and Palm 2006) adjusted for organelle (mtDNA) data (Larsson, et al. 2009) to

detect genetic differentiation at various levels of F_{ST} using an effective population size (N_e) of 10,000 and varying the number of generations (t) accordingly as described in Chapter 2.3.1. Where suitable, population structure was investigated using F_{ST} and exact tests of population differentiation calculated in Arlequin 3.5.1.2 (Excoffier and Lischer 2010) (see Chapter 2.3.1) as described in Chapter 2.3.1. Where multiple tests were conducted, significance levels were adjusted according to a Bonferroni correction (Rice 1989).

Demographic history

Fu's Fs (Fu 1997), Tajima's D (Tajima 1989) and Ewens-Watterson test (F) (Slatkin 1994a) neutrality tests were performed using Arlequin 3.5.1.2 (Excoffier and Lischer 2010). Statistical significance was assessed after 1000 permutations as described in Chapter 2.3.1.

Mismatch distribution was calculated using NCR sequences for a pure demographic expansion in Arlequin 3.5.1.2 (Excoffier and Lischer 2010) as described in Chapter 2.3.1.

There are no mutation rate estimates for any closely related species of *O. cyanea*. At present the only mutation rate for octopods is for COI mtDNA; 3.81 substitutions per site per billion years (Strugnell, et al. 2008), based on phylogenetic analysis of an Antarctic octopus *Pareldone turqueti* using fossil calibration. Recent theory on the molecular clock, however, suggests that the rate of the molecular clock varies according to metabolic rate, which is dependent on body size and temperature (Gillooly, et al. 2005). Due to the massive temperature difference between the Antarctic and the SWIO (2°C versus +24°C), this mutation rate may not be applicable to tropical octopods.

Overall mutation rates of 0.356% per million years (Myr) and 0.056% per Myr, have been estimated for the entire mitochondrial genome of the giant squid *Architeuthis* spp.(Winkelmann, et al. 2013). However, as mentioned earlier, the molecular clock is thought to be body size and temperature dependent (Gillooly, et al. 2005). As *Architeuthis* spp. are an extremely large, deep-sea species, it may be that this mutation rate may not be particularly suitable for shallow-water octopods. Additionally, the time dependence hypothesis of mutation rates suggests rates estimated using ancient fossil (as in this case) calibrations may significantly underestimate the mutation rate when applied to relevant demographic history (Ho and Phillips 2009).

Ultimately, mutation rates for COI and 16S rRNA were not required for this study as the uninformative nature of these markers meant that they were not suitable for analysis.

Substitution rates of NCR were therefore derived from a well-studied past oceanographic event in the SWIO which may have had significant impact of *O. cyanea* populations. Additionally substitution rates from published studies on gastropod evolution, the most closely related Molluscan class to Octopoda (Marko 2002), were also used.

Estimates of mutation rates of Gastropoda COI range from 0.335 to 1.2 % Myr (Marko 2002) but no estimates exist for the non-coding region, which is expected to have a higher mutation rate due to its non-coding nature. Mutation rates of 1.2% per Myr, the higher end of the estimates of COI mutation rates and 1.8% per Myr, (mutation rate is half the divergence rate) a mutation rate around the order observed in the control region of tropical marine fish (Donaldson and Wilson 1999) were therefore additionally used, equivalent to 1.2×10^{-8} and 1.8×10^{-8} mutations per site per year respectively.

Additionally, the mutation rate in NCR was calculated using the flooding of the Seychelles Bank. The flooding of the Seychelles Bank, an area of 42,000 km^2, occurred approximately 9,840 years ago (9.84 KA) ± 250 years (Badyukov, et al. 1989) and may have provided a large increase in available habitat to *O. cyanea*, resulting in a population expansion. The increase in flat, shallow seabed and subsequent reef building which began 8 - 9 KA, would have represented a massive habitat expansion for *O. cyanea*. The expansion time of *O. cyanea* in the Seychelles was therefore fixed at 9.0 KA. Using this calibration, a substitution rate of 1.17×10^{-7} per site per year was calculated for NCR and was used to calculate expansion time estimates across the SWIO.

Calculations of time since expansion using mismatch analyses were checked using the online spreadsheet http://www.uni-graz.at/zoowww/mismatchcalc/index.php (Schenekar and Weiss 2011).

The most suitable model of nucleotide substitution to account for the observed sequence polymorphism in samples was calculated using jModelTest version 2.1.3 (Darriba, et al. 2012; Posada 2008). Bayesian skyline plots were conducted in BEAST v.1.7.5 (Drummond, et al. 2012) to investigate past effective population sizes as described in Chapter 2.3.1. The substitution model HKY and a substitution rate of 1.17×10^{-4} mutations per site per thousand years were used. Results were examined in TRACER (Rambaut and Drummond 2009) as described in Chapter 2.3.1.

4.3 Results

4.3.1 Microsatellite analyses

Nine hundred and fifty six individuals from 14 locations were genotyped for 10 microsatellite loci. Kizim Kazi, Tanzania was not included in the microsatellite

screening due to the small sample size (n = 34). .After the removal of replicate samples and linked loci (see below), the final dataset comprised of seven hundred and thirty individuals from 12 localities genotyped for 7 microsatellite loci.

Tests for Hardy Weinberg Equilibrium (HWE) and Linkage Disequilibrium (LD)

HWE tests performed in Genepop 4.1 (Rousset and Raymond 1995) found significant deviation from HWE after Bonferroni correction in all sampling locations (Table 4.3) due to heterozygote deficiency.

Table 4.3 Chi-squared values and significance of deviation from HWE in the 14 sampling sites used to assess population structuring in *O. cyanea* across the SWIO. Statistically significant estimates ($p<0.05$) are highlighted in bold. Significance after Bonferroni correction is indicated by *.

Site	**Loci showing deviation**	**Overall deviation from HWE**
MO	OC31, OC15, OC18	χ^2 = Infinity, p **= 0.0001***
KU	OC15	χ^2 = Infinity, **p = 0.0001***
ZZ	OC31, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
PE	OC22, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
VI	OC22, OC15	χ^2 = Infinity, **p = 0.0001***
AN	OC22, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
AK	OC22, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
BE	OC22, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
I	OC22, OC19, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
NR	ROC6, OC31, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
SR	ROC6, OC22, OC19, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
MA	ROC6, OC31, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
MS	ROC6, OC22, OC31, ROC28, OC15, OC18	χ^2 = Infinity, **p = 0.0001***
PS	ROC6, OC22, OC15, OC18	χ^2 = Infinity, **p = 0.0001***

Locus OC15 significantly deviated from HWE at all sampling sites. Global tests of linkage disequilibrium revealed significant linkage between 6 pairs of loci due to replicates in the samples (see below).

Removal of replicates

Significant deviation from HWE is caused by a number of factors such as the Wahlund effect, null alleles, inbreeding and selection (Beebee and Rowe 2001). Additionally, replications are also known to cause significant deviations from HWE and LD. All sample sites were therefore checked for replicates and any replicates were removed from the dataset. In this case, replicates were defined as multiple arm tips collected from the same octopus. A large number of replicates were found in the Ankiembe

sample and also the Vamizi Island sample which were collected by collaborators. The original sample size for Vamizi Island was 80 individuals. However, an error during the collection of these samples by a collaborator meant that a large number of the samples were replicates from the same octopus. This ranged from two to eight replicates from the same individual (one for each arm). Once all replicates had been removed from the dataset the final sample size was 13. The sample size was therefore too small to be used for microsatellite analysis and the Vamizi Island sample was removed from the dataset. The original sample size for Ankiembe was also 80. However after removal of a number of replicates, final sample size was reduced to 20. This sample size was too small for microsatellite analysis and therefore the Ankiembe sample was also removed from the dataset.

After replicate removal, the dataset was tested again for HWE and linkage equilibrium. All sites displayed significant multilocus deviation from HWE after Bonferroni correction (X^2 = infinity, $p < 0.0001$) and a number of loci showed significant deviations from HWE: OC31, OC22, ROC6, OC18, OC15. The test for linkage disequilibrium found significant linkage between three pairs of loci ($p < 0.05$); OC31 and ROC1, OC19 and OC17, OC22 and OC15. Linkage of these loci was observed at all sample locations.

To test for large allele drop-out, stuttering and null alleles in the dataset, MICRO-CHECKER and FreeNA were used. Five loci displayed null alleles using MICRO-CHECKER; OC22, OC31, ROC1, OC18 and OC15. FreeNA also identified OC22, OC31, OC18 and OC15 as null alleles. However, in FreeNA ROC1 was borderline null allele, displaying a lower null allele frequency than ROC32 (0.1342 versus 0.1369 respectively), which was not considered a null allele. No large allele drop-out or stuttering was observed.

To remove the linkage disequilibrium and also to reduce the number of loci with null alleles, OC15, OC19 and OC31 were removed from the dataset. Both OC15 and OC31 displayed null allele characteristics and OC19 showed greater deviation from HWE than ROC17 which it was linked to.

The final microsatellite dataset contained 7 loci; ROC32, ROC6, ROC1, OC22, ROC17, OC18 and ROC28. Of these loci, OC22, OC18 and possibly ROC1 contain null alleles.

Tests for HWE and linkage disequilibrium were then repeated for the 7 remaining loci (Tables 4.4, 4.5). No loci or sites displayed any evidence of linkage disequilibrium.

Table 4.4. Chi-squared values and significance of deviation from HWE of the 7 microsatellite loci used to investigate population structuring of *O. cyanea* in the SWIO. Statistically significant estimates (p<0.05) are highlighted in bold. Significance after Bonferroni correction is indicated by *.

Locus	Overall Deviation from HWE
ROC32	χ^2 = 30.7440, p = 0.2379
Oc22	χ^2 = Infinity, **p = 0.0001***
ROC28	χ^2 = 35.0619, p = 0.1103
ROC1	χ^2 = 20.6091, p = 0.7618
OC18	χ^2 = Infinity, **p = 0.0001***
ROC17	χ^2 = 37.0782, p = 0.0736
ROC6	χ^2 = Infinity, **p = 0.0001***

Table 4.5. Chi-squared values and significance of deviation from HWE in the 12 sampling sites used to assess population structuring in *O. cyanea* across the SWIO. Statistically significant estimates (p<0.05) are highlighted in bold. Significance after Bonferroni correction is indicated by *.

Site	Loci showing deviation	Overall Deviation from HWE
MO	OC18	χ^2 = Infinity, **p = 0.0001***
KU	ROC17	χ^2 = 33.04, **p = 0.0028**
ZZ	OC18	χ^2 = Infinity, **p = 0.0001***
PE	Oc18, OC22	χ^2 = Infinity, **p = 0.0001***
AN	OC18, OC22	χ^2 = Infinity, **p = 0.0001***
BE	OC18, OC22	χ^2 = Infinity, **p = 0.0001***
I	Oc18, OC22	χ^2 = Infinity, **p = 0.0001***
NR	Oc18	χ^2 = Infinity, **p = 0.0001***
SR	OC18, OC22, ROC6	χ^2 = Infinity, **p = 0.0001***
MA	OC18, ROC6	χ^2= Infinity, **p = 0.0001***
MS	OC18, OC22, ROC6, ROC28	χ^2 = 65.07, **p = 0.0001***
PS	Oc18, OC22, ROC6	χ^2 = Infinity, **p = 0.0001***

Three loci are principally responsible for the remaining deviations from HWE; OC18, OC22 and ROC6. OC18 and OC22 are null alleles and it is therefore likely that the deviation from HWE is caused by a heterozygote deficit at sites where departure from HWE is observed. ROC6 is not a null allele but deviates from HWE at South Rodrigues, Mauritius, North island and Praslin.

Tests for detection of microsatellites under selection

Runs in LOSITAN (Antao, et al. 2008) using all loci and all sampling sites and runs using pairs of loci for all sample sites failed to detect any non-neutrality of markers. Testing using all loci at pairs of sample sites indicated that the locus ROC6 was an outlier and subject to positive selection when one of the sample locations was Mauritius (Figure 4.2). Consequentially, tests of population differentiation were carried out using the 6 neutral loci in addition to the full set of 7 loci.

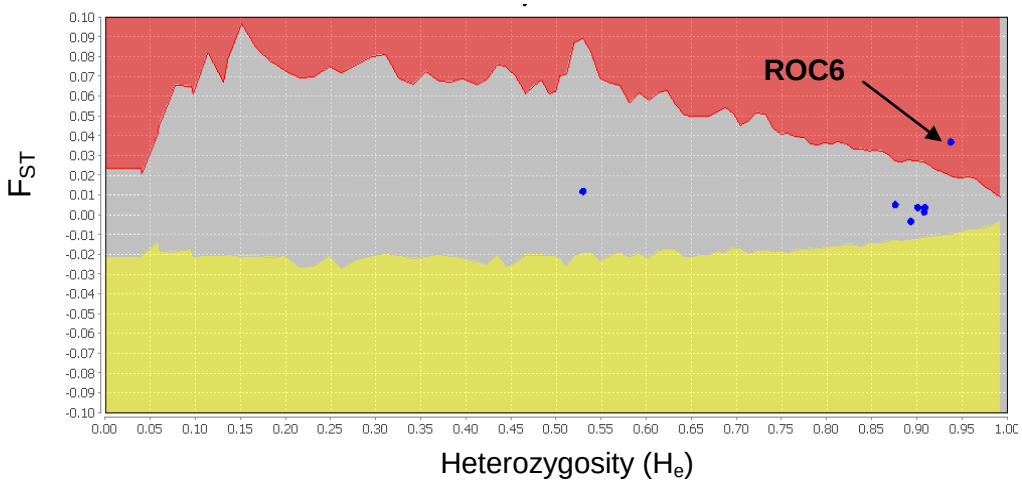

Figure 4.2. Identification of outlier loci using the program LOSITAN (Antao et al. 2008) The central grey area represents candidates for neutral loci, the yellow area represents candidates for balancing selection and the upper red area represents candidates for positive selection. Markers are indicated by ●. Marker ROC6 represents the single outlier locus in the 7 loci tested.

Population diversity

The average number of alleles per locus found at each sampling location was high, ranging from 13.57 – 14.52 (Table 4.6). Locus OC22 exhibited the lowest number of alleles (8-13 per sampling location). This decreased number of alleles compared to other loci is likely to be due to OC22's null allele characteristics. This is also reflected in decreased allelic richness, higher FIS values (-0.144 – 0.570) and lower observed heterozygosities (0.25 - 0.45) for OC22 at all sampling sites compared to the other loci. Allelic richness ranged from 9.3 - 10.02 across sampling locations. Mauritius displayed the lowest allelic richness of 9.3 compared to 9.63 - 10.02 at other sampling sites. Pairwise tests of significance of FIS, AR, Na H_E and H_O between sampling locations were not significant.

Table 4.6. Genetic diversity of 8 species-specific microsatellite loci in *O. cyanea* samples: n number of individuals genotyped; Na number of alleles; AR allelic richness; H_E expected heterozygosity, H_O observed heterozygosity; FIS inbreeding coefficient.

		MO	KU	ZZ	PE	AN	BE	I	NR	SR	MA	MS	PS
ROC 32	**n**	56	58	76	52	74	68	53	40	48	73	47	60
	Na	13	15	14	15	14	14	14	12	14	15	14	14
	AR	9.71	10.07	9.58	10.65	9.91	9.83	10.07	9.66	10.63	9.60	9.73	10.44
	H_E	0.90	0.90	0.89	0.91	0.90	0.90	0.89	0.87	0.91	0.89	0.89	0.91
	H_O	0.98	0.88	0.85	0.94	0.88	0.90	0.82	0.74	0.9	0.88	0.90	0.91
	FIS	-0.09	0.02	0.05	-0.03	0.02	0.01	0.09	0.16	0.02	0.02	-0.01	-0.01
OC 22	**n**	56	58	76	52	74	68	53	40	48	73	47	60
	Na	9	11	13	8	13	11	9	10	8	13	11	12
	AR	3.97	5.27	5.83	4.78	6.41	6.33	5.17	5.09	5.25	5.70	6.41	6.97
	H_E	0.29	0.42	0.46	0.48	0.53	0.60	0.44	0.40	0.46	0.45	0.51	0.58
	H_O	0.25	0.40	0.40	0.31	0.43	0.47	0.34	0.46	0.38	0.41	0.33	0.45
	FIS	0.15	0.05	0.14	**0.36**	0.19	**0.22**	**0.23**	-0.14	0.19	0.09	**0.36**	0.22
ROC 28	**N**	56	58	76	52	74	68	53	40	48	73	47	60
	Na	16	17	16	16	17	17	17	16	15	18	14	16
	AR	11.26	10.92	10.39	10.85	11.38	10.73	11.30	11.97	11.58	10.80	10.61	11.27
	H_E	0.92	0.91	0.91	0.91	0.92	0.91	0.92	0.93	0.92	0.90	0.91	0.92
	H_O	0.89	0.86	0.89	0.88	0.86	0.91	0.84	0.93	0.96	0.82	0.89	0.85
	FIS	0.03	0.06	0.02	0.03	0.06	0.01	0.08	0.01	-0.05	0.09	0.03	0.08
ROC 1	**N**	56	58	76	52	74	68	53	40	48	73	47	60
	Na	13	15	16	15	14	16	13	14	14	15	16	13
	AR	9.85	10.15	10.15	9.74	10.17	10.92	9.63	10.37	9.32	10.09	10.75	9.48
	H_E	0.90	0.88	0.90	0.90	0.90	0.91	0.90	0.90	0.88	0.90	0.90	0.87
	H_O	0.80	0.93	0.92	0.85	0.82	0.92	0.87	0.90	0.92	0.92	0.88	0.91
	FIS	0.11	-0.06	-0.02	0.06	0.10	-0.01	0.03	-0.01	-0.05	-0.02	0.03	-0.47
OC 18	**N**	56	58	76	52	74	68	53	40	48	73	47	60
	Na	16	16	17	17	13	13	18	15	14	16	13	19
	AR	10.37	10.11	9.77	10.09	9.30	9.63	10.36	10.20	10.23	9.66	8.53	10.78
	H_E	0.90	0.88	0.88	0.85	0.88	0.87	0.87	0.89	0.87	0.87	0.82	0.87
	H_O	0.56	0.77	0.60	0.42	0.43	0.61	0.62	0.49	0.55	0.60	0.68	0.66
	FIS	**0.37**	0.12	**0.33**	**0.51**	**0.52**	**0.29**	**0.29**	**0.46**	**0.37**	**0.32**	0.18	**0.24**
ROC 17	**N**	56	58	76	52	74	68	53	40	48	73	47	60
	Na	16	15	14	17	15	14	16	14	14	16	13	12
	AR	9.80	8.87	9.44	10.30	9.38	9.09	10.09	9.87	9.45	10.49	10.26	9.16
	H_E	0.89	0.87	0.88	0.89	0.89	0.88	0.90	0.89	0.91	0.91	0.91	0.89
	H_O	0.86	0.91	0.92	0.89	0.91	0.97	0.96	0.83	0.92	0.93	1.0	0.95
	FIS	0.04	-0.06	-0.04	0.01	-0.02	-0.09	-0.07	0.08	-0.04	-0.03	-0.11	-0.06
ROC 6	**N**	56	58	76	52	74	68	53	40	48	73	47	60
	Na	19	19	18	18	18	19	16	14	17	13	15	18
	AR	12.72	12.01	13.13	12.50	12.78	12.71	11.86	11.21	11.74	8.78	11.37	12.04
	H_E	0.92	0.92	0.94	0.93	0.94	0.94	0.93	0.93	0.93	0.87	0.92	0.93
	H_O	0.92	0.82	0.87	0.90	0.88	0.90	0.90	0.81	0.88	0.82	0.9	0.75
	FIS	0.02	0.11	0.08	0.03	0.06	0.05	0.03	0.124	0.05	0.06	0.02	**0.19**
Av. over all sites	**N**	56	58	76	52	74	68	53	40	48	73	47	60
	Na	14.57	15.42	15.42	15.14	14.86	14.86	14.71	13.57	13.71	15.14	13.71	14.86
	AR	9.67	9.63	9.76	9.84	9.90	9.89	9.78	9.77	9.74	9.30	9.67	10.02
	HE	0.74	0.83	0.84	0.84	0.85	0.86	0.84	0.83	0.84	0.83	0.84	0.85
	HO	0.75	0.80	0.78	0.74	0.74	0.81	0.76	0.74	0.79	0.77	0.80	0.78
	FIS	**0.08**	0.04	**0.07**	**0.12**	0.06	**0.07**	**0.09**	**0.12**	0.06	**0.08**	0.05	**0.08**

Population structure

Estimation of the power of the dataset to detect significant population structuring was estimated using POWSIM (Ryman and Palm 2006). The data had considerable power to detect even very low levels of differentiation (Table 4.7) and had a 100% power to detect F_{ST} values of 0.01 and above. The estimated occurrence of type 1 error in the dataset was 5% and 6% respectively for Chi^2 and Fisher tests (Table 4.7).

Table 4.7. Statistical power of *O. cyanea* microsatellite loci to detect population differentiation at various levels of F_{ST} using tests based on Chi^2 and Fisher exact methods.

	Statistical Power		**Type 1 error Estimate**	
F_{ST}	Chi^2	Fisher	Chi^2	Fisher
0.001	0.89	0.85	0.05	0.06
0.01	1.00	1.00	0.04	0.05
0.05	1.00	1.00	0.04	0.05

Due to ROC6 displaying selection, F_{ST} values were calculated first including ROC6 and then subsequently omitting ROC6. F_{ST} values including ROC6 showed 30 significant population differentiations (Table 4.8), 7 of which remained significant after Bonferroni correction (Rice 1989). F_{ST} values were then calculated using 6 loci; ROC32, OC18, OC22, ROC28, ROC1 and ROC17, omitting ROC6. No significant population structuring was observed after Bonferroni correction (Table 4.9). No population differentiation across sample sites was observed using the single loci OC18, OC22, ROC28, ROC1 and ROC17. After applying Bonferroni correction ROC32 revealed Mauritius was significantly differentiated from Andavadoaka. F_{ST} calculated using locus ROC6 revealed 27 instances of significant population differentiation, 10 of which remained significant after Bonferroni correction (Rice 1989). The omission of ROC6, a locus under selection, led to no significant population structuring between sample sites across the SWIO.

Table 4.8. Pairwise F_{ST} values between *O. cyanea* samples based on 7 microsatellite loci ROC1, OC18, ROC17, ROC28, OC32 and OC22 and ROC6. Statistically significant estimates ($p<0.05$) are highlighted in bold. Significance after Bonferroni correction is indicated by *

	MO	**KU**	**ZZ**	**PE**	**AN**	**BE**	**I**	**NR**	**SR**	**MA**	**MS**
MO	-										
KU	0.002	-									
ZZ	0.003	**0.004**	-								
PE	0.002	**0.006**	0.001								
AN	0.001	0.003	-0.001	**0.002**							
BE	0.003	**0.005**	0.000	0.001	-0.002	-					
I	0.000	0.001	**0.002**	**0.003**	**0.001**	0.000	-				
NR	0.001	0.002	0.001	0.003	0.001	0.000	0.000	-			
SR	-0.003	0.002	0.002	0.002	-0.002	0.000	**0.002**	-0.001	-		
MA	**0.007**	**0.009***	**0.008***	**0.006**	**0.006***	**0.008***	**0.005**	**0.008**	**0.008***	-	
MS	0.004	0.003	**0.004***	**0.002**	**0.004**	**0.006**	**0.002**	0.002	**0.007***	**0.007**	-
PS	0.006	0.001	**0.004**	**0.003**	0.002	**0.003**	**0.003**	0.000	-0.001	**0.009**	**0.003**

Table 4.9. Pairwise F_{ST} values between *O. cyanea* samples based on 6 microsatellite loci ROC1, OC18, ROC17, ROC28, OC32 and OC22.

	MO	**KU**	**ZZ**	**PE**	**AN**	**BE**	**I**	**NR**	**SR**	**MA**	**MS**
MO	-										
KU	0.002	-									
ZZ	0.003	0.003	-								
PE	0.002	0.006	0.000	-							
AN	0.001	0.003	0.000	0.002	-						
BE	0.004	0.005	0.000	0.001	-0.002	-					
I	-0.001	0.001	0.001	0.0032	0.001	0.000	-				
NR	0.002	0.002	0.001	0.002	0.001	0.001	0.001	-			
SR	-0.003	-0.001	0.001	-0.001	-0.003	0.000	0.001	-	-		
MA	0.001	0.005	0.002	0.002	0.002	0.003	-0.001	0.006	0.002	-	
MS	0.004	0.003	0.003	0.001	0.004	0.004	0.001	0.001	0.003	0.003	-
PS	0.006	0.001	0.004	0.0021	0.003	0.002	0.003	0.001	-0.002	0.004	0.000

Genic exact test of differentiation (Raymond and Rousset 1995a) using all 7 loci showed significant population structuring (Table 4.10) in 45 pairwise comparisons, 18 of which were significant after Bonferroni correction (Rice 1989). Mauritius was significantly differentiated from every other sample site. However, when genic tests of differentiation were rerun omitting locus ROC6, significant population structuring was observed in 32 pairwise comparisons, only 2 of which were significant after Bonferroni correction; Andavadoaka and Ivovona and Andavadoaka and Pemba (Table 4.11).

Table 4.10. Genic exact test of differentiation between *O. cyanea* sample sites based on 7 microsatellite loci. Statistically significant estimates ($p<0.05$) are highlighted in bold. Significance after Bonferroni correction is indicated by *.

	MO	KU	ZZ	PE	AN	BE	I	NR	SR	MA	MS
MO											
KU	0.572										
ZZ	**0.018**	**0.004**									
PE	0.051	**0.001 ***	**0.003**								
AN	**0.049**	**0.012**	0.131	**<0.001***							
BE	0.147	**0.002**	0.135	0.168	0.388						
I	0.068	**0.025**	**0.004**	**0.006**	**<0.001***	0.201					
NR	0.311	**0.090**	**0.054**	**0.012**	0.175	0.393	0.092				
SR	0.455	**0.001**	**<0.001***	**<0.001***	**0.074**	**0.019**	**<0.001***	0.430			
MA	**<0.001***	**<0.001 ***	**<0.001***	**<0.001***	**<0.001***	**<0.001 ***	**<0.001***	**<0.001***	**<0.001 ***		
MS	**0.014**	0.006	**0.002**	**0.003**	**<0.001***	**0.001**	**0.003**	0.064	**<0.001***	**<0.001**	
PS	0.092	0.182	**0.001**	**0.002**	**0.005**	**0.003**	**0.009**	0.295	0.411	**<0.001***	**0.009**

Table 4.11. Genic exact test of differentiation between *O. cyanea* sample sites calculated using 6 microsatellite loci ROC1, OC18, ROC17, ROC28, OC32 and OC22 with no type 1 error correction. Statistically significant estimates ($p<0.05$) are highlighted in bold. Significance after Bonferroni correction is indicated by *.

	MO	KU	ZZ	PE	AN	BE	I	NR	SR	MA	MS
MO											
KU	9.972										
ZZ	**0.013**	0.087									
PE	0.106	**0.002**	0.123								
AN	**0.035**	**0.009**	0.072	**<0.001***							
BE	0.009	**0.003**	0.207	0.165	0.522						
I	0.122	**0.040**	**0.015**	**0.009**	**<0.001***	0.194					
NR	0.327	0.123	0.167	0.055	0.156	0.443	0.104				
SR	0.389	**0.015**	**0.004**	**0.006**	0.170	**0.035**	**0.004**	0.477			
MA	0.330	0.123	**0.011**	0.052	**0.001**	0.076	**0.050**	**0.021**	**0.016**		
MS	**0.198**	0.068	**0.012**	**0.040**	**0.002**	**0.018**	0.0876	0.124	**0.002**	**0.019**	
PS	0.094	0.194	**0.018**	**0.013**	**0.005**	**0.005**	**0.025**	0.318	0.649	0.146	0.066

Both exact tests of genic differentiation and F_{ST} tests show very little population structuring in *O. cyanea* across the SWIO using 6 loci.

Bayesian clustering analysis in STRUCTURE (Pritchard, et al. 2003) using 7 loci, with locprior and no admixture including ROC6, the locus under selection, indicated two

genetic sub-populations as the most likely structure (K = 2, ΔK = 624, Figure 4.3) and seemed to indicate some structuring of populations across the SWIO. However STRUCTURE analysis using 6 loci, excluding ROC6, with locprior and no admixture indicated 3 genetic sub-populations using the Evanno method (K = 3, ΔK = 6.944, Figure 4.4). However, the plot of this run indicates no structure as all individuals from all populations belong to each cluster equally. The Evanno method is unable to estimate the probability of K = 1. Therefore the highest value of LnP(K) was used to estimate the number of populations. Highest LnP(K) was found at K = 1 (LnP(K) = -18041) suggesting that one population *O. cyanea* was present across the SWIO. The structure observed in Figure 4.3 is due to the presence of locus ROC6. Runs of STRUCTURE using 6 loci without locprior and with admixture did not reveal any significant structuring across the SWIO.

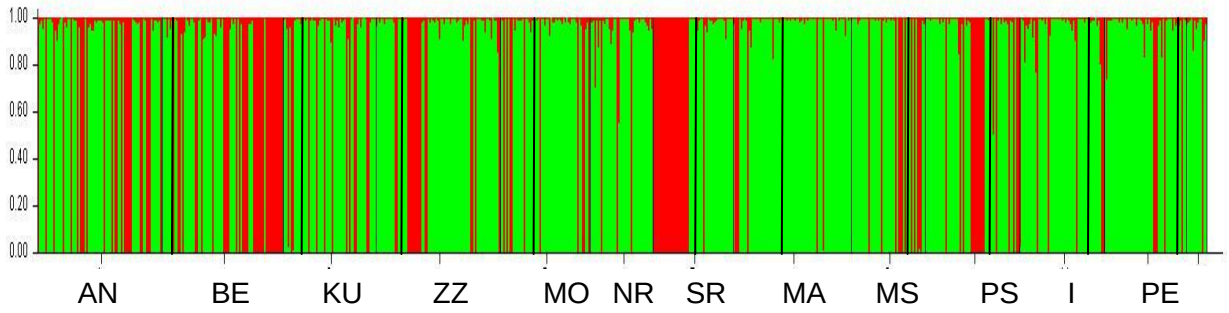

Figure 4.3. Bayesian analysis of individual multi-locus genotype clustering (STRUCTURE) cryptic population structure, and admixture levels using 7 loci, including ROC6, for *O. cyanea* across the Southwestern Indian Ocean.

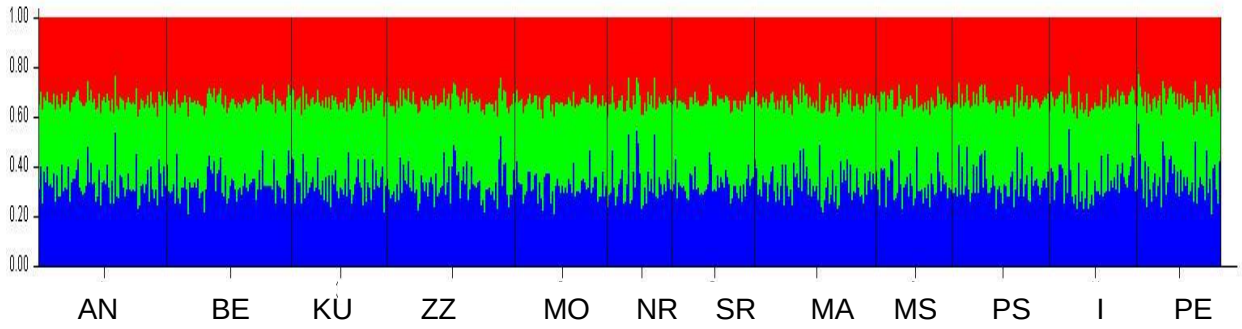

Figure 4.4. Bayesian analysis of individual multi-locus genotype clustering (STRUCTURE) cryptic population structure, and admixture levels using 6 loci, excluding ROC6, for *O. cyanea* across the Southwestern Indian Ocean.

Testing of isolation by distance indicated non-significant correlation between genetic distance and geographic distance, or between genetic distance and oceanographic distance (r = 0.01, p = 0.49 and r = 0.01, p = 0.52 respectively).

Demographic history

To test the theory of a population bottleneck the M ratio was calculated in M_P_VAL (Garza and Williamson 2001) and heterozygosity tests were conducted in BOTTLENECK (Piry, et al. 1999). Values for M_P_VAL (Garza and Williamson 2001) were as follows; effective population size (N_e) was estimated to be 5000 pre-bottleneck. Since the absolute rate of microsatellite mutation in our study species is unknown θ was calculated using a range of mutation rates 1×10^{-4}, 3×10^{-4} and 5×10^{-4}, which are typical of mutation rates of microsatellites for fishes, which are of the order of 10^{-4} mutations per microsatellite locus per generation (Angers and Bernatchez 1998), corresponding to a θ of 2, 6 and 10 respectively. The calculated M ratio was 0.877 (Table 4.12)

Table 4.12. M-ratio values for varying values of theta and Critical M ratio.

Theta (θ)	**M ratio**	**Critical M**
10	0.877 (66.1% of the time you expect a smaller ratio at equilibrium)	0.72
6	0.877 (59.3% of the time you expect a smaller ratio at equilibrium	0.68
2	0.877 (82.8% of the time you expect a smaller ratio at equilibrium)	0.74

Both heterozygosity tests and M-ratio tests reject the hypothesis of a population bottleneck (Tables 4.12, 4.13).

Table 4.13. Demographic history of *O. cyanea* across the SWIO: probability (p-values) of a bottleneck in populations under various mutational models.

Location	**IAM**	**TPM**	**SMM**
MO	0.148	0.148	0.656
KU	0.148	0.289	1.00
ZZ	0.148	0.188	0.711
PE	0.148	0.469	0.852
AN	0.148	0.148	0.406
BE	0.167	0.149	0.619
I	0.148	0.406	0.711
NR	0.148	0.188	0.594
SR	0.148	0.289	0.813
MA	0.148	0.188	0.973
MS	0.148	0.344	0.766
PS	0.188	0.469	0.711

4.3.2 **Mitochondrial DNA analyses**

Population diversity

One hundred and twenty four individuals were sequenced for 658bp of COI, resulting in 3 haplotypes, one common to all sampling sites and two private haplotypes. Due to the very low diversity in this dataset, COI was uninformative as a marker to investigate population structuring. Sample sizes were therefore not increased and the dataset was not used in any further analyses.

The same one hundred and twenty four individuals were sequenced for 480bp of 16S rDNA, resulting in 7 haplotypes. Of the mutations observed 5 were substitutions, all of which were transitions. Haplotype and nucleotide diversity were very low (0.20 - 0.56 and 0.0004 - 0.0021) respectively. Of the three high frequency haplotypes observed, two were found at all sites, and the third was found in Tanzania, Madagascar and the Seychelles (Figure 4.5). Due to the uninformative nature of this marker, no further analysis of this dataset was carried out and sample size was not increased.

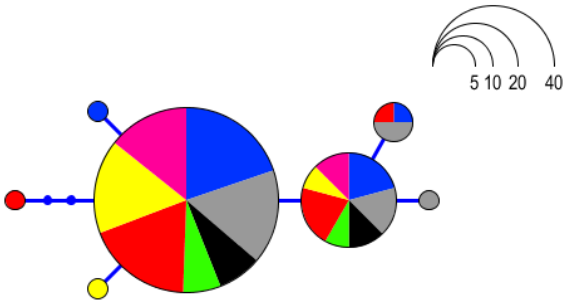

Figure 4.5. Haplotype network of *O. cyanea* based on 480bp of 16S rDNA mtDNA. Branch lengths are proportional to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation. dark blue circles represent missing haplotypes. Tanzania, Madagascar, Kenya, Mozambique, Rodrigues, Seychelles, Mauritius.

Two hundred and six individuals were sequenced for 480bp of NCR resulting in 16 haplotypes. Nineteen polymorphic sites were identified. Mutations consisted of 15 transitions, 1 transversion and 3 insertion/deletions. No significant structuring was observed between samples within countries and data were subsequently pooled by country to increase sample size.

Table 4.14. Mitochondrial genetic diversity statistics for NCR sequences of *O. cyanea*, pooled per country. Sample size (n), number of haplotypes (H), haplotype richness after rarefaction to 18 individuals (H'), haplotype diversity (h), nucleotide diversity (π).

Country	n	H	H'	h	Π
Kenya	19	3	3.0	0.57	0.0012
Tanzania	36	6	4.5	0.67	0.0019
Mozambique	18	4	4.0	0.67	0.0018
Madagascar	42	7	4	0.62	0.0017
Rodrigues	27	4	3.3	0.59	0.0014
Mauritius	26	5	4.0	0.41	0.0012
Seychelles	29	7	4.6	0.66	0.0025
Average	206	-	3.91	0.599	0.0016

NCR haplotype diversity ranged from 0.41 – 0.79. Nucleotide diversity was low π = 0.0012 – 0.0025 (Table 4.14). The Seychelles had the highest nucleotide diversity of any site, over twice the diversity of Mauritius and Kenya. The Seychelles also had the highest haplotype diversity after rarefaction.

The haplotype network revealed two common haplotypes, haplotypes 3 and 4 which were present at all sample sites across the SWIO. (Figure 4.6, Figure 4.7). Haplotype 10 was present at all sites with the exception of Mozambique and Rodrigues. Haplotype 11 was present at Madagascar, Rodrigues, Tanzania and Mauritius. The proportion of common haplotypes at sample sites varied; at Madagascar, Tanzania, Kenya, Praslin and Mauritius, haplotype 3 was more common than haplotype 4, whilst in the Mozambique and Mahe samples, haplotype 4 is more common. This was particularly pronounced in Mauritius, where 20 individuals possessed haplotype 3 but only 3 individuals possessed haplotype 4.

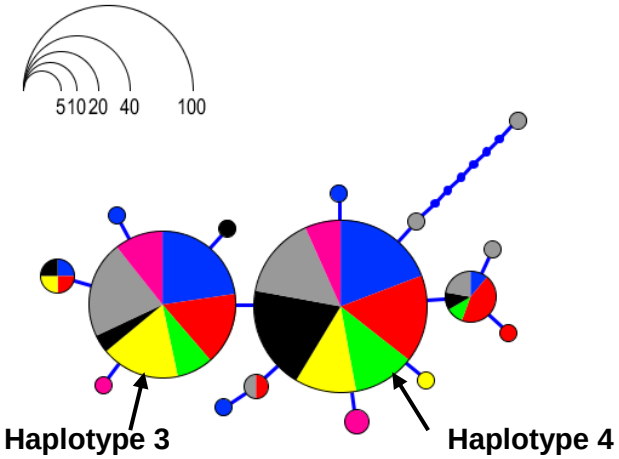

Figure 4.6. Haplotype network of *O. cyanea* based on 480bp of NCR mtDNA. Branch lengths are proportional to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation. Tanzania, Madagascar, Kenya, Mozambique, Rodrigues, Seychelles, Mauritius.

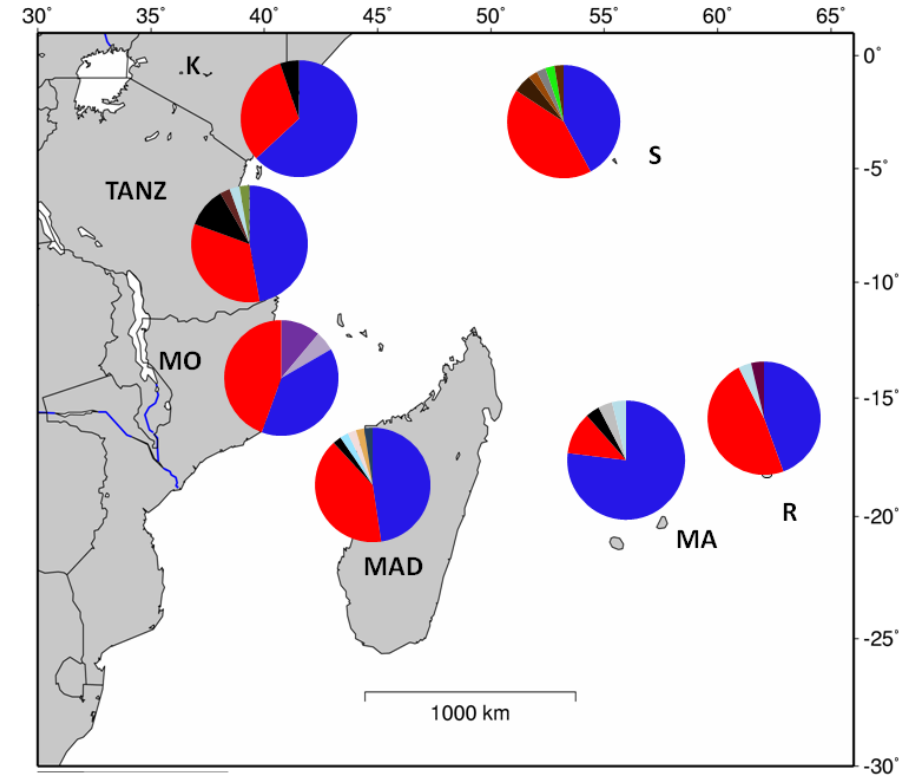

Figure 4.7. Map of SWIO with *O. cyanea* NCR haplotype frequencies by country. K – Kenya, TANZ – Tanzania, MO – Mozambique, MAD – Madagascar, MA – Mauritius, R – Rodrigues, S – Seychelles. Haplotypes: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16.

Population Structure

POWSIM results indicated that the NCR dataset had a 9% and 6% chance, based on Chi^2 and Fisher tests respectively, of detecting an F_{ST} of 0.01. At an F_{ST} of 0.05 the NCR dataset had a 89% and 86% chance of detecting population structuring (Table 4.15).

Table 4.15. Statistical power of *O. cyanea* NCR mtDNA samples to detect population differentiation at various levels of F_{ST} using tests based on Chi^2 and Fisher exact methods.

	Statistical Power	
F_{ST}	**Chi^2**	**Fisher**
0.001	0.09	0.06
0.01	0.27	0.18
0.05	0.89	0.86

F_{ST} and exact tests indicated the Mauritian population of *O. cyanea* was differentiated from Mozambique, Madagascar, the Seychelles and Rodrigues but was not differentiated from Tanzania and Kenya (Tables 4.16, 4.17). These results became non-significant after Bonferroni correction. POWSIM estimation of a type 1 error was 4%.

Table 4.16. Pairwise F_{ST} values based on *O. cyanea* NCR mtDNA samples. Statistically significant estimates ($p<0.05$) are highlighted in bold.

	MA	**K**	**TANZ**	**MO**	**MAD**	**S**
K	-0.013	-				
TANZ	0.022	-0.026	-			
MO	**0.110**	0.020	0.022	-		
MAD	**0.066**	-0.007	0.003	-0.014	-	
S	**0.051**	-0.011	-0.008	-0.009	-0.012	-
R	**0.125**	0.029	0.026	-0.022	-0.019	-0.007

Table 4.17. Exact test of differentiation values using *O. cyanea* mtDNA NCR samples. Statistically significant estimates ($p<0.05$) are highlighted in bold.

	MA	**K**	**TANZ**	**MO**	**MAD**	**S**
K	0.322 ±0.005	-				
TANZ	0.076 ±0.006	0.935 ±0.002	-			
MO	**0.006 ±0.001**	0.188 ±0.006	0.155 ±0.003	-		
MAD	**0.032 ±0.005**	0.878 ±0.003	0.502 ±0.009	0.352 ±0.013	-	
S	**0.012 ±0.004**	0.934 ±0.003	0.779 ±0.008	0.403 ±0.016	0.774 ±0.008	-
R	**0.010 ±0.002**	0.415 ±0.005	0.287 ±0.006	0.349 ±0.010	0.944 ±0.003	0.752 ±0.009

Demographic History

Significant departures from neutrality were observed in at least one test for all island populations of *O. cyanea* (Table 4.18). Tajima's D statistic indicated a significant departure from neutrality for Seychelles and Rodrigues, while Ewens-Watterson test was significant for Mauritius.

Table 4.18. Mitochondrial neutrality tests for *O. cyanea* NCR dataset, samples pooled per country. Sample size (n), number of haplotypes, Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs). Statistically significant estimates ($p<0.05$) are highlighted in bold.

Country	n	F	D	Fs
Madagascar	42	**0.39**	**-2.09**	**-3.06**
Tanzania	36	0.35	-1.40	-1.77
Kenya	19	0.50	-1.16	-0.03
Mozambique	18	0.36	-0.53	-0.49
Seychelles	29	0.36	**-1.95**	-1.80
Rodrigues	27	0.43	**-1.51**	-0.65
Mauritius	26	**0.61**	**-1.73**	**-2.37**

To investigate whether departures from neutrality had been limited to island populations, samples were pooled into two groups; (i) African mainland – Tanzania, Kenya and Mozambique and (ii) surrounding islands (Table 4.19). Both Zanzibar and Vamizi Island were included as African mainland because although they are islands, they are extremely close to the African mainland (Vamizi Island 5 km, Zanzibar 40 km) and samples from Zanzibar were likely to have been fished off the African coast. Both African mainland and surrounding islands showed significant deviation from neutral expectations (Table 4.19).

Mismatch distribution analysis of NCR dataset by country indicated no significant departure from the hypothesis of a population expansion (Tables 4.20, Figure 4.8).

Time since expansion was calculated using the equation $T = 2ut$, where T is the time in mutational units, u is the mutation rate and t is generation time. Mutation rates of 1.17×10^{-7}, 1.2×10^{-8} and 1.8×10^{-8} per site per year were used to estimate time since expansion. Mismatch calculations for Praslin and Mahe were calculated separately, as evidence for the flooding of the Seychelles Bank is from cores taken near Mahe, not Praslin. Praslin is on the edge of the Seychelles Bank and it is possible that as the sea

level rose, the Seychelles Bank off Praslin was flooded prior to the bank near Mahe and therefore colonised at an earlier date by *O. cyanea*.

Table 4.19. Mitochondrial genetic diversity levels and neutrality tests for *O cyanea*, samples pooled per country NCR. Sample size (n), number of haplotypes (H), haplotype diversity (h), nucleotide diversity (π), Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs). Statistically significant estimates ($p<0.05$) are highlighted in bold.

Geographic Area	n	H	h	π	F	D	Fs
African mainland	62	8	0.65	0.0018	0.36	**-1.64**	**-3.42**
Islands off African mainland	107	12	0.62	0.0019	**0.39**	**-2.33**	**-7.04**

Table 4.20. Mismatch analysis of *O. cyanea* NCR mtDNA by country. T time since expansion in mutational units with 95% confidence intervals in brackets, θ_0 population size before expansion θ_1 population size after expansion, SSD sum of squares deviation, TSE1 time since expansion in thousands of years using a mutation rate of 1.17×10^{-7} per site per year with 95% confidence intervals in brackets, TSE2 time since expansion in thousands of years using a mutation rate of 1.8×10^{-8} per site per year with 95% confidence intervals in brackets, TSE3 time since expansion in thousands of years using a mutation rate of 1.2×10^{-8} per site per year with 95% confidence intervals in brackets.

Location	T	θ_0	θ_1	SSD	TSE1 (KA)	TSE2 (KA)	TSE3 (KA)
Madagascar	0.9 (0.44-1.60)	0.00	99999.00	0.017	8.0 (3.9-14.2)	52 (25.5-96.6)	78 (38.2-138.9)
Tanzania	1.0 (0.39-1.88)	0.00	99999.00	0.017	8.9 (3.5-16.7)	59 (22.6-108.8)	87 (33.9-163.2)
Kenya	0.7 (0.00-1.53)	0.00	99999.00	0.028	6.2 (0.0-13.6)	40 (0.0-88.5)	61 (0.0-132.8)
Mozambique	1.0 (0.12-2.03)	0.00	99999.00	0.025	8.9 (1.0-18.0)	59.0 (6.9-117.5)	87 (10.4-176.2)
Mahe	0.9 (0.44-1.66)	0.00	99999.00	0.023	8.0 (3.9-14.8)	52 (25.5-96.0)	78 (38.2-144.1)
Praslin	1.1 (0.48-1.70)	0.00	99999.00	0.026	9.8 (4.3-15.1)	64 (27.7-98.4)	95 (41.7-147.6)
Rodrigues	0.8 (0.27-1.66)	0.00	99999.00	0.032	7.1 (2.4-14.8)	46 (15.6-96.0)	69 (23.4-144.1)
Mauritius	0.5 (0.01-1.06)	0.00	99999.00	0.005	4.5 (0.08-9.4)	29 (0.6-61.3)	43 (0.87-92.0)

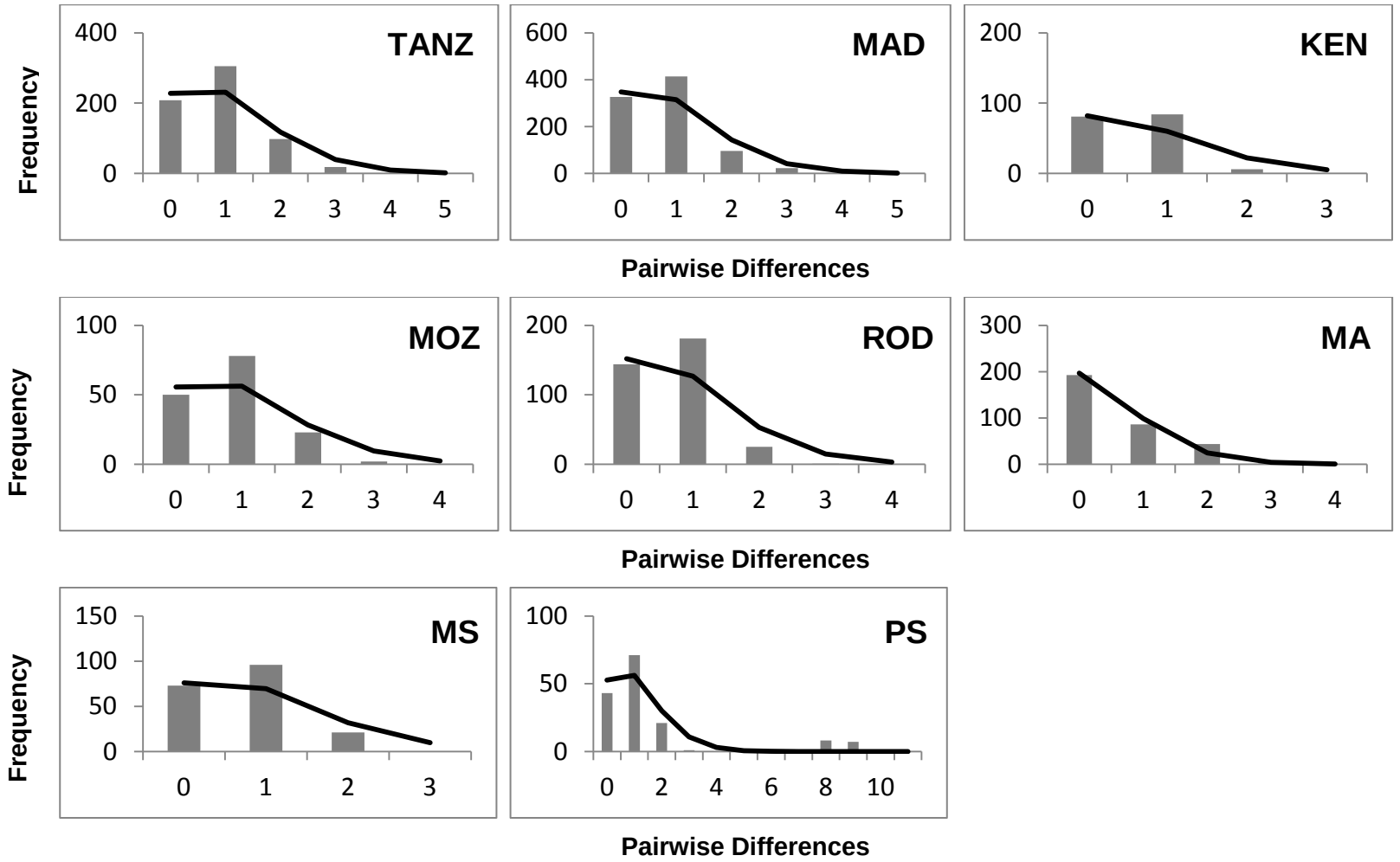

Figure 4.8. Mismatch distribution graphs for countries of the SWIO based on *O. cyanea* 480bp NCR mtDNA sequences. Filled bars indicate the observed frequency of pairwise distribution, black lines indicate the expected distribution under a model of sudden demographic expansion. TANZ - Tanzania, MAD-Madagascar , KEN – Kenya, MOZ – Mozambique, ROD - Rodrigues, MA - Mauritius, MS - Mahe, PS – Praslin.

Mismatch analysis from NCR data (Figure 4.8) shows a unimodal distribution and the lack of significant deviation from the model of population expansion. The Mauritian population appears to have undergone a more recent population expansion than the rest of the SWIO.

Bayesian skyline plots (BSP) were constructed using the NCR dataset to further explore past population sizes in the SWIO. However the confidence intervals of the BSP were extremely large and consequently were not suitable for inclusion in this thesis.

4.4 **Discussion**

Genetic diversity

Mitochondrial genetic diversity was extremely low for COI (*h* = 0.00 - 0.25, π= 0.00 - 0.0005) and 16S rDNA (h = 0.20 - 0.61, π = 0.0005 - 0.0022). Haplotype and nucleotide diversity for NCR were (h = 0.41 - 0.67, π = 0.0012 - 0.0025) considerably lower than those seen in other octopus species; *O. variabilis*, h = 0.41 - 0.79, π = 0.0011 - 0.0031 (Lü, et al. 2013) and *Pareldone turqueti*, π = 0.000 – 0.0053 (Strugnell, et al. 2012). Similarly low haplotype and nucleotide diversity has however been observed in the giant squid *Architeuthis* spp. where average NCR region (control region) π = 0.00166, h = 0.613 (Winkelmann, et al. 2013)

The low diversity of the mitochondrial data contrasts sharply with microsatellite genetic diversity (H_o = 0.25 - 0.96). When loci containing null alleles were excluded, microsatellite genetic diversity was considerably higher (H_o = 0.74 - 0.96), similar in range to those reported for other octopus species; *O. maya* H_o = 0.31 - 0.80 (Juárez, et al. 2010), *P. turqueti*, H_e = 0.77 - 0.88 (H_o not reported) (Strugnell, et al. 2012), *O. vulgaris* H_o = 0.484 – 0.914 (Cabranes, et al. 2008) and *O. maorum* H_o = 0.62 – 0.92 (Doubleday, et al. 2009).

The contrast in genetic diversity between the microsatellite and mitochondrial data may be partly explained by the fact that microsatellites have a higher mutation rate than mtDNA and a four times larger effective population size than that of mtDNA. Additionally, it is likely that a process has occurred whereby mitochondrial diversity has been drastically reduced, either by a selective sweep or by a population bottleneck followed by a population expansion. Research on the giant squid *Architeuthis* spp. recently revealed comparable levels of haplotype and nucleotide diversity in the NCR region most likely caused by either a selective sweep or population bottleneck followed by an expansion (Winkelmann, et al. 2013).

Potential selection on locus ROC6

A salient feature of the nuclear data was the differentiation reported for the Mauritius sample that was caused by a single locus (ROC6). ROC6 was identified as a positive outlier in all comparisons involving Mauritius using the program LOSITAN (Antao, et al. 2008), suggesting that this locus is under selection. An increasing number of population genetics studies have identified microsatellites which were assumed to be neutral to actually be under the effects of selection (Nielsen, et al. 2006; Nielsen, et al. 2009; Rhode, et al. 2013).

Microsatellites are a popular choice in population genetic studies as they are generally non-coding DNA and are assumed to be a neutral marker. However, microsatellites are commonly used for linkage mapping of important quantitative trait loci (QTLs) under selection in crops, livestock and model organisms (Slate 2005) meaning that neutrality cannot be automatically assumed for natural populations. “Genetic Hitchhiking”, (Smith and Haigh 1974) the mechanism by which a microsatellite is selectively favoured due to its being linked to a gene under selection is fairly common in natural populations (André, et al. 2011; Nielsen, et al. 2006; Nielsen, et al. 2009; O'Malley, et al. 2007). In this case, the average heterozygosity of the microsatellite loci will be reduced, the extent of which varies depending on the distance from the locus under selection (Smith and Haigh 1974). Additionally, recent studies have identified functional microsatellites that affect the fitness of an individual (Gemayel, et al. 2010; Kashi and King 2006). Putatively functional microsatellites are primarily located in or near genic regions, where a change in the number of times the motif is repeated is thought to modify gene expression or change protein sequence (Gemayel, et al. 2010; Haasl and Payseur 2013).

ROC6 may be linked to a specific gene under positive selection in the Mauritian population of *O. cyanea.* Microsatellites under selection have been used to detect divergent selection in marine fish species (André, et al. 2011). Many marine fish species show little population structuring because populations are large, evolutionarily young and have high gene flow (André, et al. 2011). However, recent studies have shown that organisms can undergo adaptive phenotypic evolution over a few generations (Carroll, et al. 2007). Loci under selection are therefore valuable as population markers on these ecological timescales where isolated populations have not yet diverged at neutral loci (André, et al. 2011; Nielsen, et al. 2006; Nielsen, et al. 2009; O'Malley, et al. 2007) which may be the case here.

Population structuring and gene flow

Interpreting patterns of non-neutral genetic differentiation in terms of population connectivity is challenging. Does the divergence of Mauritius at ROC6 reflect (a) differential selection of genotypes from a panmictic gene flow (i.e. no breakdown in connectivity) or (b) population isolation not detected at the other microsatellite loci? NCR mtDNA tests of F_{ST} showed population structuring between Mauritius and Mozambique, Madagascar, Seychelles and Rodrigues which were non-significant after Bonferroni correction. However Bonferroni correction is often considered too conservative and may result in an increase in type 2 error (Cabin and Mitchell 2000; Moran 2003; Nakagawa 2004; Narum 2006; Ryman and Jorde 2001). It is therefore possible that the mtDNA detected some structuring between Mauritius and other sampling locations in the SWIO.

POWSIM results of the mtDNA dataset indicated that at low levels of F_{ST} (0.001), the power to detect population differentiation was limited, 9% and 6% using Chi^2 and Fisher exact tests respectively. The microsatellite dataset however had high statistical power, at an F_{ST} of 0.001 the microsatellite dataset had a 90% and 85% chance of detecting structuring, using Chi^2 and Fisher exact tests respectively and at an F_{ST} of 0.01 a 100% probability of detecting any structuring. Due to the high statistical power of the microsatellite dataset it therefore seems unlikely that failure to detect population structure at other loci may be due to insufficient statistical power.

The combination of some limited structuring in the mtDNA dataset and the high statistical power of the microsatellite dataset suggest that the selection acting upon ROC6 represents some isolation between Mauritius and other locations in the SWIO that is not currently detected in neutral loci, perhaps due to mutation drift disequilibrium. Mutation drift disequilibrium may mean that the population isolation of Mauritius is not detected at neutral loci due to the retention of ancestral patterns of connectivity at these loci. These retained historical patterns of connectivity may overestimate current connectivity of populations. There is some support for this theory in the dataset as mtDNA demographic tests indicated evidence of non-equilibrium of populations through significant deviations from neutrality. Mismatch analysis also indicated a more recent population expansion for Mauritius, ranging from 5 - 43 KA, compared to 7 - 95 KA at other sampling locations. If high levels of gene flow between other locations and Mauritius were occurring and the same processes were occurring in Mauritius and the rest of SWIO, then it could be expected that population expansion at Mauritius would have occurred at the same time as the rest of the SWIO. This

therefore may be a further indication that Mauritius is at least partially isolated from the rest of the SWIO.

The mtDNA demographic tests reported evidence of non-equilibrium populations, indicating equilibrium of populations has not yet been reached. The rate of approach to equilibrium is determined by the inverse of the effective population size or migration rate - whichever is greatest (Slatkin 1994b). Due to its smaller effective population size, mtDNA might be expected to attain equilibrium sooner than diploid nuclear markers, which may explain why some limited structuring was indicated between Mauritius and other sampling locations in the SWIO which was not observed in the microsatellite dataset. Additionally, microsatellites may experience greater levels of homoplasy (Angers and Bernatchez 1998; Angers, et al. 1999). Size homoplasy may frequently occur at microsatellite loci and underestimate the real diversity which may limit their ability to detect population structuring.

To date only one previous study has found any evidence of population structuring of Mauritian populations. Muths, et al. (2012) discovered some evidence of local population structuring in snapper *Lutjanus kasmira* between Mauritius and the Mozambique Channel using mtDNA. They suggest that this is due to local processes rather than permanent geographic isolation. During this thesis, two other fish species, *L. harak* and *L. nebulosus* indicate structuring between Mauritius and other locations in the SWIO (which is further discussed in Chapter 8). It may be, therefore, that there is some local process in Mauritius, perhaps an undetected oceanographic feature for example, which is causing at least partial isolation of the Mauritian population from the rest of the SWIO. As more studies emerge on the SWIO, a currently understudied area, more species may exhibit this pattern of emerging isolation (Ridgway and Sampayo 2005).

Excluding the partial isolation of Mauritius, the microsatellite and mtDNA datasets reveal very little significant population structuring of *O. cyanea* across the rest of the SWIO, consistent with the initial hypothesis of little population structuring in *O. cyanea* across the SWIO. Microsatellite and mtDNA F_{ST} tests were non-significant after Bonferroni correction and microsatellite STRUCTURE runs using 6 loci indicated no population structuring. Genic exact tests using 6 microsatellite loci, however, indicated significant structuring between Andavadoaka and Pemba and Andavadoaka and Ivovona after Bonferroni correction ($\chi^2 = 33.95$, $p = 0.00069$, $\chi^2 = 36.09$, $p = 0.00031$ respectively). Genic (allelic) exact tests have been shown to be more powerful as estimators of population differentiation than F_{ST} when sampling size is unbalanced (Goudet, et al. 1996) However, two loci of the dataset contain null alleles which have

been shown to inflate measures of population differentiation where significant population structuring exists (Balloux and Lugon-Moulin 2002; Chapuis and Estoup 2007). Genic exact tests were therefore repeated excluding loci OC18 and OC22 which have been identified as containing null alleles and ROC6 which shows selection, another factor which may inflate estimates of population structure. The resulting genic exact test indicated no structuring between Andavadoaka and Pemba and Andavadoaka and Ivovona previously seen when OC18, OC22 and ROC6 were included in the analysis but subsequently displayed structuring between Andavadoaka and Mauritius and Andavadoaka and Mahe after Bonferroni correction (χ^2= 29.62, p = 0.00024, χ^2= 27.10, p = 0.00068 respectively). This differentiation was not detected by other population measures and it is therefore likely that if there is structuring between these sites it is minimal.

The utilisation of a microsatellite locus which is under positive selection can considerably inflate population structuring (Nielsen, et al. 2006). This is clearly visible in the microsatellite F_{ST} values, exact tests and STRUCTURE results when comparing 7 loci including ROC6 and 6 loci omitting ROC6. Due to a rigorous approach and testing of neutrality in this study, ROC6 was identified as being under positive selection and this was taken into consideration when investigating the population structuring of *O. cyanea* across the rest of the SWIO excluding Mauritius.

The panmixia of populations of *O. cyanea* across the remaining sites in the SWIO represents sufficient gene flow between populations to maintain homogenous populations across an area of over 9000 km^2. This is possibly the result of the combination of the considerable pelagic larval duration of *O. cyanea* and oceanographic currents of the region which, when combined, provide a significant dispersal potential and consequently gene flow, for *O. cyanea* larvae. These results are in keeping with a number of studies which have demonstrated high gene flow across the SWIO (Dorenbosch, et al. 2006; Muths, et al. 2013; Visram, et al. 2010).

Alternatively, since populations are not at migration drift equilibrium, indicated by significant departures from neutrality in the mtDNA, and there is evidence of isolation of Mauritius, an alternative explanation may be that neutral loci are overestimating the current connectivity across the SWIO due to the retainment of historical patterns of connectivity. In view of this, future genetic studies of population structure in *O. cyanea* may benefit from the analysis of Type 1 markers (i.e. markers under selection) as neutral markers may not have attained equilibrium and therefore may be overestimating current connectivity of populations of *O. cyanea* across the SWIO.

Past demographic history

In the marine environment, mechanisms shaping population structure may be contemporary (Nielsen, et al. 2009) and/or historical, such as large-scale climate-induced changes (Grant and Bowen 1998; Janko, et al. 2007; Marko 2004). An initial pilot study of mitochondrial DNA of *O. cyanea* revealed strikingly low haplotype diversity in the cytochrome oxidase 1 region (COI). Low levels of mtDNA diversity may be suggestive of selection or of a past population bottleneck perhaps caused by climate variability.

The mtDNA neutrality tests indicated a significant departure from neutrality which may be due to a population expansion or selective sweep acting on the mitochondrial genome. No bottleneck (on a scale of 1-500 generations) was detected in the dataset. However, the “star-like” shape of the haplotype networks for COI, 16S rRNA and NCR are suggestive of a population bottleneck followed by a population expansion. It is therefore possible that a population bottleneck has occurred, which is not detectable using the current dataset, followed by a population expansion. However, the power to resolve past population changes is severely limited by lack of variation in mtDNA of *O. cyanea* and so neither a population bottleneck nor a mitochondrial selective sweep can be ruled out as a driver of the low mitochondrial diversity observed in the dataset.

Mismatch analysis and neutrality tests show evidence of a population expansion. The dating of this expansion is difficult to estimate as suitable mutation rates for octopus NCR are extremely limited. The selection of three reasonable mutation rates resulted in population expansion estimates ranging from 5 - 95KA. However, possibly due to the low variability of the mtDNA NCR dataset there are large 95% confidence intervals around the Tau values and the resulting estimates of population expansion (Table 4.20). It may therefore be that there is insufficient variability in the dataset for accurate calculation of the mismatch distribution.

Assuming that these mutation rates are reasonable, it is possible that *O. cyanea* populations have undergone a population expansion within the last 177,000 years (Table 4.20). Estimates of population expansion using the least conservative mutation rate and taking account of confidence intervals places the most recent population expansion within the last 20,000 years, coinciding with the end of the Last Glacial Maximum (LGM). Using the two more conservative mutation rates, estimates of population expansion range from 29 - 95KA, although when confidence intervals are taken into account, the date of population expansion ranges from present day to 177 KA.

A number of phylogeographic studies on a range of species in the Indian Ocean have shown evidence of recent population expansion (Fratini, et al. 2010; Gopal, et al. 2006; Neethling, et al. 2008; Tolley, et al. 2005) over a large temporal range. Spiny lobster *Palinurus gilchristi* underwent a population expansion 5.3 – 10.6 KA (Tolley, et al. 2005), spiny lobster *Palinurus delagoae* underwent a population expansion 30 - 40 KA, or 9 - 13 KA dependent on mutation rate (Gopal, et al. 2006), *Scylla serrata* underwent a population expansion 1.0 - 1.5 million years ago (MYA) (Fratini, et al. 2010) and reef fish *Myripristis berndti* underwent a population expansion 300 KA – 1 MYA (Craig, et al. 2007).

In the studies involving *P. gilchristi* (Tolley, et al. 2005), the date of population expansion was thought to be linked to both the increase in available habitat after the flooding of the continental shelf as sea levels rose and increased sea temperature post-LGM (Gopal, et al. 2006; Tolley, et al. 2005). If the mutation rate of 1.17×10^{-7} per site per year was confirmed by further research to be the most suitable mutation rate for *O. cyanea* mtDNA NCR, then it may be that the population expansion of *O. cyanea* coincided with the increasing sea level and sea temperature post-LGM. During the LGM, sea levels in the SWIO were up to 120 m lower compared to present day (Camoin, et al. 1997; Camoin, et al. 2004) and a decrease in sea temperature of approximately 2°C occurred in the outer areas of *O. cyanea's* current range (Prell, et al. 1980). Following the LGM, *O. cyanea* populations may have expanded their range by recent colonisation, creating largely homogenous populations of *O. cyanea* across the SWIO that all still retain the gene frequencies present in the ancestral pre-expansion population.

Interestingly, mismatch analysis suggests that the population expansion of Mauritius occurred more recently than population expansion in the rest of the SWIO (Table 4.20). The reasons for this difference in expansion date is unclear and although an extensive literature review has been undertaken, no particular reason for this difference in expansion times can be found. However, the differences in population expansion estimates may be partly due to the isolation of the Mauritian population from the rest of the SWIO. It may be that in other areas of the SWIO, population expansion occurred in a number of locations where conditions were favoured to *O. cyanea* populations, perhaps due to increasingly available habitat for example. As these populations increased in size, increased gene flow from these populations may have increased population sizes across the rest of the SWIO. However, if as suggested in this study, Mauritius is isolated from the rest of the SWIO, the Mauritian population would not benefit from the immigration of individuals. Therefore it may have taken Mauritian population longer to increase its population size. Alternatively, perhaps local conditions in Mauritius did not become more favourable to populations of *O. cyanea* until

significantly later than elsewhere in the SWIO, perhaps accounting for why a more recent population expansion was observed in Mauritius.

Implications for fisheries of *O. cyanea*

Effective fisheries management and conservation are limited by the lack of knowledge of demographic history, gene flow and adaption to environmental factors. The results of this study indicate that the current fisheries management system, which assumes *O. cyanea* to be a single panmictic stock, is inaccurate. The positive selection observed in ROC6 at Mauritius suggests that there is local adaption occurring in these populations. This local adaption may be important to the survival of *O. cyanea* in Mauritius and care should therefore be taken to maintain the genetic diversity of the Mauritian *O. cyanea* population.

The effect of fishing on *O. cyanea* genetic diversity is difficult to estimate as there are no measurements of genetic diversity of nuclear and mitochondrial markers prior to fishing. However, the high genetic diversity of the microsatellites may suggest that to date fishing has not overly impacted population diversity, despite reports of a significant decrease in catch in *O. cyanea* in Mauritius (FAO 2013).

The apparent lack of decreased genetic diversity in response to fishing in *O. cyanea* is perhaps due to the high gene flow between populations in the SWIO (excluding Mauritius) resulting in the replenishment of locally fished populations by continual immigration from under-exploited areas of the SWIO. These under-exploited areas are in effect acting as unofficial marine protected areas (MPAs), which have been shown to assist in maintaining local genetic diversity in areas surrounding the MPA (Christie, et al. 2010; Harrison, et al. 2012). Alternatively, it may be that since *O. cyanea* populations are not at migration drift equilibrium, genetic diversity of *O. cyanea* may not actually reflect contemporary diversity. This may be particularly relevant to fisheries of *O. cyanea* in Mauritius. Mauritian catch statistics of *O. cyanea* show a sharp decrease in catch from 335 tons in 2002 to 84 tons in 2006, with catches remaining low from 2006 up to the latest FAO catch data in 2011. The significant decrease in catch of *O. cyanea* in Mauritius suggests that populations in Mauritius are being overfished (FAO 2013), despite no significantly reduced genetic diversity at Mauritius being observed in comparison to other sampling locations across the SWIO. Additionally, the lack of overlapping generations in *O. cyanea*, meaning that if one generation is heavily exploited then recruitment to the next generation will be severely reduced, combined with the isolation of Mauritius, which may limit replenishment of populations by migration from other areas of the SWIO, may have led to the severe decline of *O. cyanea* stocks in Mauritius. We suggest however, that this considerable decrease in

stocks is not yet detectable in neutral markers utilised in this study due to migration drift disequilibrium which has led to the retainment of ancestral connectivity and diversity levels.

Chapter 5: Population structuring and phylogeography of *Lethrinus harak* across the SWIO

5.1 Introduction

Historical processes associated with past climatic events have been shown to dramatically influence patterns of species distribution, diversity, and speciation (Fraser, et al. 2009; Janko, et al. 2007; Koizumi, et al. 2012; Palero, et al. 2008). The Pleistocene climate fluctuations of the last two million years dramatically affected distribution and diversity of terrestrial and aquatic organisms through habitat alteration (Avise and Walker 1998; Carstens and Knowles 2007; Hewitt 1999; Hoareau, et al. 2012). Pleistocene climate fluctuations affected population diversity, structure and speciation of both temperate (Avise and Walker 1998; Carstens and Knowles 2007; Hewitt 1999) and tropical species (Gaither, et al. 2011; Gaither, et al. 2010; Hoareau, et al. 2012). For species in temperate regions, expansions and contractions of the ice sheets rendered large areas uninhabitable. As a result many species survived glacial maxima by retreating to refugia at lower latitudes, profoundly affecting population diversity and structure (Provan and Bennett 2008). Ice sheets did not extend to tropical regions, such as the Southwestern Indian Ocean. However other effects such as changes in sea surface temperature, sea level and oceanographic features, caused by climate oscillations, may have been causal factors in the structuring of tropical marine populations during the Pleistocene (Hoareau, et al. 2012).

Lethrinus harak is a hermaphroditic emperor fish with a wide distribution across the Indian Ocean and West Pacific from the Red Sea to Fiji and Samoa (Carpenter and Allen 1989). The life cycle of *L. harak* is divided into a pelagic larval phase (for 29 days) (Nakamura, et al. 2010) and an adult life which is spent inshore in shallow coral rubble, mangroves, lagoons and adjacent to coral reefs (Carpenter and Allen 1989). The maximum age for this fish is 15 years (Carpenter and Allen 1989). The wide distribution of *L. harak* across the Indian Ocean and large population size make it a good model to investigate the link between evolutionary events and sea level changes in the Southwestern Indian Ocean. According to coalescent theory, a large population size results in a longer time to coalescence, allowing more ancient genetic signatures of demographic evolutionary events to be discovered (Hoareau, et al. 2012). In addition, a wide distribution increases the probability of a taxon experiencing biogeographic barriers within its range which may promote divergence and speciation events (Hoareau, et al. 2012).

The aim of this study was to reconstruct the evolutionary history of *L. harak* and assess the effects of Pleistocene climate oscillations on the demographic history of *L. harak*.

5.2 Methods

5.2.1 Sampling and DNA extraction

Sampling was carried out at 6 locations across the SWIO (Figure 5.1, Table 5.1) from 2010 to 2012. Samples were obtained from local fish markets and subsistence fishermen. Samples from Madagascar were collected by Blue Ventures, a non-governmental organisation based in Madagascar. Samples were preserved in absolute ethanol. Total genomic DNA was extracted as described in Chapter 2.2.1 using a phenol/chloroform extraction as described by Sambrook, et al. (1989).

Table 5.1. Sampling records for *L. harak* across the SWIO listing sample site, country and sample size and date of collection.

Country	Region	Code	Original Sample size	No. of samples sequenced	Collection Date
Zanzibar	Stone town fish market	ZZ	78	18	Sept 2010
Mauritius	Mahebourg fish markets	MA	43	22	May 2011
Madagascar	Belo sur Mer	B	51	22	June 2011
Madagascar	Andavadoaka	AN	23	20	June 2011
Mozambique	Pemba fish market	PE	50	21	Sept 2012
Kenya	Mida Creek	DA	24	19	June 2009

5.2.2 Genetic screening

Wherever possible, 20 individuals from each sample were sequenced for COI and CR mtDNA fragments. Samples were PCR amplified for COI using universal fish COI primer pair FishF1, FishR1 (Ward, et al. 2005) and primer pairs Lh COI F, Lh COI R (designed during this study, see Chapter 2.2, Table 2.3). Samples were PCR amplified for CR using the primer pair L CR1F, L CR1R designed during this study (Table 5.2) using the methods described in Chapter 2.2.2.

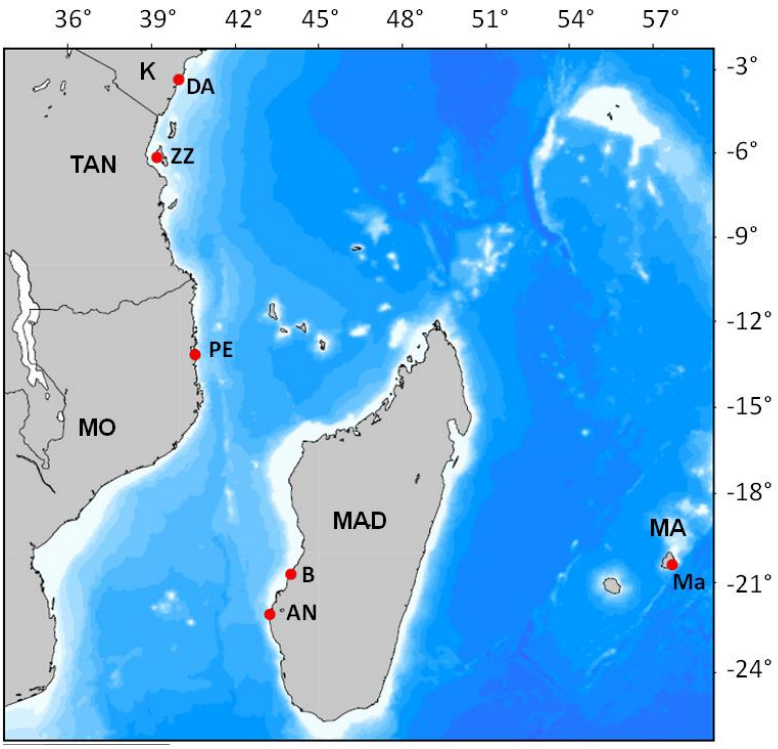

Figure 5.1. Sampling strategy of *L. harak* across the SWIO. Sampling locations are indicated by red circles. Countries; K – Kenya, TAN-Tanzania, MO-Mozambique, MAD-Madagascar, MA-Mauritius. Sampling locations; DA-Mida Creek, ZZ-Zanzibar, PE-Pemba, B-Belo Sur Mer, AN-Andavadoaka, Ma–Mahebourg.

Table 5.2. Primers and annealing temperatures used in this study to amplify various regions of mtDNA of *L. harak*.

Amplified Region	**Primer name (Reference)**	**Primer sequence**	**Annealing Temp (°C)**
COI	FishF1 FishR1 (Ward, et al. 2005)	TCAACCAACCACAAAGACATTGGCAC TAGACTTCTGGGTGGCCAAAGAATCA	55
COI	Lh COIF Lh COIR (This study)	CGAACTTAGTCAGCCCGGA TGCTGATAGAGGATTGGGTC	55
CR	LCR1F LCR1R (This study)	CGGTCTTGTAAACCGGATGT GTCATGGCCCTGAAATAGGA	57

5.3 **Mitochondrial DNA analyses**

Phylogenetic Reconstruction

Phylogenetic relationships among sequences were inferred using network and tree based approaches as described in Chapter 2.2.2. Haplotype networks were constructed for COI, CR and concatenated (COI & CR) mtDNA using maximum likelihood methods in PHYLIP DNAml (Felsenstein 1989) and subsequently constructed in Hapview (Ewing 2010). Sequence divergence between clades was calculated using the minimum p-distance in MEGA6 (Tamura, et al. 2013).

The reconstruction of phylogenetic relationships between individuals was carried out using Bayesian methods as described in Chapter 2.2.2. Maximum likelihood methods were not used in this case as they failed to retrieve a consistent phylogeny.

The most suitable model of nucleotide substitution to account for the observed sequence polymorphism in samples was calculated using Bayesian Information Criterion (BIC) in jModelTest version 2.1.3 (Darriba, et al. 2012; Guindon and Gascuel 2003).

Bayesian Markov Chain Monte Carlo (MCMC) analyses were carried out in MrBayes v 3.2 (Ronquist, et al. 2012) as described in Chapter 2.3.1.

Initially trees were rooted with the sister species *L. lentjan*, identified from BLAST searching as the most closely related species to *L. harak*. Sequence similarity was 93% and 85% for COI and CR respectively. However, use of *L. lentjan* as an outgroup for both CR and COI phylogenetics reconstruction led to the placing of the root of the tree in an unexpected position, which altered ingroup topology. Rooting of the tree using outgroup *L. lentjan* meant that haplotypes which were associated with older clades 1 and 2 (according to other genetic analyses), were assumed to be evolutionarily more recent than clade 3.

Rooting of the COI mtDNA phylogenetic tree was therefore carried using *L. harak* COI sequence from Japan (GenBank accession number JF952781) as an outgroup which had a 96% similarity to the SWIO COI sequences. No suitable sequence could be identified to root the CR or the concatenated mtDNA dataset. However, since mtDNA is non-recombining, the position of the root in the COI phylogenetic tree could be used to infer the correct root in the CR mtDNA phylogenetics tree, allowing reconstructions of all datasets to be carried out.

Trees were constructed in Figtree v.1.4 (Rambaut 2012).

Population diversity and population structure

Arlequin 3.5.1.2 (Excoffier & Lischer, 2010) was used to calculate number of haploptypes (H), haplotype diversity (*h*) and nucleotide diversity (π) as described in Chapter 2.3.1.

The sample-size dependent statistical power of the data to detect genetic differentiation was assessed using the simulation method implemented in POWSIM (v 4.0) (Ryman and Palm 2006) adjusted for organelle (mtDNA) data (Larsson, et al. 2009) as described in Chapter 2.3.1

Population structure was investigated using F_{ST} (haplotypes considered equidistant), Φ_{ST} (divergence between haplotypes considered) and exact tests of population differentiation calculated in Arlequin 3.5.1.2 (Excoffier and Lischer 2010) (see Chapter 2.3.1). Significance levels were adjusted for multiple tests using Bonferroni correction (Rice 1989).

Partitioning of genetic diversity among and within groups of samples was assessed using a hierarchical analysis of molecular variance (AMOVA) performed in Arlequin 3.5.1.2, with 1000 permutations (Excoffier and Lischer 2010).

Demographic history

Tajima's D test (Tajima 1989), Fu's Fs test (Fu 1997) and Ewens-Watterson (Slatkin 1994a) neutrality tests were calculated to test for neutrality. Mismatch distribution was carried out to test for past demographic expansion (see Chapter 2.3.1).

To estimate time since expansion, substitution rates were calculated from divergence times. Substitution rates of 6×10^{-3} per million years (MYR) and 0.018 per MYR were used for COI and CR respectively. Substitution rates were calculated from divergence rates of 1.2% per MYR (Bermingham, et al. 1997) and 3.6% per MYR (Donaldson and Wilson 1999) for COI and CR respectively, calculated from the closure of the Isthmus of Panama. A generation time of 4 years was used for *L. harak* (Ebisawa and Ozawa 2009) and calculations of time since expansion using mismatch analyses were checked using the online spreadsheet http://www.uni-graz.at/zoowww/mismatchcalc/index.php (Schenekar and Weiss 2011).

Time since divergence

Time since divergence was estimated by calculating the percentage divergence between clades and then applying a divergence rate of 3.6% per MYR for CR mtDNA (Donaldson and Wilson 1999) (see Chapter 2.3.1).

5.4 **Results**

Haplotype phylogeny

One hundred and twenty two individuals were sequenced for a 510bp fragment of COI and a 390bp fragment of HVR1 producing 6 haplotypes for COI and 24 haplotypes for CR, resulting in 27 haplotypes for the concatenated dataset of 900bp (Figure 5.2).

The haplotype network for COI (Figure 5.2A) reveals two haplogroups, within which are two dominant haplotypes; haplotype 1 (73 individuals) and haplotype 2 (44 individuals), which are present in varying proportions at all sites across the SWIO (Table 5.3, Figure 5.3). Haplotype 2 is dominant at Mauritius with 73% of the individuals sequenced possessing haplotype 2 (Figure 5.3), whilst haplotype 1 is the most common haplotype at all other sites across the SWIO, on average comprising 70% of the haplotypes at each sampling location (excluding Mauritius) and accounting for 60% of the total haplotypes in the dataset. Haplotype 5 was found at both Pemba and Andavadoaka (2 individuals) whilst the remaining 3 haplotypes were private haplotypes. Outgroup rooting of the tree identified revealed haplogroup 2 as the basal node (Figure 5.4).

Table 5.3. *L. harak* haplotype frequencies at sampling locations across the SWIO based on the 510bp COI mtDNA dataset. ZZ – Stonetown fish market, Zanzibar, MA – Mahebourg, Mauritius, B – Belo sur Mer, Madagsacar, AN – Andavadoaka, Madagascar, Pe – Pemba, Mozambique, DA – Mida Creek, Kenya.

	Haplotype frequency						
Haplotype	**ZZ**	**MA**	**B**	**AN**	**PE**	**DA**	**Overall**
1	9	6	16	12	17	14	74
2	7	16	6	7	4	4	44
3	1	0	0	0	0	0	1
4	1	0	0	0	0	0	1
5	0	0	0	1	1	0	2
6	0	0	0	0	0	1	1

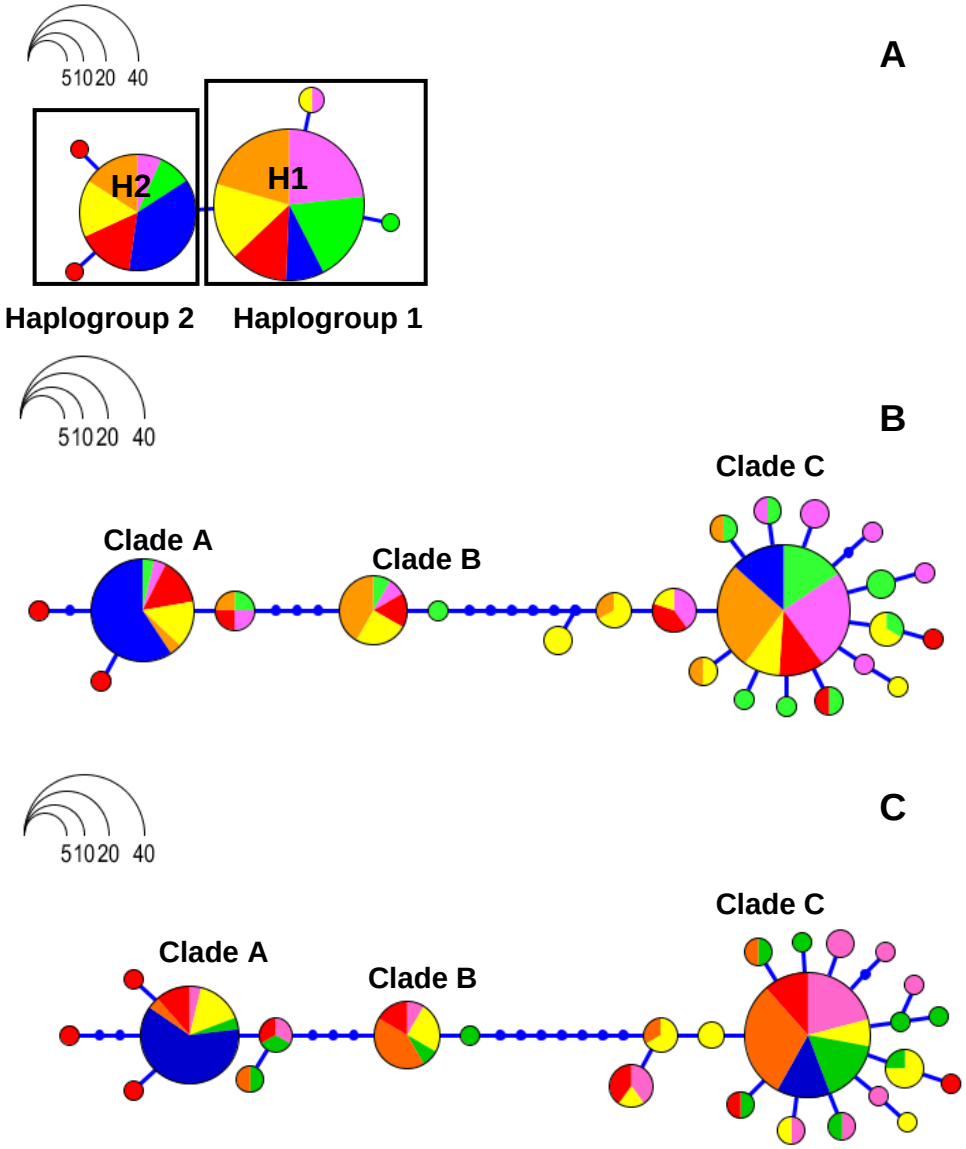

Figure 5.2. Haplotype networks of *L. harak* based on **A**. 510bp of COI mtDNA. Branch lengths are proportional to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation. H1 – haplotype 1, H2 - haplotype 2 **B**. Haplotype network of *L. harak* based on 390bp of CR mtDNA Branch lengths are proportional to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation **C**. Haplotype network of *L. harak* based on the 900bp concatenated mtDNA dataset. Branch lengths are proportional to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation; Tanzania ●, Mauritius ●, Andavadoaka, Madagascar ●, Belo sur Mer, Madagascar ●, Kenya ●, Pemba, Mozambique ●.

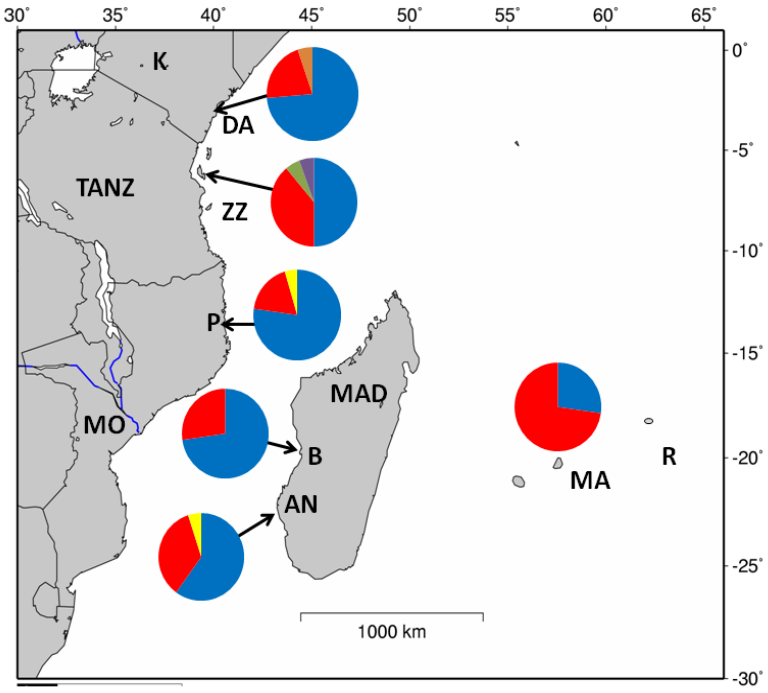

Figure 5.3. Map of SWIO with proportion of *L. harak* COI mtDNA haplotypes at each sampling location; haplotype 1 ●, haplotype 2 ●, haplotype 3 ●, haplotype 4 ●, haplotype 5 ●, haplotype 6 ●.

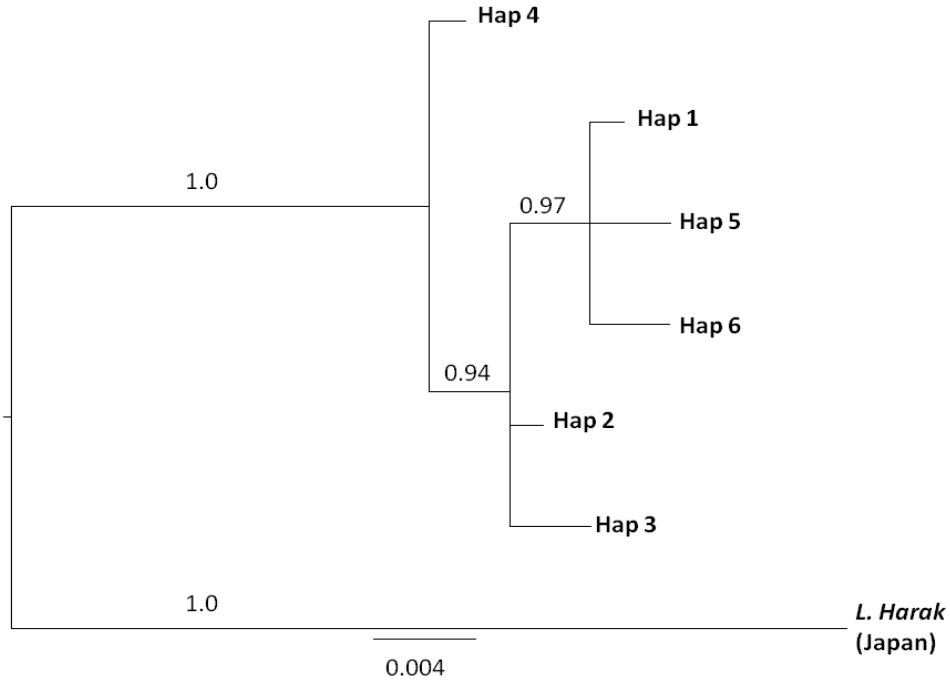

Figure 5.4. Reconstruction of phylogenetic relationships of haplotypes within *Lethrinus harak* using the COI mtDNA dataset. Statistical support for nodes is given for Bayesian posterior probabilities. Outgroup = *L. harak* (Japan).

The CR mtDNA haplotype network revealed finer scale, but complementary phylogenetic diversity to COI mtDNA. Three distinct CR clades were identified (Figure 5.2B, Table 5.4) with CR Clades B and C representing a divergence occurring within

haplogroup 1 of the COI mtDNA whilst clade A corresponded to haplogroup 2 of the COI mtDNA. Phylogenetic reconstruction of the CR dataset supported the existence of three clades (Figure 5.5). Rooting of the CR mtDNA phylogenetic tree was carried out by positioning the root in the same place as for the COI phylogenetic tree as no suitable sequence was available to root the CR mtDNA tree (Figure 5.6). For the concatenated sequences three clades were resolved, mirroring the CR phylogeny (Figure 5.2C, Table 5.5).

Table 5.4. *L. harak* haplotype frequencies at sampling locations across the SWIO based on the 390bp CR mtDNA dataset.

Haplotype	**Haplotype frequency**							**Clade**
	ZZ	**MA**	**B**	**AN**	**PE**	**DA**	**Overall**	
1	1	0	0	0	0	1	2	C
2	2	0	0	1	2	0	5	C
3	4	16	1	4	1	1	27	A
4	1	0	0	0	0	0	1	C
5	2	0	5	3	1	1	12	B
6	5	6	12	4	11	7	45	C
7	1	0	1	0	1	1	4	A
8	1	0	0	0	0	0	1	A
9	1	0	0	0	0	0	1	A
10	0	0	0	0	1	0	1	C
11	0	0	0	0	1	0	1	C
12	0	0	0	0	1	0	1	C
13	0	0	0	0	2	0	2	C
14	0	0	0	0	1	1	2	C
15	0	0	0	2	0	1	3	C
16	0	0	0	0	0	1	1	C
17	0	0	0	0	0	2	2	C
18	0	0	0	0	0	1	1	B
19	0	0	1	0	0	1	2	C
20	0	0	0	0	0	1	1	C
21	0	0	1	2	0	0	3	C
22	0	0	1	1	0	0	2	C
23	0	0	0	1	0	0	1	C
24	0	0	0	2	0	0	2	C

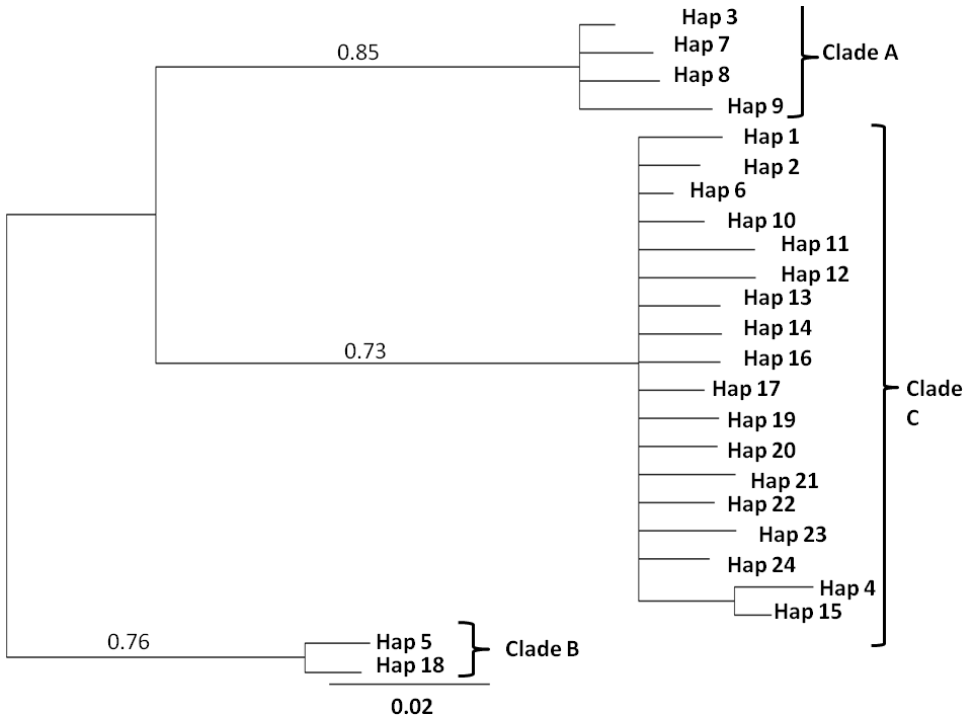

Figure 5.5. Reconstruction of phylogenetic relationships of haplotypes within *L. harak* based on the 390bp CR mtDNA dataset. Statistical support for nodes is given for Bayesian posterior probabilities.

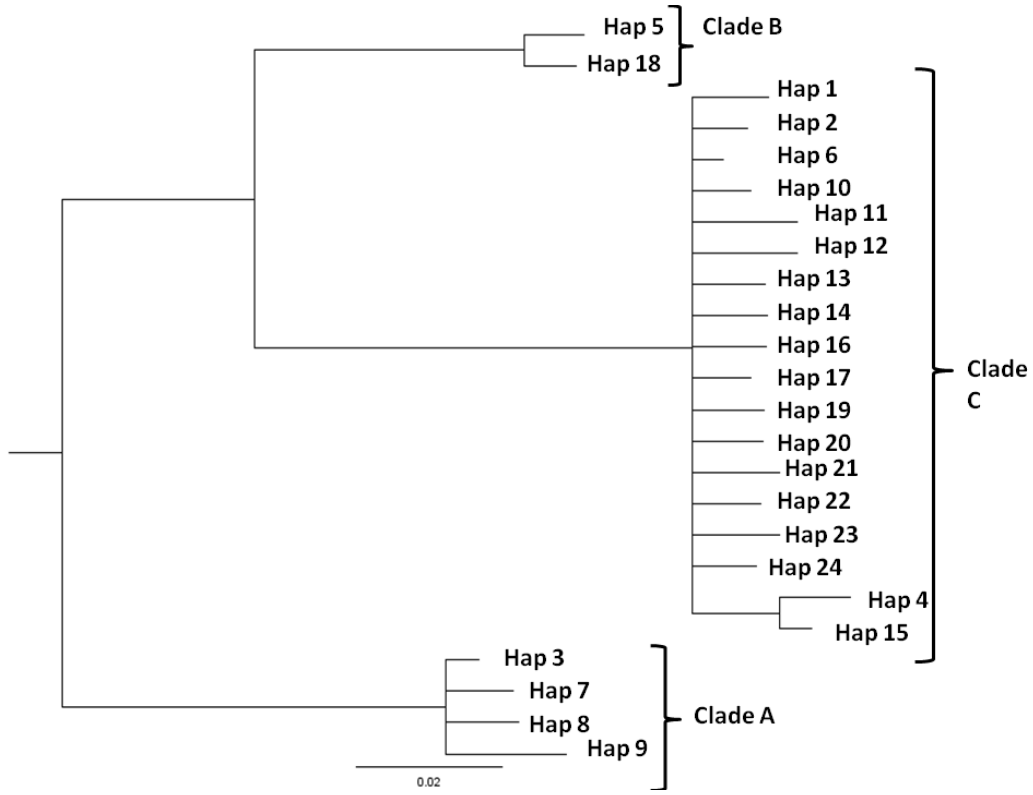

Figure 5.6. Reconstruction of phylogenetic relationships of haplotypes within *L. harak* using the 390bp CR mtDNA dataset. Statistical support for nodes is given for Bayesian posterior probabilities.

Table 5.5. *L. harak* haplotype frequencies at sampling locations across the SWIO based on the 900bp concatenated mtDNA sequences.

Haplotype	Haplotype frequency							Clade
	ZZ	MA	B	AN	PE	DA	Overall	
1	1	0	0	0	0	1	2	C
2	2	0	0	1	2	0	5	C
3	3	16	1	4	1	1	26	A
4	1	0	0	0	0	0	1	C
5	2	0	5	3	1	1	12	B
6	5	6	12	3	9	7	42	C
7	1	0	1	0	1	1	4	A
8	1	0	0	0	0	0	1	A
9	1	0	0	0	0	0	1	A
10	0	0	0	0	1	0	1	C
11	0	0	0	0	1	0	1	C
12	0	0	0	0	1	0	1	C
13	0	0	0	0	2	0	2	C
14	0	0	0	0	1	1	2	C
15	0	0	0	2	0	1	3	C
16	0	0	0	0	0	1	1	C
17	0	0	0	0	0	1	1	C
18	0	0	0	0	0	1	1	B
19	0	0	1	0	0	1	2	C
20	0	0	0	0	0	1	1	C
21	0	0	1	2	0	0	3	C
22	0	0	1	1	0	0	2	C
23	0	0	0	1	0	0	1	C
24	0	0	0	2	0	0	2	C
25	1	0	0	0	0	0	1	A
26	0	0	0	1	1	0	2	C
27	0	0	0	0	0	1	1	C

The concatenated dataset and the CR dataset confirmed the presence of 3 clades A, B and C, across the SWIO, complementary to the two haplogroups found in the COI dataset (Figure 5.2). Distribution of the clades across regions in the SWIO varied considerably. Clade A is dominant at Mauritius with 73% of the individuals sequenced possessing clade A haplotypes (Figure 5.7), whilst clade C is most common at all other sites across the SWIO. Clade C comprised 50% - 85.5% of the haplotypes at all sampling locations excluding Mauritius and accounted for 63% of the total haplotypes

in the dataset (Table 5.6). Clade B was present at all sampling locations at low frequencies (5% - 27%) except at Mauritius. Highest frequency of clade B was at Belo sur Mer where it comprised 27% of the sample.

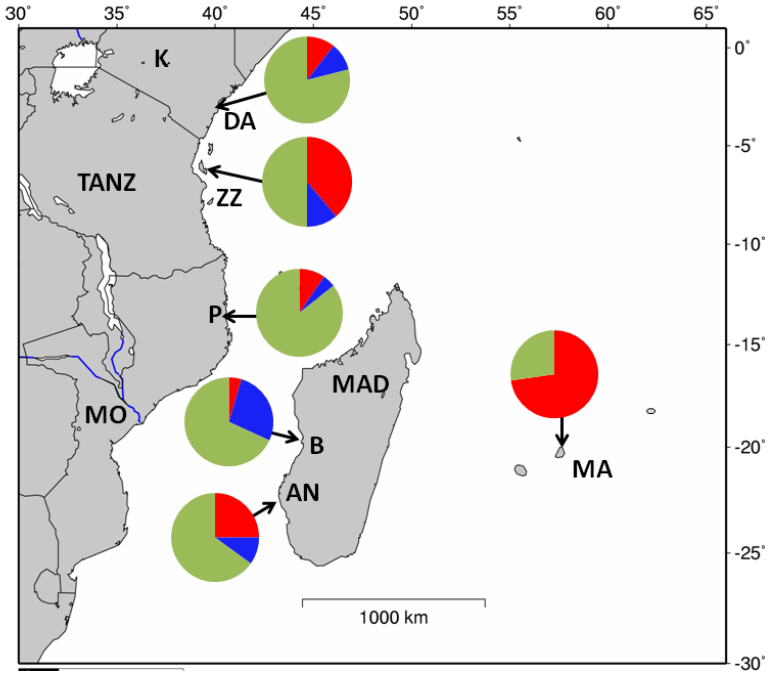

Figure 5.7. A. Map of SWIO showing the proportion of *L. harak* individuals belonging to concatenated mtDNA clades 1-3 at each sampling site. Clade A = ●, Clade B = ●, Clade C = ●.

Table 5.6. Table displaying proportion of *L. harak* clades present at each sampling location across the SWIO.

Sampling Location	**Proportion of Clades (%)**		
	A	**B**	**C**
ZZ	39.0	11.0	50.0
MA	73.0	0.0	27.0
B	4.5	27.0	68.5
AN	25.0	10.0	65.0
PE	9.5	5.0	85.5
DA	10.5	10.5	79.0

Haplotype diversity per sampling location

Haplotype diversity ranged from h = 0.39 - 0.61 and h = 0.42 - 0.91 for COI and CR mtDNA respectively (Table 5.7, 5.8). Nucleotide diversity of COI (π = 0.0008 - 0.0015) was considerably lower than CR nucleotide diversity (π = 0.008 - 0.017). The highest haplotype and nucleotide diversity was seen at Zanzibar for COI mtDNA (h = 0.63, π = 0.0015). Andavadoaka had the highest haplotype diversity for CR mtDNA (h = 0.91) and the highest nucleotide diversity was observed at Zanzibar (π = 0.0015). Mida creek

in Kenya and Andavadoaka in Madagascar had considerably higher haplotype diversity than Mauritius (h = 0.042) for both COI and CR mtDNA.

Low COI haplotype diversity was observed at Pemba, Mozambique (h = 0.0039) followed by Belo sur Mer and Mauritius (h = 0.0042). Lowest COI nucleotide diversity was observed at Pemba, Mauritius and Belo sur Mer, (π = 0.008, Table 5.7).

Sequences were then separated into the three clades according to the CR haplotype network and genetic diversity levels were calculated for each clade (Table 5.9). Clade C had considerably higher haplotype and nucleotide diversity than clades B and A (h = 0.65, π = 0.0023). Very low haplotype and nucleotide diversity were observed in clade B (h = 0.15, π = 0.0004). Clade C exhibits significant deviations from neutrality for all three neutrality tests.

Table 5.7. Mitochondrial genetic diversity levels and neutrality tests for the 510bp *L. harak* COI mtDNA; Sample size (n), number of haplotypes (H), number of private haplotypes (Ph), haplotype diversity (h), nucleotide diversity (π), Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs).

Location Name	Country	n	H	pH	h	π	F	D	Fs
Stone Town fish market	Zanzibar	18	4	2	0.63	0.0015	0.41	-0.38	-0.76
Mahebourg fish markets	Mauritius	22	2	0	0.42	0.0008	0.60	0.90	1.12
Belo sur Mer	Madagascar	22	2	0	0.42	0.0008	0.60	0.90	1.12
Andavadoaka	Madagascar	20	3	0	0.54	0.0011	0.49	0.06	0.07
Pemba fish market	Mozambique	22	3	0	0.39	0.0008	0.63	-0.60	-0.55
Dabaso beach	Kenya	19	4	1	0.43	0.0009	0.59	-0.48	-0.42
Overall	-	123	6	3	0.51	0.0011	0.49	-0.83	-2.06

Table 5.8. Mitochondrial genetic diversity levels and neutrality tests for the 390bp *L. harak* CR mtDNA; Sample size (n), number of haplotypes (H), number of private haplotypes (pHap), haplotype diversity (h), nucleotide diversity (π), Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs). Statistically significant estimates (p<0.05) are highlighted in bold.

Location Name	Country	n	H	pHap	h	π	F	D	Fs
Stone Town fish market	Zanzibar	18	10	3	0.88	0.017	0.17	0.71	0.442
fish markets	Mauritius	22	2	0	0.42	0.011	0.60	1.70	9.71
Belo sur Mer	Madagascar	22	7	1	0.67	0.012	0.36	0.59	1.76
Andavadoaka	Madagascar	20	10	2	0.91	0.015	0.14	1.47	0.38
Pemba fish market	Mozambique	22	9	4	0.75	0.008	**0.28**	-1.39	-1.92
Dabaso beach	Kenya	19	13	4	0.87	0.011	**0.17**	-0.83	**-3.68**
Overall	-	123	27	14	0.81	0.014	**0.20**	0.41	-2.42

Table 5.9. Genetic diversity tests and neutrality tests for the 390bp *L. harak* CR mtDNA separated into clades; Sample size (n), number of haplotypes (H), number of private haplotypes (pHap), haplotype diversity (h), nucleotide diversity (π), Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs). Statistically significant estimates ($p<0.05$) are highlighted in bold.

Clade	**n**	**H**	**pHap**	**h**	**π**	**F**	**D**	**Fs**
A	33	4	2	0.32	0.0010	0.69	-1.50	-1.64
B	13	2	1	0.15	0.0004	**0.86**	-1.14	-0.54
C	76	18	7	0.65	0.0025	**0.35**	**-1.95**	**-17.67**

Population Structure

POWSIM results indicate that the COI dataset has a 5% chance of detecting low levels of population structuring (F_{ST} = 0.001) but a 62% likelihood of detecting considerable population structuring (Table 5.10). POWSIM results also indicated that the CR and concatenated datasets had a 97% chance of detecting population structuring of F_{ST} = 0.05 (Table 5.11, 5.12).

Table 5.10. Statistical power of *L. harak* COI mtDNA sequences to detect population differentiation at various levels of F_{ST} using tests based on Chi^2 and Fisher exact methods.

	Statistical Power	
F_{ST}	Chi^2	Fisher
0.00	0.03	0.06
0.001	0.02	0.05
0.01	0.07	0.10
0.05	0.64	0.62

Table 5.11. Statistical power of *L. harak* CR mtDNA sequences to detect population differentiation at various levels of F_{ST} using tests based on Chi^2 and Fisher exact methods.

	Statistical Power	
F_{ST}	Chi^2	Fisher
0.00	0.01	0.06
0.001	0.04	0.08
0.01	0.19	0.27
0.05	0.97	0.96

Table 5.12. Statistical power of *L. harak* concatenated mtDNA sequences to detect population differentiation at various levels of F_{ST} using tests based on Chi^2 and Fisher exact methods.

	Statistical Power	
F_{ST}	Chi^2	Fisher
0.00	0.01	0.06
0.001	0.02	0.06
0.01	0.16	0.26
0.05	0.99	0.98

Assessment of population structure using F_{ST}, Φ_{ST} and exact tests of differentiation on the COI, CR and the concatenated mtDNA dataset reveal two significantly differentiated populations; a Mauritian population and a SWIO population comprising all other sample sites (Tables 5.13 - 5.18). Some of these values become non-significant after Bonferroni correction. After Bonferroni correction, F_{ST}, and exact tests for the concatenated dataset still indicate that Mauritius is significantly differentiated from all other sampling locations (Tables 5.17 - 5.18).

Table 5.13. Estimates of genetic differentiation of *L. harak* based on analysis of 510bp fragment of COI mtDNA. F_{ST} and Φ_{ST} estimates are presented below and above the diagonal respectively. Statistically significant estimates ($p<0.05$) are highlighted in bold. Values with * indicate significance after Bonferroni correction.

	ZZ	**MA**	**B**	**AN**	**PE**	**DA**
ZZ	-	**0.049**	0.049	-0.006	0.127	0.091
MA	**0.106**	-	**0.311***	**0.197**	**0.407***	**0.364***
B	0.019	**0.311**	-	-0.032	-0.021	-0.035
AN	-0.037	**0.178**	-0.223	-	-0.010	-0.004
PE	**0.071**	**0.390***	-0.031	0.0153	-	-0.037
DA	0.036	**0.344**	-0.043	-0.007	-0.042	-

Table 5.14. Exact test of differentiation p values based on *L. harak* 510bp COI mtDNA samples. Statistically significant estimates ($p<0.05$) are highlighted in bold. Values with * indicate significance after Bonferroni correction.

	ZZ	MA	B	AN	PE	DA
MA	0.05854 ±0.0010					
B	0.20727 ±0.0019	**0.00568 ±0.0002**				
AN	0.63738 ±0.0017	**0.02870 ±0.0006**	0.61469 ±0.0013			
PE	0.10215 ±0.0017	**0.00065 ±0.0003***	0.72240 ±0.0009	0.49908 ±0.0014		
DA	0.22245 ±0.0019	**0.00227 ±0.0006***	0.71740 ±0.0011	0.43169 ±0.0019	0.91403 ±0.0006	-

Table 5.15. Estimates of genetic differentiation of *L. harak* based on analysis of 390bp CR mtDNA. F_{ST} and Φ_{ST} estimates are presented below and above the diagonal respectively. Statistically significant estimates ($p<0.05$) are highlighted in bold. Values with * indicate significance after Bonferroni correction.

	ZZ	MA	B	AN	PE	DA
ZZ	**-**	0.090	0.048	-0.002	**0.168***	**0.103**
MA	**0.097**	-	**0.347***	**0.267***	**0.479***	**0.412***
B	0.055	**0.367***	-	-0.031	0.041	-0.012
AN	0.001	**0.278***	-0.029	-	0.060	0.008
PE	**0.145**	**0.464***	-0.035	0.050	-	-0.027
DA	0.010	**0.418***	-0.010	0.010	-0.029	-

Table 5.16. Exact test of differentiation values based on *L. harak* 390 bp CR mtDNA. Statistically significant estimates (p<0.05) are highlighted in bold. Values with * indicate significance after Bonferroni correction.

	ZZ	MA	B	AN	PE
MA	**0.00423* ±0.0011**				
B	0.08291 ±0.0069	**0.00000* ±0.0000**			
AN	0.52264 ±0.0104	**0.00145* ±0.0005**	0.07510 ±0.0066		
PE	0.33820 ±0.0071	**0.00000* ±0.0000**	0.23811 ±0.0075	**0.02980 ±0.0022**	
DA	0.48108 ±0.0111	**0.00004* ±0.0000**	0.29593 ±0.0066	0.16682 ±0.0070	0.60528 ±0.0126

Table 5.17. Estimates of genetic differentiation of *L. harak* based on analysis of 900bp of concatenated mtDNA. F_{ST} and Φ_{ST} estimates are presented below and above the diagonal respectively. Statistically significant estimates (p<0.05) are highlighted in bold. Values with * indicate significance after Bonferroni correction.

	ZZ	MA	B	AN	PE	DA
ZZ	-	0.088	0.052	-0.002	**0.161**	**0.104**
MA	**0.189***	-	**0.351***	**0.265***	**0.460***	**0.409***
B	0.053	**0.359***	-	-0.028	0.028	-0.013
AN	-0.017	**0.178***	0.088	-	0.051	0.013
PE	-0.007	**0.274***	0.018	0.032	-	-0.030
DA	-0.001	**0.256***	0.021	0.013	-0.020	-

Samples were then grouped into clades to assess the partitioning of genetic variation using a hierarchical analysis of molecular variance (AMOVA) in Arlequin (Table 5.19). Pairwise differences indicated that 90.9% of the variation was explained by the partitioning of clades A-C and only 9.1% of the variance was found within clades, whilst haplotype frequencies indicated that 50% of the variation was explained by partitioning of the clades whilst the other 50% was explained by structuring within the clades.

Table 5.18. Exact test of differentiation values based on *L harak* 900bp concatenated mtDNA samples. Statistically significant estimates (p<0.05) are highlighted in bold. Values with * indicate significance after Bonferroni correction.

	ZZ	MA	B	AN	PE
MA	**0.00273* ±0.0063**				
B	0.09359 ±0.0063	**0.00000* ±0.0000**			
AN	0.51945 ±0.0112	**0.00150* ±0.0006**	**0.02503 ±0.0023**		
PE	0.65861 ±0.0134	**0.00011* ±0.0000**	0.15797 ±0.0067	0.10024 ±0.0031	
DA	0.78520 ±0.0071	**0.00000* ±0.0000**	0.28343 ±0.0115	0.24519 ±0.0099	0.82921 ±0.0079

Table 5.19. Analysis of molecular variance (AMOVA) based on 900bp concatenated mtDNA displaying the total variance explained by the grouping of clades A, B & C of *L. harak* populations in the SWIO. Statistically significant estimates (p<0.05) are highlighted in bold.

	Pairwise differences		Haplotype frequencies	
	Variance	% Total	Variance	% Total
Among clades	4.08	**90.9**	17.66	**50**
Within clades	0.41	**9.1**	31.66	**50**

Demographic History

Fu's Fs (Fu 1997), Tajima's D (Tajima 1989), and Ewens-Watterson test (Slatkin 1994a) showed no deviation from neutrality per site or globally for COI mtDNA (Table 5.7). Ewens-Watterson test (Slatkin 1994a) displayed deviation from neutrality at Kenya, Pemba and overall for CR mtDNA and Fu's Fs test was also significant for Kenya (Table 5.8). All three neutrality tests displayed significant deviations from neutrality for *L. harak* clade C based on the CR mtDNA (Table 5.9). Clade A displayed no deviation from neutrality and clade B displayed deviation from neutrality for Ewens-Watterson test (Slatkin 1994a).

To test the theory of a past population expansion, mismatch analysis was conducted for the COI and CR mtDNA datasets (Tables 5.20, 5.21). However, when using CR data Tau varied widely and Theta values at time 0 and 1 were unrealistic. In addition, the Mauritius SSD value was significant, not supporting the hypothesis of a population expansion. When the mismatch graph was observed for CR mtDNA, a multimodal population can be seen (Figure 5.8), consistent with a number of expansions as

opposed to a single population expansion which mismatch analysis assumes. The CR dataset was therefore partitioned into its three respective clades and mismatch analysis repeated (Table 5.22, Figure 5.9-5.11). Mismatch analysis for clades per sampling site could not be carried out using site specific data for clades 1 and 2 as often only one haplotype was present at each sampling location.

Mismatch analysis of clades A, B and C using CR mtDNA suggests the existence of two separate population expansions, one 215,000 years ago and a more recent population expansion 71,500 years ago (Table 5.22, Figure 5.9 - 5.11). Ninety five percent confidence intervals for estimates of clade B population expansion are particularly large, ranging from 0 – 3,247,900 years ago. No significant deviation from the hypothesis of a population expansion was observed for COI mtDNA with the date of the expansion ranging from 83,000-150,000 years ago (Table 5.20).

Table 5.20. Mismatch analysis of *L harak* using COI mtDNA. T time since expansion in mutational units with 95% confidence intervals in brackets below, θ_0 population size before expansion θ_1 population size after expansion, SSD sum of squared deviations, TSE time since expansion in thousands of years with 95% confidence intervals in brackets below. Statistically significant estimates ($p<0.05$) are highlighted in bold.

Location	T	θ_0	θ_1	SSD	TSE (KA)
ZZ	0.9 (0.21-1.91)	0.000	99999.00	0.0212	147 (34.3-312.1)
MA	0.6 (0.02-1.21)	0.000	99999.00	0.0087	100 (3.3-197.7)
B	0.6 (0.02-1.27)	0.000	99999.00	0.0087	100 (3.3-207.5)
AN	0.8 (0.15-2.12)	0.004	99999.00	0.0322	130 (24.5-346.4)
PE	0.5 (0.01-1.17)	0.000	99999.00	0.0077	8 (1.63-191.1)
DA	0.6 (0.00-2.01)	0.002	99999.00	0.0012	100 (0.00-326.8)

Table 5.21. Mismatch analysis of *L. harak* using CR mtDNA. T time since expansion in mutational units, θ_0 population size before expansion θ_1 population size after expansion, SSD sum of squared deviations, TSE time since expansion in thousands of years. Statistically significant estimates ($p<0.05$) are highlighted in bold.

Location	T	θ_0	θ_1	SSD	TSE (KA)
Zanzibar	11.8 (2.07-22.23)	0.002	12.74	0.0421	842 (147.4-1,583.3)
Mauritius	3.0 (0.39-3.50)	0.000	0.45	**0.2302**	214 (27.8-249.3)
Belo	10.5 (0.26-93.45)	0.000	3.53	0.1111	750 (18.5-6656.0)
Andavadoaka	11.5 (0.00-61.77)	0.004	10.01	0.0387	840 (0.0-4399.6)
Pemba	0.5 (0.00-4.00)	1.143	99999.00	0.0264	36 (0.00-284.9)
Kenya	1.4 (0.43-2.52)	0.000	99999.00	0.0621	100 (30.6-194.5)

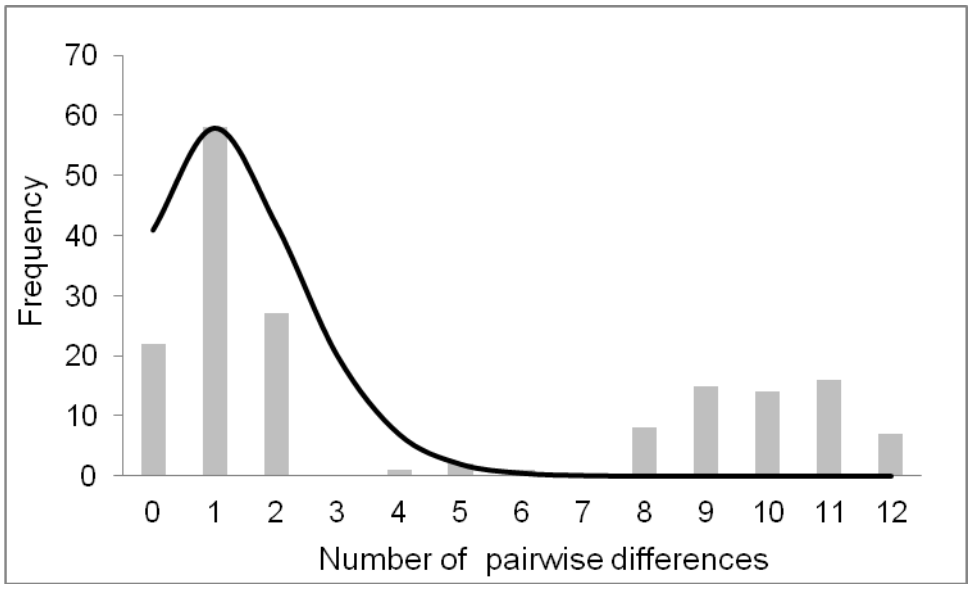

Figure 5.8. Mismatch distribution histogram for *L. harak* CR mtDNA. Filled bars indicate the observed frequency of pairwise distribution, the black line indicates the expected distribution under a model of sudden demographic expansion.

Table 5.22. Mismatch analysis of *L. harak* using CR mtDNA partitioned into 3 clades as indicated by CR haplotype network. T - Time since expansion mutational units, θ_0 population size before expansion θ_1 population size after expansion, SSD sum of square deviations, TSE time since expansion in thousands of years.

Clade	T	θ_0	θ_1	SSD	TSE (KA)
A	3.0 (0.00-4.20)	0.000	9999.99	0.002	215 (0.0-299.1)
B	3.0 (0.00-45.60)	0.000	9999.99	0.028	215 (0.0-3247.9)
C	1.0 (0.54-1.63)	0.000	9999.99	0.002	71.5 (38.5-116.1)

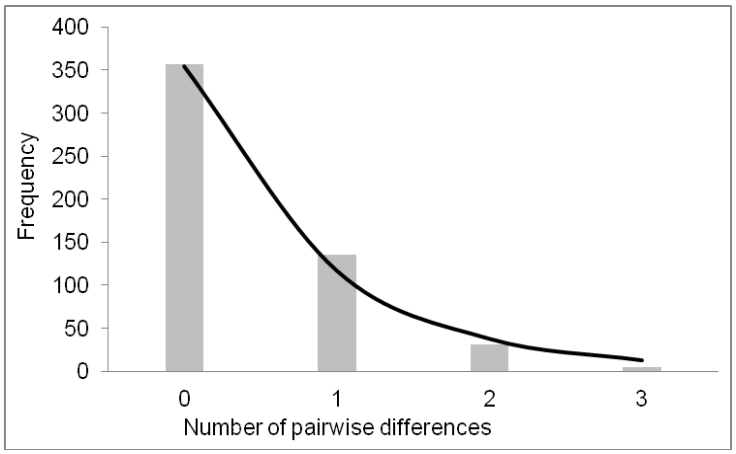

Figure 5.9. Mismatch distribution histogram for clade A of *L. harak* CR mtDNA. Filled bars indicate the observed frequency of pairwise distribution, the black line indicates the expected distribution under a model of sudden demographic expansion.

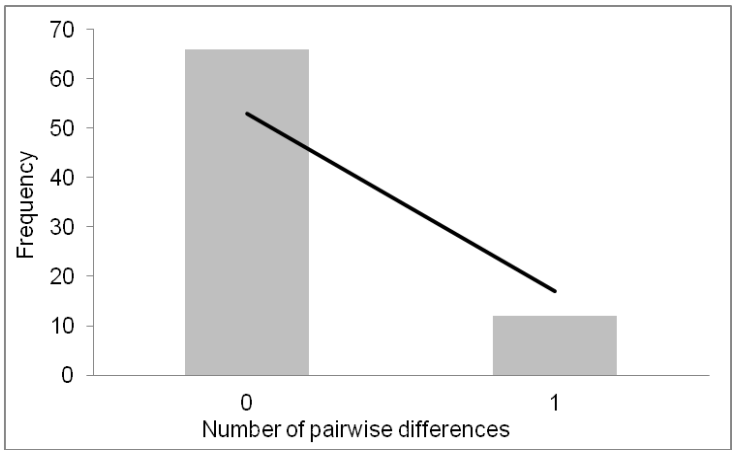

Figure 5.10. Mismatch distribution histogram for clade B of *L. harak* CR mtDNA. Filled bars indicate the observed frequency of pairwise distribution, the black line indicates the expected distribution under a model of sudden demographic expansion.

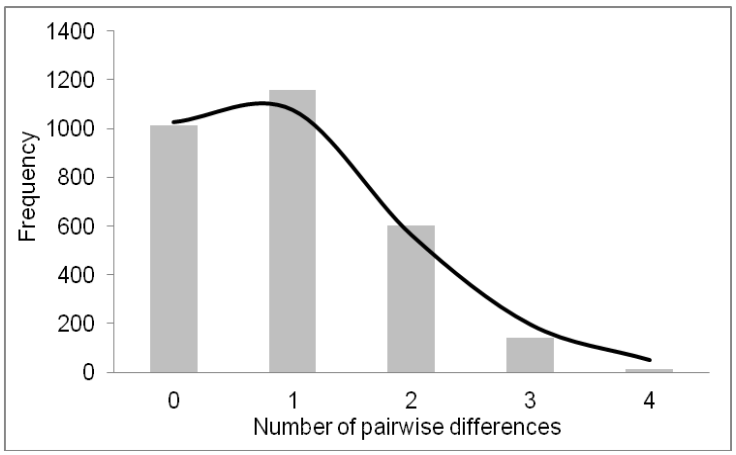

Figure 5.11. Mismatch distribution histogram for clade C of *L. harak* CR mtDNA. Filled bars indicate the observed frequency of pairwise distribution, the black line indicates the expected distribution under a model of sudden demographic expansion.

Time since divergence

Phylogenetic reconstruction indicates that clade A diverged from clades B and C, followed by the divergence of clades B from C. Genetic divergence between clade A and clades B and C was 2.0%, resulting in an estimated divergence date of 550,000 years ago. Genetic divergence between clades B and C was 1.8%, resulting in an estimate of divergence 500,000 years ago.

5.5 **Discussion**

Genetic diversity

Population assessment of genetic diversity for COI mtDNA *L. harak* revealed average values for haplotype diversity (overall h = 0.51) and considerably lower values for nucleotide diversity (overall π = 0.0011). Values of haplotype and nucleotide diversity were consistently higher for Zanzibar compared with the remaining sampling locations. Genetic diversity for mtDNA CR for *L. harak* revealed high haplotype diversity (overall h = 0.81) and average values for nucleotide diversity (π = 0.014). Unlike the COI mtDNA diversity statistics, Zanzibar did not have consistently higher haplotype and nucleotide diversity for CR mtDNA. Haplotype and nucleotide diversity reported for COI and CR in *L. harak* are considerably lower than other *Lethrinus* species. COI haplotype and nucleotide diversity levels of h = 0.69 - 0.80, π = 0.004 - 0.0076 have been reported for *L. nebulosus* as part of this study (see Chapter 6). Additionally, CR haplotype and nucleotide diversity statistics for *L. miniatus* (h = 0.61 - 0.98, π = 0.004 - 0.02) and *L. nebulosus* (h = 0.98 - 1.00, π = 0.015 - 0.044) are considerably higher than those reported for *L. harak*.

The considerably lower haplotype and nucleotide diversity in *L. harak* than other *Lethrinus* species may be due to a number of factors. Within-species genetic diversity

is thought to reflect a species population size, demographic history, ecology and its adaptive ability (Bazin, et al. 2006). However, it has recently been argued that population size does not influence mtDNA genetic diversity (Bazin, et al. 2006). Significant deviation from neutrality was observed for clade C (Tajima's D = -17.67 $p<0.05$) of the CR mtDNA, indicating either a past population expansion or a selective sweep. Mismatch analysis confirmed the existence of a past population expansion suggesting that the reduced diversity statistics may be in part due to a recent population expansion. However the possibility of a selective sweep having occurred cannot be excluded.

Phylogeography and population structure

The CR mtDNA dataset revealed the presence of three divergent clades, A - C, across the SWIO. The COI mtDNA dataset revealed complementary haplogroups, with clade A corresponding to haplogroup 2 and clades B and C corresponding to haplogroup 1. The proportion of clades defined by the CR mtDNA present at sampling locations varied considerably across the SWIO. Clade C was the largest clade, comprising 73 individuals and was predominant at all sampling sites except Mauritius where Clade A was dominant. Clade B was present at all sampling sites except Mauritius.

Significant population structuring of the Mauritius population

The presence of the three clades in varying proportions across the SWIO led to considerable population structuring between Mauritius and the other locations, confirmed by exact tests, F_{ST} and Φ_{ST}. Clade A is the dominant clade in the Mauritius sample, as opposed to Clade C which is dominant at all other sampling sites. Therefore F_{ST}, which measures haplotype frequencies to estimate population differentiation, found significant population structuring between Mauritius and other sampling locations. Additionally, there is considerable nucleotide divergence between clades A and C explaining why Φ_{ST}, which accounts for divergence between haplotypes, also indicated structuring between Mauritius and all other sampling locations.

The high proportion of clade A haplotypes at Mauritius, the low proportion of clade C haplotypes and the absence of clade B haplotypes suggest that gene flow between Mauritius and the rest of the sampling locations in the SWIO has been intermittent. The high proportion of clade A at Mauritius compared to other sampling locations may have a number of causes; one hypothesis is that this clade evolved here. However, the lack of low frequency haplotypes in clade A from Mauritius does not support this theory (Hewitt 1999). However, since the Mauritian population of *L. harak* is likely to be considerably smaller than the African mainland populations (as it is a small island population with limited habitat, in comparison to the large habitat available at the African mainland population) it may be that low haplotype frequencies of clade A have

been lost by genetic drift. Alternatively, a population expansion of clade A may have occurred in the eastern Indian Ocean/Pacific and spread westwards across the Indian Ocean via the Southern Equatorial Current. If this were the case, colonisers would first reach Mauritius, followed by the African mainland, accounting for the higher proportion of clade A observed at Mauritius. The more recent population expansion of clade C is likely to have originated from the African mainland. This theory is supported by the high number of low frequency haplotypes belonging to clade C which originate from the African mainland. Current oceanographic conditions may have limited the spread of this population expansion to Mauritius thus accounting for the proportion of clade A and C observed at Mauritius. Again however, the lack of low frequency haplotypes at Mauritius in clade A, (which would be expected if individuals from a population expansion in the Indo-Pacific had reached Mauritius first) does not support this theory. It is also possible that the skewed clade frequencies may be a result of genetic drift post isolation or amplified as the result of a founder effect.

The most likely explanation for the difference in proportion of clades between Mauritius and the rest of the sampling locations across the SWIO is contemporary isolation. The predominant current flow in the region is from east to west via the South Equatorial Current which flows across the Mascarene Plateau north of Mauritius towards the East African coastline. Adults of *L. harak* are demersal and do not migrate large distances. Therefore most dispersal and subsequent gene flow occurs during the pelagic larval phase. Pelagic larvae may employ a range of physiological and behavioural characteristics to facilitate favourable dispersal, whether that be dispersal over large areas, or alternatively remaining close to their natal reef (Leis 2007). Despite any larval behavioural characteristics, it is likely that the strong east to west current in the form of the SEC limits the number of larvae which reach Mauritius from the African mainland. However, if larvae from Mauritius were entering the SEC, it could be expected that some larvae that originated from Mauritius would reach the African coastline, particularly as no structuring was observed between the Seychelles, the furthest east sampling location, and the African mainland. The results of this study however indicate contemporary isolation, suggesting that gene flow from Mauritius to the African mainland is not occurring, perhaps due to some local process at Mauritius which is preventing larval dispersal. This has also been suggested to be the case for *Lutjanus kasmira* in the SWIO which found structuring between Mauritius and the Mozambique Channel due to local processes (Muths, et al. 2012). However, despite comprehensive review of the current literature on the SWIO, the cause of this isolation remains elusive.

Population structure in the rest of the SWIO

Population structuring in the rest of the SWIO, excluding Mauritius, was minimal. F_{ST} tests for both the COI and CR datasets indicated some structuring between Pemba and Zanzibar which was non-significant after Bonferroni correction. The concatenated dataset showed no structuring for $F_{ST,}$ Φ_{ST} and exact tests after Bonferroni correction. CR mtDNA Φ_{ST} displayed structuring between Pemba and Zanzibar after Bonferroni correction, perhaps indicating some structuring between Pemba and Zanzibar.

No other significant structuring was observed in the dataset, indicating that there is high levels of connectivity across the remaining populations sampled in the SWIO. The maintenance of this connectivity over distances of up to 3000 km, is likely to be a result of both life history characteristics of *L. harak* and the oceanography of the region. Adult individuals of *L. harak* are demersal and do not migrate large distances. Dispersal of *L. harak* therefore predominantly occurs during the pelagic larval phase. The SWIO oceanography consists of a number of large scale currents. (see Chapter 1.4). The length of the larval duration of *L. harak*, 28 days, combined with these numerous strong currents may result in a considerable dispersal potential for *L. harak*, thereby maintaining connectivity between populations in the SWIO.

The comparison of this study's results with previous studies of *Lethrinus* population structure and phylogeography perhaps highlights the significance of the isolation of the Mauritian population. It is currently thought that the range of *L. harak* across the SWIO is continuous (Carpenter and Allen 1989). However, no previous studies on *Lethrinus* have found evidence of an isolated population within a continuous range. The majority of population genetic studies of *Lethrinus* have been carried out in Australia. *L. nebulosus* populations were found to be panmictic across distances of 1400 – 2080 km in Northwestern Australia using allozyme variability (Johnson, et al. 1993). Similarly, no genetic structuring of *L. miniatus* was observed using eight microsatellite loci and six sampling locations from three geographic regions of the Great Barrier Reef (GBR) over distances of approximately 500-1050 km.

The only study to date which displays population structuring in *Lethrinus* populations is a study into *L. miniatus* populations in East and west Australia using HVR1 sequencing (van Herwerden, et al. 2009). Populations from East and West Australia were found to be discrete, indicating isolation between populations on the East coast and the West coast. *L. miniatus*. However, this is perhaps not surprising as the geographic range of *L. miniatus* across the coast of Australia is known to be discontinuous, with populations in East and West Australia but no known populations along the North coast of Australia.

Population demography

Mismatch analysis of mtDNA CR suggests the occurrence of two past population expansions, 215 KA and 71.5 KA. Confidence intervals of these dates however are large, particularly for clade B most likely because the low diversity of this clade means that Arlequin is unable to calculate an accurate mismatch. Excluding clade B for this reason and taking into account 95% confidence intervals, it is possible that two population expansions have occurred, most likely within the past 300,000 years. The reasons for these population expansions cannot be definitively explained but it is perhaps possible that they are linked to the Pleistocene climate cycles.

Recent studies have shown that Pleistocene sea-level changes have affected the demography of marine organisms both in temperate (Marko, et al. 2010) and tropical (Gaither, et al. 2011; Gaither, et al. 2010; Hoareau, et al. 2012) marine environments. Decreased sea level during glacial periods increases fragmentation of the marine habitat, and reduces the area of suitable habitat available for a species leading to reduced connectivity and gene flow and reduced population sizes), both of which may contribute to genetic drift and differentiation among populations. Rises in sea level during inter-glacial periods favoured demographic expansions as populations recovered a larger distribution range resulting in a contraction-expansion dynamic of populations (Hoareau, et al. 2012; Marko, et al. 2010).

Over the past 900,000 years, seaways in the SWIO have remained open and no ice cover has occurred (Hoareau, et al. 2012). However, significant changes in sea level in of up to 140 m have occurred (Hansen, et al. 2013; Lisiecki and Raymo 2005) in the SWIO, associated with glacial and interglacial periods. It is therefore possible that population expansions of *L. harak* may be linked to increased sea levels in the SWIO during interglacial periods. Increased sea levels would have increased the available habitat for *L. harak*, perhaps causing favourable conditions for a population expansion.

In addition to sea level affecting population divergence, sea surface temperature may also have affected the distribution of *L. harak* populations during the Pleistocene. Information on SST in the SWIO during the Pleistocene is fairly sparse. During the last 150 KA, the tropical Indian ocean remained 1.5 – 2.5 °C within present day temperatures (Sonzogni, et al. 1998). However, during the past 60 KA SST around northern Madagascar varied from 20 - 22°C. South of northern Madagascar the position of the subtropical convergence zone fluctuated during the late Pleistocene, during glacial periods migrating north by up to 5° (Flores, et al. 1999). The current temperature range of *L. harak* is 22 - 32°C (this study) and its current range corresponds to the boundary of the subtropical zone (Carpenter and Allen 1989; Smith 1959). It is therefore likely that it would have been colder than this in the southern limit

of *L. harak's* distribution, resulting in a northward contraction in range of *L. harak* during particularly strong glacial periods to more tropical areas such as Kenya, Tanzania, northern Mozambique and northern Madagascar. There is perhaps some suggestion of this in the haplotype network for clades B and C of the CR mtDNA. The most common haplotype in clades B and C are dominated by samples from Belo sur Mer, the most southerly sampling location in this study. The existence of a genetically depauperate peripheral population at Belo sur Mer perhaps suggests that the population expansion 71.5 KA ago originated in the more tropical regions of the Indian Ocean and then spread southwards as the sea level and sea temperature increased (Eckert, et al. 2008).

In this study we have shown that *L. harak* exhibits genetically diverse populations with high connectivity across the majority of the SWIO, with the exception of Mauritius, which appears to be isolated from the rest of the SWIO. The high connectivity across the majority of the SWIO suggests a lack of major biogeographic boundaries in the region although there may be some local processes occurring at Mauritius.

Chapter 6: Genetic structure of *Lethrinus nebulosus* across the Southwestern Indian Ocean (SWIO).

6.1 Introduction

Predicting the structuring of marine populations is a complex task. Historically, the potential for dispersal over large distances and the absence of apparent barriers in the marine environment led to the theory of demographically open marine populations over hundreds and even thousands of kilometres (Féral 2002). However, in the last 20 years, research has revealed highly differentiated marine populations over relatively small spatial scales (Ayre and Hughes 2004; Hilbish 1996; Reeb and Avise 1990).

Structuring of populations may be the result of historical factors such as past climatic fluctuations (Janko, et al. 2007; Koizumi, et al. 2012; Palero, et al. 2008), palaeoceanography (Fraser, et al. 2009), biogeographic boundaries (Floeter, et al. 2008; Lessios, et al. 1999; von der Heyden, et al. 2008) or contemporary factors such as life history characteristics (Gerlach, et al. 2010; Purcell, et al. 2009; Purcell, et al. 2006), geographic distances (isolation by distance) and oceanographic features (Doubleday, et al. 2009; García-Rodríguez and Perez-Enriquez 2006; Olivares-Banuelos, et al. 2008; Perez-Losada, et al. 2007; Selkoe, et al. 2006).

Molecular genetic data have widespread application in the identification of population and conservation units of marine fish species (Waples, et al. 2008). Knowledge of population structure is essential for sustainable fisheries management (Waples, et al. 2008). Despite this, structuring of many fished species has yet to be investigated using genetic techniques, particularly in less studied areas of the ocean such as the Southwestern Indian Ocean. There is therefore a requirement for molecular studies into population structure of a number of fished species.

Lethrinus nebulosus is a reef-associated fish widespread across the Indo-West Pacific from the East African coast through the Gulf of Aden, Red Sea, Arabian Gulf, along the west and south coasts of India and around Sri Lanka, through to southern Japan and southwards in Samoan, Tongan and Australian waters (Fischer and Bianchi 1984; Grandcourt, et al. 2006; Lo Galbo, et al. 2002). *L. nebulosus* inhabits coral and rocky reefs in shallow inshore waters up to 80 m deep (Fischer and Bianchi 1984).

Like most Lethrinidae, *L. nebulosus* is a long- lived fish, 14 – 26 years with a maximum reported size of 70 – 80 cm in length (Carpenter and Allen 1989; Ebisawa and Ozawa 2009; Fischer and Bianchi 1984). *L. nebulosus* is a protogynous hermaphrodite (Ebisawa and Ozawa 2009; Marriott, et al. 2010) and age at 50% sexual maturity is 4 years (Ebisawa and Ozawa 2009). Reproduction is seasonal and timing of spawning

varies with geographical location (Ebisawa and Ozawa 2009; Grandcourt, et al. 2006). Adults are resident and spawning occurs multiple times (Berry, et al. 2012; Ebisawa and Ozawa 2009). Spawning produces pelagic eggs and larvae which are distributed considerable distances by currents and are predicted to maintain connectivity over large spatial scales, thus maintaining connectivity between populations (Berry, et al. 2012).

L. nebulosus is an important component of commercial and subsistence fisheries in the SWIO (Carpenter and Allen 1989). The life history characteristics of *L. nebulosus*; long life span, slow development and protogyny, make this species particularly vulnerable to over-exploitation (Huntsman and Schaaf 1994). In the Western Indian Ocean, *L. nebulosus* is caught from reefs and coastal waters using bottom trawls, seine nets, traps, hand-lines and gill nets. The species is commercially important in Mauritius, India and Sri Lanka (Fischer and Bianchi 1984; van der Elst, et al. 2005). Despite its importance in the capture component of fisheries in this region, exact landing quantities of *L. nebulosus* are not available as most countries in the SWIO do not report species-level catch statistics to the Fisheries and Agriculture Organisation of the United Nations (FAO). However, in the Western Indian Ocean (FAO Fishing Zone 51) catches of Lethrinidae vary from 61,031 tonnes in 2002 to 38,413 tonnes in 2011 with a general trend of decreasing catches from 2002 to 2011 (FAO 2013). Despite this, no information is available on population structure and basic management of stocks in the region making *L. nebulosus* vulnerable to over-fishing.

Within the range of *L. nebulosus* in the SWIO is the Mozambique Channel, a site of complex oceanographic currents. The Mozambique Channel circulation consists of southward moving anticyclonic eddies on the western side of the channel and poorly defined weak currents on the east side of the channel (Quartly and Srokosz 2004; Schouten, et al. 2003). Oceanographic eddies have recently been identified as a possible factor governing population sub-structuring as they may disrupt gene flow between adjacent populations and induce local larval retention (Galarza, et al. 2009; Hoffman, et al. 2012; Jackson, et al. 2014; Mokhtar-Jamai, et al. 2011; Selkoe, et al. 2010). Additionally, previous studies have shown genetically differentiated lineages north and south of the Mozambique Channel in spiny lobster *Palinurus delagoae* and mangrove crab *Perisesarma guttatum* (Gopal, et al. 2006; Silva, et al. 2010a). The southern limit of *L. nebulosus*' distribution is Algoa Bay (34 °S) in South Africa. Abundance of *L. nebulosus* increases northwards along the East African coast and the core distribution of *L. nebulosus* in the SWIO may be considered northwards of 26 °S (Fischer and Bianchi 1984; Smith 1959), encompassing both the channel itself and north and south of the Mozambique Channel.. The circulation in the Mozambique

Channel has been relatively constant for the past 30,000 years (Obura 2012) and may therefore have acted as a long term barrier to gene flow, ultimately generating separate phylogeographic lineages north and south of the channel.

The aim of this study therefore was to investigate the population genetic structure of *L. nebulosus* across the SWIO, with emphasis on the potential role of the Mozambique Channel as a biogeographic barrier by combining genetic markers with different evolutionary modalities to attempt to distinguish between historical and contemporary structuring processes. The implications of the findings for fisheries management are discussed.

6.2 Methods

6.2.1 Sampling and DNA extraction

Sampling was carried out at 6 locations across the SWIO (Figure 6.1, Table 6.1) from 2010 to 2012. Samples were obtained from local fish markets and subsistence fishermen. Samples from Madagascar were collected by non-governmental organisation (NGO) Blue Ventures. Samples were preserved in absolute ethanol and total genomic DNA was extracted as described in Chapter 2.2.1 using a phenol/chloroform extraction (Sambrook, et al. 1989).

Additionally, the South African Institute of Aquatic Biodiversity (SAIAB) collected a number of *L. nebulosus* samples in collaboration with this study (Figure 6.1, Table 6.2). These samples were extracted and sequenced for control region (CR) by SAIAB and were kindly made available to this study.

The smaller dataset collected by the PhD student was sequenced for COI and CR and revealed similar patterns of diversity and structuring as the dataset collected by SAIAB. The SAIAB dataset was sequenced for CR only. This chapter therefore analyses the combined CR sequences of the smaller dataset collected by the PhD student and the larger dataset collected by SAIAB. Additionally, analysis of the smaller dataset was conducted and the results are contained in the appendix to this chapter.

Table 6.1. Sampling records of *L. nebulosus* collected across the SWIO, listing sampling location, country and sample size.

Location Name	Country	Code	Sample size
Stone town fish market	Zanzibar, Tanzania	ZT	10
Mahebourg fish market	Mauritius	Ma	9
Mahe fish market	Seychelles	S	13
Belo sur Mer	Madagascar	B	14
Andavadoaka	Madagascar	AN	8
Maputo fish market	Mozambique	MM	7

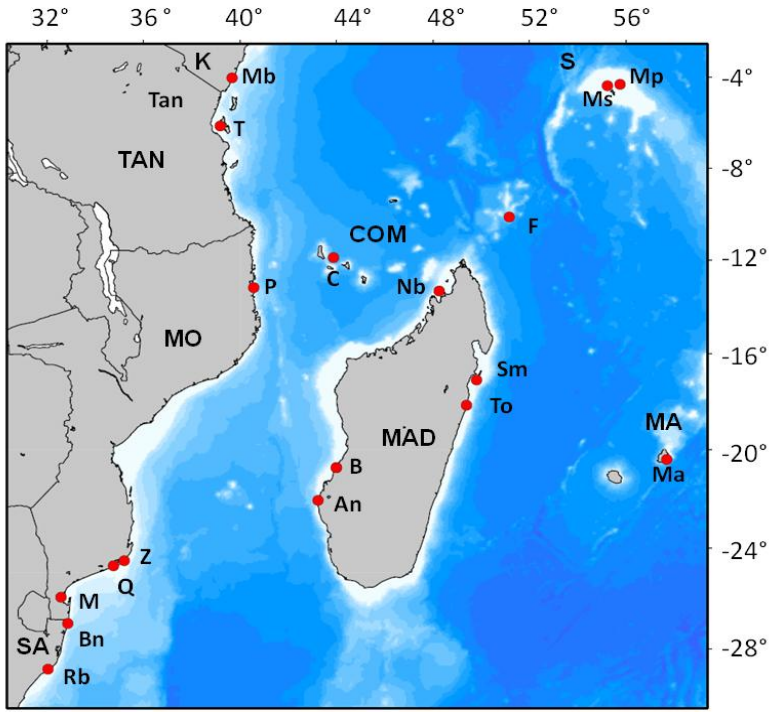

Figure 6.1. Sampling strategy for *L. nebulosus* across the SWIO. Sampling locations are indicated by red circles. Sampling locations include both sampling locations used during this PhD study and additional sampling locations used by SAIAB. Countries; K – Kenya, TAN-Tanzania, MO-Mozambique, SA-South Africa, COM-Comoros, MAD-Madagascar, MA-Mauritius, S-Seychelles. Sampling locations; Mb- Kenya, T-Tanzania, C-Comoros Island, Z-Zavora, Q-Quissico, M-Maputo, Bn-Banga Nek, Rb-Richard's Bay, B-Belo Sur Mer, An-Andavadoaka, Nb-Nosy Bé, F- Farquhar atoll, Ma–Mauritius, Ms-Mahe, Mp-Mahe plateau, To-Toamasina, Sm-St Marie.

Table 6.2. Sampling records of *L. nebulosus* collected across the SWIO by SAIAB listing sampling location, country and sample size. * indicate details of specific sampling locations were not provided for these localities.

Location Name	Country	Sample size
Mahe Plateau	Seychelles	7
Farquhar	Seychelles	30
Kenya*	Kenya	29
Belo sur Mer	West Madagascar	30
Comoros*	Comoros	5
Nosy Bé	West Madagascar	6
St Marie	East Madagascar	9
Tanzania*	Tanzania	30
Toamasina	East Madagascar	3
Zavora	Mozambique	4
Quissico	Mozambique	28
Richard's Bay	South Africa	55
Banga Nek	South Africa	3
Mauritius*	Mauritius	30

6.2.2 **Genetic screening**

Samples were sequenced for a 500bp fragment of the mitochondrial DNA COI gene using universal fish primers (Ward, et al. 2005) and 380bp of the control region using species-specific primers (Table 6.3) as described in Chapter 2.2.2.

Table 6.3. Mitochondrial DNA primers and annealing temperatures for amplification of COI mtDNA and HVR1 CR mtDNA in *L.nebulosus*.

Amplified Region	Primer name	Primer sequence	Annealing Temperature (°C)
COI	Fish F1	TCAACCAACCACAAAGACATTGGCAC	55
	Fish R1	TAGACTTCTGGGTGGCCAAAGAATCA	
	Fish R2	ACTTCAGGGTGACCGAAGAATCAGAA	
CR	LCR1F	CGGTCTTGTAAACCGGATGT	54
	LCR1R	GTCATGGCCCTGAAATAGGA	

Sequences were checked to confirm that samples were *L. nebulosus* using the BLAST search engine on GenBank.

6.2.3 **Mitochondrial DNA analyses**

Phylogenetic reconstruction

Phylogenetic relationships among sequences were inferred using network and tree-based approaches. Haplotype networks were constructed for COI, CR and the concatenated (COI and CR) mtDNA for the smaller study and for the CR from both the small and large study combined, referred to as the "combined CR mtDNA dataset" using maximum likelihood methods in PHYLIP DNAml (Felsenstein 1989) and subsequently constructed in Hapview (Ewing 2010) as described in Chapter 2.3.1 Sequence divergence between clades was calculated using the minimum p-distance in MEGA6 (Tamura, et al. 2013).

Reconstruction of phylogenetic relationships between individuals was carried out using Bayesian and maximum likelihood methods as described in Chapter 2.3.1. The most suitable model of nucleotide substitution was calculated using Bayesian Information Criterion (BIC) in jModelTest version 2.1.3 (Darriba, et al. 2012; Guindon and Gascuel 2003).

Bayesian Markov chain Monte Carlo (MCMC) analyses were carried out using the appropriate model and default settings in MrBayes v 3.2 (Ronquist, et al. 2012) as described in Chapter 2.3.1.

Maximum likelihood analyses were conducted using a neighbour-joining starting tree and implemented using the online version of PhyML 3.0 (Guindon, et al. 2010) as described in Chapter 2.3.1.

Trees were constructed in Figtree v.1.4 (Rambaut 2012).

Genetic diversity

Arlequin 3.5.1.2 (Excoffier and Lischer 2010) was used to calculate number of haplotypes (H), haplotype diversity (h) and nucleotide diversity (π) as described in Chapter 2.3.1.

Population structure and phylogeography

The statistical power of the data to detect genetic differentiation was assessed using the simulation method implemented in POWSIM (v 4.0) (Ryman and Palm 2006) adjusted for organelle (mtDNA) data (Larsson, et al. 2009) as described in Chapter 2.3.1.

Genetic differentiation was examined using F_{ST} (to investigate haplotype frequency distribution), Φ_{ST} to take into account divergence between haplotypes and the exact test of population differentiation calculated in Arlequin 3.5.1.2 (Excoffier and Lischer 2010). Both F_{ST} and Φ_{ST} were used in this case because when measuring population structuring, F_{ST} uses haplotype frequency occurring in the differing sampling locations to test for structuring between sites, whereas Φ_{ST} incorporates haplotype divergence (Excoffier, et al. 1992). Where multiple tests were conducted, significance levels were adjusted according to Bonferroni correction (Rice 1989).

Demographic history

Fu's Fs (Fu 1997), Tajima's D (Tajima 1989) and Ewens-Watterson tests (Slatkin 1994a) were calculated to test for deviations from neutrality using Arlequin 3.5.1.2 (Excoffier and Lischer 2010) as described in Chapter 2.3.1

Mismatch distribution was calculated separately for COI and the combined CR sequences for demographic expansion in Arlequin 3.5.1.2 (Excoffier and Lischer 2010) as described in Chapter 2.3.1.

To estimate time since expansion, substitution rates were calculated from divergence times. Mutation rates of 6×10^{-3} per million years (MYR) and 0.018 per MYR were used for COI and CR respectively. Mutation rates were calculated from divergence rates of 1.2% per MYR (Bermingham, et al. 1997) and 3.6% per MYR (Donaldson and Wilson 1999) for COI and CR respectively (see Chapter 2.2). Calculations of time since expansion using mismatch analyses were checked using the online spreadsheet http://www.uni-graz.at/zoowww/mismatchcalc/index.php (Schenekar and Weiss 2011).

Bayesian Skyline plots (Chapter 2.3.1), implemented in BEAST v.1.7.4 (Drummond, et al. 2012), were performed to explore demographic changes through time within each population as described in Chapter 2.3.1. To incorporate variation among related lineages, a local molecular clock was enforced, using a mutation rate of 0.018 MYR for CR.

Time since divergence

Time since divergence was estimated by calculating the percentage divergence between CR clades and then applying a divergence rate of 3.6% per MYR (Donaldson and Wilson 1999) as described in Chapter 2.3.1.

Additionally, a coalescent approach was employed to estimate time since most recent common ancestor (tmrca) from the mtDNA dataset. A total of 45 individuals of *L. nebulosus* (five from each sampling location) and 1 *L. obsoletus* individual, a closely related sister species to *L. nebulosus*, were used to estimate time since most recent common ancestor using the methods described in Chapter 2.3.1. Calibration of the molecular clock was conducted by fixing the lineage mutation rate at 1.8×10^{-5} per thousand years for the CR (calculated from 3.6% divergence rate per MYR) (Donaldson and Wilson 1999). Estimates of tmrca were relative only as no fossil or biogeographical data was available to use as node calibrators.

6.3 Results

Mitochondrial DNA analyses

The combined CR dataset (Table 6.4) contained 299 individuals. These samples were then subsequently grouped by region for analysis (Table 6.5). Given their small sample sizes and their position at the northern end of the Mozambique Channel, the Comoros, Pemba and Nosy Bé (Northwest Madagascar) samples were combined and considered as the North Mozambique Channel (NMC) for the purpose of examining genetic diversity and differentiation among sampling localities.

Two hundred and ninety nine 380bp sequences of the CR mtDNA were aligned producing a final alignment 416bp in length. The alignment contained 36 positions which were characterised by 26 gaps, representing insertion/deletion events, 192 variable sites and 71 invariable sites. These polymorphic sites (gaps included) defined 226 haplotypes, 195 of which were private haplotypes. The most suitable model of sequence evolution was found to be HKY + G.

Table 6.4. Combined CR mtDNA dataset displaying sampling regions, country and sample sizes of *L. nebulosus* across the SWIO. *Details of the specific collection localities were not provided for these sampling regions.

Location Name	Sampling Region/Country	Sample size
Mahe*	Seychelles	13
Mahe Plateau	Seychelles	7
Farquhar	Seychelles	29
Kenya*	Kenya	30
Tanzania*	Tanzania	30
Stone Town*	Zanzibar, Tanzania	10
Comoros*	Comoros	5
Nosy Bé	Northwest Madagascar	6
Andavadoaka	West Madagascar	14
Belo sur Mer	West Madagascar	8
St Marie	East Madagascar	9
Toamasina	East Madagascar	4
Maputo	Mozambique	7
Zavora	Mozambique	4
Quissico	Mozambique	27
Pemba	Mozambique	1
Richard's Bay	South Africa	58
Banga Nek	South Africa	3
Mauritius*	Mauritius	36

Table 6.5. *L. nebulosus* combined CR mtDNA dataset displaying sampling size by country.

Country	Sample size
Seychelles	49
Kenya	30
Tanzania	41
NMC	12
Mauritius	36
West Madagascar	22
East Madagascar	13
Mozambique	38
South Africa	58

Phylogenetic reconstruction

The haplotype network reconstruction using the combined CR mtDNA sequences from the two studies revealed the presence of two distinct and highly divergent clades, clade

1 and clade 2 (Figure 6.2). This complements the smaller study which also found two divergent groups. The 3 individuals from the smaller study in Group B cluster in clade 2, whilst the remaining 58 individuals from the smaller study cluster in clade 1. In the combined CR dataset 23 of the 38 individuals from the Mozambique sample belong to clade 2 whilst the remaining 15 individuals belong to clade 1. In the South African sample, clade 2 was also predominant with 55 individuals belonging to clade 2 and only 3 individuals belonging to clade 1. The minimum p-distance between the two clades was 0.16.

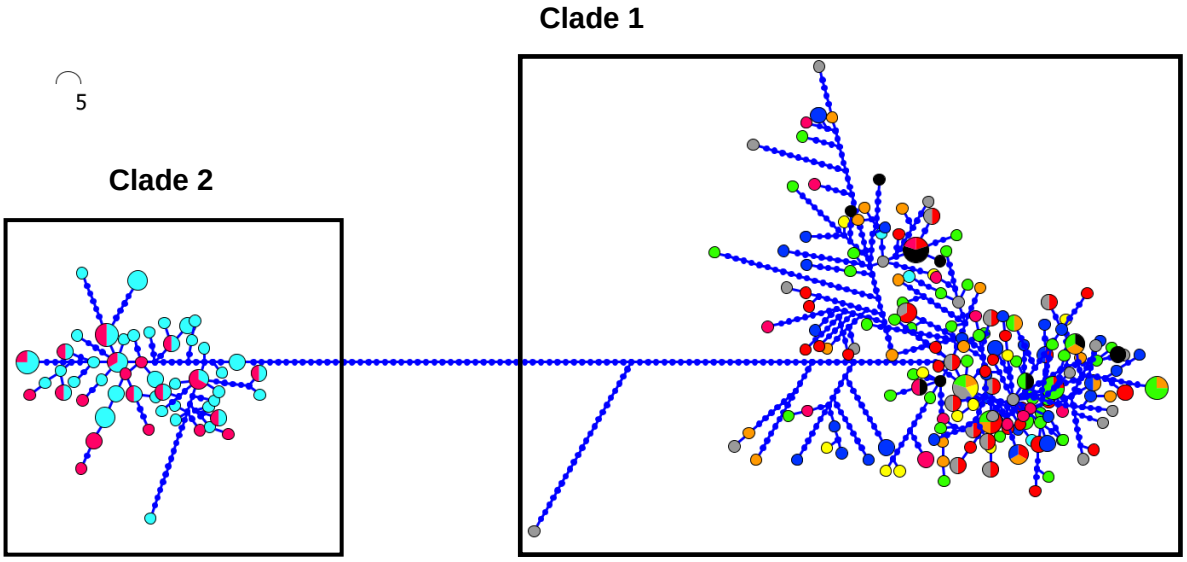

Figure 6.2. Haplotype network of *L. nebulosus* based on 380bp CR mtDNA. Branch lengths are proportional to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation; Tanzania, Mauritius, West Madagascar, East Madagascar, Seychelles, Mozambique, South Africa, Kenya, Northern Mozambique Channel (NMC).

Reconstruction of phylogenetic relationships among the CR dataset revealed two divergent monophyletic clades within *L. nebulosus* with high branch support (Figure 6.3).

Rooting was carried out using *L. obsoletus*, a closely related species (Figure 6.3). Minimum p-distance between the two clades was 0.156, equivalent to 15.6% divergence. Minimum p-distance between the two clades and *L. obsoletus* was 0.162 and 0.158 for *L. nebulosus* clades 1 and 2 respectively (Table 6.6). Observed branch lengths between clade 1 and clade 2 were of a similar length to those observed between *L. obsoletus* and *L. nebulosus*. Both branch lengths and p-distances suggest that the level of divergence between these clades is of a similar magnitude to that between two distinct species.

Table 6.6. Minimum p-distance between *L. obsoletus* and *L. nebulosus* clades 1 and 2, calculated based on 380bp of CR mtDNA.

	L. obsoletus	***L. nebulosus* clade 1**	***L. nebulosus* clade 2**
L. obsoletus	-		
***L. nebulosus* clade 1**	0.162	-	
***L. nebulosus* clade 2**	0.158	0.156	-

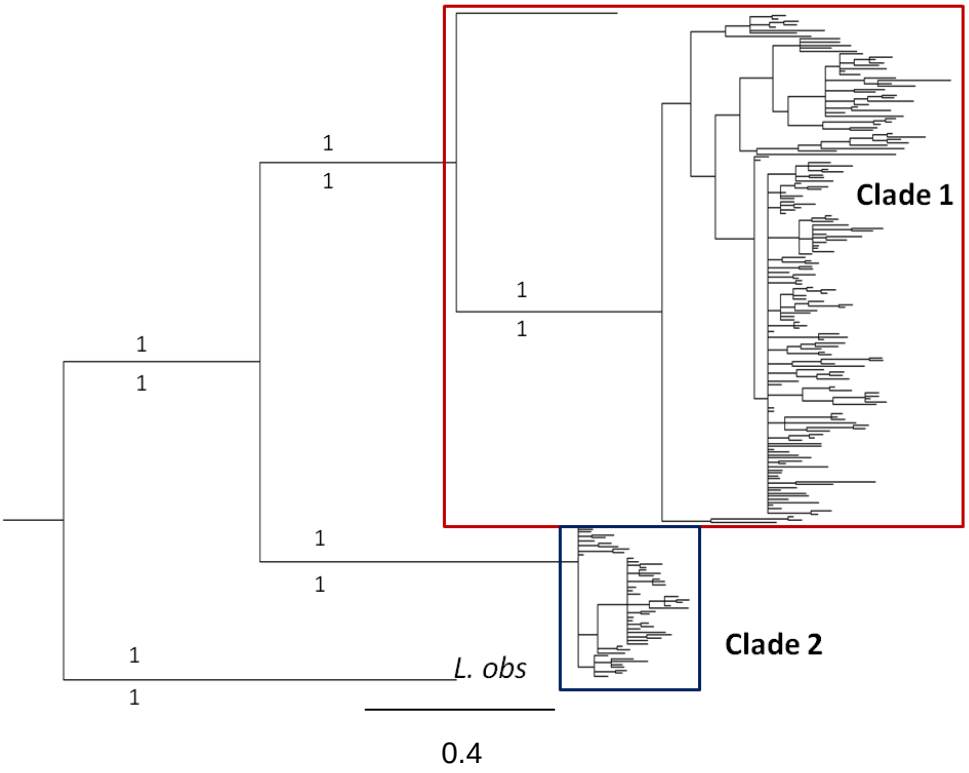

Figure 6.3. Reconstruction of phylogenetic relationships between clade 1 and clade 2 of *L. nebulosus* and *L. obsoletus* using CR mtDNA. Posterior probabilities above branches, maximum likelihood aLRT below branches.

Population diversity

Samples were split into their respective clades and population diversity statistics calculated.

The HKY model is not implemented in Arlequin 3.5.2 (Excoffier and Lischer 2010), so the more inclusive Tamura-Nei (TrN) model was therefore used to calculate genetic diversity statistics.

Clade 1 haplotype diversity was high h = 0.94 - 1.00 (Table 6.7). Nucleotide diversity was also high, (π = 0.037 – 0.050). West Madagascar had the highest haplotype and nucleotide diversity (h = 1.00, π = 0.05) (Table 6.7) of clade 1. Global tests indicated significant departure from neutrality for two out of three neutrality tests (Table 6.7).

Haplotype diversity of clade 2 was similar to that of clade 1, h = 0.94 - 1.00 for clade 1 and h = 0.99 for clade 2 (Table 6.8). Nucleotide diversity of both clades were fairly high with clade 1 having a considerably higher nucleotide diversity than clade 2; π = 0.037 - 0.050 and π = 0.018 - 0.020 (for clades 1 and 2 respectively). Clade 2 showed a significant departure from neutrality for the Ewens-Watterson test overall and at all sampling locations.

Table 6.7. Mitochondrial genetic diversity levels and neutrality tests for *L. nebulosus* clade 1 CR mtDNA; Sample size (n), number of haplotypes (H), number of private haplotypes (pHap), haplotype diversity (h), nucleotide diversity (π), Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs). Statistically significant estimates (p <0.05) are highlighted in bold.

Location	**n**	**H**	**pHap**	**h**	**π**	**F**	**D**	**Fs**
Tanzania	41	37	19	1.00	0.038	0.65	-1.26	**-19.42**
West Madagascar	22	22	15	1.00	0.050	Na	-0.87	**-9.09**
East Madagascar	13	13	13	1.00	0.038	Na	-0.74	**-4.25**
Seychelles	49	44	37	0.99	0.037	0.95	**-1.47**	**-24.24**
Kenya	30	29	17	1.00	0.047	**1.00**	**-1.50**	**-12.86**
NMC	12	9	3	0.94	0.042	0.90	0.73	0.84
Mozambique	15	14	11	0.99	0.043	**1.00**	-0.95	-2.98
South Africa	3	-	-	-	-	-	-	**-**
Mauritius	36	31	28	0.99	0.047	0.62	-1.05	**-11.53**
Overall	221	176	195	0.99	0.042	0.89	**-1.39**	**-23.74**

Table 6.8. Mitochondrial genetic diversity levels and neutrality tests for *L. nebulosus* clade 2 CR mtDNA; Sample size (n), number of haplotypes (H), number of private haplotypes (pHap), haplotype diversity (h), nucleotide diversity (π). Statistically significant estimates (p <0.05) are highlighted in bold.

Location	**n**	**H**	**pHap**	**h**	**π**	**F**	**D**	**Fs**
Mozambique	23	20	8	0.99	0.018	0.68	-1.16	**-11.24**
South Africa	55	43	32	0.99	0.020	0.46	-1.23	**-25.89**
Overall	78	52	40	0.99	0.019	0.90	-1.39	**-23.75**

Phylogeography and Population Structure

Statistical power of the CR dataset could not be ascertained using POWSIM because the maximum number of alleles/haplotypes per locus that the program can process is 100. The CR dataset contains 226 haplotypes.

Population structure

Tests of population differentiation F_{ST}, Φ_{ST} and the exact test showed high levels of genetic differentiation between the two clades (Table 6.9). Clade 2 was significantly differentiated from all sampling locations in clade 1 using Φ_{ST} (Table 6.10). Within clade 2 there was no significant structuring between the Mozambique and South Africa samples (F_{ST} = -0.003, Φ_{ST} = -0.019, P>0.05).

Table 6.9. Estimates of genetic differentiation based on analysis of 380bp CR mtDNA between *L. nebulosus* clade 1 and clade 2. F_{ST} and Φ_{ST} estimates are presented below and above the diagonal respectively. Statistically significant estimates (p <0.05) are highlighted in bold. Significance after Bonferroni correction is indicated by *.

	Clade1	**Clade 2**
Clade 1	-	**0.81***
Clade 2	**0.007***	-

Table 6.10. Estimates of genetic differentiation based on 380bp CR mtDNA among regional samples of the two clades, clade 1 (C1) and clade 2 (C2) of *L. nebulosus*. F_{ST} and Φ_{ST} estimates are presented below and above the diagonal respectively. Statistically significant estimates ($p < 0.05$) are highlighted in bold. Samples significant after Bonferroni correction are indicated by *. Mo-Mozambique, SA-South Africa, WM-West Madagascar, NMC-Northern Mozambique Channel, T-Tanzania, K-Kenya, EM-East Madagascar, S-Seychelles, Ma-Mauritius.

		Clade 1									**Clade 2**	
		Mo-SA	**WM**	**NMC**	**T**	**K**	**EM**	**S**	**Ma**		**Mo**	**SA**
	Mo-SA	-	0.003	0.072	0.004	0.004	0.003	**0.006**	0.007		**0.860***	**0.874***
Clade 1	**WM**	-0.021	-	**0.063**	-0.008	-0.018	-0.012	0.010	0.009		**0.832***	**0.855***
	NMC	0.014	**0.025**	-	**0.025**	**0.030**	**0.030**	**0.028***	**0.032***		**0.871***	**0.882***
	T	0.004	0.015	**0.140**	-	-0.007	0.005	0.008	0.003		**0.849***	**0.863***
	K	-0.012	-0.002	**0.089**	-0.006	-	-0.004	0.003	**0.005**		**0.827***	**0.858***
	EM	-0.022	-0.004	**0.103**	0.003	-0.014	-	0.002	0.007		**0.873***	**0.882***
	S	0.001	-0.003	**0.158**	0.003	0.008	0.013	-	**0.021**		**0.85***	**0.863***
	Ma	-0.005	0.002	**0.097**	**0.005**	0.001	0.004	**0.007**	-		**0.826***	**0.847***
Clade 2	**Mo**	0.009	0.006	**0.035**	**0.008**	**0.007**	0.006	**0.009**	**0.010**		-	-0.019
	SA	**0.008**	0.005	**0.033***	**0.007***	**0.006**	0.005	**0.008***	**0.009***		-0.003	-

Population structuring within clade 1 was then examined. Estimates of Φ_{ST} vealed some structuring within the North Mozambique Channel within clade 1. However. this structuring was not significant after Bonferroni correction (Table 6.10). Hierarchical population structuring (AMOVA) between clade 1 and clade 2 revealed 82.23% ($P < 0.05$) variation was accounted for between clades whilst 17.43% ($P = 0.019$) of variation was accounted for within clades.

Phylogeography

Clade 1 encompasses sampling locations from the entire SWIO whilst clade 2 contains individuals from Southern Mozambique and South Africa only (Figure 6.4). The proportion of the two clades in the Mozambique and South Africa sample varied, although clade 2 was the dominant clade present at each sampling location (Figure 6.4). The Southern Mozambique sampling set contained 15 individuals from clade 1, 23 individuals from clade 2 and the South Africa sampling set contained 3 individuals from clade 1 and 55 individuals from clade 2. Sampling locations northeast of these locations in West Madagascar (Andavadoaka and Belo sur Mer) possessed no individuals belonging to clade 2.

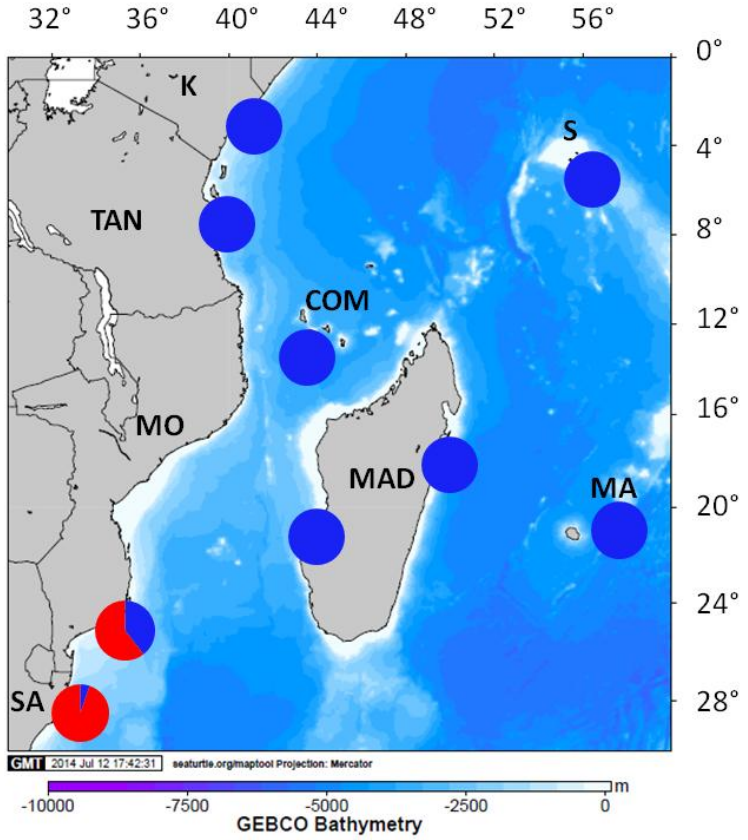

Figure 6.4. Map of the SWIO showing the relative frequency of *L. nebulosus* clade 1 and clade 2 CR mtDNA at sampling locations. Blue represents clade 1, red represents clade 2. SA- South

Africa, MAD-Madagascar, MO-Mozambique, MA-Mauritius, COM-Comoros Islands, TAN-Tanzania, K-Kenya, S-Seychelles.

Demographic history

The dataset for clade 1 and clade 2 had highly significant Fu's Fs test values, suggesting a possible past population expansion (Table 6.11).

Table 6.11. Neutrality tests and mismatch analysis for *L. nebulosus* CR mtDNA. Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs). T time since expansion in mutational units (95% confidence intervals shown in brackets below), θ_0 population size before expansion, θ_1 population size after expansion, SSD sum of squared deviations, TSE time since expansion in years (95% confidence intervals shown in brackets below). Statistically significant estimates ($p < 0.05$) are highlighted in bold.

Clade	F	D	Fs	Tau	θ_0	θ_1	SSD	TSE
1	0.007	**-1.39**	**-23.76**	8.7 (4.96-25.00)	10.90	99999.00	0.0042	630,000 (360,000-1,830,000)
2	0.025	-1.19	**-24.90**	6.3 3.31-18.20	3.33	99999.00	0.0027	444,100 (242,000-1,330,000)

Analysis of the clade 1 CR dataset revealed a multimodal mismatch analysis (Figure 6.5). The Bayesian skyline plot for clade 1 shows a two- step population expansion (Figure 6.6), in keeping with the multimodal mismatch distribution. The plot indicates a gradual expansion in population size from 750,000 – 400,000 years ago, with a second more recent and faster population expansion dating from 125,000 – 80,000 years ago.

The mismatch distribution histograms revealed very different topographies for clade 1 and 2 (Figures 6.5, 6.7). Clade 2 displayed a single gradual demographic expansion rather than multiple population expansions. Bayesian skyline plots estimate the date of this expansion from 400,000 – 100,000 years ago, slightly earlier than the expansion estimate of 444,100 years ago, generated from the mismatch distribution (Figure 6.8).

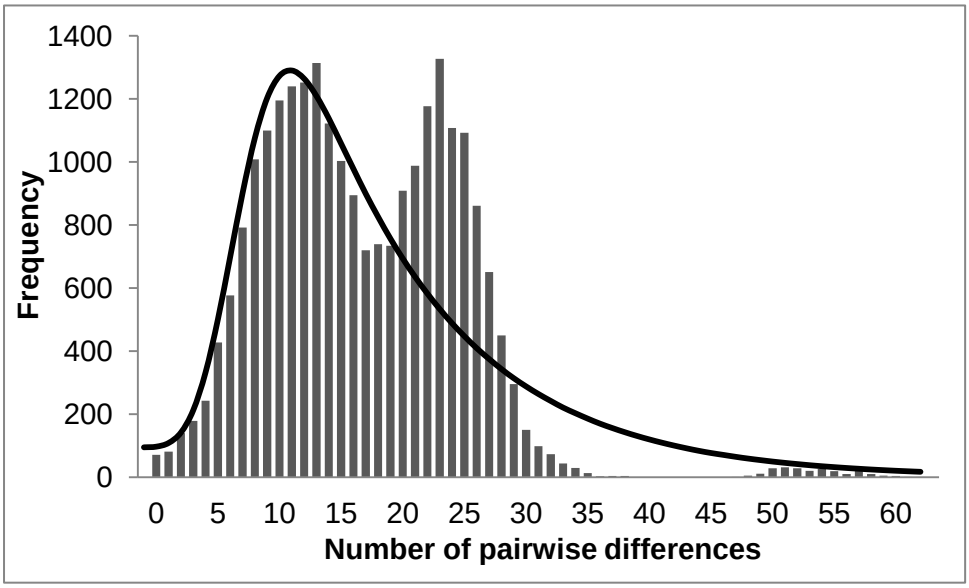

Figure 6.5. Mismatch distribution histogram for *L. nebulosus* clade 1 CR mtDNA across the SWIO. Filled bars indicate the observed frequency of pairwise distribution, black line indicates the expected distribution under a model of sudden demographic expansion.

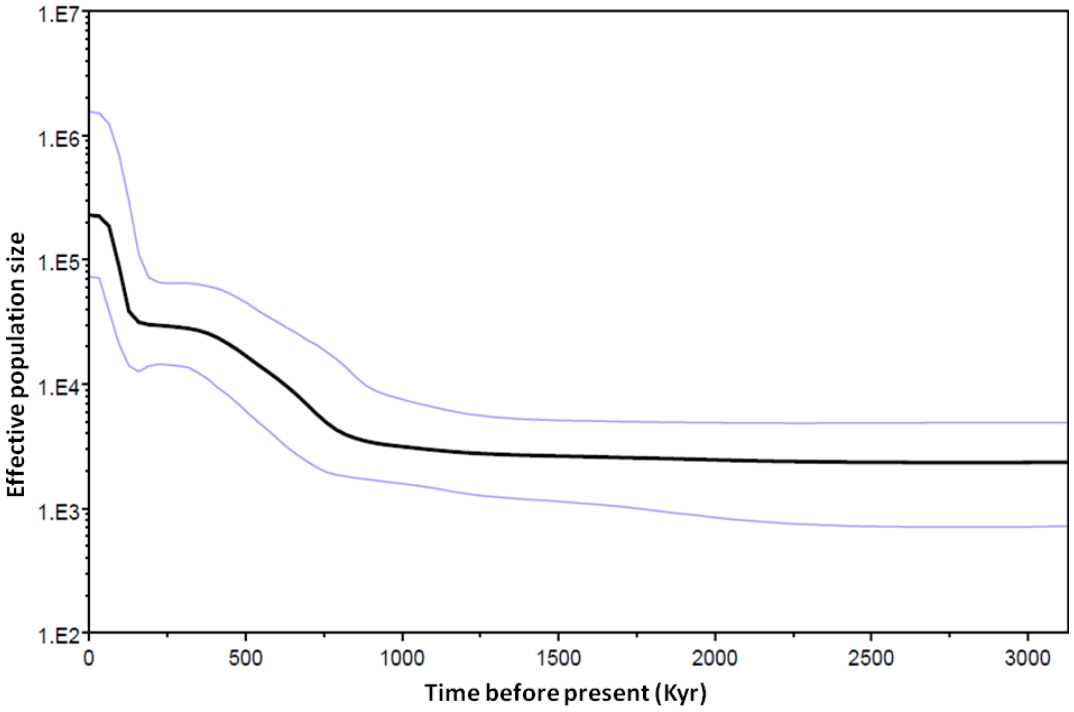

Figure 6.6. Bayesian skyline plot of *L. nebulosus* clade 1 CR mtDNA. The y-axis is the product of maternal effective size and generation time. The x-axis is the time from present in units of thousands of years. The thick solid line is the median estimate and the thin lines (blue) show the 95% highest posterior density limits.

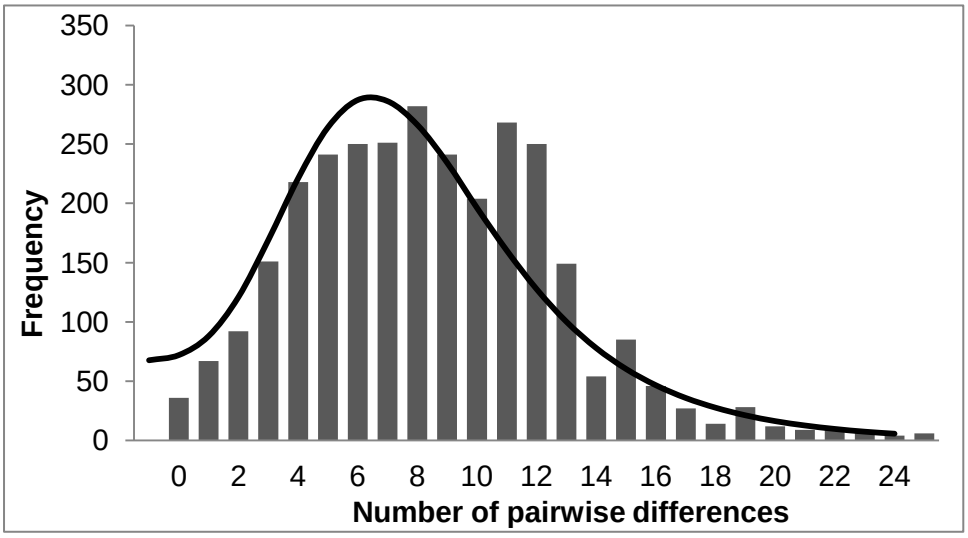

Figure 6.7. Mismatch distribution histogram for *L. nebulosus* clade 2 CR mtDNA. Filled bars indicate the observed frequency of pairwise distribution, black line indicates the expected distribution under a model of sudden demographic expansion.

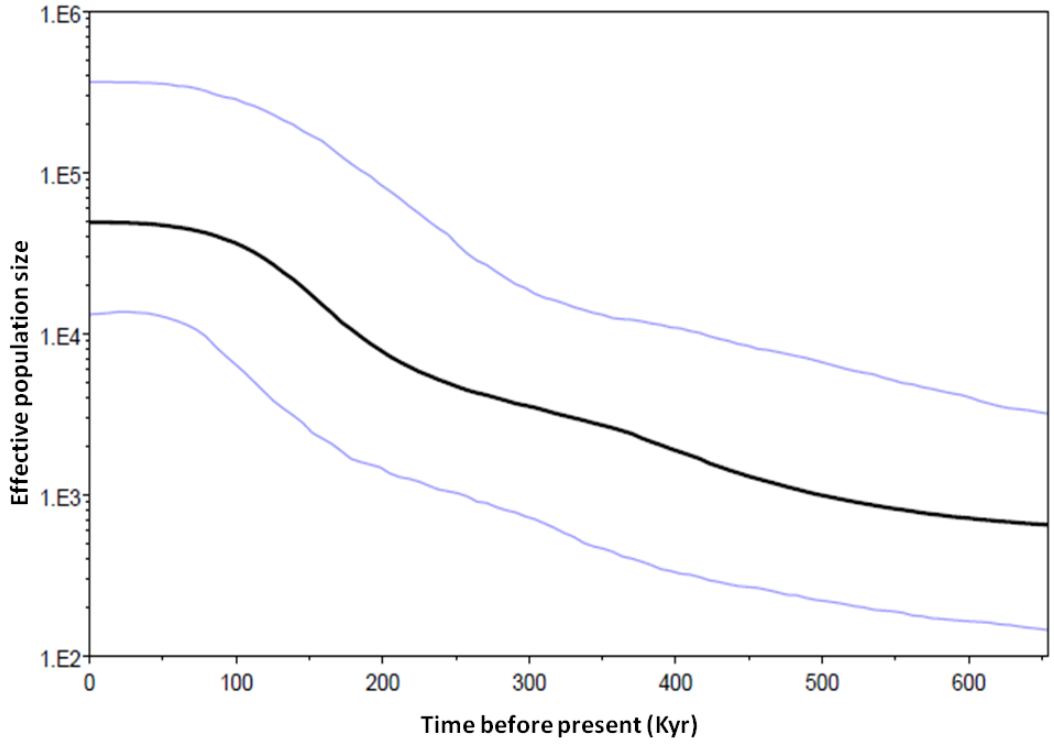

Figure 6.8. Bayesian skyline plot of *L. nebulosus* clade 2 CR mtDNA. The y-axis is the product of maternal effective size and generation time. The x-axis is the time from present in units of thousands of years. The thick solid line is the median estimate and the thin lines (blue) show the 95% highest posterior density limits.

Time since divergence

Time since divergence of clade 1 from clade 2 calculated using 3.6% sequence divergence per MYR resulted in an approximate divergence time of 4.4 MYR ago. Calculations in BEAST under a local clock revealed an estimated divergence time of

8.2 MYR ago, with a 95% highest posterior density (HPD) of 5.7 - 10.6 MYR ago. Absolute divergence times could not be calculated in BEAST due to the absence of suitable node calibrations.

6.4 Discussion

The dataset supports the existence of two highly divergent clades of *L. nebulosus* in the SWIO; clade 1 which is common across the SWIO and along the east coast of Africa to the southern end of the Mozambique Channel and clade 2, whose distribution across the SWIO is restricted to the south African and southern Mozambican coastline of east Africa. This was supported by F_{ST}, Φ_{ST} and the exact test, reconstruction of haplotype networks and reconstruction of intra-specific phylogenies.

Estimates of divergence of the two clades vary from 4.4 – 10.6 million years ago. The current distribution of clade 1 *L. nebulosus* is tropical Indian Ocean (29 - 30°C) whereas distribution of clade 2 *L. nebulosus* is subtropical (24 - 26°C). It is possible that the divergence of these lineages was caused by past sea surface temperature in the tropical Indian Ocean. Prior to 4 million years ago, New Guinea was displaced to the south of its present- day position and the island of Halmahera lay largely below sea level (Cane and Molnar 2001). The majority of water flow into the Indian Ocean was through the Indonesian seaway between New Guinea and Sulawesi, sourced from the warm South Pacific rather than the cooler North Pacific from which it is sourced today. This warm water was carried across the Indian Ocean to the Western Indian Ocean by the South Equatorial Current, which has been fairly constant in its flow for the last 10 million years (Cane and Molnar 2001). Consequently, the central Indian Ocean, which corresponds to the current distribution of clade 1, was 2 - 3°C warmer than current day sea surface temperature. In the subtropical Indian Ocean however, sea surface temperature was the same as present day sea temperature (24 - 26°C) (Cane and Molnar 2001). This considerable difference in sea surface temperature between the tropical and subtropical Indian Ocean may have driven the divergence of these two lineages. When New Guinea was displaced northwards and the flow through the Indonesian seaway was displaced to its present day position, approximately 4 million years ago, central Indian Ocean sea surface temperatures cooled by 2 - 3°C (Cane and Molnar 2001) to its present day temperature. However, from the estimates of divergence obtained in this study, it is possible that the two lineages had already diverged by this time.

Population structure across the SWIO

In the introduction it was hypothesised that the current regime of the Mozambique Channel may act as a barrier to gene flow for *L. nebulosus*, causing the divergence of lineages north and south of the Mozambique Channel. Two clades have been found; one with a more northern distribution in the channel whilst the second clade has a more southerly distribution south of the Mozambique Channel. However, analyses have shown that the divergence of these two stocks occurred considerably earlier than the establishment of the current flow regime in the channel. The current flow regime in the Mozambique Channel has therefore not acted as a driver of population differentiation in *L. nebulosus*.

No significant structuring was observed within clade 2. Within clade 1 the Northern Mozambique Channel locality was shown to be slightly differentiated from all other sampling locations except the combined South Africa/Mozambique sample using Φ_{ST} and F_{ST} although this was largely non-significant after Bonferroni correction. Structuring between NMC and the Seychelles and Mauritius was however still significant after Bonferroni correction for Φ_{ST} (Φ_{ST} = 0.028 and 0.032 respectively). Bonferroni correction is often considered too conservative (Cabin and Mitchell 2000; Narum 2006) so there may be some population structuring between the NMC site and other sampling locations within clade 1. However, sample size of the NMC sample was limited to 12 individuals. Further sample collection and analysis would therefore be necessary to elucidate the extent of structuring in this region.

Previous studies of *Lethrinus* population structure and phylogeography show minimal population structure over considerable distances. *L. nebulosus* populations were found to be panmictic across distances of 1400 – 2080 km in Northwestern Australia using allozyme variability (Johnson, et al. 1993). Similarly, no genetic structuring of *L. miniatus* was observed using eight microsatellite loci and six sampling locations from three geographic regions of the Great Barrier Reef (GBR) over distances of approximately 500-1050 km.

Differentiation between the two clades when measured by F_{ST} did not show all populations within clade 1 to be significantly differentiated from clade 2. Only 2 populations within clade 1 were genetically differentiated from clade 2 after Bonferroni correction using F_{ST}. Additionally, F_{ST} between the two clades was only 0.007 ($p < 0.05$), despite no haplotypes being shared between the two clades. This is due to the design of F_{ST} itself. F_{ST} was designed by Wright (1949) to test population structuring using low diversity markers such as allozymes. However as molecular techniques have advanced, highly diverse markers such as the control region have been discovered

which, when diversity levels are high within populations, often result in drastically reduced values of F_{ST} (Charlesworth 1998; Jost 2008). The reason for this is the way that these estimates of differentiation are calculated. Additive partitioning of diversity into mean within subpopulation diversity and between subpopulation diversity means these two are not independent. Therefore when subpopulation diversity is large, between subpopulation diversity and hence F_{ST} has to be small regardless of the differentiation between subpopulations (Jost 2008).

The southern distribution of *L. nebulosus* is recorded as occurring in KwaZulu-Natal, South Africa, but occurring less frequently south of this region (Fischer and Bianchi 1984; Smith 1959; van der Elst, et al. 2005). It is therefore likely that the core distribution of clade 2 is centred around the South African coastline, near to KwaZulu-Natal. However, no sampling locations on the west side of the Mozambique Channel from Mid-Mozambique were collected during this study and so the true extent of the clade 2 distribution along the east African coastline is still undetermined. Further sampling from along the Mozambican coastline needs to be undertaken to ascertain the northern limit of clade 2's distribution.

The present study suggests that the southern limit of clade 1 is likely to be North-eastern South Africa – KwaZulu- Natal. Clade 1 was present in both Mozambique and South African sampling localities but only 3 individuals from clade 1 were present in a sample size of 55 individuals for South Africa. Interestingly, the 3 individuals representative of clade 1 in the South African sample were all from Banga Nek, the northernmost sampling locality in South Africa. No clade 1 individuals were found in the much larger sample from Richards Bay south of Banga Nek and it may be that this is the southern limit of this clade.

Evolutionary history

The two clades showed contrasting demographic histories. Both clades displayed significant values for Fu's Fs Test (Fu 1997) and Ewens-Watterson test which are sensitive to past population expansion and mismatch analysis detected the presence of a population expansion in both clades.

A multimodal mismatch distribution for clade 1 was observed which is typically associated with a population which has been stable for a considerable period of time or has undergone multiple past population expansions. Bayesian skyline plots of clade 1 suggested the existence of at least two past demographic expansions; a gradual expansion from 750,000 – 200,000 years ago and a second more recent demographic expansion dating from 80,000 - 180,000 years ago. Mismatch analysis estimated a

single population expansion 630,000 years ago. However, 95% confidence intervals for the mismatch analysis are large, ranging from 360,000-1,830,000 years ago. This may be due to the limitations of mismatch analysis itself which assumes a single step-wise population expansion (Schneider and Excoffier 1999). BSP however suggests a multi-step expansion and it may be therefore that the assumptions of the mismatch analysis have been violated, which may explain the large confidence intervals.

The dating of the population expansions of clade 1 places both demographic expansions in the Pleistocene, a period of considerable global climate oscillation. In the SWIO during this period a number of glacial and interglacial stades occurred, causing sea-level to rise and fall approximately 120 m. It is possible that populations in the SWIO during glacial periods were limited by available suitable habitat. The SWIO has very limited continental shelf area in comparison with other ocean regions, with the majority of the shelf 70 m deep in current sea level (Qasim 1977). During low periods of past sea level the majority of continental shelf would have been subaerially exposed, such as the Seychelles Bank (Badyukov, et al. 1989). Therefore during interstadial events where sea level increased by up to 120 m, it could be expected that *L. nebulosus*, which inhabit areas of continental shelf, would experience a demographic expansion due to increased habitat availability.

In contrast to clade 1, clade 2 exhibited evidence of a single population expansion. The mismatch distribution was unimodal and expansion was estimated to have occurred 444,000 years ago (95% confidence intervals estimated expansion from 242,000 to 1,330,000 years ago). The Bayesian skyline plot supported this date of expansion which may have coincided with an interglacial period in the Pleistocene where sea levels increased by 120 m. This would represent a considerable increase in available habitat which may explain why a population expansion was observed at this time.

Pleistocene climate-induced changes have been reported to contribute to population demographic and genetic sub-structuring in several marine taxa. Climate fluctuations have been shown to have severely reduced *Engraulis mordax* and *Sardinops sagax* populations in the eastern Pacific (Lecomte, et al. 2004), influenced gene flow and connectivity in *Homarus americanus* in the northwestern Atlantic (Kenchington, et al. 2009), in *Cephalopholis argus* (Epinephelidae) across the Indo-Pacific barrier (Gaither, et al. 2011) and in *Ammodytes personatus* in the Northwestern Pacific (Han, et al. 2012). It is therefore possible that past climatic events in the Indian Ocean have shaped *L. nebulosus'* past evolutionary history.

Genetic evidence for cryptic species

The genetic markers used in this study reflect patterns of gene flow and connectivity from an evolutionary and historical perspective. MtDNA analysis indicates two distinct clades along the South African coast; clade 1 with a large geographical range across the SWIO, and clade 2 which is restricted to the coastline of South Africa and southern Mozambique. Despite the overlapping ranges of these two clades in South Africa and southern Mozambique, the ancestral divergence of these lineages has been maintained. The high level of divergence and phylogenetic reconstruction suggest these two lineages may be different species. All individuals identify most closely with *L. nebulosus* sequences on GenBank and individuals from each clade can be morphologically identified as *L. nebulosus* from the blue or white lateral scales, blue lines radiating from the eyes and no red colouration on the operculum (Fischer and Bianchi 1984).

There are no obvious morphological features that can be used to distinguish between the two clades. It is therefore likely that clade 2 represents a previously unrecognised cryptic species, a fairly common occurrence in marine fish (Bickford, et al. 2007; Colborn, et al. 2001; Griffiths, et al. 2010; Kon, et al. 2007). Cryptic species of *Lethrinus* sp have been recorded in the literature, for example in the Indo-West Pacific (Borsa, et al. 2013) and previously unrecognized species of Lethrinus have also been discovered (Carpenter and J.E. 2003) in the West Pacific and East Indian oceans.

That morphological analyses did not allow differentiation of the two lineages is not particularly unexpected as this has been found in multiple fish species (Colborn, et al. 2001; Griffiths, et al. 2010; Santos, et al. 2006; Watanabe, et al. 2010), suggesting that in many fish species morphological changes occur at a slower pace compared with genetic variation (Colborn, et al. 2001). The existence of cryptic species may be due to the relative stability of the marine environment on evolutionary timescales. This is because despite the short-term location of particular habitats changing with climatic fluctuations, these habitats still exist across a broad geographic scale for millions of years (Jackson and Sheldon 1994).

Implications for fisheries management

Results in the present study suggest that *L. nebulosus* exhibits two genetically divergent lineages across the SWIO. The absence of shared haplotypes, maintenance of these divergent lineages and phylogenetic reconstruction suggests no effective gene flow between the two lineages. As such, both *L. nebulosus* populations should be

managed independently as Evolutionary Significant Units - ESU as defined by Moritz (1994).

Both lineages of *L. nebulosus* exhibit high genetic diversity levels suggesting that over-exploitation has not yet caused a significant decrease in genetic diversity. Clade 2 of *L. nebulosus* displays lower nucleotide diversity than clade 1 which may be due to the differing demographic histories experienced by the two clades. However there are no limitations to captures of *L. nebulosus* in the SWIO which may lead to future exploitation of these populations.

Accurate identification of fish to species level is a prerequisite for fisheries management (Waples, et al. 2008). The taxonomic status of these clades requires further investigation and is currently being investigated by SAIAB using additional genetic data and morphological and morphometric studies. It is hoped that, despite no obvious distinguishing morphological features having been recognized to date between the two lineages, this further research may find morphological and morphometric characters which could be used to distinguish these two lineages. This would be the first step towards the future sustainable management of these two species.

Appendix

COI Mitochondrial DNA analyses

Sixty one individuals were sequenced for 500bp of COI and 380bp of CR mtDNA in the smaller study, resulting in 24 and 57 haplotypes respectively. The K80 nucleotide substitution model was the most suitable model of sequence evolution.

Phylogenetic reconstruction

The COI, CR and concatenated (COI and CR) mtDNA haplotype networks (Figure 6.9 A,B,C) for the smaller study each show the same 3 individuals (Group B) which all possess unique haplotypes and are separated by a considerable number of mutational steps; a minimum of 24 for the COI haplotype network, 63 for CR haplotype network and 88 for the concatenated mtDNA haplotype network, from the remaining haplotypes (Group A). These individuals display a 0.05, 0.12 and 0.089 minimum p-distance sequence divergence from the haplotypes in group A, for COI, CR and the concatenated mtDNA respectively. This distinct and highly divergent haplotype group is comprised of three individuals all collected from Maputo, Mozambique. Additional individuals collected from Maputo, Mozambique possess haplotypes which cluster in Group A.

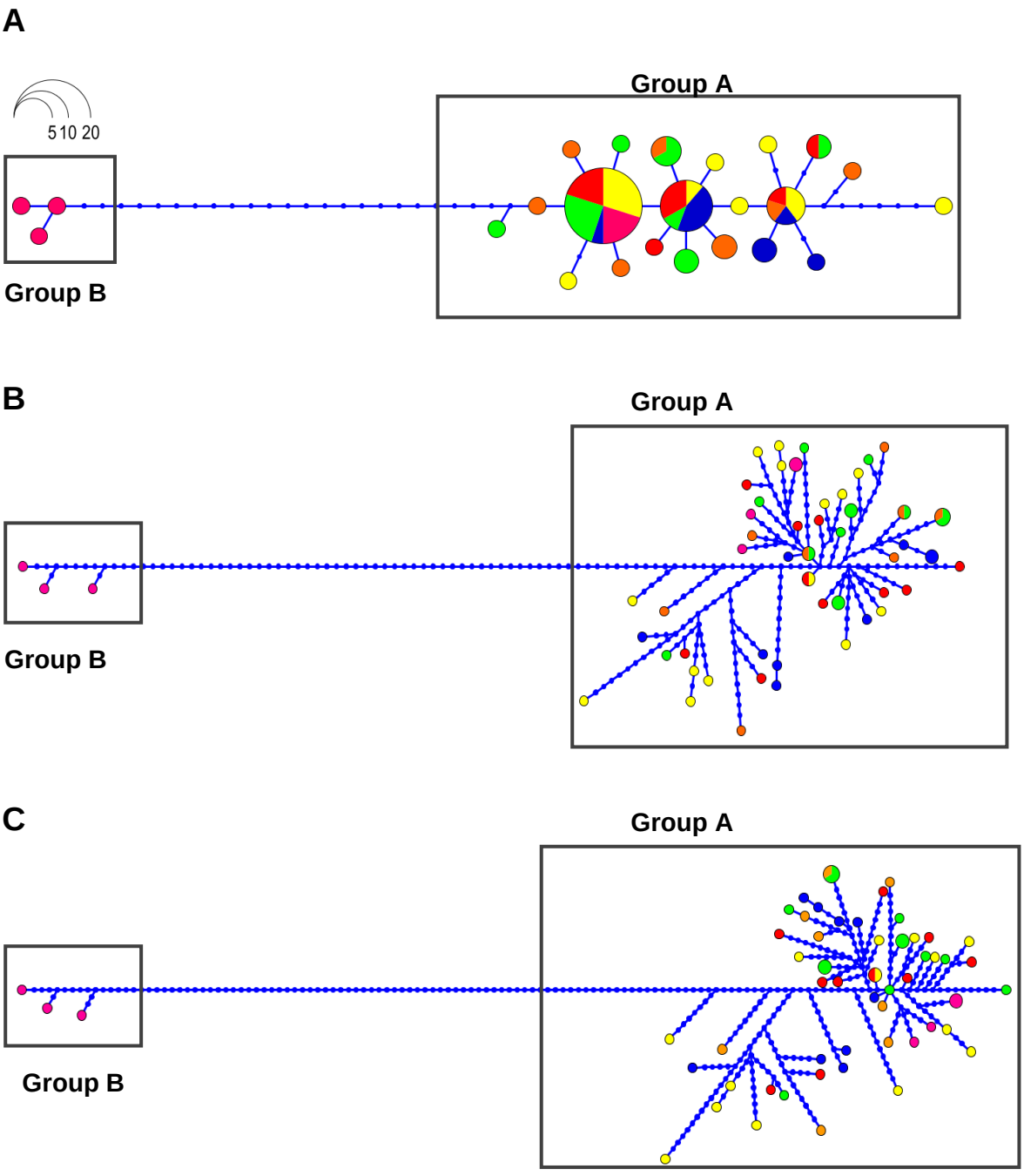

Figure 6.9. Haplotype networks of *L. nebulosus* based on **A**: 500bp of COI mtDNA. **B**: 380bp of CR mtDNA. **C:** 880bp concatenated COI and CR mtDNA, from the smaller study. Branch lengths are proportional to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation; Tanzania, Mauritius, Andavadoaka, Madagascar, Belo Sur Mer, Madagascar, Mahe, Seychelles, Maputo, Mozambique.

Haplotype networks were then constructed for COI, CR and concatenated mtDNA from the smaller study removing the 3 divergent haplotypes to look in detail at the haplotype relationships in clade 1 (Figure 6.10). The COI haplotype network reveals three main haplotype groups, separated by one to two mutational steps (Figure 6.10.A). As would be expected from mtDNA, the CR haplotype network shows CR haplotypes corresponding to the three COI haplotype groups clustering more closely together than

CR haplotypes possessing a different COI haplotype (Figure 6.10.B). The concatenated mtDNA haplotype network (Figure 6.10.C) again shows haplotypes clustered together according to their COI haplotype. Samples from Maputo are only present in one of the haplotype groupings. All other sampling locations are represented in all three haplogroups.

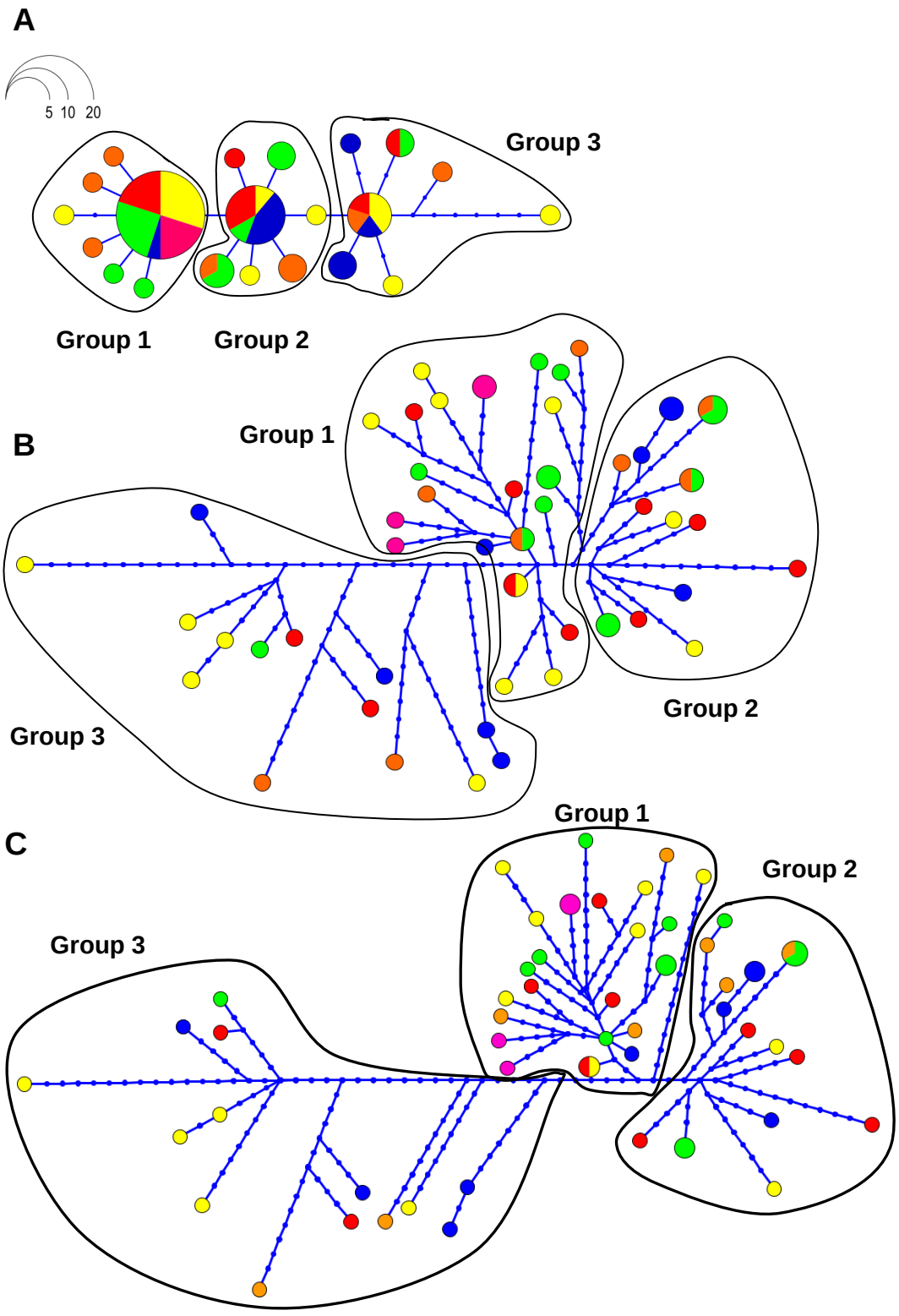

Figure 6.10. Haplotype networks of *L. nebulosus* based on **A:** 500bp COI mtDNA. **B:** 380bp CR mtDNA from the smaller study. **C** 880bp concatenated mtDNA Branch lengths are proportional

to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation; Tanzania, Mauritius, Andavadoaka, Madagascar, Belo Sur Mer, Madagascar, Mahe, Seychelles, Maputo, Mozambique.

Population Diversity

Samples were split into their respective clades and population diversity statistics calculated for each clade for COI and the combined CR dataset. Population diversity statistics are not displayed for the CR dataset of the smaller study as they are included in the combined CR dataset analysed in the main part of this chapter.

Due to the small number of individuals (3) belonging to clade 2 in the COI dataset, population diversity statistics were not calculated for these three individuals. Population diversity statistics were also not calculated for Mozambique clade 1 COI dataset as there are only 4 individuals in the sample, all of which possess the same haplotype.

Clade 1 COI haplotype diversity was high, h = 0.69 - 0.96 (Table 6.12). Nucleotide diversity was also high, (π = 0.0041 – 0.0076). Belo sur Mer had the highest haplotype and nucleotide diversity of both the COI dataset and the CR dataset in the corresponding region to Belo sur Mer - West Madagascar (h = 1.00, π = 0.05) (Table 6.12). Departure from neutrality was observed in the COI dataset for Andavadoaka and Belo sur Mer for the Ewens-Watterson test (Table 6.12). Global tests indicated significant departure from neutrality for the COI dataset for all three neutrality tests and for two out of three neutrality tests for the CR dataset (Table 6.12).

Table 6.12. Mitochondrial genetic diversity levels and neutrality tests for *L. nebulosus* COI mtDNA; Sample size (n), number of haplotypes (H), number of private haplotypes (pHap), haplotype diversity (h), nucleotide diversity (π), Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs). Statistically significant estimates ($p < 0.05$) are highlighted in bold.

Location Name	**Country**	**n**	**H**	**pHap**	**h**	**π**	**F**	**D**	**Fs**
Stone Town fish market	Zanzibar	10	5	1	0.80	0.0041	0.70	-0.74	-0.55
Mahebourg fish markets	Mauritius	9	4	2	0.69	0.0049	0.84	0.47	0.82
Mahe fish market	Seychelles	13	7	3	0.85	0.0040	0.80	-1.18	-2.26
Belo sur Mer	Madagascar	8	7	5	0.96	0.0076	**0.98**	-1.29	**-2.58**
Andavadoaka	Madagascar	14	8	5	0.82	0.0069	**1.00**	-0.88	-1.50
Overall	-	54	31	16	0.85	0.0053	**0.99**	**-1.91**	**-11.61**

Phylogeography and population structure

POWSIM results indicated that the COI dataset had a 10% and 28% chance, based on Chi^2 and Fisher tests respectively, of detecting an F_{ST} of 0.01. At an F_{ST} of 0.05 the COI dataset of both tests had a 100% likelihood of detecting population structuring (Table 6.13). Statistical power of the CR dataset could not be ascertained using POWSIM because the maximum number of alleles/haplotypes per locus that the program can process is 100. The CR dataset contains 226 haplotypes.

Table 6.13. Statistical power of the *L. nebulosus* COI mtDNA dataset to detect population differentiation at various levels of F_{ST} using tests based on Chi^2 and Fisher exact methods inferring a population size of 10,000.

	Statistical Power	
F_{ST}	**Chi^2**	**Fisher**
0.00 (type 1 error)	0.0008	0.036
0.001	0.008	0.058
0.01	0.098	0.28
0.05	1.00	1.00

Population structuring within clade 1 was then examined. COI F_{ST} values indicated significant structuring within clade 1 between Mauritius and all other sites which, after Bonferroni correction, was still significant between Mauritius and Maputo, Mozambique (Table 6.14). Belo sur Mer was also differentiated from Mozambique after Bonferroni correction. Φ_{ST} indicated some structuring between Mauritius and Maputo and between Mauritius and the Seychelles but this was not significant after Bonferroni correction. The exact test also indicated some structuring between Mauritius and the Seychelles after Bonferroni correction (Table 6.15).

Table 6.14. Estimates of genetic differentiation based on 510bp COI mtDNA within *L. nebulosus* clade 1. F_{ST} and Φ_{ST} estimates are presented below and above the diagonal respectively. Statistically significant estimates ($p<0.05$) are highlighted in bold. Values with * indicate significance after Bonferroni correction. ZT-Zanzibar, MA-Mauritius, S-Seychelles, AN-Andavadoaka, B-Belo sur Mer, MM-Maputo Mozambique.

	ZT	**MA**	**S**	**AN**	**B**	**MM**
ZT	-	0.08	-0.02	-0.41	-0.41	0.12
MA	0.09	-	**0.22**	0.05	0.04	**0.41**
S	-0.01	**0.19***	-	0.02	0.01	0.02
AN	-0.02	**0.19**	-0.01	-	-0.31	0.03
B	**0.11**	**0.16**	**0.08**	**0.01**	-	0.07
MM	0.18	**0.54***	0.14	0.11	**0.40***	-

Table 6.15. Exact test of differentiation based on *L. nebulosus* 510bp COI mtDNA. Statistically significant estimates (p <0.05) are highlighted in bold. Values with * indicate significance after Bonferroni correction.

	ZT	MA	S	AN	B	MM
ZT	0.097 ±0.001					
S	0.500 ±0.002	**0.003* ±0.001**				
AN	0.702 ±0.002	**0.009 ±0.001**	0.337, ±0.002			
B	**0.0166 ±0.002**	**0.016 ±0.001**	**0.023 ±0.001**	**0.024 ±0.001**		
MM	0.257 ±0.002	**0.002* ±0.001**	0.413 ±0.003	0.576 ±0.002	**0.040 ±0.001**	

Demographic history

Significant departures from neutrality were observed at Belo and Mahe, Seychelles for COI. suggesting a possible past population expansion (Table 6.12).

Clade 1 COI mismatch analysis was carried out excluding the three divergent Mozambican haplotypes. Mismatch analysis of all COI samples combined rejected the hypothesis of a population expansion (SSD = 0.17, P = 0.001) (Table 6.16). Examination of the mismatch histogram for COI revealed a multimodal mismatch distribution for clade 1 (Figure 6.11). By grouping of the COI samples according to their three respective haplogroups, the various modes of the mismatch distribution were separated and estimates for each population expansion were obtained (Figures 6.12, 6.13, 6.14, Table 6.17). The COI dataset indicated at least three population expansions; 84,000, 200,000 and 550,000 years ago.

Table 6.16. Mismatch analysis of *L. nebulosus* using COI mtDNA. T time since expansion in mutational units, θ_0 population size before expansion, θ_1 population size after expansion, SSD sum of squared deviations, TSE time since expansion in years. Statistically significant estimates (p<0.05) are highlighted in bold.

Location	T	$\theta 0$	$\theta 1$	SSD
ZT	0.4	2.04	99999.00	0.03
MA	4.7	0.00	99999.00	0.103
S	1.9	0.00	99999.00	0.016
AN	1.7	2.22	99999.00	0.014
B	2.8	0.52	99999.00	0.013

MM	0.8	0.00	99999.00	0.03
Overall	0.7	0.00	99999.00	**0.17**

Table 6.17. Mismatch analysis of *L. nebulosus* using COI mtDNA separated into their respective groups. T time since expansion in mutational units, θ_0 population size before expansion, θ_1 population size after expansion, SSD sum of squared deviations, TSE time since expansion in years.

Group	T	θ0	θ	SSD	TSE
1	0.5	0	18.39	0.0001	83,300
2	1.2	0	9999.99	0.0295	200,000
3	3.3	0	41.52	0.0619	550,000

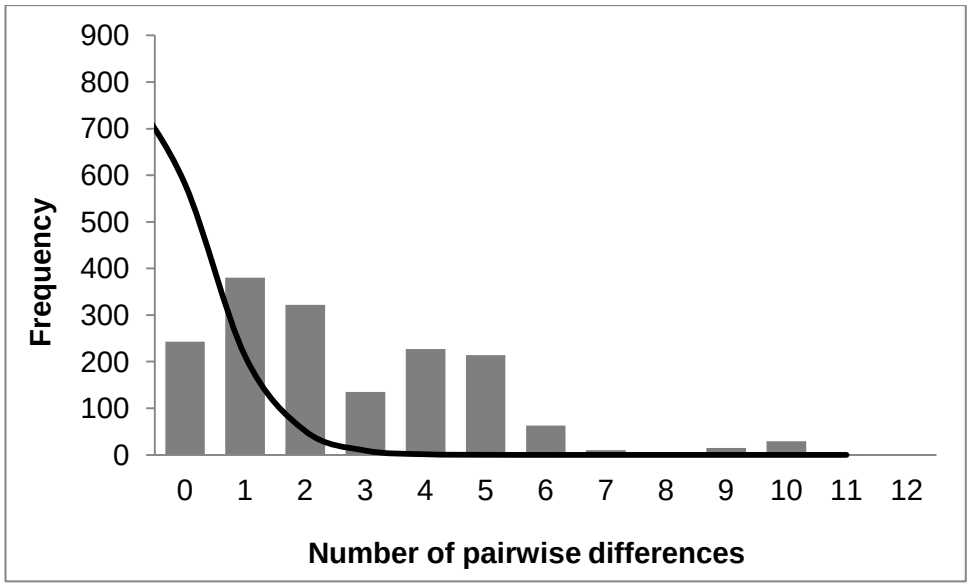

Figure 6.11. Mismatch distribution histogram for *L. nebulosus* clade 1 COI mtDNA across the SWIO. Filled bars indicate the observed frequency of pairwise distribution, black line indicates the expected distribution under a model of sudden demographic expansion.

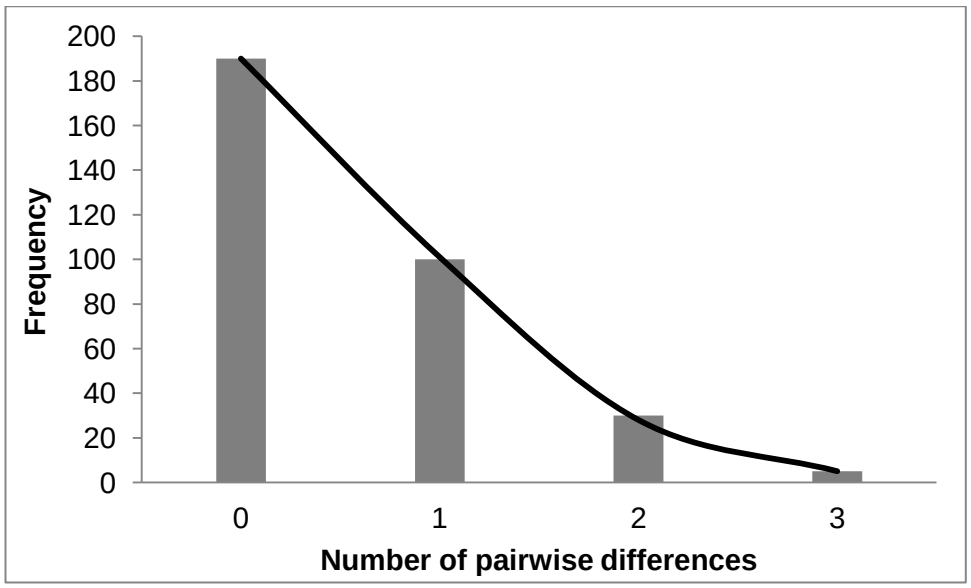

Figure 6.12. Mismatch distribution histogram for *L. nebulosus* group 1 COI mtDNA. Filled bars indicate the observed frequency of pairwise distribution, black line indicates the expected distribution under a model of sudden demographic expansion.

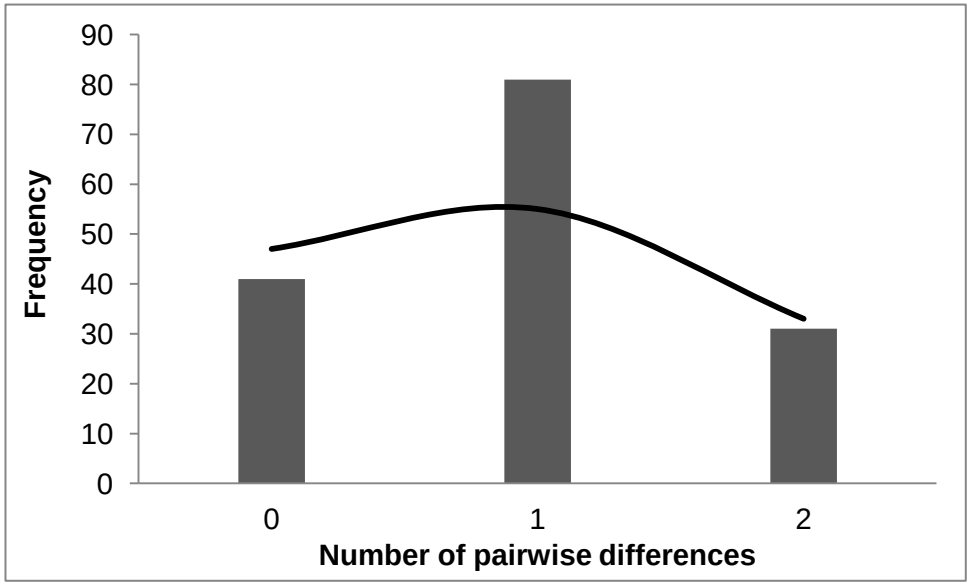

Figure 6.13. Mismatch distribution histogram for *L. nebulosus* group 2 COI mtDNA. Filled bars indicate the observed frequency of pairwise distribution, black line indicates the expected distribution under a model of sudden demographic expansion.

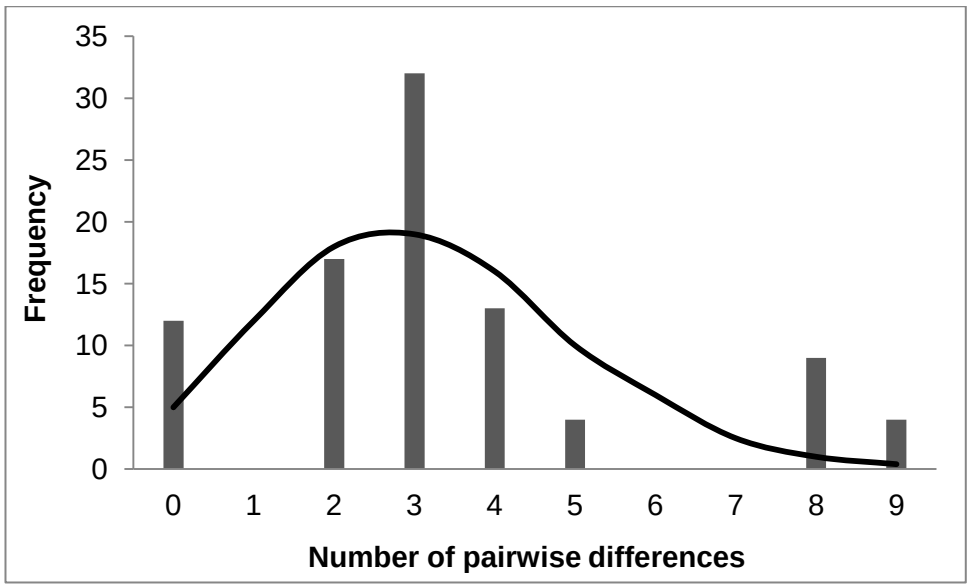

Figure 6.14. Mismatch distribution histogram for *L. nebulosus* group 3 COI mtDNA. Filled bars indicate the observed frequency of pairwise distribution, black line indicates the expected distribution under a model of sudden demographic expansion.

Chapter 7: Occurrence of a previously unrecognised species of *Lethrinus* across the Seychelles

7.1 Introduction

Molecular genetic data have found numerous applications in the field of population identification, fisheries management and elucidation of conservation units for aquatic species (Waples, et al. 2008). One of the most fundamental applications of molecular genetic data for fisheries management and conservation is accurate species identification. In fisheries, correct species identification underpins fisheries management as species catch data are used to estimate biomass sizes and to monitor quote uptake throughout the year. Incorrect species identification will therefore invalidate any fisheries monitoring data. Additionally, accurate species identification is necessary for identification of eggs and larvae for life history research and ecological monitoring (Waples, et al. 2008; Ward 2000).

The advent of inexpensive and rapid DNA sequencing has provided biologists with an additional tool with which to identify species. DNA sequencing has proved invaluable in distinguishing morphologically similar species and in identifying previously unknown cryptic species, where two or more distinct species have been classified as a single species which are morphologically very similar or in some cases identical (Bickford, et al. 2007). Cryptic species are fairly common in the marine environment, particularly in marine fish (Bickford, et al. 2007; Colborn, et al. 2001; Griffiths, et al. 2010; Kon, et al. 2007; Santos, et al. 2006; Tang, et al. 2010; Watanabe, et al. 2010).

Despite the value of genetic techniques in fisheries management, many fished species have yet to be investigated using genetic techniques, particularly in less studied areas of the ocean such as the Southwestern Indian Ocean (SWIO). Within the SWIO, the genus *Lethrinus* is an important fish in both commercial and subsistence fisheries. However, their meristic and morphometric characteristics are conservative making Lethrinidae one of the most difficult tropical marine coastal fishes to identify and classify (Carpenter and Allen 1989; Sato 1971; Smith 1959). In this study we initially aimed to investigate genetic variation in *Lethrinus mahsena*, a commercially harvested species currently managed as a single stock across the SWIO. However, initial results revealed that putative *L. mahsena* samples from the Seychelles comprised a highly divergent clade when compared to samples from elsewhere across the SWIO. The study therefore became an investigation of the degree of taxon diversification and within-species variation across the SWIO.

7.2 Methods

7.2.1 Sampling and DNA extraction

Sampling of putative *L. mahsena* was carried out at 6 locations across the SWIO (Figure 7.1, Table 7.1) from 2010 to 2012. Samples were obtained from local fish markets and subsistence fishermen (see Table 7.1). Samples from Madagascar were collected by Blue Ventures. Specimens were identified as *L. mahsena* by use of the description contained in the FAO species catalogue for *L. mahsena* (Carpenter and Allen 1989). Fish were identified by the following characteristics; head purplish grey, sometimes with a red blotch on the nape, body yellow to greenish blue with nine or ten dusky yellow-green or brown bars and a red bar at the base of the pectoral fin which sometimes extends above and below the pectoral fin base and red caudal fins (Carpenter and Allen 1989).

Fin samples were preserved in absolute ethanol and total genomic DNA was extracted as described in Chapter 2.2.1 using a phenol/chloroform extraction after Sambrook et al. (1989).

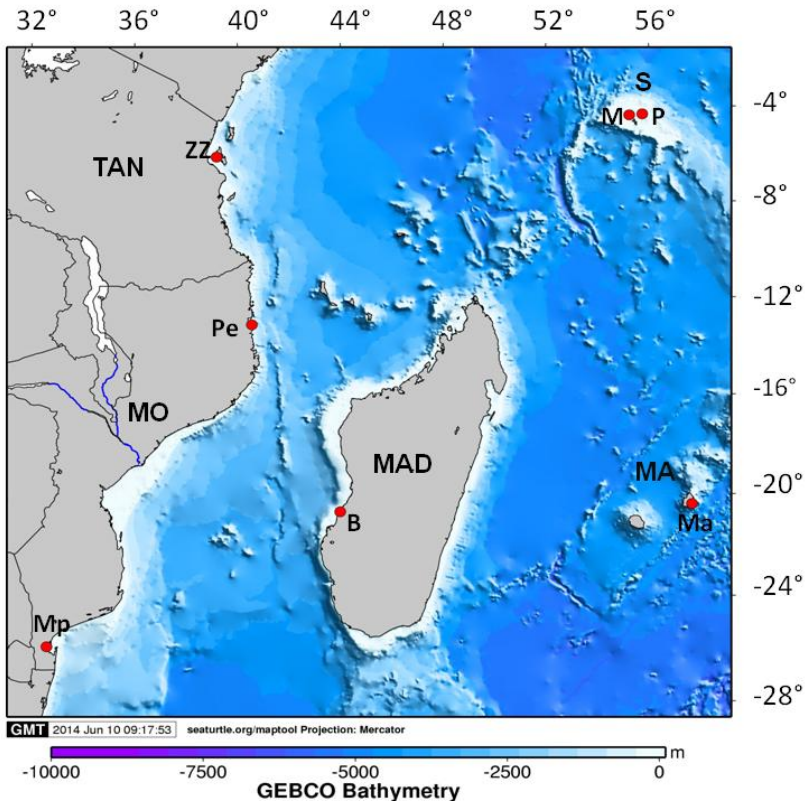

Figure 7.1. Map of sampling area of the SWIO. Sampling locations are indicated by red circles. Countries; TAN-Tanzania, MO-Mozambique, MAD-Madagascar, MA-Mauritius, S-Seychelles. Sampling locations; ZZ-Zanzibar, Mp-Maputo, Pe-Pemba, B-Belo sur Mer, Ma - Mahebourg, M – North Island, P - Praslin.

Table 7.1. Sampling records for putative *L. mahsena* across the SWIO, listing sampling location, country and sample size.

Country	Location Name	Code	Sample size
Mozambique	Maputo fish market	Mp	8
Mozambique	Pemba fish market	Pe	3
Madagascar	Belo sur Mer	B	5
Zanzibar, Tanzania	Stone town fish market	ZZ	19
Mauritius	Mahebourg fish market	Ma	18
Seychelles	Ile du Nord	M	13
Seychelles	Baie St Anne, Praslin	P	9

7.2.2 **Genetic screening**

Species-specific primers Lh COIF and LM COIR were used to amplify putative samples of *L. mahsena*. Samples were sequenced for a 489bp fragment of the mitochondrial DNA COI gene as described in Chapter 2.2.2).

7.2.3 **Mitochondrial DNA analyses**

Phylogenetic reconstruction

Haplotype networks were constructed using maximum likelihood methods in PHYLIP v.3.6 DNAml (Felsenstein 1989) and subsequently constructed in Hapview (Ewing 2010). Sequence divergence between clades was calculated using the minimum p-distance in MEGA6 (Tamura, et al. 2013).

Reconstruction of phylogenetic relationships was carried out using Bayesian and maximum likelihood methods as described in Chapter 2.3.1. The most suitable model of nucleotide substitution to account for the observed sequence polymorphism in samples was calculated using Bayesian Information Criterion (BIC) in jModelTest version 2.1.3 (Darriba, et al. 2012; Guindon and Gascuel 2003).

Bayesian Markov Chain Monte Carlo (MCMC) analyses were carried out using the appropriate model and default settings in MrBayes v 3.2 (Ronquist, et al. 2012) as described in Chapter 2.3.1.

Maximum likelihood analyses were conducted using a neighbour-joining starting tree and implemented using the online version of PhyML 3.0 (Guindon, et al. 2010) as described in Chapter 2.3.1.

Trees were constructed in Figtree v.1.4 (Rambaut 2012).

Population diversity

Arlequin 3.5.1.2 (Excoffier and Lischer 2010) was used to the calculate number of haplotypes (H), haplotype diversity (*h*) and nucleotide diversity (π) as described in Chapter 2.3.1. Haplotype networks were constructed for each mitochondrial region using maximum likelihood methods in Hapview (Ewing 2010). Fu's Fs (Fu 1997), Tajima's D (Tajima 1989) and Ewens-Watterson's F (Slatkin 1994a) were calculated to test for deviations from neutrality using Arlequin 3.5.1.2 (Excoffier and Lischer 2010) as described in Chapter 2.3.1.

Demographic history

Mismatch distribution was calculated Arlequin 3.5.1.2 (Excoffier and Lischer 2010) as described in Chapter 2.3.1.

To estimate time since expansion, substitution rates were calculated from divergence times. Substitution rates of 6×10^{-3} per million years (MYR) were used for COI mtDNA calculated from divergence rates of 1.2% per MYR estimated from the closure of the Isthmus of Panama (Bermingham, et al. 1997). Calculations of time since expansion using mismatch analyses were checked using the online spreadsheet http://www.uni-graz.at/zoowww/mismatchcalc/index.php (Schenekar and Weiss 2011).

Time since divergence

Time since divergence was estimated by calculating the mean percentage divergence between clades or species and then applying a divergence rate of 1.2% per MYR for COI mtDNA, calculated from the closure of the Isthmus of Panama (Donaldson and Wilson 1999).

7.3 Results

Phylogenetic reconstruction

Seventy five individuals of putative *L. mahsena* were sequenced for 490bp of COI mtDNA. Haplotype network and phylogenetics reconstruction revealed the existence of two highly divergent clades (Figure 7.2, 7.3), separated by 21 mutational steps compared to a maximum of 5 mutational steps within clade 1 and 3 mutational steps within clade 2. The minimum p-distance between the two clades was 0.043, considerably higher than the minimum p-distance of 0.000 within *L. mahsena* clade 1 and clade 2.

Clade 1 contained 56 individuals of *L. mahsena* and representatives of this clade were present at all sampling locations except Praslin (Figure 7.4). Clade 1 was predominant at all sites except those in the Seychelles where clade 2 was dominant (Figure 7.4).

Clade 2 contained 19 individuals and distribution of clade 2 was restricted to sampling locations in the Seychelles (Figure 7.4).

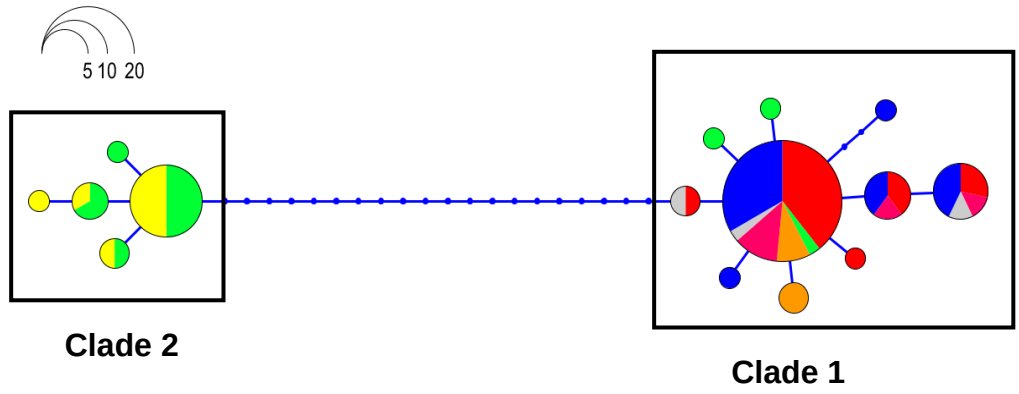

Figure 7.2. Haplotype network of *L. mahsena* based on 480bp of COI mtDNA. Branch lengths are proportional to the number of differences. The node size is proportional to the haplotype frequency. Each small blue circle represents a single mutation; Zanzibar, Mauritius, Praslin, Belo sur Mer, Madagascar, Ile du Nord, Maputo, Pemba.

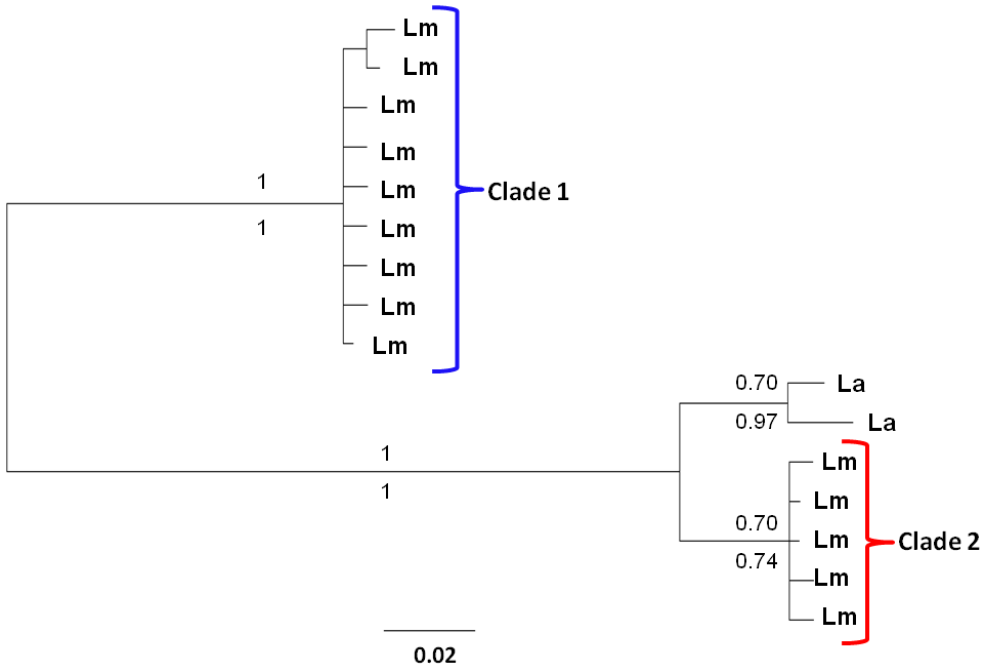

Figure 7.3. Reconstruction of phylogenetic relationships within *Lethrinus* sp. using 490bp COI mtDNA. Statistical support for nodes is given for both Bayesian (posterior probabilities) above branches) and ML analyses (aLRT values) below branches. Lm – *L. mahsena*, La – *L. atkinsoni*.

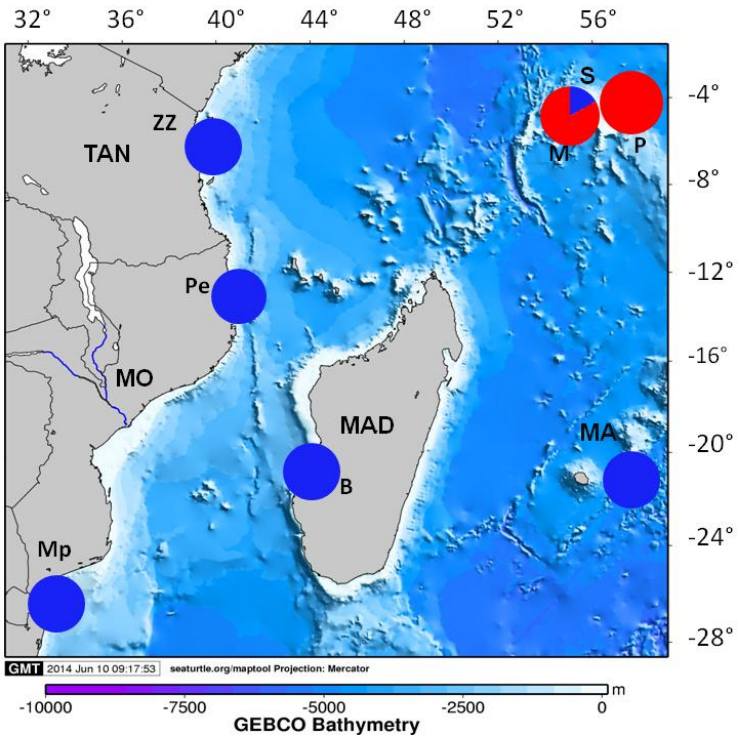

Figure 7.4. Map of the SWIO showing the relative frequency of the *L. mahsena* haplogroups. Blue represents group 1, red represents group 2. MAD-Madagascar, MO-Mozambique, MA-Mauritius, TAN-Tanzania, K-Kenya, S-Seychelles, Mp-Maputo, Pe-Pemba, ZZ-Zanzibar, M-Ile du Nord, P-Baie St Anne, B-Belo sur Mer.

To further investigate the two highly divergent clades, an interspecific phylogeny of putative *L. mahsena* was constructed using a 489bp fragment of COI mtDNA together with *L. atkinsoni*, a closely related sister species of *L. mahsena*, using *L. atkinsoni* sequences from GenBank (KF009603.1, EF609384.1, KC970391.1), (Figure 7.4).

The phylogeny shows strong support for the divergence of the two putative *L. mahsena* clades, with clade 2 appearing to be more closely related to *L. akinsoni* (Figure 7.3) than to clade 1. To further evaluate the taxonomic position of these two divergent clades, a phylogenetic tree using a number of related *Lethrinus* species was constructed (Figure 7.5). Once again clade 2 clusters more closely to *L. atkinsoni* than to any other *Lethrinus* species, including clade 1 of *L. mahsena*, with high branch support.

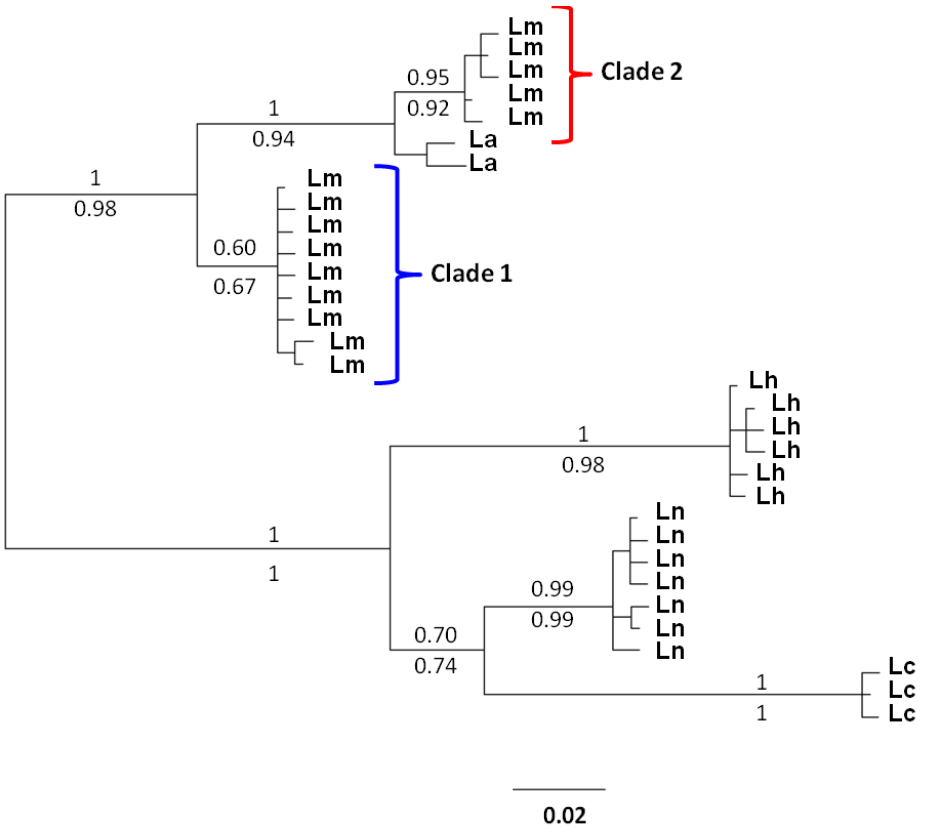

Figure 7.5. Reconstruction of phylogenetic relationships within *Lethrinus* sp. using 490bp COI mtDNA. Statistical support for nodes is given for both Bayesian analyses (posterior probabilities) above branches and ML analyses (aLRT values) below branches. Lm – *L. mahsena*, La – *L. atkinsoni*, Ln – *L. nebulosus*, Lh – *L. harak*, Lc – *L. crocineus*.

Searches using the Basic Local Alignment Search Tool (BLAST) on clade 1 individuals revealed a 99% sequence similarity to 3 individuals of *L. mahsena* on GenBank (JF493750.1, JF493751.1, JF493752.1). Three other entries on GenBank which were labelled as *L. mahsena* displayed 96% similarity (EF609387.1) and 92% similarity (JQ350088.1, JQ350089.1) to clade 1.

BLAST searches using individuals from clade 2 indicated that they were most closely identified (99% sequence similarity) to one sample of *L. mahsena* (EF609387.1) collected from India followed by three samples of *L. atkinsoni* at 98% similarity. Three other *L. mahsena* sequences (JF493751.1, JF493750.1 JF493752.1) displayed 96% sequence similarity, whilst another two *L. mahsena* samples showed 91% similarity (JQ350088.1, JQ350089.1) on GenBank.

It is likely that a number of samples which are labelled *L. mahsena* in GenBank are incorrectly identified. *L. mahsena* GenBank samples JQ350088.1, JQ350089.1 both have 100% sequence similarity to *L. lentjan* samples on GenBank and it is therefore likely that these samples are not *L. mahsena* but *L. lentjan*, which has a similar appearance to *L. mahsena* in that it often has red markings around the pectoral fins (Carpenter and Allen 1989). GenBank sample EF609387.1 displayed 98% sequence

similarity to *L. atkinsoni* samples on GenBank. It is therefore likely that these samples have been incorrectly identified as *L. mahsena* and should be excluded from the BLAST analysis. Interestingly, all of these three GenBank entries are from the same research study.

Clade 1 therefore showed 99% similarity to true *L. mahsena* samples on GenBank (JF493751.1, JF493750.1 JF493752.1), which corresponded to the majority of the samples taken during this study, whilst clade 2 displayed 98% sequence similarity to individuals of *L. atkinsoni*, a species commonly found in the West Pacific, and only 96% sequence similarity to *L. mahsena* samples on GenBank (JF493751.1, JF493750.1 JF493752.1).

It is apparent from the combination of the phylogenetic reconstruction and GenBank BLAST results that clade 2 represents a separate taxon from *L. mahsena* and so was analysed separately. It is likely that this taxon is part of or closely related to *L. atkinsoni* (see Discussion) and will be referred to as *Lethrinus* sp. A from this point forward.

The K80 + I nucleotide substitution model was found to be the most suitable model of sequence evolution.

Population Diversity

The two samples from Pemba and Maputo for *L. mahsena* were combined to make one Mozambique sample due to their small sample size. Diversity indices were calculated for both *L. mahsena* and *Lethrinus* sp. A. Haplotype and nucleotide diversities were similar for the two species. Haplotype diversity for *L. mahsena* ranged from 0.53 - 0.69 and haplotype diversity of *Lethrinus* sp. A ranged from 0.58 - 0.64 (Table 7.2, 7.3). Nucleotide diversity ranged from 0.0012 - 0.0023 and 0.0016 - 0.0017 for *L. mahsena* and *Lethrinus* sp. A respectively (Table 7.2, 7.3). *L. mahsena* displayed highest haplotype and nucleotide diversity at Mozambique, perhaps indicating slight structuring in this sample of *L. mahsena*. This is perhaps to be expected as this is a combination of samples from North and South Mozambique with considerable distance between the two sampling locations.

Departure from neutrality was observed in the mtDNA dataset for Ewens-Watterson test (F) for both samples of *L. atkinsoni* sp (Table 7.3). Only one departure from neutrality of Ewens-Watterson test was observed in *L. mahsena* (Tables 7.2).

Table 7.2. Mitochondrial genetic diversity levels and neutrality tests for *L. mahsena* COI; Sample size (n), number of haplotypes (H), number of private haplotypes (pHap), haplotype diversity (h), nucleotide diversity (π), Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs).

Country	**Location Name**	**n**	**H**	**pHap**	**h**	**π**	**F**	**D**	**Fs**
Zanzibar, Tanzania	Stone town fish market	19	5	1	0.53	0.0016	0.50	-0.99	-1.89
Mauritius	Mahebourg fish market	18	4	1	0.54	0.0017	0.49	-0.13	-0.54
Seychelles	Ile du Nord	3	3	2	-	-	-	-	-
Madagascar	Belo sur Mer	5	2	1	0.60	0.0012	0.52	1.22	0.62
Mozambique	Maputo & Pemba markets	9	4	0	0.69	0.0023	0.38	0.03	-0.82
	Overall	54	9	7	-	-	0.44	-0.04	-0.35

Table 7.3. Mitochondrial genetic diversity levels and neutrality tests for *Lethrinus* sp. A COI; Sample size (n), number of haplotypes (H), number of private haplotypes (pHap), haplotype diversity (h), nucleotide diversity (π), Ewens-Watterson test (F), Tajima's D test (D) and Fu's Fs test (Fs).

Location Name	**n**	**H**	**pHap**	**h**	**π**	**F**	**D**	**Fs**
North Island	10	4	1	0.64	0.0016	0.42	-1.03	-1.46
Praslin	9	4	1	0.58	0.0017	0.48	-0.94	-1.42
Overall	19	5	2	0.61	0.0017	0.45	0.98	-1.44

Demographic history

Fu's Fs test displayed significant departures from neutrality for *Lethrinus* sp. A COI mtDNA (Table 7.3) suggesting a past population expansion. Both *L. mahsena* and *Lethrinus* sp. A appear to have undergone a relatively recent population expansion, 190,000 years ago for *L. mahsena* and 150,000 years ago for *L. atkinsoni* sp (Table 7.4, Figures 7.6, 7.7).

Table 7.4. Mismatch analysis for *L. mahsena* and *Lethrinus* sp. A COI mtDNA. T time since expansion in mutational units with 95% confidence intervals in brackets, θ_0 population size before expansion, θ_1 population size after expansion, SSD sum of squared deviations, TSE time since expansion in years with 95% confidence intervals in brackets.

Species	**T**	**θ_0**	**θ_1**	**SSD**	**TSE**
L. mahsena	1.1 (0.1-2.6)	0.0439	99999.00	0.002	190,000 (17,000-442,000)
Lethrinus sp. A .	0.9 (0.2-2.1)	0.000	99999.00	0.007	150,000 (34,000-357,000)

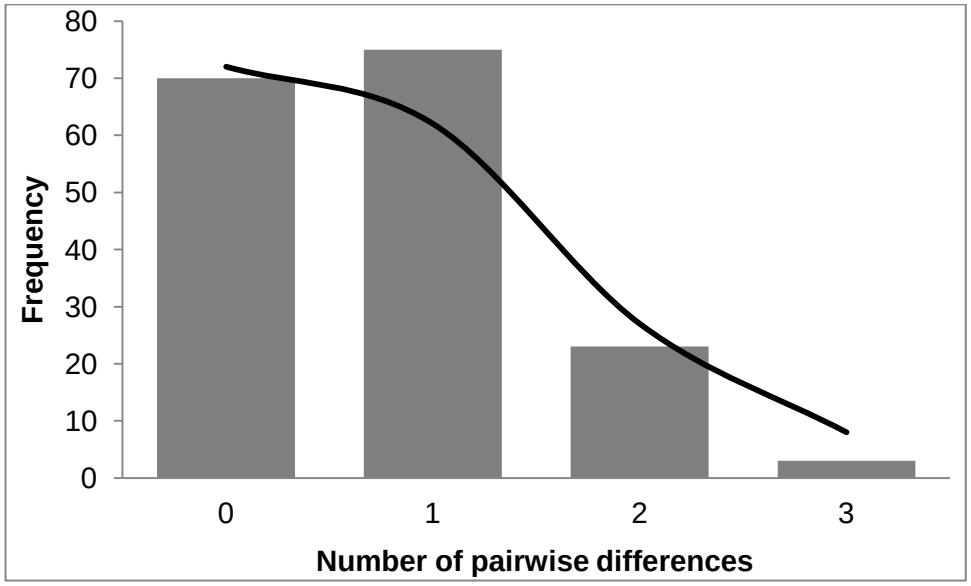

Figure 7.6. Mismatch distribution histogram for *L.aff. atkinsoni* in the Seychelles based on COI mtDNA sequences. Filled bars indicate the observed frequency of pairwise distribution, the black line indicates the expected distribution under a model of sudden demographic expansion.

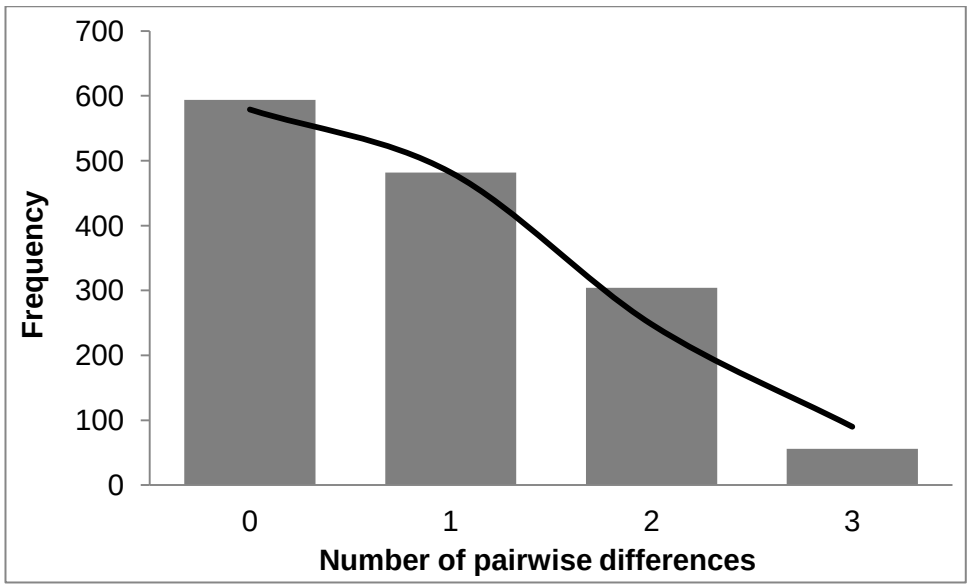

Figure 7.7. Mismatch distribution histogram for *L. mahsena* across the SWIO based on COI mtDNA sequences. Filled bars indicate the observed frequency of pairwise distribution, the black line indicates the expected distribution under a model of sudden demographic expansion.

Time since Divergence

The average sequence divergence between *L. atkinsoni* and *Lethrinus* sp. A is 2.2% and minimum sequence divergence is 1.9%. Using these sequence divergence estimates, time since divergence was calculated using a 1.2% sequence divergence per MYR resulting in an approximate divergence time of 1.60 - 1.80 MYR ago.

The average sequence divergence between *Lethrinus* sp. A and *L. mahsena* is 4.6% and minimum sequence divergence is 4.3%. Using these sequence divergence estimates, time since divergence was calculated using a 1.2% sequence divergence per MYR resulting in an approximate divergence time of 3.60 - 3.80 MYR ago.

7.4 **Discussion**

Analysis of putative *L. mahsena* samples across the SWIO revealed the presence of two highly divergent clades, one of which, clade 2, was found solely in the Seychelles. Haplotype networks, phylogenetic reconstruction and GenBank BLAST searches revealed that this Seychelles clade clustered more closely with GenBank sequences for *L. atkinsoni* than *L. mahsena*, revealing a possible cryptic admixture of species within what is currently regarded as a single *L. mahsena* species unit across the SWIO.

Clade 2 can be conclusively identified as a separate species from *L. mahsena*. High COI sequence divergence was observed between the two clades, with a minimum p-distance of 0.043, equivalent to a minimum 4.3% sequence divergence, compared to minimum p-distances of 0.000 within the two clades. Phylogenetic reconstruction

further confirmed the divergence between the two clades, with clade 2 being more closely related to *L. atkinsoni* than to *L. mahsena*. Branch support for this divergence was high using both Bayesian probability and maximum likelihood methods. Additionally GenBank BLAST searches further supported the hypothesis that clade 2 was a separate species from *L. mahsena*, indicating only a 96% similarity of clade 2 to *L. mahsena* based on COI mtDNA sequences. Clade 2 was therefore determined to be distinct from clade 1.

It may be therefore that *Lethrinus* sp. A is a cryptic species in the SWIO. Cryptic species are defined as two or more distinct species that were classified as a single species due to their morphological similarity (Bickford, et al. 2007). The genetic data collected provide evidence for the differentiation of *Lethrinus* sp. A from *L. mahsena*. This differentiation has not previously been identified by any previous research, or by local fishermen, who use the same name for both species, suggesting that the two species are similar in appearance. This suggests that *Lethrinus* sp. A and *L. mahsena* may be a cryptic species complex in the SWIO. However, due to the lack of morphological data in this study, it is also possible that *Lethrinus* sp. A may be a recognised species of *Lethrinus* which has not previously been sequenced for COI mtDNA. This possibility seems unlikely however, considering that the FAO species catalogue for Lethrinidae of the world (Carpenter and Allen 1989) does not appear to contain any recognised *Lethrinus* species of similar appearance to *L. mahsena* in the SWIO. Additionally, the meristic and morphometric characteristics of Lethrinidae are conservative, making them one of the most difficult tropical marine coastal fishes to identify and classify (Carpenter and Allen 1989; Sato 1971; Smith 1959) and perhaps increasing the likelihood of cryptic species complexes and taxonomic misidentification. Recent studies have shown the existence of a cryptic species of *Lethrinus* using morphometric and genetic techniques (Borsa, et al. 2013), *L. nebulosus* samples from South Africa and Australia have displayed 5.68% COI sequence divergence, suggesting underestimated species diversity in *L. nebulosus* (Zemlak, et al. 2009). Additinally a new species of *Lethrinus* has also been recorded in the literature (Carpenter and J.E. 2003).

Lethrinus sp. A may be a separate species from *L. atkinsoni*. Phylogenetic reconstruction displayed divergence between *Lethrinus* sp. A and *L. atkinsoni* with high branch support. The minimum p-distance between *Lethrinus* sp. A and *L. atkinsoni* was 0.019, equivalent to a minimum 1.9% sequence divergence. whilst minimum p-distance within and *L. atkinsoni* was 0.000. It is therefore possible that *Lethrinus* sp. A is a separate species to *L. atkinsoni. T*his cannot be confirmed however without further research, due to the limited data that this study collected.

In this study we have not used a specific cut-off threshold to define marine fish species although this is a common technique in previous studies. For example, Zemlak, et al. (2009) use a 3.5% mean sequence divergence using the mutational model K2P as a screening threshold for marine fish species. This is based on the principle that an appropriate cut-off threshold for provisional species is 10x that of the mean intra-species variation for a particular group (Hebert, et al. 2004). In marine fish and *Lethrinus* specifically, 3.5% is approximately 10x that of average intra-specific sequence variation (Hebert, et al. 2004; Ward, et al. 2005; Zemlak, et al. 2009; Zhang and Hanner 2011). However, the method of using K2P sequence divergence has been criticised as K2P is often not the most appropriate model of substitution, as is the case here, does not perform better than uncorrected distances and this method is not supported by any empirical evidence (Srivathsan and Meier 2012). Additionally, the use of mean sequence divergence instead of the smallest interspecific distance has been criticised as exaggerating the "barcode gap" and leading to species misidentification (Meier, et al. 2008). We have therefore reported both the minimum p-distance and average sequence divergence and have not employed a threshold cut-off value in this study.

Phylogenetic reconstruction and divergence estimates indicate that *Lethrinus* sp. A is more closely related to *L. atkinsoni* than *L. mahsena*. This is perhaps unexpected as the present recognised distribution of *L. atkinsoni* is widespread in the central and West Pacific including Indonesia, northern Australia, the Philippines to southwest Japan and throughout the West Pacific to the Tuamotus, but has no recorded distribution in the Indian Ocean. However, examination of the divergence date of *Lethrinus* sp. A from *L. atkinsoni* may provide some explanation for this.

In the present study the divergence of *Lethrinus* sp. A from *L. atkinsoni* has been estimated to have occurred 1.6 – 1.8 MYR ago. This is obviously an approximate date, as it does not take into account either the most suitable model of mutation or the variation of substitution rates between lineages and so should be interpreted with caution. However, the estimate of this divergence coincides with the Pleistocene epoch, (2.58 MYR – 11,700 years ago), a period of considerable climate variation and oscillation. Although details of the climate of the Indian Ocean are not available for this period of the Pleistocene, it is likely that sea level and sea surface temperature in the SWIO varied considerably. It may therefore be that at some point Pleistocene climate conditions, such as raised sea levels during interglacial periods, enabled the colonisation of part of the Indian Ocean, including the Seychelles, by *L. atkinsoni* from the western Pacific. Subsequent variation in Pleistocene climate isolated the Seychelles population, perhaps through the mechanism of the Indo-Pacific Barrier (IPB), a land bridge between Asia and Australia (Voris 2000) during periods of lowered

sea level which severely restricted dispersal between the Indian and Pacific oceans. Isolation of conspecific populations across the IPB may have led to allopatric speciation of *Lethrinus* sp. A from *L atkinsoni*. It therefore seems that the most likely explanation for this speciation is Pleistocene driven vicariance. This has been observed in a number of fish species including flathead mullet and eel gobies (Tang, et al. 2010) in the NW Pacific (Shen, et al. 2011) and sleek unicornfish in the Indo-Pacific (Horne and van Herwerden 2013)

The current study only found samples of *Lethrinus* sp. A in the Seychelles. A number of *Lethrinidae* are present throughout the Indian and Pacific Oceans, including *L. harak*, *L. nebulosus*, *L. mahsena* and *L. rubriopercolatus* (Carpenter and Allen 1989; Ebisawa and Ozawa 2009; Nanami and Yamada 2009), although an endemic species of *Lethrinus*, *L. enigmaticus*, is described from the Seychelles. The range of *Lethrinus* sp. A, however, is probably not restricted solely to the Seychelles. During this study, GenBank sequence EF609387.1, which was sampled off the coast of India and identified as *L. mahsena*, exhibited 100% similarity to one of the *L. atkinsoni* sp sequences collected from Praslin during this study. It is therefore very likely that *Lethrinus atkinsoni* sp. is present in other locations across the Indian Ocean, including off the coast of India but to date has been misidentified as *L. mahsena*.

At present, two varying colour forms of *L. mahsena* are recognised within the species description (Carpenter and Allen 1989). The name *L. sanguineus* is sometimes used to identify the form of this species with a bright red oblique streak from above to below the pectoral fin base (Carpenter and Allen 1989). Carpenter and Allen (1989) examined both *L. mahsena* and *L. sanguineus* colour forms and found no morphological differences between the two forms other than the colour variation, thus classifying them both as *L. mahsena*. They suggest no definite hypothesis for the cause of variation in the red colour forms but suggest it may be due to population, environment or reproductive mechanisms (Carpenter and Allen 1989). At the time of this classification, no DNA sequences were available to compare the two colour forms. It is therefore possible that these two colour forms correspond to the *L. mahsena* and *Lethrinus* sp. A found in this study. However, as morphological data were not recorded in the present study, further work would be required to investigate this hypothesis.

Accurate identification of fished species and of this species in particular, is a prerequisite for stock management and conservation (Waples, et al. 2008). *Lethrinus* sp. A and *L. mahsena* are popular fisheries catches across the SWIO including the Seychelles and are currently sold commercially and in fish markets as *L. mahsena*, or Dame Biri in the local language. Further morphological and genetic work is therefore required to formally classify the taxonomic position of *Lethrinus* sp. A and identify

populations of this species across the Indian Ocean. This information may then be used to effect the sustainable fisheries management of these two species.

Chapter 8: Conclusion

8.1 Introduction

Despite the SWIO's recognised importance as a biodiversity hotspot there are limited data on genetic diversity (intra- and inter-specific) within the region. In recent years, environmental degradation and declines in natural resources and biodiversity have been observed in the SWIO (Berg, et al. 2002; Ridgway and Sampayo 2005). Further research in the form of conservation genetic studies are therefore urgently required to investigate population structure, identify unrecognised genetic diversity (such as cryptic species) and develop species identification techniques. This research would greatly advance current scientific knowledge of biodiversity in the SWIO and would provide valuable information which could then be utilised for the effective management and conservation of biodiversity within the SWIO. To this end, Ridgway and Sampayo (2005) outlined three research directions for the SWIO which this study set out to address:

1. To describe spatial patterns of genetic diversity for a number of taxa;
2. To study the influence of historical biogeographical events on present day populations; and
3. To use more realistic models to understand the genetics of natural populations.

This thesis has investigated intra-specific genetic diversity across the SWIO in a number of species and examined the underlying factors, including historical and contemporary drivers associated with the partitioning of this genetic diversity.

8.2 Key findings

8.2.1 Identification of cryptic species

The discovery of two possible cryptic species of *Lethrinus* in the SWIO was unexpected but is perhaps not surprising given the limited research on SWIO taxa to date (Ridgway and Sampayo 2005). *Lethrinidae* are one of the hardest tropical fish genus to identify taxonomically (Carpenter and Allen 1989) and cryptic species of Lethrinus have been recorded in the Indo-Pacific (Borsa, et al. 2013). The advent of DNA sequencing technology has led to the discovery of many previously unrecognised cryptic species in marine fish (Bickford, et al. 2007; Colborn, et al. 2001; Griffiths, et al. 2010; Kon, et al. 2007; Santos, et al. 2006; Watanabe, et al. 2010). The Western Indian Ocean has been identified as a biodiversity hotspot (Roberts, et al. 2002) but less than half the species present in East Africa have been described (Griffiths 2005) and knowledge of the fauna of the Western Indian Ocean remains limited in comparison to other well-studied regions such as the Indo-Pacific (Ridgway and Sampayo 2005).

Lethrinus nebulosus displayed two highly divergent lineages, in keeping with two separate species, which are estimated to have diverged 4.5 - 10 million years ago (MYA). These two species have distinct and slightly overlapping ranges. The recognised *L. nebulosus* was widespread across the SWIO including Seychelles, Mauritius, Kenya and Tanzania. The cryptic species was more geographically limited within the studied range, occurring along the coast of southern Mozambique and South Africa.

Putative samples of *L. mahsena* collected during this research revealed the presence of cryptic species in the Seychelles, classified as *Lethrinus* sp. A Phylogenetic reconstruction suggested it to be to be more closely related to *L. atkinsoni*, a species found in the West Pacific, than *L. mahsena*. The divergence of *Lethrinus* sp. A from *L. atkinsoni* is estimated to have occurred approximately 1.6 - 1.8 MYA, during the Pleistocene epoch.

The discovery of two cryptic *Lethrinus* species has a number of implications with regard to biodiversity and fisheries management which are discussed in further detail in the next section.

8.2.2 **Population structuring in the SWIO**

At the outset of this research it was hypothesised that the physical oceanography of the SWIO may present a barrier to intra-specific gene flow, specifically it was predicted that the circulation in the Mozambique Channel may restrict gene flow between populations north and south of the Channel. The Mozambique Channel circulation consists of southward moving anticyclonic eddies on the western side of the channel and poorly defined weak currents on the eastern side of the channel (Quartly and Srokosz 2004; Schouten, et al. 2003). Circulation in the Mozambique Channel has been relatively constant for the past 30,000 years (Obura 2012) and it was therefore predicted that it may have acted as a long term barrier to gene flow, ultimately generating separate phylogeographic lineages north and south of the channel. However, in species which had sufficient sampling locations to test this hypothesis, *L. nebulosus* and *O. cyanea*, no structuring was observed due to the physical oceanography of the Mozambique Channel.

The results revealed high population connectivity across the SWIO in *L. harak*, *L. nebulosus* and *O. cyanea*, with some structuring observed between Mauritius and other sampling locations in the SWIO. The *O. cyanea* microsatellite dataset (excluding ROC6) was characterized by very low and non-significant pairwise neutral F_{ST}, with no relationship existing between geographic and genetic distances. Neither was any clear genetic structure recognized by the method implemented in STRUCTURE. These

results are in keeping with the few studies previously conducted in the SWIO which reported high levels of connectivity between localities (Dorenbosch, et al. 2006; Muths, et al. 2009; Muths, et al. 2013; Silva, et al. 2010b; Visram, et al. 2010).

It is likely that both life history characteristics of these species and physical oceanographic properties of the SWIO contribute to the genetic homogeneity observed across most of the SWIO. The pelagic larval phase is the agent of connectivity between populations for most reef fishes and *O. cyanea* and therefore is the mechanism by which genetic homogeneity may be maintained within these species. The pelagic larval duration (PLD) of these species range from 28 days for *Lethrinus* sp to ~50 days for *O. cyanea* (Ebisawa and Ozawa 2009; van Heukelem 1973). Pelagic larval durations of this length may allow larvae to passively cover distances of several hundred kilometres in an area where marine currents are of the order of 20-30 Sverdrup (1 Sverdrup = 10^6 m^3 s^{-1}) (Schott, et al. 2009). Additionally, *O. cyanea* produce pelagic paralarvae which are specially adapted to pelagic life (Villanueva and Norman 2008) and may possess considerable swimming abilities. Swimming ability in *Lethrinu*s sp. is also considerable, with late stage swimming speeds of 38±3 cm s^{-1} (Fisher, et al. 2005) which may well enhance dispersal ability. Therefore both life history characteristics and physical oceanography of the SWIO may explain why genetic connectivity was maintained across distances of over 3000 km.

However, despite the findings of high genetic connectivity across the SWIO, the results indicate that populations of *O. cyanea*, *L. harak* and *L. nebulosus* are not at migration-drift equilibrium. Molecular markers examine genetic differentiation over evolutionary timescales and may therefore be affected by both past as well as contemporary patterns of dispersal (Hauser and Carvalho 2008). Migration-drift disequilibrium may mean that current population structuring is not detected at neutral loci such as mtDNA and microsatellites due to the retainment of ancestral patterns of connectivity at these loci. Recently separated large population may therefore appear genetically homogenous even in the absence of contemporary gene flow (Hauser and Carvalho 2008). Retained historical patterns of connectivity in the study species may therefore be overestimating the current connectivity of populations in the SWIO.

Significant structuring was observed between Mauritius and other locations in the SWIO in *L. harak*, *L. nebulosus* and *O. cyanea* (Table 8.1). Population structure of *L. harak* indicates that the Mauritian population is differentiated from all other populations across the SWIO and *L. nebulosus* displays some population structuring between Mauritius and the Northern Mozambique Channel. The positive selection of microsatellite locus ROC6 in *O. cyanea* is also interesting. The combination of some limited structuring in the *O. cyanea* mtDNA dataset and the high statistical power of the

microsatellite dataset suggest that the selection acting upon ROC6 represents some isolation between Mauritius and other locations in the SWIO that is not currently detected in neutral loci, perhaps due to migration-drift disequilibrium. Recent studies have shown that organisms can undergo adaptive phenotypic evolution over a few generations (Carroll et al. 2007). Loci under selection may be valuable as population markers on these ecological timescales where isolated populations have not yet diverged at neutral loci (André, et al. 2011; Nielsen, et al. 2006; Nielsen, et al. 2009; O'Malley, et al. 2007). That the locus ROC6 is under selection only in Mauritius perhaps suggests the beginnings of population divergence between Mauritius and other sites across the SWIO.

A review of all the literature on population structuring of marine species in the SWIO revealed that many of the studies demonstrate panmictic populations with high levels of marine connectivity between localities; such as mud crab *Scylla serrata* (Fratini, et al. 2010), fiddler crab *Uca annulipes* (Silva, et al. 2010b), swordfish *Xiphias gladius* (Muths, et al. 2009; Muths, et al. 2013), *Lutjanus fulviflamma* (Dorenbosch, et al. 2006). However a couple of studies display small genetic differences that indicate some isolation of peripheral areas. In both *Scarus ghobban* (Visram, et al. 2010) and the mangrove crab *Neosarmatium meinerti* (Ragionieri, et al. 2010) the Seychelles populations were isolated from East African localities.

Only one other previous study has found restricted differentiation between Mauritius and other locations in the SWIO (Muths, et al. 2012). mtDNA sequencing of *L. kasmira* found significant differentiation between Moroni, Comoros Island, in the Northern Mozambique Channel and Mauritius. This differentiation was greater than the differentiation between Mauritius and the most distant localities in the data set, ruling out the likelihood of peripheral isolation. This also appears to be the case for our data. The level of differentiation between Mauritius and Northern Mozambique Channel (NMC) in *L. nebulosus* is greater than the differentiation between NMC and the Seychelles, the most distant location sampled. We therefore conclude, in agreement with Muths, et al. (2012) that there is some local process occurring at Mauritius which is responsible for the population structuring observed.

Table 8.1. Summary table displaying population structuring of study species across the SWIO.

Species	Population structure in SWIO
L. nebulosus	Φ_{ST} based on CR mtDNA indicated population structuring in clade 1 between Mauritius and northern Mozambique Channel after Bonferroni correction. F_{ST} based on COI mtDNA indicated population structuring between Mauritius and Maputo, Mozambique and the Seychelles after Bonferroni correction.
L. harak	Concatenated mtDNA dataset indicated Mauritius populations differentiated from all other sampling locations in SWIO using F_{ST}, Φ_{ST} and exact test of population differentiation after Bonferroni correction.
O. cyanea	One positive outlier microsatellite locus ROC6 reporting differentiation of Mauritius from all other samples. F_{ST} and exact tests on NCR mtDNA indicate some structuring between Mauritius and Mozambique, Madagascar, Seychelles and Rodrigues which is non-significant after Bonferroni correction.

8.2.3 **The influence of Pleistocene climate fluctuations on the demographic history of species in the SWIO**

The Pleistocene climate fluctuations of the last two million years dramatically affected distribution and diversity of terrestrial and aquatic organisms through habitat alteration (Avise and Walker 1998; Carstens and Knowles 2007; Hewitt 1999; Hoareau, et al. 2012). Pleistocene climate fluctuations affected intra- and inter-specific diversity of both temperate (Avise and Walker 1998; Carstens and Knowles 2007; Hewitt 1999) and tropical species (Gaither, et al. 2011; Gaither, et al. 2010; Hoareau, et al. 2012). Signatures of the effects of Pleistocene sea level changes have been revealed for marine organisms both in temperate (Marko, et al. 2010) and tropical environments (Gaither, et al. 2011; Han, et al. 2012; Hoareau, et al. 2012).

During the Pleistocene in the SWIO, sea levels rose and fell by up to 120 m (Pillans, et al. 1998) and it is possible that these sea-level changes affected the demographic history of *Lethrinus* sp. A and possibly *O. cyanea* in two ways: Firstly, the decreased sea level during glacial periods may have increased fragmentation of marine habitat and decreased areas of suitable available habitat. This increased fragmentation of habitat may have led to reduced gene flow between populations across the SWIO, driving divergence of populations during glacial periods. Secondly, interglacial periods resulted in raised sea levels (within 10m of present day sea level) facilitating population expansions due to increased habitat availability. Additionally, increased sea surface temperature at the edges of species' ranges may have also encouraged population expansion. It is therefore possible that in *L. nebulosus*, *L. harak* and *L. mahsena* the Pleistocene sea-level fluctuations stimulated a contraction-expansion dynamic of populations (Provan and Bennett 2008)

Table 8.2. Summary table displaying the possible effect of Pleistocene climate oscillations on the demographic history of the study species.

Species	Pleistocene climate effect on demographic history of study species
L. nebulosus	CR mtDNA revealed demographic expansion during the Pleistocene
L. mahsena	COI mtDNA indicated a population expansion approximately 190 KA during the late Pleistocene.
L. harak	Divergence of three CR mtDNA and population expansions, all occurring during the Pleistocene epoch.
O. cyanea	Population expansion of *O. cyanea* 5 - 189 KA (dependent on mutation rate), coinciding with the late Pleistocene climate fluctuations.

8.3 Implications of work

8.3.1 Underestimated biodiversity

Tropical reefs such as those in the SWIO are the most biologically diverse of shallow water marine ecosystems (Roberts, et al. 2002). Despite the SWIO's importance as a biodiversity hotspot, the SWIO remains relatively understudied (Ridgway and Sampayo 2005). The identification of two previously unknown cryptic species during this thesis supports this idea. It is highly likely that there are a number of cryptic and possibly endemic marine species in the SWIO which have not yet been documented. The biodiversity of the SWIO may therefore be significantly higher than current estimates suggest.

8.3.2 Fisheries management and biodiversity conservation

In keeping with the rest of the world, the SWIO is under ever increasing fishing pressure (van der Elst, et al. 2005). More than 60 million people reside within 100km of the coast in the Western Indian Ocean (excluding South Africa) and there is great dependence on marine resources for food and employment. In Mozambique alone 50% of the population's protein intake is from marine sources. However, current records suggest that the Western Indian Ocean is at or above its maximum sustainable yield of circa 4 million tonnes per annum and a number of stocks are in decline.

Lethrinus are a slow growing, long-lived fish species, making them vulnerable to overexploitation, particularly as more popular fish species become increasingly overfished. For this reason, previous studies into *Lethrinus* have highlighted the importance of species-specific knowledge of stock structure in the development of management strategies (van Herwerden, et al. 2009). aResearch on *L. miniatus* found isolated populations in East and West Australia with the West Australian populations displaying genetic impoverishment possibly due to selection. Consequently, the West Australian population will be less resilient to perturbations such as fishing and climate

change than the East Australian populations which have a high genetic diversity. The authors therefore suggest managing these two populations as discrete stocks and carefully monitoring the genetic diversity of the West Australian population as an indicator of population health.

Similarly, results obtained in the present work revealed that populations of *L. nebulosus* and *L. mahsena* are not panmictic populations across the SWIO, as was previously assumed and therefore should not be managed as a single unit. In fact, the population of *L. nebulosus* found off the coast of South Africa and southern Mozambique is likely to constitute a different species and should certainly be managed as a separate unit from the remaining *L. nebulosus* stocks in the SWIO. Unlike the van Herwerden, et al. (2009), the populations of *L. nebulosus* in the SWIO are not genetically impoverished. However, the *L. nebulosus* population found off the coast of South Africa and southern Mozambique displays considerably lower nucleotide diversity than the *L. nebulosus* population in the rest of the SWIO, making it more vulnerable to overexploitation than the SWIO population.

The taxonomic position of *Lethrinus* sp. A in the Seychelles is less clear. However, it is likely to be a separate species to *L. mahsena* and so should also be managed as a discrete stock.

As the SWIO's marine fauna is understudied (Ridgway and Sampayo 2005), further research is likely to reveal a number of cryptic species. The detection of cryptic species and accurate species identification is crucial for the preservation and maintenance of genetic biodiversity and management of sustainable fisheries. Without further research, unrecognised cryptic species and isolated populations will be particularly vulnerable to exploitation and extirpation.

Fisheries management of Mauritius should also be carefully considered if, as this research suggests, local processes are limiting gene flow between Mauritian populations and populations across the rest of the SWIO. Mauritian stocks may be more vulnerable to overexploitation than other locations across the SWIO. If a resource is overexploited in Mauritius, restocking of the Mauritian population from other locations in the SWIO may be limited, meaning that the Mauritian population may not recover. This appears to be the case in Mauritian *O. cyanea* fisheries. Mauritian catch statistics of *O. cyanea* show a sharp decrease in catch from 335 tons in 2002 to 84 tons in 2006 with catches remaining low from 2006 up to the latest FAO catch data in 2011, possibly suggesting overexploited populations in this area (FAO 2013). Despite the decrease in catch data, possible evidence of overfishing was not reflected in the data in the form of decreased diversity statistics. However, if, as hypothesised above, populations of *O. cyanea* are not at migration-drift equilibrium, genetic diversity of Mauritian *O. cyanea*

may not reflect contemporary diversity. It therefore seems possible that stocks of *O. cyanea* in Mauritius may have been overfished in recent years. The lack of overlapping generations in *O. cyanea*, meaning that if one generation is heavily exploited then recruitment to the next generation will be severely reduced, and the isolation of Mauritius, which may limit replenishment of populations by migration from other areas of the SWIO, may have led to the severe decline of *O. cyanea* stocks in Mauritius. We suggest however, that this considerable decrease in stocks is not yet detectable using the neutral markers utilised in this study due to migration-drift disequilibrium.

It is therefore suggested that fisheries of *O. cyanea* in Mauritius be scaled down considerably to allow the populations to recover. Additionally, due to the apparent isolation of Mauritian populations from the rest of the SWIO in a number of valuable fisheries species in this thesis, it would be expedient to set fairly conservative fishing quotas for Mauritian species in general.

Due to the relative instability and underdeveloped nature of the majority of the countries adjacent to the SWIO, fishing industries are not as developed as those in other oceans such as the Atlantic. However, both inshore subsistence fisheries and commercial fisheries have increased in the last decade (van der Elst, et al. 2005) and the total landings for the Indian Ocean are thought to have reached or exceeded their fully fished status (van der Elst, et al. 2005). Therefore, the initiation and designation of marine protected areas (MPAs) could be a useful method to maintain the genetic biodiversity of the SWIO.

8.3.3 **Climate change**

On-going climate change is an increasing threat and one of the major factors governing the survival of marine species (Briggs 2011; Dawson, et al. 2011). To date, climate change has been responsible for range shifts in a number of species (Cheung, et al. 2009; Perry, et al. 2005) and if warming continues it is predicted to lead to local extinctions in sub-polar regions, tropics and semi-enclosed seas, species invasions and further range shifts (Cheung, et al. 2009). As most fish possess specific temperature requirements for feeding, dispersal and spawning, climatic fluctuations may be predicted to have a severe impact on fish populations. How a species reacts to climate change will depend upon multiple factors such as body size, tropic level and generation time (Lo Brutto, et al. 2011).

Reconstruction of evolutionary histories for the study species suggest multiple past population expansions and possible contractions which appear to coincide with major climatic fluctuations during the Pleistocene epoch. Glacial periods appear to have encouraged divergence, possibly due to reduced sea levels and decreases in available

habitat. Conversely, interglacial periods appear to correlate with population expansions, possibly due to increased habitat availability as a result of raised sea levels. As such, it is likely that future climatic changes will influence population demography in these species.

8.4 Limitations

8.4.1 Sampling

As with most studies, one of the main limitations of this research was obtaining samples. This was due to difficulties in physically accessing some locations, particularly sites in Mozambique. In addition, obtaining sufficient sample sizes for each marker used was also challenging.

It was originally planned that samples would be collected from Somalia. Unfortunately due to the presence of Somali pirates in this region of the Indian Ocean it was considered unsafe to attempt collection of samples from this country.

8.4.2 Selection of markers

The studies on *Lethrinus* species utilised COI and CR mtDNA markers to investigate phylogeography and past demographic history and reconstruct phylogenies. However, as mtDNA is non-recombining, this study essentially used only one locus and these results and interpretations should therefore be treated as a tentative explanation of the demographic history and phylogeography of these species. The utilisation of several different single locus markers, including both nuclear and mtDNA markers is therefore preferable in studies of population biology (Ballard and Whitlock 2004; Sunnucks 2000). The value of obtaining additional data from independent markers is illustrated by the number of studies which have displayed conflicts between mitochondrial and nuclear data (Ballard and Whitlock 2004; Rognon and Guyomard 2003; Shaw 2002; Sota 2002) Analysis of nuclear markers should therefore be employed to further investigate the demographic history, phylogeography and phylogenetic reconstruction of these species. In particular, gene trees may not accurately reflect species trees due to introgression or selection acting upon mtDNA (Ballard and Whitlock 2004). Therefore for *L. nebulosus* and *Lethrinus* sp. A, where results suggested a possible speciation event, use of additional nuclear markers would be particularly valuable (Ballard and Whitlock 2004).

8.5 Future research

As mentioned in the previous section, further research in the form of independent nuclear markers should be carried out to confirm the phylogeography, demographic history and phylogenies of *Lethrinus* species used in this study.

The differentiation of Mauritius from all other populations across the SWIO in *L. harak* requires additional investigation. The results suggest contemporary isolation of Mauritius from the rest of the SWIO. To address this hypothesis further, microsatellite markers could be used to investigate in detail population structuring and gene flow of *L. harak* across the SWIO. The results obtained from this study would also be valuable for the management of *L. harak* fisheries across the SWIO if microsatellites confirmed the existence of an isolated Mauritian population. However, since the results suggest that populations are not at equilibrium, microsatellites may overestimate current gene flow in *L. harak* populations in the SWIO.

The results obtained in this thesis suggest that population structuring in *O. cyanea* and *L. harak* may be very subtle and beyond the resolution of traditional neutral markers. Although current molecular markers are extremely powerful in investigating neutral genetic variation, allowing insights into demographic processes, they are unable to reveal adaptive genetic differentiation. Most technological developments have focused on neutral genetic variation, treating selection as a nuisance factor that is difficult to estimate quantitatively and should therefore be avoided (Hauser and Seeb 2008). This simplification has allowed the application of theoretical models of neutral evolution but has ignored a major part of the biodiversity of these organisms, which may be used to identify biologically important differences among demes.

The use of markers under selection to infer adaptive genetic diversity between populations is still in its infancy (Hauser and Seeb 2008). However with the advent of genome scans which are able to screen genome-wide patterns of DNA polymorphism to detect the locus-specific signature of positive directional selection, the use of multiple markers to investigate adaptive population divergence is well under way.

8.6 **Summary**

This thesis represents the first population genetic study of a heavily fished species *O. cyanea*, across the SWIO. Results revealed high gene flow across the SWIO, with little population structuring. However, the results also revealed that populations were not at migration-drift equilibrium and hence current connectivity of populations across the SWIO may be overestimated. This thesis has also highlighted the possible isolation of Mauritius through local processes in a number of species, identified two possible cryptic species and revealed that past population fluctuations of the study species may be linked to Pleistocene climate fluctuations.

8.7 **References**

André C, Larsson LC, Laikre L, Bekkevold D, Brigham J, Carvalho GR, Dahlgren TG, Hutchinson WF, Mariani S, Mudde K, Ruzzante DE, Ryman N 2011. Detecting

population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity 106: 270-280. doi: 10.1038/hdy.2010.71

Avise JC, Walker D 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings of the Royal Society of London. Series B: Biological Sciences 265: 457-463. doi: 10.1098/rspb.1998.0317

Ballard JWO, Whitlock MC 2004. The incomplete natural history of mitochondria. Molecular Ecology 13: 729-744.

Berg H, Francis J, Souter P 2002. Support to marine research for sustainable management of marine and coastal resources in the western Indian Ocean. Ambio 31: 597-601. doi: 10.2307/4315314

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22: 148-155. doi: 10.1016/j.tree.2006.11.004

Borsa P, Hsiao DR, Carpenter KE, Chen WJ 2013. Cranial morphometrics and mitochondrial DNA sequences distinguish cryptic species of the longface emperor (Lethrinus olivaceus), an emblematic fish of Indo-West Pacific coral reefs. Comptes Rendus Biologies 336: 505-514. doi: 10.1016/j.crvi.2013.09.004

Briggs JC 2011. Marine extinctions and conservation. Marine Biology 158: 485-488. doi: 10.1007/s00227-010-1596-0

Carpenter KE, Allen GR. 1989. FAO Species Calalogue Vol 9. Emperor fishes and large-eye breams of the world (family Lethrinidae). In. Food and Agricultural Organisation of the United Nations. Rome: FAO.

Carstens BC, Knowles LL 2007. Shifting distributions and speciation: species divergence during rapid climate change. Molecular Ecology 16: 619-627. doi: 10.1111/j.1365-294X.2006.03167.x

Cheung WWL, Lam VWY, Sarmiento JL, Kearney K, Watson R, Pauly D 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish and Fisheries 10: 235-251.

Colborn J, Crabtree RE, Shaklee JB, Pfeiler E, Bowen BW 2001. The evolutionary enigma of bonefishes (*Albula* spp.): cryptic species and ancient separations in a globally distributed shorefish. Evolution 55: 807-820.

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM 2011. Beyond predictions: Biodiversity conservation in a changing climate. Science 332: 53-58. doi: 10.1126/science.1200303

Dorenbosch M, Pollux BJA, Pustjens AZ, Rajagopal S, Nagelkerken I, van der Velde G, van der Staay SYM 2006. Population structure of the Dory snapper, *Lutjanus fulviflamma*, in the western Indian Ocean revealed by means of AFLP fingerprinting. Hydrobiologia 568: 43-53. doi: 10.1007/s10750-006-0020-8

Ebisawa A, Ozawa T 2009. Life-history traits of eight *Lethrinus* species from two local populations in waters off the Ryukyu Islands. Fisheries Science 75: 553-566. doi: 10.1007/s12562-009-0061-9

FAO. 2013. Fishery and Aquaculture Statistics 2011 In. FAO Yearbook 2011. Rome: Food and Agriculture Organisation of the United Nations.

Fisher R, Leis JM, Clark DL, Wilson SK 2005. Critical swimming speeds of late-stage coral reef fish larvae: variation within species, among species and between locations. Marine Biology 147: 1201-1212. doi: 10.1007/s00227-005-0001-x

Fratini S, Ragionieri L, Cannicci S 2010. Stock structure and demographic history of the Indo-West Pacific mud crab *Scylla serrata*. Estuarine, Coastal and Shelf Science 86: 51-61. doi: 10.1016/j.ecss.2009.10.009

Gaither MR, Bowen BW, Bordenave TR, Rocha LA, Newman SJ, Gomez JA, van Herwerden L, Craig MT 2011. Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle. BMC Evolutionary Biology 11: 189. doi: 10.1186/1471-2148-11-189

Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW 2010. Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). Journal of Biogeography 37: 133-147. doi: 10.1111/j.1365-2699.2009.02188.x

Griffiths AM, Sims DW, Cotterell SP, El Nagar A, Ellis JR, Lynghammar A, McHugh M, Neat FC, Pade NG, Queiroz N, Serra-Pereira B, Rapp T, Wearmouth VJ, Genner MJ 2010. Molecular markers reveal spatially segregated cryptic species in a critically endangered fish, the common skate (*Dipturus batis*). Proceedings of the Royal Society B: Biological Sciences 277: 1497-1503. doi: 10.1098/rspb.2009.2111

Griffiths CL 2005. Coastal marine biodiversity in East Africa. Indian Journal of Marine Science 34: 35-41.

Han Z, Yanagimoto T, Zhang Y, Gao T 2012. Phylogeography study of *Ammodytes personatus* in Northwestern Pacific: Pleistocene isolation, temperature and current conducted secondary contact. PLoS ONE 7: e37425. doi: 10.1371/journal.pone.0037425

Hauser L, Carvalho GR 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9: 333-362. doi: 10.1111/j.1467-2979.2008.00299.x

Hauser L, Seeb JE 2008. Advances in molecular technology and their impact on fisheries genetics. Fish and Fisheries 9: 473-486. doi: 10.1111/j.1467-2979.2008.00306.x

Hewitt GM 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68: 87-112. doi: 10.1111/j.1095-8312.1999.tb01160.x

Hoareau TB, Boissin E, Berrebi P 2012. Evolutionary history of a widespread Indo-Pacific goby: the role of Pleistocene sea-level changes on demographic contraction/expansion dynamics. Molecular Phylogenetics and Evolution 62: 566-572. doi: 10.1016/j.ympev.2011.10.004

Kon T, Yoshino T, Mukai T, Nishida M 2007. DNA sequences identify numerous cryptic species of the vertebrate: a lesson from the gobioid fish *Schindleria*. Molecular Phylogenetics and Evolution 44: 53-62. doi: 10.1016/j.ympev.2006.12.007

Lo Brutto S, Arculeo M, Stewart Grant W 2011. Climate change and population genetic structure of marine species. Chemistry and Ecology 27: 107-119. doi: 10.1080/02757540.2010.547486

Marko PB, Hoffman JM, Emme SA, McGovern TM, Keever CC, Cox LN 2010. The 'Expansion-Contraction' model of Pleistocene biogeography: rocky shores suffer a sea change? Molecular Ecology 19: 146-169.

Muths D, Gouws G, Mwale M, Tessier E, Bourjea J, Moran P 2012. Genetic connectivity of the reef fish *Lutjanus kasmira* at the scale of the western Indian Ocean. Canadian Journal of Fisheries and Aquatic Sciences 69: 842-853.

Muths D, Grewe P, Jean C, Bourjea J 2009. Genetic population structure of the swordfish (*Xiphias gladius*) in the southwest Indian Ocean: Sex-biased differentiation, congruency between markers and its incidence in a way of stock assessment. Fisheries Research 97: 263-269. doi: 10.1016/j.fishres.2009.03.004

Muths D, Le Couls S, Evano H, Grewe P, Bourjea J 2013. Multi-genetic marker approach and spatio-temporal analysis suggest there is a single panmictic population

of Swordfish *Xiphias gladius* in the Indian Ocean. PLoS ONE 8: e63558. doi: 10.1371/journal.pone.0063558

Nielsen EE, Hansen MM, Meldrup D 2006. Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua L.*): implications for inferring population structure in nonmodel organisms. Molecular Ecology 15: 3219-3229. doi: 10.1111/j.1365-294X.2006.03025.x

Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger GL, Ogden R, Carvalho GR 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). BMC Evolutionary Biology 9: 276. doi: 10.1186/1471-2148-9-276

O'Malley KG, Camara MD, Banks Ma 2007. Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). Molecular Ecology 16: 4930-4941. doi: 10.1111/j.1365-294X.2007.03565.x

Obura D 2012. The diversity and biogeography of Western Indian Ocean reef-building corals. PLoS ONE 7: e45013. doi: 10.1371/journal.pone.0045013

Perry AL, Low PJ, Ellis JR, Reynolds JD 2005. Climate change and distribution shifts in marine fishes. Science 308: 1912-1915. doi: 10.1126/science.1111322

Pillans B, Chappell J, Naish TR 1998. A review of the Milankovitch climatic beat: template for Plio–Pleistocene sea-level changes and sequence stratigraphy. Sedimentary Geology 122: 5-21. doi: 10.1016/S0037-0738(98)00095-5

Provan J, Bennett KD 2008. Phylogeographic insights into cryptic glacial refugia. Trends in Ecology and Evolution 23: 564-571. doi: 10.1016/j.tree.2008.06.010

Quartly GD, Srokosz MA 2004. Eddies in the southern Mozambique Channel. Deep-Sea Research Part II-Topical Studies in Oceanography 51: 69-83. doi: 10.1016/j.dsr2.2003.03.001

Ragionieri L, Cannicci S, Schubart CD, Fratini S 2010. Gene flow and demographic history of the mangrove crab *Neosarmatium meinerti*: A case study from the western Indian Ocean. Estuarine, Coastal and Shelf Science 86: 179-188. doi: 10.1016/j.ecss.2009.11.002

Ridgway T, Sampayo EM 2005. Population genetic status of the Western Indian Ocean: What do we know? Western Indian Ocean Journal of Marine Science 4: 1-9.

Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, Vynne C, Werner TB 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science 295: 1280-1284. doi: 10.1126/science.1067728

Rognon X, Guyomard R 2003. Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. Molecular Ecology 12: 435-445.

Santos S, Hrbek T, Farias IP, Schneider H, Sampaio I 2006. Population genetic structuring of the king weakfish, *Macrodon ancylodon* (Sciaenidae), in Atlantic coastal waters of South America: deep genetic divergence without morphological change. Molecular Ecology 15: 4361-4373. doi: 10.1111/j.1365-294X.2006.03108.x

Schott FA, Xie S-P, McCreary JP 2009. Indian Ocean circulation and climate variability. Reviews of Geophysics 47: RG1002. doi: 10.1029/2007RG000245

Schouten MW, de Ruijter WPM, van Leeuwen PJ, Ridderinkhof H 2003. Eddies and variability in the Mozambique Channel. Deep-Sea Research Part II-Topical Studies in Oceanography 50: 1987-2003. doi: 10.1016/S0967-0645(03)00042-0

Shaw KL 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation

in Hawaiian crickets. Proceedings of the National Academy of Sciences 99: 16122-16127. doi: 10.1073/pnas.242585899

Silva IC, Mesquita N, Paula J 2010. Lack of population structure in the fiddler crab *Uca annulipes* along an East African latitudinal gradient: genetic and morphometric evidence. Marine Biology 157: 1113-1126. doi: 10.1007/s00227-010-1393-9

Sota T 2002. Radiation and reticulation: extensive introgressive hybridization in the carabid beetles *Ohomopterus* inferred from mitochondrial gene genealogy. Population Ecology 44: 145-156. doi: 10.1007/s101440200018

Sunnucks P 2000. Efficient genetic markers for population biology. Trends in Ecology and Evolution 15: 199-203. doi: 10.1016/S0169-5347(00)01825-5

van der Elst R, Everett B, Jiddawi N, Mwatha G, Afonso P, Boulle D 2005. Fish, fishers and fisheries of the Western Indian Ocean: their diversity and status. A preliminary assessment. Philosophical transactions. Series A, Mathematical, Physical, and Engineering Sciences 363: 263-284. doi: 10.1098/rsta.2004.1492

van Herwerden L, Aspden WJ, Newman SJ, Pegg GG, Briskey L, Sinclair W 2009. A comparison of the population genetics of Lethrinus miniatus and Lutjanus sebae from the east and west coasts of Australia: Evidence for panmixia and isolation. Fisheries Research 100: 148-155. doi: 10.1016/j.fishres.2009.07.003

van Heukelem WFV 1973. Growth and life-span of Octopus cyanea (Mollusca: Cephalopoda). Journal of Zoology 169: 299-315. doi: 10.1111/j.1469-7998.1973.tb04559.x

Villanueva R, Norman MD 2008. Biology of the planktonic stages of benthic octopuses. Oceanography and Marine Biology: An Annual Review 46: 105-202.

Visram S, Yang M-C, Pillay RM, Said S, Henriksson O, Grahn M, Chen CA 2010. Genetic connectivity and historical demography of the blue barred parrotfish (*Scarus ghobban*) in the Western Indian Ocean. Marine Biology 157: 1475-1487. doi: 10.1007/s00227-010-1422-8

Watanabe K, Kawase S, Mukai T, Kakioka R, Miyazaki J-I, Hosoya K 2010. Population divergence of *Biwia zezera* (Cyprinidae: Gobioninae) and the discovery of a cryptic species, based on mitochondrial and nuclear DNA sequence analyses. Zoological Science 27: 647-655. doi: 10.2108/zsj.27.647

8.8

Abercrombie D, Clarke S, Shivji M 2005. Global-scale genetic identification of hammerhead sharks: Application to assessment of the international fin trade and law enforcement. Conservation Genetics 6: 775-788. doi: 10.1007/s10592-005-9036-2

Akasaki T, Nikaido M, Tsuchlya K, Segawa S, Hasegawa M, Okada N 2006. Extensive mitochondrial gene arrangements in coleoid Cephalopoda and their phylogenetic implications. Molecular Phylogenetics and Evolution 38: 648-658. doi: 10.1016/j.ympev.2005.10.018

Aldonov VK, Druzhinin AD 1979. Some data on scavengers (family Lethrinidae) from the Gulf of Aden region. Journal of Ichthyology 18: 527-535.

Ali JR, Huber M 2010. Mammalian biodiversity on Madagascar controlled by ocean currents. Nature 463: 653-656.

André C, Larsson LC, Laikre L, Bekkevold D, Brigham J, Carvalho GR, Dahlgren TG, Hutchinson WF, Mariani S, Mudde K, Ruzzante DE, Ryman N 2011. Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*):

direct, simultaneous evaluation of neutral vs putatively selected loci. Heredity 106: 270-280. doi: 10.1038/hdy.2010.71

Angers B, Bernatchez L 1998. Combined use of SMM and non-SMM methods to infer fine structure and evolutionary history of closely related brook charr (*Salvelinus fontinalis*, Salmonidae) populations from microsatellites. Molecular Biology and Evolution: 143-159.

Angers B, Magnan P, Plante M, Bernatchez L 1999. Canonical correspondence analysis for estimating spatial and environmental effects on microsatellite gene diversity in brook charr (*Salvelinus fontinalis*). Molecular Ecology 8: 1043-1053. doi: 10.1046/j.1365-294x.1999.00669.x

Anisimova M, Gascuel O 2006. Approximate likelihood ratio test for branches: a fast, accurate, and powerful alternative. Systematic Biology 55: 539-552. doi: 10.1080/10635150600755453

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G 2008. LOSITAN: A workbench to detect molecular adaptation based on a F_{ST} -outlier method. BMC Bioinformatics 9: 323. doi: 10.1186/1471-2105-9-323

Ardill JD, Sanders MJ. 1991. Priorities for fisheries management and development in the South West Indian Ocean. In. FAO Fisheries Rome: FAO.

Asensio L, Samaniego L 2009. Rapid identification of grouper and wreck fish meals by ELISA: a field study in restaurants. International Journal of Food Science & Technology 44: 1585-1589. doi: 10.1111/j.1365-2621.2008.01857.x

Avise JC, Walker D 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proceedings of the Royal Society of London. Series B: Biological Sciences 265: 457-463. doi: 10.1098/rspb.1998.0317

Ayre DJ, Hughes TP 2004. Climate change, genotypic diversity and gene flow in reef-building corals. Ecology Letters 7: 273-278. doi: 10.1111/j.1461-0248.2004.00585.x

Badyukov DD, Demidenko EL, Kaplin PA 1989. Paleogeography of the Seychelles Bank and the northwest Madagascar Shelf during the last glacio-eustatic regression (18,000 a B.P.). Chinese Journal of Oceanology and Limnology 7: 89-92. doi: 10.1007/BF02842661

Ballard JWO, Whitlock MC 2004. The incomplete natural history of mitochondria. Molecular Ecology 13: 729-744.

Balloux F, Lugon-Moulin N 2002. The estimation of population differentiation with microsatellite markers. Molecular Ecology 11: 155-165. doi: 10.1046/j.0962-1083.2001.01436.x

Baum BR 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. Taxon 41: 3-10.

Bautil B, Samboo CR. 1988. Preliminary stock assessment for the mahsena emperor (*Lethrinus mahsena*) on the Nazareth Bank of Mauritius. In: Sanders MJ, Sparre P, Venema SC, editors. Proceedings of the workshop on the assessment of the fishery resources in the Southwest Indian Ocean. Albion, Mauritius. p. 260-267.

Bazin E, Glémin S, Galtier N 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312: 570-572. doi: 10.1126/science.1122033

Beaumont MA 2005. Adaptation and speciation: what can F_{ST} tell us? Trends in Ecology and Evolution 20: 435-440. doi: 10.1016/j.tree.2005.05.017

Beaumont MA, Nichols RA 1996. Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society B: Biological Sciences 263: 1619-1626. doi: 10.1098/rspb.1996.0237

Beebee T, Rowe G. 2001. An introduction to molecular ecology. Oxford, UK: Oxford University Press.

Bennett P 2000. Demystified ... microsatellites. Molecular Pathology 53: 177-183.

Benson G 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27: 573-580.

Berg H, Francis J, Souter P 2002. Support to marine research for sustainable management of marine and coastal resources in the western Indian Ocean. Ambio 31: 597-601. doi: 10.2307/4315314

Bermingham E, McCafferty SS, Martin AP. 1997. Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: Kocher TD, Stepien, C A, editor. Molecular Systematics of Fishes. San Diego: Academic Press. p. 113-128.

Berry O, England P, Marriott RJ, Burridge CP, Newman SJ 2012. Understanding age-specific dispersal in fishes through hydrodynamic modelling, genetic simulations and microsatellite DNA analysis. Molecular Ecology 21: 2145-2159. doi: 10.1111/j.1365-294X.2012.05520.x

Bertrand J 1986. Data on the reproduction of *Lethrinus mahsena* on the Banks of the Saya-de-Malha Indian Ocean. Cybium 10: 15-29.

Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22: 148-155. doi: 10.1016/j.tree.2006.11.004

Borsa P, Hsiao DR, Carpenter KE, Chen WJ 2013. Cranial morphometrics and mitochondrial DNA sequences distinguish cryptic species of the longface emperor (Lethrinus olivaceus), an emblematic fish of Indo-West Pacific coral reefs. Comptes Rendus Biologies 336: 505-514. doi: 10.1016/j.crvi.2013.09.004

Briggs JC 2011. Marine extinctions and conservation. Marine Biology 158: 485-488. doi: 10.1007/s00227-010-1596-0

Brothers EB, Williams DMB, Sale PF 1983. Length of larval life in twelve families of fishes at “One Tree Lagoon”, Great Barrier Reef, Australia. Marine Biology 76: 319-324. doi: 10.1007/BF00393035

Cabin RJ, Mitchell RJ 2000. To Bonferroni or not to Bonferroni: when and how are the questions. Bulletin of the Ecological Society of America 81: 246-248. doi: 10.2307/20168454

Cabranes C, Fernandez-Rueda P, Martínez JL 2008. Genetic structure of *Octopus vulgaris* around the Iberian Peninsula and Canary Islands as indicated by microsatellite DNA variation. ICES Journal of Marine Science 65: 12-16. doi: 10.1093/icesjms/fsm178

Camoin GF, Colonna M, Montaggioni LF, Casanova J, Faure G, Thomassin BA 1997. Holocene sea level changes and reef development in the southwestern Indian Ocean. Coral Reefs 16: 247-259. doi: 10.1007/s003380050080

Camoin GF, Montaggioni LF, Braithwaite CJR 2004. Late glacial to post glacial sea levels in the western Indian Ocean. Marine Geology 206: 119-146. doi: 10.1016/j.margeo.2004.02.003

Cane MA, Molnar P 2001. Closing of the Indonesian seaway as a precursor to east African aridification around 3-4 million years ago. Nature 411: 157-162.

Carpenter KE, Allen GR. 1989. FAO Species Calalogue Vol 9. Emperor fishes and large-eye breams of the world (family Lethrinidae). In. Food and Agricultural Organisation of the United Nations. Rome: FAO.

Carpenter KE, J.E. R 2003. *Lethrinus ravus*, a new species of emperor fish (Perciformeds: Lethrinidae) from the western Pacific and eastern Indian Oceans. Zootaxa 240: 1-8.

Carroll SP, Hendry AP, Reznick DN, Fox CW 2007. Evolution on ecological time-scales. Functional Ecology 21: 387-393. doi: 10.1111/j.1365-2435.2007.01289.x

Carstens BC, Knowles LL 2007. Shifting distributions and speciation: species divergence during rapid climate change. Molecular Ecology 16: 619-627. doi: 10.1111/j.1365-294X.2006.03167.x

Cawthorn D-M, Steinman HA, Witthuhn RC 2012. DNA barcoding reveals a high incidence of fish species misrepresentation and substitution on the South African market. Food Research International 46: 30-40. doi: 10.1016/j.foodres.2011.11.011

Céspedes A, García T, Carrera E, González I, Fernández A, Asensio L, Hernández PE, Martín R 2000. Genetic differentiation between sole (*Solea solea*) and Greenland halibut (*Reinhardtius hippoglossoides*) by PCR–RFLP analysis of a 12S rRNA gene fragment. Journal of the Science of Food and Agriculture 80: 29-32. doi: 10.1002/1097-0010

Chapuis M-P, Estoup A 2007. Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24: 621-631. doi: 10.1093/molbev/msl191

Charlesworth B 1998. Measures of divergence between populations and the effect of forces that reduce variability. Molecular Biology and Evolution 15: 538-543.

Cheung WWL, Lam VWY, Sarmiento JL, Kearney K, Watson R, Pauly D 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish and Fisheries 10: 235-251.

Chiang H-C, Hsu C-C, Wu GC-C, Chang S-K, Yang H-Y 2008. Population structure of bigeye tuna (*Thunnus obesus*) in the Indian Ocean inferred from mitochondrial DNA. Fisheries Research 90: 305-312. doi: 10.1016/j.fishres.2007.11.006

Christie MR, Tissot BN, Albins MA, Beets JP, Jia Y, Ortiz DM, Thompson SE, Hixon MA 2010. Larval connectivity in an effective network of marine protected areas. PLoS ONE 5: e15715. doi: 10.1371/journal.pone.0015715

Civera T 2003. Species identification and safety of fish products. Veterinary Research Communications 27: 481-489. doi: 10.1023/B:VERC.0000014205.87859.ab

Colborn J, Crabtree RE, Shaklee JB, Pfeiler E, Bowen BW 2001. The evolutionary enigma of bonefishes (*Albula* spp.): cryptic species and ancient separations in a globally distributed shorefish. Evolution 55: 807-820.

Coscia I, Mariani S 2011. Phylogeography and population structure of European sea bass in the North-East Atlantic. Biological Journal of the Linnean Society 104: 364-377. doi: 10.1111/j.1095-8312.2011.01712.x

Courchamp F, Angulo E, Rivalan P, Hall RJ, Signoret L, Bull L, Meinard Y 2006. Rarity value and species extinction: The anthropogenic allee effect. PLoS Biology 4: e415. doi: 10.1371/journal.pbio.0040415

Cowen RK, Sponaugle S 2009. Larval dispersal and marine population connectivity. Annual Review of Marine Science 1: 443-466. doi: 10.1146/annurev.marine.010908.163757

Craig MT, Eble JA, Bowen BW, Robertson DR 2007. High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). Marine Ecology Progress Series 334: 245-254.

Cronin MA, Palmisciano DA, Vyse ER, Cameron DG 1991. Mitochondrial DNA in wildlife forensic science: species identification of tissues. Wildlife Society Bulletin 19: 94-105. doi: 10.2307/3782423

Darriba D, Taboada GL, Doallo R, Posada D 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9: 772. doi:

http://www.nature.com/nmeth/journal/v9/n8/abs/nmeth.2109.html#supplementary-information

Dawnay N, Ogden R, McEwing R, Carvalho GR, Thorpe RS 2007. Validation of the barcoding gene COI for use in forensic genetic species identification. Forensic Science International 173: 1-6. doi: 10.1016/j.forsciint.2006.09.013

Dawson TP, Jackson ST, House JI, Prentice IC, Mace GM 2011. Beyond predictions: Biodiversity conservation in a changing climate. Science 332: 53-58. doi: 10.1126/science.1200303

de Ruijter WPM, Ridderinkhof H, Lutjeharms JRE, Schouten MW, Veth C 2002. Observations of the flow in the Mozambique Channel. Geophysical Research Letters 29: 1401-1403. doi: 10.1029/2001GL013714

de Ruijter WPM, van Aken HM, Beier EJ, Lutjeharms JRE, Matano RP, Schouten MW 2004. Eddies and dipoles around South Madagascar: formation, pathways and large-scale impact. Deep-Sea Research Part I-Oceanographic Research Papers 51: 383-400. doi: 10.1016/j.dsr.2003.10.011

DeSalle R, Egan MG, Siddall M 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society B: Biological Sciences 360: 1905-1916. doi: 10.1098/rstb.2005.1722

Dew B 1959. Some observations on the development of two Australian octopuses. Proceedings of the Royal Society of New South Wales 1957-1958: 44-52.

Donaldson KA, Wilson RR 1999. Amphi-panamic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA control region of fishes. Molecular Phylogenetics and Evolution 13: 208-213. doi: 10.1006/mpev.1999.0625

Dorenbosch M, Pollux BJA, Pustjens AZ, Rajagopal S, Nagelkerken I, van der Velde G, van der Staay SYM 2006. Population structure of the Dory snapper, *Lutjanus fulviflamma*, in the western Indian Ocean revealed by means of AFLP fingerprinting. Hydrobiologia 568: 43-53. doi: 10.1007/s10750-006-0020-8

Doubleday ZA, Semmens JM, Smolenski AJ, Shaw PW 2009. Microsatellite DNA markers and morphometrics reveal a complex population structure in a merobenthic octopus species (*Octopus maorum*) in south-east Australia and New Zealand. Marine Biology 156: 1183-1192. doi: 10.1007/s00227-009-1160-y

Drummond AJ, Suchard MA, Xie D, Rambaut A 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29: 1969-1973. doi: 10.1093/molbev/mss075

Duda Jr TF, Palumbi SR 1999. Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations. Marine Biology 134: 705-710. doi: 10.1007/s002270050586

Earl D, vonHoldt B 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361. doi: 10.1007/s12686-011-9548-7

Ebisawa A 2006. Reproductive and sexual characteristics in five *Lethrinus* species in waters off the Ryukyu Islands. Ichthyological Research 53: 269-280.

Ebisawa A, Ozawa T 2009. Life-history traits of eight *Lethrinus* species from two local populations in waters off the Ryukyu Islands. Fisheries Science 75: 553-566. doi: 10.1007/s12562-009-0061-9

Eckert CG, Samis KE, Lougheed SC 2008. Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. Molecular Ecology 17: 1170-1188. doi: 10.1111/j.1365-294X.2007.03659.x

Edwards RRC, Bakhader A, Shaher S 1985. Growth, mortality, age composition and fisheries yields of fish from the Gulf of Aden. Journal of Fish Biology 27: 13-21.

Estoup A, Cornuet JM. 1999. Microsatellite evolution: inferences from population data. In: Goldstein DB, Schlotterer C, editors. Microsatellites - Evolution and Applications. Oxford, UK: Oxford University Press. p. 49-65.

Evanno G, Regnaut S, Goudet J 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620. doi: 10.1111/j.1365-294X.2005.02553.x

Ewing G. 2010. Hapview http://www.cibiv.at/~%20greg/haploviewer. Version 1.6.

Excoffier L, Lischer HEL 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10: 564-567. doi: 10.1111/j.1755-0998.2010.02847.x

Excoffier L, Smouse PE, Quattro JM 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application. Genetics 131: 479-491.

FAO. 2013. Fishery and Aquaculture Statistics 2011 In. FAO Yearbook 2011. Rome: Food and Agriculture Organisation of the United Nations.

Felsenstein J 1989. PHYLIP - Phylogeny inference package (Version 3.2). Cladistics 5: 164-166.

Féral JP 2002. How useful are the genetic markers in attempts to understand and manage marine biodiversity? Journal of Experimental Marine Biology and Ecology 268: 121-145. doi: citeulike-article-id:195460

Filonzi L, Chiesa S, Vaghi M, Nonnis Marzano F 2010. Molecular barcoding reveals mislabelling of commercial fish products in Italy. Food Research International 43: 1383-1388. doi: 10.1016/j.foodres.2010.04.016

Fischer W, Bianchi G. 1984. FAO species identification sheets for fishery purposes. Western Indian Ocean; (Fishing Area 51). In. Food and Agricultural Organisation of the United Nations. Rome: FAO.

Fisher R, Leis JM, Clark DL, Wilson SK 2005. Critical swimming speeds of late-stage coral reef fish larvae: variation within species, among species and between locations. Marine Biology 147: 1201-1212. doi: 10.1007/s00227-005-0001-x

Fleming K, Johnston P, Zwartz D, Yokoyama Y, Lambeck K, Chappell J 1998. Refining the eustatic sea-level curve since the Last Glacial Maximum using far- and intermediate-field sites. Earth and Planetary Science Letters 163: 327-342. doi: 10.1016/S0012-821X(98)00198-8

Floeter SR, Rocha LA, Robertson DR, Joyeux JC, Smith-Vaniz WF, Wirtz P, Edwards AJ, Barreiros JP, Ferreira CEL, Gasparini JL, Brito A, Falcón JM, Bowen BW, Bernardi G 2008. Atlantic reef fish biogeography and evolution. Journal of Biogeography 35: 22-47. doi: 10.1111/j.1365-2699.2007.01790.x

Flores J-A, Gersonde R, Sierro FJ 1999. Pleistocene fluctuations in the Agulhas Current retroflection based on the calcareous plankton record. Marine Micropaleontology 37: 1-22.

Frankham R 2005. Genetics and extinction. Biological Conservation 126: 131-140. doi: 10.1016/j.biocon.2005.05.002

Fraser CI, Spencer HG, Waters JM 2009. Glacial oceanographic contrasts explain phylogeography of Australian bull kelp. Molecular Ecology 18: 2287-2296. doi: 10.1111/j.1365-294X.2009.04201.x

Fratini S, Ragionieri L, Cannicci S 2010. Stock structure and demographic history of the Indo-West Pacific mud crab *Scylla serrata*. Estuarine, Coastal and Shelf Science 86: 51-61. doi: 10.1016/j.ecss.2009.10.009

Fu Y-X 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915-925.

Gaither MR, Bowen BW, Bordenave TR, Rocha LA, Newman SJ, Gomez JA, van Herwerden L, Craig MT 2011. Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle. BMC Evolutionary Biology 11: 189. doi: 10.1186/1471-2148-11-189

Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW 2010. Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). Journal of Biogeography 37: 133-147. doi: 10.1111/j.1365-2699.2009.02188.x

Galarza JA, Carreras-Carbonell J, Macpherson E, Pascual M, Roques S, Turner GF, Rico C 2009. The influence of oceanographic fronts and early-life-history traits on connectivity among littoral fish species. Proceedings of the National Academy of Sciences 106: 1473-1478. doi: 10.1073/pnas.0806804106

García-Rodríguez FJ, Perez-Enriquez R 2006. Genetic differentiation of the California spiny lobster *Panulirus interruptus* (Randall, 1840) along the west coast of the Baja California Peninsula, Mexico. Marine Biology 148: 621-629. doi: 10.1007/s00227-005-0101-7

Garcia-Vazquez E, Perez J, Martinez JL, Pardiñas AF, Lopez B, Karaiskou N, Casa MF, Machado-Schiaffino G, Triantafyllidis A 2010. High level of mislabeling in Spanish and Greek hake markets suggests the fraudulent introduction of African species. Journal of Agricultural and Food Chemistry 59: 475-480. doi: 10.1021/jf103754r

Garland ED, Zimmer, C.A 2002. Techniques for identification of bivalve larvae. Marine Ecology Progress Series 225: 299-310.

Garza JC, Williamson EG 2001. Detection of reduction in population size using data from microsatellite loci. Molecular Ecology 10: 305-318.

Gemayel R, Vinces MD, Legendre M, Verstrepen KJ 2010. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annual Review of Genetics 44: 445-477. doi: 10.1146/annurev-genet-072610-155046

Gerlach G, Atema J, Kingsford MJ, Black KP, Miller-Sims V 2007. Smelling home can prevent dispersal of reef fish larvae. Proceedings of the National Academy of Sciences 104: 858-863. doi: 10.1073/pnas.0606777104

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P 2010. Calculations of population differentiation based on G_{ST} and D: forget G_{ST} but not all of statistics! Molecular Ecology 19: 3845-3852. doi: 10.1111/j.1365-294X.2010.04784.x

Gillooly JF, Allen AP, West GB, Brown JH 2005. The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proceedings of the National Academy of Sciences of the United States of America A 102: 140-145. doi: 10.1073/pnas.0407735101

Glenn TC, Schable NA 2005. Isolating microsatellite DNA loci. Methods in Enzymology 395: 202-222. doi: 10.1016/s0076-6879(05)95013-1

Goffredi S, Jones W, Scholin C, Marin R, III, Vrijenhoek R 2006. Molecular detection of marine invertebrate larvae. Marine Biotechnology 8: 149-160. doi: 10.1007/s10126-005-5016-2

Gopal K, Tolley KA, Groeneveld JC, Matthee CA 2006. Mitochondrial DNA variation in spiny lobster *Palinurus delagoae* suggests genetically structured populations in the southwestern Indian Ocean. Marine Ecology Progress Series 319: 191-198.

Goudet J 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity 86: 485-486.

Goudet J, Raymond M, de Meeüs T, Rousset F 1996. Testing differentiation in diploid populations. Genetics 144: 1933-1940.

Gourlan AT, Meynadier L, Allègre CJ 2008. Tectonically driven changes in the Indian Ocean circulation over the last 25 Ma: Neodymium isotope evidence. Earth and Planetary Science Letters 267: 353-364. doi: 10.1016/j.epsl.2007.11.054

Gouws G. 2012. Genetic stock assessment of the blue emperor (*Lethrinus nebulosus*) in the South West Indian Ocean. In. South African Institute for Aquatic Biodiversity National Research Foundation. p. 1-32.

Grandcourt EM 2002. Demographic characteristics of a selection of exploited reef fish from the Seychelles: preliminary study. Marine and Freshwater Research 53: 123-130. doi: 10.1071/MF01123

Grandcourt EM, Al Abdessalaam TZ, Al Shamsi AT, Francis F 2006. Biology and assessment of the painted sweetlips (*Diagramma pictum* (Thunberg, 1792)) and the spangled emperor (*Lethrinus nebulosus* (Forsskål, 1775)) in the southern Arabian Gulf. Fishery Bulletin 104: 75-88.

Grant WS, Bowen BW 1998. Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. Journal of Heredity 89: 415-426. doi: 10.1093/jhered/89.5.415

Griffiths AM, Sims DW, Cotterell SP, El Nagar A, Ellis JR, Lynghammar A, McHugh M, Neat FC, Pade NG, Queiroz N, Serra-Pereira B, Rapp T, Wearmouth VJ, Genner MJ 2010. Molecular markers reveal spatially segregated cryptic species in a critically endangered fish, the common skate (*Dipturus batis*). Proceedings of the Royal Society B: Biological Sciences 277: 1497-1503. doi: 10.1098/rspb.2009.2111

Griffiths CL 2005. Coastal marine biodiversity in East Africa. Indian Journal of Marine Science 34: 35-41.

Guard M. 2009. Biology and fisheries status of octopus in the western Indian Ocean and the suitability for marine stewardship council certification. In: Guard M, editor: United Nations Environment Programme Institute for Security Studies (ISS).

Guard M, Mgaya YD 2002. The artisanal fishery for *Octopus cyanea* Gray in Tanzania. AMBIO A Journal of the Human Environment 31: 528-536. doi: 10.2307/4315304

Guerra Á, Roura Á, González ÁF, Pascual S, Cherel Y, Pérez-Losada M 2010. Morphological and genetic evidence that *Octopus vulgaris* Cuvier, 1797 inhabits Amsterdam and Saint Paul Islands (southern Indian Ocean). ICES Journal of Marine Science 67: 1401-1407. doi: 10.1093/icesjms/fsq040

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59: 307-321. doi: 10.1093/sysbio/syq010

Guindon S, Gascuel O 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology 52: 696-704.

Haasl RJ, Payseur BA 2013. Microsatellites as targets of natural selection. Molecular Biology and Evolution 30: 285-298. doi: 10.1093/molbev/mss247

Hall BG. 2011. Phylogenetic Trees made easy: a how-to manual. Sunderland, Massachussetts: Sinauer Associations, Inc.

Hall TA 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41: 95-98.

Han Z, Yanagimoto T, Zhang Y, Gao T 2012. Phylogeography study of *Ammodytes personatus* in Northwestern Pacific: Pleistocene isolation, temperature and current conducted secondary contact. PLoS ONE 7: e37425. doi: 10.1371/journal.pone.0037425

Hansen J, Sato M, Russell G, Kharecha P 2013. Climate sensitivity, sea level and atmospheric carbon dioxide. . Philosophical Transactions of the Royal Society of London. Series A 371: 2012-2024.

Harrison Hugo B, Williamson David H, Evans Richard D, Almany Glenn R, Thorrold Simon R, Russ Garry R, Feldheim Kevin A, van Herwerden L, Planes S, Srinivasan M, Berumen Michael L, Jones Geoffrey P 2012. Larval export from marine reserves and the recruitment benefit for fish and fisheries. Current Biology 22: 1023-1028. doi: 10.1016/j.cub.2012.04.008

Hauck M, Sweijd NA 1999. A case study of abalone poaching in South Africa and its impact on fisheries management. ICES Journal of Marine Science 56: 1024-1032. doi: 10.1006/jmsc.1999.0534

Hauser L, Carvalho GR 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries 9: 333-362. doi: 10.1111/j.1467-2979.2008.00299.x

Hauser L, Seeb JE 2008. Advances in molecular technology and their impact on fisheries genetics. Fish and Fisheries 9: 473-486. doi: 10.1111/j.1467-2979.2008.00306.x

Hebert PD, Ratnasingham S, de Waard JR 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society B: Biological Sciences 270 Suppl 1: S96-99. doi: 10.1098/rsbl.2003.0025

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM 2004. Identification of birds through DNA barcodes. PLoS Biology 2: e312. doi: 10.1371/journal.pbio.0020312

Heddrick PW. 2005. Genetics of Populations. USA: Jones and Bartlett.

Hewitt GM 1999. Post-glacial re-colonization of European biota. Biological Journal of the Linnean Society 68: 87-112. doi: 10.1111/j.1095-8312.1999.tb01160.x

Hilbish TJ 1996. Population genetics of marine species: the interaction of natural selection and historically differentiated populations. Journal of Experimental Marine Biology and Ecology 200: 67-83. doi: 10.1016/S0022-0981(96)02645-7

Ho SYW, Phillips MJ 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology 58: 367-380. doi: 10.1093/sysbio/syp035

Ho SYW, Shapiro B 2011. Skyline-plot methods for estimating demographic history from nucleotide sequences. Molecular Ecology Resources 11: 423-434. doi: 10.1111/j.1755-0998.2011.02988.x

Hoareau TB, Boissin E, Berrebi P 2012. Evolutionary history of a widespread Indo-Pacific goby: the role of Pleistocene sea-level changes on demographic contraction/expansion dynamics. Molecular Phylogenetics and Evolution 62: 566-572. doi: 10.1016/j.ympev.2011.10.004

Hoffman JI, Clarke A, Clark MS, Fretwell P, Peck LS 2012. Unexpected fine-scale population structure in a broadcast-spawning Antarctic marine mollusc. PLoS ONE 7: e32415. doi: 10.1371/journal.pone.0032415

Horne JB, van Herwerden L 2013. Long-term panmixia in a cosmopolitan Indo-Pacific coral reef fish and a nebulous genetic boundary with its broadly sympatric sister species. Journal of Evolutionary Biology 26: 783-799. doi: 10.1111/jeb.12092

Hsieh YW, Hwang PA, Pan HH, Chen JB, Hwang DF 2003. Identification of tetrodotoxin and fish species in an adulterated dried mullet roe implicated in food poisoning. Journal of Food Science 68: 142-146. doi: 10.1111/j.1365-2621.2003.tb14130.x

Hubisz MJ, Falush D, Stephens M, Pritchard JK 2009. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9: 1322-1332. doi: 10.1111/j.1755-0998.2009.02591.x

Hunt DJ, Parkes HC, Lumley IC 1997. Identification of the species of origin of raw and cooked meat products using oligonucleotide probes. Food Chemistry 60: 437-442. doi: 10.1016/S0308-8146(96)00364-0

Huntsman GR, Schaaf WE 1994. Simulation of the impact of fishing on reproduction of a protogynous grouper, the Graysby. North American Journal of Fisheries Management 14: 41-52. doi: 10.1577/1548-8675

Jackson AM, Semmens BX, Sadovy de Mitcheson Y, Nemeth RS, Heppell SA, Bush PG, Aguilar-Perera A, Claydon JAB, Calosso MC, Sealey KS, Schärer MT, Bernardi G 2014. Population structure and phylogeography in Nassau grouper *Epinephelus striatus*, a mass-aggregating marine fish. PLoS ONE 9: e97508. doi: 10.1371/journal.pone.0097508

Jackson JBC, Sheldon PR 1994. Constancy and change of life in the sea Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 344: 55-60. doi: 10.1098/rstb.1994.0051

Jacquet JL, Pauly D 2008. Trade secrets: Renaming and mislabeling of seafood. Marine Policy 32: 309-318. doi: http://dx.doi.org/10.1016/j.marpol.2007.06.007

Janko K, Lecointre G, DeVries A, Couloux A, Cruaud C, Marshall C 2007. Did glacial advances during the Pleistocene influence differently the demographic histories of benthic and pelagic Antarctic shelf fishes? Inferences from intraspecific mitochondrial and nuclear DNA sequence diversity. BMC Evolutionary Biology 7: 220. doi: 10.1186/1471-2148-7-220

Jensen JL, Bohonak AJ, Kelley ST 2005. Isolation by distance, web service. BMC Genetics 6: 13. doi: 10.1186/1471-2156-6-13

Johnson MS, Hebbert DR, Moran MJ 1993. Genetic analysis of populations of North-Weatern Australian fish species. Australian Journal of Marine and Freshwater Research 44: 673-685.

Jones WJ, Preston CM, Marin III R, Scholin CA, Vrijenhoek RC 2008. A robotic molecular method for in situ detection of marine invertebrate larvae. Molecular Ecology Resources 8: 540-550. doi: 10.1111/j.1471-8286.2007.02021.x

Jost L 2008. G(ST) and its relatives do not measure differentiation. Molecular Ecology 17: 4015-4026. doi: 10.1111/j.1365-294X.2008.03887.x

Juárez OE, Rosas C, Arena L 2010. Heterologous microsatellites reveal moderate genetic structure in the *Octopus maya* population. Fisheries Research 106: 209-213. doi: 10.1016/j.fishres.2010.08.011

Kashi Y, King DG 2006. Simple sequence repeats as advantageous mutators in evolution. Trends Genet 22: 253-259. doi: 10.1016/j.tig.2006.03.005

Kenchington EL, Harding GC, Jones MW, Prodohl PA 2009. Pleistocene glaciation events shape genetic structure across the range of the American lobster, *Homarus americanus*. Molecular Ecology 18: 1654-1667. doi: 10.1111/j.1365-294X.2009.04118.x

Koizumi I, Usio N, Kawai T, Azuma N, Masuda R 2012. Loss of genetic diversity means loss of geological information: The endangered Japanese crayfish exhibits remarkable historical footprints. PLoS ONE 7: e33986. doi: 10.1371/journal.pone.0033986

Kon T, Yoshino T, Mukai T, Nishida M 2007. DNA sequences identify numerous cryptic species of the vertebrate: a lesson from the gobioid fish *Schindleria*. Molecular Phylogenetics and Evolution 44: 53-62. doi: 10.1016/j.ympev.2006.12.007

Kulmiye AJ, Ntiba MJ, Kisia SM 2002. Some aspects of the reproductive biology of the thumbprint emperor *Lethrinus harak* Forsskal, 1775), in Kenyan coastal waters. Western Indian Ocean Journal of Marine Science 1: 135-144.

Kuo CH, Avise JC 2005. Phylogeographic breaks in low-dispersal species: the emergence of concordance across gene trees. Genetica 124: 179-186.

Larsson LC, Charlier J, Laikre L, Ryman N 2009. Statistical power for detecting genetic divergence - organelle versus nuclear markers. Conservation Genetics 10: 1255-1264. doi: 10.1007/s10592-008-9693-z

Lecomte F, Grant WS, Dodson JJ, Rodriguez-Sanchez R, Bowen BW 2004. Living with uncertainty: genetic imprints of climate shifts in East Pacific anchovy (*Engraulis mordax*) and sardine (*Sardinops sagax*). Molecular Ecology 13: 2169-2182. doi: 10.1111/j.1365-294X.2004.02229.x

Leis JM 2007. Behaviour as input for modelling dispersal of fish larvae: behaviour, biogeography, hydrodynamics, ontogeny, physiology and phylogeny meet hydrography. Marine Ecology Progress Series 347: 185-193. doi: 10.3354/meps06977

Lessios HA, Kessing BD, Robertson DR, Paulay G 1999. Phylogeography of the pantropical sea urchin *Eucidaris* in relation to land barriers and ocean currents. Evolution 53: 806-817. doi: 10.2307/2640720

Lindgren AR, Katugin ON, Amezquita E, Nishiguchi MK 2005. Evolutionary relationships among squids of the family Gonatidae (Mollusca : Cephalopoda) inferred from three mitochondrial loci. Molecular Phylogenetics and Evolution 36: 101-111. doi: 10.1016/j.ympev.2004.12.009

Lisiecki LE, Raymo ME 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18} O records. Paleoceanography 20: 1003. doi: 10.1029/2004PA001071

Lo Brutto S, Arculeo M, Stewart Grant W 2011. Climate change and population genetic structure of marine species. Chemistry and Ecology 27: 107-119. doi: 10.1080/02757540.2010.547486

Lo Galbo AM, Carpenter KE, Reed DL 2002. Evolution of trophic types in emperor fishes (*Lethrinus*, Lethrinidae, Percoidei) based on cytochrome B gene sequence variation. Journal of Molecular Evolution 54: 754-762. doi: 10.1007/s0023901-0076-z

Lü Z-M, Liu L-Q, Li H, Wu C-W, Zhang J-S 2013. Deep phylogeographic break among *Octopus variabilis* populations in China: evidence from mitochondrial and nuclear DNA analyses. Biochemical Systematics and Ecology 51: 224-231. doi: 10.1016/j.bse.2013.09.003

Lutjeharms JRE 2006. The ocean environment off southeastern Africa: a review. South African Journal of Science 102: 419-426.

Mann BQ 2000. South Africa linefish status reports. Oceanographic Research Institute Special Publication 7: 1-260.

Mantel N 1967. The detection of disease clustering and a generalised regression approach. Cancer Research 27: 200-220.

Marko PB 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the isthmus of panama. Molecular Biology and Evolution 19: 2005-2021.

Marko PB 2004. 'What's larvae got to do with it?' Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. Molecular Ecology 13: 597-611.

Marko PB, Hoffman JM, Emme SA, McGovern TM, Keever CC, Cox LN 2010. The 'Expansion-Contraction' model of Pleistocene biogeography: rocky shores suffer a sea change? Molecular Ecology 19: 146-169.

Marriott RJ, Jarvis ND, Adams DJ, Gallash AE, Norriss J, Newman SJ 2010. Maturation and sexual ontogeny in the spangled emperor *Lethrinus nebulosus*. Journal of Fish Biology 76: 1396-1414. doi: 10.1111/j.1095-8649.2010.02571.x

Martin AK, Goodlad SW, Salmon DA 1982. Sedimentary basin in-fill in the northernmost Natal Valley, hiatus development and Agulhas Current palaeo-oceanography. Journal of the Geological Society 139: 183-201. doi: 10.1144/gsjgs.139.2.0183

McKeown NJ, Shaw PW 2008. Polymorphic nuclear microsatellite loci for studies of brown crab, *Cancer pagurus* Molecular Ecology Resources 8: 653-655. doi: 10.1111/j.1471-8286.2007.02033.x

McMillan WO, Palumbi SR 1995. Concordant evolutionary patterns among Indo-West Pacific butterflyfishes. Proceedings of the Royal Society of London Series B: Biological Sciences 260: 229-236. doi: 10.1098/rspb.1995.0085

Meier R, Zhang G, Ali F 2008. The use of mean instead of smallest interspecific distances exaggerates the size of the “barcoding gap” and leads to misidentification. Systematic Biology 57: 809-813. doi: 10.1080/10635150802406343

Miller DD, Mariani S 2010. Smoke, mirrors, and mislabeled cod: poor transparency in the European seafood industry. Frontiers in Ecology and the Environment 8: 517-521. doi: 10.1890/090212

Mokhtar-Jamai K, Pascual M, Ledoux JB, Coma R, Feral JP, Garrabou J, Aurelle D 2011. From global to local genetic structuring in the red gorgonian *Paramuricea clavata*: the interplay between oceanographic conditions and limited larval dispersal. Molecular Ecology 20: 3291-3305. doi: 10.1111/j.1365-294X.2011.05176.x

Moran MD 2003. FORUM Arguments for rejecting the sequential Bonferroni in ecological studies. Nordic Society Oikos 100: 403-405.

Moretti VM, Turchini GM, Bellagamba F, Caprino F 2003. Traceability issues in fishery and aquaculture products. Veterinary Research Communications 27: 497-505. doi: 10.1023/B:VERC.0000014207.01900.5c

Moritz C 1994. Defining 'Evolutionarily Significant Units' for conservation. Trends in Ecology and Evolution 9: 373-375. doi: 10.1016/0169-5347(94)90057-4

Muths D, Gouws G, Mwale M, Tessier E, Bourjea J, Moran P 2012. Genetic connectivity of the reef fish *Lutjanus kasmira* at the scale of the western Indian Ocean. Canadian Journal of Fisheries and Aquatic Sciences 69: 842-853.

Muths D, Grewe P, Jean C, Bourjea J 2009. Genetic population structure of the swordfish (*Xiphias gladius*) in the southwest Indian Ocean: Sex-biased differentiation, congruency between markers and its incidence in a way of stock assessment. Fisheries Research 97: 263-269. doi: 10.1016/j.fishres.2009.03.004

Muths D, Le Couls S, Evano H, Grewe P, Bourjea J 2013. Multi-genetic marker approach and spatio-temporal analysis suggest there is a single panmictic population of Swordfish *Xiphias gladius* in the Indian Ocean. PLoS ONE 8: e63558. doi: 10.1371/journal.pone.0063558

Nakagawa S 2004. A farewell to Bonferroni: the problems of low statistical power and publication bias. Behavioral Ecology 15: 1044-1045. doi: 10.1093/beheco/arh107

Nakamura Y, Shibuno T, Suzuki N, Nakamori J, Kanashiro K, Watanabe Y 2010. Interspecific variations in age and size at settlement of 8 emperor fishes (Lethrinidae) at the southern Ryukyu Islands, Japan. Fisheries Science 76: 503-510. doi: 10.1007/s12562-010-0225-7

Nanami A, Yamada H 2009. Site fidelity, size, and spatial arrangement of daytime home range of thumbprint emperor *Lethrinus harak* (Lethrinidae). Fisheries Science 75: 1109-1116. doi: 10.1007/s12562-009-0136-7

Narum SR 2006. Beyond Bonferroni: Less conservative analyses for conservation genetics. Conservation Genetics 7: 783-787. doi: 10.1007/s10592-005-9056-y

Neethling M, Matthee CA, Bowie RCK, von der Heyden S 2008. Evidence for panmixia despite barriers to gene flow in the southern African endemic, *Caffrogobius caffer* (Teleostei: Gobiidae). BMC Evolutionary Biology 8: 325. doi: 10.1186/1471-2148-8-325

Nei M 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences 70: 3321-3323.

New AL, Alderson SG, Smeed DA, Stansfield KL 2007. On the circulation of water masses across the Mascarene Plateau in the South Indian Ocean. Deep Sea Research Part I: Oceanographic Research Papers 54: 42-74. doi: http://dx.doi.org/10.1016/j.dsr.2006.08.016

Nielsen EE, Hansen MM, Meldrup D 2006. Evidence of microsatellite hitch-hiking selection in Atlantic cod (*Gadus morhua L.*): implications for inferring population structure in nonmodel organisms. Molecular Ecology 15: 3219-3229. doi: 10.1111/j.1365-294X.2006.03025.x

Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Mittelholzer C, Taranger GL, Ogden R, Carvalho GR 2009. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). BMC Evolutionary Biology 9: 276. doi: 10.1186/1471-2148-9-276

Norman MD 1992. *Octopus cyanea* Gray, 1849 (Mollusca: Cephalopoda) in Australian waters: description, distribution and taxonomy. Bulletin of Marine Science 49: 20-38.

O'Malley KG, Camara MD, Banks Ma 2007. Candidate loci reveal genetic differentiation between temporally divergent migratory runs of Chinook salmon (*Oncorhynchus tshawytscha*). Molecular Ecology 16: 4930-4941. doi: 10.1111/j.1365-294X.2007.03565.x

Obura D 2012. The diversity and biogeography of Western Indian Ocean reef-building corals. PLoS ONE 7: e45013. doi: 10.1371/journal.pone.0045013

Ogden R 2008. Fisheries forensics: The use of DNA tools for improving compliance , traceability and enforcement in the fishing industry. 462-472.

Olivares-Banuelos NC, Enriquez-Paredes LM, Ladah LB, De la Rosa-Velez J 2008. Population structure of purple sea urchin *Strongylocentrotus purpuratus* along the Baja California peninsula. Fisheries Science 74: 804-812. doi: 10.1111/j.1444-2906.2008.01592.x

Palero F, Abelló P, Macpherson E, Gristina M, Pascual M 2008. Phylogeography of the European spiny lobster (*Palinurus elephas*): influence of current oceanographical features and historical processes. Molecular Phylogenetics and Evolution 48: 708-717. doi: 10.1016/j.ympev.2008.04.022

Palumbi SR. 1996. Nucleic acids II: the polymerase chain reaction. In: Hillis DM, Moritz C, Mable BK, editors. Molecular Systematics. Sunderland, Massachussetts: Sinauer & Associates Inc. p. 205-247.

Peery MZ, Kirby R, Reid BN, Stoelting R, Doucet-Bëer E, Robinson S, Vásquez-Carrillo C, Pauli JN, Palsbøll PJ 2012. Reliability of genetic bottleneck tests for detecting recent population declines. Molecular Ecology 21: 3403-3418. doi: 10.1111/j.1365-294X.2012.05635.x

Perez-Losada M, Guerra A, Carvalho GR, Sanjuan A, Shaw PW 2002. Extensive population subdivision of the cuttlefish *Sepia officinalis* (Mollusca : Cephalopoda) around the Iberian Peninsula indicated by microsatellite DNA variation. Heredity 89. doi: 10.1038/sj.hdy.6800160

Pérez-Losada M, Guerra A, Carvalho GR, Sanjuan A, Shaw PW 2002. Extensive population subdivision of the cuttlefish *Sepia officinalis* (Mollusca: Cephalopoda)

around the Iberian Peninsula indicated by microsatellite DNA variation. Heredity 89: 417-424. doi: 10.1038/sj.hdy.6800160

Perez-Losada M, Nolte MJ, Crandall KA, Shaw PW 2007. Testing hypotheses of population structuring in the Northeast Atlantic Ocean and Mediterranean Sea using the common cuttlefish *Sepia officinalis*. Molecular Ecology 16: 2667-2679. doi: 10.1111/j.1365-294X.2007.03333.x

Perrier C, Grandjean F, Gentil J, Cherbonnel C, Evanno G 2010. A species-specific microsatellite marker to discriminate European Atlantic salmon, brown trout, and their hybrids. Conservation Genetics Resources 3: 131-133. doi: 10.1007/s12686-010-9307-1

Perry AL, Low PJ, Ellis JR, Reynolds JD 2005. Climate change and distribution shifts in marine fishes. Science 308: 1912-1915. doi: 10.1126/science.1111322

Pillans B, Chappell J, Naish TR 1998. A review of the Milankovitch climatic beat: template for Plio–Pleistocene sea-level changes and sequence stratigraphy. Sedimentary Geology 122: 5-21. doi: 10.1016/S0037-0738(98)00095-5

Piry S, Luikart G, Cornuet JM 1999. Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. Journal of Heredity 90: 502-503. doi: 10.1093/jhered/90.4.502

Posada D 2008. jModelTest: phylogenetic model averaging. Molecular Biology and Evolution 25: 1253-1256. doi: 10.1093/molbev/msn083

Posada D, Buckley TR 2004. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53: 793-808. doi: 10.1080/10635150490522304

Posada D, Crandall KA 2001. Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proceedings of the National Academy of Sciences 98: 13757-13762. doi: 10.1073/pnas.241370698

Prell WL, Hutson WH, Williams DF, Allan WHB, Geitzenauer K, Molfino B 1980. Surface circulation of the Indian Ocean during the last glacial maximum, approximately 18,000 yr B.P. Quaternary Research 14: 309-336.

Preu B, Spieß V, Schwenk T, Schneider R 2011. Evidence for current-controlled sedimentation along the southern Mozambique continental margin since Early Miocene times. Geo-Marine Letters 31: 427-435. doi: 10.1007/s00367-011-0238-y

Pritchard JK. 2010. Documentation for structure software : Version 2.3. In.

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

Pritchard JK, Wen W, Il C 2003. Documentation for structure software: Version 2 .

Provan J, Bennett KD 2008. Phylogeographic insights into cryptic glacial refugia. Trends in Ecology and Evolution 23: 564-571. doi: 10.1016/j.tree.2008.06.010

Purcell JF, Cowen RK, Hughes CR, Williams DA 2009. Population structure in a common Caribbean coral-reef fish: implications for larval dispersal and early life-history traits. Journal of Fish Biology 74: 403-417. doi: 10.1111/j.1095-8649.2008.02078.x

Purcell JF, Cowen RK, Hughes CR, Williams DA 2006. Weak genetic structure indicates strong dispersal limits: a tale of two coral reef fish. Proceedings of the Royal Society B: Biological Sciences 273: 1483-1490. doi: 10.1098/rspb.2006.3470

Qasim SZ 1977. Biological productivity of the Indian Ocean. Indian Journal of Marine Science 6: 122-137.

Quartly GD, Srokosz MA 2004. Eddies in the southern Mozambique Channel. Deep-Sea Research Part II-Topical Studies in Oceanography 51: 69-83. doi: 10.1016/j.dsr2.2003.03.001

Raemaekers S, Hauck M, Bürgener M, Mackenzie A, Maharaj G, Plagányi ÉE, Britz PJ 2011. Review of the causes of the rise of the illegal South African abalone fishery and consequent closure of the rights-based fishery. Ocean & Coastal Management 54: 433-445. doi: 10.1016/j.ocecoaman.2011.02.001

Ragionieri L, Cannicci S, Schubart CD, Fratini S 2010. Gene flow and demographic history of the mangrove crab *Neosarmatium meinerti*: A case study from the western Indian Ocean. Estuarine, Coastal and Shelf Science 86: 179-188. doi: 10.1016/j.ecss.2009.11.002

Rambaut A. 2012. Figtree v. 1.4.0. http://tree.bio.ed.ac.uk/software/figtree/.

Rambaut A, Drummond AJ. 2009. Tracer v. 1.5. http://tree.bio.ed.ac.uk/software/tracer/

Rasmussen RS, Morrissey MT 2008. DNA-based methods for the commercial fish and seafood species. Comprehensive Reviews in Food Science and Food Safety 7: 280-295.

Raymond M, Rousset F 1995a. An exact test for population differentiation. Evolution 49: 1280-1283.

Raymond M, Rousset F 1995b. GENEPOP Version 1.2: population genetics software for exact tests and ecumenicism. Journal of Heredity 86: 248-249.

Read CI, Bellwood DR, van Herwerden L 2006. Ancient origins of Indo-Pacific coral reef fish biodiversity: a case study of the leopard wrasses (Labridae: Macropharyngodon). Molecular Phylogenetics and Evolution 38: 808-819. doi: 10.1016/j.ympev.2005.08.001

Reeb CA, Avise JC 1990. A genetic discontinuity in a continuously distributed species: mitochondrial DNA in the American oyster *Crassostrea virginica*. Genetics 124: 397-406.

Rhode C, Vervalle J, Bester-van der Merwe AE, Roodt-Wilding R 2013. Detection of molecular signatures of selection at microsatellite loci in the South African abalone (*Haliotis midae*) using a population genomic approach. Marine Genomics 10: 27-36. doi: 10.1016/j.margen.2013.03.001

Rice WR 1989. Analyzing tables of statistical tests. Evolution 43: 223-225.

Ridgway T, Sampayo EM 2005. Population genetic status of the Western Indian Ocean: What do we know? Western Indian Ocean Journal of Marine Science 4: 1-9.

Riginos C, Harrison R 2005. Cryptic vicariance in Gulf of California fishes parallels vicariant patterns found in Baja California mammals and reptiles. Evolution 59: 2678-2690. doi: 10.1554/05-257.1

Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding M, Wells F, Vynne C, Werner TB 2002. Marine biodiversity hotspots and conservation priorities for tropical reefs. Science 295: 1280-1284. doi: 10.1126/science.1067728

Rogers AR, Harpending H 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9: 552-569.

Rognon X, Guyomard R 2003. Large extent of mitochondrial DNA transfer from *Oreochromis aureus* to *O. niloticus* in West Africa. Molecular Ecology 12: 435-445.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539-542. doi: 10.1093/sysbio/sys029

Roper CFE, Sweeney MJ, Nauen CE. 1984. FAO Species Catalogue. Volume 3, Cephalopods of the World. In. Food and Agriculture Organisation of the United Nations. Rome: FAO.

Ryman N, Jorde PE 2001. Statistical power when testing for genetic differentiation. Molecular Ecology 10: 2361-2373.

Ryman N, Palm S 2006. POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. Molecular Ecology Notes 6: 600-602. doi: 10.1111/j.1471-8286.2006.01378.x

Saigusa K, Matsumasa M, Yashima Y, Takamiya M, Aoki Y 2009. Practical applications of molecular biological species identification of forensically important flies. Legal Medicine 11, Supplement 1: S344-S347. doi: 10.1016/j.legalmed.2009.01.026

Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular Cloning a laboratory manual: Cold Spring Harbor Laboratory Press New York.

Santos S, Hrbek T, Farias IP, Schneider H, Sampaio I 2006. Population genetic structuring of the king weakfish, *Macrodon ancylodon* (Sciaenidae), in Atlantic coastal waters of South America: deep genetic divergence without morphological change. Molecular Ecology 15: 4361-4373. doi: 10.1111/j.1365-294X.2006.03108.x

Sato T 1971. A revision of the Japanese sparoid fishes of the genus *Lethrinus*. Journal of the Faculty of Science University of Tokyo Section IV: Zoology 12: 117-144.

Schenekar T, Weiss S 2011. High rate of calculation errors in mismatch distribution analysis results in numerous false inferences of biological importance. Heredity 107: 511-512. doi: 10.1038/hdy.2011.48

Schneider S, Excoffier L 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics 152: 1079-1089.

Schott FA, McCreary JP 2001. The monsoon circulation of the Indian Ocean. Progress in Oceanography 51: 1-123.

Schott FA, Xie S-P, McCreary JP 2009. Indian Ocean circulation and climate variability. Reviews of Geophysics 47: RG1002. doi: 10.1029/2007RG000245

Schouten MW, de Ruijter WPM, van Leeuwen PJ, Ridderinkhof H 2003. Eddies and variability in the Mozambique Channel. Deep-Sea Research Part II-Topical Studies in Oceanography 50: 1987-2003. doi: 10.1016/S0967-0645(03)00042-0

Schuelke M 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18: 233-234. doi: 10.1038/72708

Selkoe KA, Gaines SD, Caselle JE, Warner RR 2006. Current shifts and kin aggregation explain genetic patchiness in fish recruits. Ecological Society of America 87: 3082-3094.

Selkoe KA, Toonen RJ 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9: 615-629. doi: 10.1111/j.1461-0248.2006.00889.x

Selkoe KA, Watson JR, White C, Horin TB, Iacchei M, Mitarai S, Siegel DA, Gaines SD, Toonen RJ 2010. Taking the chaos out of genetic patchiness: seascape genetics reveals ecological and oceanographic drivers of genetic patterns in three temperate reef species. Molecular Ecology 19: 3708-3726. doi: 10.1111/j.1365-294X.2010.04658.x

Shaw KL 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences 99: 16122-16127. doi: 10.1073/pnas.242585899

Shen KN, Jamandre BW, Hsu CC, Tzeng WN, Durand JD 2011. Plio-Pleistocene sea level and temperature fluctuations in the northwestern Pacific promoted speciation in the globally-distributed flathead mullet *Mugil cephalus*. BMC Evolutionary Biology 11: 83. doi: 10.1186/1471-2148-11-83

Siegert MJ. 2002. Ice sheets and Late Quaternary environmental changes. Chichester, UK: John Wiley & Sons, Ltd.

Silva IC, Mesquita N, Paula J 2010a. Genetic and morphological differentiation of the mangrove crab *Perisesarma guttatum* (Brachyura: Sesarmidae) along an East African latitudinal gradient. Biological Journal of the Linnean Society 99: 28-46. doi: 10.1111/j.1095-8312.2009.01338.x

Silva IC, Mesquita N, Paula J 2010b. Lack of population structure in the fiddler crab *Uca annulipes* along an East African latitudinal gradient: genetic and morphometric evidence. Marine Biology 157: 1113-1126. doi: 10.1007/s00227-010-1393-9

Slate J 2005. Quantitative trait locus mapping in natural populations: progress, caveats and future directions. Mol Ecol 14: 363-379. doi: 10.1111/j.1365-294X.2004.02378.x

Slatkin M 1994a. An exact test for neutrality based on the Ewens sampling distribution. Genetical Research 64: 71-74.

Slatkin M. 1994b. Gene flow and population structure. In. Ecological Genetics. Princeton, NJ: Princeton University Press. p. 3-17.

Slatkin M 1995. A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457-462.

Smith JLB 1959. Fishes of the family Lethrinidae from the Western Indian Ocean. Ichthyological Bulletin of Rhodes University 17: 285-295.

Smith JM, Haigh J 1974. The hitch-hiking effect of a favourable gene. Genetics Research 23: 23-35.

Smith MM, Heemstra PC. 1986. Smiths' sea fishes. Berlin Springer Berlin Heidelberg.

Smith WHF, Sandwell DT 1997. Global sea floor topography from satellite altimetry and ship depth soundings. Science 277: 1956-1962.

Sonzogni C, Bard E, Rostek F 1998. Tropical sea-surface temperatures during the last galcial period: A view based on alkenones in Indian Ocean sediments. Quaternary Science Reviews 17: 1185-1201. doi: 10.1016/S0277-3791(97)00099-1

Sota T 2002. Radiation and reticulation: extensive introgressive hybridization in the carabid beetles *Ohomopterus* inferred from mitochondrial gene genealogy. Population Ecology 44: 145-156. doi: 10.1007/s101440200018

Srivathsan A, Meier R 2012. On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. Cladistics 28: 190-194. doi: 10.1111/j.1096-0031.2011.00370.x

Strugnell JM, Lindgren AR 2007. A barcode of life database for the Cephalopoda? Considerations and concerns. Reviews in Fish Biology and Fisheries 17: 337-344. doi: 10.1007/s11160-007-9043-0

Strugnell JM, Rogers AD, Prodöhl PA, Collins MA, Allcock AL 2008. The thermohaline expressway: the Southern Ocean as a centre of origin for deep-sea octopuses. Cladistics 24: 853-860. doi: 10.1111/j.1096-0031.2008.00234.x

Strugnell JM, Watts PC, Smith PJ, Allcock AL 2012. Persistent genetic signatures of historic climatic events in an Antarctic octopus. Molecular Ecology 21: 2775-2787. doi: 10.1111/j.1365-294X.2012.05572.x

Sunnucks P 2000. Efficient genetic markers for population biology. Trends in Ecology and Evolution 15: 199-203. doi: 10.1016/S0169-5347(00)01825-5

Sweijd NA, Bowie RCK, Evans BS, Lopata AL 2000. Molecular genetics and the management and conservation of marine organisms. Hydrobiologia 420: 153-164. doi: 10.1023/A:1003978831621

Tajima F 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

Takumiya M, Kobayashi M, Tsuneki K, Furuya H 2005. Phylogenetic relationships among major species of Japanese coleoid cephalopods (Mollusca: Cephalopoda) using three mitochondrial DNA sequences. Zoological Science 22: 147-155. doi: 10.2108/zsj.22.147

Tamura K, Nei M 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10: 512-526.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30: 2725-2729. doi: 10.1093/molbev/mst197

Tang Q, Liu H, Mayden R, Xiong B 2006. Comparison of evolutionary rates in the mitochondrial DNA cytochrome b gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes). Molecular Phylogenetics and Evolution 39: 347-357. doi: 10.1016/j.ympev.2005.08.007

Tang W, Lshimatsu A, Fu C, Yin W, Li G, Chen H, Wu Q, Li B 2010. Cryptic species and historical biogeography of eel gobies (Gobioidei: Odontamblyopus) along the northwestern Pacific coast. Zoological Science 27: 8-13. doi: 10.2108/zsj.27.8

Teske PR, Oosthuizen A, Papadopoulos I, Barker NP 2007. Phylogeographic structure of *Octopus vulgaris* in South Africa revisited: identification of a second lineage near Durban harbour. Marine Biology 151: 2119-2122. doi: 10.1007/s00227-007-0644-x

Thomas L, Bell JJ 2013. Testing the consistency of connectivity patterns for a widely dispersing marine species. Heredity 111: 345-354. doi: 10.1038/hdy.2013.58

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG 1997. The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 25: 4876-4882. doi: 10.1093/nar/25.24.4876

Tolley KA, Grol MGG, Gopal K, Matthee CA 2005. Mitochondrial DNA panmixia in spiny lobster *Palinurus gilchristi* suggests a population expansion. Marine Ecology Progress Series 297: 225-231.

Tomita K, Yokobori S, Oshima T, Ueda T, Watanabe K 2002. The cephalopod *Loligo bleekeri* mitochondrial genome: multiplied noncoding regions and transposition of tRNA genes. Journal of Molecular Evolution 54: 486-500. doi: 10.1007/s00239-001-0039-4

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG 2012. Primer3—new capabilities and interfaces. Nucleic Acids Research 40: e115. doi: 10.1093/nar/gks596

Van de Putte A, Van Houdt JKJ, Maes GE, Janko K, Koubbi P, Rock J, Volckaert FAM 2009. Species identification in the trematomid family using nuclear genetic markers. Polar Biology 32: 1731-1741. doi: 10.1007/s00300-009-0672-8

van der Elst R, Everett B, Jiddawi N, Mwatha G, Afonso P, Boulle D 2005. Fish, fishers and fisheries of the Western Indian Ocean: their diversity and status. A preliminary assessment. Philosophical transactions. Series A, Mathematical, Physical, and Engineering Sciences 363: 263-284. doi: 10.1098/rsta.2004.1492

van Herwerden L, Aspden WJ, Newman SJ, Pegg GG, Briskey L, Sinclair W 2009. A comparison of the population genetics of *Lethrinus miniatus* and *Lutjanus sebae* from the east and west coasts of Australia: Evidence for panmixia and isolation. Fisheries Research 100: 148-155. doi: 10.1016/j.fishres.2009.07.003

van Heukelem WF 1976. Growth, bioenergetics and life-span of *Octopus cyanea* and *Octopus maya*. [[Hawaii]: University of Hawaii.

van Heukelem WFV 1973. Growth and life-span of *Octopus cyanea* (Mollusca: Cephalopoda). Journal of Zoology 169: 299-315. doi: 10.1111/j.1469-7998.1973.tb04559.x

Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538. doi: 10.1111/j.1471-8286.2004.00684.x

Vasemägi A, Nilsson J, Primmer CR 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar*). Molecular Biology and Evolution 22: 1067-1076. doi: 10.1093/molbev/msi093

Villanueva R 1995. Experimental rearing and growth of planktonic *Octopus vulgaris* from hatching to settlement. Canadian Journal of Fisheries and Aquatic Sciences 52: 2639-2650. doi: 10.1139/f95-853

Villanueva R, Norman MD 2008. Biology of the planktonic stages of benthic octopuses. Oceanography and Marine Biology: An Annual Review 46: 105-202.

Vincze T, Posfai J, Roberts RJ 2003. NEBcutter: a program to cleave DNA with restriction enzymes. Nucleic Acids Research 31: 3688-3691. doi: 10.1093/nar/gkg526

Visram S, Yang M-C, Pillay RM, Said S, Henriksson O, Grahn M, Chen CA 2010. Genetic connectivity and historical demography of the blue barred parrotfish (*Scarus ghobban*) in the Western Indian Ocean. Marine Biology 157: 1475-1487. doi: 10.1007/s00227-010-1422-8

von der Heyden S, Prochazka K, Bowie RCK 2008. Significant population structure and asymmetric gene flow patterns amidst expanding populations of *Clinus cottoides* (Perciformes, Clinidae): application of molecular data to marine conservation planning in South Africa. Molecular Ecology 17: 4812-4826. doi: 10.1111/j.1365-294X.2008.03959.x

Voris HK 2000. Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. Journal of Biogeography 27: 1153-1167. doi: 10.1046/j.1365-2699.2000.00489.x

Waples R 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity 89: 438-450. doi: 10.1093/jhered/89.5.438

Waples RS, Punt AE, Cope JM 2008. Integrating genetic data into management of marine resources: how can we do it better? Fish and Fisheries 9: 423-449. doi: 10.1111/j.1467-2979.2008.00303.x

Ward RD 2000. Genetics in fisheries management. Hydrobiologia 420: 191-201. doi: 10.1023/A:1003928327503

Ward RD, Costa FO, Holmes BH, Steinke D 2008. DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergence for most taxa, but *Zeus faber* and *Lepidopus caudatus* each probably constitute two species. Aquatic Biology 3: 71-78. doi: 10.3354/ab00068

Ward RD, Elliott NG, Grewe PM 1995. Allozyme and mitochondrial DNA separation of Pacific northern bluefin tuna, *Thunnus thynnus orientalis* (Temminck and Schlegel), from southern bluefin tuna, *Thunnus maccoyii* (Castelnau). Australian Journal of Marine and Freshwater Research 46: 921-930.

Ward RD, Holmes BH 2007. An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. Molecular Ecology Notes 7: 899-907. doi: 10.1111/j.1471-8286.2007.01886.x

Ward RD, Woodwark M, Skibinski DOF 1994. A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. Journal of Fish Biology 44: 213-232. doi: 10.1111/j.1095-8649.1994.tb01200.x

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN 2005. DNA barcoding Australia's fish species. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 360: 1847-1857. doi: 10.1098/rstb.2005.1716

Watanabe K, Kawase S, Mukai T, Kakioka R, Miyazaki J-I, Hosoya K 2010. Population divergence of *Biwia zezera* (Cyprinidae: Gobioninae) and the discovery of a cryptic species, based on mitochondrial and nuclear DNA sequence analyses. Zoological Science 27: 647-655. doi: 10.2108/zsj.27.647

Webb KE, Barnes DKA, Clark MS, Bowden DA 2006. DNA barcoding: A molecular tool to identify Antarctic marine larvae. Deep-Sea Research Part II: Topical Studies in Oceanography 53: 1053-1060. doi: 10.1016/j.dsr2.2006.02.013

Wegmann D, Leuenberger C, Neuenschwander S, Excoffier L 2010. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. BMC Bioinformatics 11: 116. doi: 10.1186/1471-2105-11-116

Weir BS, Cockerham CC 1984. Estimating F-Statistics for the analysis of population-structure. Evolution 38: 1358-1370. doi: 10.2307/2408641

Wells MJ, Wells J 1970. Observations on the feeding, growth rate and habits of newly settled *Octopus cyanea*. Journal of Zoology 161: 65-74. doi: 10.1111/j.1469-7998.1970.tb02170.x

Winkelmann I, Campos PF, Strugnell J, Cherel Y, Smith PJ, Kubodera T, Allcock L, Kampmann M-L, Schroeder H, Guerra A, Norman M, Finn J, Ingrao D, Clarke M, Gilbert MTP 2013. Mitochondrial genome diversity and population structure of the giant squid Architeuthis: genetics sheds new light on one of the most enigmatic marine species. Proceedings of the Royal Society B: Biological Sciences 280. doi: 10.1098/rspb.2013.0273

Winnepenninckx B, Backeljau T, De Wachter R 1993. Extraction of high-molecular-weight DNA from Mollusks. Trends in Genetics 9: 407.

Wright S 1949. The genetical structure of populations. Annals of Eugenics 15: 323-354. doi: 10.1111/j.1469-1809.1949.tb02451.x

Wyrtki K. 1973. Physical oceanography of the Indian Ocean. Berlin: Springer-Verlag.

Yarnell J 1969. Aspects of the behaviour of *Octopus cyanea*. Animal Behaviour 17: 747-754.

Young PC, Martin RB 1982. Evidence for protogynous hermaphroditism in some Lethrinid fishes. Journal of Fish Biology 21: 475-484.

Zardoya R, Castilho R, Grande C, Favre-Krey L, Caetano S, Marcato S, Krey G, Patarnello T 2004. Differential population structuring of two closely related fish species, the mackerel (*Scomber scombrus)* and the chub mackerel (*Scomber japonicus*), in the Mediterranean Sea. Molecular Ecology 13. doi: 10.1111/j.1365-294X.2004.02198.x

Zemlak TS, Ward RD, Connell AD, Holmes BH, Hebert PDN 2009. DNA barcoding reveals overlooked marine fishes. Molecular Ecology Resources 9: 237-242. doi: 10.1111/j.1755-0998.2009.02649.x

Zhang J-B, Hanner R 2011. DNA barcoding is a useful tool for the identification of marine fishes from Japan. Biochemical Systematics and Ecology 39: 31-42. doi: 10.1016/j.bse.2010.12.017