

**GENETIC STRUCTURE AND BIOGEOGRAPHY OF
THREE WRASSE SPECIES (LABRIDAE) WITHIN
THE WESTERN INDIAN OCEAN**

Submitted in fulfilment of the requirement for the degree of

MASTER OF SCIENCE

at RHODES UNIVERSITY

By

SISANDA MAYEKISO

2015

DECLARATION

I am aware of the University Policy on Plagiarism. This work contained herein is my original work. This thesis does not include other person's ideas, information and illustrations without specifically acknowledging their authorship.

Name: Sisanda Mayekiso

Student number: g11m7325

ABSTRACT

The biogeography of wrasses within the Western Indian Ocean (WIO) is poorly understood, with regard to origins and genetic differentiation or connectivity among the regions of the WIO. This region is a good model for studying the influence of physical complexities and biogeographic breaks in shaping patterns of differentiation in wrasses. Three reef-associated fish species, *Cheilio inermis*, *Thalassoma hebraicum* and *T. lunare*, were selected to examine the factors that have influenced patterns of differentiation across the WIO. Each species was sampled from various localities of the WIO, the Red Sea and Indo-West Pacific Islands. Sequence data were generated from two mitochondrial gene fragments (cytochrome *b* and ATPase 6) and one nuclear locus (the first intron of the ribosomal protein S7 gene). Genetic analyses were used to calculate genetic diversity indices within species, which were then compared among species. The relationships among haplotypes and alleles were constructed using median-joining networks. Where necessary, neighbour-joining trees (NJ) were constructed to examine relationships among haplotypes and alleles for the *Thalassoma* species. Population structure was analysed using AMOVA and pairwise Φ_{ST} to compare and calculate differentiation between the WIO localities. Mismatch distributions were used to examine population growth and decline or stability, and demographic parameters were used to calculate time of population expansion.

There was high haplotype ($h = 0.88$ to 0.98) and low nucleotide diversities ($\pi = 0.003$ to 0.008) among all species for mitochondrial markers. For S7 intron I, high allelic ($A = 0.95$ to 0.98) and low nucleotide diversities ($\pi = 0.002$ to 0.014) were observed for all species. The pairwise Φ_{ST} values revealed little to great ($\Phi_{ST} = -0.02$ to 0.67) genetic differentiation between localities, across all species for the three gene regions. The pairwise comparisons indicated the differentiation in *C. inermis* of Tanzania and Kenya from Mozambique and Nosy Be (Madagascar). The widespread *C. inermis* also revealed the differentiation of Kenya and Tanzania. For *C. inermis*, the AMOVAs of ATPase 6 and cytochrome *b* data indicated high differentiation among defined locality groups. The groups were defined according to geographic proximity. However, the AMOVA of the nuclear gene (S7 intron I) did not find variation among defined locality groups. *Cheilio inermis* revealed a sequence divergence of 0.4%. The divergence that was found in *C. inermis* was not enough to suggest a cryptic species within the WIO. Overall, the widespread and monotypic *C. inermis* revealed genetic differentiation within the WIO. *Thalassoma hebraicum* generally revealed little genetic

differentiation across the WIO. The AMOVAs of the three gene regions showed no variation among specimens of the defined locality groups. However, some differentiation was found between localities. The pairwise comparisons of *T. hebraicum* revealed the differentiation of Seychelles from the African mainland and Madagascar. Southern Africa was observed to be differentiated from Nosy Be and Zanzibar. The observed differentiation could be caused by oceanic barriers such as the South Equatorial Current (SEQC), East African Coastal Current (EACC), and the Comoros Gyres and eddies in the Mozambique Channel, and Agulhas Current. *Thalassoma lunare* revealed genetic isolation between the WIO and the Red Sea as well as within the WIO. The genetic isolation between the WIO and the Red Sea is probably due to the historical isolation by the Bab al Mandab and contemporary barriers such as the cold upwelling cells in Somalia. The differentiation of Maldives from the African mainland and Seychelles could be due to distance and the upwelling cells created by monsoon winds. Mismatch distributions suggested that *C. inermis* and *T. hebraicum* had undergone demographic expansion during the Pleistocene (92 678 to 40 219 years ago).

The results of the current study are similar to those from previous studies of WIO reef fish species, and the results of the present study could have potential implications for conservation and fisheries management. Single genetic markers and single species studies do not detect all barriers to dispersal in the WIO, thus they are insufficient to inform conservation management. Thus, the use of multispecies and genetic markers in the current study can be adopted by other studies of the marine taxa of the WIO.

TABLE OF CONTENTS

ABSTRACT	ii
LIST OF TABLE	vi
LIST OF FIGURES	viii
ACKNOWLEDGEMENTS.....	x
CHAPTER 1: GENERAL INTRODUCTION	1
Western Indian Ocean and WIO biogeography	3
Oceanographic characteristics	6
Molecular genetic markers.....	8
Rationale for the study	9
Aim and objectives of the study.....	9
CHAPTER 2: MATERIAL AND METHODS	11
Study area and sampling	11
Selection of molecular genetic markers	12
Mitochondrial DNA	13
Nuclear DNA (nDNA).....	14
DNA extraction and sequencing	15
Data analysis	16
Demographic history.....	18
CHAPTER 3: REGIONAL DIFFERENTIATION OF <i>CHEILO INERMIS</i> IN THE WESTERN INDIAN OCEAN.....	20
Introduction.....	20
Materials and methods	23
Sampling	23
DNA amplification.....	23
Genetic data analysis.....	24
Results.....	25
Genetic diversity analysis	25
Discussion	37
CHAPTER 4: GENETIC DIFFERENTIATION OF <i>THALASSOMA SPECIES</i> WITHIN THE WESTERN INDIAN OCEAN.....	43
Introduction.....	43
Materials and methods	47
Sampling	47
DNA amplification.....	47
Genetic data analysis.....	48
Results.....	49
Genetic diversity analysis	50
Relationships among haplotypes and alleles.....	54
Discussion	71
<i>Thalassoma hebraicum</i>	71
<i>Thalassoma lunare</i>	74
CHAPTER 5: CONCLUDING DISCUSSION	79

REFERENCES	82
APPENDIX ONE: HAPLOTYPE AND ALLELE FREQUENCIES	101

LIST OF TABLES

Table 2.1: Primers sets and primer references used for the amplification and sequencing of the two mitochondrial fragments and nuclear DNA fragment. The forward primer is indicated with an asterisk (*) while the reverse primer is denoted with double asterices (**).....	16
Table 3.1: Thermocycling conditions for the amplification of the respective mitochondrial (ATPase 6 and cytochrome <i>b</i>) and nuclear (S7 intron I) gene regions from <i>Cheilio inermis</i> ...	24
Table 3.2: Sampling localities, locality codes and sample sizes for <i>Cheilio inermis</i> collected from different Western Indian Ocean localities, the Red Sea and the Philippines	25
Table 3.3: Sample sizes (N) and diversity estimates at each of the localities from the Philippines (Pacific) and Western Indian Ocean and for the total sample of <i>Cheilio inermis</i> for the ATPase 6 and cytochrome <i>b</i> (A) and nuclear S7 intron I (B) datasets. Genetic diversity indices included the number of haplotypes (Nh) or alleles (NA), number of private haplotypes (NPH) or private alleles (NPA), haplotype diversity (<i>h</i>) or allelic diversity (<i>A</i>), nucleotide diversity (π), number of polymorphic sites (<i>S</i>) and average pairwise differences (<i>k</i>) among individuals. Standard deviations are presented for haplotype or allelic and nucleotide diversities	27
Table 3.4: Pairwise Φ_{ST} comparison values between localities of <i>Cheilio inermis</i> based on mitochondrial (A) ATPase 6 and nuclear S7 intron I data sets (ATPase 6: below diagonal; S7 intron I: above diagonal) and (B) the mitochondrial cytochrome <i>b</i> data set. Significant ($P < 0.05$) values are indicated in bold	32
Table 3.5: The results of the AMOVAs examining genetic variation at different hierarchical levels for <i>Cheilio inermis</i> sampled from several localities across the Western Indian Ocean for the ATPase 6, cytochrome <i>b</i> and nuclear S7 intron I data sets. The hierarchical arrangement of the data sets was based on proximity as described in the text. Significant ($P < 0.05$) values are indicated in bold.....	33
Table 3.6: Neutrality test indices, including Tajima's (1989) <i>D</i> and Fu's (1997) <i>F_s</i> , and the mismatch distribution statistics, including Harpending's (1994) raggedness index (<i>r</i>) and the Sum of Squared deviations (SSD), for the (A) mitochondrial ATPase 6 and cytochrome <i>b</i> , and (B) nuclear S7 intron I datasets. Significance is indicated by asterices: * $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$	34
Table 3.7: Demographic parameters of τ , θ_0 , and θ_1 from the mismatch distributions of the (A) ATPase 6 and (B) cytochrome <i>b</i> datasets for <i>Cheilio inermis</i> were used to estimate the time of expansion and the effective female population sizes at time zero (N_{ef0}) and the present day (N_{ef1}).....	36
Table 4.1: Polymerase Chain Reaction cycle profiles for the generation of the two mitochondrial (ATPase 6 and cytochrome <i>b</i>) and nuclear S7 intron I datasets from each of the <i>Thalassoma</i> species (<i>T. hebraicum</i> and <i>T. lunare</i>)	48

Table 4.2: Localities, locality codes and sample sizes of <i>Thalassoma hebraicum</i> and <i>T. lunare</i> collected from various Western Indian Ocean localities, the Red Sea and Indo-Pacific Islands	50
Table 4.3: Sample sizes (<i>N</i>) and genetic diversity indices derived from the mitochondrial ATPase 6 and cytochrome <i>b</i> (A) and nuclear S7 intron I (B) datasets for <i>Thalassoma hebraicum</i> for each locality and the overall sample. Indices included the number of haplotypes (<i>N_h</i>) or alleles (<i>N_A</i>), number of private haplotypes (<i>N_{PH}</i>) or private alleles (<i>N_{PA}</i>), haplotype diversity (<i>h</i>) or allelic diversity (<i>A</i>), nucleotide diversity (π), number of polymorphic sites (<i>S</i>) and average pairwise differences (<i>k</i>) among individuals. Standard deviations are presented for haplotype or allelic and nucleotide diversities.....	51
Table 4.4: Sample sizes (<i>N</i>) and genetic diversity indices for sampling localities in the Western Indian Ocean localities, the Red Sea and the Indo-West Pacific Islands and the overall sample of <i>Thalassoma lunare</i> for the two gene regions, ATPase 6 and S7 intron I. Indices of genetic diversities included the number of haplotypes (<i>N_h</i>) and alleles (<i>N_A</i>), the number of private haplotypes (<i>N_{PH}</i>) and alleles (<i>N_{PA}</i>), haplotype diversity (<i>h</i>) and allelic diversity (<i>A</i>), nucleotide diversity (π), the number of polymorphic sites (<i>S</i>) and the average pairwise differences among samples (<i>k</i>). Standard deviations are presented for haplotype, allelic and nucleotide diversities	53
Table 4.5: Pairwise Φ_{ST} values between WIO localities for (A) <i>Thalassoma hebraicum</i> and (B) between WIO localities and the Red Sea for <i>T. lunare</i> based the analyses of ATPase 6 sequence data. Significant estimates ($P < 0.05$) are indicated in bold.....	61
Table 4.6: Pairwise Φ_{ST} values between WIO localities for the <i>Thalassoma hebraicum</i> based on the mitochondrial cytochrome <i>b</i> data set	63
Table 4.7: Comparison of pairwise Φ_{ST} values between WIO localities for (A) <i>Thalassoma hebraicum</i> and (B) between WIO localities and the Red Sea for <i>T. lunare</i> based on nuclear S7 intron I data sets	65
Table 4.8: The results of AMOVAs examining the partitioning of differentiation at various hierarchical levels for <i>Thalassoma hebraicum</i> from across the WIO based on the ATPase 6, cytochrome <i>b</i> and nuclear S7 intron I data sets. The hierarchical arrangements are described in the text.....	66
Table 4.9: The results of neutrality tests, including Tajima's (1989) <i>D</i> and Fu's (1997) <i>F_s</i> , and the statistics from the mismatch distribution, including Harpending's (1994) raggedness (<i>r</i>) index and the Sum of Squared Deviations (SSD), for the (A) mitochondrial ATPase 6 and cytochrome <i>b</i> , and (B) nuclear S7 intron I data sets. Significant estimates are indicated: ** $P < 0.01$, *** $P < 0.001$	68
Table 4.10: Demographic parameters τ , θ_0 , and θ_1 from the mismatch distributions of the mitochondrial (A) ATPase 6 and (B) cytochrome <i>b</i> gene regions for <i>T. hebraicum</i> used to calculate time of expansion and effective population size for time zero (<i>N_{ef0}</i>) and the present day (<i>N_{ef1}</i>)	70

LIST OF FIGURES

- Figure 1.1:** The Western Indian Ocean region, modified from Santini and Winterbottom (2002), includes the Somali Basin (2), Natal Basin (3), Arabian Basin (4), Chagos Archipelago (include Maldives and Laccadives) (5), as well as the various islands of the Mascarene Plateau (6). The dashed circle indicates the Plateau. The solid lines indicate a potential physical barrier separating the different basins 4
- Figure 1.2:** Major oceanographic characteristics of the Western Indian Ocean, adapted from Lutjeharms and Bornman (2010) 8
- Figure 2.1:** Sampling localities across the Western Indian Ocean for the three wrasse species, and the sampling locality of *Cheilio inermis* in the Philippines. The numbers represent the Western Indian Ocean and West Pacific (Philippines) biogeographic regions corresponding to Santini and Winterbottom (2002) (see Figure 1.1). The Philippines Sea (11), Perolus Island and French Polynesia (16) are indicated corresponding to Santini and Winterbottom (2002). Solid lines and broken lines indicate potential physical barriers (Santini & Winterbottom, 2002) 12
- Figure 2.2:** The mitochondrial gene order of fish, showing the positions (shaded) of the ATPase synthase 6 (ATP 6) and cytochrome *b* (cyt *b*) genes. Figure taken from Meyer (1993) 14
- Figure 3.1:** *Cheilio inermis*, cigar wrasse. Photographed specimen: SAIAB 197031, 235 mm standard length, Shimoni landing site, Kenya (photo: SAIAB).....20
- Figure 3.2:** Distribution of *Cheilio inermis* (red) in the Indo-Pacific, including the Western Indian Ocean (modified from <http://www.fishbase.org>) 21
- Figure 3.3:** Median-joining networks of the ATPase 6 (A) and cytochrome *b* (B) haplotypes for *Cheilio inermis* from the Pacific (Philippines) and nine localities of the Western Indian Ocean. The circles represent unique haplotypes with the size reflecting frequency of haplotype (the smallest circle indicates a frequency of one). The colours are the geographical origin of the haplotypes. Connecting lines represent one mutational step unless indicated otherwise. Missing haplotypes (mv) are indicated in small white circles. Loops represent alternative hypotheses.....29
- Figure 3.4:** Median-joining network of genealogical relationships among nuclear S7 intron I alleles found in *Cheilio inermis* from six localities across Western Indian Ocean. Pie charts are coloured according to the geographical origin of the allele. The size of nodes is proportional to number of individuals observed with a given allele. Connecting lines represent one mutational step unless indicated otherwise. Missing alleles are represented by small white circles. Loops represent alternative hypotheses 30
- Figure 3.5:** Mismatch distribution curves of pairwise differences among individuals for the (A) ATPase 6, (B) cytochrome *b* and (C) nuclear S7 intron I gene regions for the overall population of *Cheilio inermis*. The observed curve of the number of pairwise differences among individuals is indicated with a black line. The expected curve under a model of population growth is indicated with a grey line 35

Figure 4.1: Distribution of *Thalassoma hebraicum* in the Western Indian Ocean (modified from <http://www.fishbase.org>). Inset image: adult/terminal male phase *Thalassoma hebraicum*, the gold bar wrasse. Photographed specimen: SAIAB 197032, 133.8 mm standard length, Shimoni landing site, Kenya (photo: SAIAB) 44

Figure 4.2: Adult (terminal male) *Thalassoma lunare*, the crescent-tail wrasse. Photographed specimen: SAIAB 87107, 160 mm standard length, Zanzibar, Tanzania (photo: SAIAB) ..44

Figure 4.3: Distribution of *Thalassoma lunare* in the Indo-Pacific, including the Western Indian Ocean (modified from <http://www.fishbase.org>). 45

Figure 4.4: Neighbour-joining tree derived from ATPase 6 sequence data for the two *Thalassoma* species (*T. hebraicum* and *T. lunare*) depicting relationships among the WIO and the Red Sea haplotypes. Nodal support (bootstrap percentage %) is shown on the branches. The scale bar indicates the estimated evolutionary distance..... 55

Figure 4.5: Neighbour-joining tree derived from S7 intron I sequence data for the two *Thalassoma* species (*T. hebraicum* and *T. lunare*) depicting relationships among the WIO, Red Sea and Pacific alleles. Nodal support (bootstrap percentage %) is shown on the branches. The scale bar indicates the estimated evolutionary distance..... 56

Figure 4.6: Median-joining network depicting genealogical relationships of the mitochondrial cytochrome *b* haplotypes for the two *Thalassoma* species (*Thalassoma hebraicum* and *Thalassoma lunare*). Pie charts are coloured according to the species identification of the individuals.....57

Figure 4.7: Median-joining networks of genealogical relationships among the ATPase 6 haplotypes of (A) *Thalassoma hebraicum* and (B) *T. lunare* from several Western Indian Ocean localities and the Red Sea. The size of the circles shows the frequency of each haplotype. The pie charts are shaded with a colour that represents the geographic origin of the haplotype. Connecting lines between the circles represent one mutational step unless indicated otherwise. Missing haplotypes are indicated by small circles 59

Figure 4.8: Median-joining network depicting genealogical relationships of the mitochondrial cytochrome *b* haplotypes for the *Thalassoma hebraicum*. The size of the node corresponds to the frequency of that haplotype and the colours represent the occurrence of that haplotype at different WIO localities. The numbers on the branches indicate mutational differences where greater than one 62

Figure 4.9: Median-joining networks of genealogical relationships among S7 intron I alleles of (A) *Thalassoma hebraicum* and (B) *Thalassoma lunare* from WIO localities, the Red Sea and Pacific. Connecting lines between nodes represent one mutational step unless indicated otherwise. The size of nodes is proportional to number of individuals observed with a given haplotype. Pie charts are coloured according to the geographical origins of the allele. Missing haplotypes (mv) are shown in small white circles 64

Figure 4.10: Mismatch distributions of pairwise differences among (A) ATPase 6 and (B) cytochrome *b* and (C) S7 intron I sequences of all *Thalassoma hebraicum* individuals. The observed curve is indicated with a black line. The expected model under a population growth model is illustrated with a grey line 69

ACKNOWLEDGEMENTS

I would first like to acknowledge my supervisors, Dr Gavin Gouws and Dr Monica Mwale, for their energy and time that they have dedicated in this study. I am grateful for their encouragement and motivation that kept me on my toes on this study. This study would not have been completed successfully without their patience and constructive criticism. Special thanks must also be extended to the Marine Group meetings for their invaluable support, paper discussions and comments on presentations.

This study could not have been possible without the financial support from African Coelacanth Ecosystem Programme (ACEP). Without this funding, field trips and laboratory work at the South African Institute for Aquatic Biodiversity (SAIAB) would not have been possible. I would like to thank Kenya Marine and Fisheries Research Institute (KMFRI) for assisting in the fieldtrip in Kenya and monitoring at the Shimoni landing site. Special thanks to Ofer Gon for his support and assisting in sampling in Kenya and Durban. I appreciate the help of other institutes and individuals for providing me with samples from the WIO and Philippines. Special thanks to Monica Mwale, Murray Duncan, Christine Coppinger and Devin Isemonger for collecting samples for me in Mozambique. For sampling along the South African coast, I would like to extend my sincere thanks to Nosiphiwo Springbok, Khanyisile Ngomane, Moqebelo Morallana and my supervisors for their assistance. Special thanks to Poogendri Reddy for my genetics training and extra help in the laboratory.

Special thanks to my friends at SAIAB: Nosiphiwo Springbok, Pholoshi Abram Maake, Moqebelo Morallana, Dr Tshoanelo Miya, Mpho Ramoejane and Amanda Gura. They, together with my other close friends at SAIAB, too many to be mentioned by name, that were supporting and encouraging me. Last but not least, biggest thanks to my parents and for their love and sacrifices they have made to provide me with the best education. I would love to thank them for their support in my studies even if they were not sure what I was doing. I love you all!

CHAPTER ONE

GENERAL INTRODUCTION

The Western Indian Ocean (WIO) forms part of the greater marine assemblage of the Indo-Pacific, and is generally not regarded as a separate zoogeographic unit (Van der Elst *et al.*, 2005). The WIO is characterised by high diversity and substantial endemism among marine fishes with at least 2200 fish species, comprising about 15 % of the total marine fish fauna in the world (Richmond, 2000; Van der Elst *et al.*, 2005; Costello *et al.*, 2010). The WIO consists of the Indian Ocean, the Arabian Sea and two evaporative basins, the Red Sea and the Arabian Gulf (Costello *et al.*, 2010). The Red Sea and Arabian Gulf are known as areas of endemism with isolated openings to the other regions of the WIO (Van der Elst *et al.*, 2005; Briggs & Bowen, 2013). The South Western Indian Ocean (SWIO) is also recognised as an area of endemism (Hoareau *et al.*, 2013). The high levels of endemism found off South Africa can be largely attributed to relative geographic isolation and to the temperate climate associated with large currents which preclude incursion of species from other regions (Beckley *et al.*, 2002; Van der Elst *et al.*, 2005). Although the taxonomy and knowledge of the diversity of the WIO fauna are advancing, there is a lack of knowledge on biogeography of the WIO. The origins and relationships of the WIO fauna, and the processes that have led to establishment of regional faunas within the WIO remain unresolved.

Various researchers have reported that physical factors of the environment, such as oceanographic features, and the life history characteristics of species may influence genetic structure and patterns of genetic connectivity in marine fishes and invertebrates (Palumbi, 1994; Neethling *et al.*, 2008; Ragionieri *et al.*, 2010; Knutsen *et al.*, 2013). The genetic structure among populations of reef fish species has also been observed to correspond to geographic barriers (Patarnello *et al.*, 2007; Bernardi *et al.*, 2004). However, there is no consensus view on the role of the factors mentioned above in shaping genetic structure of marine organisms. Geographic barriers and oceanographic features of the WIO, as well as life history characteristics of species, should be taken into consideration when elucidating the biogeographic patterns of the WIO.

Bellwood and Wainwright (2002) stated that the biogeographic patterns of coral reef fishes, such as wrasses, observed today reflect a long and complex history of vicariant and dispersal

events. Genetic studies in the Indo-Pacific and Atlantic Oceans have successfully used reef fishes, such as wrasses (family Labridae), to understand vicariant and dispersal biogeography (Bernardi *et al.*, 2004; Beldade *et al.*, 2009). The aforementioned genetic studies of wrasses have used sequence data from one nuclear and two to three mitochondrial DNA (mtDNA) gene regions. Bernardi *et al.* (2004) studied the biogeography of the genus *Thalassoma* in the Indo-Pacific region. This study revealed vicariance, which might have occurred approximately 8-13 million years (Ma) ago, showing that the Tethyan Event and the closing of the Isthmus of Panama were the major historical factors shaping the evolution of this genus. The genus *Clepticus* was examined to determine whether speciation resulted from dispersal or vicariant events (Beldade *et al.*, 2009). Genetic isolation of the populations of the Creole wrasse was not influenced by the two recognised barriers in the Atlantic Ocean, the Amazon-Orinoco and Mid-Atlantic Ridge (Beldade *et al.*, 2009). Life history features were suggested to maintain gene flow in the genus *Clepticus*; the coastal pelagic habitat and long larval duration facilitated dispersal across these biogeographical barriers (Beldade *et al.*, 2009).

The target fish species of the current study are wrasses from the family Labridae. The Labridae (wrasses) is the fifth largest fish family, the second largest marine fish family and one of the most ecologically- and morphologically-diverse marine faunas (Westneat & Alfaro, 2005). The family includes hogfish, tamarins, scarines and odacines (Heemstra & Heemstra, 2004) and comprises about 600 species in 82 genera (Westneat & Alfaro, 2005). These are found predominantly in tropical and subtropical seas in coastal waters, while fewer species occur in temperate zones (Heemstra & Heemstra, 2004; Kuitert, 2010). Wrasses are among the most abundant and conspicuous fishes on coral reefs, occurring in a wide diversity of colours, and have vastly different morphologies (Thresher, 1984). Van der Elst (1990) noted that these species are not always easy to identify because of the frequent changes in colour and patterns that occur between sexes and different life phases. Wrasses are reported to be protogynous hermaphrodites, with the fish first maturing as females (initial phase), undergoing sex change and becoming terminal males (Heemstra & Heemstra, 2004). Breeding has been reported to occur year round during the daytime (Thresher, 1984).

Three labrid fishes, goldbar wrasse *Thalassoma hebraicum* (Lacepède, 1801), crescent-tail wrasse *Thalassoma lunare* (Linnaeus, 1758) and cigar wrasse *Cheilodactylus inermis* (Forsskål, 1775), were selected in the current study to examine patterns of genetic differentiation and connectivity within the WIO. These wrasses are good models for biogeographic studies

because they are reef-associated fishes that have discrete distributions on patchy reefs across the WIO (Shima, 2001; Bernardi *et al.*, 2004). The selected species are discussed in detail in individual chapters (see Chapters 3 and 4).

Western Indian Ocean biogeography

Geographically, the WIO includes a combination of the continental mainland, oceanic and continental islands, and a complex system of ocean currents (Arthurton, 2003; Van der Elst *et al.*, 2005; Lutjeharms, 2007). The WIO proper (Figure 1.1) extends from the African tropical waters of Somalia in the north to the warm-temperate region of South Africa in the south (Kairu & Nyandwi, 2000; Beckley *et al.*, 2002; Gullström *et al.*, 2002), and stretches to the central Indian Ocean up to the tip of India. This encompasses the tropical coastal areas of East Africa (Somalia, Kenya, Tanzania and Mozambique) and the subtropical and temperate coasts of South Africa. The island state of Madagascar is surrounded by the various islands of the Mascarene Plateau (Mauritius, La Réunion, Seychelles and Rodrigues) and the Comoros islands (Klausewitz, 1972; Gullström *et al.*, 2002; New *et al.*, 2005). The Chagos Archipelago and Maldives are found in the centre of the Laccadive-Ridge in the tropical Indian Ocean (Sheppard, 2000). The peripheral areas of the WIO consist of the Red Sea, the Arabian Peninsula/Persian Gulf, Gulf of Oman and Gulf of Aden (Klausewitz, 1972).

The WIO has a diverse marine fauna that may have originated from the Indo-West Pacific province (Briggs 1999, Briggs & Bowen, 2012). Although species inventories are patchy or fragmented (Turpie *et al.*, 2000), this diversity is less than that of the presumed centre of origin or radiation, the Indo-Malayan region (Santini & Winterbottom, 2002). The distribution of several fish species of the Indo-West Pacific, including those of the WIO, are also believed to be maintained by dispersal from this centre of origin into less diverse areas (Briggs, 1992, 2005; Bellwood & Wainwright, 2002; Heads, 2005). This evidence is supported by the pattern of dispersal tracks that radiate outward from the centre of origin, and a decrease in diversity from this centre (Briggs, 2005), as well as the existence of the older taxa to the periphery of this region (Briggs, 1999).

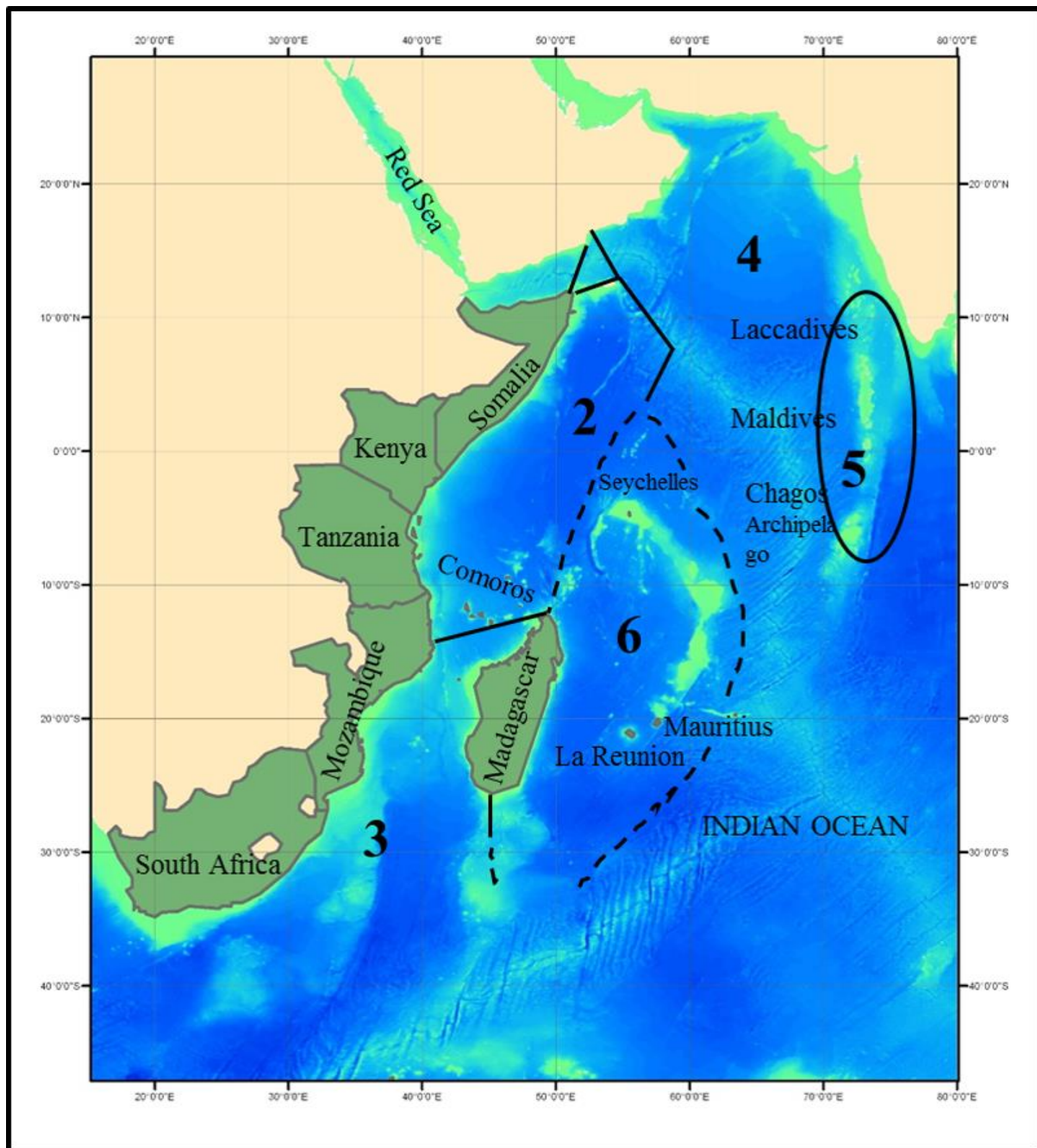


Figure 1.1: The Western Indian Ocean region, modified from Santini and Winterbottom (2002), includes the Somali Basin (2), Natal Basin (3), Arabian Basin (4), Chagos Archipelago (include Maldives and Laccadives) (5), as well as the various islands of the Mascarene Plateau (6). The dashed circle indicates the Plateau. The solid lines indicate a potential physical barrier separating the different basins.

The endemism and the origin of WIO diversity may be attributed to the formation of barriers and subsequent vicariance (Bellwood & Wainwright, 2002; Heads, 2005). Santini and Winterbottom (2002) demonstrated the separation of the coral faunas of six adjacent regions

(see Figure 1.1), the Red Sea, Arabian Basin, Somali Basin, Natal Basin, Mascarene Plateau and Chagos Ridge, through a series of vicariant events, resulting from a complex tectonic history. The Red Sea and Arabian Basin do not form part of the WIO, according to Santini and Winterbottom (2002).

Cold seasonal upwellings off Somalia and southern Oman were highlighted as barriers that isolated the Somali Basin and Arabian Sea from the rest of the Indo-West Pacific (see Figure 1.1). This separation may have occurred before or during the Miocene period (Santini & Winterbottom, 2002). The Mozambique Current was affected by land bridges, which could also have isolated the Natal Basin from Madagascar and its plateau. The Mascarene Plateau was separated from the African plate, and confined to the north-west by the Somali Basin and to the south-west by the Madagascar Plateau, while the Chagos Ridge was formed by series of eruptions as the Indian Plate moved to the north over a hotspot (Santini & Winterbottom, 2002). Santini and Winterbottom (2002) suggested that the central Indian Ridge separated the Mascarene Plateau and Chagos Ridge.

Various regions have been proposed for the WIO using different characteristics. Longhurst (1998), using hydrodynamics and water colour data, divided the WIO into four hydrogeographic areas: the Eastern African Coastal Province (EAFR), Northwest Arabian Upwelling Province (ARAB), Indian Monsoon Gyre Province (MONS) and Indian South Subtropical Gyre Province (ISSG). The EAFR consists of the coastal regions of the Indian Ocean from northern Kenya southwards to South Africa, including the Mozambique Channel and the entire west coast of Madagascar. The ARAB province contains the coastal areas of the Arabian Sea and Somalia, while MONS includes the Comoros and Seychelles (Longhurst, 1998). MONS is suggested to be similar to the Northwest Arabian Upwelling Province (ARAB) in that it is influenced by monsoon winds, but seasonality is stronger in the latter (Longhurst, 1998). The ISSG is a large oceanographic province in the south (Longhurst, 1998) and contains the Mascarene Islands (La Réunion, Mauritius and Rodrigues).

Briggs and Bowen (2012) recently defined the Tropical Indo-West Pacific region based on endemism, fish distributions and taxonomic partitions. The Tropical Indo-West Pacific region includes the provinces of the WIO, Eastern Indian Ocean, Red Sea (including the Gulf of Aden), Indo-Polynesia, Hawaii, Marquesas and Easter Island (Briggs & Bowen, 2012). The Western Indian Ocean province includes the east African mainland, Madagascar, Mascarene

Plateau, Seychelles and the Comoros (Briggs & Bowen, 2013), and comprises of about 1000 fish species and 142 (14.2 %) endemics. The Eastern Indian Ocean includes the Laccadives, Maldives and Chagos (Briggs & Bowen, 2012). According to Briggs and Bowen (2012), the Red Sea and Gulf of Aden do not form part of WIO province. The province of the Red Sea includes the Gulf of Aden because the fishes of the Red Sea extend into the Gulf of Aden (Briggs & Bowen, 2012), and has 14% endemism in fishes. The high endemism in the Red Sea could have been caused by the formation of land bridges during Pleistocene glaciations, which restricted the entrance to the Red Sea from the WIO (Sirocko, 2003; Van der Elst *et al.*, 2005).

Oceanographic characteristics

The central part of the Indian Ocean is dominated by the South Equatorial Current (SEQC), which flows continuously from east to the west throughout the year, drawing water from the Indo-Pacific and Indonesian region (Bock, 1978) (Figure 1.2). In the WIO, the SEQC flows and splits as it reaches the Mascarene Ridge and Madagascar, and again when it reaches the African mainland (Swallows *et al.*, 1988; de Ruijter *et al.*, 2004). When the SEQC reaches Madagascar, it branches northwards as the Somali Current (Lutjeharms, 2006). The Somali Current is strongly influenced by monsoon winds, which change direction in different seasons (Lutjeharms, 2006; Magori, 2008). In summer, the north-west monsoon wind causes the Somali Current to flow northward. In winter, the south-west monsoon wind causes the Somali Current to reverse its direction and flows southwards off the coast of Somalia, forming the Equatorial Counter Current (Sheppard & Wells, 1988; Kairu & Nyandwi, 2000; Magori, 2008). The SEQC also flows southwards into the East Madagascar Current and forms anticyclonic gyres in the Comoros Basin (Lutjeharms, 2006). This current then bifurcates again when it approaches the African mainland (de Ruijter *et al.*, 2004), where it continues to flow southwards through the Mozambique Channel as the Mozambique Current (de Ruijter *et al.*, 2004; Lutjeharms, 2006; Magori, 2008). The Mozambique Channel is fed by series of eddies (Quartly & Srokosz, 2004), which flow southwards to form the Agulhas Current along the coast of South Africa (Lutjeharms, 2006, 2007). The Agulhas Current retroflects in the Cape region of South Africa to complete the subtropical gyre (Schouten *et al.*, 2002; New *et al.*, 2005).

Oceanographic processes are recognised as possible factors responsible for shaping regional genetic differentiation or promoting connectivity within the WIO. For example, the South Equatorial Current and Equatorial Counter Current were found to influence the genetic structuring of a reef fish parrotfish *Scarus ghobban* (see Visram *et al.*, 2010), whereas the

oceanographic features of WIO aided dispersal in a reef fish, the Dory snapper *Lutjanus fulviflamma* (Dorenbosch *et al.*, 2006).

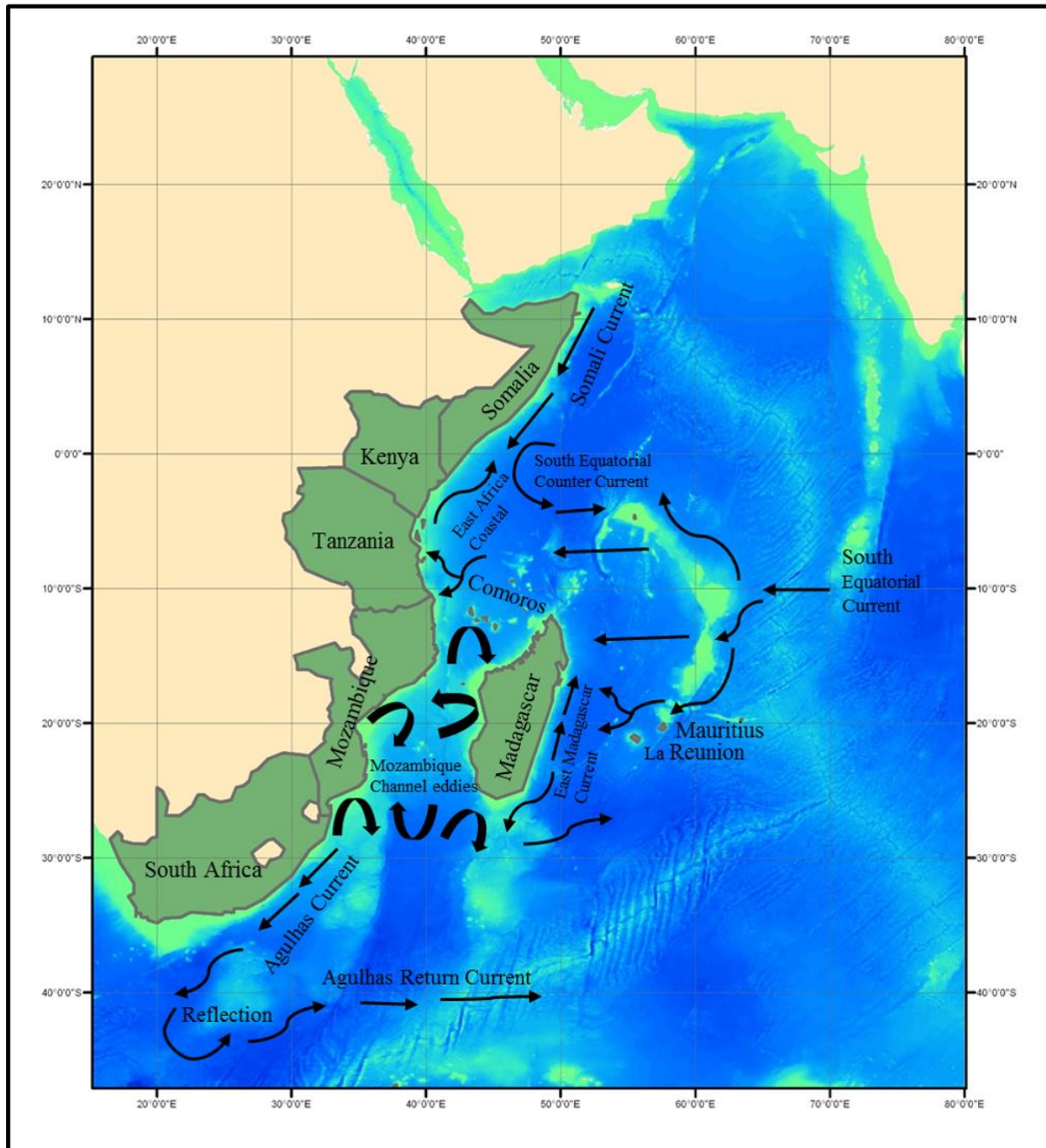


Fig. 1.2: Major oceanographic characteristics of the Western Indian Ocean, adapted from Lutjeharms and Bornman (2010).

Molecular genetic markers

Marine larvae are very difficult to monitor directly, and genetic markers can be used to infer the dispersal of larvae and interactions between populations (Visram *et al.*, 2010), as larvae are more dispersive than sedentary adults (Leis, 2002). Connectivity in reef fishes can be assessed using indirect methods, such as the analysis of variation in genetic markers (Visram *et al.*,

2010). Molecular genetic markers are recognised as powerful tools and have been used successfully for analysing the genetic variation of marine species (Chauhan & Rajiv, 2010). Various studies in the WIO have used genetic markers to investigate genetic differentiation in reef fishes and crustaceans (Fratini & Vannini, 2002; Froukh & Kochzius, 2007; Ragionieri *et al.*, 2010), and to detect the various processes or features that have influenced genetic structure (Gaither *et al.*, 2010; Visram *et al.*, 2010). The selection of particular markers is motivated in detail in Chapter 2.

Rationale for the study

The WIO represents a good model to study biogeography because of the physical complexities of the environment, which might influence the distribution of genetic variation of the three wrasse species. There are few comparable studies of genetic variation in fishes (Chow *et al.*, 1997; Froukh & Kochzius, 2007; Visram *et al.*, 2010; Muths *et al.*, 2012) and invertebrates (Fratini & Vannini, 2002; Ragionieri *et al.*, 2010) in the WIO. The available scientific literature is limited and this makes it difficult to draw conclusions on the biogeography of this region and to compare WIO patterns with findings from other regions in the world, especially the Indo-Pacific and Caribbean Islands (e.g. Bernardi *et al.*, 2001; De Grave, 2001; Williams *et al.*, 2002; Rocha *et al.*, 2008; Stuart *et al.*, 2008; Gaither *et al.*, 2010).

Information on the genetic variation of marine species is invaluable because it has the potential to reflect the historical and contemporary interactions among a complex set of ecological, demographic and life history characteristics, as well as oceanographic and biogeographic processes (Hedgecock, 1986; Benzie, 1999; Bohonak, 1999). A molecular genetic approach will be used in the current study to determine patterns of differentiation or connectivity of the three selected wrasses and to identify the features or processes that are responsible for the observed genetic structure. There is a need to close this large gap of knowledge and provide a clear understanding of the biodiversity, and the origins and interactions of the regional fish faunas in the WIO. This information is also invaluable as baseline data for conservation management.

Aim and objectives of the study

The aim of the study is to examine regional differentiation within the three target wrasse species across the WIO and to identify the processes that might have contributed to the observed patterns. It is expected that the genetic structure will reflect contemporary or historical patterns

of connectivity and colonisation, or isolation of the WIO regions. The observed patterns will shed light on whether the regional faunas originated from the establishment of vicariant barriers or from dispersal events. This has potential for elucidating the processes that are responsible for shaping the patterns of biogeography in this region.

Specific objectives of the study are:

1. To examine genetic diversity in and genetic differentiation among different geographic regions of the WIO for each species. This allows the study to determine whether there is spatial genetic structuring among populations. If genetic structure is detected, the study may reveal distinct biogeographic provinces within WIO. This will also determine whether there is connectivity among the regions of the WIO for each wrasse species.
2. To determine the relationships among the different areas of the WIO for each species. This will offer an understanding of whether the localities share a common history of vicariance or dispersal.
3. To identify the processes that have shaped regional relationships. This will provide an understanding of which factors assist dispersal and maintain genetic homogeneity or isolate the populations of these three wrasses and influence the genetic structure.

CHAPTER TWO

MATERIALS AND METHODS

Study area and sampling

Samples of the cigar wrasse *Cheilio inermis* (Forsskål, 1775), goldbar wrasse *Thalassoma hebraicum* (Lacepède, 1801) and crescent-tail wrasse *T. lunare* (Linnaeus 1758) were collected from numerous localities in the Western Indian Ocean (WIO) (Figure 2.1). These localities were selected to span a range of physical and oceanographic features, and known biogeographic boundaries that may have an impact on the genetic structure of fishes. The samples of the wrasse species were obtained from coastal habitats from oceanic and continental islands, and the continental mainland. Samples were also obtained from the biomaterial collection of the South African Institute for Aquatic Biodiversity (SAIAB) in Grahamstown and from collaborators in the WIO. Samples were purchased at landing sites, fish markets or from ornamental fish collectors. Samples were also obtained by active sampling, using rotenone in rock pools in tidal areas, snorkelling and the use of hand nets, and from seagrass beds using seine nets.

Fin clips or a small piece of muscle tissue were taken from each specimen and preserved in 95% ethanol for storage before genetic laboratory analysis. Specimens of each species were measured and photographed in the field. Tissue samples, images of the specimens and voucher specimens were accessioned in the National Fish Collection at SAIAB. The donated tissue samples of *T. lunare* from the Red Sea were not accompanied by photographs to confirm the species identification. The identity of these samples was confirmed using BLAST (Basic Local Alignment Search Tool) after DNA sequencing (Altschul *et al.*, 1990). The BLAST program compares the nucleotide or sequence query of each individual to sequences of the same gene region from known species on GenBank (<http://www.ncbi.nlm.nih.gov/genbank>). Additional sequences of *C. inermis* and *T. hebraicum* were downloaded from GenBank for use in the present study.

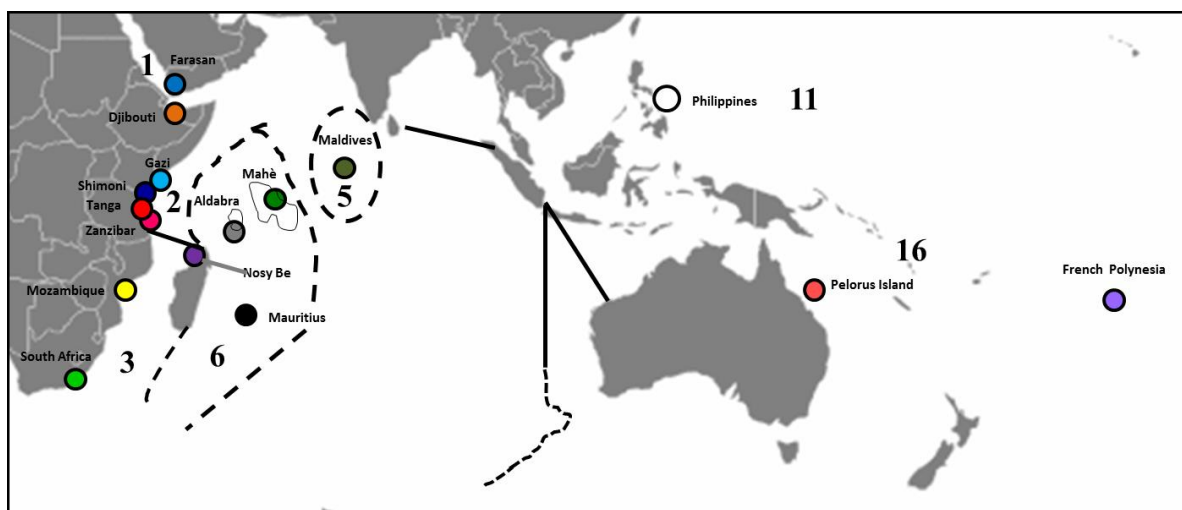


Fig. 2.1: Sampling localities across the Western Indian Ocean for the three wrasse species, and the sampling locality of *Cheilio inermis* in the Philippines. The numbers represent the Western Indian Ocean and West Pacific (Philippines) biogeographic regions corresponding to Santini and Winterbottom (2002) (see Figure 1.1). The Philippines Sea (11), Pelorus Island and French Polynesia (16) are indicated corresponding to Santini and Winterbottom (2002). Solid lines and broken lines indicate potential physical barriers (Santini & Winterbottom, 2002).

Selection of molecular genetic markers

Molecular markers are useful tools that can provide critical insight into the degree of differentiation among populations of widely-distributed species and into the mechanisms that are responsible for the observed genetic structure (Bowen *et al.*, 2001; Rocha, 2003; Klanten *et al.*, 2007). Genetic markers have also been used as a tool for analysing patterns of colonisation and gene flow in marine organisms over varying geographic scales (Palumbi, 1995; Duran *et al.*, 2004). Many of the studies of genetic differentiation in the WIO (Dorenbosch *et al.*, 2006; Gopal *et al.*, 2006; Craig *et al.*, 2007; Klanten *et al.*, 2007; Ragionieri *et al.*, 2010; Visram *et al.*, 2010) have only used a single genetic marker, in most cases a mitochondrial gene fragment. However, mitochondrial DNA (mtDNA) sequence data do not capture the whole genetic history in the case of detecting population genetic structure (Sunnucks, 2000; Muths *et al.*, 2012). The present study will therefore use multiple markers, which will provide different perspectives because of the different modes of inheritance, rates of evolution and levels of genetic variability, and resolution at different time scales (Hoffman *et al.*, 2009). The selected markers should detect the processes operating on both ecological and evolutionary time scales.

Mitochondrial DNA

The analysis of mitochondrial DNA can provide insight into hybridisation, biogeography and phylogenetic relationships (Wilson, 1985). Mitochondrial DNA (mtDNA) has traditionally been the marker of choice (see below) for intra- and interspecific genetic studies of fishes (Avisé *et al.*, 1987; Neethling *et al.*, 2008; Gaither *et al.*, 2010; Visram, 2010). The mitochondrial genome is a small, circular, double-stranded molecule located in mitochondria and contains genes that support aerobic respiration (Moritz *et al.*, 1987; Meyer, 1993; Rokas *et al.*, 2003). Mitochondrial DNA is compressed or compacted with 15 kilobases and has 37 genes: two ribosomal RNA genes (12S rDNA and 16S rDNA), 22 transfer RNAs (tRNA) and 13 protein-coding genes (Wilson *et al.*, 1985; Avisé *et al.*, 1987; Moritz *et al.*, 1987). The mitochondrial gene order of piscines is generally the same as other vertebrates, which is depicted in Figure 2.2. The animal mitochondrial genome mostly shows maternal inheritance, without recombination, has a high mutational rate, a shorter coalescence time and, thus, greater sensitivity in reflecting the genetic impact of population subdivision compared to nuclear genes (Avisé *et al.*, 1987; Moritz *et al.*, 1987; Rokas *et al.*, 2003; Duran *et al.*, 2004).

The ATP synthase subunit 6 (ATPase 6) and cytochrome *b* mtDNA gene regions were selected for this study to determine connectivity among regions as they have been used in many genetic studies (e.g., Bernardi *et al.*, 2004; Lessios and Robertson, 2006, Craig *et al.*, 2007; Beldade *et al.*, 2009; Gaither *et al.*, 2010; Muths *et al.*, 2012). These markers can also determine patterns of genetic differentiation and identify the different mechanisms that operate on evolutionary time-scales. For example, Bernardi *et al.* (2004) found differentiation in wrasses and showed that the Tethyan Event and the closing of the Isthmus of Panama were the major factors shaping the evolution of wrasses.

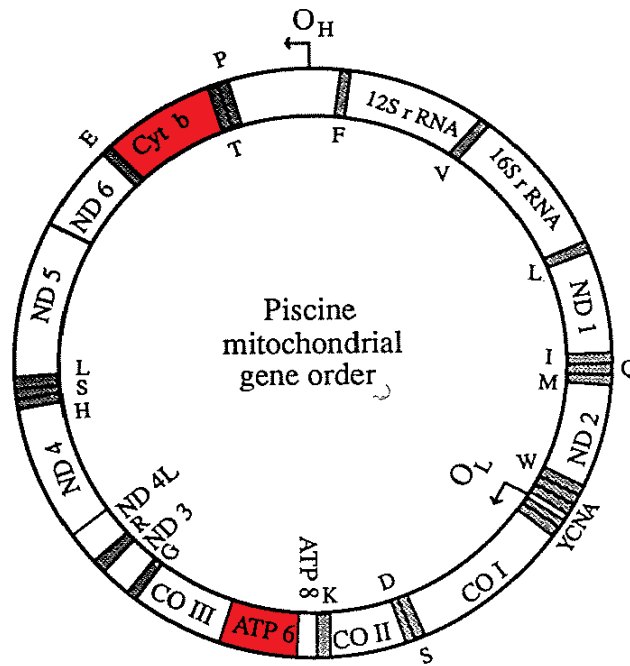


Fig. 2.2: The mitochondrial gene order of fish, showing the positions (shaded) of the ATPase synthase 6 (ATP 6) and cytochrome *b* (cyt *b*) genes. Figure taken from Meyer (1993).

Nuclear DNA (nDNA)

The nuclear genome is located in the nucleus of eukaryotic organisms (Zhang & Hewitt, 2004). Nuclear genetic markers have coding and non-coding regions (Sunnucks, 2000), with most of the nuclear genes employed for population genetic analysis being non-coding introns (Zhang & Hewitt, 2004). These nuclear introns show a higher level of genetic polymorphism within and between species than the coding exons (Chow & Hazama, 1998; Zhang & Hewitt, 2004), which makes them more versatile (Hellberg *et al.*, 2002; Zhang & Hewitt, 2004). Recombination is found regularly in nuclear DNA, varying from locus to locus (Zhang & Hewitt, 2004). Nuclear markers provide paternal and maternal information for the analysis of population genetic structure (Gadagkar *et al.*, 2005). The insertions and deletions (indels) of nuclear genes, especially of introns, contain a large amount of phylogenetic information that should not be ignored (Zhang & Hewitt, 2004). Nuclear DNA polymorphism can be a useful tool for studying the mechanisms of evolution (Zhang & Hewitt, 2004). For example, the analyses of microsatellites revealed high levels of genetic structure in the blotcheye soldierfish *Myripristis berndti* (Muths *et al.*, 2011), whereas mtDNA data for this species showed patterns of historical connectivity (Craig *et al.*, 2007).

Other studies of wrasses have used both mitochondrial and nuclear genes to answer their research questions (Bernardi *et al.*, 2004; Beldade *et al.*, 2009). As mentioned earlier, the nuclear and mitochondrial markers provide alternative perspectives and reflect a more complete genetic history when combined. Therefore, the first intron of the S7 ribosomal protein gene was also selected to address the stated aims of the present study. This nuclear marker has been used by Bernardi *et al.* (2004) to study the biogeography of the genus *Thalassoma*, and by Beldade *et al.* (2009) to evaluate historical biogeography in Creole wrasses.

DNA extraction and sequencing

Total genomic DNA was extracted from muscle tissue or fin clips of each specimen using the Promega Wizard® Genomic DNA Purification Kit (Madison, New York) and by following the manufacturer's protocol. Extracted DNA was stored in a -20 °C freezer until analyses. Fragments of the three genes, the mitochondrial ATPase synthase 6 (ATPase6) and cytochrome *b* (cyt *b*) genes, and the first intron of the nuclear ribosomal protein S7 gene, were amplified by polymerase chain reaction (PCR). PCRs were carried out in 25 µL volumes, containing 6 to 8 µL of template DNA, 3.0 µL of 25 mM MgCl₂, 2 µL of a 8 mM dNTP solution, 2.5 µL of 10X buffer, 0.1 µL of Super-therm *Taq* DNA polymerase, 0.5 µL of 10 pmol of the forward and reverse primers. Hypure™ molecular biology grade (nuclease free) water was added to make up the final PCR volume to 25 µL. Negative controls, without a DNA template, were included in each series of PCR reactions for all three markers. The specific primers were selected from published literature (Table 2.1). PCR amplifications were performed in either an Eppendorf Mastercycler Gradient or a Thermo Hybaid PX2 thermocycling machine. The thermocycling regimes for the amplification of the mitochondrial and nuclear gene regions for each species are presented in the respective chapters.

PCR products were electrophoresed in 1% agarose gels and stained with ethidium bromide for visualization using an ultra-violet transilluminator to confirm successful amplification. The successfully-amplified PCR products were purified using the QIAquick purification kit (QIAGEN). The sequences of the respective gene regions were generated in the forward direction using the BigDye V 3.1 Terminator cycle sequence kit (Applied Biosystems) and analysed on Applied Biosystems ABI 3730XL automated sequencer by a commercial facility (Macrogen, South Korea). Sequences for each locus were edited and misreads were corrected manually using ChromasLITE (Technylesium). The sequences were edited further to a common length using Lasergene 10 SeqMan Pro (DNASTAR). The aligned nuclear S7 intron

I dataset was phased with the PHASE option in DnaSP 5.0 (Librado & Rozas, 2009) to determine the alleles of each specimen from the ambiguities that are present in the sequence.

Table 2.1: Primers sets and primer references used for the amplication and sequencing of the two mitochondrial fragments and nuclear DNA fragment. The forward primer is indicated with an asterix (*) while the reverse primer is denoted with double asterices (**).

Gene	Primer name	Primer sequence (5' to 3')	Source
ATPase 6	8.2_L8331*	GCR TYR GCC TTT TAA GC	Taylor <i>et al.</i> (2002)
	CO3.2_H9236**	GTT AGT GGT CAK GGG GA GGR TC	Taylor <i>et al.</i> (2002)
Cyt b	CB3H**	GGCAAATAGGAARTATCATTC	Beldade <i>et al.</i> (2009)
	CYTTHAL*	AACGGAGCATCNTTCTTCTTT	Bernardi <i>et al.</i> (2004)
S7 intron I	S7RPEX1F*	TGG CCT CTT CCT TGG CC GTC	Chow & Hazama (1998)
	S7RPEX2R**	AAC TCG TCT GGC TTTT CG	Chow & Hazama (1998)

Data analysis

General genetic diversity indices, such as the number of polymorphic or segregating (S) sites, the average number of pairwise differences among individuals (k), and haplotype (h) and nucleotide (π) diversities were estimated for each sampling locality and for the overall sample for each gene region for each species, using DnaSP 5.0. Haplotype diversity (h) describes the number and frequency of different haplotypes in a sample (Nei & Tajima, 1980; Kvist, 2000). The value ranges from 0 to 1 and indicates the probability that two individuals drawn randomly from a sample will have different haplotypes. Nucleotide diversity (π) is the probability that two homologous nucleotides are different (Nei, 1987). The diversity indices also included the number of haplotypes or alleles and the number of private haplotypes or alleles, which is number of haplotypes or alleles restricted to a particular sampling locality.

Bird *et al.* (2011) stated that Wright's (1951) fixation F_{ST} is the most widely used metric to calculate genetic variation. The Φ_{ST} is the analogue of F_{ST} and was used to identify, calculate and compare the degree of genetic differentiation among populations (Bird *et al.*, 2011), using Arlequin version 3.5 (Excoffier & Lischer, 2010). The differentiation was corrected based on suitable model of nucleotide substitution. The most appropriate model of nucleotide

substitution was determined for the respective data sets using MEGA version 6.6 (Tamura *et al.*, 2013), with model selection based on Akaike's (1974) Information Criterion (AIC), which is used to find the optimal model and the best combination of parameters from the models assessed. Values of Φ_{ST} range from 0, indicating genetic homogeneity, and are maximised at 1, indicating genetic differentiation or the fixation of different haplotypes being compared in a sample (Bird *et al.*, 2011). Wright's (1978) interpretation of Φ_{ST} is based on allozyme loci from a range of species. According to Wright (1978), pairwise Φ_{ST} values ranging from 0 to 0.05 indicate little differentiation, 0.05 to 0.15 moderate differentiation, and 0.15 to 0.25 indicate high differentiation. Ward and Grewe (1994) revised population genetic data for fishes and found that the mean Φ_{ST} was estimated at 0.062 for a sample of different marine fish species, considerably less than for anadromous ($\Phi_{ST} = 0.108$) and freshwater ($\Phi_{ST} = 0.222$) fish species. Values less than the mean Φ_{ST} for marine fish indicate a lack of differentiation (Ward & Grewe, 1994).

Analyses of Molecular Variance (AMOVA) were performed to examine genetic structure at various hierarchical levels using Arlequin. AMOVA is based on a matrix of genetic distances among sampled localities and yields F-statistic analogues (Φ_{CT} , Φ_{SC} and Φ_{ST}) for the comparison of variation within and among groups (Excoffier *et al.*, 1992; Bird *et al.*, 2011). The analogue Φ_{CT} is estimated among groups, Φ_{SC} among populations within groups (Meirmans, 2006) and Φ_{ST} is based on genetic structure among localities. AMOVA was performed by grouping the sampled localities according to geographic proximity and biogeographic regions (e.g., Santini & Winterbottom, 2002). The localities were not grouped according to hydrogeographic regions and biogeographic provinces (e.g., Longhurst, 1998; Briggs & Bowen, 2012) due to the small sample size. The details and arrangements of the localities grouped in the AMOVA are discussed in the individual chapters for each species (see Chapters 3 and 4). The significance of pairwise Φ_{ST} and AMOVAs was tested using a permutation procedure (Excoffier *et al.*, 1992), with 10000 permutations. The nearest neighbour statistic S_{nn} (Hudson, 2000) was calculated in DnaSP to investigate the geographic distribution of genetic variation. The statistic measures how frequently closely related sequences are from the same geographic locality (Hudson, 2000). Values of S_{nn} below 0.5 indicate panmixia and values closer to 1 indicate genetic differentiation (Hudson, 2000). The nearest neighbour statistic S_{nn} was calculated for the overall sample for each gene region for each species.

Networks were selected for the present study to investigate and present non-bifurcating, genealogical relationships among haplotypes (Clement *et al.*, 2000). Median-joining networks (Bandelt *et al.*, 1999) were constructed using Network version 4.6 (www.fluxus-technology.com). Each species was analysed with the respective gene fragments treated separately, because each gene region had different sample sizes and the samples in each data set were not always from the exact same sampling localities. The same samples were also not included in each data set. Neighbour joining trees were constructed in MEGA 6.6 (Tamura *et al.*, 2013) for the two *Thalassoma* species (*T. hebraicum* and *T. lunare*) to examine relationships among haplotypes and alleles, respectively. Nodal support for the relationships was determined by bootstrapping (Felsenstein, 1985), using 1000 replicates for neighbour joining trees.

Demographic history

Tajima's (1989) D and Fu's (1997) F_s neutrality tests, and mismatch distributions and Harpending's raggedness statistics (r) (Harpending, 1994) were also calculated in Arlequin to identify the historical signatures of demographic expansion, population decline or growth, or stability. The neutrality tests and mismatch statistics were calculated for each sampling locality and the overall sample, using each gene region, for each wrasse species

Neutrality tests can determine whether the markers in question fit the expectations of neutral evolution or selection. Tajima's D test and Fu's F_s are used to test the selective neutrality of a random sample of DNA sequences and can provide evidence of demographic changes brought about by population decline and expansion (Tajima, 1989). Tajima's D is based on the number of segregating sites (S) and the average number of nucleotide differences (k) between DNA sequences in a sample under a neutral mutation model (Tajima, 1989). Negative values of Tajima's D indicate excesses of low frequency haplotypes, generally reflecting a population expansion following a severe reduction in population size (Tajima, 1989). A positive D value indicates balancing selection or mixture of isolated populations (Rand, 1996). Fu's F_s test detects an excess of recent mutations (rare alleles) and is more powerful than other tests in cases of population expansion (Fu, 1997). A mismatch distribution is the distribution of pairwise differences among sequences in a sample and has been widely used to estimate the demographic parameters of past population expansions (Schneider & Excoffier, 1999). A unimodal mismatch distribution indicates that the population has undergone population expansion, while a population at mutation-drift equilibrium produces a multimodal distribution

(Rogers & Harpending 1992; Excoffier *et al.*, 2004). Harpending's (1994) raggedness statistic (r) and the Sum of Squared Deviation (SSD) were used to measure the smoothness of the mismatch distribution curve and to determine whether it fits to models of sudden population expansion or constant population size by means of parametric bootstrapping (Rogers & Harpending, 1992; Schneider & Excoffier, 1999).

Demographic changes in the population were calculated by the parameters of Theta (θ_0 , θ_1) and Tau (τ), derived from the mismatch distribution using a non-linear least squares approach (Schneider & Excoffier, 1999). Tau (τ) is defined as the expansion time, expressed in units of mutational time, and was used to calculate the actual time of expansion (Rogers & Harpending, 1992) with the equation $T = \tau/2u$ (Rogers & Harpending, 1992). Here, T is the time since expansion, u = number of base pairs $\times$ mutation rate (μ) $\times$ generation time in years. The generation time is taken as the age of sexual maturity for females (Gaither *et al.*, 2010). The generation times of the three wrasse species, *Thalassoma hebraicum*, *T. lunare* and *C. inermis*, are unknown. The bluehead wrasse (*Thalassoma bifasciatum*) matures at 3 years (yrs) (Warner, 1998); therefore, a generation time of 3 years was used in the present study to calculate the time of expansion and effective female population size for each species, using the mitochondrial gene regions. The sequence divergence rate of ATPase 6 is 1.3 % per million years (Ma) (Bermingham *et al.*, 1997) and the mutation rate of this marker is 0.65 % per Ma. Cytochrome *b* has a sequence divergence rate of 2% per Ma (Bowen *et al.*, 2001, 2006; Craig *et al.*, 2007; Muths *et al.*, 2011) and the mutation rate is 1% per Ma. The mutation rate of nuclear S7 intron I is not documented and transmission dynamics are different; therefore, time of expansion was not calculated for S7 using this approach. Effective female population size (N_{ef}) was estimated for time zero (θ_0) and the current population size (θ_1) using the equation $\theta = 2N_{ef} \mu$ (Rogers & Harpending, 1992).

CHAPTER THREE

REGIONAL DIFFERENTIATION OF *CHEILIO INERMIS* IN THE WESTERN INDIAN OCEAN

Introduction

Cheilio inermis (Forsskål, 1775), commonly known as the cigar wrasse (Figure 3.1), belongs to a monotypic genus in the family Labridae. This species was initially described from the Red Sea (Forsskål 1775) and occurs throughout the Indo-Pacific, from the Hawaiian and Easter Islands, northwards to southern Japan and south to Lord Howe Island (Australia) in tropical waters (Figure 3.2; Heemstra & Heemstra, 2004; Kuitert, 2010). In the Western Indian Ocean (WIO), *C. inermis* ranges from the Red Sea and Oman south to the Kei River (South Africa) (Figure 3.2; Heemstra & Heemstra, 2004; Kuitert, 2010). Although *C. inermis* is considered a single widely-distributed species, distinct colour variation is observed between geographic populations (see Kuitert, 2010). The young have different colour patterns to the adults, with the juveniles being mottled brown and green, sometimes with a broad lateral stripe (Myers, 1991). Rare individuals or adults are uniformly yellow and large males are orange, black and white, with some of the larger males having a multi-coloured patch behind their pectoral fin (Myers, 1991). Adults of *C. inermis* occur in loose aggregations, but form schools to spawn (Kuitert, 2010).



Fig. 3.1: *Cheilio inermis*, cigar wrasse. Photographed specimen: SAIAB 197031, 235 mm standard length, Shimoni landing site, Kenya (photo: SAIAB).

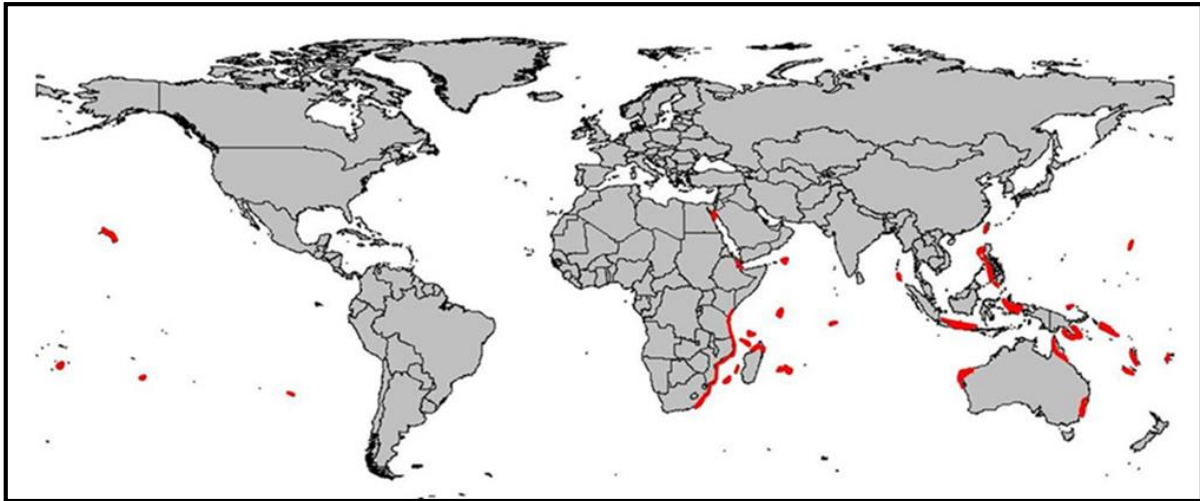


Fig. 3.2: Distribution of *Cheilio inermis* (red) in the Indo-Pacific, including the Western Indian Ocean (modified from <http://www.fishbase.org>)

Cheilio inermis is not a well-studied species and there is little information available on its reproductive biology. There is also no information on the genetic diversity and structure of *C. inermis*. Studies on the regional differentiation or genetic structure of widespread marine organisms have revealed that biological processes and various physical features of the environment (such as ocean current circulation) influence these patterns (Doherty *et al.*, 1995; Riginos & Nachman, 2001; Bourjea *et al.*, 2007). For example, water movements can either disperse larvae over long distances (Groeneveld, 2012), to promote dispersal and genetic connectivity, or return the larvae to near their parental origin (Chiswell & Booth 1999), causing genetic differentiation (Dorenbosch *et al.*, 2006).

For coral reef fishes, the different life history stages (adults, juveniles, larvae and eggs) have varying dispersal capabilities (Bernardi *et al.*, 2001) and, therefore, play a role in genetic structuring. For example, most adult reef fishes are sedentary (Sale, 1980), restricting adult dispersal among populations. Thus, dispersal for such species is accomplished through other life stages, such as the pelagic larval stage (Leis, 1991). The pelagic eggs and larvae of most coral reef fishes are thought to be dispersed by ocean currents (Roberts, 1997). However, it has been reported that larvae may not passively drift away with ocean currents but may actively choose to settle near their parental origin and this reduces dispersal and promotes genetic differentiation (Planes *et al.*, 1996, 1998; Leis, 2002). The larvae of some reef fishes have also been described to swim against ocean currents (Leis, 2002). Reef fish species have larvae that

may sense reefs from a distance through smell or sound, which helps them navigate to settlement habitats (Leis, 2010). This means that the larvae of the reef fish do not necessarily rely on the ocean currents to find settlement habitat.

The pelagic larval duration (PLD) of some marine species can last from days to several months (Brothers *et al.*, 1983). Reef fishes that have a long pelagic phase are assumed to have high dispersal and to show genetic connectivity over a large geographic range (Shulman & Bermingham, 1995; Klanten *et al.*, 2007). For example, the long PLD (around 30 days) of the bluestripe snapper *Lutjanus kasmira* may have assisted dispersal and maintained genetic connectivity across the WIO (Muths *et al.*, 2012). *Cheilio inermis* is reported to have a long pelagic larval phase of 43 to 73 days (Victor, 1986) and it would be reasonable to expect greater connectivity and little genetic differentiation across its distribution. A fish species with a shorter larval duration is expected to show low levels of dispersal and become genetically structured (Riginos & Victor, 2001).

Cheilio inermis is a benthopelagic species that is associated with seagrass beds, coastal bays and algal-covered flats. This species is also occasionally found in lagoons and seaward reefs to a depth of 30 m, and also enters estuaries (Heemstra & Heemstra, 2004; Kuiter, 2010). Juveniles live in seagrass beds or at the bottom of brown algae (*Sargassum*) (Kuiter, 2010). Lagoons and coral reefs are naturally patchy habitats (Ragionieri *et al.*, 2010), such that populations on patches of suitable habitat might only be connected through the exchange of pelagic larvae. High genetic differentiation could therefore be expected because the distance between the patches of suitable habitat has the potential to limit dispersal among populations (Gaither *et al.*, 2010; Muths *et al.*, 2011). For example, populations of the blennioid fish *Axoclinus nigricaudus* were found to be genetically distinct because of the distance between rocky shores (Riginos & Nachman, 2001). Therefore, populations of *C. inermis* could be genetically isolated because of the patchy distribution of their preferred habitats.

As mentioned above, *C. inermis* is regarded as a single species, but has different colour patterns across its wide distribution. This species lives in isolated habitats and adults are sedentary. Thus, it may be reasonable to expect evidence of genetic differentiation and the possible existence of a cryptic species complex across its wide distribution. Cryptic species are two or more different species that are mistakenly classified under one species name (Bickford *et al.*, 2006) and they are common in marine species (Knowlton, 1993). For example, the goliath

grouper *Epinephelus itajara* was regarded as a single species, but was found to be genetically-structured across its Indo-Pacific and western Atlantic distribution, containing two or more species (Craig *et al.*, 2008). On the other hand, the pelagic larval phase of *C. inermis* is long; therefore, ocean currents and larval dispersal can assist in connecting different populations, maintaining genetic homogeneity across its full distribution.

The aims of the current study are:

1. To determine the extent of spatial genetic differentiation among separated localities in the WIO, as well as Djibouti and Philippines, for *C. inermis* using multiple markers.
2. To determine whether *C. inermis* represents a single species or a species complex across the sampled distribution, using genetic data.
3. To determine the factors that might maintain or affect the genetic structure of *C. inermis* in the WIO.

Materials and methods

Sampling

Samples of *Cheilio inermis* were collected from several localities using different methods and approaches described in Chapter 2. Some of the WIO sequences were downloaded from GenBank.

DNA amplification

Total DNA extraction (i.e. from fin clips and muscle tissue) was carried out using a commercial extraction kit (see Chapter 2). The mitochondrial ATPase 6 and cytochrome *b* fragments and the S7 intron I nuclear gene fragment were amplified for *C. inermis* by Polymerase Chain Reaction (PCR). The PCR conditions for the amplification of the three gene fragments are detailed below (Table 3.1). Successful PCRs were confirmed by electrophoresis in 1% agarose gel and sequenced in one direction by Macrogen (South Korea) (see Chapter 2).

Table 3.1: Thermocycling conditions for the amplification of the respective mitochondrial (ATPase 6 and cytochrome *b*) and nuclear (S7 intron I) gene regions from *Cheilio inermis*.

Gene region	PCR profile cycle					
	Stage 1		Stage 2		Stage 3	
	Initial denaturation	Denaturation	Annealing	Extension	Cycles	Final extension
ATPase 6	94 °C, 3 min	94 °C, 45 s	56 °C, 1 min	72 °C, 1 min	35	72 °C, 10 min
Cytochrome <i>b</i>	94 °C, 3 min	94 °C, 45 s	54-56 °C, 45 s	72 °C, 1 min	35	72 °C, 10 min
S7 intron I	94 °C, 3 min	95 °C, 30 s	56 °C, 1 min	72 °C, 2 min	30	72 °C, 10 min

Genetic data analysis

Genetic diversity indices for individual localities and the overall sample were obtained and median-joining networks were constructed as described in Chapter 2. For cytochrome *b* data, MEGA version 6.6 (Tamura *et al.*, 2013) was used to calculate sequence divergence between the localities of the south (Mozambique and Madagascar) and north (Tanzania, Kenya and Djibouti) (see below), to evaluate the existence of cryptic diversity within the widespread *C. inermis*. The most appropriate model of nucleotide substitution for each dataset was selected using MEGA version 6.6 as described in Chapter 2. Pairwise population comparisons, Analyses of Molecular Variance (AMOVA) and nearest-neighbour statistics were also calculated as described in Chapter 2. The arrangement that maximised the values of Φ_{CT} among groups in the AMOVA was selected as the most likely hypothesis. Sampling localities that consisted of single individuals were excluded from the pairwise Φ_{ST} and AMOVAs. The neutrality tests, mismatch statistics, mismatch distribution and demographic parameters were examined for each sampling locality and overall sample as discussed in Chapter 2. As discussed previously, the mismatch distribution parameters (τ , θ_0 , θ_1) were used to calculate the time of expansion and effective population sizes (Rogers & Harpending, 1992; see Chapter 2). The formula $T = \tau/2\mu$ (Rogers & Harpending, 1992) was used to calculate time of expansion for the mitochondrial markers. The equation $\theta = 2N_{ef}\mu$ was used to calculate the effective population size (N_{ef}) for time zero (θ_0) and the present population size (θ_1). Generation time is represented by the age of sexual maturity of females (see Chapter 2). The females of *C. inermis* mature at a length of 121 to 150 mm (Anand & Pillia, 2002), but the generation time of *C. inermis* is

unclear. The bluehead wrasse *Thalassoma bifasciatum* was estimated to mature at 3 years (yrs) (Warner, 1998). The generation time of 3 yrs and mutation rates of the mitochondrial markers were used in the present study to calculate the time of expansion and effective female population size for *C. inermis* as discussed in Chapter 2.

Results

In total, 82 fin clips or tissue samples of *C. inermis* were collected from various localities in the WIO and Pacific (Table 3.2). Sequences from Nosy Be-Antsiranana (Madagascar) were downloaded from GenBank for the cytochrome *b* data set (Accession numbers: JF458026, JF458027 and JF458028). The localities that had one individual were combined with the nearest locality in the same biogeographic region (i.e. Santini and Winterbottom, 2002). Pemba (Mozambique) had one individual therefore Pemba and Inhaca Island were combined as one locality (Mozambique). Distant localities that had more than one individual were treated separately (Table. 3.2).

Table 3.2: Sampling localities, locality codes and sample sizes for *Cheilodactylus inermis* collected from different Western Indian Ocean localities, the Red Sea and the Philippines.

Sampling localities	Code	Sample size	GPS coordinates	
			Latitude	Longitude
Philippines	PP	1	-8.4035	120.9824
Djibouti, Red Sea	DJ	1	-11.7160	143.9751
Gazi, Kenya	GZ	4	-4.4545	39.5157
Shimoni, Kenya	SM	26	-4.6478	39.3810
Tanga, Tanzania	TN	4	-5.1000	39.1000
Zanzibar Island, Tanzania	ZN	4	-6.1295	39.4667
Aldabra Atoll, Seychelles	AD	1	-9.4047	46.2094
Pemba, Mozambique	PM	1	-12.9602	40.9702
Inhaca Island, Mozambique	MZ	36	-29.867	31.048
Nosy Be Antsiranana, Madagascar	NB	3	-13.4546	48.1525
La Morne, Mauritius	MA	1	-20.4692	54.3438
Overall		82		

Genetic diversity analysis

The genetic diversity indices were estimated for the respective gene regions for each locality and the overall sample (Table 3.3A and B). A total of 748 base pairs (bp) of the ATPase 6 gene fragment amplified for 65 individuals (Table 3.3). The ATPase 6 gene revealed a number of polymorphic sites ($S = 2$ to 18) and the average number of pairwise differences (k) ranged from

1.33 to 2.61 per locality. The 1036 bp cytochrome *b* dataset from 66 individuals yielded numbers of polymorphic sites (*S*) from 2 to 40, and *k* within the sampled localities ranged from 1.33 to 8.67. The 684 bp nuclear S7 intron I data of 32 individuals (64 phased sequences) had the number of polymorphic sites (*S*) ranging from 3 to 12 and values of *k* from 3.00 to 4.67 per locality.

Table 3.3: Sample sizes (N) and diversity estimates at each of the localities from the Western Indian Ocean and the Philippines (Pacific) and for the total sample of *Cheilodactylus inermis* for the ATPase 6 and cytochrome *b* (A) and nuclear S7 intron I (B) datasets. Genetic diversity indices included the number of haplotypes (N_h) or alleles (N_A), number of private haplotypes (N_{PH}) or private alleles (N_{PA}), haplotype diversity (h) or allelic diversity (A), nucleotide diversity (π), number of polymorphic sites (S) and average pairwise differences (k) among individuals. Standard deviations are presented for haplotype or allelic and nucleotide diversities.

A

Locality	ATPase 6 (748 bp)							Cytochrome <i>b</i> (1036 bp)						
	N	N_h	N_{PH}	$h \pm SD$	$\pi \pm SD$	S	K	N	N_h	N_{PH}	$h \pm SD$	$\pi \pm SD$	S	k
Mozambique	34	17	12	0.87 ± 0.05	0.003 ± 0.0004	18	2.30	31	22	15	0.96 ± 0.02	0.005 ± 0.0007	40	5.38
Tanga	3	2	1	0.67 ± 0.31	0.002 ± 0.0002	2	1.33	3	2	1	0.67 ± 0.31	0.001 ± 0.0006	2	1.33
Zanzibar	4	4	1	1.00 ± 0.03	0.003 ± 0.0007	4	2.00	4	4	2	1.00 ± 0.18	0.003 ± 0.0009	6	3.33
Gazi	4	4	1	1.00 ± 0.18	0.003 ± 0.0008	4	2.17	4	3		0.83 ± 0.02	0.001 ± 0.0005	3	1.50
Shimoni	17	10	4	0.90 ± 0.05	0.003 ± 0.0005	11	2.61	20	12	7	0.85 ± 0.08	0.003 ± 0.0004	19	2.98
Nosy Be	-	-		-	-	-	-	3	2	1	0.67 ± 0.31	0.008 ± 0.0039	13	8.67
Mauritius	1	-		-	-	-	-	-	-	-	-	-	-	-
Djibouti	1	-		-	-	-	-	1	-	1	-	-	-	-
Philippines	1							-	-	-	-	-		
Overall	65	25	19	0.88 ± 0.03	0.003 ± 0.0001	26	2.29	66	36	27	0.94 ± 0.02	0.004 ± 0.0005	56	4.52

B

Locality	S7 intron I (684 bp)						
	N	N_A	N_{PA}	$A \pm SD$	$\pi \pm SD$	S	K
Mozambique	18	22	11	0.93 ± 0.12	0.006 ± 0.0004	11	4.67
Tanga	2	4	-	1.00 ± 0.18	0.005 ± 0.0012	7	3.38
Zanzibar	3	5	1	0.93 ± 0.12	0.005 ± 0.0014	9	3.80
Gazi	3	6	1	1.00 ± 0.01	0.005 ± 0.0009	9	4.13
Shimoni	5	9	4	0.98 ± 0.05	0.005 ± 0.0009	12	3.89
Aldabra Atoll	1	2	1	1.00 ± 0.50	0.004 ± 0.0022	3	3.00
Overall	32	33	18	0.97 ± 0.01	0.006 ± 0.0003	21	4.41

Overall haplotype (h) or allelic diversities (A) were high for all gene markers: ATPase 6 ($h = 0.88 \pm 0.03$), cytochrome b ($h = 0.94 \pm 0.02$) and S7 intron I ($A = 0.97 \pm 0.01$) (Table 3.3A and B). These values were also high within localities ($h = 0.67$ to 1.00) across all markers. Nucleotide diversities (π) for the overall sample were low for ATPase 6 ($\pi = 0.003 \pm 0.0001$), cytochrome b ($\pi = 0.004 \pm 0.0005$) and S7 intron I ($\pi = 0.006 \pm 0.0003$).

The median-joining networks for the two mitochondrial DNA genes are presented in Figure 3.1. ATPase 6 data set had 25 unique haplotypes (66 individuals), with six of these haplotypes being shared among individuals or localities. Four of the shared haplotypes were widespread among localities and found in the centre of the network (H1, H2, H4 and H5). Haplotype (H1) was shared among WIO localities (Mozambique, Mauritius, Gazi, and Shimoni) and the Philippines (Pacific). The four common haplotypes represented 46% of the sample. The central widespread haplotypes (H1, H2, H4 and H5) probably indicates a common ancestral haplotype in the species (Templeton *et al.*, 1995). There were 19 singleton haplotypes (Posada & Crandall, 2001), represented by a single individual. These were from Mozambique, Tanga, Zanzibar, Gazi and Shimoni. The missing intermediates (mV) represent extant haplotypes that were not sampled (Posada & Crandall, 2001) during the course of the study.

The cytochrome b data set had 36 unique haplotypes of *C. inermis* from seven localities (Table 3.3; Figure 3.3). The network revealed a general star-like shape with one haplotype (H22) at the centre, from which low frequency haplotypes radiated. Some of the haplotypes from Mozambique and Nosy Be-Antsiranana (Madagascar) were divergent from the main network (at least seven mutational steps) and were also highly divergent from other haplotypes from the same localities. There were three haplotypes (H11, H12 and H22) that were widespread across WIO localities. Haplotype (H2) was shared between Nosy Be (Madagascar) and Gazi. Haplotypes H3 and H8 were shared between Mozambique and Shimoni (Kenya). Three haplotypes (H5, H10 and H27) were shared among individuals within the same locality. This data set had 27 singleton haplotypes which were found in some of the included localities. The mean sequence divergence between the north and south localities is 0.4%. The minimum sequence divergence between haplotypes is zero and the maximum sequence divergence between haplotypes is 1.5%.

The phasing of the 32 aligned nuclear S7 intron I sequences (684 bp) yielded 33 unique alleles in the species from six localities (Figure 3.4). These localities were Mozambique, Tanga, Zanzibar, Gazi, Shimoni and Aldabra Atoll. There were 15 shared alleles in the network (Figure 3.2). There were four widespread alleles in the network (H2, H3, H4 and H5). Allele H3 is central in the network probably indicating a persistent ancestral allele. Two (H3 and H5) common alleles represented 11% of the sample. Three alleles (H20, H24 and H26) were shared between Mozambique and Gazi. Alleles H19 and H25 were shared between Mozambique and Shimoni, while allele H29 was shared among Shimoni and Gazi. There were also alleles that were shared among individuals within the same locality (H8, H9, H14, and H16). Singletons were found in Mozambique, Zanzibar, Gazi, Shimoni and Aldabra Atoll. There were no singletons for Tanga in the network.

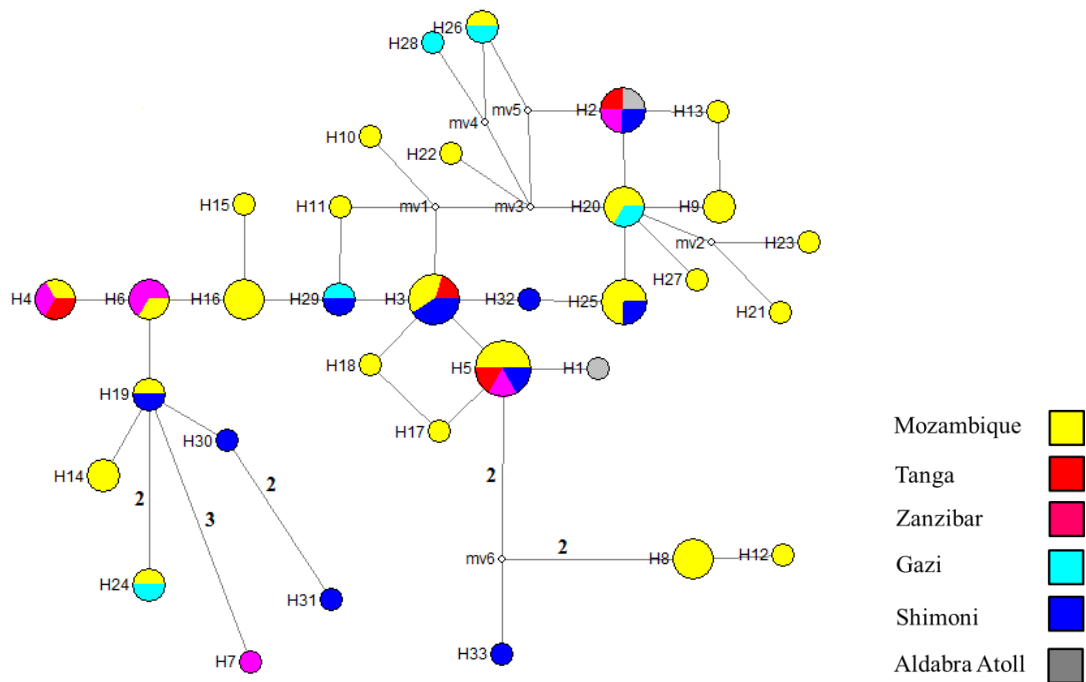


Fig. 3.4: Median-joining network of genealogical relationships among nuclear S7 intron I alleles found in *Cheilio inermis* from six localities across Western Indian Ocean. Pie charts are coloured according to the geographical origin of the allele. The size of nodes is proportional to number of individuals observed with a given allele. Connecting lines represent one mutational step unless indicated otherwise. Missing alleles are represented by small white circles. Loops represent alternative hypotheses.

The estimated best-fitting model of nucleotide substitution for both the ATPase 6 and cytochrome *b* mtDNA gene regions was a Hasegawa-Kishino-Yano (HKY+ G) (Hasegawa *et al.*, 1985) model, with a gamma distribution ($\alpha = 0.05$) of rate variation for the ATPase 6 gene only. The best model for nuclear S7 intron I data was a Jukes-Cantor (JC) (Jukes & Cantor, 1969) model, with a gamma distribution of rate variation ($\alpha = 0.05$). As the estimated model for the mitochondrial gene regions was not available in Arlequin, the model of Tamura and Nei (1993) was most similar and was selected for the mitochondrial markers.

Pairwise Φ_{ST} estimates of population differentiation are displayed in Table 3.4. The ATPase 6 data set revealed non-significant ($P > 0.05$) pairwise Φ_{ST} comparison values, which ranged from -0.09 to 0.19. This data set revealed some pairwise Φ_{ST} values that were lower than 0.05 and 0.062 which have been suggested to reflect little differentiation (Wright, 1978), particularly in marine fish species (Ward & Grewe, 1994). ATPase 6 data highlighted the differentiation of Tanga (Tanzania) from Mozambique ($\Phi_{ST} = 0.15$) and Shimoni ($\Phi_{ST} = 0.12$). Gazi was differentiated from Mozambique ($\Phi_{ST} = 0.15$) and Zanzibar ($\Phi_{ST} = 0.19$).

The mitochondrial cytochrome *b* data set revealed little to great genetic differentiation among localities (Table 3.4B). The cytochrome *b* data set showed significant ($P < 0.05$) differentiation for comparisons ($\Phi_{ST} > 0.27$) between Tanga and the three sampling localities (Mozambique, Shimoni and Gazi). Zanzibar was moderately differentiated from Tanga ($\Phi_{ST} = 0.09$). Cytochrome *b* data suggested differentiation of Nosy Be from the three localities: Tanga ($\Phi_{ST} = 0.32$), Shimoni ($\Phi_{ST} = 0.18$), and Zanzibar ($\Phi_{ST} = 0.10$).

The nuclear S7 intron I revealed non-significant ($P > 0.05$) pairwise Φ_{ST} values that ranged from 0.00 to 0.14, which indicate little to moderate genetic differentiation (Table 3.4A). The S7 intron I gene detected differentiation of Gazi from two localities, Zanzibar ($\Phi_{ST} = 0.14$) and Shimoni ($\Phi_{ST} = 0.11$). Hudson's (2002) nearest-neighbour statistic revealed low values for all gene markers (ATPase 6: $S_{nn} = 0.39$; cytochrome *b*: $S_{nn} = 0.43$, S7 intron I: $S_{nn} = 0.39$), indicating little genetic differentiation among localities.

Table 3.4: Pairwise Φ_{ST} comparison values between localities of *Cheilio inermis* based on mitochondrial (A) ATPase 6 and nuclear S7 intron I data sets (ATPase 6: below diagonal; S7 intron I: above diagonal) and (B) the mitochondrial cytochrome *b* data set. Significant ($P < 0.05$) values are indicated in bold.

A

Locality	Mozambique	Tanga	Zanzibar	Gazi	Shimoni
Mozambique	-	-0.10	0.01	0.03	0.03
Tanga	0.15	-	-	-0.03	-0.09
Zanzibar	0.02	0.01	-	0.14	0.00
Gazi	0.15	0.01	0.19	-	0.11
Shimoni	0.02	0.12	0.06	-0.09	-

B

Locality	Nosy Be	Mozambique	Tanga	Zanzibar	Gazi
Nosy Be	-				
Mozambique	0.07	-			
Tanga	0.32	0.27	-		
Zanzibar	0.10	0.04	0.09	-	
Gazi	0.01	-0.02	0.58	0.03	-
Shimoni	0.18	0.04	0.39	0.06	-0.12

Sampling localities with sample sizes of one were not included in the AMOVAs. The grouping that best explained the variation in the ATPase 6 and S7 intron I data was defined as: Gazi and Shimoni (1), Tanga and Zanzibar (2) and Mozambique (3). The arrangement of localities that best described the cytochrome *b* data was as follows: Gazi and Shimoni (1), Tanga and Zanzibar (2), Mozambique and Nosy Be (3). The Φ_{CT} values among regions were maximized when the localities were grouped according to proximity.

The mitochondrial gene regions detected genetic variation among regions (Table 3.5, ATPase 6: 9.31%, $\Phi_{CT} = 0.093$, $P = 0.05$; cytochrome *b*: 9.50%, $\Phi_{CT} = 0.095$, $P < 0.05$). By contrast, S7 intron I revealed a negative component of variation (-3.41, $\Phi_{CT} = -0.034$) among regions, and this was not significant ($P > 0.05$). There was no variation found among localities within regions for the two mitochondrial gene regions (ATPase 6: -8.06%, $\Phi_{SC} = -0.089$; cytochrome *b*: -0.57%, $\Phi_{SC} = 0.006$) and these values were not significant ($P > 0.05$). S7 intron I data revealed little variation that was attributed to the differences among localities within regions (2.75%, $\Phi_{SC} = 0.026$, $P < 0.05$). Most of the variation was found within localities (ATPase 6: 98.75%; cytochrome *b*: 91.07%, $P < 0.05$; S7 intron I: 100.67%) for the three gene regions.

Table 3.5: The results of the AMOVAs examining genetic variation at different hierarchical levels for *Cheilio inermis* sampled from several localities across the Western Indian Ocean for the ATPase 6, cytochrome *b* and nuclear S7 intron I data sets. The hierarchical arrangement of the data sets was based on proximity as described in the text. Significant ($P < 0.05$) values are indicated in bold.

Gene region	Source of variation	d.f.	Sum of squares	% of variation	<i>P</i> value
ATPase 6	Among regions	2	3.64	9.31	0.05
	Among localities within regions	3	2.36	-8.06	0.73
	Within localities	57	66.35	98.75	0.41
Cytochrome <i>b</i>	Among regions	2	14.27	9.50	0.04
	Among localities within regions	5	9.54	-0.57	0.21
	Within localities	59	144.56	91.07	0.03
S7 intron I	Among regions	2	3.07	-3.41	0.80
	Among localities within regions	2	4.76	2.75	0.21
	Within localities	57	116.30	100.67	0.51

Tajima's D values and Fu's F_s were negative and significant for the overall sample for the three gene regions (Table 3.6). These statistics indicate an excess of low-frequency haplotypes, which provides evidence of population expansion (Tajima, 1989). All of these gene regions revealed unimodal mismatch distributions, indicating that the population had undergone a recent population expansion (Figure 3.5). Harpending's raggedness index (r) and the Sum of Squared Deviations (SSD) were low and significant for the overall sample, for the three gene regions (see Table 3.6). These statistics indicate a good fit to the model distribution of population expansion for all of the gene regions

Table 3.6: Neutrality test indices, including Tajima's (1989) D and Fu's (1997) F_s , and the mismatch distribution statistics, including Harpending's (1994) raggedness index (r) and the Sum of Squared deviations (SSD), for the (A) mitochondrial ATPase 6 and cytochrome b , and (B) nuclear S7 intron I datasets. Significance is indicated by asterices: $*P < 0.05$, $**P < 0.01$ and $***P < 0.001$.

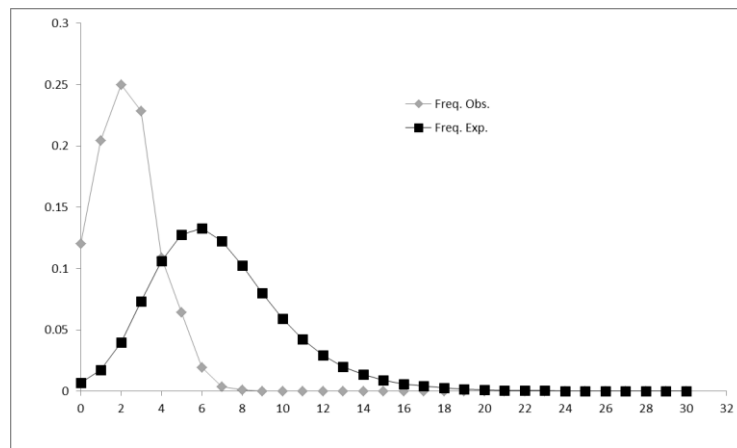
A

Locality	ATPase 6								Cytochrome b							
	D	P value	Fu F_s	P value	r	P value	SSD	P value	D	P value	Fu F_s	P value	r	P value	SSD	P value
Mozambique	-1.60	0.03*	-11.29	0.00***	0.43	0.02*	0.34	0.00***	-1.69	0.02*	-11.82	0.00***	0.56	0.01**	0.79	0.00***
Tanga	0.00	0.93	-1.06	0.59	0.24	0.57	0.11	0.26	0.00	0.93	1.06	0.58	1.00	0.48	0.26	0.11
Zanzibar	-0.78	0.19	-1.87	0.03*	0.47	0.22	0.43	0.04*	0.18	0.71	-1.08	0.09	0.68	0.22	0.65	0.06
Gazi	-0.06	0.58	-1.74	0.04*	0.66	0.17	0.53	0.02*	-0.75	0.23	-0.28	0.21	0.08	1.00	0.95	0.00***
Shimoni	-0.73	0.26	-4.09	0.01*	0.58	0.05*	0.56	0.00***	-1.67	0.03*	-5.25	0.00***	0.13	0.05*	0.04	0.07
Nosy Be	-	-	-	-	-	-	-	-	0.00	0.70	3.92	0.92	1.00	0.50	0.45	0.00***
Overall	-1.84	0.01**	-19.49	0.00***	0.01	0.02*	0.03	0.00***	-1.98	0.00***	-25.51	0.00***	0.03	0.02*	0.05	0.00***

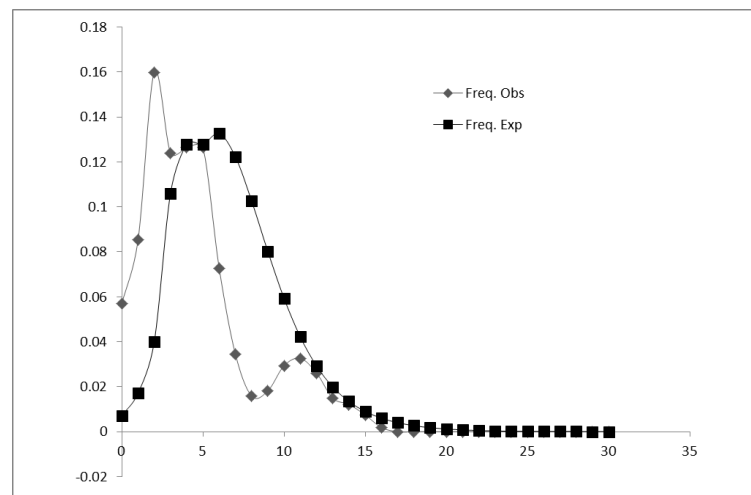
B

Locality	S7 intron I							
	D	P value	Fu F_s	P value	r	P value	SSD	P value
Mozambique	-0.41	0.38	-6.20	0.00***	0.77	0.00***	0.88	0.00***
Tanga	0.18	0.70	-0.66	0.16	0.33	0.48	0.41	0.10
Zanzibar	1.41	0.93	0.00	0.40	0.66	0.08	0.64	0.03*
Gazi	0.16	0.58	-1.53	0.40	0.45	0.11	0.34	0.06
Shimoni	0.70	0.24	-1.29	0.20	3.38	3.05	0.03	0.35
Overall	-0.57	0.02*	-8.50	0.01**	0.04	0.01**	0.07	0.00***

A



B



C

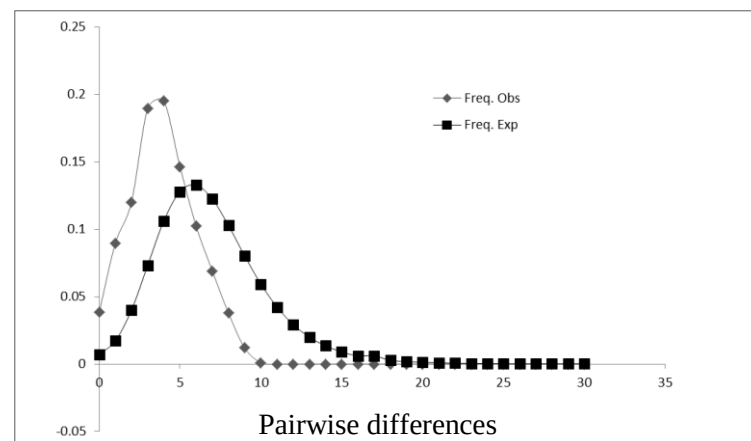


Fig. 3.5: Mismatch distribution curves of pairwise differences among individuals for the (A) ATPase 6, (B) cytochrome *b* and (C) nuclear S7 intron I gene regions for the overall population of *Cheilio inermis*. The observed curve of the number of pairwise differences among individuals is indicated with a black line. The expected curve under a model of population growth is indicated with a grey line.

The demographic parameters based on the mismatch analyses of the ATPase 6 and cytochrome *b* data are shown for each locality and the overall sample in Table 3.7. Both data sets yielded a Tau (τ) for the overall sample of 2.50, which led to estimated expansion times of 85 698 (ATPase 6) and 40 219 (cytochrome *b*) years ago. Although the cytochrome *b* gene data set showed an expansion time that is more recent than ATPase 6, both of these data sets suggests an expansion that has taken place during the Pleistocene. The expansion times based on the ATPase 6 gene 95% confidence interval (0.962–3.396) ranged from 32 976 to 116 412 years ago, while cytochrome *b* (95% confidence interval: 1.900-2.936) was dated between 30 556 to 47 232 years ago. The initial effective population size (N_{ef0}) of ATPase 6 was estimated at zero and the present effective population size (N_{ef1}) was estimated at 79 112. For the cytochrome *b* data set, N_{ef0} was zero and N_{ef1} was estimated at $N_{ef1} = 1\ 608\ 735$. Both of these data sets suggest population growth.

Table 3.7: Demographic parameters of τ , θ_0 , and θ_1 from the mismatch distributions of the (A) ATPase 6 and (B) cytochrome *b* datasets for *Cheilio inermis* were used to estimate the time of expansion and the effective female population sizes at time zero (N_{ef0}) and the present day (N_{ef1}).

A

Locality	τ	Time of expansion	θ_0	N_{ef0}	θ_1	N_{ef1}
Mozambique	2.63	90 154	0	0	15.74	539 559
Tanga	2.28	78 157	0	0	11.01	377 417
Zanzibar	2.81	96 325	0	0	99 999	3 427 910
Gazi	2.23	76 443	0	0	99 999	3 427 910
Shimoni	2.68	91 868	0	0	46.09	1 579 939
Overall	2.50	85 698	0	0	23.08	79 112

B

Locality	τ	Time of expansion	θ_0	N_{ef0}	θ_1	N_{ef1}
Mozambique	1.84	29 601	4.20	67 566	62.73	1 009 170
Tanga	2.30	37 001	0	0	11.01	117 124
Zanzibar	2.32	37 323	1.21	19 466	99 999	1 608 735
Gazi	1.76	28 314	0	0	99 999	1 608 735
Shimoni	3.51	56 467	0	0	15.21	244. 927
Nosy Be	4.26	58 532	0	0	99 999	1 608 735
Overall	2.50	40 219	0	0	99 999	1 608 735

Discussion

Both mitochondrial markers (ATPase 6 and cytochrome *b*) indicated high levels of genetic differentiation within the WIO. The nuclear S7 intron I revealed connectivity and little differentiation, involving only some localities, across the sampling area. The mitochondrial genes had fewer widespread haplotypes and these were mainly central in the networks, representing ancestral haplotypes and indicating historical connectivity within the WIO (Templeton *et al.*, 1995). There were many private haplotypes in the networks of the mitochondrial genes, indicating high diversity. For S7 intron I, the network showed shared alleles among sampling localities and few private alleles.

There was only one individual from the Philippines that was included in the present study. The one individual from Philippines did not amplify for the cytochrome *b* and S7 intron I gene markers. The Philippines (Pacific) sample was not differentiated from the WIO localities (Mozambique, Mauritius, Tanga, Zanzibar, Gazi and Shimoni) in the ATPase 6 network. More samples from the Pacific (including the Philippines) need to be added to confirmed connectivity or differentiation between the Indian and Pacific Oceans. Klanten *et al.* (2007) suggested a pattern of connectivity between the Indian Ocean and Pacific localities (including the Philippines) in the bignose unicornfish *Naso vlamingii* using the mitochondrial control region. The long pelagic larval phase and swimming abilities of *N. vlamingii* were suggested to promote genetic homogeneity between the two ocean basins (Klanten *et al.*, 2007). By contrast, other authors found differentiation between the WIO and Pacific (Gaither *et al.*, 2010; Muths *et al.*, 2011). Gaither *et al.* (2010) observed genetic differentiation in blacktail snappers *Lutjanus fulvus* between the Indian Ocean and the Pacific Ocean and suggested isolation by distance. Muths *et al.* (2011) showed genetic differentiation between the Western Pacific and WIO populations of the soldierfish *Mypristis berndti* and this was attributed to the Indo-Pacific Barrier.

There were low and high pairwise Φ_{ST} values that were observed for all markers. The ATPase 6 data revealed the differentiation of Mozambique from Tanzania (Tanga) and Kenya (Gazi). ATPase 6 also revealed the differentiation of Tanzania from Kenya. Cytochrome *b* data indicated the differentiation of Tanzania from Mozambique and Madagascar (Nosy Be), differentiation was also found between Madagascar and Kenya. Cytochrome *b* data revealed the differentiation of Tanga from Kenya (Shimoni and Gazi). Most of the pairwise Φ_{ST} values for S7 intron I were low and non-significant; this marker only indicated the differentiation of

Gazi from Shimoani and Zanzibar. The Analyses of Molecular Variance revealed different results for the mitochondrial fragments and the nuclear gene. The Analyses of Molecular Variance showed most of the variation was within localities for all gene fragments. The Analyses of Molecular Variance revealed high levels of genetic differentiation among defined regional groups for the two mitochondrial data sets, with 9% of all variation being found among three areas: Gazi and Shimoani (1), Tanga and Zanzibar (2) and Mozambique (3). In the case of cytochrome *b*, Nosy be (Madagascar) was grouped with Mozambique. However, the nuclear S7 intron I gene revealed a negative Φ_{CT} value among regions, which suggests a lack of spatial differentiation among regions. The nearest neighbour statistics, suggests little genetic differentiation among localities for all gene fragments.

The mitochondrial and nuclear markers showed different results for the present study because they have different levels of genetic variability and rates of evolution (Hoffman *et al.*, 2009). The mitochondrial markers offered sufficient resolution to detect genetic differentiation in *C. inermis* within the WIO. The nuclear marker was not variable enough to detect differentiation. Future studies need to use highly variable markers, such as microsatellites, to identify contemporary genetic structure (Hellberg, 2009). The use of microsatellites would add great confidence and reveal patterns of genetic differentiation, as seen in the study of *Lutjanus kasmira* (Muths *et al.*, 2012).

Marine fish species with a long larval duration are expected to show widespread dispersal through ocean currents (Ward, 2000). For example, the pelagic larval duration (PLD) of bluestripe snapper *Lutjanus kasmira* is around 30 days, which could permit the larvae to passively cover distances of several hundreds of kilometres in an area (WIO) where ocean currents are 20-30 Sverdrup (Muths *et al.*, 2012). The pelagic larval duration of *C. inermis* is 43 to 73 days (Victor, 1986) and the species was expected to show little to no genetic differentiation across the WIO. The WIO was not differentiated from the Philippines but that was only based on one sample, but the connectivity of *C. inermis* in the WIO was found to be more restricted than expected. The different life history stages of coral reef fishes, such as *C. inermis*, might have contributed to the genetic differentiation of this species. Adults of coral reef fishes are sedentary (Sale, 1980); therefore, genetic connectivity between distant populations is not expected (Kochuis & Blohm, 2005). Although larvae of reef fish are assumed to be dispersive, they can settle near the parental origin (Planes *et al.*, 1996), and that might be the case for *C. inermis*, this reduces genetic connectivity and promotes genetic

differentiation. The present study is in agreement with what was previously observed in the blotcheye soldierfish *Myripristis berndti*. Muths *et al.* (2011) considered differentiation of *M. berndti* across the WIO, based on mitochondrial cytochrome *b* and nuclear microsatellites. *Myripristi berndti* had a long PLD but high genetic differentiation was observed (Muths *et al.*, 2011). The genetic differentiation in the present study does not support the prediction of Riginos and Victor (2001) that marine fish species that have a long pelagic larval phase will show high dispersal abilities and panmixia over a large geographic range. Nevertheless, other studies have supported the prediction and provided evidence that PLD is a good predictor of genetic structure (Beldade *et al.*, 2009; Muths *et al.* 2012). Muths *et al.* (2012) suggested that the long pelagic phase of *L. kasmira* and the swimming capabilities of the fish have helped to maintain genetic connectivity across the WIO. Beldade *et al.* (2009) indicated that dispersal of the genus *Clepticus* across biogeographic barriers is aided by the coastal pelagic habitats and long larval phase (see Chapter 1).

The widespread *C. inermis* lives in patches of suitable habitat, such as coral reefs, lagoons, seagrass beds and algal-covered flats (Heemstra & Heemstra, 2004; Kuiter, 2010). Another possible explanation for the differentiation in the current study could be the function of distance between the patches of suitable habitat (Gaither *et al.*, 2010). Henriksson and Grahm (2013) used AFLPs to investigate genetic structure of the endemic spinefoot shoemaker *Siganus sutor* within the WIO. Most of the differentiation in the *S. sutor* was observed in the separation of the east coast Africa (Tanzania and Kenya) from the Mascarene Islands (Mauritius and Seychelles) and from the Comoros populations, and this was attributed to differences in habitat continuity (Henriksson & Grahm, 2013).

Oceanographic features are one of the major factors that are suggested to be responsible for genetic differentiation (Tsang *et al.*, 2012). The selected markers for *C. inermis* showed the genetic isolation of Kenya from Tanzania, and the isolation of Kenya and Tanzania from Mozambique and Madagascar. In the WIO, the South Equatorial Current (SEQC) splits into two ocean currents when reaching the African mainland, one of the ocean currents flows to the north forming the East African Coastal Current (EACC) along Tanzania and Kenya (Swallows *et al.*, 1988; Magori, 2008; see Chapter 1). The SEQC might have separated the individuals of Tanzania and Kenya from Mozambique and Madagascar. This study is consistent with the recent study of Visram *et al.* (2010) that revealed a pattern of isolation in the blue barred parrotfish *Scarus ghobban* within the WIO, and a separation of Kenya and Seychelles from

Tanzania and Mauritius. This genetic separation was attributed to the SEQC and EACC (Visram *et al.*, 2010). The EACC might have limited connectivity between Tanzania and Kenya. A similar pattern was observed in the study of the mud crab *Scylla serrata* by Fratini and Vannini (2002), which used sequence data of the mitochondrial DNA cytochrome oxidase subunit I (COI) gene, and found differentiation along the African east coast between Kenya and Tanzania which is attributed to local ocean currents and life history features (Fratini & Vannini, 2002). However, Fratini and Vannini (2002) were not specific about which ocean currents and life history characteristics were responsible for the patterns observed.

Genetic connectivity of the widespread *C. inermis* seems to be restricted between the northern and southern localities. The southward-flowing component of the SEQC forms the Comoros Gyre off the northern tip of Madagascar, (Lutjeharms, 2006; Obura, 2012) and continues to flow to the south as large cyclonic and anti-cyclonic eddies in the Mozambique Channel (Schouten *et al.*, 2002). Despite the southward-flowing current, the Comoros Gyre and eddies could limit the dispersal and exchange of individuals of the widespread *C. inermis*. Eddies in the Mozambique Channel were identified as oceanographic barriers to connectivity in the green turtle *Chelonia mydas* (Bourjea *et al.*, 2007) and blotcheye soldierfish *Mypristis berndti* (Muths *et al.*, 2011). The study on green turtle, based on the mitochondrial control region, revealed genetic differentiation, with populations of the north and south of the Mozambique Channel belong to separate genetic stocks, with the south Mozambique Channel population also being a mix of two separate stocks (Bourjea *et al.*, 2007). It is important to note that this differentiation in *C. inermis* might have been due to historical factors. The land bridges that interrupted the Mozambique Channel (Santini & Winterbottom, 2002) might have prevented the widespread *C. inermis* from crossing and exchanging genes with other populations, and caused differentiation within the WIO (see above).

Cryptic species are more successfully identified by using molecular genetic markers than morphology (Knowlton, 1993). The present study found differentiation in the widespread and monotypic *C. inermis*. The sequence divergence that was found between the north and south localities for the cytochrome *b* data is 0.4%. The widespread and monotypic goliath grouper *Epinephelus itajara* revealed a sequence divergences of 6.05 to 6.46% for cytochrome *b*, suggesting that the populations of *E. itajara* in the Pacific and western Atlantic represent two or more distinct species, and that these distinct species require separate management and conservation strategies (Craig *et al.*, 2008). The hawkfish *Cirrhitops fasciatus* is considered as

a single species with a disjunct population in the Mascarene Islands and the Hawaiian Islands, but the species was found to be highly divergent by 11% (based on cytochrome *b*) (Randall & Shultz, 2008). The population of hawkfish in the Mascarene Islands was described as a new species *Cirrhitops mascarenensis*. The divergence of 0.4% in the present study is far less than 6.05 to 6.46% and 11%. The differentiation that was found in the widespread *C. inermis* is not enough to suggest the existence of a cryptic species complex. Future studies should increase the sample size of *C. inermis*. The widespread and monotypic *C. inermis* (Forsskål 1775) was initially described in the Red Sea, it would be interesting to add samples from the Red Sea to see whether *C. inermis* would show a similar pattern to the crescent-tail wrasse *Thalassoma lunare* (see Chapter 4).

Both mitochondrial and nuclear sequence data analyses revealed high haplotype or allelic diversities and low nucleotide diversities. Grant and Bowen (1998) suggested that a fish species characterised by high haplotype and low nucleotide diversities would have experienced population expansion after a period of low effective population size. The unimodal mismatch distributions observed for the three gene fragments showed evidence of population expansion. The negative values of the neutrality tests were consistent with the hypothesis of a recent population expansion in *C. inermis*. The population expansion of *C. inermis* was estimated to have been initiated approximately 85 698 years ago for ATPase 6 and 40 219 years ago for cytochrome *b*, and this expansion appears to have taken place in the late Pleistocene. From 110 000 to 75 000 years ago, the sea level fluctuated between 0 and -75m, and from 50 000 to 35 000 years ago, sea level fluctuated between -75 and -110m (Toms *et al.*, 2014). The mitochondrial gene fragments have showed different times of expansion and overlapping confidence interval dating for *C. inermis*. The different expansion times could be due to the different levels of variability, resolution and coalescence times of the mitochondrial gene fragments (ATPase 6 and cytochrome *b*). The mutation rates might not be realistic enough for the markers and species in question. After the period of low effective population size, the population growth of *C. inermis* might be the result of post-glacial sea level rise (Hewitt, 2000). The sea level rise might have increased suitable habitats for marine species (Bellwood & Wainwright, 2002; Gaither *et al.*, 2011). Tolley *et al.* (2005) observed a unimodal mismatch distribution and negative neutrality test values that indicated a recent population expansion for the spiny lobster *Palinurus gilchristi* in the southwestern Indian Ocean (SWIO). They suggested that the expansion of *P. gilchristi* was a consequence of sea level rise, which caused the continental shelf of South Africa to become flooded to provide new and larger suitable

habitats for this species (Tolley *et al.*, 2005). Bowen *et al.* (2006) observed a unimodal mismatch distribution that indicated recent population expansion during the late Pleistocene in the Atlantic squirrelfish *Holocentrus ascensionis* and the possible explanation of that expansion was an increase of suitable habitat when the sea level rose (Bowen *et al.*, 2006). This explanation might have been the same case for the widespread *C. inermis* within the WIO.

CHAPTER FOUR

GENETIC DIFFERENTIATION OF *THALASSOMA* SPECIES WITHIN THE WESTER INDIAN OCEAN

Introduction

The origins of the Western Indian Ocean (WIO) fauna have been related to biogeographic hypotheses that are dependent on vicariance (Santini & Winterbottom, 2002; see Chapter 1). However, the specific biogeographic factors that might have influenced the regional genetic differentiation of reef fish species, such as the goldbar wrasse *Thalassoma hebraicum* (Lacepède 1801) and the crescent-tail wrasse *T. lunare* (Linnaeus 1758), in the WIO are unknown. These two *Thalassoma* species (*T. hebraicum* and *T. lunare*) were included in the present study to investigate genetic differentiation and connectivity across the identified biogeographical regions of the WIO. *Thalassoma hebraicum* is an endemic WIO species and has a narrow distribution, ranging from Kenya to Algoa Bay in South Africa (Figure 4.1; Heemstra & Heemstra, 2004). *Thalassoma lunare* (Figure 4.2), on the other hand, has a much broader geographic range, occurring throughout the tropical zones of the Indo-Central Pacific (Figure 4.3; Kuiter, 2010). In the WIO region, the species occurs from Oman, through to the Red Sea and to Algoa Bay in South Africa (Figure 4.3; Heemstra & Heemstra, 2004). *Thalassoma lunare* shows almost no colour variation over its extensive geographical range (Kuiter, 2010).

Population genetic structuring of reef fishes may derive from habitat preferences, life history of the species and oceanographic features (Tsang *et al.*, 2012; Knutsen *et al.*, 2014). The two *Thalassoma* species are both strongly associated with reefs, occupying the same distinct patches of suitable habitat as the cigar wrasse *Cheilio inermis* (see Chapter 3). Adults of *T. hebraicum* reside on coral and rocky reefs at 1-30 m depth and the juveniles are found in rock and tide-pools (Heemstra & Heemstra, 2004). *Thalassoma lunare* lives along coral and rocky reefs, outer reefs and shallow lagoons, at depths of 1-30 m (Heemstra & Heemstra, 2004), and has been reported to enter estuaries (Kuiter, 2010). As discussed previously, the patchy and isolated distributions of suitable habitat for reef species limit dispersal and influence the genetic structure of fish species (Muths *et al.*, 2011). Genetic differentiation in the corkwing wrasse

Symphodus melops has been caused by the distance between patches of suitable habitat (Knutsen *et al.*, 2013).

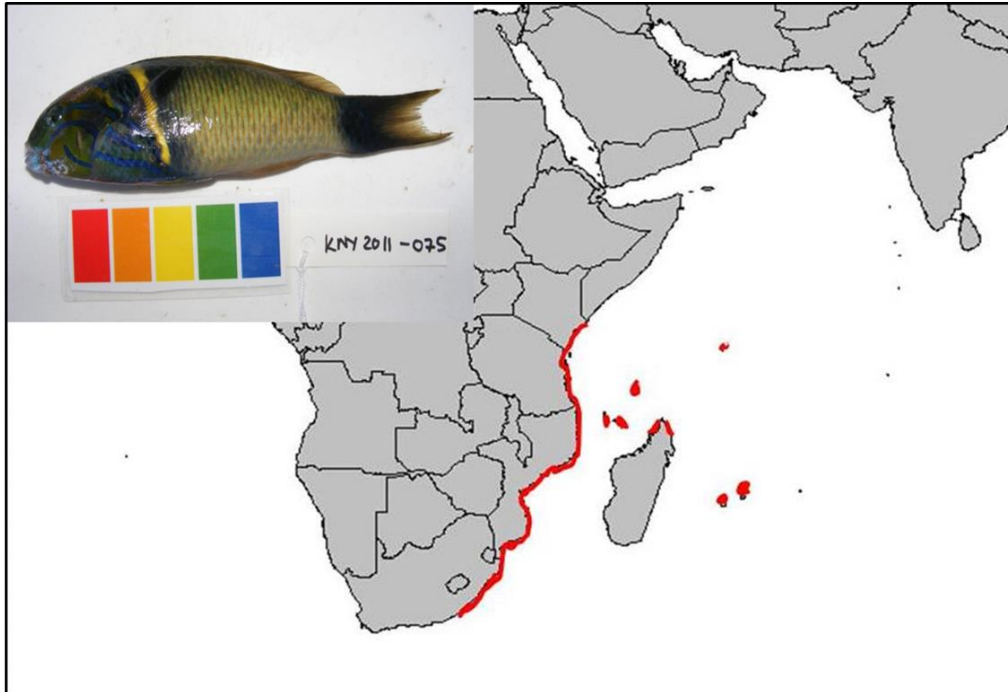


Fig. 4.1: Distribution of *Thalassoma hebraicum* in the Western Indian Ocean (modified from <http://www.fishbase.org>). Inset image: adult/terminal male phase *Thalassoma hebraicum*, the gold bar wrasse. Photographed specimen: SAIAB 197032, 133.8 mm standard length, Shimoni landing site, Kenya (photo: SAIAB).



Fig. 4.2: Adult (terminal male) *Thalassoma lunare*, the crescent-tail wrasse. Photographed specimen: SAIAB 87107, 160 mm standard length, Zanzibar, Tanzania (photo: SAIAB)

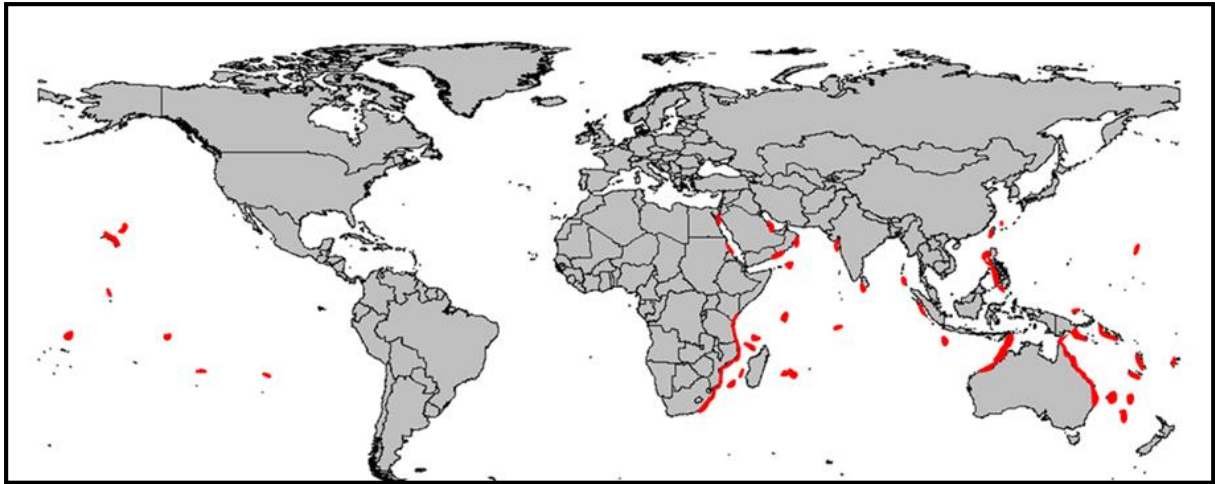


Fig 4.3: Distribution of *Thalassoma lunare* in the Indo-Pacific, including the Western Indian Ocean (modified from <http://www.fishbase.org>)

There is, however, very little information on the biology of both species. As discussed in Chapter 3, the general assumption is that these species, as coral reef fish, may have a biphasic life cycle in which the adults are less mobile (Salas *et al.*, 2010) and dispersal is mainly accomplished during the pelagic larval phase (Riginos & Victor, 2001). Both of these *Thalassoma* species spawn on outgoing tides to facilitate the transport of eggs away from the spawning site to maximise dispersal (Bardach *et al.*, 1980; Thresher, 1984). Although the larval duration of *T. hebraicum* is not documented, *T. lunare* has a larval phase that lasts from 39 to 55 days (Victor, 1986). These species, therefore, have the potential to disperse widely among reefs during the pelagic larval phase and to connect different populations (see Chapter 3). However, the larvae of reef fish may not disperse with ocean currents, but can return to the spawning grounds (Dorenbosch *et al.*, 2006; Chapter 3). This restricts dispersal and allows genetic differentiation to occur (see Chapter 3).

Marine species that have a pelagic larval phase have generally been considered to be panmictic or genetically homogenous over a wide geographic range (Henriksson *et al.*, 2012). However, recent studies have observed genetic differentiation even among marine species that have a pelagic larval phase (Bernardi *et al.*, 2001; Bay *et al.*, 2004; Froukh & Kochzius, 2007). The fourline wrasse *Larabicus quadrilineatus* is an endemic fish in the Red Sea and Gulf of Aden, and revealed genetic differentiation between the northern and southern Red Sea, with the

species found to be comprised of separate genetic stocks, and the separation was attributed to turbidity and reef development (Froukh & Kochzius, 2007).

The oceanographic features of the WIO seem to be the most influential parameter of larval movement during the pelagic phase (Masterson *et al.*, 1997). Chiswell and Booth (1999) found that eddies can return the phyllosoma larvae of the rock lobster *Jasus edwardsii* to their coastal habitats after being dispersed far offshore. Similarly, the high dispersal capability of some marine reef fish is brought about by ocean currents (Craig *et al.*, 2007). For example, the East African Coastal Current (EACC) and South Equatorial Current (SEQC) were found to maintain genetic homogeneity of the Dory snapper *Lutjanus fulviflamma* (Dorenbosch *et al.*, 2006).

The expectation is to find similar patterns between the two *Thalassoma* species within the WIO. The two species (*T. hebraicum* and *T. lunare*) have similar life history features and are exposed to similar environmental and oceanographic features in the WIO. The two species are expected to show regional differentiation based on known biogeographic boundaries and oceanographic circulation patterns of the WIO. *Thalassoma hebraicum* may have different genetic compositions, produced by the disjunct habitats, the sedentary adults and the returning of the larvae to the parental origin as described in Chapter 3. On the other hand, the pelagic larvae of *T. hebraicum* might drift passively with ocean currents, resulting in high levels of genetic connectivity across the WIO. As *T. lunare* has a long larval phase, it would be reasonable to find great dispersal and little to no genetic differentiation across the WIO (Grant & Bowen, 1998). *Thalassoma lunare* is expected to show genetic differentiation between the WIO and Red Sea, because the species might not be able to cross physical barrier of Bab al Mandab and upwellings that are influenced by the monsoon system in the WIO (Di Battista *et al.*, 2013). The ecological characteristics (e.g. temperature or turbidity) of the Red Sea (Roberts *et al.*, 1992) might also play a role in restricting the dispersal of *T. lunare* (see above).

The specific aims of the present study were

1. To investigate genetic differentiation among the different localities of the WIO, the Red Sea and certain islands of the Indo-West Pacific within the two *Thalassoma* species, using a molecular genetic approach.
2. To determine the factors that might be responsible for maintaining genetic homogeneity or influencing the distribution of genetic variation in the two *Thalassoma* species.

Materials and methods

For an outline of the overall materials and methods refer to Chapter 2.

Sampling

Samples of the two *Thalassoma* species (*T. hebraicum* and *T. lunare*) were collected from numerous localities using different approaches as discussed in Chapter 2. Some of the WIO and the Indo-West Pacific islands sequences were downloaded from GenBank.

DNA amplification

DNA extractions, Polymerase Chain Reactions (PCRs), purification and sequencing were described and discussed in Chapter 2. The thermocycling regimes for the amplification of each gene fragment in each of the *Thalassoma* species are presented in Table 4.1. Successful PCRs were confirmed by electrophoresis in 1% agarose gels and sequencing in the forward direction was done by Macrogen (South Korea) (see Chapter 2).

Table 4.1: Polymerase Chain Reaction cycle profiles for the generation of the two mitochondrial (ATPase 6 and cytochrome *b*) and nuclear S7 intron I datasets from each of the *Thalassoma* species (*T. hebraicum* and *T. lunare*).

Gene region	PCR profile cycle					
	Stage 1		Stage 2		Stage 3	
	Initial denaturation	Denaturation	Annealing	Extension	Cycles	Final extension
ATPase 6	94 °C, 3 min	94 °C, 45 s	48°C, 1 min	72 °C, 1 min	35	72 °C, 10 min
Cytochrome <i>b</i>	94 °C, 4 min	94 °C, 45 s	56 °C, 45 s	72 °C, 1 min	35	72 °C, 10 min
S7 intron I	94 °C, 3 min	95 °C, 45 s	56 °C, 1 min	72 °C, 2 min	30	72 °C, 10 min

Genetic data analysis

The genetic diversity indices were calculated for each locality and for the overall sample of each species, for the respective gene regions (see Chapter 2). Median-joining networks were constructed separately for the two species for the selected markers. The sample size of *T. lunare* was very small and this made it unfeasible to draw a median-joining network for the cytochrome *b* data for this species. The cytochrome *b* data of the two species were combined in the median-joining network. Neighbour-joining trees were constructed in MEGA 6.6 (Tamura *et al.*, 2013; see Chapter 2) for the two species together to examine phylogenetic relationships among ATPase 6 haplotypes and S7 intron I alleles, respectively.

MEGA version 6.6 (Tamura *et al.*, 2013) was used to choose the most suitable model of nucleotide substitution for each dataset (see Chapter 2). Pairwise population comparisons and nearest-neighbour statistic (S_{nn} : Hudson, 2000) were determined for the selected gene regions for each species as described in Chapters 2 and 3. For *T. hebraicum*, Analyses of Molecular Variance (AMOVA) were performed for the three gene regions by grouping the localities as discussed in Chapter 2. The arrangement that described the variation best maximised genetic divergences between defined groups. For *T. lunare*, AMOVAs were not done due to the small sample size.

For *T. hebraicum*, neutrality tests, mismatch distributions, mismatch statistics, and demographic parameters were examined for each sampling locality and the overall sample. The time of a possible population expansion and effective population sizes were calculated for *T. hebraicum* for the two mitochondrial gene fragments. The mutation rates and generation time used for *T. hebraicum* were discussed in Chapter 2 and 3. The data set for *T. lunare* had inadequate samples to allow the investigation of neutrality tests, mismatch distributions and statistics, or demographic parameters.

Results

Specimens of the two *Thalassoma* species were collected from numerous localities across the WIO and the Red Sea (Table 4.2). The localities that had single individuals were combined with the nearest locality. Pemba, Mozambique was combined with the three South Africa localities (see below) and these were denoted as southern Africa. Some *T. hebraicum* and *T. lunare* sequences were downloaded from GenBank for the cytochrome *b* data set. For *T. hebraicum*, two of those sequences were from Nosy Be-Antsiranana (Madagascar) and their accession numbers were JF458280 and JF458281. One sequences from Zanzibar (Tanzania) was also downloaded (accession number AY328869). For *T. lunare*, two S7 intron I sequences from Pelorus Island were downloaded and the accession numbers were DQ443836 and DQ443837. A sequence (AY329656) from an individual from French Polynesia was also included in the S7 intron I data set.

Table 4.2: Localities, locality codes and sample sizes of *Thalassoma hebraicum* and *T. lunare* collected from various Western Indian Ocean localities, the Red Sea and Indo-Pacific Islands.

Sampling localities	Code	<i>T. hebraicum</i>	<i>T. lunare</i>	GPS coordinates	
				Latitude	Longitude
Pelorus Island	PI	-	2	-18.5618	146.5051
French Polynesia	FP	-	1	-17.6249	149.2461
Farasan, southern Red Sea	RD	-	15	-17.1312	42.0352
Kafu Atoll, Maldives	MD		4	-4.9625	73.4612
Shimoni, Kenya	SM	19	-	-4.6478	39.3807
Tanga, Tanzania	TN		2	-5.1000	39.1000
Zanzibar, Tanzania	ZN	9	5	-6.1667	39.1833
Pemba, Mozambique	MZ	1	-	-12.9602	40.9702
Cape Vidal, South Africa	SA	3	-	-28.1501	32.6071
Aliwal Shoal, South Africa	SA	2	-	-30.2500	30.8167
Ballito, South Africa	SA	3	-	-29.5421	31.2305
Nosy Be Antsiranana, Madagascar	NB	2	-	-13.4546	48.1525
Mahè, Seychelles	MH	2	2	-4.6167	55.4500
Aldabra Atoll, Seychelles	AD	4	3	-9.4047	46.2094
Overall		45	34		

Genetic diversity analysis

The edited ATPase 6 sequences for the 27 individuals of *T. hebraicum* were 819 base pairs (bp) long and yielded 18 unique haplotypes (Table 4.3A). The average number of pairwise differences (k) ranged from 2.56 to 7.00 for each locality and the number of polymorphic sites (S) within the localities ranged from 4 to 13. Cytochrome *b* sequences were included for 34 individuals of *T. hebraicum* and were 472 nucleotides in length. The value of k was estimated to vary from 1.40 to 3.80 per locality and the data yielded a number of polymorphic sites (S) ranging from 3 to 10 for each sampling locality. The phased S7 intron I data yielded an alignment of 312 bp from 16 individuals (Table 4.3B). Once phased the number of sequences doubled (to 32) and there were 26 different alleles. The values of k ranged from 2.83 to 6.24 and the number of polymorphic sites (S) ranged from 3 to 24 per locality.

Table 4.3: Sample sizes (N) and genetic diversity indices derived from the mitochondrial ATPase 6 and cytochrome b (A) and nuclear S7 intron I (B) datasets for *Thalassoma hebraicum* for each locality and the overall sample. Indices included the number of haplotypes (N_h) or alleles (N_A), number of private haplotypes (N_{PH}) or private alleles (N_{PA}), haplotype diversity (h) or allelic diversity (A), nucleotide diversity (π), number of polymorphic sites (S) and average pairwise differences (k) among individuals. Standard deviations are presented for haplotype or allelic and nucleotide diversities.

A

Locality	ATPase 6 (819 bp)							Cytochrome b (472 bp)						
	N	N_h	N_{PH}	$h \pm SD$	$\pi \pm SD$	S	k	N	N_h	N_{PH}	$h \pm SD$	$\pi \pm SD$	S	k
Southern Africa	5	4	-	0.90 ± 0.16	0.004 ± 0.0001	8	3.40	5	3	1	0.70 ± 0.21	0.003 ± 0.0012	3	1.40
Zanzibar	3	3	2	1.00 ± 0.27	0.003 ± 0.0012	4	2.67	6	2	3	0.93 ± 0.31	0.004 ± 0.0003	10	3.80
Shimoni	13	10	7	0.95 ± 0.05	0.003 ± 0.0005	13	2.56	16	8	3	0.91 ± 0.12	0.008 ± 0.0026	8	2.02
Nosy Be	-	-	-	-	-	-	-	2	1	-	-	-	-	-
Mahè	2	2	2	1.00 ± 0.50	0.008 ± 0.0042	7	7.00	1	-	-	-	-	-	-
Aldabra Atoll	4	4	2	1.00 ± 0.17	0.005 ± 0.0015	9	4.50	4	4	2	0.90 ± 0.16	0.004 ± 0.0010	4	1.80
Overall	27	18	13	0.94 ± 0.03	0.004 ± 0.0006	28	3.28	34	14	9	0.88 ± 0.04	0.004 ± 0.0008	18	2.14

B

Locality	S7 intron I (312 bp)						
	N	N_A	N_{PA}	$A \pm SD$	$\pi \pm SD$	S	k
Southern Africa	4	7	-5	0.96 ± 0.07	0.017 ± 0.0023	15	5.50
Zanzibar	1	1	-	-	-	-	-
Shimoni	9	16	14	0.98 ± 0.02	0.020 ± 0.0021	24	6.24
Mahè	1	1	1				
Aldabra Atoll	1	2	-	0.83 ± 0.22	0.009 ± 0.0024	3	2.83
Overall	16	26	20	0.98 ± 0.01	0.002 ± 0.0017	29	5.76

The 18 individuals of *T. lunare* produced an ATPase 6 alignment of 819 bp, which represented ten unique haplotypes (Table 4.4). The values of k ranged from 0.67 to 4.00 for each locality, and the number of segregating sites (S) within the localities ranged from 1 to 14. The phased S7 intron I data yielded an alignment of 312 bp from 10 individuals. Once phased the number of sequences doubled (to 20) and there were 14 unique alleles. The values of k was estimated to vary from 1.00 to 5.05 per locality and the data yielded a number of polymorphic sites (S) ranging from 2 to 12 for each sampling locality. Overall haplotype (h) or allelic diversities (A) of *T. hebraicum* were high for all markers: ATPase 6 ($h = 0.94 \pm 0.03$), cytochrome *b* ($h = 0.88 \pm 0.04$) and ($A = 0.98 \pm 0.01$; Table 4.3). Nucleotide diversities (π) for the overall sample were low for the ATPase 6 ($\pi = 0.004 \pm 0.0006$), cytochrome *b* ($\pi = 0.004 \pm 0.0008$) and S7 intron I ($\pi = 0.002 \pm 0.0017$) gene fragments. The ATPase 6 data set of *T. lunare* yielded a high haplotype diversity ($h = 0.89 \pm 0.05$) and low nucleotide diversity ($\pi = 0.008 \pm 0.0008$) for the overall sample (Table 4.4). Overall allelic and nucleotide diversities of nuclear S7 intron I were high at $A = 0.95 \pm 0.03$ and $\pi = 0.014 \pm 0.0001$, respectively. For *T. lunare*, genetic diversity indices were not calculated for cytochrome *b* data because there was only one haplotype (see below).

Table 4.4: Sample sizes (N) and genetic diversity indices for sampling localities in the Western Indian Ocean localities, the Red Sea and the Indo-West Pacific Islands and the overall sample of *Thalassoma lunare* for the two gene regions, ATPase 6 and S7 intron I. Indices of genetic diversities included the number of haplotypes (N_h) and alleles (N_A), the number of private haplotypes (N_{PH}) and alleles (N_{PA}), haplotype diversity (h) and allelic diversity (A), nucleotide diversity (π), the number of polymorphic sites (S) and the average pairwise differences among samples (k). Standard deviations are presented for haplotype, allelic and nucleotide diversities.

Locality	ATPase 6 (819 bp)							S7 intron I (312 bp)						
	N	N_h	N_{PH}	$h \pm SD$	$\pi \pm SD$	S	k	N	N_h	N_{PH}	$A \pm SD$	$\pi \pm SD$	S	k
Southern Africa	3	1	-	1.00 ± 0.01	0.001 ± 0.0001	4	1.17	-	-	-	-	-	-	-
Zanzibar	3	2	-	0.67 ± 0.31	0.001 ± 0.0001	1	0.67	-	-	-	-	-	-	-
Shimoni	-	-	-	-	-	-	-	1	1	-	0.95 ± 0.55	0.001 ± 0.0023	2	1.00
Mahè	1	1	1	-	-	-	-	-	-	-	-	-	-	-
Aldabra Atoll	1	1	1	-	-	-	-	-	-	-	-	-	-	-
Kafu Atoll	3	3	3	1.00 ± 0.27	0.003 ± 0.0012	4	2.67	1	2	1	1.00 ± 0.50	0.006 ± 0.0032	2	2.00
Red Sea	7	3	1	0.52 ± 0.21	0.005 ± 0.0031	14	4.00	5	8	8	0.97 ± 0.06	0.016 ± 0.0024	12	5.05
French														
Polynesia	-	-	-	-	-	-	-	1	1	1	0.98 ± 0.10	0.002 ± 0.0041	2	1.00
Pelorus Island	-	-	-	-	-	-	-	2	2	1	0.50 ± 0.26	0.003 ± 0.0017	2	1.00
Overall	18	10	6	0.89 ± 0.05	0.008 ± 0.0008	23	6.66	10	14	10	0.95 ± 0.03	0.014 ± 0.0001	16	4.66

Relationships among haplotypes and alleles

Neighbour-joining trees (NJ) were constructed to examine relationships among all included individuals of the two *Thalassoma* species using the ATPase 6 and nuclear S7 intron I gene fragments, respectively. The estimated-best fitting model of nucleotide substitution for the ATPase 6 data set was a Tamura and Nei (1993) model. This model was included in the construction of the neighbour-joining tree. The NJ trees from the ATPase 6 identified two distinct lineages, corresponding to *T. hebraicum* and *T. lunare*, respectively (Figure 4.4). The two lineages were separated by a mean sequence divergence of 15.9%. *Thalassoma hebraicum* revealed (27 individuals, 819 bp) a widespread lineage comprised of WIO individuals. *Thalassoma lunare* (18 individuals, 819 bp) revealed two distinct lineages (i.e. WIO and Red Sea lineages) which were separated by a mean sequence divergence of 1.2%. The first lineage contained samples from the Red Sea. The second lineage was widespread, including samples from the WIO and one individual from the Red Sea

A Jukes and Cantor (1969) model was identified as the most appropriate model for the S7 intron I data. The NJ tree of the phased S7 intron I data revealed two distinct lineages corresponding to the two species (Figure 4.5). The two lineages were separated by a mean sequence divergence of 6.3%. *Thalassoma hebraicum* (16 individuals, 312 bp) showed a widespread lineage with alleles occurring at all the sampling localities across the WIO. *Thalassoma lunare* (10 individuals, 312 bp) revealed two distinct lineages, which were separated by a mean sequence divergence of 0.4%. The first lineage contained only Red Sea alleles. The second lineage contained alleles from the WIO and the West Pacific Islands (French Polynesia and Pelorus Island), with some of the alleles were found in the Red Sea.

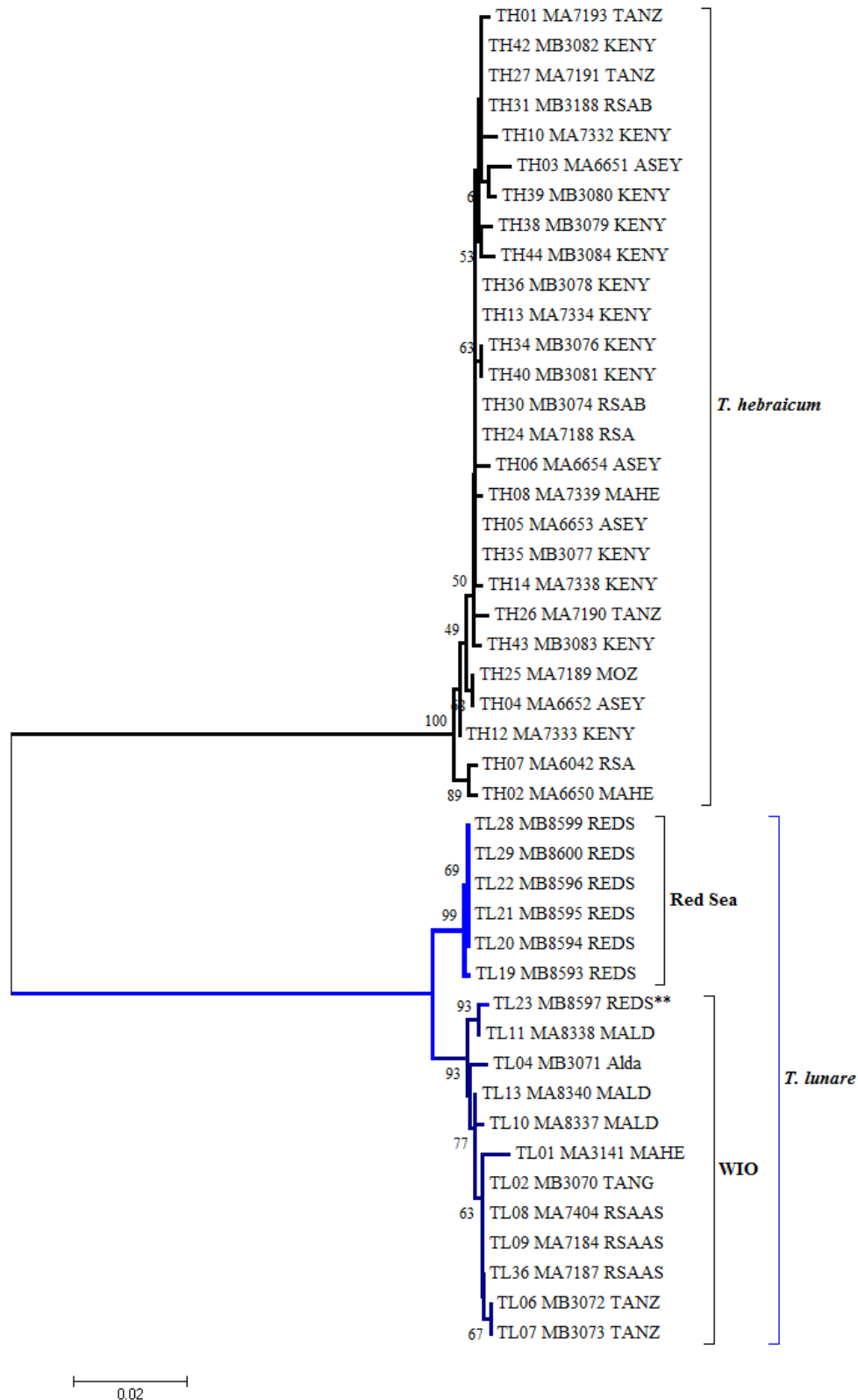


Fig. 4.4: Neighbour-joining tree derived from ATPase 6 sequence data for the two *Thalassoma* species (*T. hebraicum* and *T. lunare*) depicting relationships among the WIO and the Red Sea haplotypes. Nodal support (bootstrap percentage %) is shown on the branches. The scale bar indicates the estimated evolutionary distance.

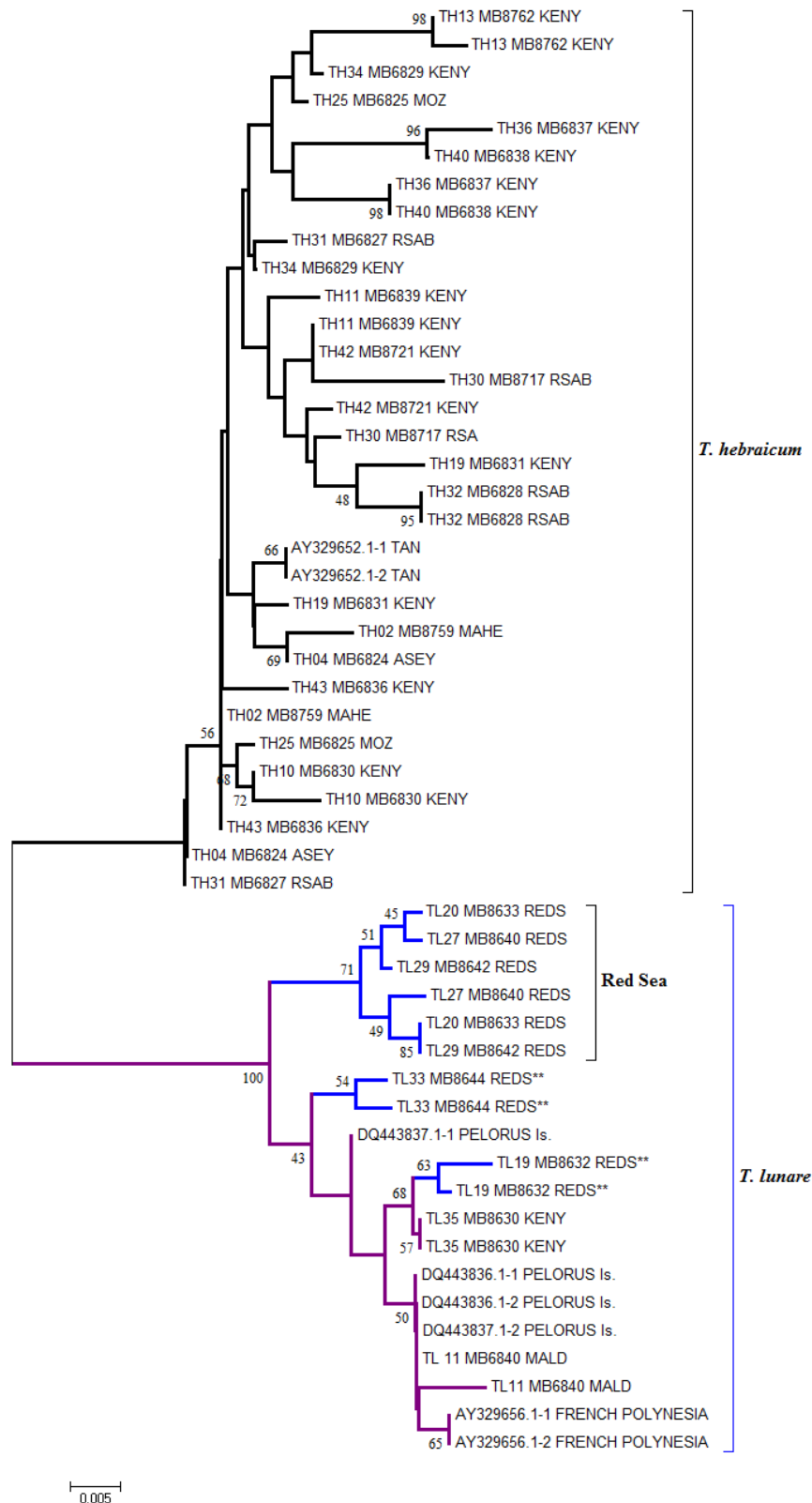


Fig. 4.5: Neighbour-joining tree derived from S7 intron I sequence data for the two *Thalassoma* species (*T. hebraicum* and *T. lunare*) depicting relationships among the WIO, Red Sea and Pacific alleles. Nodal support (bootstrap percentage %) is shown on the branches. The scale bar indicates the estimated evolutionary distance.

The median-joining network for the cytochrome *b* data set constructed for the two *Thalassoma* species is presented in Figure 4.6. There was only one cytochrome *b* haplotype among the *T. lunare* samples which was shared with *T. hebraicum*. There was no separation of the two species in the network. This network was not in agreement with the NJ trees of ATPase 6 and S7 intron I.

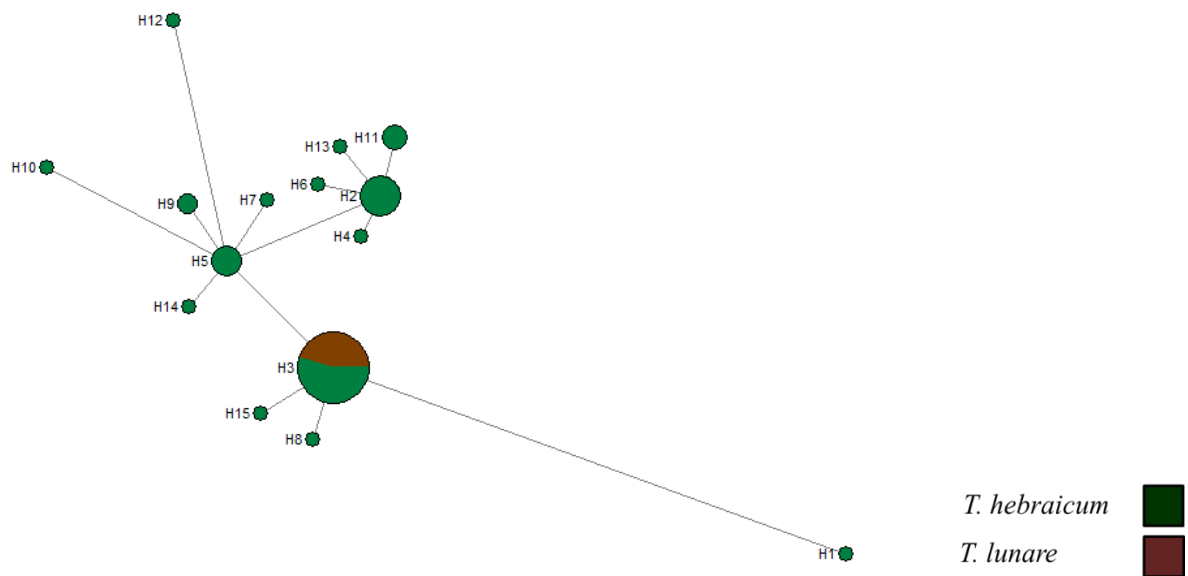


Fig 4.6: Median-joining network depicting genealogical relationships of the mitochondrial cytochrome *b* haplotypes for the two *Thalassoma* species (*Thalassoma hebraicum* and *Thalassoma lunare*). Pie charts are coloured according to the species identification of the individuals.

The median joining networks of the ATPase 6 data for the two *Thalassoma* species are depicted in Figure 4.7. ATPase 6 data had 18 unique haplotypes with three of those haplotypes being shared among localities. Only the two central haplotypes (H3 and H4) were widespread among localities. Their central position indicates the persistence of ancestral haplotypes (Posada & Crandall, 2001). Haplotype H1 was shared between southern Africa and Aldabra Atoll. Some of the southern Africa, Mahè and Aldabra Atoll (Seychelles) haplotypes were divergent (4 mutation steps) from the main network and were also divergent from other haplotypes from the same locality. Haplotype H14 from Shimoni was shared among individuals within the same locality. The remaining haplotypes were each restricted to a specific locality.

The *T. lunare* individuals were from South Africa (Mozambique was not included), Zanzibar, Mahè, Aldabra Atoll, Kafu Atoll (Maldives) and the Red Sea (Figure 4.7B). The ATPase 6 network showed two distinct and highly divergent haplogroups. The WIO and the Red Sea haplogroups were separated by nine mutational steps from each other. Mahè (Seychelles) was highly divergent (four mutational steps) from the rest of the WIO haplotypes. There was one haplotype of the WIO haplogroup present in the Red Sea. Within the WIO, there were only two shared haplotypes H9 and H10, of which H9 was shared between South Africa and Zanzibar. Haplotype H10 from Zanzibar was shared among individuals from the same locality.

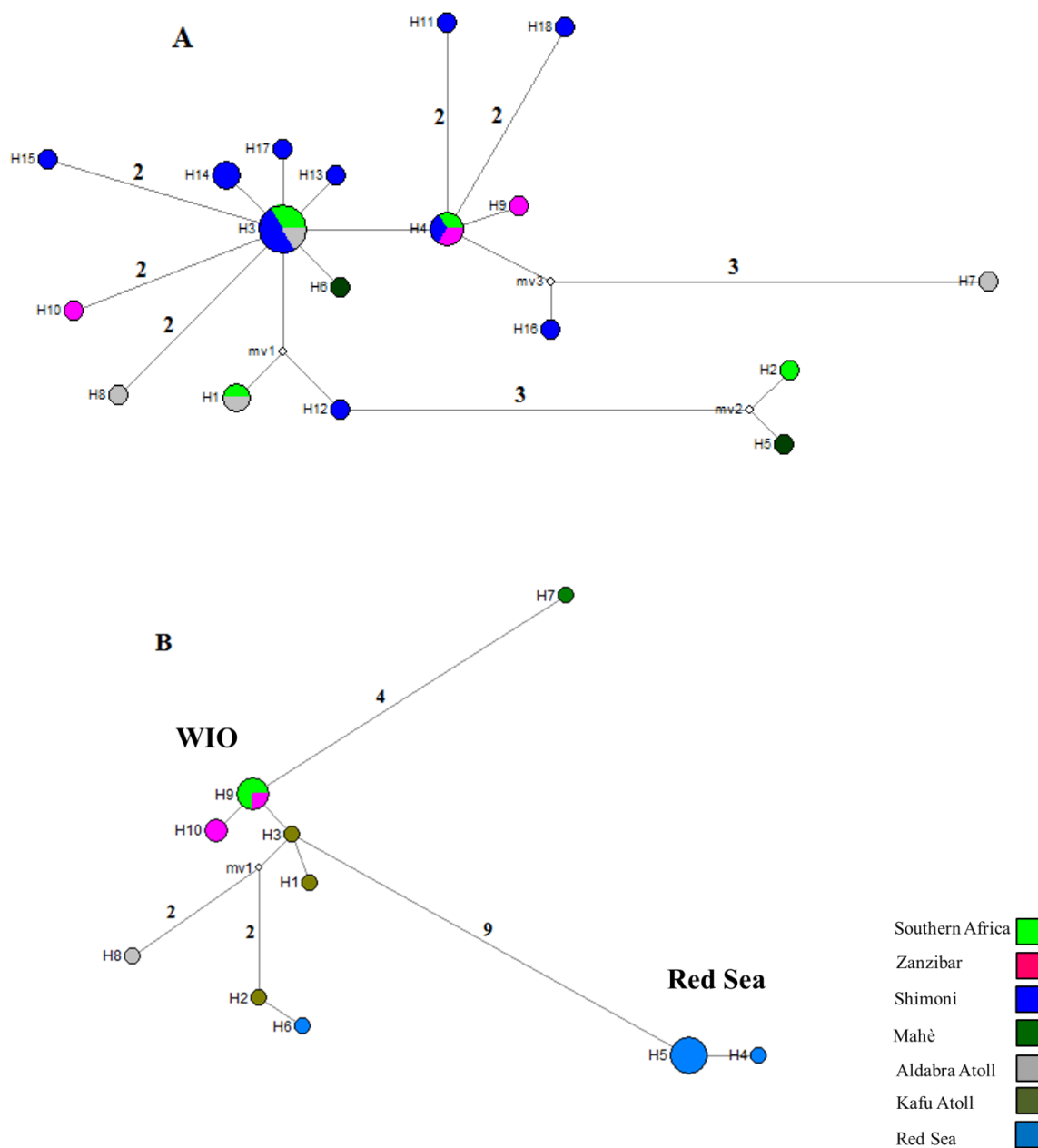


Fig. 4.7: Median-joining networks of genealogical relationships among the ATPase 6 haplotypes of (A) *Thalassoma hebraicum* and (B) *T. lunare* from several Western Indian Ocean localities and the Red Sea. The size of the circles shows the frequency of each haplotype. The pie charts are shaded with a colour that represents the geographic origin of the haplotype. Connecting lines between the circles represent one mutational step unless indicated otherwise. Missing haplotypes are indicated by small circles

The most appropriate model of nucleotide evolution for the mitochondrial ATPase 6 gene was the Tamura and Nei (1993) model for the two *Thalassoma* species. None of the pairwise (Φ_{ST}) values between the localities of the WIO (Table 4.5A) for *T. hebraicum* were significant (all $P > 0.05$). The data revealed Φ_{ST} values that were mostly less than 0.05 which indicate little differentiation (Wright 1978; Ward & Grewe, 1994). The data suggested great but non-significant ($P > 0.05$) differentiation of Mahè (Seychelles) from Zanzibar ($\Phi_{ST} = 0.17$) and from Shimoni ($\Phi_{ST} = 0.25$).

For *T. lunare*, the Red Sea was differentiated from the WIO localities, yielding Φ_{ST} values ranging from 0.61 to 0.70 (Table 4.5B). Only the comparisons between the individuals from Kafu Atoll (Maldives) and the Red Sea revealed significant ($P < 0.05$) genetic differentiation ($\Phi_{ST} = 0.62$). Great differentiation ($\Phi_{ST} = 0.25$ to $\Phi_{ST} = 0.50$) was observed in the WIO localities. The two Seychelles localities (Mahè and Aldabra Atoll) were combined together because they consisted of one individual. Little differentiation was only observed between Maldives and Seychelles ($\Phi_{ST} = 0.05$). The nearest neighbour statistic for *T. hebraicum* yielded $S_{nn} = 0.28$, and *T. lunare* yielded S_{nn} of 0.78 for the overall sample. For *T. hebraicum*, this indicates panmixia for the overall sample because the observed S_{nn} value was lower than 0.5. The nearest neighbour statistic of *T. lunare* indicated spatial genetic structure because the observed value is closer to one.

Table 4.5: Pairwise Φ_{ST} values between WIO localities for (A) *Thalassoma hebraicum* and (B) between WIO localities and the Red Sea for *T. lunare* based the analyses of ATPase 6 sequence data. Significant estimates ($P < 0.05$) are indicated in bold.

A

	Southern Africa	Zanzibar	Mahè	Aldabra Atoll
Southern Africa	-			
Zanzibar	0.03	-		
Mahè	-0.12	0.17	-	
Aldabra Atoll	-0.07	-0.03	0.03	-
Shimoni	0.02	0.02	0.25	0.02

B

	Southern Africa	Zanzibar	Seychelles	Maldives
Southern Africa	-			
Zanzibar	0.50	-		
Seychelles	0.25	0.27	-	
Maldives	0.43	0.44	0.05	-
Red Sea	0.70	0.70	0.61	0.62

The cytochrome *b* network for *T. hebraicum* is depicted in Figure 4.8. Two widespread haplotypes (H2 and H5) were central in the network. These central haplotypes may indicate common ancestral haplotypes in the species (Templeton *et al.*, 1995). Haplotype H3 was the most widespread and common. Haplotype H10 was shared between Shimoni and Zanzibar. Haplotype H13 from Shimoni was shared among individuals within the same locality. Haplotype H1 from Zanzibar (AY328869) was highly divergent and five mutational steps from the next closest haplotype in the network. This data set had singletons which were represented by one individual.

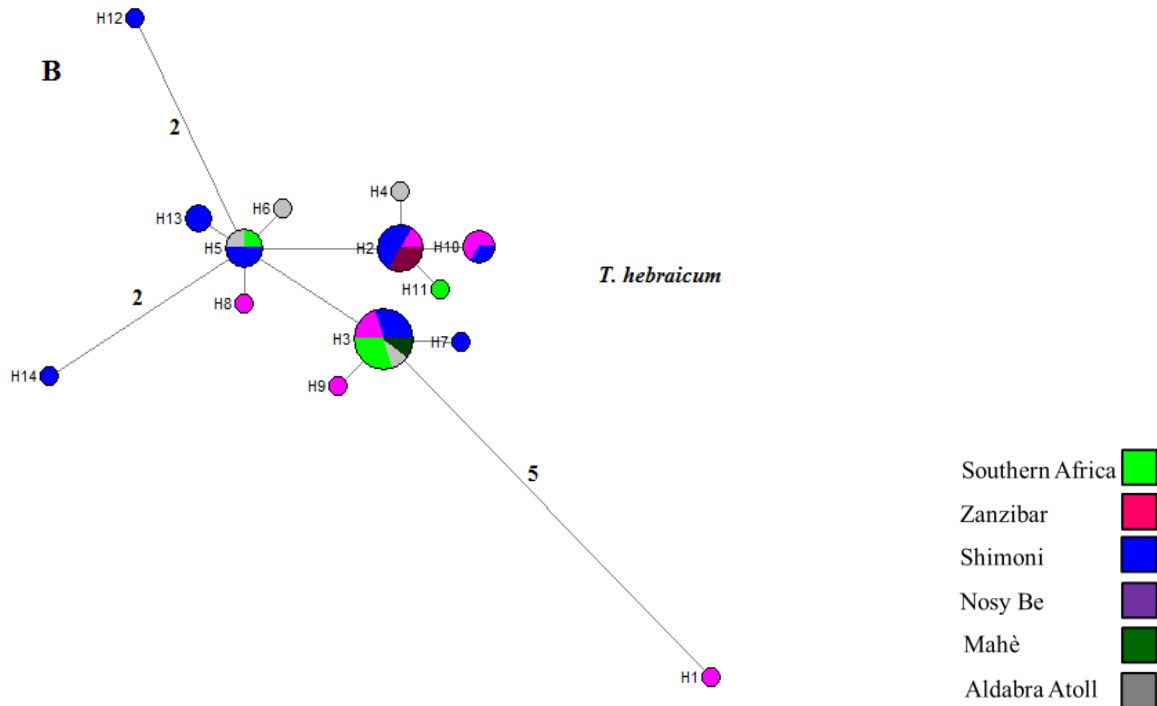


Fig. 4.8: Median-joining network depicting genealogical relationships of the mitochondrial cytochrome *b* haplotypes for the *Thalassoma hebraicum*. The size of the node corresponds to the frequency of that haplotype and the colours represent the occurrence of that haplotype at different WIO localities. The numbers on the branches indicate mutational differences where greater than one.

Estimates of Φ_{ST} derived from cytochrome *b* for *T. hebraicum* indicated differentiation of Nosy Be from southern Africa ($\Phi_{ST} = 0.37$) and Aldabra Atoll ($\Phi_{ST} = 0.21$; Table 4.6). The remaining estimates of Φ_{ST} suggested little to no differentiation between WIO localities. Hudson's (2000) nearest neighbour statistic S_{nn} was estimated to be 0.27 for the overall sample. For *T. lunare*, pairwise Φ_{ST} was not done because there was only one cytochrome *b* haplotype.

Table 4.6: Pairwise Φ_{ST} values between WIO localities for the *Thalassoma hebraicum* based on the mitochondrial cytochrome *b* data set.

	Southern Africa	Zanzibar	Shimoni	Nosy Be
Southern Africa	-			
Zanzibar	-0.10	-		
Shimoni	0.00	0.01	-	
Nosy Be	0.37	0.01	-0.03	-
Aldabra Atoll	-0.11	-0.07	-0.04	0.21

The median-joining networks of the nuclear marker are presented in Figure 4.9. For *T. hebraicum*, the S7 intron I data (312 bp) resulted in 26 unique alleles. These alleles H3 and H4 were shared among localities. Allele H3 was shared between Aldabra Atoll and Mahè (both Seychelles). Allele H4 was shared between Aldabra Atoll and southern Africa. Shimoni (H7 and H21) was shared among different individuals within the same locality. Some of the alleles from Zanzibar H1 and Southern Africa H18 were found in the same individual. The remaining alleles were each restricted to a specific locality.

The network for *T. lunare* was constructed with alleles from the WIO localities, the Red Sea and West Pacific Islands (Fig. 4.9B). After phasing, the S7 intron I data set had 14 unique alleles with only one allele H2 shared between localities, between Kafu Atoll (Maldives) and Pelorus Island. There were singletons present in the network. The network has two distinct lineages. The Red Sea alleles (lineage 1) are differentiated from the WIO and the West Pacific Islands (French Polynesia and Pelorus Island), and some of the alleles are found in the Red Sea. This network is in agreement with the neighbour-joining tree (see above).

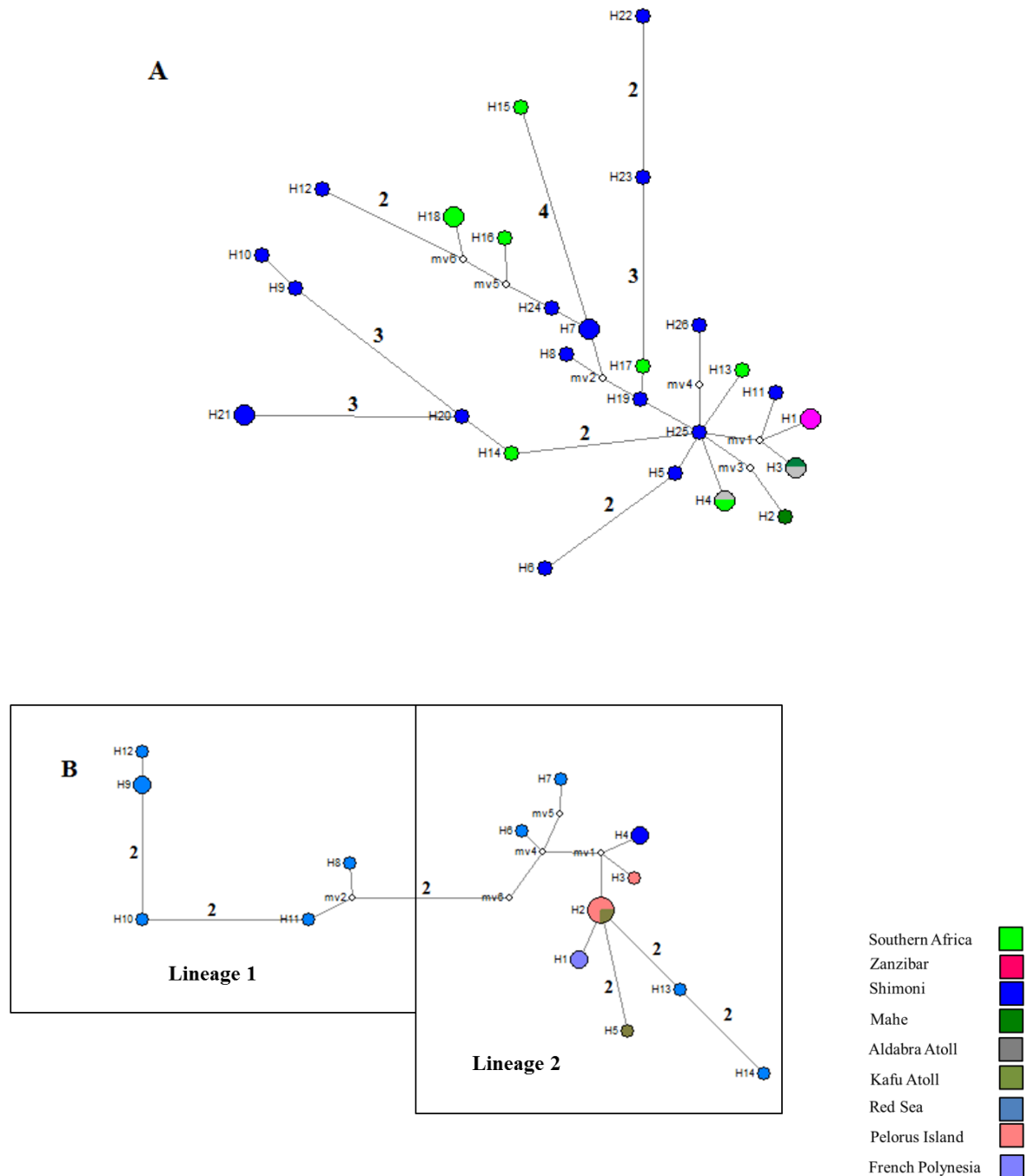


Fig. 4.9: Median-joining networks of genealogical relationships among S7 intron I alleles of (A) *Thalassoma hebraicum* and (B) *Thalassoma lunare* from WIO localities, the Red Sea and Pacific. Connecting lines between nodes represent one mutational step unless indicated otherwise. The size of nodes is proportional to number of individuals observed with a given haplotype. Pie charts are coloured according to the geographical origins of the allele. Missing haplotypes (mv) are shown in small white circles.

The most suitable model of nucleotide substitution for the nuclear S7 intron I gene was the Tamura and Nei (1993) model for *T. hebraicum*, and the Jukes and Cantor (1969) model was used for *T. lunare*. For *T. hebraicum*, among-population comparisons obtained from the S7 intron I yielded Φ_{ST} values from 0.06 to 0.30 (Table 4.7A). Shimoni showed little differentiation from all of the WIO localities. Aldabra Atoll (Seychelles) revealed differentiation from southern Africa ($\Phi_{ST} = 0.15$) and Zanzibar ($\Phi_{ST} = 0.30$). Moderate differentiation was identified between southern Africa and Zanzibar ($\Phi_{ST} = 0.15$). Estimates of Φ_{ST} for *T. lunare* indicated differentiation of the Red Sea from Shimoni ($\Phi_{ST} = 0.24$) and Kafu Atoll (Maldives) ($\Phi_{ST} = 0.21$) (Table 4.7B). The S7 intron I suggested great differentiation of Kafu Atoll from Shimoni ($\Phi_{ST} = 0.67$). No comparisons were statistically significant for the two species. The two species showed differentiation with the nearest neighbour statistic of $S_{nn} = 0.60$ for *T. hebraicum*, and $S_{nn} = 0.67$ for *T. lunare* for the overall sample.

Table 4.7: Comparison of pairwise Φ_{ST} values between WIO localities for (A) *Thalassoma hebraicum* and (B) between WIO localities and the Red Sea for *T. lunare* based on nuclear S7 intron I data sets.

A

	Southern Africa	Zanzibar	Aldabra Atoll
Southern Africa	-		
Zanzibar	0.15	-	
Aldabra Atoll	0.15	0.30	-
Shimoni	0.06	0.06	0.07

B

	Kafu Atoll	Shimoni	Red Sea
Kafu Atoll	-		
Shimoni	0.67	-	
Red Sea	0.21	0.24	-

For *T. hebraicum*, the most suitable model of nucleotide substitution for each gene fragment (ATPase 6, cytochrome *b* and S7 intron I) was a Tamura and Nei (1993) model. The Analysis of Molecular Variance (AMOVA) of *T. hebraicum* was examined with localities grouped based proximity and on Santini & Winterbottom's (2002) biogeographic regions. There were three regions for the ATPase 6 and cytochrome *b* gene regions: Natal Basin (southern Africa and Nosy Be), Somali Basin (Zanzibar and Shimoni) and Mascarene Plateau (Seychelles). The

nuclear S7 intron I data was arranged with two groups as follows: southern Africa, Shimoni and Zanzibar versus Seychelles. The arrangements provided the best description of differentiation, maximising the Φ_{CT} values attributed to differentiation among groups. The components of variance among regions for the three gene fragments (ATPase 6: -4.54%, Φ_{CT} = -0.045; cytochrome *b*: -8.82%, Φ_{CT} = -0.088; S7 intron I: 0.77%, Φ_{CT} = 0.007) were not significant ($P > 0.05$; Table 4.8). There were low and non-significant ($P > 0.05$) components of variance found among localities within regions for all datasets (ATPase 6: 7.59%, Φ_{SC} = 0.073; cytochrome *b*: 5.03%, Φ_{SC} = 0.046; S7 intron I: 7.56%, Φ_{SC} = 0.076). Most of the variation (96.95%, 103.79%, 91.67% and for ATPase 6, cytochrome *b* and S7 intron, respectively) was partitioned within localities. The components associated with differentiation within localities were not significant ($P > 0.05$) for the mitochondrial gene regions, but significant ($P > 0.05$) for the nuclear gene region. For *T. lunare*, AMOVA could not be done for all markers due to small sample sizes. There one cytochrome *b* haplotype for *T. lunare* may suggest a lack of variation.

Table 4.8: The results of AMOVAs examining the partitioning of differentiation at various hierarchical levels for *Thalassoma hebraicum* from across the WIO based on the ATPase 6, cytochrome *b* and nuclear S7 intron I data sets. The hierarchical arrangements are described in the text.

Gene region	Source of variation	d.f.	Sum of squares	% of variation	<i>P</i> value
ATPase 6	Among regions	2	3.48	-4.54	0.53
	Among localities within regions	2	4.17	7.59	0.38
	Within localities	22	35.41	96.95	0.20
Cytochrome <i>b</i>	Among regions	2	1.18	-8.82	1.00
	Among localities within regions	2	2.84	5.03	0.16
	Within localities	29	31.60	103.79	0.55
S7 intron I	Among regions	2	4.15	0.77	0.50
	Among localities within regions	2	8.77	7.56	0.09
	Within localities	29	77.90	91.67	0.04

For *T. hebraicum*, the estimates from the neutrality tests and mismatch distributions for three gene markers (ATPase 6, cytochrome *b* and S7 intron I) are displayed in Table 4.9. Tajima's *D* and Fu's *F_s* statistics for the overall sample were negative and significant ($P < 0.01$) for the three gene markers. Both of these statistics suggest an excess of low-frequency variants, which indicates a population expansion. The mismatch distributions of the three genes are presented in Figure 4.10. The ATPase 6 gene revealed two distinct peaks, and this mismatch distribution might have picked up the differentiation of certain Seychelles haplotypes in the network. The mismatch distributions of the cytochrome *b* and S7 intron I data were multimodal. The multimodal distributions can represent a demographic equilibrium (Zhang *et al.*, 2014). However, Harpending's raggedness index (*r*) and the Sum of Squared Deviations were low and non-significant for the overall sample of the three gene markers, showing a good fit to the expected population expansion distribution.

The estimated expansion time for *T. hebraicum*, based on the ATPase 6 data set, using a Tau (τ) of 3.00, was 92 678 years ago (Table 4.10). The 95% confidence interval (CI: 1.57-4.23) produced rough dates of expansion from 48 501 to 130 676 years ago. The analyses of cytochrome *b* showed a Tau (τ) of 2.00, which resulted in expansion time of 69 444 years ago, and the 95% confidence interval (CI: 1.36-2.58) revealed expansion times between 47 222 to 89 583 years ago. The time of expansion was estimated to have taken place during the Pleistocene for both gene fragments. Both of these gene fragments indicated an initial effective population size (N_{ef0}) of 0 and the current effective population size (N_{ef1}) was 1 073 833 for ATPase 6 and 3 472 187 for cytochrome *b*, respectively.

Table 4.9: The results of neutrality tests, including Tajima's (1989) D and Fu's (1997) F_s , and the statistics from the mismatch distribution, including Harpending's (1994) raggedness (r) index and the Sum of Squared Deviations (SSD), for the (A) mitochondrial ATPase 6 and cytochrome b , and (B) nuclear S7 intron I data sets. Significant estimates are indicated: ** $P < 0.01$, *** $P < 0.001$.

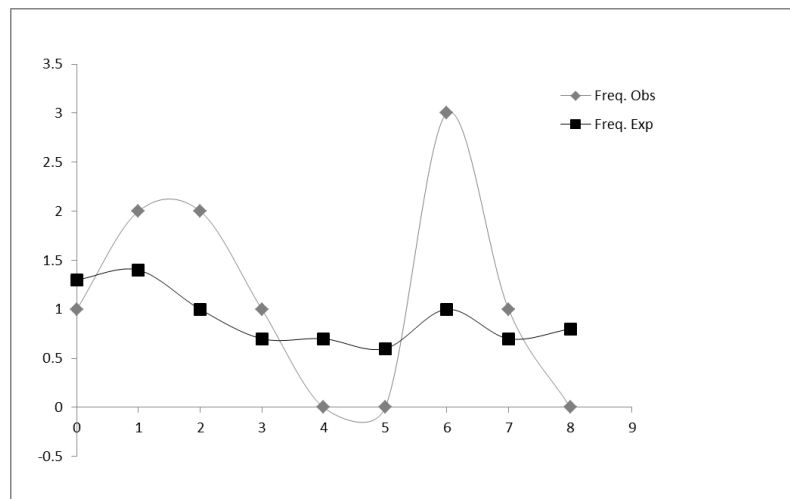
A

Locality	ATPase 6								Cytochrome b							
	D	P value	Fu's F_s	P value	r	P value	SSD	P value	D	P value	Fu's F_s	P value	r	P value	SSD	P value
Southern Africa	-0.81	0.32	-0.13	0.36	0.17	0.78	0.08	0.56	-0.17	0.45	0.93	0.13	0.11	0.84	0.04	0.65
Tanzania	0.00	0.84	-0.34	0.16	0.44	0.78	0.12	0.50	-1.05	0.19	-0.79	0.28	0.19	0.37	0.03	0.59
Shimoni	-1.60	0.04	-5.90	0.00***	0.04	0.68	0.00	0.65	-0.88	0.01**	-3.89	0.01**	0.08	0.35	0.01	0.20
Seychelles	-0.83	0.12	-0.52	0.00***	0.56	0.32	0.17	0.14	-0.41	0.09	-1.19	0.09	0.13	0.73	0.01	0.73
Overall	-2.00	0.01**	-11.59	0.00***	0.03	0.70	0.01	0.32	-1.79	0.00***	-8.14	0.00***	0.06	0.31	0.01	0.12

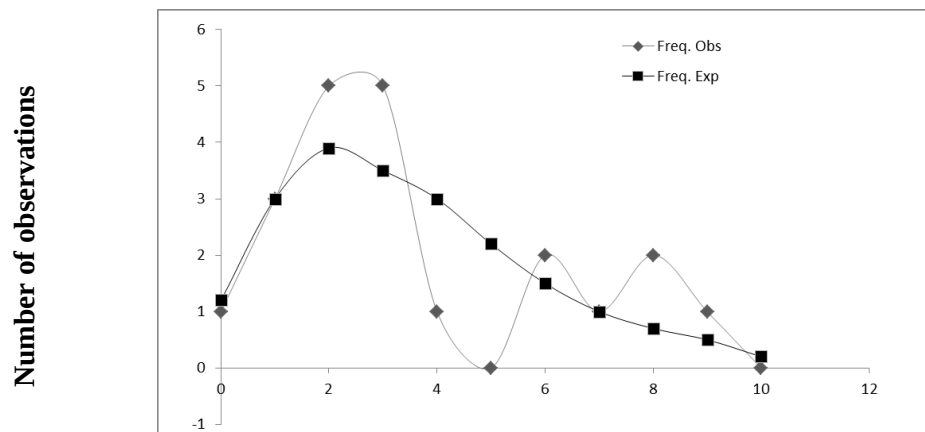
B

Locality	S7 intron I							
	D	P value	Fu F_s	P value	r	P value	SSD	P value
Southern Africa	-0.17	0.45	-0.06	0.31	0.17	0.82	0.04	0.65
Tanzania	-1.81	0.06	-1.81	0.06	0.07	0.80	0.03	0.59
Shimoni	-1.05	0.06	-3.04	0.01**	0.08	0.24	0.01	0.20
Seychelles	-0.41	0.44	-1.19	0.09	0.13	0.77	0.02	0.74
Overall	-0.71	0.01**	-18.49	0.00***	0.02	0.42	0.01	0.37

A



B



C

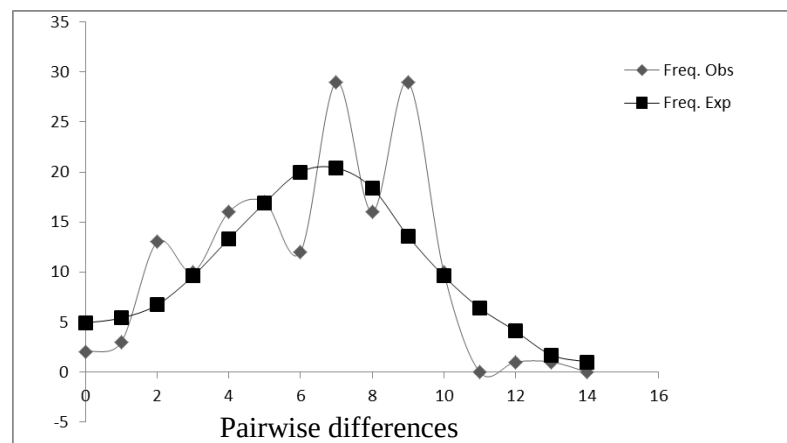


Fig. 4.10: Mismatch distributions of pairwise differences among (A) ATPase 6 and (B) cytochrome *b* and (C) S7 intron I sequences of all *Thalassoma hebraicum* individuals. The observed curve is indicated with a black line. The expected model under a population growth model is illustrated with a grey line.

Table 4.10: Demographic parameters τ , θ_0 , and θ_1 from the mismatch distributions of the mitochondrial (A) ATPase 6 and (B) cytochrome *b* gene regions for *T. hebraicum* used to calculate time of expansion and effective population size for time zero (N_{ef0}) and the present day (N_{ef1}).

A

Locality	τ	Time of expansion	θ_0	N_{ef0}	θ_1	$N_{ef1} (10^9)$
Southern Africa	7.20	222 428	0	0	6.33	195 551
Tanzania	3.40	105 035	0	0	99 999.00	3 089 249
Shimoni	2.80	86 499	0	0	99 999.00	3 089 249
Seychelles	7.90	244 053	0	0	17.56	542 477
Overall	3.00	92 678	0	0	34.76	1 073 833

B

Locality	τ	Time of expansion	θ_0	N_{ef0}	θ_1	$N_{ef1} (10^9)$
Southern Africa	0.02	694 44	0.89	30 902	99 999.00	3 472 187
Tanzania	1.50	52 083	2.03	70 486	99 999.00	3 472 187
Shimoni	2.30	79 861	0	0	99 999.00	3 472 187
Seychelles	2.20	76 388	0	0	99 999.00	3 472 187
Overall	2.00	69 444	0	0	99 999.00	3 472 187

Discussion

Thalassoma hebraicum

Both the mitochondrial (ATPase 6 and cytochrome *b*) and nuclear S7 intron I sequence data analyses indicated little genetic structuring in *T. hebraicum* across the WIO. Both of the mitochondrial networks were characterised by a few widely distributed haplotypes, which represented ancestral haplotypes, indicating historical connectivity among localities. The S7 intron I network showed many private alleles indicating high diversity. The neighbour-joining trees of the ATPase 6 and S7 intron I genes showed a *T. hebraicum* lineage that was widespread across the WIO localities.

Pairwise comparisons yielded Φ_{ST} values that were low and high for all markers. The pairwise Φ_{ST} values for ATPase 6 suggested the differentiation of Mahé (Seychelles) from East African localities (Zanzibar and Shimoani). Cytochrome *b* indicated the differentiation of Nosy Be (Madagascar) from Aldabra Atoll and southern Africa. The S7 intron I data showed the differentiation of Seychelles (Aldabra Atoll) from the African mainland (Zanzibar and southern Africa). This marker also indicated differentiation between Zanzibar and southern Africa. The Analyses of Molecular Variance (AMOVA) based on the three genes detected most of the differentiation within localities. Across the markers, the AMOVAs did not find differentiation among grouped localities. The nearest neighbour statistics of the two mitochondrial genes were low, indicating panmixia for the overall sample, while the nearest neighbour statistics of S7 intron I was closer to one indicating differentiation.

The two mitochondrial gene fragments revealed connectivity from historical perspective, and the nuclear S7 intron I gene reflected patterns of genetic structure from contemporary perspective (Hauser & Ward, 1998; Hellberg, 2009). These three markers may not be sufficient to detect fine scale structure. Highly variable markers need to be considered in the future. The mitochondrial control region has a faster mutation rate than all other mitochondrial genes in fish (Lee *et al.*, 1995; Kochzius & Blohm, 2005) and it is a suitable marker for the investigation of genetic structure of populations (Avise *et al.*, 1987). The selection of nuclear markers for future studies is discussed in Chapter 3 and 5. There is little information on the reproductive biology of *T. hebraicum*. The larval duration of *T. hebraicum* is unknown. Bardach *et al.* (1980) reported that *Thalassoma* species spawn on outgoing tides to aid dispersal. The spawning behaviour of *Thalassoma* species may be the perfect strategy to increase dispersal and help

maintain connectivity among the WIO localities as the larvae of *T. hebraicum* are dispersed away from the spawning grounds to connect disjunct populations. Larvae of reef fish species are not all passive and some can swim faster than the average speeds of ocean currents (Leis, 2010). For example, dispersal in *Lutjanus kasmira* (i.e. by pelagic larvae) is suggested to be enhanced by the swimming capabilities of the larvae of the species (Muths *et al.*, 2012; see Chapter 3).

Some of the genetic analyses of the present study have shown differentiation between localities. Pairwise comparisons showed the differentiation of Seychelles from the African mainland (southern Africa, Zanzibar and Shimoni) and Nosy Be (Madagascar). The ATPase 6 network also showed the differentiation of certain Seychelles haplotypes from those from other WIO localities. The differentiation of *T. hebraicum* could be explained by the influence of regional ocean current circulation, such as the South Equatorial Current (SEQC) and East Africa Coastal Current (EACC) (see Chapter 3). The results of the present study are in agreement with the study of the mangrove crab *Neosarmatium meinerti* within the WIO. Ragionieri *et al.* (2010) considered differentiation in *N. meinerti* using mitochondrial cytochrome oxidase I (COI) data. In *N. meinerti*, clear genetic differentiation of Seychelles from the East African coast (Tanzania and Kenya) and southern Africa was detected, which was suggested to be influenced by the South Equatorial Current (SEQC) and East Africa Coastal Current (EACC) (Ragionieri *et al.*, 2010). Another possible explanation for the differentiation in *T. hebraicum* might be the geographic distance from Seychelles to the African mainland. Seychelles is separated from coastal East Africa by more than 1500 km of open sea (Ragionieri *et al.*, 2010). The pelagic larvae of *T. hebraicum* might not be able to disperse across hundreds of kilometres by means of ocean currents to connect Seychelles with the African mainland.

The nuclear S7 intron I revealed the differentiation of southern Africa from Zanzibar and Nosy Be. Genetic connectivity of *T. hebraicum* is restricted between the northern and southern localities, and this differentiation can be attributed to SEQC and the Comoros Gyre (Obura, 2012; see Chapter 3). The SEQC has been identified as major oceanographic barriers to genetic connectivity in the blue barred parrotfish *Scarus ghobban* studied to date (Visram *et al.*, 2010; see Chapter 3). This species showed that connectivity is restricted between southern localities. The seasonal anticyclones in the northern entrance of the Mozambique Channel (Donguy & Piton, 1991), and the anti-cyclonic and cyclonic eddies in the central part of the channel (Schouten *et al.*, 2002; see Chapter 1) could have restricted the dispersal of *T. hebraicum*. The

observed differentiation of *T. hebraicum* supports the hypothesis that larval dispersal across the Mozambique Channel is restricted (Berry & Plante, 1973; Gopal *et al.*, 2006). Gopal *et al.* (2006) used mitochondrial control region data to investigate differentiation in the spiny lobster *P. delagoae* within the southwestern Indian Ocean. They found differentiation of South Africa from Mozambique and Madagascar, which was attributed to the anti-cyclonic eddies along the Mozambique Shelf and East Madagascar Current water masses reaching the Agulhas Current (Gopal *et al.*, 2006).

The significant, negative Tajima's D and Fu's F_s values for the three markers illustrate that the population of *T. hebraicum* had undergone a recent population expansion. The three markers (ATPase 6, cytochrome b and S7 intron I) revealed high haplotype and low nucleotide diversities, indicating population expansion after a period of low effective population size (Grant & Bowen, 1998). The mismatch distribution of the ATPase 6 data showed a bimodal shape because of the divergence of certain Seychelles individuals from the remaining samples. However, the model of population expansion was not rejected. The expansion time was estimated to have been initiated 92 678 years ago for ATPase 6. Mismatch distributions of cytochrome b (approximately 69 444) and S7 intron I data showed multimodal curves. The population expansion of *T. hebraicum* took place in the late Pleistocene, as it did for *C. inermis* (see Chapter 3). During the Pleistocene (from 110 000 to 75 000 years ago) the sea level was lowered between 0 and -75m (Toms *et al.*, 2014; see Chapter 3). The mitochondrial markers have shown different expansion times and overlapping confidence interval dating for *T. hebraicum*. The different expansion times of the mitochondrial markers are explained in Chapter 3. Sea level fluctuations (during the Pleistocene) might have reduced suitable habitats for marine fish such *T. hebraicum* in tropical seas (Voris, 2000). The most likely explanation for the expansion of *T. hebraicum* is when the post-glacial sea level rise (Camoin *et al.*, 2004) increased suitable habitats. Klanten *et al.* (2007) suggested that population expansion of *Naso vlamingii* took place during the Pleistocene in the Indo-Pacific. They suggested that the sea level rise might have increased available local habitats for this species to expand (Klanten *et al.*, 2007).

Thalassoma lunare

Both mitochondrial ATPase 6 and nuclear S7 intron I data revealed genetic structure between the Red Sea and WIO, as well as within the WIO. The networks and neighbour joining trees of the two markers (ATPase 6 and nuclear S7) showed two highly-divergent haplogroups of the WIO and the Red Sea for the widespread *T. lunare*. In the networks, the two haplogroups (i.e. WIO and Red Sea) were separated from each other by nine mutational steps (ATPase 6), while the S7 intron I was separated by three mutational steps. ATPase 6 network and neighbour joining tree showed that one haplotype of the WIO haplogroup was present in the Red Sea. Nuclear S7 intron I network and neighbour joining tree revealed four alleles of the WIO and Pacific that were found in the Red Sea. These Red Sea individuals reflect contemporary spatial connectivity rather than the historical patterns of connectivity. The two lineages differed by 1.2% sequence divergence for ATPase 6 and 0.4% divergence for the S7 intron I data. The damselfish *Chromis dimidiata* revealed a sequence divergence of 1.9% that was based on cytochrome *b*, suggesting that there are two distinct species in the central Red Sea and in the Indian Ocean (Randall & Di Battista, 2013). The type locality of *Chromis dimidiata* is the Red Sea, the Indian Ocean population was described as a new species, *Chromis fieldi* (Randall & Di Battista, 2013). Mitochondrial cytochrome *b* data revealed a sequence divergence of 0.3% between morphologically distinct species of Creole wrasse, *Clepticus brasiliensis* and *C. africanus* (Beldade *et al.*, 2009). The gene markers (ATPase 6 and S7 intron I) of the present study are not comparable, but there was no cytochrome *b* data. There was only one cytochrome *b* haplotype for the widespread *T. lunare*. The sequence divergence of other species (see above) were used to see how variable ATPase 6 and cytochrome *b* markers are when they compared to each other. The sequence divergence of 0.3% is far less than the 1.2% sequence divergence of ATPase 6 between the haplogroups of *T. lunare*, and the sequence divergence of 1.9 % is not far less than 1.2%, this could suggest a possibility of two different species occurring in the Red Sea and WIO. Future studies need to add samples from the type locality Indonesia (Kuitert & Tono-zuka, 2001) to confirm the possibility of two different species occurring in the WIO and Red Sea.

The two distinct lineages of *T. lunare* could be caused by the restricted connectivity between the WIO and the Red Sea, and this restricted connectivity might have occurred for a long period of time. The geological history of the Red Sea may have created historical barriers to reef fish such as *T. lunare* (Di Battista *et al.*, 2013). During the Pleistocene, the sea level dropped by 120 m and produced land bridges between the WIO and the Red Sea (Tsang *et al.*, 2012). This

restricted the exchange of water between the WIO and Red Sea, owing to the shallow strait at the Bab al Mandab (the only gateway from the WIO to the Red Sea) (Siddall *et al.*, 2003). This might have prevented genetic connectivity of the widespread *T. lunare* and led to speciation (i.e. the distinct lineages) between the WIO and the Red Sea. Di Battista *et al.* (2013) also noted that the gateway of the Red Sea limited the connectivity of several reef fish species between the WIO and the Red Sea.

The oceanographic features of the Red Sea and Gulf of Aden are strongly influenced by the Indian Ocean monsoon winds (Biton *et al.*, 2010). The monsoon winds influence the Somali Current and reverse its direction, according to the seasons (Lutjeharms, 2006). The area between the Horn of Africa (Somalia) and the Arabian Gulf has been described as a biogeographic barrier (Kemp, 1998). Ngoile and Linden (1997) stated that there is cold surface water and low salinity produced by massive upwelling in Somalia during the southwest monsoon. Santini and Winterbottom (2002) mentioned that contemporary separation of the Red Sea from the WIO is caused by the seasonal cold upwelling cells. The cold upwelling cells might form an effective barrier to the movements of tropical marine reef fishes such as *T. lunare*. The salinity in the Somali Current might be a barrier, causing the isolation of the widespread *T. lunare* between the two ocean basins (WIO and Red Sea). The turbid-water in the south of Red Sea is also suggested to limit larval dispersal of reef fishes such as *T. lunare* (Roberts *et al.*, 1992). Froukh & Kochzius (2007) provided evidence of differentiation which was caused by turbidity in the endemic fourline wrasse *Larabicus quadrilineatus*.

The distinction between the WIO and Red Sea in the current study is supported by studies of mud crab and multiple reef fishes (Gopurenko *et al.*, 1999; Fratini & Vannini, 2002; Di Battista *et al.*, 2013). Gopurenko *et al.* (1999) and Fratini & Vannini (2002) found genetic differentiation among populations of the mangrove crab *Scylla serrata*, using mitochondrial markers, with the Red Sea being completely isolated from WIO localities (South Africa and Mauritius). Recently, Di Battista *et al.* (2013) used two mitochondrial markers (cytochrome *b* and COI), to reveal restricted genetic connectivity between the WIO and Red Sea for the following reef fishes: brown surgeonfish *Acanthurus nigrofuscus*, peacock hind *Cephalopholis argus*, threadfin butterflyfish *Chaetodon auriga*, checkerboard wrasse *Halichoeres hartulanus* and bluestripe snapper *Lutjanus kasmira*. By contrast, the lionfish *Pterois miles* revealed a different picture. Kochzius and Blohm (2005) found no genetic differentiation between *P. miles* populations in the Indian Ocean and Red Sea using the mitochondrial control gene region.

The median joining network and neighbour joining tree for the S7 intron I data showed that Kafu Atoll (Maldives) is not differentiated Pelorus Island. The Indo-Pacific Barrier (IPB) is a barrier that separates the Indian and Pacific Oceans (Bellwood & Wainwright, 2002; Gaither *et al.*, 2010). The Indo-Pacific Barrier was enhanced during the low sea-level stands (during the Pleistocene, as below as -120 m) when there were fewer passages between islands, and by strong upwelling cells that were likely to reduce the availability of suitable habitat for tropical species, such as *T. lunare* (Voris, 2000; Barber *et al.*, 2002). The IPB seems not to have restricted the connectivity of the widespread *T. lunare* by causing isolation or genetic diversification between the two oceans (WIO and Pacific). The lack of differentiation observed between Maldives and Pelorus Island (Pacific) could be the result of life history characteristics of the species. The SEQC carries waters from the Indonesian region across the Indian Ocean (Obura, 2012). The widespread *T. lunare* has long pelagic larval duration (PLD) of 39 to 55 days (Victor, 1986), and the SEQC may be able to carry pelagic larvae across the Indo-Pacific Barrier. Similarly, Gaither *et al.* (2010) confirmed that the Indo-Pacific Barrier was not effective in restricting genetic connectivity in *Lutjanus fulvus* and *L. kasmira*. By contrast, other authors (Duda & Palumbi, 1999; Bay *et al.*, 2004; Craig *et al.*, 2007) found differentiation of the WIO from other Pacific localities. Craig *et al.* (2007) used cytochrome *b* to consider differentiation of the Indian Ocean and Pacific Oceans in the reef fish *Myripristis berndti*. They found differentiation between the Indian and Pacific Oceans across the Indo-Pacific Barrier, and little differentiation between the Eastern and Central Pacific which was separated by the Eastern Pacific Barrier (EPB). Duda and Palumbi (1999) found marked genetic differentiation between the WIO and Pacific Ocean in the black tiger prawn *Penaeus monodon*, using an elongation factor 1-alpha intron data. This differentiation was attributed to the Indo-Australian Archipelago (Duda & Palumbi, 1999).

The ATPase 6 and S7 intron I data indicated genetic differentiation within the WIO. The ATPase 6 data indicated differentiation of Kafu Atoll (Maldives) and Seychelles from the African mainland (southern Africa and Zanzibar), and Zanzibar was differentiated from southern Africa. The variable nuclear S7 intron I data revealed differentiation between Kafu Atoll (Maldives) and Shimoni (Kenya). The nearest neighbour statistic for S7 intron I provided evidence of geographic genetic structure. However, the nearest neighbour statistic for ATPase 6 was less than 0.5, indicating little genetic differentiation for the overall sample. The sample size of *T. lunare* for was insufficient to examine differentiation using AMOVA.

There are eddies that circulate around the Maldives, which are driven by the seasonal monsoon winds (Shankar *et al.*, 2002). The differentiation of Maldives from the African mainland and Seychelles is probably due to the upwelling cells created by monsoon winds (Sarkar & Gupta, 2009). Sinha (1994) stated that eddy-like, circulatory motions of waters keep pelagic larvae of fishes within the vicinity of the islands for a long period of time, and this might have restricted dispersal of the widespread *T. lunare*. Besides ocean currents, the differentiation of Maldives (the distant locality in the north-east) from the African mainland and Seychelles is probably due to isolation by distance, but tests of isolation by distance were not done due to small sample sizes.

The differentiation of Seychelles from the African mainland was observed in the widespread *T. lunare*. This differentiation has been also observed in the WIO species *T. hebraicum*. The differentiation of Seychelles is also associated with the influence of the SEQC and distance, as discussed for the WIO species *T. hebraicum*. *Thalassoma lunare* revealed differentiation between southern Africa and Zanzibar, and this differentiation is attributed to the Comoros Gyre and eddies in the Mozambique Channel, as discussed previously for *T. hebraicum* and *C. inermis* (see above and Chapter 3).

The mitochondrial cytochrome *b* analysis showed connectivity (i.e. a single widespread haplotype) among localities which were found to be distinct in the analysis of ATPase 6 and S7 intron I. The cytochrome *b* data of the two species were combined and the median-joining network showed sharing of this haplotype between *T. hebraicum* and *T. lunare*. Certain individuals of *T. lunare* that were separated from *T. hebraicum* by the ATPase 6 and S7 intron I markers were included in the cytochrome *b* analysis. The Red Sea samples were not represented in the analysis of cytochrome *b*. There are potential explanations for the existence of a shared cytochrome *b* haplotype between the two species. There is the possibility of incomplete lineage sorting in that cytochrome *b* might not be less variable compare to ATPase 6 and S7 intron I. The cytochrome *b* data might thus reveal ancestral polymorphism (variants that arose by mutation before the divergence event). Hybridization is normally expected between closely related species (Bernardi *et al.*, 2004). *Thalassoma lunare* is not closely related to *T. hebraicum* (Bernardi *et al.*, 2004). However, *T. lunare* appears to hybridise with species that are not closely related. *Thalassoma lunare* and *T. rueppelli* are not closely related, but they were seen hybridizing together in the Red Sea (Randall & Miroz, 2001; Bernardi *et al.*, 2004). *Gomphosus varius* and *T. lunare* hybridized in north-western Australia (Randall & Miroz,

2001), and they are not closely related either. As the mitochondrial ATPase 6 did not show the same pattern as the cytochrome *b* data, hybridization could not be a possible explanation for this pattern. If there was hybridization in the two species, one would expect to find haplotypes of one species within individuals of the other species (Allendorf *et al.*, 2001). Future studies should add more samples in the cytochrome *b* data to confirm this observation, and maybe other markers (see above and Chapter 3) should also be used to confirm this pattern.

CHAPTER FIVE

CONCLUDING DISCUSSION

The primary aim of this study was to evaluate genetic differentiation of the selected wrasse species within the WIO, which may be related to the biogeographic hypotheses of dispersal and vicariance. In the present study, three reef-associated labrid fishes were selected to study biogeography, namely *Cheilio inermis*, *Thalassoma hebraicum* and *T. lunare* (refer to Chapters 3 and 4). Multiple markers (mitochondrial ATPase 6, cytochrome *b* and nuclear S7 intron I) were used to assess genetic structure of the three wrasses, to provide information on variability and resolution at different time scales (refer to Chapter 2). The selected markers have different means of transmission and different evolutionary rates, which are important for distinguishing ecological and evolutionary processes.

In summary, the present study showed varying levels of differentiation and some connectivity among the three reef-associated labrid fishes within the WIO. The particular results of the present study are dependent on the markers that were utilised. This study showed the differentiation of Kenya and Tanzania from Mozambique and Nosy Be (Madagascar) in the monotypic and widespread *C. inermis* (see Chapter 3). *Cheilio inermis* also revealed the differentiation of Kenya and Tanzania (Chapter 3). Only the nuclear locus of *C. inermis* did not reveal variation among defined groups. This pattern of differentiation was attributed to the life history features of the species, and physical features such as the South Equatorial Current (SEQC), East Africa Coastal Current (EACC), Comoros Gyres and the land bridges that interrupted the Mozambique Channel (Obura, 2012; Santini & Winterbottom, 2002; Chapter 3). Sequence divergence was 0.4% therefore the possibility of cryptic species was not found in *C. inermis* across the WIO. Overall, the widespread *C. inermis* indicated differentiation and little connectivity within the WIO.

The WIO species *T. hebraicum* generally revealed little genetic differentiation within the WIO. The AMOVAs of *T. hebraicum* showed no differentiation among the defined locality groups for the three markers. Most of the differentiation was found within localities for the three markers. However, some genetic differentiation was detected between localities. *Thalassoma hebraicum* showed differentiation between Seychelles and the African mainland, and indicated the differentiation of southern Africa from Zanzibar and Nosy Be (see Chapter 4). This pattern

of differentiation was associated with the SEQC and the geographic distance between Seychelles to the African mainland (Chapter 4). The large anti-cyclonic and cyclonic eddies in the Mozambique Channel, as well as the upper Agulhas Current might have restricted larval dispersal between Madagascar and southern Africa.

The widespread *T. lunare* revealed genetic differentiation between the WIO and Red Sea. This differentiation could be either due to the historical isolation by the Bab al Mandab and present-day barrier such as seasonal monsoon winds creating thermal and salinity barriers through upwellings (Chapter 4). The WIO localities (Maldives, Seychelles, and Shimoni) were differentiated from each other (Chapter 4). Maldives was genetically differentiated from the WIO localities in *T. lunare*. The differentiation of Maldives might be caused by isolation by distance and upwellings cells (Chapter 4). Maldives was not genetically differentiated from the Indo-West Pacific Island (Pelorus Island) in *T. lunare*. The Indo-Pacific Barrier (IPB) seems not to limit connectivity of *T. lunare*, thereby causing genetic diversification between the two oceans (WIO and Pacific). The SEQC might carry the pelagic larvae of *T. lunare* across the Indo-Pacific Barrier. The results of the study support the hypothesis that biogeographic barriers among areas and physical factors (oceanographic processes) influence the distribution of genetic variation of wrasse fish species within the WIO. The life history features of *T. hebraicum* were suggested to have assisted dispersal and connectivity within WIO. The present study supports both biogeographic hypotheses of dispersal and vicariance. The patterns that were observed by the current study are consistent with or similar to the previously reported isolation of localities in fish species and invertebrates of the WIO (see Chapter 3 and 4).

Caution should be taken when interpreting the results of the present study due to small sample sizes, especially for the widespread *T. lunare*, and due to the fact that some of the samples amplified for the mitochondrial gene markers and did not amplify for the nuclear marker. Another reason is that each marker had different number of samples and the samples were not always from the exact same localities. Future studies needs to carefully consider more sampling localities and larger sample sizes in order to understand the biogeography of the WIO and do more analyses for better confidence in the patterns observed.

This study has used and compared results obtained from different markers. This provides greater confidence in the patterns observed (Guo & Chen, 2010). The selected markers were sufficient to investigate biogeographic patterns; the markers detected genetic structure (i.e. in

C. inermis and *T. lunare*) and could separate patterns that were due to historical processes from contemporary ones (e.g. *T. lunare*). Mitochondrial markers specifically reveal historical rather than contemporary connectivity (Hauser & Ward, 1998). A marker with a faster mutation rate, such as the mitochondrial control region (Meyer, 1993; Lee *et al.*, 1995), should be considered because this marker can pick up recent divergences and fine scale structure. Future genetic analyses should be done using microsatellites, because they are highly variable and sensitive (Hellberg, 2009). Microsatellites are good markers to detect genetic structure and patterns of contemporary genetic flow (Hauser & Ward, 1998).

Biological information for the selected wrasse species (*T. hebraicum*, *T. lunare* and *C. inermis*) is limited; for example, the larval duration of *T. hebraicum* is unknown. Patterns of genetic differentiation might be a consequence of intrinsic characteristics of the fish species; therefore, biological and ecological studies are needed to understand the reproductive biology and ecology of the species. Genetic data can provide a viewpoint of connectivity over thousands to hundreds of thousands of years, but might not necessary show present-day connectivity (Hellberg *et al.*, 2002; Reece *et al.*, 2010). Thus, the investigation of connectivity in the three wrasse species should not only use genetic markers. Selected markers should be combined with isotopes or otolith microchemistry approaches for studying movement and connectivity (Fowler *et al.*, 1995; Jones *et al.*, 1999).

Muths *et al.* (2011) stated that single-species studies are unsatisfactory as tools to inform managers, because single-species case studies are unable to detect all barriers to dispersal (Toonen *et al.*, 2010). Toonen *et al.* (2010) used a multispecies approach to evaluate population connectivity across the Hawaiian Archipelago, and they revealed four unknown barriers to dispersal that define five isolated regions of the Hawaiian Archipelago. The present study has used multispecies approach to investigate connectivity across the WIO; thus, the genetic information of the study can be used by conservation and fisheries managers to understand fish connectivity and identify the features that influence genetic connectivity within the WIO.

References

- Akaike, H. (1974). A new look at the statistical model identification. *Institute of Electrical and Electronics Engineers Transactions on Automatic Control* 19: 716-723
- Allendorf, F.W., Leary, R.F., Spruell, P. & Wenburg, J.K. (2001). The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution* 16: 614-622
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410
- Anand, P.E.V. & Pillia, N.G.K. (2002). Reproductive biology of some common coral reef of the Indian EEZ. *Journal of Marine Biological Association of India* 44: 122-135
- Arthurton, R. (2003). The fringing reef coasts of Eastern Africa-present processes in their long term context. *Western Indian Ocean Journal of Marine Science* 2: 1-13
- Awise, J.C., Arnold, J., Martin B.R., Bermingham, E., Lamb, T., Niegel, T.E., Reeb, C.A. & Saunders, N.C. (1987). Intraspecific phylogeography: The mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics* 18: 489-522
- Bandelt, H.J., Forster, P. & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16: 37-48
- Barber, P.H., Palumbi S.R., Erdmann, M.V. & Moosa, M.K. (2002). Sharp genetic breaks among populations of *Haptosquilla pulchella* (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology* 11: 659-674
- Bardach, J.E., Magnuson, J.J., May, R.C. & Reinhart, J.M. (1980). Fish behaviour and its use in capture and culture of fishes. Manila, Philippines. 473 pp.

- Bay, L.K., Choat, J.H., van Herwerden, L. & Robertson, D.R. (2004). High genetic diversities and complex genetic structure in an Indo-Pacific tropical reef fish (*Chlorurus sordidus*): evidence of an unstable evolutionary past? *Marine Biology* 144: 757–767
- Beckley, L.E., Hulley, P. & Skelton, P.H. (2002) Synoptic overview of marine ichthyology in South Africa. *Marine and Freshwater Research* 53: 99-105
- Beldade, R., Heiser, J.B., Robertson, D.R., Gasparini, J.L., Floeter, S.R. & Bernardi, G. (2009). Historical biogeography and speciation in the Creole wrasses (Labridae, *Clepticus*). *Marine Biology* 156: 679-687
- Bellwood, D.R. & Wainwright, P.C. (2002). The history and biogeography of fishes on coral reefs. In: Sale PF (ed.) Coral reef fishes. Dynamic and diversity in a complex ecosystem. Academic Press, San Diego. Pp. 2-15.
- Benzie, J.A.H. (1999). Genetic structure of coral reef organisms: Ghosts of dispersal past. *American Zoology* 39: 131-145
- Bermingham, E., McCafferty, S.S. & Martin A.P (1997). Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: Kocher T, Stepien C (eds) Molecular Systematics of Fishes. Academic Press, New York. Pp. 113-126
- Bernardi, G., Burcciarelli, D., Costagliola, D., Robertson, D.R., & Heiser, J.B. (2004). Evolution of coral reef fish *Thalassoma* spp. (Labridae). 1. Molecular phylogeny and biogeography. *Marine Biology* 144: 369-375
- Bernardi, G., Holbrook, S.J. & Schmitt, R.J. (2001). Gene flow at three spatial scales in a coral reef fish, the three spot *Dascyllus trimaculatus*. *Marine Biology* 138: 457-465
- Berry, P.F. & Plante, R. (1973) Revision of the spiny lobster genus *Palinurus* in the South-West Indian Ocean. *Transactions of the Royal Society of South Africa* 40:373–380

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., Ingram, K.K. & Das, I. (2006). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 22:148-155

Bird, C.E., Karl, S.A., Smouse, P.E. & Toonen, R.J. (2011). Detecting and measuring genetic differentiation. In: Held, C., Koenemann, S., Schubart, C.D. (eds). *Crustacean Issues 19: Phylogeography and population genetics in Crustacea*. CRC Press, Boca Raton, Florida. Pp. 31-55.

Biton, E., Gildor, H., Trommer, G., Siccha, M., Kucera, M., van der Meer, M.T.J. & Schouten, S. (2010). Sensitivity of Red Sea circulation to monsoonal variability during the Holocene: an integrated data and modelling study. *Paleoceanography* 25: PA4209

Bock, K. (1978). *A guide to common reef fishes of the Western Indian Ocean*. Macmillan, Hong Kong. 122 pp.

Bohonak, A.J. (1999). Dispersal, gene flow and population structure. *The Quarterly Review of Biology* 74: 21-45

Bourjea, J., Lapegue, S., Gagnevin, L., Broderick, D., Mortimer, J.A., Ciccione, S., Roos, D., Taquet, C. & Grizel, H. (2007). Phylogeography of the green turtle, *Chelonia mydas* in the Southwest Indian Ocean. *Molecular Ecology* 16: 175–186

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S. & Robertson, D.R. (2001). Phylogeography of the trumpetfishes (*Aulostomus*): Ring species complex on a global scale. *Evolution* 55: 1029-1039

Bowen, B.W., Muss, A. Rocha, L.A. & Grant, W.S. (2006). Shallow mtDNA coalescence in Atlantic pygmy angelfishes (genus *Centropyge*) indicates a recent invasion from the Indian Ocean. *Journal of Heredity* 97: 1-12

Briggs, J.C. (1992). The marine East Indies: centre of origin. *Global Ecology and Biogeography Letters* 2: 149-156

- Briggs, J.C. (1999). Coincident biogeographic patterns: Indo-West Pacific Ocean. *Evolution* 53: 326-335
- Briggs, J.C. (2005). The marine East Indies: diversity and speciation. *Journal of Biogeography* 32: 1517–1522
- Briggs, J.C. & Bowen, B.W. (2012). A realignment of marine biogeographic provinces with particular reference to fish distributions. *Journal of Biogeography* 39: 12-30
- Briggs, J.C. & Bowen, B.W. (2013). Marine shelf habitat: biogeography and evolution. *Journal of Biogeography* 40: 1023-1035
- Brothers, E.B., Williams, D. McB. & Sale P.F. (1983). Length of larval life in twelve families of fishes at “One Tree Lagoon” Great Barrier Reef Australia. *Marine Biology* 76: 319-324
- Camoin, G.F., Montaggioni, L.F. & Braithwaite, C.J.R. (2004). Late glacial to post glacial sea level in the Western Indian Ocean. *Marine Ecology* 206: 119-146
- Chauhan, T. & Rajiv, K. (2010). Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology* 1: 281-291
- Chiswell, S.M. & Booth, J.D. (1999). Rock lobster *Jasus edwardsii* larval retention by the Wairarapa Eddy off New Zealand. *Marine Ecology Progress Series* 183: 227-240
- Chow, S & Hazama, K. (1998). Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology* 7: 1255-1256
- Chow, S., Okamoto, H., Uozumi, Y., Takeuchi, Y. & Takeyama, H. (1997). Genetic stock structure of the swordfish (*Xiphias gladius*) inferred by PCR-RFLP analysis of the mitochondrial DNA control region. *Marine Biology* 127: 359-367
- Clement, M., Posada, D. & Crandall, K.A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657-1659

Costello, M.J., Coll, M., Danovaro, R., Halpin, P., Ojaveer, P. (2010) A census of marine biodiversity knowledge, resources and future challenges. *PLoS One* 5: e12110

Craig, M.T., Eble, J.A., Bowen, B.W. & Robertson, D.R. (2007). High genetic connectivity across the Indian and Pacific Oceans in the reef fish coral-reef fish. *Myripristis berndti* (Holocentridae). *Marine Ecology Progress Series* 334: 245-254

Craig, M.T., Graham, R.T., Torres, R.A., Hyde, J.R. Hyde., Freitas, M.O., Ferreira. B.P., Hostim-Silva, H., Gerhardinger, L.C., Bertoncini, A.A. & Robertson, D.R. (2008). How many species of goliath grouper are there? Cryptic genetic divergence in a threatened marine fish and the resurrection of a geopolitical species. *Endangered Species Research* 13: 1-8

De Grave, S. (2001). Biogeography of Indo-Pacific Pontoniinae (Crustacea, Decapoda): a PAE analysis. *Journal of Biogeography* 28: 1239-1253

de Ruijter, W.P.M., van Aken, H.M., Beier, E.J., Lutjeharms, J.R.E., Matano, R.P. & Schouten, M.W. (2004). Eddies and dipoles around South Madagascar: formation pathways and large-scale impact. *Deep-Sea Research I* 51: 383-400

Di Battista, J.D., Berumen, M.L., Gaither, M.R., Rocha, L.A., Eble, J.E., Choat, J.H., Craig, M.T., Skillings, D.J. & Bowen, B.W. (2013). After continents divide: comparative phylogeography of reef fishes from the Red Sea and Indian Ocean. *Journal of Biogeography* 40: 1170-1181

Doherty, P.J., Planes, S. & Mather, P. (1995). Gene flow and larval duration in seven species of fish from the Great Barrier Reef. *Ecology* 76: 2373–2391

Donguy, J. & Piton, B. (1991). The Mozambique Channel revisited. *Oceanologica Acta* 14: 549–558

Dorenbosch, M., Pollux, B.J.A., Pustjens, A.Z., Rajagopal, S., Nagelkerken, I., van der Velde, G. & Moon van der Staay, S.Y. (2006). Population structure of Dory snapper, *Lutjanus fulvivflamma*, in the western Indian Ocean revealed by means of AFLP fingerprinting. *Hydrobiologia* 568: 43-53

- Duda, Jr., T.F. & Palumbi, S.R. (1999). Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations. *Marine Biology* 134: 705-710
- Duran, S., Pascaul, M. & Turon, X. (2004). Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge *Crambe crambe*. *Marine Biology* 144: 31-35
- Excoffier, L., Laval, G. & Schneider, S. (2004). Arlequin (version 3.0): an integrated software package for population data analysis. *Evolutionary Bioinformatics Online* 1: 47-50
- Excoffier, L. & Lischer, H. (2010). Arlequin suite (version 3.5): a new series of programs to perform to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567
- Excoffier, L., Smouse, P.E. & Quattro, J.M. (1992). Analysis of Molecular Variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791
- Fowler, A.J., Campana, S.E., Jones, C.M., Thorrold, S.R. (1995). Experimental assessment of the effect of temperature and salinity on elemental composition of otoliths using laser ablation ICPMS. *Canadian Journal of Fisheries and Aquatic Science* 52: 1431-1441
- Fratini, S. & Vannini, M. (2002). Genetic differentiation in the mud crab *Scylla serrata* (Decapoda: Portunidae) within the Indian Ocean. *Journal of Experimental Marine Biology and Ecology* 27: 103-116
- Froukh, T. & Kochzuys, M. (2007). Genetic population structure of the endemic fourline wrasse (*Larabicus quadrilineatus*) suggests limited larval dispersal distances in the Red Sea. *Molecular Ecology* 16: 1359-1367

Fu, X.Y. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915-925

Gadagkar, S., Resenberg, M.S. and Kurmar, S. (2005). Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. *Journal of Experimental Zoology* 304: 64-74

Gaither, M.R., Bowen, B.W., Bordenave, T., Rocha, L.A., Newman, S.J., Gomez, J.A., van Herwerden, L. & Craig, M.T. (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

Gaither, M.R., Toonen, R.J., Robertson, D.R., Planes, S. & Bowen, B.W. (2010). Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific (*Lutjanus kasmira* and *Lutjanus fulvus*). *Journal of Biogeography* 37: 133-174

Gopal, K., Tolley, K.A., Groeneveld, J.C. & Matthee, C.A. (2006). Mitochondrial DNA variation in spiny lobster *Palinurus delagoae* suggest genetically structured populations in the southwestern Indian Ocean. *Marine Ecology Progress Series* 319: 191-198

Gopurenko, D., Hughes, J.M. & Keenan, C.P. (1999). Mitochondrial DNA evidence for rapid colonisation of the Indo-West Pacific by the mudcrab *Scylla serrata*. *Marine Biology* 134: 227–233

Grant, W.S. & Bowen, B.W. (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

Groeneveld, J.C. (2012). Retrospective Analysis of existing data on deep-water trawl-fisheries for crustaceans in the South West Indian Ocean. Specialist Report prepared for the South West Indian Ocean Project. SWIOFP RMU, KMFRI, Mombasa, Kenya, 65 pp.

Gullström, M., de la Torre Castro, M., Bandeira, S.O., Björk, M., Dahlberg, M., Kautsky, N., Rönnbäck, P. & Öhman, M.C. (2002). Seagrass ecosystems in the Western Indian Ocean. *Royal Swedish Academy of Sciences* 31: 588-596

Guo, X. & Chen, D. (2010). Comparative evolution of the mitochondrial cytochrome b gene and nuclear S7 ribosomal protein gene intron I in sinipercid and their relatives. *Hydrobiologia* 649:139-156

Harpending, H.C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* 66: 591-600

Hasegawa, M., Kishino, H. & Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 160-174

Hauser, L & Ward, R.D. (1998). Population identification in pelagic fish: the limits of molecular markers. In: Carvalho GR (ed.). *Advances in Molecular Ecology*. IOS Press: Amsterdam. Pp. 191-224.

Heads, M. (2005). Towards a panbiogeography of the seas. *Biological Journal of the Linnean Society* 84:675-723

Hedgecock, D. (1986). Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? *Bulletin of Marine Sciences* 39: 550-564

Heemstra, P. & Heemstra, E. (2004). *Coastal fishes of Southern Africa*. National Inquiry Service Centre (NISC), South Africa. 488 pp.

Hellberg, M.E. (2009). Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution and Systematics* 40: 291-310

Hellberg, M.E, Burton, R.S., Neigel, J.E., & Palumbi, S.R. (2002). Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* 70: 273-290

Henriksson, O., Mwandya, A.W., Gullström, M., Thorberg, M. & Grahn, M. (2012). Genetic identification and population structure of juvenile mullet (Mugilidae) collected for aquaculture in East Africa. *Western Indian Ocean Journal of Marine Science* 11: 41-54

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature* 405: 907–913

Hoareau, T.B., Boisson, E., Paulay, G. & Bruggemann, J.H. (2013). The Southwestern Indian Ocean as a potential marine evolutionary hotspot: perspectives from comparative phylogeography of reef brittle-stars. *Journal of Biogeography*. 40: 2167-2197

Hoffman, J.I., Dasmahapatra, K.K., Amos, W., Phillips, C.D., Gelatt & Bickham, J.W. (2009). Contrasting patterns of genetic diversity at three different genetic markers in marine mammal metapopulation. *Molecular Ecology* 18: 2961-2978

Hudson, R.R. (2000). A new statistic for detecting genetic differentiation. *Genetics* 155: 2011-2014

Jones, G.P., Millcich, M.J., Erosile, M.J. & Lunow, C. (1999). Self-recruitment in a coral fish population. *Nature* 402: 802-804

Jukes, T.H. & Cantor, C.R. (1969). Evolution of protein molecules. In Munro, H.N. (ed.). *Mammalian protein metabolism, III*. Academic Press, New York. Pp. 21-132

Kairu, K. & Nyandwi, N. (2000). Guidelines for the study of shoreline change in the Western Indian Ocean region. IOC Manuals and Guides No. 40. UNESCO. 55 pp.

Kemp, J.M. (1998). Zoogeography of the coral reef fishes of the Socotra Archipelago. *Journal of Biogeography* 25: 919–933

Klanten, O.S., Choat, H.J. & van der Herwerden, L. (2007). Extreme genetic diversity and temporal rather than spatial partitioning in a widely distributed coral reef fish. *Marine Biology* 150: 659-670

Klausewitz, W. (1972). The zoogeographical and paleogeographical problem of the Indian Ocean and Red Sea according to the ichthyofauna of the littoral. *Journal of the Marine Biological Association of India* 4: 697-706

Knowlton, N. (1993). Sibling species in the sea. *Annual Review of Ecology, Evolution and Systematics* 24:189–216

Knutsen, H., Jorde P.E., Gonzalez, E.B., Robalo, J., Albretsen, J. & Amada, V. (2013). Climate change and genetic structure of leading edge and rear end population in a northwards shifting marine fish species, the corkwing wrasse (*Symphodus melops*). *PLoS One* 8: e67492

Kochzius, M. & Blohm, D. (2005) Genetic population structure of the lionfish *Pterois miles* (Scorpaenidae, Pteroinae) in the Gulf of Aqaba and northern Red Sea. *Gene* 347: 295–301

Kuiter, R.H. (2010). Labridae Fishes: Wrasses. First Edition: Seaford, Australia. Pp. 195-365.

Kuiter, R.H. & Tono-zuka, T. (2001). Pictorial guide to Indonesian reef fishes. Part 2. Fusiliers-Dragonets, Caesionidae-Callionymidae. Zoonetics, Australia. Pp. 304-622.

Kvist, L. (2000). Phylogeny and phylogeography of European sparids. MSc Thesis. University of Oulu. 51 pp.

Lee, W.J., Conroy, J.C., Howell, W.H. & Kocher, T.D. (1995). Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution* 41: 54-66

Leis, J.M. (1991). The pelagic phase of coral reef fishes: larval biology of coral reef fishes. In: P.F Sale (ed.). *The Ecology of Fishes on Coral Reefs*. Academic Press, San Diego. Pp 183-230

Leis, J.M. (2002). Pacific coral-reef fishes: the implications of behaviour and ecology of larvae for biodiversity and conservation, and a reassessment of the open population paradigm. *Environmental Biology of Fishes* 65: 199-208

Leis, J.M. (2010). Ontogeny of behaviour in larvae of marine demersal fishes. *Ichthyological Research* 57: 325-342

Lessios, H.A & Robertson, D.R. (2006). Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings of the Royal Society B* 273: 2201-2208

Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452

Longhurst, A. (1998). *Ecological Geography of the Sea*. Academic Press. United State of America. 398 pp.

Lutjeharms, J.R.E. (2006). The coastal oceans of south-eastern Africa. In: Robinson, A.R., Brink, K.H. (eds) *The Sea. The Global Coastal Ocean. Interdisciplinary Regional Studies and Syntheses*, vol. 14B. Harvard University Press, Cambridge, Massachusetts. Pp. 783-834

Lutjeharms, J.R.E. (2007). Three decade of research on the great Agulhas Current. *Ocean Science* 3: 129-147

Lutjeharms, J.R.E. & Bornman, T.G. (2010). The importance of the greater Agulhas Current is increasingly being recognised. *South African Journal of Science* 106: 1-4

Magori, C. (2008). Tidal analysis and prediction in the Western Indian Ocean. Western Indian Ocean Marine Science Association (WIOMSA) Intergovernmental Oceanographic Commission (of UNESCO), Regional Report. Kenya Marine and Fisheries Research Institute, Mombasa, Kenya. 43 pp.

Masterson, C., Danilowicz, B. & Sale, P. (1997). Yearly and inter-island variation in the recruitment dynamics of the bluehead wrasse (*Thalassoma bifasciatum*, Bloch). *Journal of Experimental Marine Biology and Ecology* 214: 149-166

Meirmans, R.G. (2006). Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60: 2399-2402

Meyer, A. (1993). Evolution of mitochondrial DNA in fishes. In: Hochachka P.W., & Mommesen, T.P. (Ed): Biochemistry and molecular biology of fishes. Molecular Biology Frontiers, Elsevier, Amsterdam. Pp. 1-38

Moritz, C., Dowling, T.E. & Brown, W.M. (1987). Evolution of animal mitochondrial DNA: Relevance for population biology and systematics. *Annual Review of Ecology and Systematics* 18: 269-292.

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

Muths, D., Tessier, E., Gouws, G., Mwale, M., Mwaluma, J., Mwandya, A. & Bourjea, J. (2011). Restricted dispersal of the reef fish *Myripristis berndti* at scale of the SW Indian Ocean. *Marine Ecology Progress Series* 443:167-180

Myers, R.F. (1991). Micronesian reef fishes. Second Edition. Coral Graphics, Barrigada Guam. 298 pp.

Neethling, M., Matthee, C.A., Bowie, R.C.K. & von der Heyden, S. (2008). Evidence for panmixia despite barriers to gene flow in the southern African endemic, *Caffrogobius caffer* (Teleostei: Gobidae). *BMC Evolutionary Biology* 8: 325

Nei, M. (1987). Molecular Evolutionary Genetics. Columbia University Press: New York. 512 pp.

Nei, M. & Tajima, F. (1980). DNA polymorphism detectable by restriction endonucleases. *Genetics* 97: 145-163

New, A.L., Stansfield, K., Smythe-Wright, D., Smeed, D.A., Evans, A.J. & Alderson, S.G. (2005). Physical and biochemical aspects of the flow across the Mascarene Plateau in the Indian Ocean. *Philosophical Transactions of the Royal Society* 363: 151-168

- Ngoile, M. A. & Linden, O. (1997). Lesson learned from Eastern Africa: the development of policy on ICZM at national and regional levels. *Ocean and Coastal Management* 37: 295-318
- Obura, D. (2012). The diversity and biogeography of Western Indian Ocean reef building corals. *PLoS One* 7: e45013
- Palumbi, S.R. (1994). Genetic divergence, reproductive isolation and speciation. *Annual Review of Ecology and Systematics* 25: 547-572
- Palumbi, S.R. (1995). Using genetics as an indirect estimator of larval dispersal. In: McEdward, L. (ed.) *Ecology of Marine Invertebrate Larvae*. CRC Press, Boca Raton, Florida. Pp. 369-388
- Patarnello, T., Volckaert, F.A.M.J. & Castilho, R. (2007). Is the Atlantic-Mediterranean transition a phylogeographical break. *Molecular Ecology* 16: 4426-4444
- Planes, S., Galzin, R. & Bonhomme, F. (1996). A genetic metapopulation model for reef fishes in oceanic islands: the case of the surgeonfish, *Acanthurus triostegus*. *Journal of Evolutionary Biology* 9: 103–117
- Planes, S., Parroni, M. & Chauvet, C. (1998). Evidence of limited gene flow in three species of coral reef fishes in the lagoon of New Caledonia. *Marine Biology* 130: 361–368
- Posada, D. & Crandall, K.A. (2001). Intraspecific gene genealogies: Trees grafting into networks. *Trends in Ecology and Evolution* 16: 37-45
- Quartly, G.D. & Srokosz, M.A. (2004). Eddies in the southern Mozambique Channel. *Deep Sea Research II* 51: 69-83
- Ragionieri, L., Cannici, S., Schubart, S.D. & 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
- Rand, D.M. (1996). Neutrality tests of molecular markers and the connection between DNA polymorphism, demography, and conservation biology. *Conservation Biology* 10: 665-671

Randall, J.E. & Di Battista, J.D. (2013). A new species of damselfish (Pomacentridae) from the Indian Ocean. *Aqua International Journal of Ichthyology* 19:1-21

Randall, J.E. & Miroz, A. (2001). *Thalassoma lunare* x *Thalassoma ruppellii*, a hybrid labrid from the Red Sea. *Journal of Ichthyology and Aquatic Biology* 4: 131–134

Randall, J.E. & Schultz, J.K. (2008). *Cirrhitops mascarenensis*, a new species of hawkfish from the Mascarene Islands, southwestern Indian Ocean. *Smithiana Bulletin* 9: 15-20

Reece, J.S., Bowen, B.W. Joshi, K., Goz, V. & Larson, A. (2010). Phylogeography of two moray eels indicates high dispersal throughout the Indo-Pacific. *Journal of Heredity* 10: 391-402

Richmond, M.D. (2000). The marine biodiversity of the Western Indian Ocean and its biogeography: How much do we know?. Marine Education Awareness and Biodiversity Programme Development. *Western Indian Ocean Marine Science Association*. Zanzibar, Tanzania. Pp. 241-262

Riginos, C. & Victor, B.C. (2001). Larval spatial distribution and other early life-history characteristics predict genetic differentiation in eastern Pacific blennoid fishes. *Proceedings of the Royal Society of London Series B* 268: 1931-1936

Roberts, C.M. (1997). Connectivity and management of Caribbean coral reefs. *Science* 278: 1454–1457

Roberts, C.M., Shepherd, A.R.D. & Ormond, R.F.G. (1992). Large-scale variation in assemblage structure of Red Sea butterflyfishes and angelfishes. *Journal of Biogeography* 19: 239–250

Rocha, L.A. (2003). Pattern of distribution and process of speciation in Brazilian reef fishes. *Journal of Biogeography* 30: 1161-1171

- Rocha, L.A., Rocha, L.S., Robertson, D.R. & Bowen, B.W. (2008). Comparative phylogeography of Atlantic reef fishes indicates both origin and accumulation of diversity in the Caribbean. *BMC Evolutionary Biology* 8: 157
- Rogers, A.R. & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* 9: 552-569
- Rokas, A., Ladoukakis, E. & Zoures, E. (2003). Animal mitochondrial DNA recombination revisited. *Trends in Ecology and Evolution* 18: 411-417
- Salas, E., Molina-Uréna, H., Walter, R.P. & Heath, D.D. (2010). Local and regional genetic connectivity in a Caribbean coral reef fish. *Marine Biology* 157: 437-445
- Sale, P.F. (1980). The ecology of fishes on coral reefs. *Oceanography and Marine Biology* 18: 367-421
- Santini, F. & Winterbottom, R. (2002). Historical biogeography of Indo-western Pacific coral reef biota: is the Indonesian region a centre of origin? *Journal of Biogeography* 29: 189-205
- Sarkar, S & Gupta, A. K. (2009). Deep-sea paleoceanography of the Maldives Islands (ODP Hole 716A), equatorial Indian Ocean during MIS 12- 6. *Journal of Bioscience* 34: 749-764
- Schneider, S. & Excoffier, L. (1999). Estimation of past demographic parameters from the distribution of pairwise difference when mutation rates among sites: application to human mitochondrial DNA. *Genetics* 152: 1079-1089
- Schouten, M.W., de Ruijter, W.P.M. & van Leeuwen, P.J. (2002). Upstream control of Agulhas ring shedding. *Journal of Geophysical Research* 107: 3109
- Shankar, D., Vinayachandran, P.N. and Unnikrishnan, A.S. (2002). The monsoon currents in the north Indian Ocean. *Progress in Oceanography* 52: 63-120
- Sheppard, C. (2000). Seas at the millennium: an environmental evaluation. Pergamon Press, Netherlands. 240 pp.

Sheppard, C. & Wells, S.M. (1988). Coral reefs of the world, Indian Ocean, Red Sea, and Gulf. United Nation Environmental Programme. Nairobi, Kenya. 389 pp.

Shima, J.S. (2001). Regulation of local populations of a coral reef fish via joint effects of density and number dependent mortality. *Oecologia* 126: 58-65

Shulman, M.J. & Bermingham, E. (1995). Early life histories, ocean currents, and the population genetics of Caribbean reef fishes. *Evolution* 49: 897-910

Siddall, M., Rohling, E. J., Almogi-Labin, Hemleben, C., Meischner, D., Schmelzer, I & Smeed, D. A (2003). Sea level fluctuations during the last glacial cycle. *Nature* 423: 853-858

Sinha, B.N. (1994). Geo-economic survey of Lakshadweep. Concept Publishing Company: New Delhi, India. 256 pp.

Sirocko, F. (2003). Ups and downs in the Red Sea. *Nature* 423: 813–814

Stuart, A.S., Vermeij, M.J.A. & Hurlbert, A.H. (2008). Island biogeography of Caribbean coral reef fish. *Global Ecology and Biogeography* 17: 770-777

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution* 15: 200-205

Swallows, J.F., Fieux, M. & Schott, M. (1988). The boundary currents east and north of Madagascar 1. Geostrophic currents and transports. *Journal of Geophysical Research* 93: 4951-4962

Tajima, F. (1989). Statistical method for testing the neutral mutation hypotheses by DNA polymorphism. *Genetics* 123: 585-595

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

Taylor, M.I., Fox, C., Rico, I. & Rico, C. (2002). Species-specific TaqMan probes for simultaneous identification of (*Gadus morhua* L.), haddock (*Melanogrammus aeglefinus* L.) and whiting (*Merlangius merlangus* L.). *Molecular Ecology Notes* 2: 599-601

Templeton, A.R., Routman, E. & Phillips, C.A. (1995). Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the Tiger Salamander. *Ambystoma tigrinum*. *Genetics* 140: 767-782

Thresher, R.E. (1984). Reproduction in Reef Fishes. T.F.H. Publishers: Neptune City, New Jersey, USA. 399 pp.

Tolley, K.A, Groeneveld, J.C, Gopal, K. & Matthee, C.A. (2005). Mitochondrial DNA panmixia in spiny lobster *Palinurus gilchristi* suggests a population expansion. *Marine Ecology Progress Series* 297: 225–231

Toms, J.A., Compton, J.S., Smale, M. & von der Heyden, S. (2014). Variation in palaeo-shorelines explains contemporary population genetic pattern of rocky shore species. *Biology Letters* 10: 2-4

Toonen, R.J., Andrews K.R., Baums, I.B., Bird, C.E., Concepcion, G.T., Daly-Engel, T.S., Eble, J.A., Faucci, A., Gaither, M.R., Iacchei, M., Puritz, J.B., Schultz, J.K., Skillings, D.J., Timmers, M, & Bowen, B.W. (2011). Defining boundaries for ecosystem-based management: a multispecies case study of marine connectivity across the Hawaiian Archipelago. *Journal of Marine Biology* doi 10.1155/2011/460173

Tsang, L., Achituv, Y., Chu, K. & Kan Chan, B.K. (2012). Zoogeography of intertidal communities in the West Indian Ocean as determined by ocean circulation systems: patterns from the *Tetraclita* barnacles. *PLoS One* 7: e45120

Turpie, J.K., Beckley, L.E. and Katua, S.M. (2000). Biogeography and the selection of priority areas for the conservation of South African coastal fishes. *Biological Conservation* 92: 59-72

Van der Elst, R. (1990). Everyone's Guide to the Sea Fishes of Southern Africa. CNA, . 112 pp.

Van der Elst, R., Everett, B., Jiddawi, N., Mwatha, G., Afonso, P.S. & Boulle, D. (2005). Fish, fishers, fisheries of the Western Indian Ocean: their diversity and status. A preliminary assessment. *Philosophical Transactions of the Royal Society* 363: 263-284

Victor, B.C. (1986). Duration of the plankton larval stage of one hundred species of Pacific and Atlantic wrasses (family Labridae). *Marine Biology* 90: 317-326

Visram, S., Cheng M. C., Pillay, R.M., Said, S., Henriksson, O., Grahn., M. & Chen., C.A. (2010). Genetic connectivity and historical demography of the blue barred parrotfish (*Scarus ghobban*) in the western Indian Ocean. *Marine Biology* 157: 1475-1487

Voris, H.K. (2000). Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography* 27:1153–1167

Ward, R.D. (2000). Genetics in fisheries management. *Hydrobiologia* 420:191-201

Ward, R.D. & Grewe, P.M. (1994). Appraisal of molecular genetic techniques in fisheries. *Reviews in Fish Biology and Fisheries* 4: 300-325

Warner, R.R. (1998). The role of extreme iteroparity and risk avoidance in the evolution of mating systems. *Journal of Fish Biology* 55: 82-93

Westneat, M.W. & Alfaro, M.E. (2005). Phylogenetic relationship and evolutionary history of reef fish family Labridae. *Molecular Phylogenetic and Evolution* 36: 370-390

Williams, S.T., Jara, J., Gomez, E. & Knowlton, N. (2002). The marine Indo-West Pacific break: Contrasting the resolving power of mitochondrial and nuclear genes. *Integrative and Comparative Biology* 42: 941-952

Wilson, A.C., Cann, R.L., Carr, S.M., George, M., Gyllensten, U.B., Kathleen, M., Bychowski, H., Higuchi R.G., Palumbi, S.R., Prager, E.M., Sage, R.D. & Stoneking, M. (1985).

Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnaean Society* 26: 367-400

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics* 15: 323-354.

Wright, S. (1978) Evolution and the genetics of populations. A treatise in four volumes. Volume 4. Variability within and among natural populations. University of Chicago Press, Chicago, London, 590 pp.

Zhang, Y., Pham, N.K., Zang, H., Lin, J. & Qiang, L. (2014). Genetic variation in two seahorse species (*Hippocampus mohnikei* and *Hippocampus trimaculatus*): evidence for middle Pleistocene population expansion. *PLoS One* 9: e105494

Zhang, Z. & Hewitt, G.M. (2004). Nuclear DNA analyses in genetic studies of populations: practice problems and prospects. *Molecular Ecology* 12: 563-584

APPENDIX ONE

HAPLOTYPE AND ALLELE FREQUENCIES

A1: Distribution of (A) ATPase 6 and (B) cytochrome *b* haplotypes, and (C) S7 intron I alleles of *Cheilio inermis* from several localities across Western Indian Ocean and Philippines (Pacific). Locality abbreviations are according to Table 3.2. Haplotypes with the highest frequency are highlighted in bold.

A

Haplotype	MA	TN	ZN	MZ	GZ	SM	DJ	PP	N
H1	1			3	1	2	-	1	8
H2	-	2	1	3	1	1	-	-	8
H3	-	1			-	-	-	-	1
H4	-	-	1	3	1	2	-	-	6
H5	-	-	1	12	-	5	1	-	19
H6	-	-	1	-	-	-	-	-	1
H7	-	-	-	1	-	-	-	-	1
H8	-	-	-	1	-	-	-	-	1
H9	-	-	-	1	-	-	-	-	1
H10	-	-	-	1	-	-	-	-	1
H11	-	-	-	1	-	-	-	-	1
H12	-	-	-	1	-	-	-	-	1
H13	-	-	-	1	-	-	-	-	1
H14	-	-	-	1	-	-	-	-	1
H15	-	-	-	1	-	-	-	-	1
H16	-	-	-	1	-	-	-	-	1
H17	-	-	-	1	-	-	-	-	1
H18	-	-	-	1	-	1	-	-	2
H19	-	-	-	1	-	-	-	-	1
H20	-	-	-	-	1	-	-	-	1
H21	-	-	-	-	-	2	-	-	2
H22	-	-	-	-	-	1	-	-	1
H23	-	-	-	-	-	1	-	-	1
H24	-	-	-	-	-	1	-	-	1
H25	-	-	-	-	-	1	-	-	1
Overall (N)	1	3	4	34	4	17	1	1	65

B

Haplotype	MD	MZ	GZ	SM	TN	ZN	DJ	<i>N</i>
H1	1	-	-	-	-	-	-	1
H2	2	-	1		-	-	-	3
H3	-	2	-	1	-	-	-	3
H4	-	1	-	-		-	-	1
H5	-	2	-	-	-		-	2
H6	-	1	-	-	-	-	-	1
H7	-	1	-		-	-	-	1
H8	-	1	-	1	-	-	-	2
H9	-	1	-	-	-	-	-	1
H10	-	2	-	-	-	-	-	1
H11	-	1	-	-	2	1	-	4
H12	-	2	2	8	-	1	-	13
H13	-	1	-	-	-	-	-	1
H14	-	1	-	-	-	-	-	1
H15	-	1	-	-	-	-	-	1
H16	-	1	-	-	-	-	-	1
H17	-	1	-	-	-	-	-	1
H18	-	1	-	-	-		-	1
H19	-	1	-	-	-	-	-	1
H20	-	1	-	-	-	-	-	1
H21	-	1	-	-	-	-	-	1
H22	-	6	1	1	-	-	-	9
H23	-	1	-	-	-	-	-	1
H24	-	1	-	-	-	-	-	1
H25	-	-	-	1	-	-	-	1
H26	-	-	-	1	-	-	-	1
H27	-	-	-	2	-	-	-	2
H28	-	-	-	1	-	-	-	1
H29	-	-	-	1	-	-	-	1
H30	-	-	-	1	-	-	-	1
H31	-	-	-	1	-	-	-	1
H32	-	-	-	1	-	-	-	1
H33	-	-	-	-	1	-	-	1
H34	-	-	-	-	-	1	-	1
H35	-	-	-	-	-	1	-	1
H36	-	-		-	-	-	1	1
Overall (<i>N</i>)	3	31	4	20	3	4	1	66

C

Alleles	SY	TN	ZN	MZ	GZ	SM	N
A1	1	-	-	-	-	-	1
A2	1	1	1		1	-	4
A3	-	1	-	2	-	2	5
A4	-	1	1	1	-	-	3
A5	-	1	1	3	-	1	4
A6	-	-	2	1	-	-	3
A7	-	-	1	1	-	-	1
A8	-	-	-	3	-	-	3
A9	-	-	-	2	-	-	2
A10	-			1	-	-	1
A11	-		-	1	2	-	4
A12	-			1	-	-	1
A13	-		-	1	-	-	1
A14	-		-	2	-	-	2
A15	-		-	1	-	-	1
A16	-		-	3	-	-	3
A17	-		-	1	-	-	1
A18	-		-	1	-		1
A19	-		-	1	-	1	2
A20	-		-	2	-	-	2
A21	-		-	1	-	-	1
A22	-		1	1	-	-	2
A23	-		-	1	-	-	1
A24	-		-	1	1	-	2
A25	-	-	-	3	-	1	4
A26	-	-	-	1	-	-	1
A27	-	-	-	1	-	-	1
A28	-	-	-	1	-	-	1
A29	-	-	-	1	-	1	2
A30	-	-	-	1	-	1	2
A31	-	-	-	1	-	1	2
A32	-	-	-	1	-	1	2
A33	-	-	-	-	1	1	2
Overall (N)	2	4	6	36	6	10	64

A2: Distribution of (A) ATPase 6 and (B) cytochrome *b* haplotypes, and (C) S7 intron I alleles of *Thalassoma hebraicum* from numerous localities across Western Indian Ocean. Locality abbreviations are according to Table 4.1. Haplotypes with the highest frequency are highlighted in bold.

A

Haplotype	SA	MH	AD	ZN	SM	<i>N</i> per haplotype
H1	1	-	1	-	-	2
H2	1	-	-	-	-	-
H3	2	-	1	-	3	6
H4	1			1	1	3
H5	-	1	-	-	-	1
H6	-	1	-	-	-	1
H7	-	-	1	-	-	1
H8	-	-	1	-	-	1
H9	-	-		1	-	1
H10	-			1	-	1
H11	-		-		1	4
H12	-			-	1	1
H13	-		-	-	1	1
H14	-		-	-	2	2
H15	-		-	-	1	1
H16	-		-	-	1	1
H17	-		-	-	1	1
H18	-		-	-	1	1
Overall (<i>N</i>)	5	2	4	3	13	27

B

Haplotypes	SA	MH	AD	ZN	SM	NB	<i>N</i> per haplotype
H1	-	-	-	1	-	-	1
H2	-	-	-	1	3	2	6
H3	3	1	1	2	3	-	10
H4	-	-	1	-	-	-	1
H5	1	-	1	-	2	-	4
H6	-		1	-	-	-	1
H7	-	-		-	1	-	1
H8	-	-		1	-	-	1
H9	-	-		1	-	-	1
H10	-	-		1	2	-	3
H11	1	-	-	-	-	-	1
H12	-	-		-	1	-	1
H13	-	-	-	-	2	-	2
H14		-	-	-	1	-	1
Overall (<i>N</i>)	5	1	4	7	15	2	34

C

Alleles	SA	MH	AD	ZN	SM	N
A1	-	-	-	2	-	2
A2	-	1		-	-	1
A3		1	2	-	-	3
A4	1	-	-	-	-	1
A5	-	-	-	-	1	-
A6	-	-	-	-	1	1
A7	-	-	-	-	2	2
A8	-	-	-	-	1	1
A9	-	-	-	-	1	1
A10	-	-	-	-	1	1
A11	-	-	-	-	1	1
A12	-	-	-	-	1	1
A13	1	-	-	-	-	1
A14	1	-	-	-	-	1
A15	1	-	-	-	-	1
A16	1	-	-	-	-	1
A17	1	-	-	-	-	1
A18	2	-	-	-	-	2
A19	-	-	-	-	1	1
A20	-	-	-	-	1	1
A21	-	-	-	-	2	2
A22	-	-	-	-	1	1
A23	-	-	-	-	1	1
A24	-	-	-	-	1	1
A25	-	-	-	-	1	1
A26	-	-		-	1	1
Overall (N)	8	2	2	2	18	32

A3: Distribution of (A) ATPase 6 haplotypes and (B) S7 intron I alleles of *Thalassoma lunare* from localities across Western Indian Ocean, the Red Sea and Pacific. Localities abbreviations are according to Table 4.1. Haplotypes with the highest frequency are highlighted in bold.

A

Haplotype	SA	MH	AD	ZN	RD	N
H1	-	-	-	-	-	-
H2	-	-		-	-	-
H3	-	-	-	-	-	-
H4	-	-	-	-	1	1
H5	-	-	-	-	5	5
H6	-	-	-	-	1	1
H7	-	1	-		-	1
H8	-	-	1	-	-	1
H9	3	-	-	1	-	1
H10	-	-	-	2	-	2
Overall (N)	3	1	1	3	7	15

B

Haplotype	SM	MD	RD	PI	FP	<i>N</i>
H1	-	-	-	-	2	2
H2	-	1		3	-	4
H3	-	-	-	1	-	1
H4	2	-	-	-	-	3
H5	-	1	-	-	-	6
H6	-	-	1	-	-	2
H7	-	-	1	-	-	1
H8	-	-	1	-	-	1
H9	-	-	2	-	-	2
H10	-	-	1	-	-	1
H11	-	-	1	-	-	1
H12	-	-	1	-	-	1
H13	-	-	1	-	-	1
H14	-	-	1	-		1
Overall (<i>N</i>)	2	2	10	4	2	20