Comparative phylogeography of four Indo–Pacific scarine labrids: an insight into the evolutionary patterns of reef fish

Thesis submitted by
Emadch Beck (MAppSc), James Cook University, Qld,
BA(Hons), Scripps College,
in January 2013

For the degree of Doctor of Philosophy
School of Marine and Tropical Biology
James Cook University

Statement of access

I, the undersigned and the author of this work, understand that James Cook University will make it available for use within the University Library and, via the Australian Digital Theses Network, for use elsewhere.

I understand that, as an unpublished work, a thesis has significant protection under the *Copyright Act 1968*.

I do not wish to place any restrictions on access to this work.

_____ _____

Signature Date

Statement of sources declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text, and a list of references is given.

_____ _____

Signature Date

Electronic copy declaration

I, the undersigned and author of this work, declare that the electronic copy of this thesis provided to the James Cook University Library is an accurate copy of the print thesis submitted, within the limits of the technology available.

_____ _____

Signature Date

Statement on the contribution of others

This thesis would not have been possible without the help of numerous individuals and organisations.

Research funding:

- Queensland-Smithsonian Institution collaboration
- National Geographic Society (Cocos [Keeling] and Marquesas islands)
- Research Program on Reef Fishes
- Smithsonian Institution
- Seychelles Fishing Authority
- Cocos Keeling Islands National Park and Christmas Island National Park
- Australian Government Department of the Environment, Water, Heritage and the Arts
- Taiwan National Museum
- Australian Institute of Marine Science
- Lizard Island Research Station
- University of Guam Marine Laboratory
- University of Hawaii Marine Laboratory
- James Cook University internal funding scheme

Research in-kind support:

- Winifred Violet Scott Trust

Specimen collections:

- J Ackerman
- M Bariche
- D Bellwood
- L Chen
- JH Choat
- D Clumpner
- M Emslie
- OS Klanten
- M McCormick
- M Meekan
- K Rhodes
- W Robbins
- DR Robertson
- A Sabetian
- L van Herwerden

Permits:

- Collections on the GBR were carried out under GBRMPA Permit number G03/3871.1.
- This work was carried out under James Cook University Ethics Approval No. A503.

Laboratory and analytical support:

- OS Klanten
- L van Herwerden
- JP Robinson
- JCU Molecular Ecology and Evolution Laboratory

Editorial assistance:

- L van Herwerden
- OS Klanten
- JH Choat
- Biotext

Supervision:

- Dr L van Herwerden
- Dr JH Choat
- Dr OS Klanten

Acknowledgements

I would like to thank my supervisors Dr. Lynne van Herwerden, Dr. O Selma Klanten and Professor J Howard Choat for their invaluable knowledge and support during my PhD. Many thanks to my parents Stuart and Tulik Beck for their financial and moral support, and for allowing me to fly so far away from my nest to follow my dreams. I am also grateful to my friends and family for their continuing support and motivation, and for keeping me sane throughout this (longer-than-expected) journey. My sincerest apologies to anyone I have neglected during this journey, especially my friends back home and my beloved family in Palau. I would also like to express my appreciation for my husband Andy MacNee for providing me with endless capsicums and unwavering support, and for reminding me of life's finer pleasures. And thanks to my two beautiful daughters Coco Bella and India Rose for helping me to put things into perspective – you have been most welcome distractions. Finally, I would like to express my gratitude to the late Dr. D John Faulkner, who believed in me even when I didn't, and who continues to inspire me even in his afterlife; without him, I would never have had the courage to embark on this amazing journey.

Abstract

This thesis examines the genetic structure of four widely distributed Indo–Pacific parrotfish populations using a combination of phylogenetic and phylogeographic analyses. These data were used to identify spatial and temporal patterns of population structure, genetic diversity and the underlying historical processes for each species. The four species *Chlorurus sordidus*, *Scarus ghobban*, *S. rubroviolaceus* and *S. psittacus* have relatively recent evolutionary origins, ranging between 2, (*C. sordidus*), 3 (*S. rubroviolaceus*), and 4 million years (*S. ghobban* and *S. psittacus*) (Alfaro et al. 2009). These species also differ in their association with coral reefs; of the four species, *S. psittacus* is most capable of occupying shallow water habitats devoid of coral reefs, while *S. ghobban* can extend into much deeper water (up to 250 m) than the other three species. Sequences from the mitochondrial control region for the four species collected from their Indo–Pacific distributions were compared. The sampling data from Indo-Pacific wide collections are: *C. sordidus*, 354 bp, 351 individuals, 18 locations; *S. rubroviolaceus*, 378 bp, 292 individuals, 15 locations; *S. ghobban*, 350bp, 239 individuals, 12 locations; and *S. psittacus*, 322 bp, 164 individuals, 12 locations. The comparative analyses include data from Bay et al. (2004) on *C. sordidus* and Winters et al. (2010) on *S. psittacus*.

For each species, we used Bayesian and neighbour-joining analyses to generate the best tree topology and trees were outgroup rooted using relevant sister taxa. The phylogenetic relationships were also represented as a minimum-spanning haplotype tree to visualise genetic diversity and its spatial distribution. To determine the level of genetic differentiation between individuals from different sampling locations, we used pairwise F_{ST} comparisons. Pairwise F_{ST} comparisons were also made between clades identified in the phylogenetic analyses. Estimates of genetic and geographical distances between populations were used to assess isolation by distance. We analysed the molecular variance to determine the source of greatest genetic variation with the analyses structured as follows: (1) geographic location with populations grouped into their respective ocean basin, (2) populations partitioned into geographic regions and (3) groupings based on clade structure established in the phylogenetic analysis. The coalescence-based program Migrate 2.3 was used to infer migration rates and directions between sampled locations. The timing of divergence between geographically defined populations, as well as phylogenetically determined clades, was determined using coalescence analyses that were calculated in Arlequin 3.01.

Phylogenetic analyses revealed evidence of genetic partitioning between western Indian Ocean and Pacific Ocean populations for three of the four species. This was confirmed by significant

pairwise F_{ST} comparisons for *C. sordidus* ($F_{ST\ IO–PO}$ = 0.594–0.762, $P < 0.001$), *S. rubroviolaceus* ($F_{ST\ IO–PO}$ = 0.5290–0.8528, $P < 0.0001$) and *S. ghobban* ($F_{ST\ IO–PO}$ = 0.748–0.803, $P \leq 0.001$), but not for *S. psittacus* ($F_{ST\ IO–PO}$ = 0.046–0.166, $P < 0.01$). For all species, individuals from Western Australia were associated with the Pacific Ocean clade, despite geographically located in the Indian Ocean. A genetic break associated with the central Pacific barrier was also found in *C. sordidus* and evidence of a third genetically distinct population at Cocos Keeling Island was detected in *S. ghobban* ($F_{ST\ IO–CK}$ = 0.754–0.762, $P < 0.001$; $F_{ST\ PO–CK}$ = 0.766–0.851, $P \leq 0.009$). We report the location of the genetic break between western Indian Ocean and Pacific Ocean populations to be at Cocos Keeling Island for *S. ghobban*, Christmas Island for *C. sordidus* and further west again for *S. rubroviolaceus*.

Localised population structure was identified at peripheral locations for all four species. Specifically, Hawaiian populations were genetically differentiated in *C. sordidus*, *S. rubroviolaceus* and *S. psittacus*; Marquesan/French Polynesian populations were differentiated in *C. sordidus*, *S. rubroviolaceus* and *S. psittacus*; east Pacific populations were differentiated in *S. rubroviolaceus* and *S. ghobban*; Western Australian populations were differentiated in *S. ghobban*; and finally, Arabian Gulf and northern Oman populations were differentiated in *C. sordidus*. Isolation by distance was only detected in *S. rubroviolaceus* and was only significant within the ocean basin (Z_{IO} = 290.5452, r = 0.6856, R^2 = 0.470, $P < 0.05$; Z_{PO} = 51568.0352, r = 0.5428, R^2 = 0.295, $P < 0.05$). Migration estimates indicate largely uneven gene flow that was predominantly from east to west for all but *S. psittacus* and could be generally explained by present-day oceanographic currents. Migration analysis was not possible for *S. psittacus*. Coalescence calculations for the three species indicate that the timing of divergence between extant western Indian Ocean and Pacific Ocean populations of *C. sordidus* and *S. rubroviolaceus* both took place during the Pleistocene approximately 1 million years ago (mya), and that the divergence of extant populations of *S. ghobban* took place prior to that, approximately 2.4 mya. These coalescent ages were an order of magnitude older than the coalescent age of *S. psittacus*.

Aspects of genealogical concordance were demonstrated in three of the four species in this study. Aspect I was evidenced by strong bootstrap support for at least two distinct lineages. Aspect II was evidenced by support for the lineages in other studies employing independent molecular markers. Aspect III of genealogical concordance was evidenced by congruent patterns of phylogenetic structure across the three codistributed species which distinguished all west Indian Ocean individuals from the east Indo-Pacific individuals. Aspect IV was evidenced by the same strongly supported lineages separated at the same biogeographic area: the Cocos Keeling and Christmas Islands.

Despite a lack of phylogenetic structure at the largest spatial scale in *S. psittacus*, we noted that all four species had the following in common: a central population containing individuals from the eastern Indian Ocean and the central Pacific Ocean, and several smaller populations containing individuals from peripheral locations. Therefore, we inferred that isolation at peripheral locations is responsible for lowered genetic diversity and is a major force behind generating the genetic differences that ultimately lead to evolutionary novelty and reduced genetic diversity in widespread scarine labrids. Furthermore, overlap at central locations, such as at Christmas and Cocos Keeling islands, contributes to the increase in genetic diversity, which is the building block for adaptation under selection during times of environmental flux. The contrasting pattern of population structure at the level of ocean basins observed between *S. psittacus* and the other three species can be explained by smaller ancestral populations during Plio-Pleistocene low sea level stands owing to its higher level of habitat specificity compared with the other three species. Given enough time, and a large enough population, we would expect to see similar levels of partitioning at the Indo – West Pacific Barrier. Finally, we conclude that populations fluctuate substantially over evolutionary timescales, and different species experience these fluctuations at different times, irrespective of the evolutionary ages of the species, as evident from the variable coalescent ages of extant populations. These differences are likely associated with inherent differences in ecology.

Contents

Tables

Figures

Chapter 1 General Introduction

1.1 Biodiversity in the marine environment

Tropical coral reefs host more marine species than any other marine environment; however, the origin of that biodiversity remains a topic of much debate. In the tropical Indo–Pacific, the area known as the 'coral triangle' or the Indo–Australian Archipelago, has received much attention among coral-reef scientists due to its unprecedented marine biodiversity; this area hosts more species of corals (Hughes *et al.*, 2002), gastropods, crustaceans (Hoeksema, 2007) and reef fishes (Bellwood, Hughes, 2001) than anywhere else in the world. Since this pattern of increasing species richness in the Indo–Australian Archipelago was first observed in reef-building corals (Stehli, Wells, 1971), numerous models have been proposed, and can be summed up into three major categories: centre of origin, centre of overlap/accumulation and centre of survival. These models are described below:

i. **Centre of origin model**: speciation occurs within the centre (i.e. the Indo–Australian Archipelago) with subsequent radiation to peripheral areas. This model predicts that short-range neo-endemic species from the centre of the distribution (i.e. the Indo–Australian Archipelago) will be younger than those from the periphery or edge of the distribution (Briggs, 1999; Briggs, 2000; Ekman, 1953; Mora *et al.*, 2003). This hypothesis also assumes that Pleistocene glaciations are responsible for this pattern, and thus endemics should have originated less than 2 million years ago (mya) (Barber, Bellwood, 2005; Mora *et al.*, 2003).

ii. **Centre of accumulation/overlap model**: speciation occurs outside the central region followed by inward dispersal resulting in species accumulating in the Indo–Australian Archipelago from neighbouring areas (Jokiel, Martinelli, 1992; Ladd, 1960). Alternatively, vicariance separates previously widespread populations into Indian Ocean and Pacific Ocean populations, which are subsequently reunited in the Indo–Australian Archipelago following range expansions (Barber *et al.*, 2000; Bellwood, Wainwright, 2002; Santini, Winterbottom, 2002; Woodland, 1983). The centre of accumulation model predicts that new endemics in the Indo–Australian Archipelago will be older than those outside the Indo–Australian Archipelago. As with the centre of origin model, new endemics should be younger than 2 million years (my) (Barber, Bellwood, 2005).

iii. **Centre of survival/refuge model**: species accumulate in the centre as a result of greater extinction rates in peripheral locations outside the Indo–Australian Archipelago (Barber, Bellwood, 2005; McCoy, Heck, 1976). This model does not make specific predictions about relative ages of species inside and outside the Indo–Australian Archipelago.

Many early attempts to test these models used endemicity as a proxy for speciation events in the marine environment (Briggs, 2000; Briggs, 2005; reviewed in Hoeksema, 2007; Mora *et al.*, 2003). However, peripheral areas and isolated oceanic islands seemed to harbour more endemics than central locations for a number of marine organisms (Bellwood, Wainwright, 2002; Connolly *et al.*, 2003; Hughes *et al.*, 2002; Roberts *et al.*, 2002). Moreover, Bellwood and Meyer (2009) found that endemicity was not always a useful measure for species origination, particularly for species with high dispersal abilities and widespread geographic ranges, a characteristic in many reef fish species. Bellwood and Meyer (2009) concluded that the Indo–Australian Archipelago was important in the survival rather than the origin of marine species, and that the study of endemics would be more useful in testing the three models of increasing species richness if it combined phylogenetic with phylogeographic analyses of cosmopolitan sister species across their ranges.

While there have been many studies published in support of the Indo–Australian Archipelago centre of origin (Mora *et al.*, 2003) and overlap theories (Bellwood, Wainwright, 2002; Santini, Winterbottom, 2002), there is increasing empirical evidence that supports peripheral and isolated habitats as cradles of marine biodiversity (Bellwood, Wainwright, 2002; Budd, Pandolfi, 2010; Dawson, Hamner, 2005; Hardie, Hutchings, 2010). The role of peripheral isolation in generating evolutionary novelty has been reported for hermit crabs (Malay, Paulay, 2009), coconut crabs (Lavery *et al.*, 1996b), the marine gastropod *Conus miliaris* (Duda, Lee, 2009), reef corals (Budd, Pandolfi, 2010; Pandolfi, 1992), reef fishes (Rocha, Bowen, 2008) as well as a number of temperate marine species (Johannesson, André, 2006). It has been suggested that the reduced gene flow, in addition to different and often greater selective pressures in these peripheral habitats, may play a role in the evolution and maintenance of biological diversity (Carson, Templeton, 1984; Templeton, 1980). Furthermore, many of the ‘peripheral’ habitats in the tropical marine environment are marginal, nonreefal habitats, which are considered to be vital in the evolutionary novelty of reef fish (Bellwood, Wainwright, 2002). Despite lowered levels of genetic diversity in peripheral and isolated populations, these are the areas that host the greatest number of endemic species per unit area (Allen, 2007), and therefore merit conservation attention (Budd, Pandolfi, 2010; Nunes *et al.*, 2009).

In order to test adequately the theories of Indo–Australian Archipelago biodiversity, it is necessary to understand present-day and historical gene flow between populations of dispersive species both inside and outside the Indo–Australian Archipelago. Using a comparative phylogeographic approach—as proposed by Funk and Omland (2003), which incorporates closely related and widely distributed species with similar distributions spanning known biogeographic barriers—will enable us to generate comparable datasets while helping to reduce some of the noise that would be generated by innate differences between species.

1.2 Comparative phylogeography of marine organisms

Many recent comparative phylogeographic studies on widespread marine species have demonstrated conflicting patterns of population structure among widely distributed tropical marine fauna (Avise, 2000). Some species exhibit distinctive populations on either side of an established marine biogeographic barrier (Lessios *et al.*, 2001; McMillan, Palumbi, 1995) and others exhibit panmictic populations that appear to be unperturbed by those barriers (Lessios *et al.*, 1998; Lessios, Robertson, 2006). These contrasting patterns of population structure have been observed even among closely related members of the same family in the Indo–Pacific, including gastropods (Crandall *et al.*, 2007), lutjanids (Gaither *et al.*, 2010), *Echinolittorina* spp. snails (Reid *et al.*, 2006), tropical sea urchins from the genus *Echinometra* (Palumbi *et al.*, 1997) and scarine labrids (Bay *et al.*, 2004; Winters *et al.*, 2010). Two separate studies on the seahorse *Hippocampus kuda* using a localised (Lourie *et al.*, 2005) and widespread sampling scheme (Teske *et al.*, 2005) produced vastly different results, emphasising the importance of a complete sampling design in a comparative phylogeographic approach.

In addition to these conflicting patterns, some studies have shown purportedly widespread species to comprise a complex of cryptic species, or species that are morphologically indistinguishable, but genetically distinct (Knowlton, 1993). Cryptic speciation has been demonstrated for a number of species, including crabs (Gopurenko, 1999; Lavery *et al.*, 1996b), mantis shrimps (Barber *et al.*, 2006), seastars (Benzie, 1999b; Vogler *et al.*, 2008; Williams, Benzie, 1998), bonefishes (Colborn *et al.*, 2001) and several species of reef fish (Bernardi *et al.*, 2002; Drew *et al.*, 2008; Hyde *et al.*, 2008; Kon *et al.*, 2007; Leray *et al.*, 2010). These studies, which increasingly document the incidence of cryptic speciation, undermine the previous generalisation that marine species with dispersive larval phases have the capacity to be widespread (Scheltema, 1986; Shanks *et al.*, 2003; Siegel *et al.*, 2003).

Molecular genetics have proven extremely useful in assessing population structure and genetic connectivity between widespread populations; however, few studies integrate phylogenetic

approaches with phylogeographic and population genetic approaches (Avise *et al.*, 1987; Brito, Edwards, 2008; Templeton, 2001). While the goal of phylogeography is to determine the underlying historical processes that are responsible for present day patterns of distribution, the aim of phylogenetics is to determine the presence of genetically distinct lineages and the evolutionary relationship between them (Avise, 2000). Although these two fields have their own sets of questions and methodologies, they share the underlying goals of determining the relationships between individuals and how they are affected by historical processes.

Given that so many phylogeographic studies have shed light on the presence of population structure in widespread marine species, having a sound understanding of the evolutionary relationships between spatially partitioned populations is therefore crucial in understanding such structure. Population genetics analyses generally rely on *a priori* groupings based on present day geographic distributions; however, because of the volatile history of coral reefs, particularly over the last 1.5 million years, present distributions are not necessarily geography based, but rather a reflection of historical patterns of fragmentation and subsequent remixing. Thus, population genetics based purely on present-day geography may not accurately represent the relationship between individuals at the time of divergence. Phylogenetics has been used to identify groups of individuals separated in time, or temporal clades, thereby providing information on which haplotypes may have been derived from a single gene pool in the past without the specific constraint of spatial partitioning (Horne *et al.*, 2008; Klanten *et al.*, 2007). With this information, population genetics analyses can then be used to explore the evolutionary history and connectivity of populations in both space and time. This is particularly valuable information for marine species in light of the tremendous capacity for dispersal in the marine environment, given the history of isolation and secondary contact between widespread reef populations. In addition, connectivity at evolutionary timescales may not represent connectivity at contemporary, ecological timescales. There has been recent increasing focus on ecological approaches to identify connectivity between populations at small spatial scales. These approaches have identified unexpectedly high levels of self-recruitment within ecological time frames (Almany *et al.*, 2007; Jones *et al.*, 2005; Jones *et al.*, 1999; Swearer *et al.*, 1999). Therefore, because coral reefs are patchily distributed over wide spatial scales and have undergone markedly different histories, an evolutionary perspective is required to understand the patterns of distribution in widespread coral reef taxa.

A study conducted by Bay et al. (2004) on the widespread Indo–Pacific scarine labrid *Chlorurus sordidus*, which incorporates both phylogenetic and population genetic methods, found evidence of genetic partitioning between western Indian Ocean and Pacific Ocean populations. In contrast, Winters et al. (2010) found no evidence of phylogenetic partitioning between

western Indian Ocean and Pacific Ocean populations in a similar scarine labrid, *Scarus psittacus*, but did find distinctive populations at peripheral locations within ocean basins. This doctoral thesis is an extension of the study by Bay et al. (2004) and combines data from three additional species, including published results by Winters et al. (2010), to allow for a comprehensive comparative phylogeographic study of widely distributed Indo–Pacific scarine labrids.

1.3 Study species

Scarine labrids are one of the most diverse and recognisable families of fish on tropical coral reefs (Sale, 1991) and many are widespread throughout the Indo–Pacific (Choat, Bellwood, 1994). They comprise more than 90 species divided into 10 genera, and have undergone a very recent rapid evolutionary radiation, resulting in over 50% belonging to the most recent genus, *Scarus* (Smith *et al.*, 2008; Streelman *et al.*, 2002). The four species selected for this study (*Chlorurus sordidus*, *Scarus rubroviolaceus*, *S. ghobban*, and *S. psittacus*) represent four of the five lineages of the most recently diverged and most speciose scarine labrids—the less diverse *Chlorurus* clade and three of the four *Scarus* clades identified to date (Alfaro *et al.*, 2009; Choat *et al.*, 2012). All four study species are widespread and abundant across the Indo–Pacific, from the east coast of Africa to the central Pacific Ocean, with *S. rubroviolaceus* and *S. ghobban* extending into the tropical east Pacific. Besides being widespread throughout the Indo–Pacific, they possess four key characteristics that render them ideal for a comparative phylogeographic study. These characteristics are discussed below:

i. **Strong reef association**. In addition to the strong reef association of most members of the genera *Scarus* and *Chlorurus*, some are also found in a variety of marginal, nonreefal habitats throughout their distributions; for example, *S. ghobban* frequents deeper water habitats up to 250 m (Rees *et al.*, 1994), as well as nonreefal habitats in the Mediterranean Sea (Bariche, Saad, 2005).

ii. **Recent evolutionary origins**. From an evolutionary perspective, scarine labrids emerged relatively recently (20-36 mya, Alfaro *et al.*, 2009), and the four study species originated as recently as the last 2–4 million years (Alfaro *et al.*, 2009; Bellwood, 1994).

iii. **Similar pelagic larval durations**. On the whole, parrotfish have moderate pelagic larval durations of between 30 and 40 days (Chen, 1999), and thus have a similar and substantial capacity for dispersal.

iv. **Newly diverged sister taxa in two species**. Firstly, *S. ghobban* has a newly diverged sister, *S. compressus*, which diverged as recently as during the last 0.15–0.18 my (Choat *et al.*, 2012) and is confined to the tropical east Pacific (Bruce, Randall, 1983). Preliminary results indicate morphological dissimilarity between these two sister species (Choat, pers. comm). Secondly, in the western Indian Ocean, *S. rubroviolaceus* has two newly diverged sister species: *S. persicus*, which diverged between 2.06 and 2.28 mya (Choat *et al.*, 2012), and is confined to northern Oman and the Persian Gulf (Bruce, Randall, 1983); and *S. ferrugineus*, which diverged between 1.66 and 1.96 mya (Choat *et al.*, 2012), and is confined to the Red Sea and the Gulf of Aden (Parenti, Randall, 2000). These sister taxa have similar meristics and colour patterns in their initial phases but differ markedly in their adult phases (Choat, pers. comm.).

The similar cosmopolitan distributions of these four species, combined with their similar potential for dispersal, strong association to coral reef habitats, relatively recent origins and the presence of newly diverged sister species in two of the four species, provide us with a good opportunity to test hypotheses of Indo–Australian Archipelago biodiversity. We can also gain better insights of the underlying processes during the early stages of reef fish diversification and the evolution of population structure in widespread reef fish.

1.4 Mitochondrial markers

In this study, I will compare sequences from the mitochondrial control region to compare and contrast patterns of population structure between the four species. The decision to use mitochondrial markers is based on the fact that they evolve more rapidly than nuclear markers, thus allowing for any population differentiation to be detected more readily than would be possible if using nuclear markers. This is particularly important for comparisons of individuals within a species. Although many researchers caution the use of single markers in phylogeographic studies, the use of these markers is validated for this study (see Zink, Barrowclough, 2008). Furthermore, others have shown that comparative phylogeography of multiple confamilial species with similar evolutionary and biological characteristics will be a powerful, informative approach to examining the patterns and processes involved in generating differentiation and diversification on coral reefs (e.g. Barber *et al.*, 2006; Funk, Omland, 2003; Horne *et al.*, 2008; Palumbi *et al.*, 1997).

1.5 Aims

This thesis presents phylogenetic and phylogeographic data from three widespread parrotfish species from throughout their Indo–Pacific distribution ranges (chapters 3–5). A final comparative chapter (Chapter 6) incorporates data from a previously published study on the scarine labrid *S. psittacus* by Winters et al. (2010), with data presented in chapters 3–5 of this thesis.

In this thesis, I address the following questions:

1) Are there congruent patterns of population structure in co-distributed confamilial widespread reef fish across the Indo–Pacific? If population structure exists, is it associated with known marine biogeographic barriers?

2) What role do peripheral or isolated environments (or both) play in the differentiation and eventual diversification of reef fish?

3) Do widely distributed Indo–Pacific scarine labrids support a centre of origin, accumulation/overlap or survival/refuge model of Indo–Australian Archipelago biodiversity?

4) What can coalescent and species nodal ages tell us about the evolution of population structure and diversity in widespread species?

Chapter 2 Materials and methods

2.1 Sampling

The majority of the specimens used in this study were collected by spearfishing. Some were obtained as fin clips from local fish markets (see Table 2.1 for sampling localities). For *S. psittacus*, samples from Bali, Taiwan, Okinawa and Tahiti were obtained from markets; samples from Taiwan were caught in Peng-Hu. Some fish from Tahiti were caught in the Tuamotos, French Polynesia (Winters *et al.*, 2010). All the markets from which samples were collected were small, local markets with local suppliers, where the likelihood of obtaining fish from afar was unlikely. Details of sample numbers for each location as well as market localities are specified in the relevant chapters. All samples were collected over multiple sampling efforts, beginning in 1998. All tissues were freshly preserved in 80% ethanol or salt-saturated 20% dimethyl sulfoxide (DMSO).

2.2 Laboratory procedures

Genomic DNA was extracted from tissues using proteinase-K digestion and salt–chloroform extraction, and precipitated in 100% ethanol using a similar protocol found in Sambrook and Russell (2001). Amplification by polymerase chain reaction (PCR) was conducted in 20-µl volumes consisting of 2 µl each of: DNA template, 10× PCR buffer, 2 mM dNTPs, 1.5–5.5 mM $MgCl_2$; 1 unit Taq DNA polymerase (Qiagen); and 0.5 pmol each of forward and reverse primers. The majority of sequences were amplified using universal control region primers (L15995F and H16498R, Meyer *et al.*, 1994), and remaining samples were successfully amplified using specifically designed primers (all primer sequences are given in Table 2.2).

In addition to the 3-step PCR used by Bay et al. (2004), I also used the following standard cycling profile: an initial denaturation at 94 ºC for 2 min followed by 35 cycles of denaturing at 94 ºC for 30 seconds per cycle, 45 seconds at various annealing temperature regimes including 55–53–50 ºC, 51–49 ºC, 51 ºC or 50 ºC for different population samples, an extension at 72 ºC for 90 s minutes, and a final extension at 72 ºC for 10 minutes. PCR products were checked on a 1.5–2% agarose gel and successfully amplified products were purified using isopropanol precipitation. All sequencing was done by Macrogen (Seoul, Korea) using a 96-capillary ABI 3730*xl* DNA Analyzer (as per manufacturer's instructions). Sampling, laboratory methods and data analysis for the respective outgroup species were conducted using the same methods.

Table 2.1 Life history data of the four species studied

	C. sordidus	***S. psittacus***	***S. rubroviolaceus***	***S. ghobban***
Distribution (Myers 1999)	Western Indian Ocean/Red Sea to central Pacific Ocean; Lord Howe Island	Western Indian Ocean/Red Sea to south and central Pacific Ocean	Western Indian Ocean/Red Sea to tropical east Pacific	Western Indian Ocean/Red Sea and the Mediterranean Sea, to tropical east Pacific; Lord Howe Island
Indian Ocean sampling sites*	Egypt, Seychelles (Seychelles Plateau, Farquhar, Amirante Plateau), Arabian Gulf (Umm al Ghanam), Oman (Khwar Limah, Muscat Banda Kharan, Bandar Kharan, Al Halaniyat Islands 1 and 2), Christmas Island, Cocos Keeling Islands, Western Australia (Abrolhos Islands, Long Island, Ningaloo Reef, Rowley Shoals)	Seychelles (Amirante Plateau and Seychelles Plateau), Cocos Keeling Islands, Christmas Island, Western Australia (Beacon Island, Rowley Shoals, Scott Reef)	Oman (Oman Al Halanyat, Muscat Banda Kharan, Al Halaniyat Islands 1 and 2), Seychelles (Farquhar, Seychelles), South Africa (Sodwana Bay), Cocos Keeling Islands, Christmas Island	Lebanon, Egypt, Oman (Khwar Ma'ili, Umm al Ghanam, Muscat Banda Kharan, Mughsayl, Bar al Hickman East, Masirah Island South, Al Halaniyat Island), Seychelles (Farquhar), Cocos Keeling Islands, Western Australia (Salmon Bay, Rottnest Island, Abrolhos Islands)
Pacific Ocean sampling sites*	Taiwan (Hengchun), Japan (Okinawa), Rota, Papua New Guinea (Kavieng, Kimbe Bay), Great Barrier Reef (Palm Group, Orpheus Island, Whitsundays Islands), Hawaii, French Polynesia (Tahiti, Moorea)	Taiwan (Hengchun, Bisha, Penghu Shan), Japan (Okinawa), Rota, Indonesia (Bali), Great Barrier Reef (Lizard Island), Hawaii, French Polynesia (Papeete, Marquesas)	Taiwan (Penghu Shan, Ho-Pin-Dao), Japan (Okinawa), Palau (Koror), Federated States of Micronesia (Pohnpei), Solomon Islands (Rorumana, Roku), Great Barrier Reef (Lizard Island, Hicks Reef, Mermaid Reef), Hawaii, French Polynesia (Marquesas), Clipperton Atoll, Panama (Las Perlas, Montuosa).	Taiwan (Taiwan, Taiwan Market, Penghu Shan), Palau, Pohnpei, Solomon Islands (Rorumana, Gizo, North Solomon Islands), Great Barrier Reef (Palm Group, Lizard Island), Panama (Las Perlas)
Habitat (Myers 1999)	Shallow reef flats, lagoons and seaward reefs	Shallow reef flats, lagoons and sheltered seaward reefs	Seaward reefs	Reefs and seagrass beds up to 250 m in depth
Size (total length)	40 cm (Parenti & Randall 2000)	30 cm (Parenti & Randall 2000)	70 cm (Parenti & Randall 2000)	90 cm (Parenti & Randall 2000)
PLD	30 days (Choat pers. comm.)	28 days (Choat pers. comm.)	44 days (Choat pers. comm.)	>30 days (Choat pers. comm.)

PLD = pelagic larval duration

* specific localities in parentheses

	Western Indian Ocean	Pacific Ocean
Chlorurus sordidus		
	Male, 23.8 cm SL, Gulf of Aqaba	Male, 22.7 cm SL, Lizard Island, Great Barrier Reef
Scarus psittacus		
	Male, 16.4 cm SL, Saudi Arabia	Male, 15.1 cm SL, Hawaii
Scarus rubroviolaceus		
	Male, 65 cm TL, Maldives*	Male, 62.7 cm TL, Hawaii*
Scarus ghobban		
	Male, 44 cm SL, Sudan (Red Sea)	Male, 40.2 cm SL, French Polynesia

SL = standard length; **TL = total length**
*The photograph of *S. rubroviolaceus* from the Indian Ocean was taken underwater; the specimen from Hawaii is a preserved specimen. Therefore, slight differences in colouration may result.
Photos courtesy of JE Randall

Figure 2.1 Specimens from the western Indian Ocean and Pacific Ocean for four species of scarine labrids

Table 2.2 Primers used for genomic DNA amplification

Primer name	Sequence 5′–3′	Product length (bp)	Annealing temperature (ºC)	Reference
L15995	AACTCTCACCCCTAGCTCCCAAAG	400	67	1
H16498	CCTGAAGTAGGAACCAGATG	400	55	1
Csor-F	TTTTAACCAAAATATGCATAGCTC	400	57	2
Csor-R	AGATGCCAGTAATARTGTGAGG	400	58	2
SghDL-F	TTATCCCTGATCATCAAGGAAT	400	55	3
SruDL-F	TGATCATCAAGAAACGAAAGC	400	53	3
SghruDL-R	TAGCTCCCAAAGCTAGAATC	400	53	3

bp = base pairs

1. Meyer et al. (1994)

2. van Herwerden & Klanten, unpublished

3. Beck & van Herwerden, unpublished

2.3 Phylogenetic analyses

Successfully amplified sequences were edited in Sequencher 4.2 (GeneCodes, 2005), and manually aligned using ClustalW (Thompson *et al.*, 1997). Modeltest 3.8 (Posada, 2006) or MrModeltest 2.2 (Nylander, 2004) was used to determine the substitution model of best fit for each dataset. Bayesian analyses were conducted in MrBayes using a Markov chain Monte Carlo (MCMC) consisting of 4 chains for 1 million generations (Huelsenbeck, Ronquist, 2001) unless otherwise noted. Trees with large InL scores (approximately the first 10% of trees generated) were discarded as burn-in. A 50% majority-rule consensus tree was generated based on the best Bayesian trees and rooted using the sister taxon as the outgroup, and viewed in PAUP* (Swofford, 2003). Specific details of these analyses, are noted in the respective chapters.

Results from the phylogenetic analyses will form the basis for the structuring of populations in the subsequent population genetic analyses. A single minimum spanning tree depicting linkages among haplotypes is another way to represent the phylogenetic relationship between individuals; distances between haplotypes were computed using Arlequin v3.01 (Excoffier *et al.*, 2006). A graphical representation was then drawn up in Adobe Illustrator to determine the relationships between haplotypes. Starburst patterns, in which the oldest haplotypes are centrally positioned,

are indicative of a population expansion and identify the likely source of the most recent common ancestor (Castelloe, Templeton, 1994).

2.4 Population genetic analyses

Tests of population differentiation and molecular diversity indices were calculated in Arlequin v3.0. Nucleotide diversity (π, Tajima, 1983) provides a measure of relative evolutionary age of a species. A value of $\pi < 0.5\%$ indicates low diversity and thus recent divergence. Haplotype diversity (h, Nei, 1987) is a measure of the shared and unique haplotypes in a population. Values greater than 0.5 are considered high, and they indicate high diversity among haplotypes. Tajima's test for selective neutrality (D, Tajima, 1983) is used to distinguish between mutations occurring randomly and those occurring due to selection. Fu's (Fs, Fu, 1997) test for population growth was also estimated to determine if populations were expanding. Under the hypothesis of selective neutrality, negative values for both Tajima's D and Fu's Fs are an indication of population expansion due to an excess of recent mutations.

Pairwise genetic distances between haplotypes from sampling locations with $n > 10$ were calculated using pairwise F_{ST}-based genetic distances in Arlequin v3.1 (Excoffier *et al.,* 2006). Bonferroni corrections were calculated based on the number of comparisons at a significance of $P < 0.05$ in order to reduce the likelihood of incorrectly rejecting a true null hypothesis (Rice, 1989). These distances, in conjunction with geographic distance between populations, were used in the Mantel test implemented in Isolation By Distance Web Service v.3.15 (IBDWS) (Jensen *et al.,* 2005), a web service that calculates isolation by distance, using 30 000 randomisations. An analysis of molecular variance (AMOVA) was also implemented in Arlequin v3.01 (Excoffier *et al.*, 2005) to determine the level of genetic partitioning within and among populations between ocean basins. Three separate AMOVAs were conducted; the first consisted of populations grouped by ocean basin, the second consisted of populations grouped into geographic regions based on pairwise F_{ST} comparisons and the third consisted of individuals grouped into clades determined by the phylogenetic analysis.

Estimates of migration rate between populations were inferred using Migrate 2.3 (Beerli, Felsenstein, 1999; Beerli, Felsenstein, 2001), a coalescence-based program that estimates migration between populations using a MCMC approach. Due to the high genetic partitioning indicated by pairwise F_{ST} comparisons, the data were partitioned by ocean basin and run as two separate datasets, with east Indian Ocean individuals that grouped with Pacific Ocean individuals incorporated into the Pacific Ocean dataset, where applicable. To account for the high degree of stochasticity associated with migration rates in low population sizes, estimates were based on an average of 6 replicate runs with the

following start parameters: 10 short chains sampling 1000 genealogies and 5 long chains sampling 10 000 genealogies. Chains were heated and set to sample every 20 steps using Bayesian inference. Values for the transition:transversion (TI:TV) ratio and the gamma-distribution-shape parameter obtained from MrModeltest 2.2 were also incorporated.

Beerli (2009), addresses some of the assumptions of his program Migrate, and provides some reassurances about what happens if those assumptions are violated. There are five main assumptions of Migrate: (1) population sizes remain constant or are randomly fluctuating around an average over time, (2) mating is random, and that there is equal probability of all individuals to produce offspring, i.e. no selection, (3) mutation rate is constant, (4) immigration is symmetrical and constant through time, although this can vary among populations, and (5) populations do not diverge. While results from such analyses should be interpreted with caution, most data sets would violate these assumptions due to the inherently complex nature of biological data. The first assumption, for example, is a common assumption in many population genetics analyses, including those for the widely accepted coalescent theory used by Kingman (1982). The second assumption about selection is also most likely not met. Migrate allows for allows for the input of site rate variation among nucleotide sites. Gamma values for each species were calculated in MrModeltest and incorporated in Migrate analyses. The assumption of symmetrical and constant immigration rate is also a standard assumption for the widely reported F_{ST}-based analyses; this assumption is likely not met in Migrate as well as the F_{ST} analyses. Finally, Beerli (2009) states that the assumption that populations do not diverge not affect the results if the time since most recent ancestor (T_{MRCA}) is younger than the population divergence time. All four species in this study have T_{MRCA}s that are younger than the population divergence so this assumption is not likely to affect the results. In addition, Beerli (2009) also states that only using population pairs can lead to an overestimation of parameters (Beerli, 2004; Slatkin, 2005), although this is not an issue as the data are separated into several geographic populations which had strong support from Bayesian analyses.

2.5 Coalescence analyses

The timing of divergence between the Indian Ocean and Pacific Ocean clades was calculated using coalescence as per Schneider and Excoffier (1999). The nucleotide mutation rate (μ) was calculated based on the sum of the proportions of conserved and variable sites, where conserved sites mutate at a slow rate of 1.1% MY^{-1} and variable sites at a rapid rate of 12.9% MY^{-1}, based on previously published mutation rates for the swordfish *Xiphias gladius* (Alvarado Bremer *et al.*, 1995; Messmer *et al.*, 2005). Based on these calculated mutation rates, an overall mutation rate for the entire sequence (u) was calculated using the formula $u = 2\mu k$, where k is the length of the sequence (Harpending, 1994; Rogers, Harpending, 1992). The generation time for each species (t_2) was calculated using the formula $t_2 = (\alpha+\omega)/2$, where α is the age at first reproduction and ω is age at last reproduction (Pianka, 1978).

The expansion time (t_1) between each clade was calculated using the formula $t_1 = t_2(\tau/2u)$, where τ is equal to the divergence time between a population before and after expansion (Harpending, 1994; Rogers, Harpending, 1992). Values for these calculations are detailed for each species in the respective chapters. A mismatch distribution analysis was implemented in Arlequin v3.01 (Excoffier *et al.*, 2006; Schneider, Excoffier, 1999) to determine the demographic history of the two major clades, as well as that of any additional clades identified by the phylogenetic and population genetic analyses (Rogers, Harpending, 1992).

In addition to conducting coalescence analyses based on geographic locations, the data were also grouped based on clades identified by the phylogenetic analyses if possible. Where applicable, summary statistics, pairwise F_{ST}, AMOVAs and coalescence analyses were also generated based on geography and phylogenetic clade structure.

Chapter 3 Population structure of the widespread parrotfish *Chlorurus sordidus*: evidence of genetic partitioning at three marine biogeographic barriers

Abstract

An extensive study of the genetic connectivity of the Indo–Pacific parrotfish *Chlorurus sordidus* provided detailed phylogeographic and phylogenetic evidence of historical separations and population bottlenecks leading to present day partitioning of this putatively widespread species. This study incorporated a combined phylogenetic, phylogeographic and population genetic approach to determine the evolutionary history and population structure of *C. sordidus*. We compared a 354 base pair (bp) DNA sequence from the mitochondrial control region in 351 individuals from 18 locations across the Indo–Pacific, spanning at least three biogeographic breaks. We confirmed previous findings of significant, strong population partitioning at the largest spatial scale. We also identified the genetic break between western Indian Ocean and Pacific Ocean populations to be at the Indo – West Pacific Barrier in the vicinity of Christmas Island in the east Indian Ocean. The vast majority of Western Australian individuals grouped within the Pacific Ocean clade despite inhabiting the east Indian Ocean. F_{ST} between the Indian Ocean and Pacific Ocean populations ranged from 0.594 to 0.762, depending on how the data were structured. We also confirmed the break associated with the central Pacific barrier for this species (F_{ST} between Hawaii and other Pacific Ocean locations ranged from 0.096 to 0.286). Within both ocean basins, we also identified previously undetected partitioning close to the north Indian Ocean biogeographic break, which separates the Arabian Gulf from the Gulf of Oman. We also identified partitioning centred around French Polynesia in the Pacific Ocean. Genetic diversity was high for both Indian Ocean (h = 0.9858, π = 3.178%) and Pacific Ocean (h = 0.992, π = 3.050%) populations, but was comparatively low for Oman (h = 0.450, π = 1.354%). We determined the coalescence time between the western Indian Ocean and Pacific Ocean clades (including the east Indian Ocean clade) to be approximately 1 million years ago, which is associated with dramatic sea level changes during the Pleistocene. In addition, we found the direction and strength of gene flow between populations to be uneven and largely—but not entirely—explained by present-day oceanographic currents.

3.1 Introduction

Many marine phylogeographic studies conducted across entire ocean basins have found evidence of population structure, but less genetic connectivity than expected (Avise, 2000) in widespread

populations including soldierfish (Craig *et al.*, 2007), gobies (Kon *et al.*, 2007), damselfish (Bernardi *et al.*, 2001) and coconut crabs (Lavery *et al.*, 1996b). In the Indo–Pacific, discreet populations have been found with ocean basin affinities, with some populations separated by a barrier in the Indian Ocean (e.g., Craig *et al.*, 2007; Marie *et al.*, 2007; McMillan, Palumbi, 1995; Timm, Kochzius, 2008; Williams, Benzie, 1998; Williams *et al.*, 2002) and others separated by a barrier in the Pacific Ocean (Craig *et al.*, 2007). Some studies have even shown widespread species to comprise groups of genetically distinct, cryptic species, including species of crabs (Gopurenko, 1999; Lavery *et al.*, 1996b), starfishes (Benzie, 1999b; Williams, Benzie, 1998) and bonefishes (Colborn *et al.*, 2001). While dispersal between coral reef environments is easily possible today, even among larvae with relatively short larval durations (< 30 days) and poor swimming capabilities via oceanographic features and wind-driven surface currents, evidence of genetic partitioning points to historical rather than contemporary barriers to gene flow as the major underlying factor responsible for these patterns. Many of these studies also suggest that sea-level fluctuations and uneven cooling due to Pleistocene glaciations isolated populations and resulted in the patterns of distribution and population structure we see today (e.g., Benzie, 1999a; Lessios *et al.*, 2001).

Coral reefs in the Middle East provide a prime example of the disproportionate effects of sea-level fluctuations, resulting in the instability of reefal environments over geological time. For example, the Arabian Gulf dried up completely during the Holocene, exposing reefs during low sea levels, while reefs in Oman were drowned repeatedly during times of high sea-level stand (Sheppard *et al.*, 2000). At other times, the Arabian Gulf would have been separated from the rest of the Indian Ocean by a land bridge, while other reef systems became isolated due to changed oceanographic conditions. Likewise, the reefs in the Pacific Ocean have endured a history of drowning and emersion, and many—such as Hawaii—have been subject to ongoing volcanic activity (Fletcher *et al.*, 2008). These events presented a myriad of shifting biogeographic barriers in space and time, which were unique in their potency and duration. Although there is evidence that these biogeographic barriers reduced gene flow enough to create genetically isolated populations at specific times, not all widespread reef taxa were affected in the same way. Several recent studies on Indo–Pacific acanthurids revealed little spatial genetic structure across their Indo–Pacific distributions, while there was evidence of much stronger temporal genetic partitioning (Horne *et al.*, 2008; Klanten *et al.*, 2007). In contrast, the soldierfish *Mypristis berndti*, which has a pelagic larval duration (PLD) of 55 days, is more widely distributed (Craig *et al.*, 2007) than several acanthurid species with a PLD of 90 days (Horne *et al.*, 2008; Klanten *et al.*, 2007). Despite having a greater distribution, *M. berndti* was found to exhibit significant genetic partitioning both between ocean basins at the Indo – West Pacific Barrier and within the ocean basin at the East Pacific Barrier (Craig et al 2007). However, the acanthurids demonstrated no partitioning on the same spatial scale (Horne *et al.*, 2008; Klanten *et al.*, 2004). Likewise, the scarine parrotfish *C. sordidus*, which has a similar distributional range to the

acanthurids, but a PLD equivalent to one-third of the acanthurids and one-half of the soldierfishes, showed significant population structure between the western Indian Ocean and Pacific Ocean (Bay *et al*., 2004). This pattern demonstrates the species-specific nature of genetic partitioning across biogeographic barriers, and the importance of biological, ecological and historical factors in understanding phylogeographic patterns. Understanding connectivity in the marine environment will require a comparative phylogeographic approach using widely distributed species with similar distributions that span known biogeographic barriers. This congeneric phylogeographic approach (Funk, Omland, 2003) will enable us to tease apart the factors underlying the patterns of distribution we see today; it will also provide insight into the processes driving speciation in one of the most diverse environments.

The scarine parrotfish *C. sordidus* is a good starting point for a comparative phylogeographic study due to its widespread nature, high abundance and fecundity, relatively short PLD and close association with coral reefs (Choat, pers. comm.). Moreover, there is evidence of both morphological (Randall *et al.*, 1997; Randall, Bruce, 1983) and genetic differentiation (Bay *et al.*, 2004) at the extremes of its distribution (Figure 2.1), as well as a Plio-Pleistocene nodal (evolutionary) age based on multiple molecular markers (Choat *et al.*, 2012; Smith *et al.*, 2008). Additionally, because of its widespread nature which encompasses the Indo-Australian Archipelago, this species is ideal for testing models of species richness, i.e. centre of origin (Briggs, 1999; Briggs, 2000; Ekman, 1953; Mora *et al.*, 2003), centre of accumulation/overlap (Barber *et al.*, 2000; Bellwood, Wainwright, 2002; Jokiel, Martinelli, 1992; Ladd, 1960; Santini, Winterbottom, 2002; Woodland, 1983), and centre of survival/refuge (Barber, Bellwood, 2005; McCoy, Heck, 1976) in that area. These models are outlined in Chapter 1. Although Bay et al. (2004) were able to find structure between western Indian Ocean and Pacific Ocean populations of *C. sordidus*, sampling limitations precluded the authors from determining the precise location of the barrier to gene flow or answer questions about the evolutionary history of the species. The present study incorporates sampling from the extremes of the *C. sordidus* geographic range, including the northern and east Indian Ocean, as well as the south and central Pacific Ocean, to provide a comprehensive view of the processes underlying structure and genetic differentiation among populations of this widespread reef fish. Here we incorporate phylogenetic and population genetic analyses, together with biogeographic information, to piece together the history of spatio–temporal divergences and connectivities at evolutionary timescales. Specifically we aim to address the following four questions:

1. Where is the geographic location of the barrier between the Indian Ocean and Pacific Ocean? Is this associated with the biogeographic Indo – West Pacific Barrier in the east Indian Ocean?

2. Is there evidence of further genetic partitioning in *C. sordidus* populations within the western Indian Ocean and Pacific Ocean? If so, is such partitioning associated with any other known biogeographic barriers?

3. When and where did *C. sordidus* populations - between and within the Indian Ocean and Pacific Ocean - diverge, and were they associated with known Pleistocene vicariance events?

4. Where gene flow is detected between spatially sampled locations, what is the direction and magnitude of gene flow between populations? Does this reflect historical or present-day ocean circulation patterns?

3.2 Materials and methods

3.2.1 Sampling, DNA amplification and sequencing

In total, 351 *C. sordidus* tissue samples were collected from 18 locations throughout the Indo–Pacific, incorporating sequences from 185 individuals from 9 populations from Bay et al. (2004), Genbank accession numbers AY392560–AY392744 (Figure 3.1). Most samples were collected by spearfishing; some samples from Okinawa and all samples from Tahiti were obtained from fish markets. Genomic DNA was isolated from tissues and amplified using universal control region primers L15995F and H16498R (Meyer *et al.*, 1994), and remaining samples were amplified using specific primers for *C. sordidus*, Csor-F and Csor-R (Table 2.2). The PCR products were purified using isopropanol precipitation and then sequenced by Macrogen Inc. (Seoul, Korea) using an ABI 3730*xl* DNA Analyzer (as per manufacturer's instructions). Sampling, DNA extraction and amplification, and sequencing for the outgroup species *C. bowersi* were conducted using the same methods. Further sampling, PCR and sequencing methods were performed as indicated in Section 2.2.

3.2.2 Phylogenetic analyses

We used Modeltest 3.8 (Posada, 2006) to determine the model of best fit for the data. A 50% majority-rule consensus tree was generated based on the best 32 Bayesian trees, and was rooted using the sister taxon *C. bowersi* (Smith *et al.*, 2008) as the outgroup. Support for clades has been indicated by Bayesian posterior probabilities. Individuals from locations with multiple sampling sites (Table 2.1) including the Seychelles, Western Australia and the Great Barrier Reef have been pooled in the phylogenetic analysis due to lack of structure observed between the respective sites. Individuals from Taiwan have also been pooled in the phylogenetic tree with those from Okinawa due to small sample size. Further phylogenetic analyses are detailed in Section 2.3.

3.2.3 Population genetic analyses

Pairwise genetic distances between haplotypes from 15 sampling locations with $n > 10$ were calculated using fixation index (F_{ST})-based genetic distances in Arlequin v3.1 (Excoffier *et al.*, 2006). Bonferroni corrections were calculated based on 15 comparisons at a significance of $P < 0.05$ in order to reduce the likelihood of incorrectly rejecting a true null hypothesis (Rice, 1989). Three separate analyses of molecular variance (AMOVA) were conducted: the first consisted of 15 populations grouped by ocean basin, the second consisted of 14 populations grouped into 5 geographic regions based on pairwise F_{ST} comparisons and the third consisted of individuals grouped into 6 genetically differentiated clades determined by the phylogenetic analysis. Further detailed population genetic analyses were performed as indicated in Section 2.4.

3.2.4 Coalescence analyses

The time to the most recent common ancestor (T_{MRCA}) between the western Indian Ocean and Pacific Ocean clades was calculated by coalescence as per Schneider and Excoffier (1999), implementing previously published mutation rates (Alvarado Bremer *et al.*, 1995; Messmer *et al.*, 2005). An overall mutation rate for the entire sequence (u) was then calculated based on these mutation rates following the formula $u = 2\mu k$, where k is the length of the sequence (321 bp). The generation time for *C. sordidus* (t_2) was calculated using the formula $t_2 = (\alpha+\omega)/2$, where the age at first reproduction (α) was 1.5 years and age at last reproduction (ω) was 8.5 years (Choat, pers. comm.). Coalescence analyses

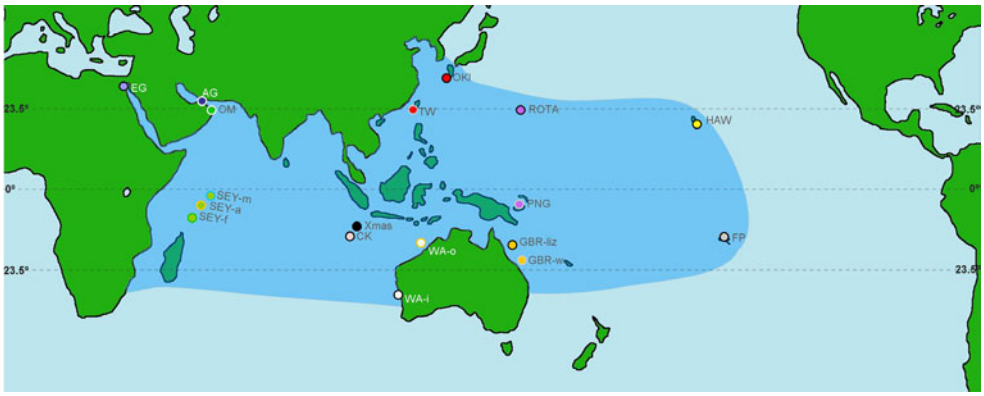


Figure 3.1 Map depicting distribution (shaded in the darker blue) and 18 total locations sampled for *C. sordidus*

Location code and numbers sampled as follows: EG = Egypt (2); AG = Arabian Gulf (27); OM = Oman (16); SEY-m = Mahe, Seychelles (15); SEY-a = Amirante, Seychelles (16); SEY-f = Farquhar, Seychelles (19); CK = Cocos Keeling Islands (20); Xmas = Christmas Island (3); WA-o = Scott and Clerk reefs, Western Australia (34); WA-i = Abrolhos islands, Western Australia (45); TW = Taiwan (3); OKI = Okinawa (16); ROTA = Rota, Micronesia (25); PNG = Kimbe Bay, Papua New Guinea (18); GBR-liz = Lizard Island, Great Barrier Reef (27); GBR-w = Whitsundays, Great Barrier Reef (15); HAW = Hawaii (25); FP = French Polynesia (27)

were conducted on data grouped both by geographic location as well as by the clades identified in the phylogenetic analyses. Further detailed coalescence analyses methods are described in Section 2.5.

3.3 Results

3.3.1 Phylogenetic analyses

We analysed the DNA sequence of 321 base pairs from the mitochondrial control region of 351 *C. sordidus* individuals from 18 locations throughout their Indo–Pacific distribution (Figure 3.1). Of the 154 polymorphic sites, 110 were parsimony informative (transition:transversion [TI:TV] ratio = 6.4052). The base frequency of the mitochondrial control region in *C. sordidus* is AT biased (A = 34.8%, C = 25.5%, G = 13.6%, T = 26.1%). Overall, the nucleotide and haplotype diversity was high (π = 5.552% and h = 0.995, respectively) for *C. sordidus*. The best-fit substitution model for the *C. sordidus* data, selected in Modeltest 3.8, was different between the hierarchical likelihood ratio tests (TrN + G, G = 0.3024) and Akaike information criterion (AIC) (TVM + I + G, I = 0.2987, G = 0.5686). The AIC model was used in subsequent analyses. A consensus of the 32 best Bayesian trees identified the same 2 major clades found in Bay et al. (2004), which correspond to the western Indian Ocean and the Pacific Ocean. Both of these clades had 100% support (Figure 3.2a). In addition to the two major clades, 6 well-supported clades within the two ocean basins also emerged. In the Indian (I) Ocean, clade I-1 comprised mainly Cocos Keeling Islands and Seychelles individuals. I-2 contained all of the north Indian Ocean individuals, including those from the Arabian Gulf and Oman, as well as individuals from the Seychelles. In the Pacific (P) Ocean, 2 of the 4 clades (P-1 and P-2) contained mostly individuals from Western Australia, the Great Barrier Reef, Rota, French Polynesia and Okinawa.

Clade P-2 had nearly 3 times as many individuals as clade P-1, but similar proportions of individuals from the respective geographic locations. Clade P-3 contained a similar composition of individuals to clades P-1 and P-2, with slightly more French Polynesian individuals. The final Pacific clade, P-4, was dominated by individuals from Hawaii, as well as a few from the Indo–Pacific region, including Western Australia, but no French Polynesian individuals. Clade P-4 was identified in Bay et al. (2004) as the Hawaiian-dominated subclade, which maintained its affinity despite the addition of more samples.

Although individuals from Western Australia grouped with the Pacific Ocean individuals, there was a single Western Australia individual (Fig. 3.2 (a) subclade I-6) that was at the base of the Indian Ocean clade. In addition, we found a single individual from Okinawa that grouped with Indian Ocean individuals (Clade I-1, Figure 3.2a) indicating limited gene flow between ocean

basins. The phylogenetic analysis also revealed geographic partitioning within each major clade, with high support. In particular, we found that Christmas Island individuals, although limited in number, were distributed between both Indian Ocean and Pacific Ocean clades.

3.3.2 Population genetic analyses

Nucleotide diversity (π) was very high within populations, ranging from 1.35% in Oman to 3.5% in Rota (Table 3.1a). It was also very high among clades, ranging from 0.939% in the Hawaiian-dominated clade, P-8, to 2.72% in the north Indian Ocean-dominated clade, I-2 (Table 3.1b). Haplotype diversity was also high for both geography-based and clade-based analyses, with values close to 1.0, with the exception of Oman (h = 0.45) (tables 3.1a–b). All values for Tajima's D were negative and none were significantly based on geography (Table 3.1a); however, based on the phylogenetic structure, two Pacific Ocean clades (P-2 and P-4) were significantly negative (Table 3.1b). Negative Tajima's D values are consistent with population expansions. Similarly, Fu's test for population growth also indicated expansion in all populations except for Oman (Table 3.1a). Fu's values were highly significant and negative for clade-based analyses (Table 3.1b). The minimum spanning network detected the same two major clades identified in the phylogenetic analysis (Figure 3.3). The Pacific Ocean clade was characterised by a few shared haplotypes and many unique haplotypes, which were distributed fairly evenly, with the exception of Hawaii. Fifty-six per cent of all Hawaiian individuals were grouped together, of which 43% were part of a major shared haplotype and the remaining samples were separated by fewer than 2 bp differences from the main Hawaii-dominated haplotype (Figure 3.3).

In the Indian Ocean clade, the Arabian Gulf haplotypes were clustered together, and branched off from an Oman-exclusive shared haplotype, which contained 69% of all Oman individuals. Western Australia populations, despite their geographic position in the east Indian Ocean and their proximity to the Cocos Keeling and Christmas islands, were more similar to Pacific Ocean individuals, with several haplotypes shared with populations at the range extremities (Figure 3.3). Partitioning within ocean basins was also evident, based on population pairwise F_{ST} comparisons between the 15 sampling locations for which adequate sample sizes were available (Table 3.2a), as well as 6 phylogenetic clades (Table 3.2b). Within the Indian Ocean, there was significant genetic partitioning between individuals from the Arabian Gulf and other Indian Ocean individuals, including those from the Indian Ocean clade. In addition, Oman showed significant genetic partitioning from the remaining east Indian Ocean individuals (Mahe, Seychelles and Cocos Keeling Islands), which were not genetically differentiated, despite more than 4600 km between the Seychelles Islands in the west Indian Ocean and the Cocos Keeling Islands in the east

(a)

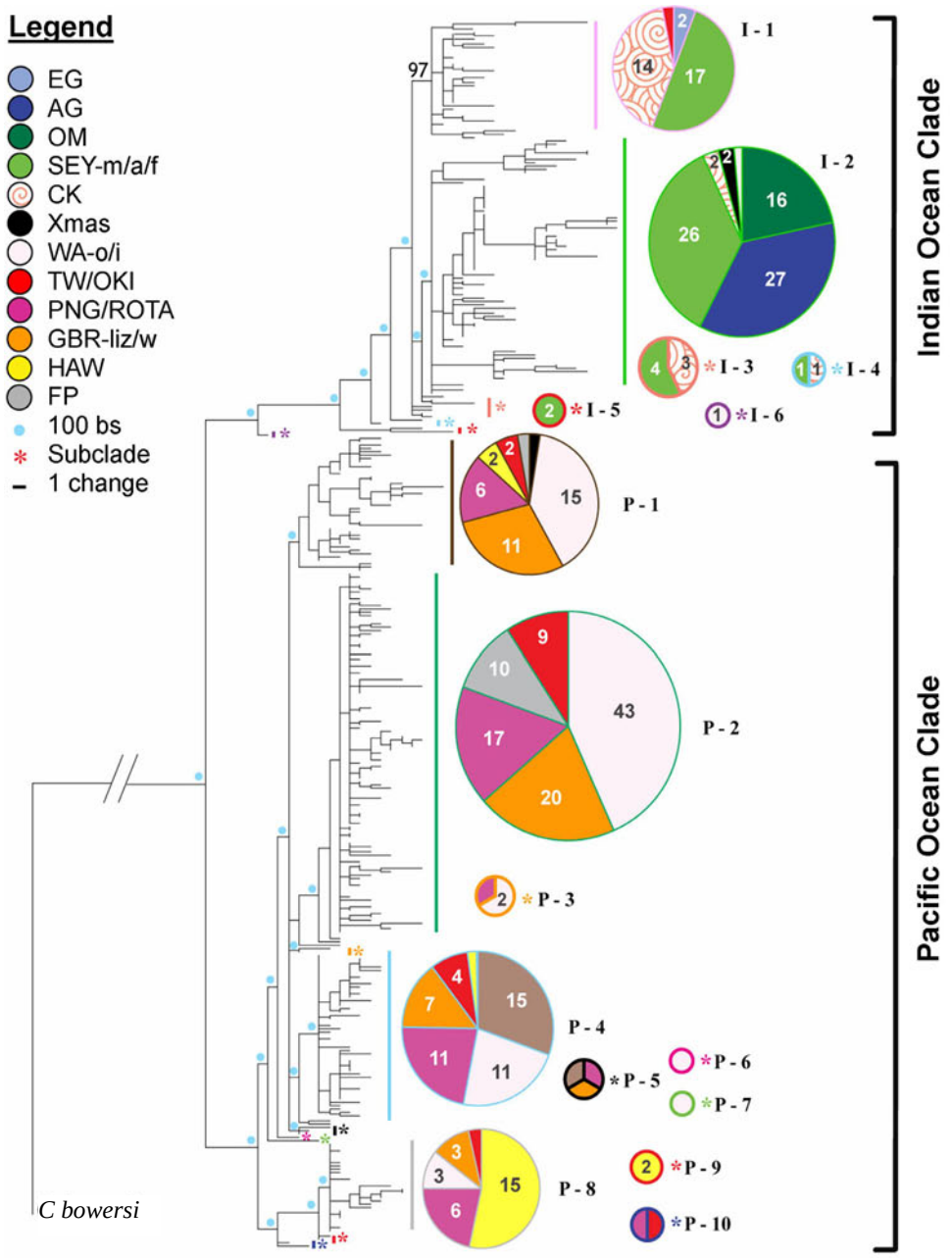


Figure 3.2(a) Consensus of 32 best Bayesian trees for *C. sordidus*

Large coloured pies indicate frequency and geographic breakdown of clades. Numbers inside pie sections indicate number of individuals. Where no numbers exist, $n = 1$. Support for nodes are shown by blue dots (bootstrap = 100). Asterisks correspond to small subclades, which are outlined in the same colour. Indian Ocean and Pacific Ocean subclades are abbreviated as I-# and P-#, respectively. Refer to Fig. 3.1 for location abbreviations.

(b)

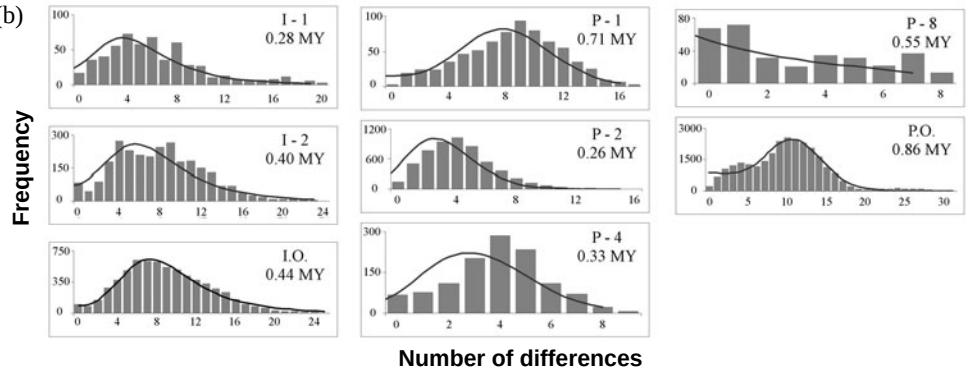


Figure 3.2(b) Mismatch distribution curves and coalescence times for larger well-supported clades (I-1, I-2, P-1, P-2, P-4, P-8) and overall clades (IO and PO)

Numbers of pairwise differences are plotted on the x-axis and frequency on the y-axis. Bars represent observed frequencies and lines represent model frequencies.

AG = Arabian Gulf; CK = Cocos Keeling Islands; EG = Egypt; FP = French Polynesia; GBR = Great Barrier Reef; HAW = Hawaii; my = million years; OKI = Okinawa; OM = Oman; PNG = Papua New Guinea; SEYm/a/f = Mahe/Amirante/Farquhar, Seychelles; TW = Taiwan; WA-i = Abrolhos islands, Western Australia; WA-o = Scott and Clerk reefs, Western Australia; Xmas = Christmas Island

Indian Ocean. Genetic partitioning within the Indian Ocean is evident in both geography-based (Table 3.2a) and clade-based population pairwise F_{ST} comparisons, all of which were highly significant (Table 3.2b).

Within the Pacific Ocean, there was much less population structure, with over 97% of comparisons yielding F_{ST} values below 0.25 (Table 3.2a). Only Hawaii showed significantly high F_{ST} values between Papua New Guinea (PNG) ($F_{ST} = 0.275$) and French Polynesia ($F_{ST} = 0.286$). Pairwise comparisons were also conducted with Moorea and Tahiti as separate populations, but no difference was observed. Thus, the populations were pooled together as French Polynesia. Between ocean basins, F_{ST} values ranged from 0.594 to 0.762. All high pairwise values ($F_{ST} > 0.25$) were significant ($P < 0.00333$). In contrast, clade-based pairwise comparisons showed high genetic differentiation between clades, with some Pacific Ocean clade comparisons exhibiting greater partitioning than observed between clades in the different ocean basins (Table 3.2b).

In particular, comparisons between the Hawaii-dominated clade, P-8, showed highly significant and high pairwise F_{ST} values when compared with clades P-2 and P-4, both of which contained a wide mix of haplotypes from different Pacific Ocean and east Indian Ocean locations. These values were higher than values recorded for genetic distances between clade P-1 and both Indian Ocean clades. Significant isolation by distance was also detected at the largest spatial scale, with a significant positive relationship between geographic and genetic distance ($Z = 528150.58$, $r^2 = 0.328$, $P < 0.0001$). Within ocean basins, this relationship was also significant, but not as strong. Within the Pacific Ocean, the relationship was positive and significant ($Z = 19052.3376$, $r^2 =$

0.240, P = 0.025) and the Indian Ocean showed a negative, non-significant, relationship (Z = 37422.2511, r^2 = 0.0175, P > 0.05).

Ocean basins accounted for the highest variation (63.2%) in the AMOVA, and there was also a substantial proportion of the variation explained within populations (31%) (Table 3.3a). Similar results were obtained from an AMOVA structure based on 5 locations with significant high F_{ST} values, including the Arabian Gulf, Oman, the central Indian Ocean (Seychelles and Cocos Keeling islands), the western/central Pacific (including French Polynesia) and Hawaii (Table 3.3b). AMOVAs based on phylogenetic clades also attributed the majority of the variation to ocean basins; however, the remaining variation was evenly divided among populations within groups and within populations (Table 3.3b).

Table 3.1a Summary statistics for all 18 *C. sordidus* sampling locations

Location	n	n_h	π ± SE (%)	h ± SE	D	PD	FS	PFS
Egypt*	2	2	*	*	*	*	*	*
Arabian Gulf	27	23	2.287 ± 1.234	0.986 ± 0.015	–0.374	0.375	–13.644	0.000
Oman	16	5	1.354 ± 0.798	0.450 ± 0.151	–0.386	0.373	2.416	0.874
Sey-Mahe	15	13	2.289 ± 1.274	0.971 ± 0.039	–0.455	0.349	–4.661	0.017
Sey-Farquhar	19	19	2.264 ± 1.246	1.000 ± 0.017	–1.462	0.069	–14.677	0.000
Sey-Amirante	16	16	3.323 ± 1.791	1.000 ± 0.022	–0.480	0.340	–8.046	0.002
CK	20	17	1.770 ± 0.990	0.968 ± 0.033	–1.178	0.121	–9.453	0.000
Christmas Island*	3	3	*	*	*	*	*	*
Total Indian Ocean	**118**	**98**	**3.178 ± 1.628**	**0.986 ± 0.006**	**–1.361**	**0.081**	**–24.496**	**0.000**
WA-offshore	34	31	3.093 ± 1.622	0.993 ± 0.010	–1.209	0.099	–16.609	0.000
WA-inshore	44	37	2.789 ± 1.456	0.991 ± 0.008	–1.015	0.156	–24.517	0.000
Taiwan*	2	2	*	*	*	*	*	*
Okinawa	16	13	3.263 ± 1.768	0.950 ± 0.049	–0.937	0.174	–2.698	0.102
Rota	25	22	3.472 ± 1.820	0.990 ± 0.014	–0.128	0.466	–8.998	0.003
PNG	18	18	2.292 ± 1.261	1.000 ± 0.019	–1.029	0.154	–12.995	0.000
GBR-Lizard Island	27	23	2.971 ± 1.568	0.989 ± 0.013	–0.797	0.230	–10.464	0.001
GBR-Whitsundays	15	14	2.530 ± 1.396	0.991 ± 0.028	–0.487	0.338	–6.092	0.004
Hawaii	25	17	2.791 ± 1.484	0.937 ± 0.037	–0.459	0.344	–3.427	0.097
French Polynesia	27	22	2.256 ± 1.218	0.969 ± 0.025	–1.176	0.125	–11.533	0.000
Total Pacific Ocean	**233**	**199**	**3.050 ± 1.552**	**0.992 ± 0.002**	**–1.416**	**0.045**	**–24.171**	**0.001**
Total	**351**	**297**	**5.552 ± 2.735**	**0.995 ± 0.001**	**–0.646**	**0.299**	**–23.571**	**0.009**

CK = Cocos Keeling Islands; D = Tajima's selective neutrality test; FS = Fu's neutrality test; GBR = Great Barrier Reef; h = haplotype diversity; n = individuals sampled; n_h = number of haplotypes sampled; π = nucleotide diversity; PD = Tajima's D and significance level; PFS = Fu's FS and significance level; PNG = Papua New Guinea; SE = standard error; Sey = Seychelles; WA = Western Australia

*Egypt, Christmas Island and Taiwan were omitted from population genetics analyses due to low sample size

The Migrate analysis suggests that the direction of gene flow between populations is largely uneven (Figure 3.4a–c). Within the Indian Ocean, there was evidence of migration in the westward direction along low latitudes, with roughly even migration between Mahe, Seychelles, and Cocos Keeling Islands (Figure 3.4a). In the northern Indian Ocean, migration was from Oman in the southerly direction as far as Mahe, Seychelles. Interestingly, migration between the Arabian Gulf and other populations was unidirectional, with only a few migrants going to the Arabian Gulf from Mahe, Seychelles, and Oman, but no migrants leaving the Arabian Gulf.

Migration along the east African coast was much greater in the southerly direction, with more than 6 times as many migrants from Mahe south to Farquhar, Seychelles. In the Pacific Ocean, migration appeared to be in the westerly direction overall, with the exception of low but unidirectional migration from Papua New Guinea to Hawaii (4.9 migrants) and over 10 times as many migrants from offshore Western Australia to French Polynesia (Figure 3.4b–c). In the south/central Pacific Ocean, migration was in the southerly direction, with no evidence of migration from French Polynesia to Hawaii. In the Indo–Australian Archipelago, migration was predominantly towards the north.

3.3.3 Coalescence analyses

Results from the coalescence analysis indicate that the coalescence time, or (T_{MRCA}) of the expanding Pacific Ocean clade (T_{MRCA} = 0.865 mya) is nearly twice as ancient as the expanding

Table 3.1b Summary statistics for *C. sordidus* phylogenetic clades

Clade	*n*	n_h	π ± SE (%)	*h* ± SE	D	PD	FS	PFS
I-1	34	27	2.011 ± 1.089	0.972 ± 0.019	–1.502	0.046	–17.952	0.000
I-2	74	54	2.722 ± 1.419	0.970 ± 0.013	–1.054	0.143	–24.856	0.000
Total Indian Ocean	**118**	**98**	**3.178 ± 1.628**	**0.986 ± 0.006**	**–1.361**	**0.081**	**–24.496**	**0.000**
P-1	38	36	2.645 ± 1.391	0.997 ± 0.007	–0.793	0.230	–24.855	0.000
P-2	106	77	1.310 ± 0.727	0.974 ± 0.010	–2.099	0.002	–25.798	0.000
P-4	49	32	1.250 ± 0.708	0.943 ± 0.022	–2.115	0.007	–25.846	0.000
P-8	26	14	0.939 ± 0.565	0.794 ± 0.084	–1.435	0.062	–6.359	0.002
Total Pacific Ocean	**233**	**199**	**3.050 ± 1.552**	**0.992 ± 0.002**	**–1.416**	**0.045**	**–24.171**	**0.001**
Total	**351**	**297**	**5.552 ± 2.735**	**0.995 ± 0.001**	**–0.646**	**0.299**	**–23.571**	**0.009**

D = Tajima's selective neutrality test; FS = Fu's neutrality test; *h* = haplotype diversity; I = Indian Ocean; *n* = individuals sampled; n_h = number of haplotypes sampled; π = nucleotide diversity; P = Pacific Ocean; PD = Tajima's D and significance level; PFS = Fu's FS and significance level; SE = standard error

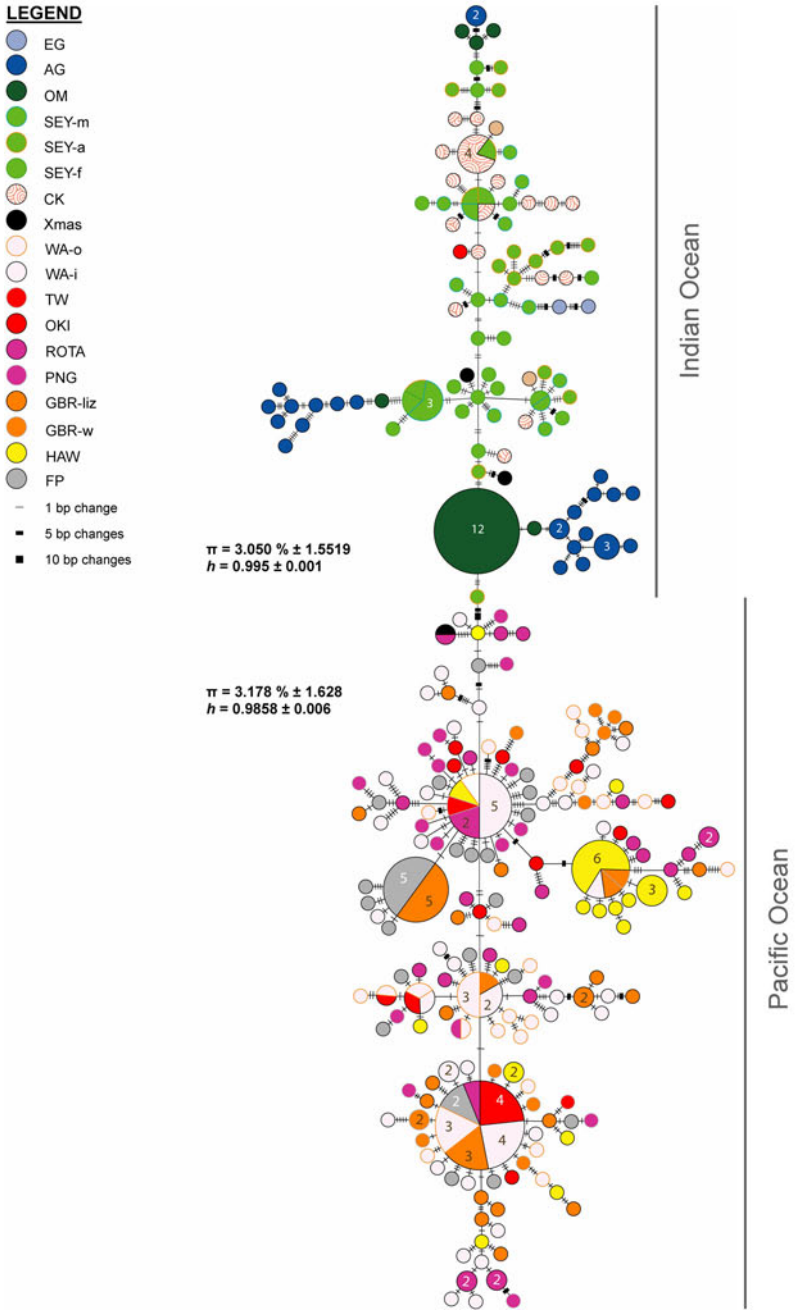


Figure 3.3 Minimum haplotype network for Pacific and Indian Ocean haplotypes, and respective nucleotide (π ± SE %) and haplotype (h ± SE) diversity indices

Small circles represent unique haplotypes and larger circles indicate shared haplotypes, with the colours and proportion of individuals from respective populations with that shared haplotype shown. Thin hash marks indicate a single base pair (bp) substitution between haplotypes, medium-sized bars indicate a 5 bp difference and thick bars represent 10 bp changes.

AG = Arabian Gulf; CK = Cocos Keeling Islands; EG = Egypt; FP = French Polynesia; GBR = Great Barrier Reef; HAW = Hawaii; OKI = Okinawa; OM = Oman; PNG = Papua New Guinea; SEYm/a/f = Mahe/Amirante/Farquhar, Seychelles; TW = Taiwan; WA-i = Abrolhos islands, Western Australia WA-o = Scott and Clerk reefs, Western Australia; Xmas = Christmas Island

Table 3.2a Population pairwise F_{ST} values between 15 sampling locations with adequate sample sizes ($n > 10$)

Location	AG	OM	SEY-m	SEY-f	SEY-a	CK	WA-o	WA-i	GBR-liz	GBR-w	OKI	ROTA	PNG	HAW	FP
AG	*	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
OM	**0.456**	*	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
SEY-m	**0.462**	**0.342**	*	0.096	0.034	0.057	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
SEY-f	**0.430**	**0.223**	0.030	*	0.066	0.001	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
SEY-a	**0.417**	**0.215**	0.050	0.035	*	0.002	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
CK	**0.482**	**0.353**	0.034	0.084	0.074	*	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
WA-o	**0.710**	**0.682**	**0.674**	**0.677**	**0.614**	**0.663**	*	0.101	0.031	0.565	0.774	0.051	0.002	**0.000**	**0.000**
WA-i	**0.733**	**0.708**	**0.703**	**0.705**	**0.649**	**0.692**	0.013	*	0.111	0.640	0.596	0.046	0.029	**0.000**	0.003
GBR-liz	**0.727**	**0.706**	**0.691**	**0.695**	**0.628**	**0.679**	0.032	0.016	*	0.293	0.280	0.066	0.019	0.000	0.008
GBR-w	**0.752**	**0.755**	**0.724**	**0.727**	**0.649**	**0.709**	–0.009	–0.011	0.007	*	0.749	0.199	0.021	0.001	0.009
OKI	**0.715**	**0.698**	**0.674**	**0.679**	**0.602**	**0.663**	–0.014	–0.007	0.006	–0.020	*	0.449	0.112	0.002	0.072
ROTA	**0.704**	**0.669**	**0.657**	**0.664**	**0.594**	**0.652**	0.029	0.028	0.030	0.016	–0.003	*	0.005	0.011	0.000
PNG	**0.753**	**0.756**	**0.727**	**0.728**	**0.657**	**0.711**	0.083	0.043	0.054	0.074	0.027	0.085	*	**0.000**	0.325
HAW	**0.741**	**0.720**	**0.703**	**0.707**	**0.644**	**0.694**	**0.211**	**0.241**	0.204	0.211	0.167	0.096	**0.275**	*	**0.000**
FP	**0.762**	**0.756**	**0.737**	**0.737**	**0.675**	**0.720**	**0.102**	0.071	0.067	0.100	0.035	0.106	0.003	**0.286**	*

AG = Arabian Gulf; CK = Cocos Keeling Islands; FP = French Polynesia; GBR-liz = Lizard Island, Great Barrier Reef; GBR = Whitsundays, Great Barrier Reef; HAW = Hawaii; OKI = Okinawa; OM = Oman; PNG = Papua New Guinea; SEYm/a/f = Mahe/Amirante/Farquhar, Seychelles; WA-i = Abrolhos islands, Western Australia WA-o = Scott and Clerk reefs, Western Australia

Notes: Values in bold indicate significant values ($P < 0.05$) after Bonferroni corrections ($P < 0.00042$)

F_{ST} values are based on 20,022 permutations

Table 3.2b Population pairwise F_{ST} values between 6 phylogenetic clades in bottom diagonal and associated *P*-values in top diagonal

Clade	I-1	I-2	P-1	P-2	P-4	P-8
I-1	*	0.000	0.000	0.000	0.000	0.000
I-2	0.310	*	0.000	0.000	0.000	0.000
P-1	0.707	0.671	*	0.000	0.000	0.000
P-2	0.835	0.793	0.514	*	0.000	0.000
P-4	0.813	0.757	0.410	0.575	*	0.000
P-8	0.822	0.761	0.543	0.743	0.715	*

I = Indian Ocean; P = Pacific Ocean

Note: Values are based on 20 022 permutations

Table 3.3a Geographic analysis of molecular variance (AMOVA), based on 15 populations

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics* (*P*)
Among groups (FCT)	1	1352.903	8.59025 Va	63.20%	0.63204 (0.05)
Among populations within groups (FSC)	13	291.168	0.78767 Vb	5.80%	0.15750 (0.05)
Within populations (F_{ST})	334	1407.236	4.21328 Vc	31.00%	0.69000 (0.05)
Total	**348**	**3051.307**	**13.59120**		

Notes: Groups correspond to the Pacific Ocean and the Indian Ocean
Significance tests are based on 20 022 permutations

Table 3.3b Regional analysis of molecular variance (AMOVA) based on 14 populations partitioned into 5 regions

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics* (*P*)
Among groups (FCT)	4	1552.019	7.52114 Va	63.26%	0.63263 (0.01)
Among populations within groups (FSC)	10	88.556	0.20472 Vb	1.72%	0.04687 (0.01)
Within populations (F_{ST})	329	1369.564	4.16281 Vc	35.01%	0.64985 (0.01)
Total	**343**	**3010.140**	**11.88867**		

Notes: The five regions are: 1. Arabian Gulf, 2. Oman, 3. central Indian Ocean (Seychelles and Cocos Keeling islands), 4. west/central Pacific Ocean + French Polynesia, 5. Hawaii.

Table 3.3c Analysis of molecular variance (AMOVA) based on phylogenetic clade structure

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics* (*P*)
Among groups (FCT)	1	1327.268	7.80843 Va	57.15%	0.57148 (0.01)
Among populations within groups (FSC)	4	592.861	3.00270 Vb	21.98%	0.51284 (0.01)
Within populations (F_{ST})	321	915.596	2.85232 Vc	20.88%	0.79124 (0.01)
Total	**326**	**2835.725**	**13.66346**		

Notes: The 6 clades have been partitioned into Indian Ocean and Pacific Ocean groups.
Significance tests are based on 20 022 permutations

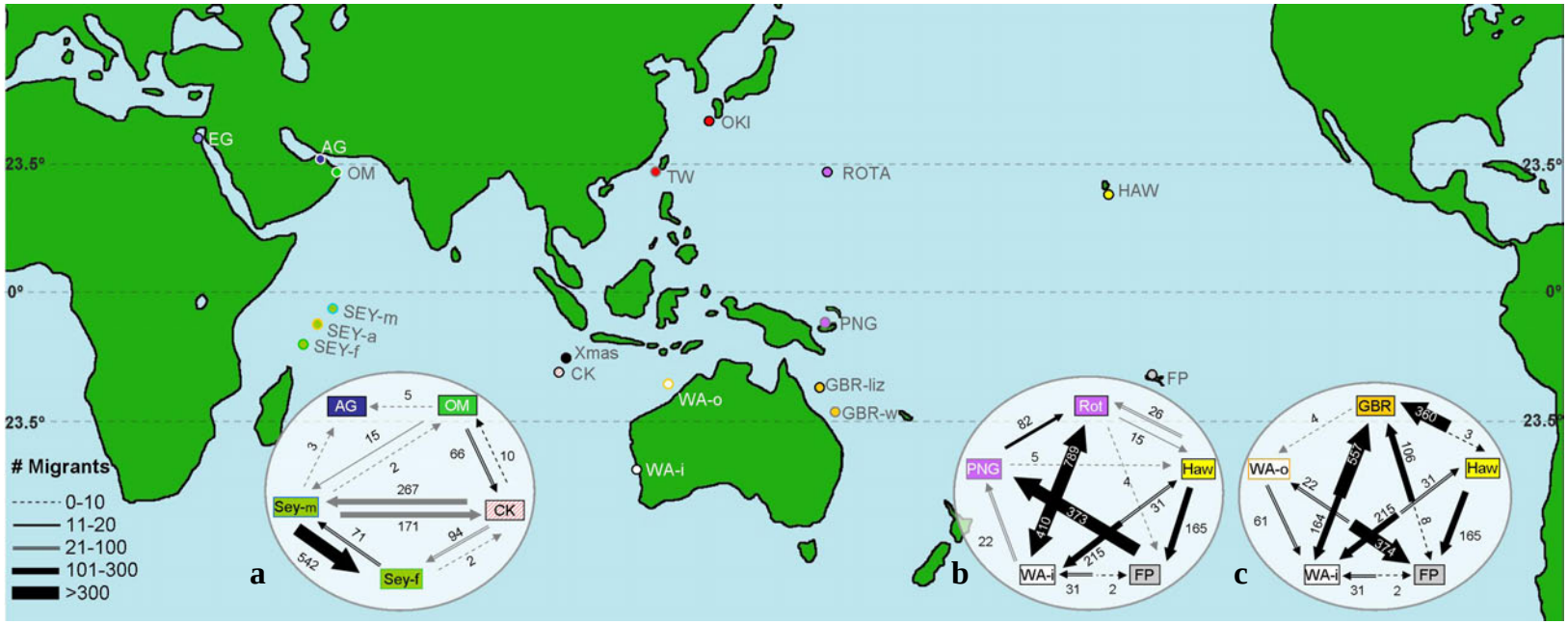


Figure 3.4(a-c) Schematic diagram of migration between populations of *Chlorurus sordidus* in the Indian Ocean (a), and the Pacific Ocean (b-c).

Abbreviations: AG = Arabian Gulf; OM = Oman; CK = Cocos Keeling Islands; Sey-m: Mahe, Seychelles; Sey-a: Amirante, Seychelles; Sey-f = Farquhar, Seychelles; EG = Egypt; Xmas = Christmas Island; OKI = Okinawa; TW = Taiwan; Rot = Rota; HAW = Hawaii; FP = French Polynesia; WA-i = Western Australia-inshore; WA-o = Abrolhos Islands, Western Australia; PNG = Papua New Guinea; GBR-liz = Lizard Island , Great Barrier Reef; GBR-w = Whitsundays, Great Barrier Reef.

*Currents in the Arabian Sea and Bay of Bengal, including the North Equatorial Current, reverse during winter months; summer currents are shown here
Notes: Black and grey arrows with numbers correspond to direction of gene flow and average number of migrants
Black arrows represent non-overlapping 95% confidence intervals (CIs) and grey arrows represent overlapping 95% CIs
*Major ocean currents have also been depicted; red arrows indicate warm water currents and blue arrows indicate cold water currents

Indian Ocean clade (T_{MRCA} = 0.442 mya), although there is a fair degree of overlap between the 95% confidence interval ranges of these coalescence times (Table 3.4a). These dates are within the time frame of the divergence of Indian and Pacific Ocean *C. sordidus* populations based on a comprehensive phylogeny of scarine labrids and a fossil-calibrated divergence date (Choat *et al.*, 2012), as well as a recent phylogenetic study based on multiple nuclear markers (Smith *et al.*, 2008). Clade-based coalescence analyses yielded similar estimates for T_{MRCA}, with clade P-1 being the oldest (T_{MRCA} = 0.71 mya), and clades P-2 and I-1 the most recently coalesced (Figure 3.2b). The mismatch distribution curve for the Indian Ocean was unimodal with a peak around 7 base-pair differences between pairs of individuals (Harpending's r = 0.003, P_r = 0.981), whereas the Pacific Ocean mismatch curve was bimodal with peaks at 4 and 11 base-pair differences (Harpending's r = 0.002, P_r = 0.968) (Figure 3.2b, Table 3.4b). Coalescence estimates place the Hawaiian lineage as the oldest (T_{MRCA} = 1.16 mya), while Oman in the western Indian Ocean appears to be the most recently expanded population (T_{MRCA} = 0.227 mya) (Table 3.4a). Coalescence date estimates have been linked to low sea-level stands, which are associated with high ^{18}O ratios (Chappell, Shackleton, 1986; Clemens *et al.*, 1996).

3.4 Discussion

In our extended analysis, we confirm that the genetic differentiation reported by Bay et al. (2004) straddles the biogeographic break at the Indo – West Pacific Barrier (e.g., Froukh, Kochzius, 2008; Marie *et al.*, 2007), and comprises a western Indian Ocean (Seychelles to Christmas and Cocos Keeling islands) and a Pacific Ocean lineage (Christmas Island, Western Australia, the Indo–Australian Archipelago and the central Pacific Ocean). The Indo – West Pacific Barrier is not a hard barrier, and is thus penetrated by other widespread marine taxa. However, for taxa that are affected by the Indo – West Pacific Barrier, the location of the barrier shifts, with some Indian Ocean and Pacific Ocean faunae overlapping at the Cocos Keeling Islands and others at Christmas Island (Hobbs, Salmond, 2008). For *C. sordidus*, the presence of Christmas Island haplotypes in both the Indian Ocean and Pacific Ocean clades suggests that the barrier for this species is located nearest Christmas Island. Despite the small number of samples from Christmas Island, the combination of the lack of haplotypes from the Cocos Keeling Islands in the Pacific Ocean clade, as well as the proximity of the Cocos Keeling Islands to Christmas Island (1000 km), provides additional support that the location for the western Indian Ocean – Pacific Ocean split for this species lies at Christmas Island. Results from the AMOVA, which indicate structure at the largest spatial scale, further support this, and are consistent with previous findings by Bay et al. (2004).

Western Indian Ocean and Pacific Ocean populations have been subject to very different evolutionary histories following the initial separation of the species 1.78–2.49 mya. The location and timing of the genetic break suggest that the differentiation between the two major clades is due to more than simply isolation by distance. The Christmas Island region has been linked to other Indian or Pacific ocean faunal divisions (e.g., McMillan, Palumbi, 1995; Palumbi, 1997), and is a known area of Pacific Ocean and Indian Ocean species overlap (Hobbs, Salmond, 2008). This geographic region has had a complex geological history—rife with volcanic activity, exposure of the Sunda Shelf during periods of low sea level stand, periodic flooding, and seasonal changes in winds and ocean currents. In addition, the timing of the diversification coincides with two glacial cycles in close succession, during which the sea levels dropped 100 m below present levels (Liu *et al.*, 1998). Paleontological data based on foraminifera also provide evidence of cold water upwelling in the east Indian Ocean during the Pliocene and Pleistocene, which resulted in the mass extinction of deep-sea foraminifera (Hayward, 2002; Kawagata *et al.*, 2005). These upwelling events may have significantly reduced dispersal between western Indian Ocean and Pacific Ocean populations in the past.

Limited dispersal between ocean basins in the past has contributed not only to population structure at the scale of the entire Indo–Pacific, but also within each ocean basin. In the northern Indian Ocean, we identify two populations, one from Oman and the other from the Arabian Gulf.

In the Pacific Ocean, there are differentiated populations from Hawaii and French Polynesia. Most of these additional differentiated populations are associated with known biogeographic breaks such as the Strait of Hormuz, which separates the gulfs of Oman and Arabia, and the Central Pacific Barrier, which separates Hawaii from the west Pacific Ocean. In addition to the Central Pacific Barrier, the East Pacific Barrier also poses a significant barrier to dispersal. The East Pacific Barrier is not an absolute barrier for the soldierfish *Myripristis berndti*, which showed structure between some, but not all, west and east Pacific populations (Craig *et al.*, 2007). However, it is an absolute barrier to dispersal for *C. sordidus* as well as two acanthurids that have longer PLDs and larvae with more advanced swimming capabilities than both *M. berndti* and *C. sordidus* (Horne *et al.*, 2008; Klanten *et al.*, 2007). This observation is evidence that while PLD is an important determinant in species dispersal capabilities, other factors are also relevant. Interestingly, the same acanthurids above did not show structure at the Indo – West Pacific Barrier; however, both *M. berndti* and *C. sordidus* did.

French Polynesian (Moorea and Tahiti) populations were less consistently differentiated from the other populations in the Pacific Ocean, and did not span any known biogeographic breaks. Other marine organisms have shown genetic partitioning at this location, including another

scarine labrid, *Scarus psittacus* (Winters *et al.*, 2010), the lutjanid *Lutjanus kasmira* (Gaither *et al.*, 2010) and a species of intertidal snail (Reid *et al.*, 2006), all of which, like *C. sordidus*, are widely distributed across the Indo–Pacific. This pattern is likely due to a combination of oceanographic currents, isolation by distance and historical separations evident by the absence of French Polynesian individuals in the Hawaii-dominated clade. It is interesting to note that in the Pacific Ocean, there are four main lineages. One is dominated by Hawaiian individuals and the remaining three contain approximately equal proportions of individuals from all other locations. This mixture of individuals from these Pacific Ocean populations appears several times in the phylogenetic tree. This is evidence of a number of temporal rather than spatial separations within the west Pacific Ocean, as has been reported for three widespread acanthurid species throughout the Indo–Pacific (Horne *et al.*, 2008; Klanten *et al.*, 2007). Analyses based on phylogenetic groupings suggest that clade P-1, which contains the widest mixture of individuals from the Pacific Ocean, including Christmas Island and Hawaii, is the oldest lineage, followed by the Hawaii-dominated clade, P-8. Location-based coalescence estimates also suggest that Hawaii is the oldest lineage, which may be an indication of a central Pacific origin of extant populations, followed by an east-to-west migration of *C. sordidus*. This is consistent with the

Table 3.4a Coalescence analysis parameters for Pacific Ocean and Indian Ocean clades based on geography

Clade	Tau τ (95% CI)	$Theta_0$ Θ_0	$Theta_1$ Θ_1	SSD	Raggedness (*P*)	T_{MRCA} my (lower–upper)
Indian Ocean	5.833 (3.3–15.7)	4.065	77.598	0.00047	0.00214968 (ns)	0.441 (0.248–1.188)
Arabian Gulf	8.369 (4.28–12.151)	0.001	23.881	0.00500	0.012 (ns)	0.634 (0.324–0.920)
Oman	3.0 (0.473–4.797)	0.375	0.625	0.06200	0.225 (ns)	0.227 (0.035–0.363)
SEY/CK	5.374 (3.688–11.886)	2.814	120.137	0.00000	0.004 (ns)	0.407 (0.279–0.9)
Pacific Ocean	11.417 (6.7–15.2)	0.002	29.865	0.00150	0.00254851 (ns)	0.864 (0.510–1.153)
W/C	10.951 (6.337–14.735)	0.000	28.565	0.00200	0.003 (ns)	0.830 (0.480–1.116)
Hawaii	15.372 (8.014–26.197)	0.000	17.264	0.02100	0.024 (ns)	1.165 (0.607–1.985)
French Polynesia	5.462 (1.687–15.941)	3.078	43.105	0.00800	0.014 (ns)	0.414 (0.128–1.208)

CI = confidence interval; CK = Cocos Keeling Islands; my = million years; ns = not significant SSD = sum of squared deviations; SEY = Seychelles; T_{MRCA} = coalescence time / time to most recent common ancestor; W/C = west/central Pacific

Notes: Tau (τ ± 95% CI) is the unit of mutational time between two populations; $theta_0$ (Θ_0) is the mutation parameter before expansion; and $theta_1$ (Θ_1) is the mutation parameter after expansion.

Harpending's raggedness index and significance based on the significance of simulated and observed raggedness are also shown

centre of accumulation model of Indo-Australian Archipelago diversity whereby species move inward following speciation outside the Indo-Australian Archipelago (Jokiel, Martinelli, 1992; Ladd, 1960).

The patterns of gene flow corroborate a westerly migration of *C. sordidus*, with significantly more migrants from Hawaii to other Pacific Ocean populations. This pattern can be explained by the North Equatorial Current, which moves warm water west past Hawaii from the Californian coast, coupled with the absence of tropical currents towards Hawaii from the west. French Polynesia is another location with highly asymmetrical gene flow: the large number of unidirectional migrants westward to Papua New Guinea from French Polynesia can be explained by the South Equatorial Current. Although many of these patterns can be attributed to oceanographic currents, there are also many instances of gene flow going against present-day currents. For example, there were more than 17 times as many migrants from offshore Western Australian reefs compared to French Polynesia. Similarly, gene flow from Rota to inshore Western Australian reefs contradicts present-day currents. This suggests that past patterns of ocean circulation were different to present-day conditions. Gene flow in the Indian Ocean also indicates a mixture of present-day and historical oceanographic conditions. Most marked is the

Table 3.4b Coalescence analysis parameters for clades determined by phylogenetic analysis

Clade	Tau τ (95% CI)	Theta$_0$ Θ_0	Theta$_1$ Θ_1	SSD	Raggedness (*P*)	T_{MRCA} my (lower–upper)
Indian Ocean	5.833 (3.3–15.7)	4.065	77.598	0.00047	0.00214968 (ns)	0.441 (0.248–1.188)
I– 1	3.716 (1.778–13.697)	3.036	31.626	0.003	0.014 (ns)	0.282 (0.135–1.038)
I – 2	5.315 (2.409–16.883)	3.904	39.648	0.003	0.005 (ns)	0.403 (0.183–1.279)
Pacific Ocean	11.417 (6.7–15.2)	0.002	29.865	0.00150	0.00254851 (ns)	0.864 (0.510–1.153)
P – 1	9.433 (5.626–12.326)	0.000	47.535	0.001	0.005 (ns)	0.715 (0.426–0.934)
P – 2	3.397 (2.256–6.442)	0.861	66.592	0.000	0.014 (ns)	0.257 (0.171–0.488)
P – 4	4.32 (2.412–5.55)	0.001	86.914	0.008	0.028 (ns)	0.327 (0.183–0.420)
P – 8	7.227 (2.628–15.102)	0.000	3.932	0.011	0.029 (ns)	0.548 (0.199–1.144)

CI = confidence interval; I = Indian Ocean; my = million years; ns = not significant; P = Pacific Ocean; SSD = sum of squared deviations; T_{MRCA} = coalescence time / time to most recent common ancestor

Notes: Tau ($\tau \pm 95\%$ CI) is the unit of mutational time between two populations; theta$_0$ (Θ_0) is the mutation parameter before expansion; and theta$_1$ (Θ_1) is the mutation parameter after expansion.

Harpending's raggedness index and significance based on the significance of simulated and observed raggedness are also shown

lack of gene flow from the Arabian Gulf to any other Indian Ocean location sampled. There were also limited numbers of migrants from Oman and Mahe, Seychelles, to the Arabian Gulf, which may be due to diverging currents between these locations (i.e. the North and South Equatorial currents). The unidirectional gene flow at this location is reflected in the high level of genetic structure between the Arabian Gulf and other Indian Ocean populations. In addition, the unidirectional migration would suggest that individuals in the Arabian Gulf came from Oman and the Seychelles. This is also supported by the haplotypic similarity between individuals from the Arabian Gulf and Oman. Gene flow from Oman to the Seychelles was limited, with more migrants moving south to Mahe, Seychelles, than north to Oman. Interestingly, migration between Oman and Farquhar, Seychelles, was non-existent, despite the large number of migrants from Mahe to Farquhar. Overall, Indian Ocean gene flow was high, and generally in the westerly direction along low latitudes (i.e. the Cocos Keeling Islands to Mahe) and in the southerly direction along the east African coast, with progressively fewer migrants with increasing distance between populations. Similarly, gene flow in the Pacific was predominantly in the westerly direction along low latitudes (i.e. French Polynesia to Papua New Guinea), in the southerly direction in the central and south Pacific Ocean, and in the northerly direction in the Indo–Australian Archipelago. These results indicate that present-day ocean circulation is generally responsible for patterns of distribution; however, historical separations and connections between populations are also evident based on gene flow, which contradict present-day ocean currents. For example, gene flow was much stronger from inshore Western Australia populations to the Great Barrier Reef despite the westerly flow of the Leeuwin Current. Likewise, gene flow from the Great Barrier Reef to French Polynesia exists despite the opposing East Australian Current. The presence of gene flow in the easterly direction in the Indian Ocean (i.e. Seychelles to the Cocos Keeling Islands) is made possible by the periodic reversal of the Northern Equatorial Current during monsoon or El Niño periods.

Historical fragmentation of populations is evident in *C. sordidus*, and is most likely responsible for the bimodal mismatches observed in populations from the west Pacific Ocean (including east Indian Ocean populations), Hawaii and Oman. A bimodal mismatch distribution curve is often attributed to stable population sizes (Rogers, Harpending, 1992), but it may also be due to historical population fragmentation (Ray *et al.*, 2003). In the latter case, the initial mode represents similarities between haplotypes of a recent ancestor and the second mode represents the haplotypes similar to the common ancestor at the time of the expansion (Ray *et al*., 2003). In contrast, a smooth unimodal mismatch curve is generally an indication of recent expansion as a result of an excess of unique haplotypes. Mismatch curves for the separate Pacific Ocean populations are bimodal. For the Pacific Ocean overall, our results support a long history of population expansion and contraction with secondary contact between allopatrically diverged

lineages. In the case of Hawaii, the negative values for Tajima's D and Fu's Fs statistic also support a history of population expansion. In addition, the combination of high nucleotide and haplotype diversity, as well as the star-shaped architecture of the Hawaiian-dominated haplotypes, suggest both an expansion within Hawaii and more recent dispersal from numerous other Pacific Ocean locations to Hawaii; for example, secondary contact on Hawaiian reefs between allopatrically differentiated lineages (Grant, Bowen, 1998).

In the Indian Ocean, Oman is the only population displaying a bimodal mismatch distribution curve, while, overall, the Indian Ocean has a clearly unimodal mismatch curve. The Oman population appears to have a similar genetic demography to the Hawaiian population, with evidence of a single common-shared haplotype shared exclusively between Omani individuals, as well as more recent recruitment to Oman from the Cocos Keeling Islands or the Seychelles (or both). Overall, the Indian Ocean coalescence time appears to be younger than that of the Pacific Ocean, which is also evident in the population statistics and the high frequency of unique haplotypes in the minimum-spanning network. Coalescence estimates indicate that Oman is the most recently diverged population and the Arabian Gulf is the most ancient among extant populations within this ocean basin, although the Arabian Gulf was most likely sourced from Oman more than once. The low haplotype diversity in Oman also suggests that this population has undergone a strong bottleneck. This is indicated by the large number of shared haplotypes in this population and the population genetic statistics, which indicate a lack of population growth. The high genetic distance between Oman and all Indian Ocean locations, except for the Seychelles, also supports a history of isolation from the rest of the Indian Ocean. This may be due to a genetic bottleneck event in recent history that caused the loss of genetic diversity, hence the presence of the Oman-exclusive shared haplotype in the minimum-spanning network. This explanation is supported by the recent age of the Oman haplotypes based on coalescence analyses. Caution should be exercised, however, when interpreting gene trees based on a single marker due to the effects of stochasticity inherent in both the mitochondrial and the nuclear genome (Ballard, Whitlock, 2004).

The Arabian Gulf appears to have had a history of numerous population contractions and expansions, which is consistent with this area being completely isolated from the rest of the Indian Ocean by a land barrier during periods of low sea-level stand. This is supported by the fact that the vast majority of the Arabian Gulf haplotypes are most similar to the common haplotype shared exclusively by individuals from Oman and by the fact that Oman appears to be a source population for the Arabian Gulf. This is one of the most obvious examples of past vicariance driving population structure in this species. Although the Arabian Gulf is very close to Oman, there is a high degree of genetic differentiation between the two populations. During

periods of low sea-level stand, temperatures and salinity in the Arabian Gulf would have dropped significantly. Even today, temperatures in the Arabian Gulf reach subtropical temperatures that many tropical coral reef species would not tolerate. It is possible that the population observed in the Arabian Gulf has maintained its genetic differentiation due to an adaptation to the more extreme thermal conditions. Somero (2002) found that ectothermic organisms, such as fish, experience external temperatures in their mitochondria. Thus, thermal adaptations, as well as the relative fitness of other haplotypes, are likely to change more rapidly than in the nuclear genome. A comparative phylogeographic study conducted on a similar species or using different (nuclear) markers would be useful in teasing apart these effects.

In conclusion, the diversification following the initial separation of *C. sordidus* has been greatly affected by historical oceanographic conditions. Based on mitochondrial evidence, extant populations of *C. sordidus* appear to have originated in the Pacific Ocean during the Pliocene, and then diverged into separate Indian Ocean and Pacific Ocean lineages following range contractions resulting from glacial sea-level changes. It is also evident that the two main lineages have undergone significantly different evolutionary histories, characterised by signals of secondary contact between lineages that have diverged allopatrically, as well as recent expansions following strong population bottlenecks. Finally, the fact that the temporal clades demonstrate very high levels of genetic partitioning, which is confirmed by pairwise F_{ST} comparisons, further supports a history of fragmentation in this species, particularly among individuals from peripheral locations, including Hawaii, French Polynesia, Oman and the Arabian Gulf.

Chapter 4 Strong population structure, zones of overlap and peripheral isolation in the widespread parrotfish, *Scarus rubroviolaceus* (Perciformes: Scarinae)

Abstract

We compared 378 bp from the mitochondrial control region of the scarine labrid *Scarus rubroviolaceus* in order to determine the population structure and genetic connectivity across its Indo–Pacific distribution. We found evidence of significant genetic differentiation between western Indian Ocean and Pacific Ocean populations based on phylogenetic and population genetic data ($0.529 \leq F_{ST\ WIO\text{-}PO} \leq 0.853$), as well as population structure between the eastern peripheral sampling locations within the Pacific Ocean, including Hawaii ($0.188 \leq F_{ST\ HAW} \leq 0.540$), the Marquesas ($0.037 \leq F_{ST\ MQS} \leq 0.540$) and the east Pacific ($0.115 \leq F_{ST\ EP} \leq 0.395$). These results were corroborated by an analysis of molecular variance (AMOVA), which showed the majority of the variance attributed to the different ocean basins (FCT = 42.62%, $P < 0.01$) and the remaining variance evenly split among populations within groups. We also found significant isolation by distance within each ocean basin ($Z_{WIO} = 290.5452$, $R^2 = 0.470$, $P < 0.05$; $Z_{PO} = 51568.0352$, $R^2 = 0.295$, $P < 0.05$), as well as highly asymmetrical gene flow between populations, with the majority of migration patterns coinciding with present-day oceanographic currents. Demographic estimates reveal that the coalescence of the two clades is consistent with Pleistocene sea-level changes, with evidence of the western Indian Ocean clade coalescing prior to the central east Pacific clade (0.988 and 0.778 million years ago [mya] respectively). Results from this study emphasised that Christmas Island (in the eastern Indian Ocean) is an important contact zone between the two ocean basins, with haplotypes from both western Indian Ocean and Pacific Ocean clades co-occurring there.

4.1 Introduction

Recent marine phylogeographic studies have demonstrated significant genetic partitioning between Indian Ocean and Pacific Ocean populations of widespread marine organisms, including the soldierfish *Myripristis berndti* (Craig *et al.*, 2007), the coconut crab *Birgus latro* (Lavery *et al.*, 1996b), several Echinolittorina snails (Reid *et al.*, 2006) and the parrotfish *Chlorurus sordidus* (Bay et al. 2004; Chapter 3). The Indo – West Pacific Barrier (Froukh, Kochzius, 2008; Marie *et al.*, 2007), a marine biogeographic barrier created by the

exposure of the Sunda Shelf and the closure of the Torres Strait during times of low sea-level stand (Voris, 2000), has been implicated in many of the genetic partitions found between Indian Ocean and Pacific Ocean fauna, including the lutjanid *Lutjanus kasmira* (Gaither *et al.*, 2010) and the parrotfish *C. sordidus* (Bay et al. 2004; Chapter 3). These are just a few examples of the importance of historical oceanographic conditions in shaping coral reef communities today. For some marine species, the area south of the Indo – West Pacific Barrier surrounding Cocos Keeling and Christmas islands is an area where fauna from both ocean basins meet (Hobbs, Salmond, 2008) and in some instances hybridise (Hobbs *et al.*, 2009). This area of overlap is known for being geologically and geomorphologically dynamic, and has had a long history of volcanic activity and upwelling (Brewer *et al.*, 2009; Woodroffe, Berry, 1994). These studies suggest that, although the capacity for broad-scale dispersal exists in many marine organisms, dispersal across biogeographic barriers may become restricted over evolutionary timescales, resulting in bifurcating lineages.

In addition to partitioning over large spatial scales across ocean basins, genetic partitioning at smaller spatial scales has been detected in a number of species, including the damselfish *Dascyllus trimaculatus* (Bernardi *et al.*, 2001). In particular, both isolated and peripheral habitats have yielded populations with lowered genetic diversity and population differentiation in a number of species including several marine invertebrates (Duda, Lee, 2009; Lavery *et al.*, 1996a; Malay, Paulay, 2009), reef corals (Budd, Pandolfi, 2010), atherinid fish *Craterocephalus capreoli* (Johnson *et al.*, 1994), as well as a number of reef fishes (Rocha, Bowen, 2008), including the parrotfishes *S. psittacus* (Winters *et al.*, 2010) and *C. sordidus* (Bay et al. 2004; Chapter 3). The increasing incidence of distinctive populations in peripheral habitats among marine species highlights the importance of peripheral or isolated habitats in reef fish biodiversity.

We provide phylogenetic and phylogeographic data from another species of scarine labrid, *S. rubroviolaceus*, whose distribution range transcends that of *C. sordidus* and includes the tropical east Pacific. Results from this chapter will allow a comparative phylogeographic analysis between members of the same family with similar distributions and life-history traits. Like *C. sordidus*, *S. rubroviolaceus* is reef-associated, but it is larger in size (Myers, 1999), has a wider range, which extends beyond the East Pacific Barrier, and has a longer pelagic larval duration (PLD) of 45–55 days (Victor 2009, pers. comm., 2 Feb.). As with *C. sordidus*, these characteristics make *S. rubroviolaceus* another good species for testing models of Indo-Australian Archipelago diversity, which are outlined in chapter 1. Given the similarities and differences between *C. sordidus* and *S. rubroviolaceus*, we wanted to determine whether

S. rubroviolaceus exhibits congruent patterns of genetic structuring in space and in time, indicative of a shared evolutionary history. Specifically, we aim to address the following questions:

1. Is there evidence of partitioning between Indian Ocean and Pacific Ocean populations? If such partitioning exists, is it associated with any known marine biogeographic barriers?

2. Is there a pattern of temporal and geographical partitioning both between and within ocean basins?

3. What is the magnitude and direction of genetic connectivity between populations within and between ocean basins?

4. When did the different lineages expand and which population acted as the likely source population for extant populations of this species?

4.2 Materials and methods

4.2.1 Sampling, DNA amplification and sequencing

To determine the population structure and connectivity of the widely distributed scarine labrid *S. rubroviolaceus*, we compared mitochondrial DNA sequences from 292 individuals from a total of 15 sampling locations throughout its distribution in the Indian and Pacific oceans (Table 2.1, Figure 4.1). Nearly all samples used in this study were collected between 2000 and 2007 by spearfishing, and some samples from Okinawa and all samples from Palau were collected from fish markets where fish were caught by local fishermen.

The samples that could not be amplified using universal control region primers were successfully amplified using control region primers designed for *C. sordidus* (Csor-F, Csor-R), as well as primers designed for *S. ghobban* and *S. rubroviolaceus* (SghDL-F, SruDL-F and SghruDL-R) (Table 2.2). Further detailed laboratory procedures were performed as indicated in Section 2.2.

4.2.2 Phylogenetic analyses

A phylogenetic analysis was conducted on all *S. rubroviolaceus* sequences in order to determine the evolutionary relationships between populations and to inform population

genetic analyses. Evolutionary relationships between populations were inferred from a consensus of the 50 best Bayesian trees generated in PAUP*4.0 (Swofford, 2003). Trees were outgroup-rooted using sequences from sister species *S. ferrugineus* and *S. persicus*. Further detailed phylogenetic analyses were performed as indicated in Section 2.3.

4.2.3 Population genetic analyses

Individuals from locations with multiple sites were pooled with their closest neighbour if sample sizes were too small (n<10), resulting in a total of 12 populations for subsequent population genetics analyses (Figure 4.1). The pooled sample locations were: Okinawa (n=3) + Taiwan (n=34), Palau (MIC1, n=3) + Pohnpei (MIC2, n=17), and Clipperton Atoll (EP1, n=2) + Panama (EP2, n=14). Locations were pooled only after initial pairwise F_{ST} comparisons confirmed high genetic connectivity between those locations. Where noted, individuals from the Cocos Keeling Islands and Christmas Island have been pooled together, depending on clades determined by phylogenetic analyses. Three separate AMOVA analyses were conducted. The first consisted of 12 populations grouped by ocean basin (all Cocos Keeling and Christmas islands individuals in the Indian Ocean group), the second consisted of 12 populations grouped into 5 geographic regions based on pairwise F_{ST} comparisons, and the third consisted of individuals grouped into 2 genetically differentiated clades determined by the phylogenetic analysis (Cocos Keeling and Christmas islands individuals separated into Indian Ocean and Pacific Ocean groups). Further details have been provided in Section 2.4.

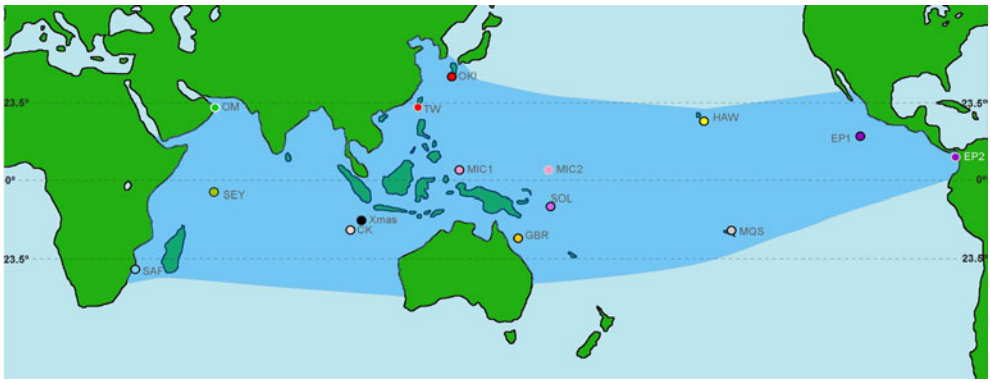


Figure **4.1 Map depicting distribution (shaded in darker blue) and locations sampled for *S. rubroviolaceus*** Location codes and numbers sampled from each location as follows: OM = Oman (25); SEY = Seychelles (60); SAF = South Africa (21); CK = Cocos Keeling Islands (6); Xmas = Christmas Island (35); OKI = Okinawa, Japan (3); TW = Taiwan (34); MIC1 = Palau, Micronesia (3); MIC2 = Pohnpei, Federated States of Micronesia (17); SOL = Solomon Islands (10); GBR = Great Barrier Reef (14); HAW = Hawaii (38); MQS = Marquesas (10); EP1 = Clipperton Atoll, east Pacific (2); EP2 = Panama, east Pacific (14).

4.2.4 Coalescence analyses

The generation time for *S. rubroviolaceus* (t_2) was calculated using the formula $t_2 = (\alpha+\omega)/2$ (Pianka, 1978), where the age at first reproduction (α) was 2.5 years and age at last reproduction (ω) was 13.5 years (Choat pers comm.). Further details have been provided in Section 2.5.

4.3 Results

4.3.1 Phylogenetic analyses

We analysed sequences of 378 base pairs from the mitochondrial control region of 292 *S. rubroviolaceus* individuals from throughout its Indo–Pacific distribution (Figure 4.1). Of the 378 base pairs, 114 were polymorphic and 82 were parsimony informative (transition:transversion [TI:TV] ratio = 15.8164). The base frequencies for *S. rubroviolaceus* were AT-rich (A = 35.33%, T = 26.50%, C = 25.33%, G = 12.84%), which is common in reef fish (McMillan, Palumbi, 1997), including other parrotfish (Bay *et al.*, 2004; Dudgeon *et al.*, 2000).

The best nucleotide substitution model selected in MrModeltest differed between hierarchical likelihood ratio tests (hLRTs) (GTR + I + G; G = 0.5464; GTR = general time reversible) and Akaike's information criterion (AIC) (HKY + I + G; G = 0.5293; HKY = Hasegawa, Kishino and Yano model). A Bayesian analysis was carried out in MrBayes via CIPRES webportal (Miller *et al.*, 2010) implementing the AIC model. A 50% consensus tree was generated after 20,000,000 generations using the best 7902 trees. There was strong bootstrap support for a division between west Indian Ocean and east Indo-Pacific populations, with evidence of both Indian and Pacific ocean affinities of Cocos Keeling Islands and Christmas Island populations (Figure 4.2a). Approximately 80% of Cocos Keeling Islands and Christmas Island individuals grouped in the Pacific Ocean clade, and the remaining minority grouped in the Indian Ocean clade. Phylogenetic analyses did not reveal further structuring at the ocean basin level as was shown in Chapter 3 (Figure 3.2a), since there was no support for subclades within the two major clades (Figure 4.2a).

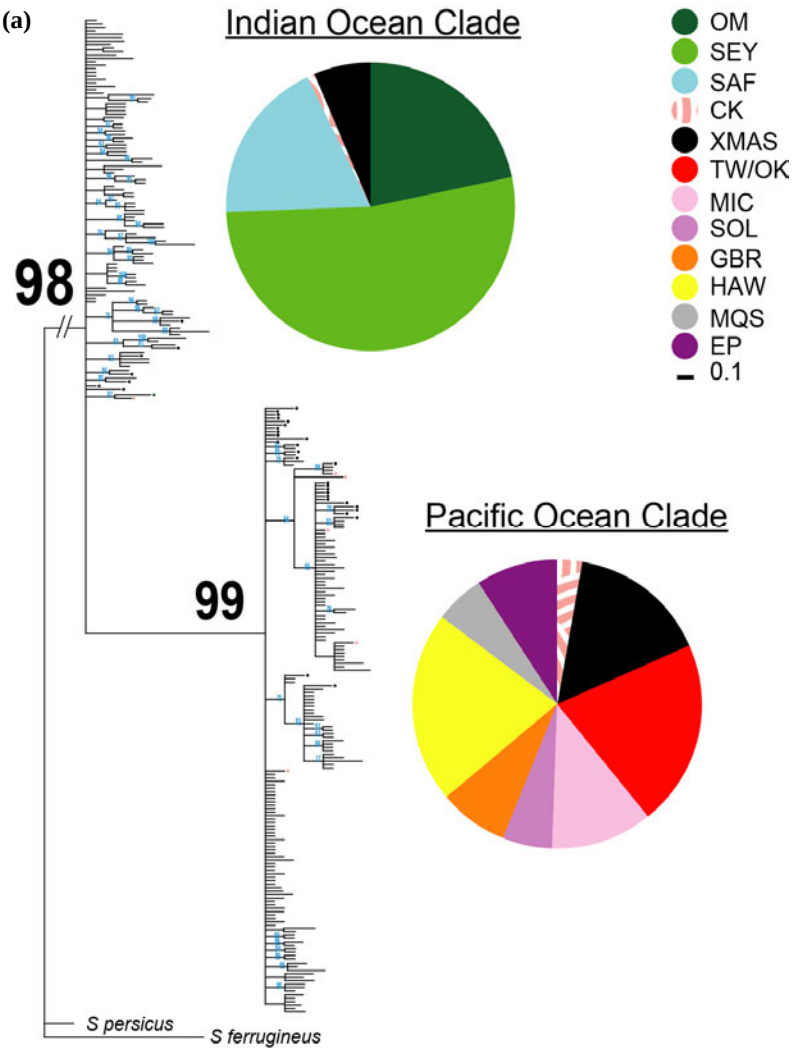
Comment [SK1]: WHERE is the support on the tree??? There is none!

Comment [CBM2]: There is strong support in the Bayesian tree now.

4.3.2 Population genetics analyses

The single minimum-spanning network revealed the same split in Pacific Ocean and Indian Ocean haplotypes, with the identical Cocos Keeling Islands and Christmas Island individuals spread between the two clades (Figure 4.3). The Pacific Ocean clade contained a larger proportion of shared haplotypes, with 8 common haplotypes and many unique haplotypes (50.8%)

(a)



(b)

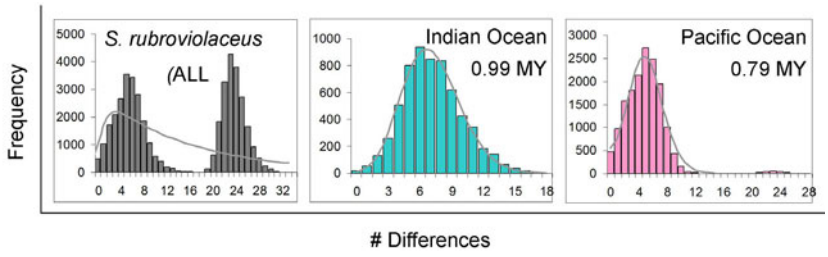


Figure 4.2(a) 50% majority rule consensus tree of all sampled trees (79 802) for *S. rubroviolaceus*.
Large coloured pies indicate frequency and geographic breakdown of Indian and Pacific Ocean clades. Posterior probabilities are indicated in blue. Tree is outgroup rooted with S. persicus and S. ferrugineus from the western Indian Ocean.

Figure 4.2(b) Mismatch distribution curves & coalescence times for all *S. rubroviolaceus clades* as well as Indian Ocean and Pacific Ocean clades
Bars indicate observed differences, curved lines indicate model frequencies.

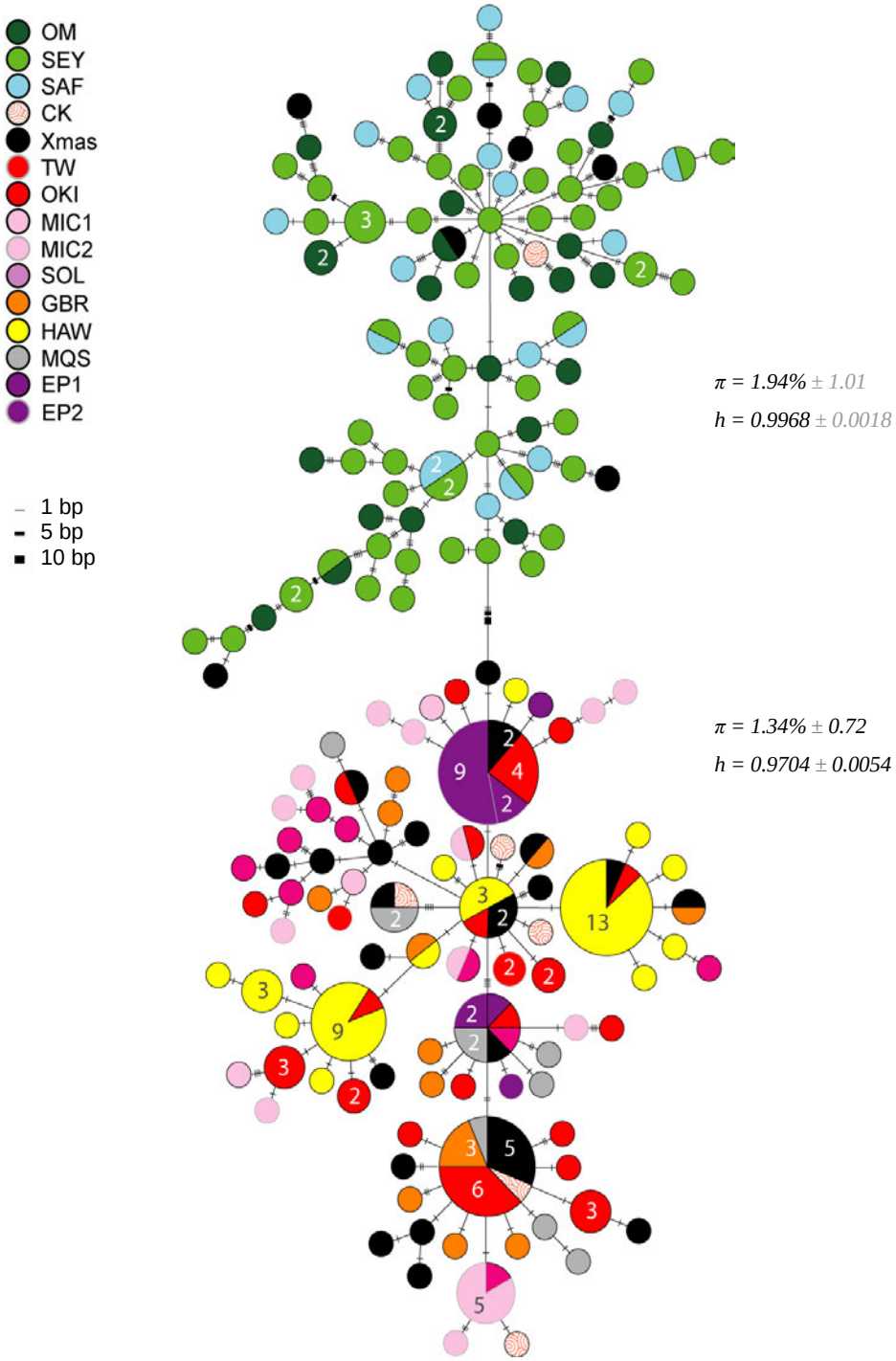


Figure 4.3 Minimum-spanning network for Pacific Ocean and Indian Ocean haplotypes, and respective nucleotide (π ± SE %) and haplotype (h ± SE) diversity indices
Small circles represent unique haplotypes and larger circles indicate shared haplotypes, with the colours and proportion of individuals from respective populations with that shared haplotype shown. Thin hash marks indicate a single base pair substitution between haplotypes, medium-sized bars indicate a 5 bp difference and thick bars represent 10 bp differences.

separated by 1 or 2 base pairs. There were very few exceptions, including one haplotype from the Cocos Keeling Islands separated by 6 substitutions. The Indian Ocean clade contained only a few shared haplotypes, which were shared by fewer than 3 individuals. There were many unique haplotypes (85%) differing between one another by 1–3 base pairs, but with at least 6 instances of more than 10 base pair changes between haplotypes. Most of the shared haplotypes in the Indian Ocean were shared between individuals from the Seychelles and South Africa.

The Indian Ocean clade was characterised by high nucleotide and haplotype diversities (π = 1.93–2.94%, h = 0.9798–0.9966) (Table 4.1). In contrast, the low proportion of unique haplotypes in the Pacific Ocean clade was reflected by more variable haplotype and nucleotide diversities. In the east Pacific, nucleotide (π) and haplotype (h) diversities ranged from 0.60% and 0.5167, respectively, to 3.17% and 1.0, respectively, at the Cocos Keeling Islands. The Cocos Keeling Islands, along with Christmas Island, had the greatest genetic diversity. Tajima's D and Fu's Fs were negative for both the Indian Ocean (D = –1.54; Fs = –24.85) and Pacific Ocean (D = –1.98; Fs = –25.27) clades; however, Tajima's D was not significant in the Indian Ocean ($P > 0.01$) (Table 4.2). A lack of significance in Tajima's D indicates that purifying or positive selection (or both) may be acting, and thus the null hypothesis of neutral evolution can be rejected. Fu's Fs was significantly negative in all

Table 4.1 Summary statistics for *S. rubroviolaceus*

Location	***n***	n_h	π ± SE (%)	***h*** ± SE	**D**	P_D	**Fs**	P_{FS}
Oman	25	23	2.33 ± 1.24	0.9933 ± 0.0134	–1.16338	0.11000	–13.72745	0.00010
Seychelles	60	55	1.94 ± 1.02	0.9966 ± 0.0040	–1.57973	0.02900	–24.98494	0.00000
South Africa	21	20	1.93 ± 1.05	0.9952 ± 0.0165	–1.05663	0.14330	–13.25405	0.00000
CK	6	6	3.17 ± 1.94	1.0000 ± 0.0962	–1.22423	0.09400	–0.72621	0.19960
Christmas Island	35	29	2.94 ± 1.52	0.9798 ± 0.0159	–0.42001	0.37650	–13.25633	0.00040
Total Indian Ocean	**147**	**133**	**1.94 ± 1.01**	**0.9968 ± 0.0018**	**–1.54**	**0.03000**	**–24.85**	**0.00000**
Okinawa/Taiwan	37	22	1.23 ± 0.69	0.9550 ± 0.0186	–1.10553	0.12790	–11.12544	0.00000
GBR	14	12	1.33 ± 0.77	0.9670 ± 0.0437	–1.15367	0.12100	–5.53256	0.00510
Micronesia	20	16	1.48 ± 0.83	0.9474 ± 0.0435	–1.11430	0.12650	–7.72941	0.00120
Solomon Islands	10	10	1.44 ± 0.86	1.0000 ± 0.0447	–0.77162	0.23280	–5.66800	0.00270
Marquesas	10	8	1.04 ± 0.65	0.9556 ± 0.0594	–0.42314	0.35570	–2.74935	0.04230
Hawaii	38	15	0.61 ± 0.38	0.8492 ± 0.0448	–0.84889	0.21870	–7.50529	0.00080
East Pacific	16	4	0.60 ± 0.39	0.5167 ± 0.1324	0.23458	0.63300	1.75111	0.83350
Total Pacific Ocean	**151**	**93**	**1.34 ± 0.72**	**0.9704 ± 0.0054**	**–1.98**	**0.00300**	**–25.27**	**0.00000**
Total	**292**	**220**	**4.00 ± *1.90***	**0.9883 ± *0.0023***	**–0.55927**	**0.33000**	**–23.74950**	**0.00300**

CK = Cocos Keeling Islands; D = Tajima's selective neutrality test; Fs = Fu's neutrality test; GBR = Great Barrier Reef; h = haplotype diversity; n = number of individuals sampled; n_h = number of haplotypes sampled; π = nucleotide diversity; PD = D and significance level; P_{FS} = Fs and significance level

Note: Micronesian (Pohnpei and Palau) and east Pacific (Clipperton Atoll and Panama) locations were pooled for these statistics and subsequent population genetics analyses.

Table 4.2 Population pairwise F_{ST} values (bottom diagonal) between 13 sampling locations and associated significance (top diagonal)

Location	OM	SEY	SAF	X-IO	X-PO	CK	OK-TW	MIC	SOL	GBR	HAW	MQS	EP
OM	*	0.3769	0.4978	0.6521	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
SEY	0.0014	*	0.7692	0.4622	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
SAF	–0.0016	–0.0073	*	0.6049	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
X-IO	–0.0153	–0.0020	–0.0098	*	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**
X-PO	**0.7297**	**0.7377**	**0.7508**	**0.7570**	*	0.1261	0.5379	0.0041	0.0337	0.6519	**0.0000**	0.0096	0.0062
CK	**0.5729**	**0.6313**	**0.6013**	**0.5290**	0.0634	*	0.0693	0.0326	0.0246	0.3242	**0.0000**	0.2043	0.0055
OK-TW	**0.7389**	**0.7419**	**0.7574**	**0.7650**	–0.0067	0.0875	*	0.0063	0.0297	0.3020	**0.0000**	0.0061	0.0072
MIC	**0.7142**	**0.7311**	**0.7363**	**0.7353**	0.1022	0.1078	0.0945	*	0.2010	0.0108	**0.0000**	**0.0001**	**0.0011**
SOL	**0.6997**	**0.7262**	**0.7270**	**0.7196**	0.0821	0.1101	0.0872	0.0245	*	0.0373	**0.0000**	**0.0019**	**0.0014**
GBR	**0.7074**	**0.7279**	**0.7342**	**0.7327**	–0.0162	0.0189	0.0036	0.1084	0.1037	*	**0.0000**	0.0730	**0.0038**
HAW	**0.8006**	**0.7846**	**0.8236**	**0.8528**	**0.2289**	**0.4149**	**0.1880**	**0.3053**	**0.2527**	**0.3261**	*	**0.0000**	**0.0000**
MQS	**0.7149**	**0.7370**	**0.7476**	**0.7533**	0.1249	0.0373	0.1465	**0.2408**	**0.2593**	0.0660	**0.5396**	*	**0.0004**
EP	**0.7392**	**0.7414**	**0.7710**	**0.8069**	0.1278	0.2265	0.1151	**0.1971**	**0.2368**	**0.2019**	**0.3493**	**0.3949**	*

CK = Cocos Keeling Islands; EP = Clipperton Atoll and Panama (east Pacific); GBR = Great Barrier Reef; HAW = Hawaii; MIC = Pohnpei and Palau (Micronesia); MQS = Marquesas; OM = Oman; SAF = South Africa; TWO = Taiwan and Okinawa; SEY= Seychelles; SOL = Solomon Islands; X-IO = Christmas Island (Indian Ocean clade); X-PO = Christmas Island (Pacific Ocean clade)
Notes: Bonferroni correction $P < 0.003846$
Values in bold indicate significant values ($P < 0.05$)
Values are based on 10 100 permutations

populations except the Cocos Keeling Islands (Pacific Ocean haplotypes), the Marquesas and the east Pacific populations, suggesting population growth in most locations (Table 4.2). Pairwise F_{ST} comparisons confirmed low genetic connectivity between western Indian Ocean and Pacific Ocean populations (F_{ST} = 0.5290–0.8528), and also revealed partitioning on a smaller spatial scale (Table 4.2). Gene flow among western Indian Ocean populations was high, as indicated by the low, non-significant western Indian Ocean pairwise F_{ST} values. An exception was the 27 Christmas Island samples that were grouped with the Pacific Ocean clade, and which were treated as distinct for the purpose of population genetic analyses on the basis of the phylogenetic findings (see above). In the Pacific Ocean, gene flow was generally high between populations, with the exception of the central and east Pacific ($F_{ST\ HAW–FP}$ = 0.5396, $P < 0.05$; $F_{ST\ HAW–CK}$ = 0.41, $P < 0.05$). Hawaii showed significantly high F_{ST} values between all other Pacific Ocean locations, with only 2 comparisons under 0.25. The Marquesas and the east Pacific generated significant F_{ST} values between 50% of Pacific Ocean locations if Bonferroni corrections are considered. If we accept that Bonferroni corrections are conservative (Moran, 2003), then we would conclude that all but one of the pairwise F_{ST} values were significant. The exception here was for the Marquesas when it was paired with the Cocos Keeling Islands. Moreover, the east Pacific sample was significantly different to all other samples when Bonferroni corrections were not applied.

There was a positive correlation between genetic distance and geographic distance at the largest spatial scale, although not statistically significant (Z = 78973.1620, R^2 = 0.120, P = 0.0646). Within ocean basins, however, the relationship between F_{ST} and geographic distance was significant (Pacific Ocean: Z = 51568.0352, R^2 = 0.295, P = 0.0417; western Indian Ocean: Z = 290.5452, R^2 = 0.470, P = 0.0489).

Table 4.3a Analysis of molecular variance (AMOVA) based on 12 populations partitioned into the Pacific and Indian ocean basins

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics* (*P*)
Among groups (F_{ct})	1	780.495	4.92257 Va	50.38%	0.50381 (0.01)
Among populations within groups (F_{sc})	10	426.192	1.75659 Vb	17.98%	0.36232 (0.01)
Within populations (F_{ST})	280	865.652	3.09162 Vc	31.64%	0.68359 (0.01)
Total	**291**	**2072.339**	**9.77077**		

Notes: Indian Ocean basin includes Oman, Seychelles, South Africa, Cocos Keeling Islands and Christmas Island, Pacific Ocean basin includes Okinawa–Taiwan, Micronesia, Solomon Islands, Great Barrier Reef, Hawaii, Marquesas, and Clipperton Atoll and Panama (east Pacific)
The groups correspond to locations in the Pacific Ocean and the Indian Ocean
Significance tests are based on 20 350 permutations

AMOVAs based on geography and phylogenetic structure showed structure at the largest spatial scale, with the greatest variation explained by ocean basin (50.38% and 74.79%, respectively, Tables 4.3a and 4.3c). Similar results were obtained when populations were grouped by locations based on pairwise F_{ST} comparisons, but with a much lower percentage of the variation explained by among populations within groups (0.89%, Table 4.3b).

Migrate 2.3 revealed gene flow that was largely asymmetrical and dictated by contemporary oceanographic currents (Figure 4.4). In the western Indian Ocean, there was high gene flow coming from the east Pacific, Hawaii and the Marquesas to other Pacific Ocean populations (Figure 4.4b). Out of the 36 Pacific Ocean Migrate comparisons, 61% were predominantly west, of which

Table 4.3b Regional analysis of molecular variance (AMOVA) based on 12 populations partitioned into 5 regions

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics* (*P*)
Among groups (F_{ct})	4	1283.866	6.42637 Va	69.77%	0.69774 (0.01)
Among populations within groups (F_{sc})	7	31.951	0.08208 Vb	0.89%	0.02948 (0.01)
Within populations (F_{ST})	280	756.522	2.70186 Vc	29.34%	0.70665 (0.01)
Total	**291**	**2072.339**	**9.21032**		

Notes: The five regions are: 1. west Indian Ocean (Oman, Seychelles, South Africa, and Christmas Island and Cocos Keeling Islands [Indian Ocean clade]); 2. east Indo – west Pacific Ocean (Christmas Island and Cocos Keeling Islands [Pacific Ocean clade], Okinawa–Taiwan, Micronesia, Solomon Islands and Great Barrier Reef); 3. Hawaii; 4. Marquesas; 5. east Pacific
Significance tests are based on 20 350 permutations

Table 4.3c Analysis of molecular variance (AMOVA) based on 12 populations divided into 2 groups based on phylogenetic analyses

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics* (*P*)
Among groups (Fct)	1	1231.659	8.77553 Va	74.79%	0.74787 (0.01)
Among populations within groups (Fsc)	10	84.158	0.25660 Vb	2.19%	0.08673 (0.01)
Within populations (F_{ST})	280	756.522	2.70186 Vc	23.03%	0.76974 (0.01)
Total	**291**	**2072.339**	**11.73400**		

Notes: The two groups are 1. Oman, Seychelles, South Africa, and Christmas Island and Cocos Keeling Islands [Indian Ocean clade]); and 2. Christmas Island and Cocos Keeling Islands [Pacific Ocean clade], Okinawa–Taiwan, Micronesia, Solomon Islands, Great Barrier Reef, Hawaii, Marquesas, and east Pacific
Significance tests are based on 20 350 permutations

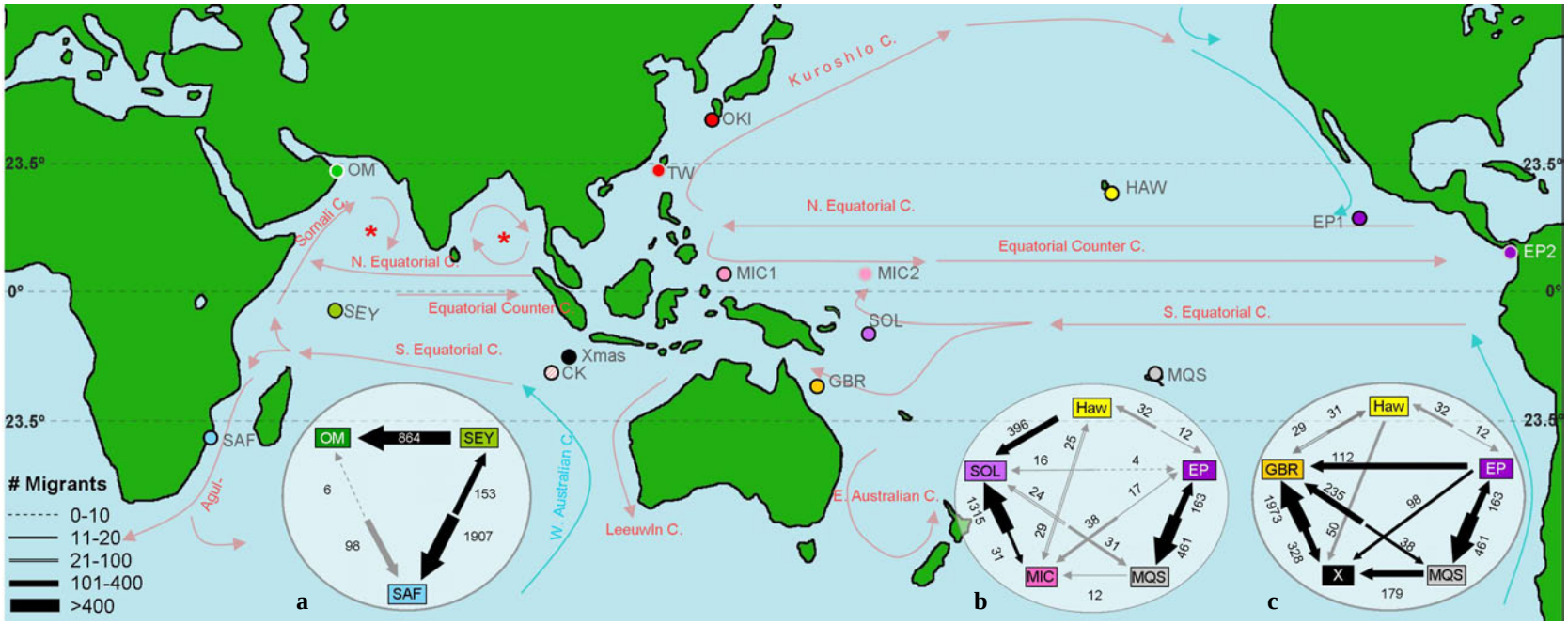


Figure 4.4(a–c) Schematic diagram of migration between *S. rubroviolaceus* populations

Between: (a) 3 populations of *Scarus rubroviolaceus* from the Indian Ocean: OM = Oman; Sey = Seychelles, SAF = South Africa;
and (b–c) populations from the Pacific Ocean; HAW = Hawaii; EP = East Pacific (1) Clipperton Atoll and (2) Panama; MQS: Marquesas; MIC: Micronesia (Palau and Pohnpei); SOL: Solomon Islands; X: Christmas Island; GBR = Great Barrier Reef.

C = current; CK = Cocos Keeling Islands; N = north; S = south; W = west; E = east.
Notes: Arrows correspond to direction and magnitude of gene flow, and values next to arrows indicate average number of migrants.
Black and grey arrows with numbers correspond to direction of gene flow and average number of migrants.
Black arrows represent non-overlapping 95% confidence intervals (CIs) and grey arrows represent overlapping 95% CIs.
Major ocean currents have also been depicted: red arrows indicate warm water currents and blue arrows indicate cold water currents.
*Currents in Arabian Sea and Bay of Bengal, including the North Equatorial Current, reverse during winter months; summer currents are shown here.

72% were exclusively in the westerly direction. There was strong, unidirectional gene flow from the east Pacific to the Great Barrier Reef and Christmas Island, 4 times as many migrants to the Marquesas and moderate gene flow to Hawaii. From Hawaii, there was strong, unidirectional gene flow to the Solomon Islands and Christmas Island, but no migration to the Marquesas. There was also strong, unidirectional migration from the Marquesas to Christmas Island, as well as more than 5 times as many migrants from Christmas Island to the Great Barrier Reef. In the Indian Ocean, migration was strongly linked with present day oceanographic currents, with significant unidirectional flow to the north from the Seychelles to Oman along the Somali current, and over 10 times as many migrants to the south from the Seychelles to South Africa along the Agulhas current.

4.3.2 Coalescence analyses

A coalescence analysis revealed that the time to the most recent common ancestor for the western Indian Ocean and Pacific Ocean lineages was very similar; the western Indian Ocean clade haplotypes (T_{MRCA} = 0.988 mya) appear older than those from the Pacific Ocean clade (T_{MRCA} = 0.778 mya) (Table 4.4). These dates are supported by fossil-calibrated dates, which place the initial diversification of *S. rubroviolaceus* between 1.6 mya and 2.23 mya (Choat *et al.*, 2012), as well as estimates based on multiple nuclear and mitochondrial markers (Choat *et al.*, 2012; Smith *et al.*, 2008). The mismatch distribution curve for all *S. rubroviolaceus* populations was bimodal, suggesting either stable population sizes (Rogers, Harpending, 1992) or historical separation of populations (Ray *et al.*, 2003) (Figure 4.2 b). The mismatch distribution curve for the Pacific Ocean was highly leptokurtic with a major peak at 5 and a smaller peak at 23 units of mutational time (Figure 4.2b). The western Indian Ocean, however, had a smooth unimodal mismatch curve that conforms to the sudden expansion model. Both populations had low, non-significant raggedness scores, suggesting that the null hypothesis of population expansion cannot be rejected.

Population-specific coalescence dates were also calculated (Table 4.4,). In the western Indian Ocean, none of the populations converged under the model of sudden expansion, and coalescence and mismatch parameters were based on the model of spatial expansion. All mismatch distribution curves showed a peak at 7 units of mutational time (Figure 4.6). Coalescence times for western Indian Ocean haplotypes were generally more ancient than Pacific Ocean haplotypes, and ranged from 0.84 to 1.3 mya. The mismatch distribution for Oman was the only bimodal curve, which indicated a second peak around 24 units of

mutational time. In the Pacific Ocean, however, coalescence times ranged between 0.47 and 0.91 mya.

4.4 Discussion

This study identified four major patterns and the most probable processes that are relevant to the observed complex phylogeography of widespread coral reef fishes in the Indo–Pacific. Firstly, there was evidence of a genetic break between western Indian Ocean and Pacific Ocean populations, which was associated with permeable biogeographic barriers and several oceanographic features. Secondly, we identified an area of overlap between eastern Indian Ocean and Pacific Ocean populations where the different lineages co-exist. This area—which lies in the eastern Indian Ocean, and includes Christmas and the Cocos Keeling islands—possesses complex oceanographic features, such as oceanic ridges and alternating currents. Thirdly, peripheral populations in the Pacific Ocean were structured into four identifiable populations that were either isolated from other reefs or comprised atypical coral reef habitats. Lastly, the peripheral populations in the western Indian Ocean do not appear to be genetically structured for this species, which contradicts the scenario for peripheral populations in the central and eastern Pacific Ocean.

4.4.1 Genetic breaks and areas of overlap between western Indian Ocean and Pacific Ocean populations

The first major finding of this study was evidence of genetic partitioning between extant *S. rubroviolaceus* populations from the western Indian Ocean and Pacific Ocean. The break between the two lineages is most likely situated to the west of the Cocos Keeling Islands. Evidence for this is provided by individual haplotypes from both the western Indian Ocean and Pacific Ocean clades overlapping at the Cocos Keeling Islands, as well as at Christmas Island, indicating bidirectional gene flow into this region. As such, the genetic (haplotype and nucleotide) diversities at these two locations were higher than in any other population, which is characteristic of secondary contact between allopatrically diverged lineages. Further sampling in the central and eastern Indian Ocean, particularly in the Bay of Bengal around the Maldives, Chagos Archipelago, and the Andaman and Java seas, will enable a more precise indication of where the break between the two *S. rubroviolaceus* lineages lies. Evidence of a genetic break that partitions western Indian Ocean and Pacific Ocean populations is strongly supported by both phylogenetic and population genetic analyses, and has also been reported in other reef fish including parrotfish (Bay et al. 2004; Chapter 3), soldierfish (Craig *et al.*, 2007) and damselfish (Bernardi *et al.*, 2001; 2002). The exact

Table 4.4. Coalescence analysis parameters for 2 main *S. rubroviolaceus* clades (Indian Ocean clade and Pacific Ocean Clade) and individual populations

Clade/ population	Mean no. differences	Tau τ (95% CI)	Θ_0	Θ_1	SSD	R (P)	T_{MRCA} (my) (0.5–0.95 CI)
Indian Ocean	7.198	6.835 (5.14–8.25)	0.323	689.3	0.00042651	0.00831 (ns)	0.988 (0.743–1.193)
Oman*	8.567	7.355 (4.40–8.90)	0.001		0.00762256	0.02095556 (ns)	1.063 (0.636–1.286)
Seychelles*	7.320	6.375 (4.83–8.54)	0.923		0.00216932	0.01301733 (*)	0.922 (0.698–1.235)
South Africa*	7.310	5.799 (3.74–9.07)	1.443		0.00268792	0.01385488 (ns)	0.838 (0.541–1.311)
Xmas/CK *	8.214	9.012	0.000				1.303
Pacific Ocean	4.92	5.378 (2.82–7.35)	0.002	31.015	0.00211774	0.0109 (ns)	0.778 (0.407–1.062)
Xmas/CK	4.704	5.477 (2.75–8.19)	0.001	25.868	0.00561252	0.0241 (ns)	0.792 (0.397–1.183)
Okinawa–Taiwan	4.568	5.556 (2.76–8.41)	0.001	21.184	0.00377861	0.0133 (ns)	0.803 (0.399–1.215)
Micronesia	5.432	6.248 (3.21–8.93)	0.000	35.312	0.01423451	0.0370 (ns)	0.903 (0.464–1.291)
Solomon Islands*	5.311	5.559	0.006				0.804
GBR	5.022	6.303 (3.07–10.00)	0.000	21.584	0.01106141	0.0233 (ns)	0.911 (0.44–1.44)
Marquesas	3.855	3.453 (2.16–7.56)	0.000	37.383	0.01013404	0.0316 (ns)	0.675 (0.312–1.09)
Hawaii	2.260	2.338 (1.32–6.71)	0.000	6.681	0.01313468	0.0467 (ns)	0.47 (0.19–0.97)
East Pacific	2.250	6.270 (1.63–15.27)	0.00	1.038	0.11763768	0.2606 (ns)	0.91 (0.235–2.2)

CI = confidence interval; CK = Cocos Keeling Islands; GBR = Great Barrier Reef; my = million years; ns = not significant ($P > 0.05$); SSD = sum of squared deviation; Xmas = Christmas Island

*Least square procedure to fit sudden expansion model mismatch distribution and observed distribution did not converge after 1800 steps and thus spatial expansion model parameters were used to calculate coalescence.

Notes: Tau ($\tau \pm$ 95% CI) is the unit of mutational time between two populations; theta$_0$ (Θ_0) is the mutation parameter before expansion; and theta$_1$ (Θ_1) is the mutation parameter after expansion. Harpending's raggedness index and significance based on the significance of simulated and observed raggedness are also shown

Fossil-calibrated dates for *S. rubroviolaceus*: 1.6–2.3 my

location of the break between western Indian Ocean and Pacific Ocean lineages, however, differs between the two parrotfish species (chapters 3 and 4).

The area around Christmas and the Cocos Keeling islands has been documented as an area where a number of Indian Ocean and Pacific Ocean species converge (Hobbs *et al.*, 2009). For some marine species, this is an area of overlap (Hobbs, Salmond, 2008), and for others, it marks the end of one ocean-basin species and the beginning of another (McMillan, Palumbi, 1997; Palumbi *et al.*, 1997). The presence of oceanic ridges to the east and west of the Cocos Keeling Islands and the Sunda Shelf probably deliver cold water upwelling, and are likely important oceanographic factors contributing to the variable history of this area.

4.4.2 Spatial structure at peripheral locations

Although phylogenetic analyses did not detect significant structure within each ocean basin, evidence from more sensitive and phylogenetically informed population genetics, and demographic analyses did. Peripheral populations in the Pacific Ocean contribute to a total of at least four genetically differentiated populations: (1) the western Pacific Ocean, including the zones of overlap in the Eastern Indian Ocean; (2) the east Pacific, including Clipperton Atoll and Panama; (3) the Hawaiian Islands; and (4) the Marquesas. Micronesia and the Solomon Islands could also be considered distinct populations if Bonferroni corrections were not strictly adhered to, since we know that these corrections are conservative and may mask biologically significant information (Garcia, 2003; Moran, 2003). Among these distinctive peripheral populations, those in the east Pacific were the most isolated based on pairwise F_{ST} comparisons and unidirectional gene flow detected by Migrate analyses.

The higher degree of population structure in the Pacific Ocean when compared to the Indian Ocean is also reflected in the minimum-spanning tree, which shows a large proportion of shared haplotypes at Hawaii and in the east Pacific. This, in conjunction with comparatively lower haplotype and nucleotide diversities, suggests that bottlenecks or selective-sweeps were associated with the colonisation of these peripheral locations. The minimum-spanning tree also reveals that the colonisation of the Hawaiian Islands was a more recent event compared to the east Pacific colonisation time, as shown by the more peripheral location of the Hawaii-dominated common haplotypes (Castelloe, Templeton, 1994; Horne *et al.*, 2008). In fact, migration analyses indicated that gene flow is nearly 3 times greater in the westerly direction from the east Pacific to Hawaii. Furthermore, a coalescence analysis confirmed that the east Pacific population is twice as old as the

Hawaiian population. The peripheral position of a large shared haplotype between distant locations (i.e. Hawaii, Taiwan–Okinawa and Christmas Island) suggests recent gene flow between these populations, and is consistent with a recent bottleneck or selective sweep event (Posada, Crandall, 2001), which resulted in the persistence of that haplotype but not others.

The patterns of genetic partitioning observed among individuals from the Pacific Ocean locations for *S. rubroviolaceus* were very similar to those observed in *C. sordidus. C. sordidus* also exhibited high levels of gene flow between Pacific Ocean locations, with the exception of individuals from peripheral locations, including Hawaii, French Polynesia and Western Australia. In contrast, there was very little genetic partitioning between *S. rubroviolaceus* individuals from the western Indian Ocean. This differs from *C. sordidus*, which exhibited low gene flow between populations from peripheral locations in the Indian Ocean, including the Arabian Gulf, Oman, Seychelles and the Cocos Keeling Islands. This is interesting, considering the similarity in geographic range and life-history characteristics between these two species. The sampling locations in the west Indian Ocean, however, were not identical for both species, and it is possible that future sampling from the Arabian Gulf for *S. rubroviolaceus* will uncover increased levels of genetic partitioning in the Indian Ocean.

Although overall Pacific Ocean and Indian Ocean statistics suggest expansion in extant populations of *S. rubroviolaceus*, evidence based on Fu's test for neutrality denotes either selection or a lack of expansion in the central and eastern Pacific Ocean. Selection is likely to be an important factor in the population structure at these peripheral locations, as realised dispersal potential of larval fishes may involve a strong selective bottleneck on which larvae arrive and survive to maturity (Gagliano, McCormick, 2007). In addition, the small second peak in the Pacific Ocean mismatch distribution may represent haplotypes from these peripheral populations that entered the population at a different time or were not part of the expansion (Joshi *et al.*, 2004).

4.4.3 Historical gene flow, population structure and contemporary ocean currents

The patterns of population structure and gene flow seen in extant *S. rubroviolaceus* populations are closely tied to present-day oceanographic features, including zones of upwelling associated with mid-ocean ridges and island arcs, and surface currents. Phylogenetic and coalescence analyses indicate that the western Indian Ocean clade is older than that from the Pacific Ocean, which is further supported by the greater genetic

diversity of the western Indian Ocean population. The population from Oman appears to have coalesced the earliest and all subsequent coalescence events coincide with periods of low sea-level stand, with the exception of South Africa. Gene flow was greatest from the Seychelles to other western Indian Ocean locations. Interestingly, this gene flow was unidirectional from the Seychelles to Oman, despite the presence of the circulating Somali Current, which flows north along the coast before circling back around the Arabian Sea towards the Seychelles. The only anomalous gene flow result for this species was the 6-fold eastward migration from Christmas Island to the Great Barrier Reef, Australia, a pattern that is most likely due to the reversal of ocean circulation through the Torres Strait during El Niño Southern Oscillation periods. This pattern is an indication of historical gene flow between Christmas Island and the western Pacific Ocean, and provides a feasible mechanism for the rise of the Pacific Ocean lineage from the east Indian Ocean. In addition, the fact that the same Christmas Island individual is basal to the Pacific Ocean clade in both the phylogenetic tree and the minimum-spanning tree is consistent with an eastward migration of extant populations from the Indian Ocean to the Pacific Ocean.

Coalescence analyses and a central positioning of the majority of east Pacific haplotypes in the minimum-spanning tree indicate that these haplotypes are older (Castelloe, Templeton, 1994), consistent with an initial eastward migration of extant populations. Once in the east Pacific, dispersal to remote populations, including the Marquesas and Hawaii, were then possible. Hawaii is situated to the north of the westward-flowing North Equatorial Current, and thus present dispersal from the west is unlikely, as evidenced by the relatively higher gene flow from Hawaii to more western populations. The remoteness of the Hawaiian Islands, the Marquesas and the east Pacific from other Pacific Ocean reef locations lends itself to speciation arising from geographic isolation. The East Pacific Barrier, a 5000 km expanse of deep open ocean between the central and eastern Pacific Ocean (Lessios *et al.*, 1998), is a well-established marine biogeographic barrier. However, the barrier is a permeable one at different times in the evolutionary history of different species, including *S. rubroviolaceus* (Lessios, Robertson, 2006), and it is no surprise that the reefs of Hawaii harbour more endemic reef fish species than most other coral reefs (Allen, 2007). Although the East Pacific Barrier appears to play a significant role in structuring populations of *S. rubroviolaceus* and many other reef fishes, it is worth noting that many widespread Indo–Pacific reef fish do not make it across this barrier into the east Pacific, including two other Indo–Pacific parrotfish *C. sordidus* and *S. psittacus* (Bay *et al.*, 2004; Winters *et al.*, 2010).

4.5 Conclusion

The earlier coalescence of the *S. rubroviolaceus* western Indian Ocean clade suggests that extant populations elsewhere were sourced from the western Indian Ocean, followed by an eastward expansion into the Pacific Ocean and the east Pacific. This is the reverse of what was observed for the closely related *C. sordidus* (Chapter 3), and contradicts conclusions about the westward direction of dispersal between these oceans, based on a modelling approach (Connolly *et al.*, 2003). Therefore, this suggests that good empirical data are likely required to validate models, as has been demonstrated for connectivity studies based on oceanographic modelling alone (Werner *et al.*, 2007). Significantly higher eastward gene flow from Christmas Island to the Great Barrier Reef supports this eastward migration. *S. rubroviolaceus* populations on the eastern periphery of its distribution appear to be the most structured and the permeable East Pacific Barrier appears to be a major contributing factor influencing population structure at these peripheral populations. Similarly, the Indo – West Pacific Barrier also proved to be a highly effective—albeit permeable—barrier, which separates western Indian Ocean and Pacific Ocean populations. These observations are consistent with the centre of accumulation (Jokiel, Martinelli, 1992; Ladd, 1960) /overlap (Barber *et al.*, 2000; Bellwood, Wainwright, 2002; Santini, Winterbottom, 2002; Woodland, 1983) models of Indo-Australian Archipelago diversity. These models postulate that speciation takes place outside the Indo-Australian Archipelago, and dispersal into the centre ensues, resulting in an accumulation of species in the Indo-Australian Archipelago (Jokiel, Martinelli, 1992; Ladd, 1960), or that vicariance causes populations that were once widespread to be separated, after which range expansions occur, leading to the populations to be reunited in the Indo-Australian Archipelago (Barber *et al.*, 2000; Bellwood, Wainwright, 2002; Santini, Winterbottom, 2002; Woodland, 1983).

In conclusion, it is evident that a combination of historical and present-day factors, as well as key life-history characteristics, are largely responsible for the patterns of distribution and connectivity observed in extant populations of *S. rubroviolaceus*. The patterns of divergence observed in this study have also been confirmed independently using nuclear microsatellite markers (Fitzpatrick *et al.*, 2011), indicating that the patterns of population structure observed here are not due to marker-associated stochasticity. In all, this study provides an interesting history of diversification that has similar and unique patterns to those observed in other reef fish with similar characteristics and habitat preferences (Winters et al. 2010; Bay et al. 2004; Chapter 3). Results from this study emphasise the species-specific nature of marine biogeographic barriers and the potential

of peripheral habitats at the edges of species' distributions to generate evolutionary novelty by their relative isolation and genetic differentiation from better connected source populations.

Chapter 5 Population structure and cryptic speciation in the widespread Indo–Pacific scarine labrid, *Scarus ghobban* (Perciformes: Scarinae)

Abstract

This study uses phylogenetic, population genetic and phylogeographic tools to investigate the population structure and genetic connectivity of the widespread parrotfish *Scarus ghobban*. A comparison of 350 bp from the mitochondrial control region between 239 individuals collected from 12 locations across its Indo–Pacific distribution revealed two genetically distinct clades that were geographically partitioned into western Indian Ocean and Pacific Ocean (including eastern Indian Ocean) populations. We also resolved a third subclade, which contained 56% of all individuals from the Cocos Keeling Islands, with the remaining 44% grouping in the Pacific Ocean clade. The same three clades were identified by population genetics analyses. Evidence from the minimum-spanning network showed that there were more differences (in base pairs) between the Cocos Keeling Islands clade and its sister Indian Ocean clade, than there were between the Indian and Pacific Ocean clades. Population pairwise F_{ST} comparisons confirmed high genetic differentiation between the three clades ($F_{ST\ IO–PO}$ = 0.748–0.803; $F_{ST\ IO–CK}$ = 0.754–0.763; $F_{ST\ PO–CK}$ = 0.766–0.851), and all were highly significant ($P < 0.001$). The division of individuals from the Cocos Keeling Islands suggests a genetic break between the east and western Indian Ocean, and Pacific Ocean populations of *S. ghobban* at the Cocos Keeling Islands, with the possibility of cryptic speciation at this location. I also revealed that the population structure within the Pacific Ocean clade is due to strong and significant population structure at peripheral locations based on population pairwise F_{ST} comparisons. Pairwise F_{ST} scores between the East Pacific and other Pacific locations ranged from 0.047 to 0.561 (p = 0.0001 - 0.765) , and pairwise F_{ST} values between Western Australia and other Pacific locations were also mostly significant ($P < 0.05$), but not as high (F_{ST} = –0.014 to 0.146.

5.1 Introduction

Recent phylogeographic studies of widely distributed marine organisms have demonstrated several unexpected results, including contrasting biogeographic patterns between co-distributed, closely related species, cryptic speciation in species considered to be widespread, and lowered genetic diversity and greater population differentiation at

peripheral habitats. A study on two co-distributed Indo–Pacific gastropods *Nerita albicilla* and *N. plicata* showed genetic partitioning between Indian Ocean and Pacific Ocean populations in one species (*N. plicata*) but not the other (*N. albicilla*) (Crandall et al., 2007), despite their ecological similarity. Similar discordance between confamilial species has also been noted in a number of widespread reef fish, including acanthurids (Horne *et al.*, 2008; Planes, Fauvelot, 2002) and parrotfishes (Bariche, Bernardi, 2009; Bay *et al.*, 2004; Winters *et al.*, 2010).

Many widespread Indo–Pacific species have exhibited genetic partitioning between western Indian Ocean and Pacific Ocean individuals (Bariche, Bernardi, 2009; Bay *et al.*, 2004; Crandall *et al.*, 2007), or between central and east Pacific Ocean populations (Duda, Lessios, 2009). Two marine biogeographic barriers are often implicated in these patterns: the Indo – West Pacific Barrier (Froukh, Kochzius, 2008; Marie *et al.*, 2007) and the East Pacific Barrier (Ekman, 1953). The Indo – West Pacific Barrier is a land barrier created during periods of low sea-level stand, which separates the Indian Ocean and Pacific Ocean basins. South of this barrier, the vicinity of Christmas and the Cocos Keeling islands in the eastern Indian Ocean is usually associated with genetic breaks between western Indian Ocean and Pacific Ocean populations of widespread marine organisms, including the starfish *Acanthaster plancii* (Benzie, 1999b), butterflyfish species (McMillan, Palumbi, 1995) and the parrotfish *Chlorurus sordidus* (Bay et al. 2004; Chapter 3). This region is also known to have cold water upwelling, which has been linked to the reduction of tropical habitat (Brewer *et al.*, 2009; Voris, 2000). Additionally, it is a known area of overlap between Indian Ocean and Pacific Ocean fauna; for some, it is the site where those two faunas converge and co-exist (Hobbs, Salmond, 2008). For others, the two faunas meet and hybridise in this area (Hobbs *et al.*, 2009; Marie *et al.*, 2007). As such, populations from this region may demonstrate higher genetic diversity than those from peripheral locations (e.g. *S. rubroviolaceus*, Chapter 4). In contrast, these latter populations often demonstrate lowered genetic diversity and distinctive population structure; for example, several parrotfish species (Chapter 4; Winters et al. 2010), the atherinid fish *Craterocephalus capreoli* (Johnson et al., 1994), several species of *Echinometra* sea urchins (Palumbi, 1997) and several species of *Calcinus* reef hermit crabs (Malay, Paulay, 2009). This phenomenon is reviewed extensively in a recent study by Hardie and Hutchings (2010).

In the Pacific Ocean, the East Pacific Barrier is an equally effective marine biogeographic barrier (Ekman, 1953), and, similar to the Indo – West Pacific Barrier, separates many marine taxa on either side into discrete lineages (Lessios *et al.*, 1998; Lessios, Robertson, 2006). However, unlike the Indo – West Pacific Barrier, which is a land barrier, the East

Pacific Barrier is a 5000 km deep water barrier, the efficacy of which lies in the fact that there are no islands in between to serve as refuges for larvae. Lessios and Robertson (2006) found that gene flow is greater in the westerly direction across the East Pacific Barrier and attributed this observation to the fact that the availability of reef habitat is greater west of the East Pacific Barrier. Therefore, despite bidirectional currents across the East Pacific Barrier, larvae have greater chances of survival to the west. One of the major outcomes of such phylogeographic studies demonstrating population structure in widespread marine species is the prevalence of cryptic speciation in the sea (Knowlton, 1993; Knowlton, 2000).

Cryptic speciation is defined as two or more groups of organisms that are morphologically very similar (and sometimes indistinguishable), but found to belong to different evolutionary lineages (Sáez, Lozano, 2005). This has been demonstrated in widespread marine taxa, including several species of fishes (Colborn *et al.*, 2001; Kon *et al.*, 2007; Lessios *et al.*, 1998), sea stars (Benzie, 1999b; Williams, Benzie, 1998) and hermit crabs (Malay, Paulay, 2009). Some of these studies have prompted scientists to revisit the taxonomic status of the organisms in question, only to find differences in their morphological features (see Bickford et al., 2007). Cryptic speciation appears to be more common than previously acknowledged, and by the same token, the tendency for marine organisms with dispersive larvae to be widespread may be less common than previously believed. The increasing recognition of cryptic speciation in the sea (Knowlton, 1993; Knowlton, 2000) is testament to the fact that present-day patterns of population structure often reflect historical, rather than present-day, oceanographic circulation patterns (Benzie, 1999a). To date, two species of moray eels, *Gymnothorax flavimarginatus* and *G. undulatus*, are the only reef fish that have demonstrated genetic homogeneity across the entire Indo–Pacific based on phylogeographic evidence (Reece et al., 2010).

In this chapter, we examine the population structure and evolutionary history of the scarine labrid *S. ghobban*, whose range extends across the entire Indo–Pacific (Parenti, Randall, 2000), and includes the Mediterranean Sea (Bariche, Saad, 2005; Goren, Aronov, 2002) as well as the east Pacific Ocean. As with the previous two species, the widespread nature of *S. ghobban* makes it ideal for testing models of Indo-Australian Archipelago diversity, which are outlined in chapter 1. Out of the four species discussed in this thesis, *S. ghobban* is the least reef-associated parrotfish and has been known to use nonreefal habitats, including the Mediterranean. Video surveillance has also shown them at depths up to 250 m (Rees et al., 1994). Crandall et al. (2007) suggests that minor differences in ecological traits, such as adult habitat preference, could lead to vastly different patterns of population structure in closely related species. In a comparison of Hawaiian chaetodontids, Craig et al. (2010)

concluded that more specialised species are at a greater risk of severe bottlenecks or extinction in response to environmental fluctuations over evolutionary timescales. Given the similarity in geographic range to *S. rubroviolaceus*, relative lack of habitat specificity in *S. ghobban*, and in light of recent findings by Bariche and Bernardi (2009) demonstrating population structure between Indian and Pacific Ocean populations of *S. ghobban*, we initially expected *S. ghobban* to exhibit the lowest level of population structure across its range, particularly within an ocean basin, when compared to the other three species. However, more recently, Visram et al. (2010) found evidence of high genetic diversity and 3 distinct clades of *S. ghobban* within the western Indian Ocean based on amplified fragment length polymorphisms (AFLPs). The authors attributed its lack of habitat specificity to the observed high genetic diversity in the western Indian Ocean. Despite this, however, a comparative phylogeographic analysis requires sampling from across a species' entire range, as differing sampling schemes from the same species can lead to very different phylogeographic patterns and potentially erroneous conclusions (Lourie et al., 2005; Teske et al., 2005). As this study is a component of a larger comparative phylogeographic study, sampling from the entire geographic range of *S. ghobban* has been used in order to address the following questions:

1. Given the evidence of population structure between Indian Ocean and Pacific Ocean populations, are any known marine biogeographic barriers implicated in the separation?
2. Is there further evidence of partitioning within an ocean basin?
3. What is the role of central and peripheral habitats in the observed patterns of population structure in this species?
4. Is there evidence of temporal (or nongeographic) partitioning between and within ocean basins?
5. What is the magnitude and direction of genetic connectivity within and between ocean basins?

5.2 Materials and methods

5.2.1 Sampling, DNA amplification and sequencing

To determine the phylogeographic patterns of the widespread Indo–Pacific scarine labrid *S. ghobban*, we compared sequences from the mitochondrial control region for 244 individuals collected from 12 locations across its distribution range, including samples from the east Pacific (tables 2.1 and 5.1, Figure 5.1). The majority of samples were collected by spearfishing, and

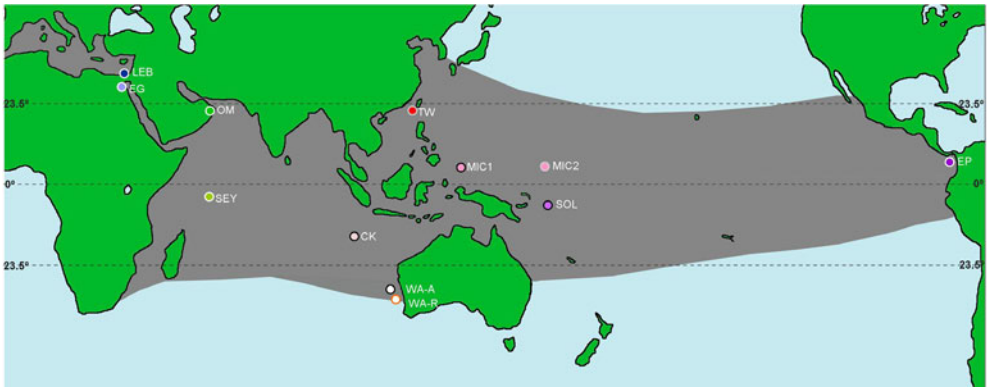


Figure 5.1 Map depicting distribution (shaded in grey) and locations sampled for *S. ghobban*

Location codes and numbers sampled from each location are as follows: LEB = Lebanon (1); EG = Egypt (3); OM = Oman (34); SEY = Seychelles (45); CK = Cocos Keeling Islands (9); WA-A = Abrolhos Island, Western Australia (11); WA-R = Rottnest Island, Western Australia (22); TW = Taiwan (37); MIC1 = Palau, Micronesia (8); MIC2 = Pohnpei, Micronesia (19); SOL = Solomon Islands (31); GBR = Great Barrier Reef (16); EP = Panama, east Pacific Ocean (8).

some samples from Taiwan and all samples from Palau were collected from fish markets. *S. compressus*, the closest sister to *S. ghobban*, was initially used as an outgroup. However, *S. compressus* grouped with other *S. ghobban* individuals in the phylogenetic analysis. Therefore, to root the phylogenetic trees, we used *S. rubroviolaceus* from the western Indian Ocean as an outgroup instead. The *S. compressus* sequence was pooled with *S. ghobban* individuals from the East Pacific. The DNA samples that could not be amplified using universal control region primers were amplified using primers developed specifically for the mitochondrial control region of *C. sordidus*, *S. ghobban* and *S. rubroviolaceus* (Table 2.2). Further detailed laboratory procedures were performed as indicated in Section 2.2.

5.2.2 Phylogenetic analyses

A Bayesian analysis was performed using Markov chain Monte Carlo (MCMC) simulation with four chains of 2 million generations, sampling trees every 100 generations. Trees began to stabilise after 10 000 generations, and a 50% majority-rule consensus tree was inferred using 32 000 of postburn-in trees from two runs. Branch support for the Bayesian tree is based on a 50% majority-rule consensus of the 15 000 shortest trees. Further detailed phylogenetic analyses were performed as indicated in Section 2.3. Our data suggest that *S. rubroviolaceus* is sister to *S. ghobban* in the Indian Ocean, which is why it

has been selected as the outgroup in the phylogenetic trees in this analysis (Choat et al., 2012). *S. compressus*, an eastern Pacific scarine labrid, has also been included in this analysis, but was not specified as an outgroup due to its genetic similarity to *S. ghobban* from the east Pacific (Choat et al., 2012).

5.2.3 Population genetic analyses

Populations with low sample numbers (i.e. Lebanon and Egypt) were not included in population genetic analyses, including population pairwise F_{ST} comparisons, Migrate analyses, analysis of molecular variances (AMOVAs) and coalescence analyses. Where noted, individuals from Cocos Keeling Islands were separated into their respective groups as defined by phylogenetic analyses. Population genetic analyses were conducted despite low sample numbers, but interpreted with caution. Migrate analyses were conducted on all populations; however, results for locations with high or significant pairwise F_{ST} values are depicted in Migrate diagrams. Three separate AMOVAs were conducted, the first consisting of individuals from 10 locations grouped by ocean basin (all Cocos Keeling Islands individuals were placed in the Indian Ocean group), the second comprised the same 10 locations grouped into 5 geographic regions based on pairwise F_{ST} comparisons and the third comprised 10 populations grouped into 3 genetically differentiated clades determined by the phylogenetic analysis (Indian Ocean clade, Cocos Keeling Islands clade and Pacific Ocean clade). For Migrate analyses, only Cocos Keeling Islands individuals from the Pacific Ocean clade are shown (Figure 5.4).

Isolation by distance analyses were based on the clade structure delineated in the phylogenetic analysis and implemented on the Isolation By Distance Web Service v.3.15 (IBDWS) (Jensen et al., 2005). Further detailed population genetic analyses were performed as indicated in Section 2.4.

5.2.4 Coalescence analyses

The generation time for *S. ghobban* (t_2) was calculated using the formula $t_2 = (\alpha+\omega)/2$ (Pianka, 1978) where the age at first reproduction (α) was 3 years and age at last reproduction (ω) was 15 years (Choat, pers. comm.). Further detailed coalescence analyses were performed as indicated in Section 2.5.

5.3 Results

5.3.1 Phylogenetic analyses

We sequenced 356 base pairs from the mitochondrial control region from 244 *S. ghobban* individuals, of which 140 sites were polymorphic and 126 were parsimony informative (transition:transversion [TI:TV] ratio = 30.5717). The base frequencies for *S. ghobban* were AT-rich (A = 34.88%, T = 23.89%, C = 23.73%, G = 17.50%), which is characteristic of reef fish mitochondrial DNA (Bay *et al.*, 2004; McMillan, Palumbi, 1997).

Modeltest results based on Akaike information criteria (AIC) determined the Hasegawa, Kishino and Yano model (HKY+I+G; $\gamma = 0.74$) as the best model, while hierarchical likelihood ratio tests (hLRTs) determined the general time reversible model (GTR+I+G; $\gamma = 0.69$) as the best model. The HKY+I+G model selected by AIC was used for the phylogenetic reconstructions using MrBayes, as it is preferred to avoid over-parameterisation of the substitution model.

S. ghobban was partitioned into 3 major clades that were geographically partitioned (Figure 5.2). The first major clade consisted of west Indian Ocean individuals with no apparent structure between locations. The second clade predominantly consisted of individuals from the Pacific Ocean, and also contained 100% of the Western Australia individuals and 44% of the Cocos Keeling Islands individuals. A third clade exclusively contained the remaining 56% of individuals from the Cocos Keeling Islands.

5.3.2 Population genetic analyses

The minimum-spanning tree identified the same geographically partitioned clades as the phylogenetic analysis (Figure 5.3), with equally high base-pair (bp) substitutions between clades. The exclusive Cocos Keeling-only clade differed from the rest of the Indian Ocean by a total of 53 bp substitutions, and the Pacific Ocean clade differed from the west Indian Ocean clade by 49 bp substitutions. The overall haplotype diversity was high ($h = 0.99$), as were location-specific haplotype diversities ($h = 0.69–1.00$) (Table 5.1a), which was reflected in the star phylogeny architecture of the minimum-spanning tree (Figure 5.3). Similar high values have been reported for pelagic marine fishes (Grant, Bowen, 1998), as well as other widespread tropical reef fish (Bay et al., 2004; Horne et al., 2008). Haplotype and nucleotide diversity was high for Indian Ocean and Pacific Ocean locations, with the exception of Palau and the east Pacific (Table 5.1a). High haplotype and nucleotide diversities are characteristic of secondary contact between isolated

Table 5.1a Summary statistics for *S. ghobban*

Location	*n*	n_h	π (%)	*h*	Fs	P_{FS}	D	P_D
Lebanon*	1	1		1.00				
Egypt*	3	3	2.46	1.00	0.99	0.440	0.000	0.704
Oman	34	28	3.95	0.96	–9.45	0.004	–0.064	0.539
Seychelles	45	37	4.10	0.97	–15.89	0.000	–0.249	0.466
Total Indian Ocean	**83**	**69**	**4.02**	**0.98**	**–24.22**	**0.000**	**–0.345**	**0.432**
Cocos Keeling Islands	9	9	11.60	1.00	0.87	0.398	0.154	0.675
Cocos Keeling Islands (PO)	(4)	(4)	4.53	1.00	0.85	0.420	0.308	0.723
WA (Rottnest Island)	22	22	3.98	1.00	–11.58	0.000	–0.969	0.168
WA (Abrolhos Island)	11	9	4.24	0.87	–2.99	0.046	–0.362	0.385
Taiwan	37	35	3.18	0.97	–24.11	0.000	–1.349	0.076
Palau (MIC1)	8	5	0.66	0.69	–0.83	0.213	–0.489	0.346
Pohnpei (MIC2)	19	18	3.15	0.94	–7.73	0.004	–0.709	0.289
Solomon Islands	31	29	2.38	0.99	–21.76	0.000	–1.513	0.046
Great Barrier Reef	16	15	3.05	0.93	–7.94	0.002	–0.555	0.316
Panama (EP)	8	6	1.71	0.78	1.20	0.718	0.222	0.592
Total Pacific Ocean	**161**	**148**	**3.26**	**0.99**	**–24.16**	**0.000**	**–1.348**	**0.056**
Total	**244**	**206**	**9.57**	**0.99**	**–23.63**	**0.011**	**1.015**	**0.885**

π = nucleotide diversity; D = Tajima's selective neutrality test; EP = east Pacific; Fs = Fu's neutrality test; *h* = haplotype diversity; MIC = Micronesia; *n* = number of individuals sampled; n_h = number of haplotypes; P_D = Tajima's D and significance level P_{FS} = Fu's Fs plus significance level; PO = Pacific Ocean; WA = Western Australia

*Samples from Egypt and Lebanon did not contain sufficient sample numbers for population genetic statistics.

lineages, while low nucleotide and haplotype diversities usually result from recent population bottlenecks (Grant, Bowen, 1998). The Cocos Keeling Islands had the highest nucleotide diversity (π = 11.6%) (Table 5.1a). This is likely due to the separate clade exclusively consisting of Cocos Keeling Islands individuals. For both ocean basins, the null hypothesis of neutrality was rejected for both Tajima's selective neutrality test (D_{IO} = –0.345, $P > 0.05$; D_{PO} = –1.348, $P > 0.05$) and Fu's neutrality test (Fs_{IO} = –24.22, $P < 0.01$; Fs_{PO} = –4.16, $P < 0.01$) (Table 5.1a), indicating that either expansion or selection is acting (Fu, 1997; Tajima, 1989; Tajima, 1996).

Pairwise F_{ST} scores between the individuals from the Pacific Ocean were low and not significant between other Pacific Ocean locations (F_{ST} = –0.014 to 0.092), whereas those from the Cocos Keeling Islands clade were high compared to all other locations (F_{ST} = 0.754–0.851) (Table 5.2). Within the Pacific Ocean, the east Pacific exhibited low levels of connectivity with all other western Pacific Ocean locations, with F_{ST} scores ranging from 0.030 to 0.569 (Table 5.2). Aside

Table 5.1b Summary statistics for *S. ghobban* supported nodes

Clade	*n*	n_h	π (%)	*h*	Fs	P_{FS}	D	P_D
I-1	39	31	2.19	0.96	–24.99	0.000	–1.423	0.054
I-2	27	22	1.90	0.94	–25.02	0.000	–1.152	0.116
CK	5	5	4.07	1.00	0.11	0.319	–0.064	0.546
P-1	14	13	1.96	0.92	–8.44	0.000	–1.415	0.070
P-2	18	17	1.37	0.94	–16.98	0.000	–1.180	0.114
P-3	10	10	1.14	1.00	–7.19	0.000	–1.104	0.148

π = nucleotide diversity; CK = Cocos Keeling Islands; D = Tajima's selective neutrality test; Fs = Fu's neutrality test; *h* = haplotype diversity; I = Indian Ocean; *n* = number of individuals sampled; n_h = number of haplotypes; P = Pacific Ocean; P_D = Tajima's D and significance level P_{FS} = Fu's Fs plus significance level
Note: JC model $\gamma = 0.74$

from peripheral Pacific Ocean locations, including Western Australia and Panama, there was relatively high connectivity between Pacific locations (Table 5.2). There were surprisingly high levels of connectivity between the Cocos Keeling Islands in the Pacific Ocean clade and other Pacific Ocean individuals; however, small sample sizes at the Cocos Keeling Islands may account for this result. Alternatively, this could suggest recent dispersal from other Pacific Ocean locations to the Cocos Keeling Islands. In the Indian Ocean clade, pairwise F_{ST} scores indicate high connectivity between Oman and the Seychelles ($F_{ST} = –0.006$) (Table 5.2).

The Migrate analysis suggested relatively even and bidirectional gene flow between Indian Ocean populations (Figure 5.3a). In contrast, gene flow appeared to be highly skewed among Pacific Ocean locations, with significantly more gene flow from east to west (Figure 5.3b–c). Very few Pacific Ocean population pairs exhibited even bidirectional gene flow; the vast majority of population pairs indicated migration to be at least twice as high westwards when compared to eastwards. Gene flow was significantly higher from the east Pacific to all other locations, with between 2 and 70 times as many migrants westwards. Migration between the Cocos Keeling Islands and other Pacific Ocean locations was predominantly towards the east, with the exception of the east Pacific. Although results from the Cocos Keeling Islands should be interpreted with caution due to low sample size, Migrate results were consistent across multiple runs, and this was consistent with high connectivity indicated in pairwise F_{ST} scores between the Pacific subset of the Cocos Keeling Islands individuals and Panamanian individuals. While there was a positive relationship between genetic and geographic distance, the relationship was not significant in either the Indian Ocean clade ($Z_{IO} = 82.73$, $R^2 = 0.83$, $P = 0.33$) or the Pacific Ocean clade ($Z_{PO} = 15312.04$, $R^2 = 0.35$, $P = 0.11$).

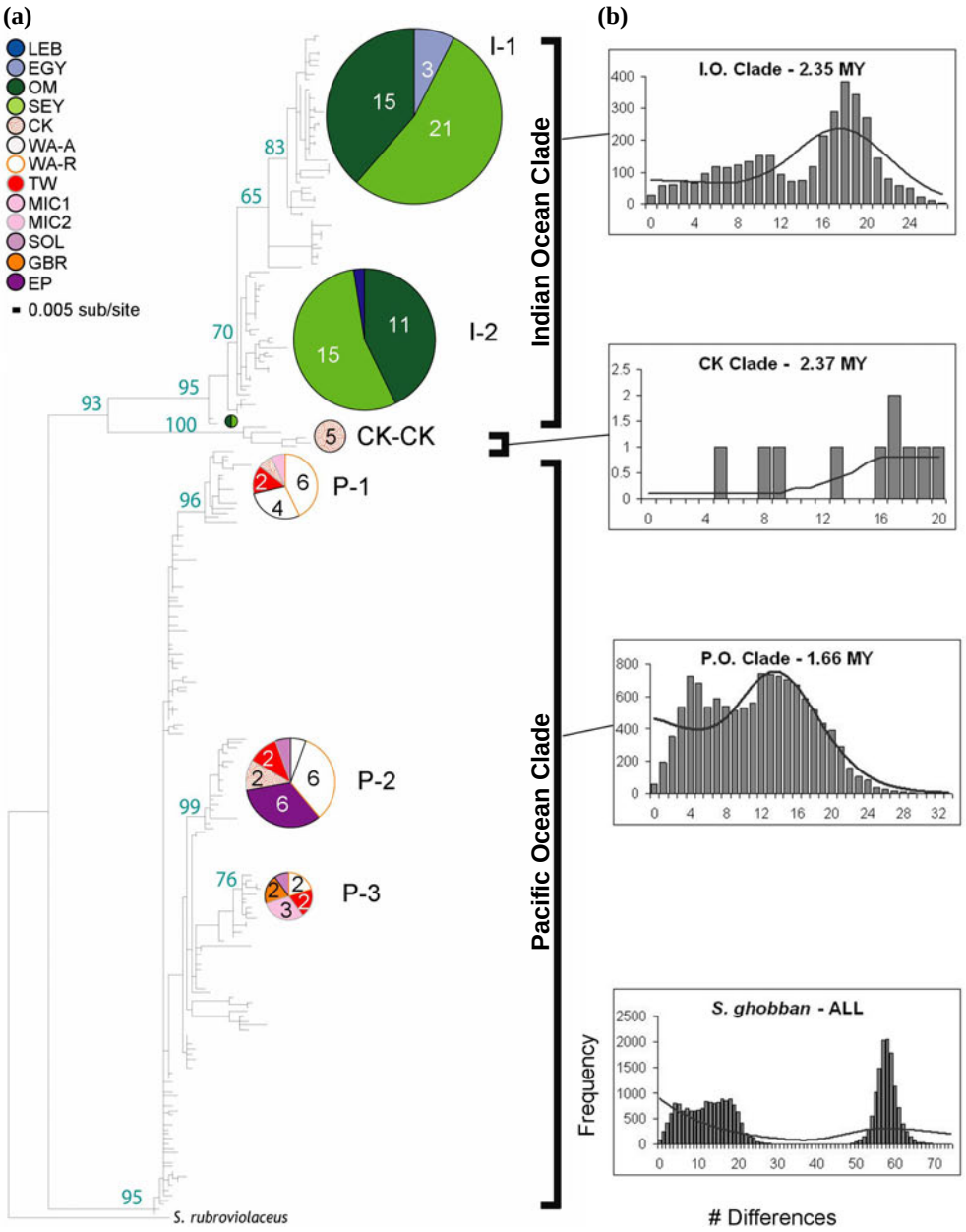


Figure 5.2(a) 50% majority-rule consensus of 15 000 best Bayesian trees for *S. ghobban*

Bayesian posterior probabilities are shown in blue. Geographic distributions of supported nodes from trees are represented by pie diagrams. Tree is out-group rooted using *S. rubroviolaceus* from the west Indian Ocean.

Figure 5.2(b) Mismatch distribution curves and coalescence times (MY) for clades determined by phylogenetic analysis, as well as all *S. ghobban* individuals

Bars indicate observed differences, curved lines indicate model frequencies.

CK = Cocos Keeling Islands; EGY = Egypt; EP = east Pacific Ocean; GBR = Great Barrier Reef; IO = Indian Ocean; LEB = Lebanon; MIC1 = Palau, Micronesia; MIC2 = Pohnpei, Micronesia; my = million years; OM= Oman; PO = Pacific Ocean; SEY = Seychelles; SOL = Solomon Islands; TW = Taiwan; WA-A = Abrolhos Island, Western Australia; WA-R = Rottnest Island, Western Australia

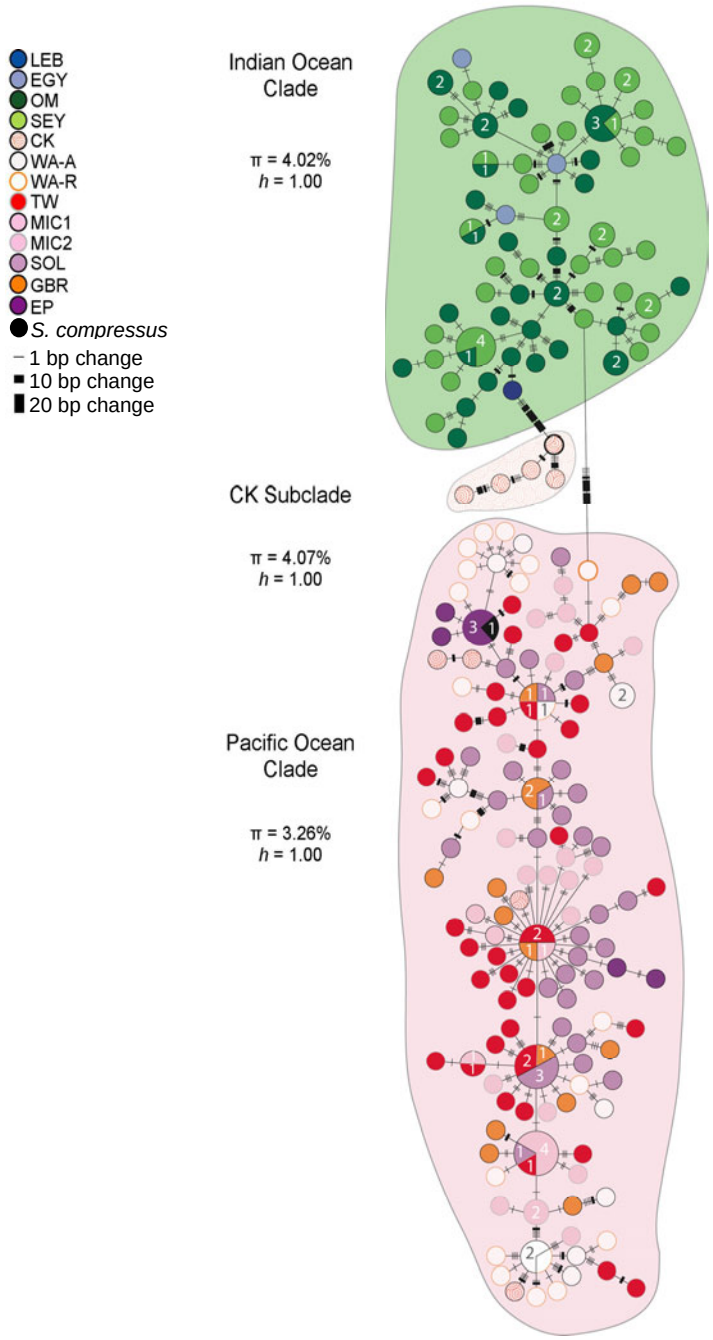


Figure 5.3 Minimum-haplotype network for *S. ghobban* haplotypes from 2 main ocean basin clades (Pacific Ocean and Indian Ocean) and the Cocos Keeling Islands subclade determined by phylogenetic analysis

Nucleotide ($\pi \pm$ SE) and haplotype ($h \pm$ SE) diversity indices are shown for Indian Ocean and Pacific Ocean clades. Small circles represent unique haplotypes and larger circles indicate shared haplotypes, with the colours and proportion of individuals from respective populations with that shared haplotype shown.

bp = base pair; CK = Cocos Keeling Islands; EGY = Egypt; EP = east Pacific; GBR = Great Barrier Reef; LEB = Lebanon; MIC1 = Palau, Micronesia; MIC2 = Pohnpei, Micronesia; my = million years; OM= Oman; SE = standard error; SEY = Seychelles; SOL = Solomon Islands; TW = Taiwan; WA-A = Abrolhos Island, Western Australia; WA-R = Rottnest Island, Western Australia

Overall, the three AMOVA analyses show that among-group variation accounts for more than 75% of the total variation and within-population variation accounts for most of the remaining variation (Tables 5.3a–b). AMOVA results of DNA sequence data partitioned based on geography showed that nearly 75% of the total variation could be explained by ocean basin (FCT = 0.74, $P < 0.01$) (Table 5.3a). When individuals were partitioned into groups based on phylogenetic analyses, the among-group variation was slightly higher (FCT = 0.82, $P < 0.01$) (Table 5.3b). A third AMOVA, in which individuals were grouped into distinct regions based on pairwise F_{ST} comparisons, revealed similar results (Table 5.3c). Despite high genetic differentiation in the Cocos Keeling Island sub population, results were nearly identical when all individuals from the Cocos Keeling Islands were removed from the analysis.

5.3.3 Coalescence analyses

Based on a total of 40% of sites mutating at a fast rate of 12.9% MY^{-1} and 60% of sites mutating at a slow rate of 1.1% MY^{-1}, the estimated rate of *S. ghobban* mitochondrial control region substitutions for the variable and conserved portions was 5.85% $bp^{-1}MY^{-1}$. Using this combined rate and a female generation time of 9 years, and assuming a constant mutation rate, estimates of mitochondrial DNA coalescence times for the Indian and Pacific ocean clades were calculated to be a mean of 2.35 mya and 1.66 mya, respectively (Table 5.4a). Although these coalescence times for the major clades covered a broad timescale and had a great deal of overlap, all coalescence ages, including upper and lower estimates, take place during the Pliocene and Pleistocene (Table 5.4a). Coalescence analyses based on supported nodes from the phylogenetic analysis add support for the Indian ocean clade I-1 being slightly older than Pacific ocean clade P-1 (Table 5.4b). As with the estimates of coalescence times based on geographic location, however, there was a great deal of overlap in the 95% confidence intervals of the clade based coalescence dates. Harpending's raggedness index was not significant for either major clade, and thus we could not reject the model of population expansion (Table 5.4a). Mismatch distributions conducted on all the data, and data divided into the respective Indian Ocean and Pacific Ocean clades, revealed strong bimodal distributions for all. This suggests the presence of multiple lineages across the distribution range and within each ocean basin (Figure 5.2). The Indian Ocean clade mismatch distribution had a small peak at 10 differences and a major peak at 18 (Figure 5.2). The mean coalescence time of the Indian Ocean clade was 2.35 mya. Mismatch distributions for both of the west Indian Ocean populations showed similar patterns of bimodality, with peaks associated with the same number of pairwise differences and similar frequencies (Figure 5.6). The mean coalescence estimates for both Indian Ocean populations were similar to the mean overall Indian Ocean (Oman = 2.33 mya, Seychelles = 2.36 mya).

Table 5.2a Population pairwise F_{ST} values (bottom diagonal) between 12 sampling locations and respective *P* values (top diagonal)

Location	OM	SEY	CK-CK	CK-PO	WA-R	WA-A	TW	MIC1	MIC2	SOL	GBR	EP
OM	*	0.567	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
SEY	–0.006	*	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
CK-CK	**0.762**	**0.754**	*	0.009	**0.000**	**0.000**	**0.000**	**0.001**	**0.000**	**0.000**	**0.000**	**0.001**
CK-PO	**0.751**	**0.748**	0.755	*	0.682	0.392	0.133	0.017	0.096	0.047	0.082	0.228
WA-R	**0.755**	**0.753**	**0.768**	–0.035	*	0.699	**0.000**	**0.000**	**0.003**	**0.000**	**0.002**	0.011
WA-A	**0.753**	**0.750**	**0.759**	0.003	–0.020	*	**0.002**	**0.000**	0.025	**0.000**	0.009	0.005
TW	**0.783**	**0.778**	**0.809**	0.053	**0.095**	**0.113**	*	0.134	0.547	0.728	0.390	**0.000**
MIC1	**0.805**	**0.794**	**0.888**	0.311	**0.215**	**0.251**	0.034	*	0.109	0.064	0.047	**0.000**
MIC2	**0.777**	**0.771**	**0.806**	0.086	**0.087**	0.085	–0.006	0.057	*	0.145	0.765	**0.000**
SOL	**0.803**	**0.794**	**0.847**	0.124	**0.127**	**0.160**	–0.008	0.060	0.017	*	0.308	**0.000**
GBR	**0.779**	**0.773**	**0.810**	0.092	**0.099**	0.105	0.000	0.093	–0.021	0.005	*	**0.000**
EP	**0.789**	**0.781**	**0.851**	0.047	0.124	0.227	**0.228**	**0.561**	**0.295**	**0.292**	**0.297**	*

CK-CK = Cocos Keeling Islands subclade; CK-PO = Cocos Keeling Islands Pacific Ocean clade; EP = Panama; GBR = Great Barrier Reef; MIC1 = Palau, Micronesia; MIC2 = Pohnpei, Micronesia; OM = Oman; SEY = Seychelles; SOL = Solomon Islands; TW = Taiwan; WA-A = Abrolhos Islands, Western Australia; WA-R = Rottnest Island, Western Australia
Notes: CK-CK, CK-PO, and MIC1 had fewer than 10 samples
Bold values indicate significant values after Bonferroni corrections ($P < 0.004$)
Based on 33,488 permutations

Table 5.2b Population pairwise F_{ST} values (bottom diagonal) between 5 well-supported nodes from the phylogenetic tree and respective *P* values (top diagonal)

Clade	I-1	I-2	P-1	P-2	P-3
I-1	*	0.000	0.000	0.000	0.000
I-2	0.646	*	0.000	0.000	0.000
P-1	0.907	0.911	*	0.000	0.000
P-2	0.911	0.923	0.700	*	0.000
P-3	0.910	0.919	0.776	0.777	*

Based on 33 488 permutations
Distance method: JC $\gamma = 0.74$

Notes: Supported nodes from Indian Ocean clade: I-1 = Egypt (3) + Seychelles (21) + Oman (15); I-2 = Lebanon (1) + Oman (11) + Seychelles (15); and Pacific Ocean clade: P-1 = Cocos Keeling Islands (1) + Rottnest Island, Western Australia (6) + Abrolhos Islands, Western Australia; (4) Taiwan (2) + Pohnpei, Micronesia; (2); P2 = Cocos Keeling Islands (2) + Taiwan (2) + Rottnest Island, Western Australia (6) + Abrolhos Islands, Western Australia (1) + east Pacific (6) + Solomon Islands (1); P-3 = Rottnest Island, Western Australia (2) + Great Barrier Reef (2) + Solomon Islands (1) + Taiwan (2) + Pohnpei, Micronesia (3)

Table 5.3a Analysis of molecular variance (AMOVA) based on 11 populations partitioned into the Indian and Pacific ocean basins

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics (P)*
Among groups (FCT)	1	3017.528	26.35838 Va	74.12%	0.74118 (0.01)
Among populations within groups (FSC)	9	463.073	2.24034 Vb	6.30%	0.24340 (0.01)
Within populations (F_{ST})	229	1594.754	6.96399 Vc	19.58%	0.80418 (0.01)
Total	**239**	**5075.355**	**35.56271**		

Notes: The two ocean basins are: 1. Indian Ocean (Oman; Seychelles; Cocos Keeling Islands); and 2. Pacific Ocean (Rottnest Island, Western Australia; Abrolhos Islands, Western Australia; Taiwan; Palau, Micronesia; Pohnpei, Micronesia; Solomon Islands; Great Barrier Reef; east Pacific)
The groups correspond to locations in the Pacific Ocean and the Indian Ocean
Significance tests are based on 33 488 permutations

Table 5.3b Regional analysis of molecular variance (AMOVA) based on 12 populations partitioned into 5 regions

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics (P)*
Among groups (FCT)	4	3585.060	22.74434 Va	78.22%	0.78224 (0.01)
Among populations within groups (FSC)	7	40.840	–0.02563 Vb	–0.09%	0.00405 (ns)
Within populations (F_{ST})	228	1449.455	6.35726 Vc	21.86%	0.78136 (0.01)
Total	**239**	**5075.355**	**29.07597**		

Notes: The 5 regions are: 1. west Indian Ocean (Oman, Seychelles), 2. Cocos Keeling Islands (CK subclade), 3. east Indian Ocean (Cocos Keeling Islands (PO subclade); Rottnest Island, Western Australia; Abrolhos Islands, Western Australia), 4. west Pacific (Taiwan; Palau, Micronesia; Pohnpei, Micronesia; Solomon Islands; Great Barrier Reef) and 5. east Pacific
Significance tests are based on 33 488 permutations

Table 5.3c Analysis of molecular variance (AMOVA) based on 12 populations divided into 3 groups based on phylogenetic analyses

Source of variation	*df*	SS	Variance components	Variation	*Φ-statistics (P)*
Among groups (FCT)	2	3505.757	30.92009 Va	82.14%	0.82135 (0.01)
Among populations within groups (FSC)	9	120.143	0.36791 Vb	0.98%	0.05471 (0.01)
Within populations (F_{ST})	228	1449.455	6.35726 Vc	16.89%	0.83113 (0.01)
Total	**239**	**5075.355**	**37.64526**		

Notes: The 3 groups are: 1. Indian Ocean clade (Oman; Seychelles), 2. Cocos Keeling Islands clade and 3. Pacific Ocean clade (Cocos Keeling Islands (PO clade); Rottnest Island, Western Australia; Abrolhos Islands, Western Australia; Taiwan; Palau, Micronesia; Pohnpei, Micronesia; Solomon Islands; Great Barrier Reef; east Pacific)
Significance tests are based on 33 488 permutations

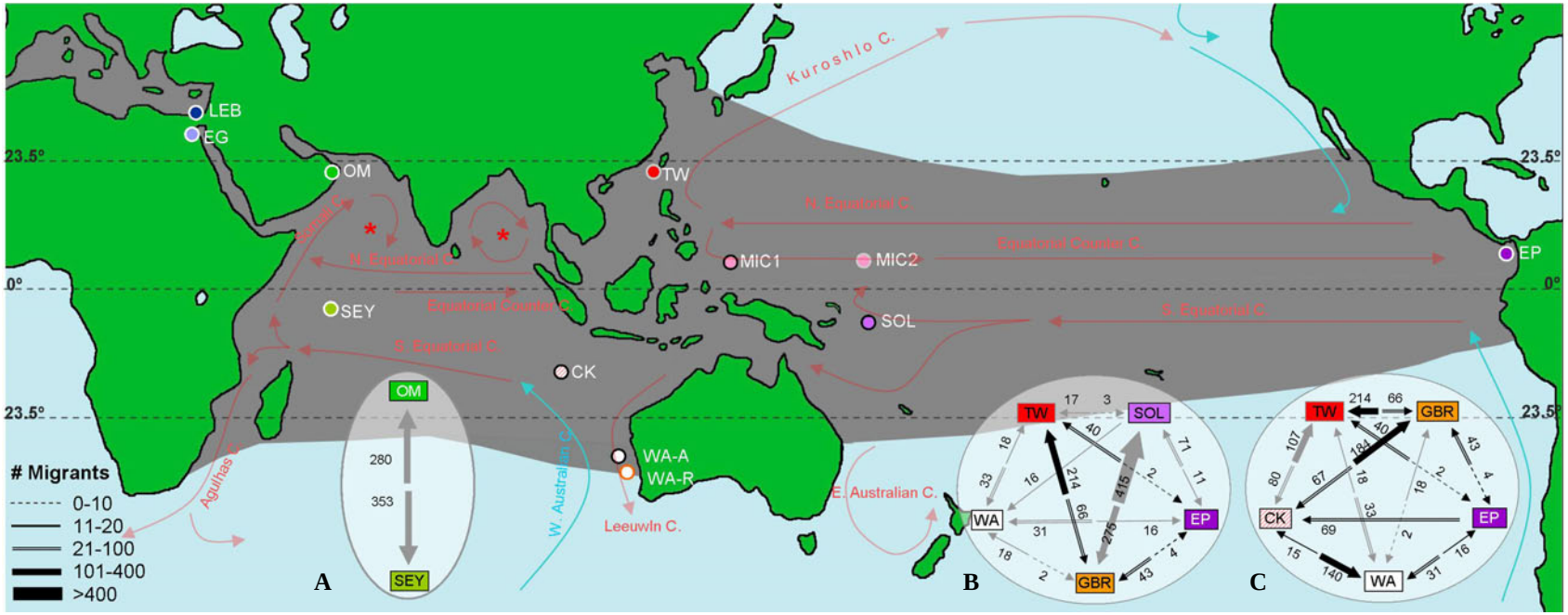


Figure 5.4 Schematic diagram of migration between populations of *Scarus ghobban*.

(a) 2 populations of *Scarus ghobban from* the Indian Ocean: OM = Oman; SEY = Seychelles; and (b and c) 5 populations from the Pacific Ocean: WA = Western Australia; TW = Taiwan; SOL = Solomon Islands; EP = east Pacific; GBR = Great Barrier Reef; CK = Cocos Keeling Islands (Pacific Island individuals only)

C = current; EG = Egypt; LEB = Lebanon; MIC1 = Palau, Micronesia; MIC2 = Pohnpei, Micronesia; N = north; S = south; W = west
Notes: Arrows correspond to direction and magnitude of gene flow, and values next to arrows indicate average number of migrants
Black arrows represent significance (i.e. overlapping migration ranges at 95%) and grey arrows indicate lack of significance between directions
Major ocean currents have also been depicted; red arrows indicate warm water currents and blue arrows indicate cold water currents.
*Currents in Arabian Sea and Bay of Bengal, including the North Equatorial Current, reverse during winter months; summer currents are shown here.

Table 5.4a Coalescence analysis parameters for 3 *S. ghobban* clades determined by phylogenetic analyses (Indian Ocean clade, Cocos Keeling Islands clade and Pacific Ocean clade), as well as individual populations

Population	Mean no. differences	Tau τ (95% CI)	Θ_0	Θ_1	SSD	*R* (*P*)	T_{MRCA} (my) (0.5–0.95 CI)
Indian Ocean clade	14.195	18.445 (12.29–23.61)	0.000	43.726	0.00912587	0.0053 (ns)	2.35 (1.56–3.00)
OM	13.930	18.322 (12.22–23.49)	0.000	44.837	0.01631568	0.0115 (ns)	2.33 (1.47–3.00)
SEY	14.463	18.523 (2.76–8.41)	0.000	47.704	0.00857747	0.0061 (ns)	2.36 (1.56–2.99)
CK-CK clade	14.200	18.658 (9.28–24.24)	0.004	142.188	0.05311931	0.1 (ns)	2.37 (1.18–3.08)
Pacific Ocean clade	11.479	13.037 (7.55–25.06)	2.526	25.147	0.00353281	0.0016 (ns)	1.66 (0.96–3.19)
CK-PO	15.667	13.896 (5.32–26.22)	5.562	115.156	0.11940124	0.2778 (ns)	1.77 (0.68–3.33)
WA-R	13.913	16.318 (10.25–20.55)	0.004	60.684	0.00402454	0.0063 (ns)	2.08 (1.30–2.61)
WA-A	14.909	16.99 (10.68–21.19)	0.00	94.98	0.02496767	0.0370 (ns)	2.16 (1.36–2.70)
TW*	11.048	2.527	11.319				0.32
MIC1	2.321	2.868 (1.03–7.39)	0.017	7.351	0.17444274	0.5893 (ns)	0.36 (0.13–0.94)
MIC2*		2.871	10.660				0.37
SOL*		2.742	5.977				0.35
GBR*	10.725	3.679 (1.85–20.56)	8.617		0.00995718	0.0144 (ns)	0.47 (0.24–2.62)
EP	6.036	8.000 (3.14–17.00)	2.703	2.739	0.14568627	0.2066 (ns)	1.02 (0.40–2.16)

CI = confidence interval; CK-CK = Cocos Keeling Islands subclade; CK-PO = Cocos Keeling Islands Pacific Ocean clade; EP = east Pacific; GBR = Great Barrier Reef; MIC1 = Palau, Micronesia; MIC2 = Pohnpei, Micronesia; my = million years; ns = not significant ($P > 0.05$); OM = Oman; SEY = Seychelles; SOL = Solomon Islands; SSD = sum of squared deviations; TW = Taiwan; WA-A = Abrolhos Islands, Western Australia; WA-R = Rottnest Island, Western Australia

*Least square procedure to fit sudden expansion model mismatch distribution and observed distribution did not converge after 1800 steps for TW, GBR, SOL, and MIC2, and coalescence dates are based on a model of spatial expansion. Not all coalescence parameters could be estimated for these populations.

Notes: Tau ($\tau \pm$ 95% CI) is the unit of mutational time between two populations; $theta_0$ (Θ_0) is the mutation parameter before expansion; and $theta_1$ (Θ_1) is the mutation parameter after expansion.

Harpending's raggedness index and significance based on the significance of simulated and observed raggedness are also shown

Fossil-calibrated dates for *S. ghobban*: 1.4–2.0 my

CK-CK, CK-PO, MIC1, and EP all had $n < 10$

Table 5.4b Coalescence analysis parameters for *S. ghobban*-supported nodes from phylogenetic analyses

Clade	Mean no. differences	Tau τ (95% CI)	Θ_0	Θ_1	SSD	R (P)	T_{MRCA} (my) (0.5–0.95 CI)
I-1	7.466	9.506 (4.72–13.50)	0.000	25.977	0.0066114	0.0116 (ns)	1.21 (0.60–1.71)
I-2	35.508	6.074 (3.35–11.27)	1.364	35.508	0.00544113	0.0149 (ns)	0.77 (0.43–1.43)
P-1	7.088	8.168 (3.44–12.01)	0.005	24.834	0.01386942	0.0376 (ns)	1.04 (0.44–1.53)
P-2	4.686	3.404 (1.46–9.78)	1.789	36.445	0.00229049	0.0118 (ns)	0.43 (0.19–1.24)
P-3	3.756	4.084 (1.90–6.03)	0.000	99999	0.02755156	0.0770 (ns)	0.52 (0.23–0.77)

CI = confidence interval; I = Indian Ocean; my = million years; ns = not significant ($P > 0.05$); P = Pacific Ocean; SSD = sum of squared deviations;
Notes: Tau ($\tau \pm$ 95% CI) is the unit of mutational time between two populations; $theta_0$ (Θ_0) is the mutation parameter before expansion; and $theta_1$ (Θ_1) is the mutation parameter after expansion.
Harpending's raggedness index and significance based on the significance of simulated and observed raggedness are also shown

The Pacific Ocean mismatch distribution was also bimodal, with one peak at 5 differences and a second at 15 differences (Figure 5.2), and location specific mismatches were also bimodal for most Pacific Ocean clade locations (Figure 5.6). Coalescence estimates placed the mean T_{MRCA} in the Pacific Ocean clade at 1.66 mya. Mean coalescence dates for individual Pacific Ocean clade populations location haplotypes ranged between 0.35 mya for Taiwan and 2.11 mya for Western Australia (Figure 5.6).

5.4 Discussion

5.4.1 Genetic partitioning between Indian and Pacific Ocean lineages

This study provides evidence of genetic partitioning between western Indian Ocean and Pacific Ocean populations of *S. ghobban* based on phylogenetic and population genetic analyses. Our findings are consistent with those by Bariche and Bernardi (2009) based on both mitochondrial and nuclear markers. In all three AMOVA analyses for this species, among group variation (FCT) accounts for the majority of the variation, and all are highly significant, providing strong support for population structure between ocean basins. The remaining variation was generally attributed to differences within populations. These ($P < 0.001$), and were highly significant even after conservative Bonferroni corrections were

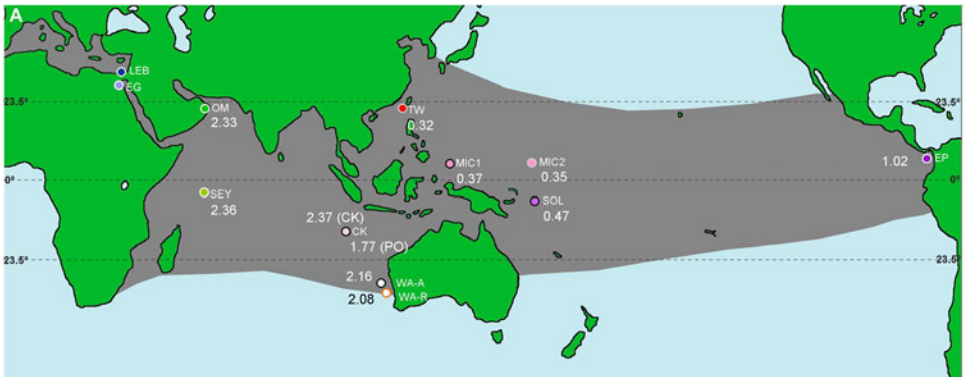


Figure 5.6(a) Map of *S. ghobban* with coalescence dates (MY) next to respective locations for which coalescence analysis was possible

CK-CK = Cocos Keeling Islands, Cocos Keeling Islands clade; CK-PO = Cocos Keeling Islands, Pacific Ocean clade; EGY = Egypt; EP = east Pacific; GBR = Great Barrier Reef; LEB = Lebanon; MIC1 = Palau, Micronesia; MIC2 = Pohnpei, Micronesia; my = million years; OM= Oman; SEY = Seychelles; SOL = Solomon Islands; TW = Taiwan; WA-A = Abrolhos Island, Western Australia; WA-R = Rottnest Island, Western Australia

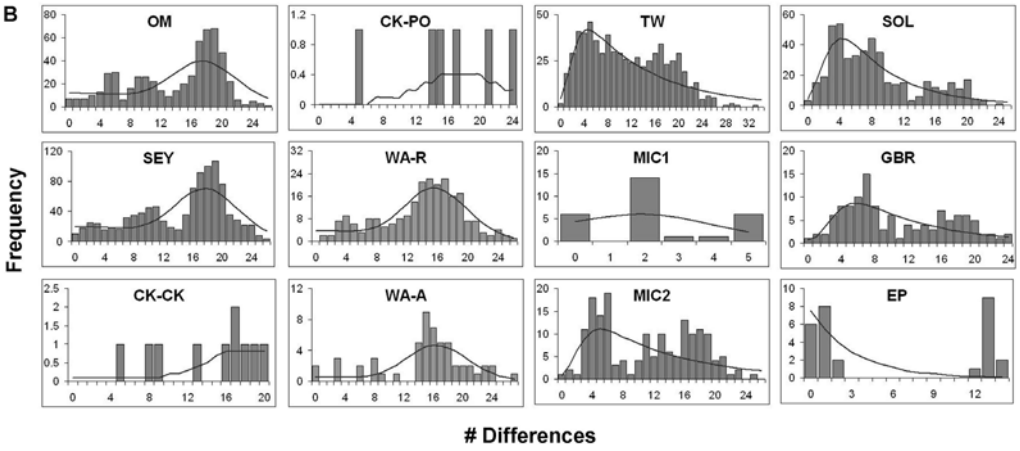


Figure 5.6(b) Mismatch distribution curves for 12 populations depicting frequency of observed pairwise differences (bars) and model frequencies (line)

Based on model of sudden expansion. For all locations, Harpending's raggedness index was not significant (P = 0.05).

AMOVA results are consistent with results for both *C. sordidus* (Bay *et al.*, 2004, Chapter 3) and *S. rubroviolaceus* (Chapter 4). Differences between western Indian Ocean and Pacific Ocean populations of *S. ghobban* have also been detected on a morphological level (Figure 2.1; Choat pers. comm.), and the classification of this species has been a long-standing debate (Cuvier, Valeciennes, 1939) since it was first described in 1775 (Forsskål, 1775). We find evidence of very limited gene flow between the two ocean basins, and all pairwise F_{ST} comparisons between Indian Ocean and Pacific Ocean populations were greater than 0.796 applied. The location of the split between Indian

Ocean and Pacific Ocean *S. ghobban* lineages is either west of the Cocos Keeling Islands, as indicated by the affiliation of nearly half of the Cocos Keeling Islands samples with the Pacific Ocean rather than the Indian Ocean clade, or at the Cocos Keeling Islands region. This latter split is based on the presence of a Cocos Keeling Islands-specific clade that consists of more than half of the sampled individuals from the Cocos Keeling Islands (Figures 5.2 and 5.3). Patterns of a genetic break occurring at the Cocos Keeling Islands region separating western Indian Ocean and Pacific Ocean populations have been well documented in a number of other reef fishes (Bay *et al.*, 2004; McMillan, Palumbi, 1995; Palumbi, 1997), including *C. sordidus* and *S. rubroviolaceus* (Chapters 3 and 4), as well as many marine invertebrates (Lavery *et al.*, 1996a; Williams, Benzie, 1996; Williams, Benzie, 1997). Although many of the species in these studies possess sufficient pelagic larval durations to disperse across the entire Indian Ocean, and despite the presence of currents that would facilitate dispersal from the eastern Indian Ocean into the western Indian Ocean, the concurrent patterns of genetic differentiation in this region are most likely attributed to historic separations due to sea-level changes associated with periods of Quaternary glacial maxima.

5.4.2 Population structure at the Cocos Keeling Islands

Although the Cocos Keeling Islands subclade comprised only 5 individuals out of a total of 9 from this location, and any results from population genetic analyses should be interpreted with caution, congruent results across all analyses do suggest that the results are not due to low sample numbers or stochasticity. The subclade had 100% bootstrap support in the phylogenetic tree, and was more different to the main Indian Ocean and Pacific Ocean clades than the Indian Ocean and Pacific Ocean clades were from one another based on the minimum-spanning tree (Figure 5.3) and pairwise F_{ST} comparisons. In addition, the time of coalescence for the Cocos Keeling subclade was very similar to other Indian Ocean locations, which is consistent with an Indian Ocean origin for extant populations of *S. ghobban*. The timing of coalescence for the Pacific Cocos Keeling Island population suggests that the haplotypes in the Pacific clade are younger than those in the Cocos Keeling Island subclade and western Indian Ocean clade, indicating that gene flow between the Cocos Keeling Islands and the Pacific Ocean is much more recent. This is supported by the high genetic connectivity indicated by pairwise F_{ST} comparisons, as well as Migrate results, which show very high gene flow from the Cocos Keeling Islands to other Pacific Ocean locations, with the exception of the east Pacific. Although the algorithms used in the Migrate program are extremely robust, these results are based on a very small sample size of 4 individuals. Coalescence dates based on well-supported

nodes in the phylogenetic analysis also indicate that the group mostly containing eastern Indian Ocean individuals (P-1) is older than the two other supported nodes in the Pacific Ocean clade. These results suggest that the Cocos Keeling Islands may be the source population for the extant Pacific Ocean clade. Until more data can be obtained from the Cocos Keeling Islands, however, our conclusions regarding this population remain speculative. Palumbi (1997) cautions interpretations based on skewed samples, stating that populations that are larger and have greater genetic diversity will appear older, even if they are not. This is an important caveat given that not only are there twice as many samples from the Pacific Ocean as there are from the Indian Ocean for *S. ghobban*, but there are only 9 samples from the Cocos Keeling Islands.

The Cocos Keeling Islands are situated near the Indonesian Throughflow, a narrow but deep passage of ocean. Here, water from the Indian and Pacific Oceans is exchanged, which supports a source population hypothesis. Although bathymetric and topographic conditions of the Indonesian Throughflow have changed little in the last 2 million years, the area around the Cocos Keeling Islands and the Indo–Australian Archipelago has undergone dramatic geological changes, including the emergence of new land masses, the formation of mountains and volcanoes, and the subsidence of submarine basins. All of these phenomena would have significantly lowered sea-surface temperatures (Kuhnt, Holbourn, 2004). Computer-based simulations during the last glacial maximum revealed that the Indonesian Throughflow, particularly through the Makassar Strait, was not significantly reduced, but that shallow-water transport from the Pacific Ocean was blocked by the Ombai and Timor straits (Zuvela-Aloise, 2005). The pattern observed at the Cocos Keeling Islands may thus be the result of allopatric speciation in which the continuous range was split into geographically isolated regions during low sea-level stands.

Given the proximity of the Cocos Keeling Islands to large seamounts and cold water upwelling zones (Brewer *et al.*, 2009; Werner, 2008), and given what we know about the ability of *S. ghobban* to withstand subtropical environments, it is feasible that the Cocos Keeling subclade represents a population of *S. ghobban* that is thermally adapted to less tropical reef-like environments, which would have resembled the environment in this area during times of low sea level. In other words, the individuals exclusively from the Cocos Keeling subclade may represent partitioning of a cryptic sister species by depth or some other relevant ecological variable. Interestingly, all *S. ghobban* from the Cocos Keeling Islands, those exclusively from the Cocos Keeling subclade and those with Pacific Ocean genetic affinities were morphologically identical to Pacific Ocean conspecifics. There are

two possible explanations of this interesting genetic pattern. Firstly, although all individuals from the Cocos Keeling Islands were speared near the surface, it is possible that the Cocos Keeling Islands individuals exclusively from the Cocos Keeling subclade and those from the Pacific clade do co-exist, but maintain genetic differentiation because the tropical (Pacific Ocean clade) forms cannot survive in the deeper, cooler environment. Secondly, the Cocos Keeling subclade may represent hybridisation of two co-occurring species, as mitochondrial DNA is inherited maternally. This geographic region near the Indo–Pacific biogeographic border is an established suture zone, where species from both ocean basins not only overlap (Hobbs, Salmond, 2008) but readily hybridise (Hobbs et al., 2009). Field observations and further sampling from different depths at the Cocos Keeling Islands and surrounding areas—including Western Australia where *S. ghobban* was observed at a depth of 250 m—would be useful in testing this hypothesis. The use of nuclear markers may also aid in resolving taxonomic ambiguities and detecting whether hybridisation is occurring here.

5.4.3 Population structure at peripheral Pacific Ocean clade populations

We also found evidence of genetic partitioning within the Pacific Ocean, particularly at peripheral locations, and migration predominantly from east to west. Locations on the periphery of the Pacific Ocean clade, including Western Australia and the east Pacific, showed significant population structure between other Pacific Ocean locations based on pairwise F_{ST} comparisons. Although pairwise F_{ST} values between Western Australia and the rest of the Pacific Ocean were not particularly high, they were all significant. The east Pacific had particularly high pairwise F_{ST} values when paired with locations in the west Pacific Ocean, which suggests low gene flow between west and east Pacific populations. Coalescence estimates suggest very recent expansion for extant west Pacific Ocean populations, ranging from 0.32 to 0.47 mya. Results from the Migrate analysis confirm highly uneven gene flow between the east Pacific and the Indo–Australian Archipelago. Westward gene flow from the east Pacific to the Indo–Australian Archipelago was the highest by between 2 and 23 times. A study on the East Pacific Barrier by Lessios & Robertson (2006) revealed very high divergence between central and east Pacific populations in *S. ghobban*, and concluded that the East Pacific Barrier poses a sufficient barrier to dispersal in *S. ghobban*. The authors also make the point that different factors are involved in an initial invasion and subsequent gene flow, and that the maintenance of gene flow across such a barrier requires well-established populations on both sides. It is interesting to note that *S. ghobban*, although abundant in the east Pacific, does not occur in Hawaii like so many scarine labrids that exist on both sides of the East Pacific Barrier.

The absence of *S. ghobban* at Hawaii may be testament to the fortuitous and stochastic nature of effective dispersal in the marine environment.

The isolation of the east Pacific from other Pacific Ocean locations may be an important factor in the rise of its sister species, *S. compressus*, which is morphologically dissimilar but has the same mitochondrial DNA as *S. ghobban* from the east Pacific location. Future studies on the phylogeography of *S. compressus* will provide valuable information on the evolutionary history of *S. ghobban*, as well as insight regarding possible hybridisation between these two species in the east Pacific and likely modes of speciation (e.g. peripatric or parapatric).

5.5.4 Historical and present-day gene flow in S. ghobban

Interestingly, the mean coalescent ages of both offshore (Abrolhos Islands) and inshore (Rottnest Island) Western Australian populations (2.08–2.16 mya) were older than that of the overall Pacific Ocean populations (1.66 mya). This is consistent with an Indian Ocean origin of extant *S. ghobban* populations, followed by an eastward migration via Western Australia. The second-oldest Pacific Ocean population was the Cocos Keeling Islands Pacific Ocean subsample (1.77 mya), followed by the east Pacific population, which was half the age of the Western Australian population (1.02 mya). This suggests that the initial direction of colonisation of the Pacific Ocean populations was from the Cocos Keeling Islands to Western Australia to the east Pacific. This is reflected in the bidirectional migration between Western Australia and the east Pacific (Figure 5.4). It is important to note that coalescence calculations generated very wide confidence intervals, and almost all dates are overlapping. More robust analyses would be helpful in resolving this. It is also interesting that these results contradict those from Migrate analyses, which are also coalescence based, and indicate that migration was unidirectional from the east Pacific Ocean to the Cocos Keeling Islands. In addition, the P-2 node in the Pacific Ocean clade contains almost all of the east Pacific individuals, as well as some from the Cocos Keeling Islands and Western Australia, providing evidence of gene flow between these populations in the past. Additional specimens and data from the Cocos Keeling Islands are required to resolve this issue, because the Cocos Keeling Pacific Ocean clade consisted of 4 only individuals. Thus, low sample numbers from this genetically bifurcated location may be introducing artefacts.

Given that the Western Australian, Cocos Keeling (Pacific Ocean) and the east Pacific are the oldest populations based on coalescence analyses, one would expect to find

haplotypes shared between these locations in a central position in the minimum-haplotype network. However, that is not the case. The majority of east Pacific haplotypes appear to be peripheral on the minimum-spanning network, with mostly Western Australian haplotypes nearby (Figure 5.3). Low sample numbers may also be responsible for some of the patterns observed in the east Pacific populations (n = 8) and further sampling from this location may also help resolve some of these apparent discrepancies. Once in the east Pacific Ocean, *S. ghobban* likely colonised the rest of the Pacific Ocean, from the Great Barrier Reef (0.47 mya) and up through the Indo–Australian Archipelago; extant populations in the Indo–Australian Archipelago all appear to have recent expansion times (0.32–0.37 mya). This is consistent with a centre of accumulation model of Indo-Australian Archipelago diversity (Jokiel, Martinelli, 1992; Ladd, 1960). The bimodal mismatch distributions, and combination of high nucleotide and haplotype diversity in the majority of populations indicate a history of fragmentation associated with changes in sea level and oceanographic conditions. Population pairwise F_{ST} comparisons between the well-supported nodes from each major clade also suggest a high degree of population structure that is temporally rather than geographically based. While fossil-calibrated dates estimate the initial diversification of this lineage to be between 1.41 and 1.97 mya, our dates are consistent with evidence from nuclear data incorporating codon rates and penalised likelihood rates, which suggest *S. ghobban* first arose between 2 and 3.5 mya (Choat *et al.*, 2012; Smith *et al.*, 2008).

5.5 Conclusions

Overall, it is clear from our results that there are a number of complex factors working together, and further detailed analyses are needed in order to resolve some of the discrepancies in this study. We conclude that the Indian Ocean lineage is more ancient than the Pacific Ocean lineage in extant populations of *S. ghobban*, appearing approximately 2.4 million years ago, followed by an eastward migration. High gene flow between the Cocos Keeling Islands and Western Australia, both in the east Indian Ocean, and other Pacific Ocean locations support the hypothesis of historical gene flow from west to east. In addition, high gene flow from the east Pacific to central and west Pacific Ocean locations is indicative of present-day gene flow between these populations. The populations in the east Pacific appear to be the most structured and the East Pacific Barrier has undoubtedly been highly instrumental in these patterns. It is also evident that a genetic break associated with the Cocos Keeling Islands played an important role in the diversification of *S. ghobban*, and most likely acted in concert with oceanographic features, including cold water upwelling and Pleistocene sea-level changes to result in the

highly structured populations in the west Indian Ocean and Pacific Ocean. More complete sampling from the Cocos Keeling Islands and adjacent locations, as well as from the east Pacific, would allow us to better understand the evolutionary history of *S. ghobban* and provide valuable insight into the different modes of cryptic speciation in widespread tropical reef fish. In addition, a further taxonomic investigation is warranted in all three evolutionary lineages in order to detect subtle differences in colour or structure that may have been previously missed and would support the phylogenetic divisions identified here.

Chapter 6 Comparative phylogeography of Indo–Pacific reef fish: a synthesis of four Indo–Pacific scarine labrids

Abstract

We examined the population structure and evolutionary history of four Indo–Pacific scarine labrids: *Chlorurus sordidus*, *Scarus rubroviolaceus*, *S. ghobban* and *S. psittacus*. Our approach combined phylogenetic, phylogeographic and population genetic analyses, together with life history information. Phylogenetic analyses identified strong population structure, with a genetic break between west Indo–Pacific and east/central Indo–Pacific extant populations in three of the four species. Pairwise F_{ST} comparisons confirmed this break in the same three species ($F_{ST} \geq 0.573$, $P < 0.00065$). This indicated that each species consists of a main, central stock, which comprised central and eastern Indo–Pacific individuals, and several smaller stocks associated with peripheral locations. Genetic diversity across species varied greatly: *S. psittacus* consistently exhibited very low (< 1%) variation in extant populations (nucleotide diversity [π] = 0.52–0.84%), while *S. ghobban* individuals exhibited the highest variation (π = 0.66–11.6%). Major migration routes were predominantly from east to west in the central main stock and from the eastern peripheral populations. Isolation at peripheral locations played an integral role in the generation of biodiversity, while overlap at central locations was important in the maintenance of biodiversity. We found genealogical concordance in the patterns of population structure at both large and small spatial scales in all species but *S. psittacus*. The shallow genetic architecture exhibited by *S. psittacus* could be attributed to differences in the size and level of population structuring in its ancestral populations compared with the other three species. Despite its shallow genetic architecture, *S. psittacus* exhibited the same basic stock structure as other three species. This suggests that population structure may be cyclical, and fluctuates between high and low genetic diversity over evolutionary timescales. The ecologically marginal conditions found at peripheral habitats may provide more opportunities for selection and adaptive change compared to central regions.

6.1 Introduction

Many widely distributed tropical marine taxa exhibit genetic partitioning over the entire Indo–Pacific, often with genetically distinct clades on either side of the Indo–West Pacific Barrier (McManus, 1985). These include a number of fish species (Bay *et al.*, 2004;

Bernardi *et al.*, 2002; Gaither *et al.*, 2010; Hickerson, Cunningham, 2005; McMillan, Palumbi, 1995; Planes, Fauvelot, 2002), sea urchins (Lessios *et al.*, 2001), gastropods (Crandall *et al.*, 2007; Reid *et al.*, 2006), and stomatopods (Barber *et al.*, 2006). There are also many species that do not exhibit a clear partition across this barrier, among them several surgeonfishes from the genus *Naso* (Horne *et al.*, 2008; Klanten *et al.*, 2007), the parrotfish *Scarus psittacus* (Winters *et al.*, 2010), two species of moray eel from the genus *Gymnothorax* (Reece *et al.*, 2010) and the lutjanid *Lutjanus kasmira* (Gaither *et al.*, 2010). Gaither *et al.* (2010) observed two co-distributed and genetically similar lutjanids with markedly different phylogeographic patterns; however, Horne *et al.* (2008) found congruent patterns of high genetic diversity and little population genetic structure across three acanthurids with similar tropical Indo–Pacific distributions. The differences in genetic partitioning of these widely distributed reef fish indicate that the efficacy of such barriers varies greatly between species and there does not appear to be a consistent barrier between Indian Ocean and Pacific Ocean fauna.

In the examples of strong population structure at both large and small spatial scales, there is a peripheral marginal habitat contributing to that structure, emphasising the importance of peripheral and isolated populations in marine speciation and biodiversity (Briggs, 2005; Mayr, 1963), as well as in structuring communities of marine organisms (e.g., Gavrilets *et al.*, 2000; Johannesson, André, 2006) including fish (Bernardi *et al.*, 2003; Bernardi *et al.*, 2002; Gaither *et al.*, 2010; Rocha, Bowen, 2008), sea urchins (Palumbi *et al.*, 1997), reef corals (Budd, Pandolfi, 2010) and hermit crabs (Malay, Paulay, 2009). In a review of isolation at peripheral habitats, Hardie and Hutchings (2010) found that populations on species' margins tend to be less stable than the core populations and more susceptible to stress adaptation. Malay and Paulay (2009) revealed that isolation on peripheral islands accounted for the majority of speciation in the *Calcinus* species of coral reef hermit crab. This contradicts the numerous theories that identify the central region known as the Indo–Australian Archipelago as the centre of biodiversity and speciation (reviewed in Palumbi, 1997). If central regions are centres of biodiversity, we would expect central lineages to be older than peripheral lineages. In contrast, if central locations were centres of accumulation, peripheral populations are expected to be older than central lineages (Barber, Bellwood, 2005). A comparison of co-distributed widespread species from the same family provides a good opportunity to address the role of peripheral and central populations in the formation and maintenance of emergent biodiversity in coral reef fishes. Scarine labrids are some of the most recognisable members of coral reef fish fauna, and combined with their high diversity, high degree of reef association and wide distribution range, they make an ideal study group for a comparison of this nature. Furthermore, two of the species studied,

S rubroviolaceus and *S. ghobban*, have sister species that are restricted to peripheral populations. *S. rubroviolaceus* has two sister species, *S. persicus* and *S. ferrugineus*, which have similar meristics and colour patterns to *S. rubroviolaceus*, but are restricted to the west Indo–Pacific (Bruce, Randall, 1983). *S. persicus* is confined to northern Oman and the Persian Gulf (Bruce, Randall, 1983), and *S. ferrugineus* is found only in the Red Sea and the Gulf of Aden (Parenti, Randall, 2000). At the eastern periphery of its distribution, *S. ghobban* has a sister species, *S. compressus*, which is restricted to the tropical east Pacific (Parenti, Randall, 2000). Although adult colouration is very distinct between the two, they have similar colour patterns in their initial phases (Rosenblatt, Hobson, 1969).

The third species in this comparison, *Chlorurus sordidus*, was found to exhibit significant population structure between west Indo–Pacific and east Indo–Pacific individuals, as well as significant population structure among individuals from peripheral locations (Bay *et al*. 2004; Chapter 3). The fourth species in this comparison, *S. psittacus*, which has similar life history characteristics and geographic distribution to *C. sordidus*, demonstrates distinctive population structure at peripheral populations within biogeographic realms, but none at the largest spatial scale (Winters *et al*., 2010). Although this appears to contradict previous patterns for this family, Winters *et al*. (2010) noticed a marked discrepancy between the species age (4.5 my, Alfaro *et al*., 2009) and the coalescent age of *S. psittacus* (<0.163 my, Winters *et al*., 2010). We expect species divergence to predate lineage coalescence because divergence and coalescence are inherently different evolutionary perspectives. On one hand, divergence occurs when populations become differentiated as we look forward in time. On the other hand, coalescence involves gene lineages becoming more similar to the point where they shared a common ancestor as we look from the present backwards in time (Rosenberg, Feldman, 2002). While we would expect species age to be older than coalescent age, the species and coalescent ages of *S. psituacus* differed by more than an order of magnitude, demonstrating that the coalescent age of extant populations can be significantly younger than the age of the species. Thus, coalescent age may not be congruent with species age, and genetic diversity may be much lower than expected if solely based on the age of a species. This incongruence also indicates that population structure in widespread reef fish is responsive to a number of physical factors including sea-level fluctuations and associated habitat gains or losses, sea-surface temperature fluctuations, isolation at peripheral or marginal habitats, overlap at central locations, and biological and ecological differences, even between closely related species within the same family.

Table 6.1 Four aspects of genealogical concordance in phylogeographic inference (after Avise, 2000).

I	*Concordance across sequene characters within a gene.* Relevance: yields statistical significance for putative gene-tree clades.
II	*Concordance in significant genealogical partitions across multiple genes within a species.* Relevance: establishes that gene-tree partitions register phylogenetic partitions at the population or species level.
III	*Concordance in the geography of gene-tree partitions across multiple codistributed species.* Relevance: implicates shared historical biogeographic factors in shaping intraspecific phylogenies.
IV	*Concordance of gene-tree partitions with spatial boundaries between traditionally recognized biogeographic provinces.* Relevance: implicates shared historical biogeographic factors in shaping intraspecific phylogenies of organismal distributions.

Avise (2000; 1990) discusses four aspects of genealogical concordance (Table 6.1) which are paramount to phylogeographic inference: Aspect I deals with agreement across characters within a gene, i.e., when lineages or clades defined by individual trees have high bootstrap support. Aspect II deals with agreement across genes, i.e., when the same phylogenetic patterns are found across independent molecular markers. Aspect III deals with agreement across codistributed species, i.e., the same well-supported phylogeographic lineages are found between species with similar life histories and geographic distributions. Finally, Aspect IV deals with agreement of gene-tree partitions between established biogeographic hotspots. Avise (2000; 1990) states that when such concordance arises, the natural stochasticity associated with Mendelian inheritance is most likely not responsible for the observed patterns. Together, these aspects of genealogical concordance provide important insights into the evolution and speciation processes of the taxa examined as well as the biogeographic regions involved. Moreover, they can be used to inform taxonomic decisions, as well as in the delimitation of management units (MU) or evolutionarily significant units (ESU) in conservation and management. In addition to genealogical concordance, genealogical **discordance** can be just as valuable in providing insights into the evolution and history of a species or a group of species.

In this study, we compare and contrast phylogenetic, biogeographic, phylogeographic and population genetic patterns of four Indo–Pacific scarine labrids with similar evolutionary ages

and life history characteristics, including data from Winters *et al*. (2010), in order to address the following broader questions:

1. Are there congruent patterns of population structure in congeneric or confamilial widespread reef fish across the Indo–Pacific?

2. What is the level of genealogical concordance (or discordance) between the four species?

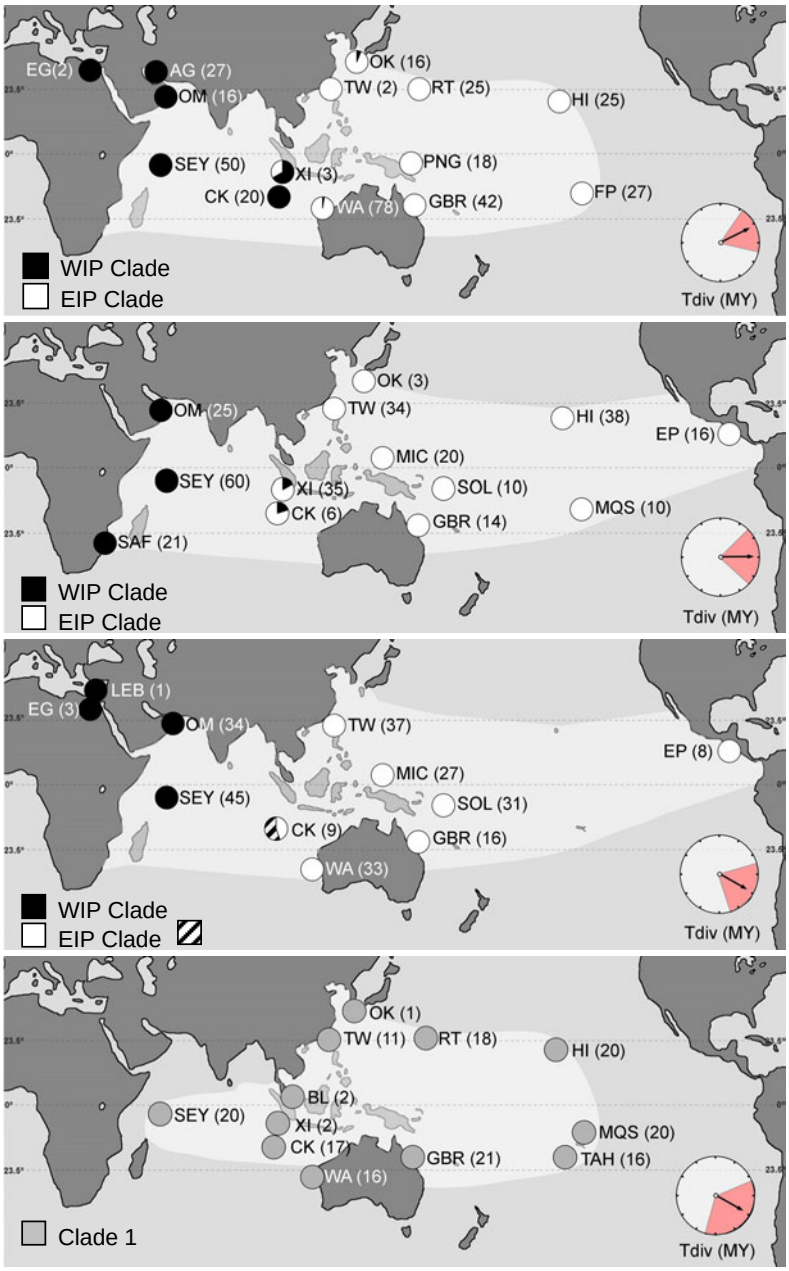
3. What is the role of peripheral and/or isolated environments in generating and maintaining biodiversity? Do these scarine labrids support a 'centre of origin', 'centre of accumulation', or 'centre of overlap' model of the Indo–Australian Archipelago centre of biodiversity?

4. Do coalescent and species nodal ages differ and, if so, what does the discrepancy tell us about the evolutionary history of populations of widespread species?

6.2 Results

6.2.1 Phylogenetic analyses

Phylogenetic analyses revealed strong population structure and evidence of a genetic break between individuals from the west Indo–Pacific and east/central Indo–Pacific in all species except *S. psittacus* (Figure 6.1a–d). In the populations sampled, the genetic break occurred at the western edge of the central Indo–Pacific marine bioregion. In *C. sordidus*, the break occurred at Christmas Island, as seen by the division of haplotypes from this location between the west Indo–Pacific and east Indo–Pacific clades (Figure 6.1a). In *S. rubroviolaceus*, this break was shifted slightly to the west, as haplotypes from both Christmas Island and the Cocos Keeling Islands were split about 20:80 between the west Indo–Pacific and east Indo–Pacific clades, respectively (Figure 6.1b). In *S. ghobban*, the break was at the Cocos Keeling Islands, shown by a third well-supported clade containing over 50% of individuals exclusively from the Cocos Keeling Islands (Figure 6.1c). Age estimates based on mitochondrial and nuclear data indicate that *C. sordidus* is the most recently diverged (1.64 MY), and *S. psittacus* is the oldest (3.15 MY) (Choat et al., 2012) (Figure 6.1a–d) .



(a) *C. sordidus*

(b) *S.rubroviolaceus*

(c) *S. ghobban*

(d) *S. psittacus*

Figure 6.1(a-d) Species distribution (light shaded area of map), proportion of individuals in different clades (pie diagrams) and expansion times in millions of years (clock diagrams) for: (a) *C. sordidus*, (b) *S. rubroviolaceus*, (c) *S. ghobban* and (d) *S. psittacus*

AG = Arabian Gulf; BL = Bali; CK = Cocos Keeling Islands; EIP = east Indo–Pacific; EG = Egypt; EP = east Pacific (Panama and Clipperton); FP = French Polynesia; GBR = Great Barrier Reef; HI = Hawaii; LEB = Lebanon; MIC = Micronesia (Palau and Pohnpei); MQS = Marquesas; my = million years; OK = Okinawa; OM = Oman; PNG = Papua New Guinea; RT = Rota; SAF = South Africa; SEY = Seychelles; SOL = Solomon Islands; TAH = Tahiti; T_{Div} = divergence time; TW = Taiwan; WA = West Australia; WIP = west Indo–Pacific; XI = Christmas Island

Note: Black arrow on clock indicates the average divergence time and the red shaded area of the clock indicates 95% confidence intervals (after Alfaro *et al*. 2009)

6.2.2 Population genetic analyses

Pairwise F_{ST} comparisons confirmed the split between west Indo–Pacific and central/east Indo–Pacific populations in all but *S. psittacus* (tables 6.2a–d). They also revealed low levels of gene flow and reduced genetic diversity in peripheral populations. All species showed peripheral isolation in the eastern margins of their distributions (tables 6.2a–d, figures 6.3a–d). Hawaii, the east Pacific and French Polynesia consistently exhibited low levels of gene flow between themselves and other Pacific Ocean locations (Tables 6.2a–d). In addition to these eastern peripheral stocks, all species had a greater central stock, consisting of the majority of individuals from the east Indo–Pacific (Figures 6.3a–d). *S. rubroviolaceus* and *S. ghobban* also had a main west Indo–Pacific stock. *Chlorurus sordidus* exhibited significant population structure in Oman and the Arabian Gulf (Figure 6.3a), and *S. psittacus* had a main stock that comprised individuals from Seychelles and another main stock that comprised individuals from the Cocos Keeling Islands (Figure 6.3d). Although *S. psittacus* did not exhibit a defined split between west Indo–Pacific and east Indo–Pacific individuals, pairwise F_{ST} comparisons revealed more genetic distance between the Marquesas and Tahiti populations (F_{ST} = 0.292, $P < 0.0014$) than between the Marquesas and Seychelles populations (F_{ST} = 0.166, $P < 0.0014$) (Table 6.2d). The nucleotide diversity (π) was very low in all populations of *S. psittacus*, ranging from 0.52% in Hawaii to 0.84% in the Seychelles (Table 6.2d, Figure 6.2). The other three species had much higher nucleotide diversity, with the highest occurring in the Cocos Keeling Islands population of *S. ghobban* (11.6%) (Figure 6.2).

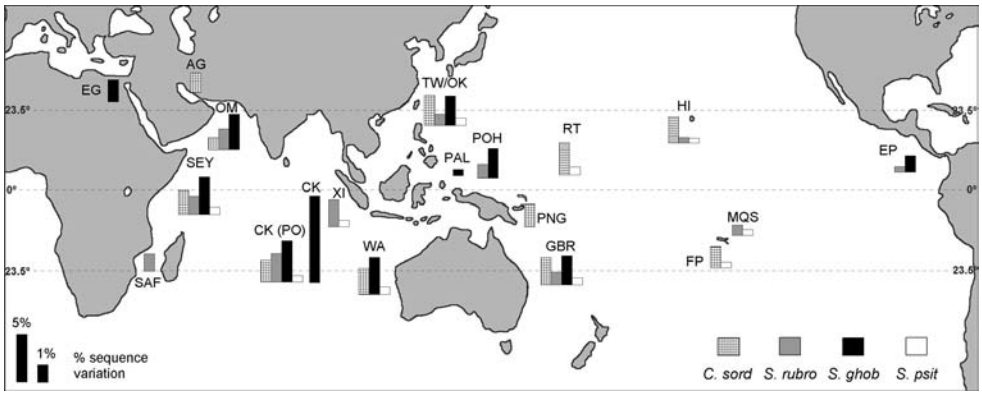


Figure 6.2 Nucleotide diversity (%) across 4 species throughout the Indo–Pacific

AG = Arabian Gulf; CK (PO) = Cocos Keeling Islands Pacific Ocean clade; CK = Cocos Keeling Islands exclusive clade; EG = Egypt; EP = east Pacific; FP = French Polynesia; GBR = Great Barrier Reef; HI = Hawaii; MQS = Marquesas; OM = Oman; PAL = Palau; PNG = Papua New Guinea; POH = Pohnpei; RT = Rota; SAF = South Africa; SEY = Seychelles; TW/OK = Taiwan/Okinawa; WA = Western Australia; XI = Christmas Island
Note: For locations where $n < 2$, sequence variation has not been reported

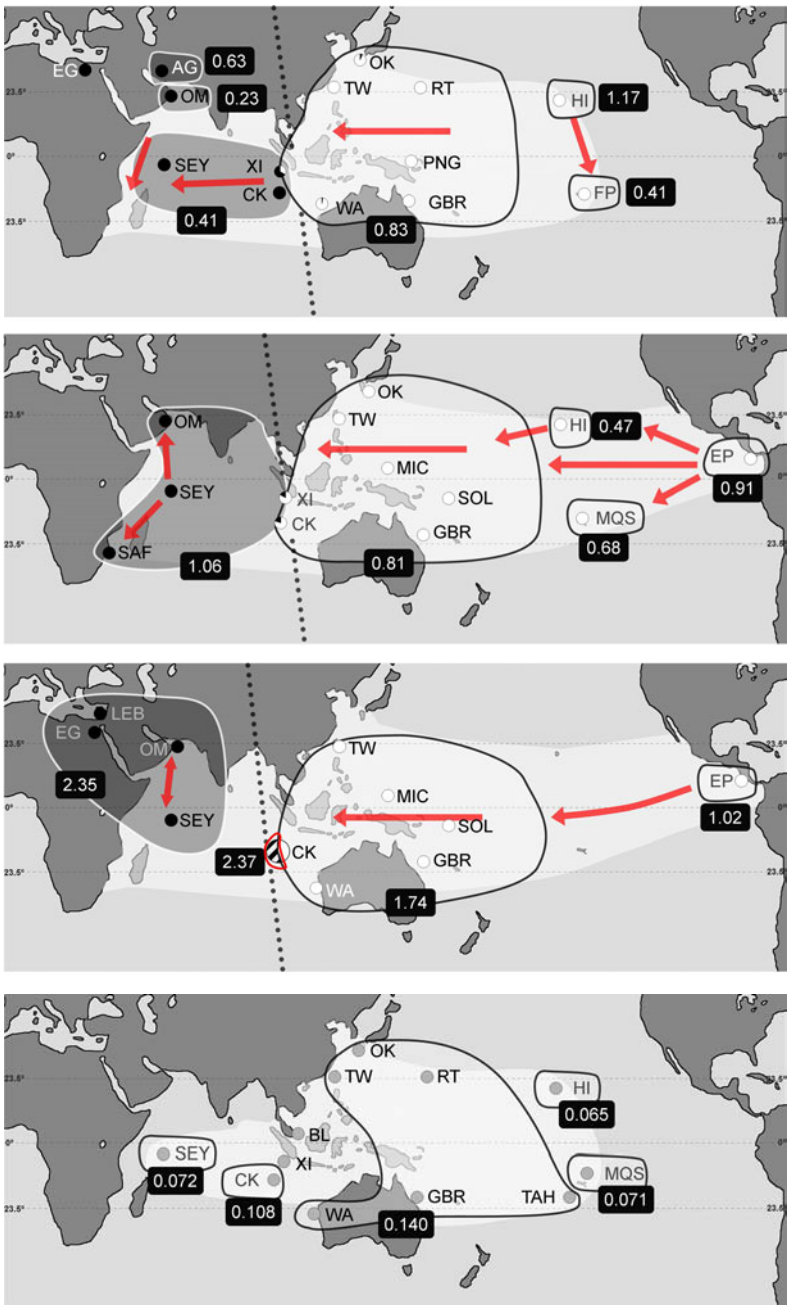
Migrate analyses (Beerli, Felsenstein, 1999) revealed largely uneven gene flow, particularly among central and eastern Pacific Ocean individuals. Among the central Pacific stock locations, the overwhelming direction of migration was to the west (Figures 6.3a–c). The *S. psittacus* data did not converge and thus Migrate results for this species were not available for this comparison. In the west Indo–Pacific populations, migration was south along the African coast and west from Christmas Island/Cocos Keeling Islands in *C. sordidus*, north and south from the Seychelles for *S. rubroviolaceus*, and bidirectional between Oman and the Seychelles for *S. ghobban* (Figure 6.3a–c).

6.2.3 Coalescence analysis

Coalescence analyses indicated fairly recent expansions for all four species, with the Hawaii stock of *S. psittacus* being the most recently expanded (0.065 mya, Figure 6.3d), and *S. ghobban* Cocos Keeling Islands and west Indo–Pacific stocks being the oldest (2.37 mya and 2.35 mya, respectively, Figure 6.3c). The oldest populations of *C. sordidus* were from Hawaii, whereas the oldest *S. rubroviolaceus* and *S. ghobban* populations were from the west Indo–Pacific (Figure 6.3a–d). For *S. psittacus*, the central stock had the oldest expansion time. Location-specific coalescence calculations also indicated that the oldest populations were in the east for *C. sordidus*, and in the west for *S. rubroviolaceus* and *S. ghobban*. The coalescence times in this study were well within the range of estimates for divergence in previous studies (Alfaro *et al.*, 2009; Smith *et al.*, 2008); however, *S. psittacus* appeared to be one of the oldest species based on evolutionary nodal age, even though its extant populations were the most recently coalesced of the four species (Figure 6.3d).

6.3 Discussion

There are four major conclusions from our four species comparison. Firstly, three of four species in this study exhibit strong population structure between western and central/eastern Indo–Pacific locations, and where a genetic break exists, it occurs on the west side of the central Indo–Pacific marine biogeographic realm (*sensu* Spalding *et al*., 2007). Secondly, all four aspects of genealogical concordance (sensu Avise 1990, 2000) were demonstrated for three of four species in this study. Thirdly, there is evidence of population structure and reduced genetic diversity at locations on either edge of the longitudinal ranges, as well as at isolated oceanic islands, including Christmas and Cocos Keeling Islands, Hawaii, the Marquesas and French Polynesia. Finally, the coalescent history of extant populations with deep divergences and shallow genetic architectures gives us a new perspective on diversification in widely distributed reef fish and what processes are involved in emergent diversification.



(a) *C. sordidus*

(b) *S. rubroviolaceus*

(c) *S. ghobban*

(d) *S. psittacus*

Figure 6.3(a-d) Coalescence times in millions of years and major migration routes (arrows) between stocks (grey and white shaded areas) for (a) *C. sordidus*, (b) *S. rubroviolaceus*, c) *S. ghobban*, and d) *S. psittacus*

West Indo-Pacific stocks are shaded in grey with white border, and East Indo-Pacific stocks are shaded in white with grey border. Stocks for *S. psittacus* are all shaded in one colour. Stock assignment is based on pairwise F_{ST} comparisons. Red arrows indicate major migration routes between and/or within different stocks based on MIGRATE analyses conducted in MIGRATE 2.3.

Table 6.2a–d Pairwise F_{ST} comparisons (bottom diagonal), within population haplotype diversity (along diagonal, in black) and significance values for pairwise comparisons (above diagonal) for (a) *Chlorurus sordidus*, (b) *Scarus rubroviolaceus*, (c) *S. ghobban* and (d) *S. psittacus*

(a) *C. sordidus* (**$P < 0.00042$)

Location	AG	OM	SEY-m	SEY-f	SEY-a	CK	WA-o	WA-i	GBR-l	GBR-w	OKI	ROTA	PNG	HAW	FP
AG	0.986	**	**	**	**	**	**	**	**	**	**	**	**	**	**
OM	**0.456**	0.450	**	**	**	**	**	**	**	**	**	**	**	**	**
SEY-m	**0.462**	**0.342**	0.971	–	*	–	**	**	**	**	**	**	**	**	**
SEY-f	**0.430**	**0.223**	0.030	1.000	–	*	**	**	**	**	**	**	**	**	**
SEY-a	**0.417**	**0.215**	0.050	0.035	1.000	*	**	**	**	**	**	**	**	**	**
CK	**0.482**	**0.353**	0.034	0.084	0.074	0.968	**	**	**	**	**	**	**	**	**
WA-o	**0.710**	**0.682**	**0.674**	**0.677**	**0.614**	**0.663**	0.993	–	*	–	–	–	*	**	**
WA-i	**0.733**	**0.708**	**0.703**	**0.705**	**0.649**	**0.692**	0.013	0.991	–	–	–	*	*	**	*
GBR-l	**0.727**	**0.706**	**0.691**	**0.695**	**0.628**	**0.679**	0.032	0.016	0.989	–	–	–	*	*	*
GBR-w	**0.752**	**0.755**	**0.724**	**0.727**	**0.649**	**0.709**	–0.009	–0.011	0.007	0.991	–	–	*	*	*
OKI	**0.715**	**0.698**	**0.674**	**0.679**	**0.602**	**0.663**	–0.014	–0.007	0.006	–0.020	0.950	–	–	*	–
ROTA	**0.704**	**0.669**	**0.657**	**0.664**	**0.594**	**0.652**	0.029	0.028	0.030	0.016	–0.003	0.990	*	*	*
PNG	**0.753**	**0.756**	**0.727**	**0.728**	**0.657**	**0.711**	0.083	0.043	0.054	0.074	0.027	0.085	1.000	**	–
HAW	**0.741**	**0.720**	**0.703**	**0.707**	**0.644**	**0.694**	**0.211**	**0.241**	0.204	0.211	0.167	0.096	**0.275**	0.937	**
FP	**0.762**	**0.756**	**0.737**	**0.737**	**0.675**	**0.720**	**0.102**	0.071	0.067	0.100	0.035	0.106	0.003	**0.286**	0.969

– = lack of significance ($P > 0.05$); * = significance at $P < 0.05$; ** and bold values = significance $P <$ Bonferroni-corrected α-value;
AG = Arabian Gulf; CK = Cocos Keeling Islands; FP = French Polynesia; GBR-l = Lizard Island, Great Barrier Reef; GBR-w = Whitsundays, Great Barrier Reef; HAW = Hawaii; OKI = Okinawa; OM = Oman; PNG = Papua New Guinea; SEY-a = Amirante, Seychelles; SEY-f = Farquhar, Seychelles; SEY-m = Mahe, Seychelles; WA-i = Beacon and Abrolhos islands, Western Australia; WA-o = Scott and Clerk reefs, Western Australia;
Note: The Bonferroni correction value is shown

(b) *S. rubroviolaceus* (**$P < 0.00056$)

Location	OM	SEY	SAF	X-IO	X-PO	CK	OK-TW	MIC	SOL	GBR	HAW	MQS	EP
OM	0.993	–	–	–	**	**	**	**	**	**	**	**	**
SEY	0.0014	0.997	–	–	**	**	**	**	**	**	**	**	**
SAF	–0.0016	–0.0073	0.995	-	**	**	**	**	**	**	**	**	**
X-IO	–0.0153	–0.0020	-0.0098	*	**	**	**	**	**	**	**	**	**
X-PO	**0.7297**	**0.7377**	**0.7508**	**0.7570**	*	–	–	–	–	–	**	–	–
CK	**0.5729**	**0.6313**	**0.6013**	**0.5290**	0.0634	1.000	–	–	–	–	**	–	–
OK-TW	**0.7389**	**0.7419**	**0.7574**	**0.7650**	–0.0067	0.0875	0.955	–	–	–	**	–	–
MIC	**0.7142**	**0.7311**	**0.7363**	**0.7353**	0.1022	0.1078	0.0945	0.947	–	–	**	**	–
SOL	**0.6997**	**0.7262**	**0.7270**	**0.7196**	0.0821	0.1101	0.0872	0.0245	1.000	–	**	–	–
GBR	**0.7074**	**0.7279**	**0.7342**	**0.7327**	–0.0162	0.0189	0.0036	0.1084	0.1037	0.967	**	–	–
HAW	**0.8006**	**0.7846**	**0.8236**	**0.8528**	**0.2289**	**0.4149**	**0.1880**	**0.3053**	**0.2527**	**0.3261**	0.849	**	**
MQS	**0.7149**	**0.7370**	**0.7476**	**0.7533**	0.1249	0.0373	0.1465	**0.2408**	**0.2593**	0.0660	**0.5396**	0.956	**
EP	**0.7392**	**0.7414**	**0.7710**	**0.8069**	0.1278	0.2265	0.1151	**0.1971**	**0.2368**	**0.2019**	**0.3493**	**0.3949**	0.517

– = lack of significance ($P > 0.05$); * = significance at $P < 0.05$; ** and bold values = significance $P <$ Bonferroni-corrected α-value; CK = Cocos Keeling Islands; EP = east Pacific; GBR = Great Barrier Reef; HAW = Hawaii; MIC = Micronesia; MQS = Marquesas; OK-TW = Okinawa/Taiwan; OM = Oman; SAF = South Africa; SEY = Seychelles; X-IO = Christmas Island, Indian Ocean; X-PO = Christmas Island, Pacific Ocean
Note: The Bonferroni correction value is shown

(c) *S. ghobban* (**$P < 0.00065$)

Location	OM	SEY	CK-CK	CK-PO	WA-R	WA-A	TW	MIC1	MIC2	SOL	GBR	EP
OM	0.99	–	**	**	**	**	**	**	**	**	**	**
SEY	–0.006	0.99	**	**	**	**	**	**	**	**	**	**
CK-CK	**0.762**	**0.754**	1.00	*	**	**	**	*	**	**	**	*
CK-PO	**0.751**	**0.748**	0.755	1.00	–	–	–	*	–	*	–	–
WA-R	**0.755**	**0.753**	**0.768**	–0.035	1.00	–	**	**	*	**	*	*
WA-A	**0.753**	**0.750**	**0.759**	0.003	–0.020	1.00	*	**	*	**	*	*
TW	**0.783**	**0.778**	**0.809**	0.053	**0.095**	**0.113**	1.00	–	–	–	–	**
MIC1	**0.805**	**0.794**	**0.888**	0.311	**0.215**	**0.251**	0.034	0.79	–	–	*	**
MIC2	**0.777**	**0.771**	**0.806**	0.086	**0.087**	0.085	–0.006	0.057	0.99	–	–	**
SOL	**0.803**	**0.794**	**0.847**	0.124	**0.127**	**0.160**	–0.008	0.060	0.017	0.99	–	**
GBR	**0.779**	**0.773**	**0.810**	0.092	**0.099**	0.105	0.000	0.093	–0.021	0.005	1.00	**
EP	**0.789**	**0.781**	**0.851**	0.047	0.124	0.227	**0.228**	**0.561**	**0.295**	**0.292**	**0.297**	0.79

– = lack of significance ($P > 0.05$); * = significance at $P < 0.05$; ** and bold values = significance $P <$ Bonferroni-corrected α-value; CK-CK = Cocos Keeling Islands, Cocos Keeling clade; CK-PO = Cocos Keeling Islands, Pacific Ocean clade; EP = east Pacific; GBR = Great Barrier Reef; MIC1 = Palau, Micronesia; MIC2 = Pohnpei, Micronesia; OM = Oman; SEY = Seychelles; SOL = Solomon Islands; TW = Taiwan; WA-A = Abrolhos Island, Western Australia; WA-R = Rottnest Island, Western Australia
Note: The Bonferroni correction value is shown

(d) *S. psittacus* (**$P < 0.0014$)

Loc	SEY	CKX	WA	TW	ROTA	GBR	HAW	TAH	MQS
SEY	0.93	**	*	**	**	**	**	**	**
CKX	**0.116**	0.98	*	**	**	**	**	**	**
WA	0.101	0.075	0.93	–	–	–	*	–	**
TW	**0.128**	**0.092**	–0.025	0.91	–	–	*	–	**
ROTA	**0.153**	**0.112**	–0.016	–0.035	0.95	*	*	–	**
GBR	**0.092**	**0.064**	0.025	0.027	0.049	0.93	–	–	**
HAW	**0.091**	**0.079**	0.061	0.090	0.083	0.025	0.87	*	**
TAH	**0.124**	**0.095**	–0.016	–0.029	–0.001	0.022	0.062	0.93	**
MQS	**0.166**	**0.233**	**0.276**	**0.276**	**0.300**	**0.240**	**0.283**	**0.292**	0.83

– = lack of significance ($P > 0.05$); * = significance at $P < 0.05$; ** and bold values = significance $P <$ Bonferroni-corrected α-value; CKX = Cocos Keeling and Christmas islands; EP = east Pacific; GBR = Great Barrier Reef; HAW = Hawaii; MQS = Marquesas; OM = Oman; SEY = Seychelles; SOL = Solomon Islands; TAH = Tahiti; TW = Taiwan; WA = Western Australia
Note: The Bonferroni correction value is shown

6.3.1 Congruent patterns of population structure

For three of the four species, there was a genetic break between west Indo–Pacific and central/east Indo–Pacific populations, and this is confirmed by both phylogenetic and population genetic analyses. With the exception of *S. psittacus*, the individuals sampled in this study exhibit a genetic break in the Cocos Keeling and Christmas Island region of the central Indo–Pacific, which is a known transitional area between west Indo–Pacific and central/east Indo–Pacific marine fauna (Hobbs *et al*., 2009). This transition can be clearly seen in Figure 6.1, and has been observed in a number of other marine taxa, including reef fishes (e.g., Craig *et al.*, 2007; Marie *et al*., 2007; McMillan, Palumbi, 1995; Timm, Kochzius, 2008) and several marine invertebrates (e.g., Williams, Benzie, 1998; Williams *et al.*, 2002). Further sampling from the Indo–Australian Archipelago may reveal where this break extends further to the north, as has been seen in several stomatopod species (Barber *et al*., 2006). Nevertheless, the combination of the geographic position at the Indo–Pacific crossroads, and the unique geomorphology and turbulent palaeohistory of the area surrounding Christmas and Cocos Keeling islands are consistent with this pattern of transitional fauna (Hobbs *et al*., 2009).

Christmas and Cocos Keeling islands are situated in the Wharton Basin, an extremely dynamic area with a history of earthquakes, cold water upwelling and cyclones. This area also experienced a period of intense volcanic activity between 3 and 5 million years ago (Brewer *et al.*, 2009; Woodroffe, Berry, 1994). The combination of features at this location, particularly over evolutionary timescales, would have made this area an effective barrier to dispersal between west Indo–Pacific and central/west Indo–Pacific populations at various points in history. Interestingly, despite the proximity of the Cocos Keeling Islands and Christmas Island, they do not necessarily harbour the

same marine communities. The difference in marine fauna between these two locations can be explained by their differing geomorphology and palaeohistory. While the Cocos Keeling Islands are atolls that emerged 4000 years ago and are subject to cycles of uplift and submergence, Christmas Island is an uplifted limestone island in an area prone to subduction and uplift (Woodroffe, Berry, 1994; Woodroffe *et al.*, 1991). The Cocos Keeling Islands are a more dynamic habitat, and host more fractured populations than Christmas Island. Furthermore, the Cocos Keeling Islands are separated from Christmas Island by a large 1800 km ridge, Investigator Ridge, the largest underwater volcanic plateau in the oceans (Werner, 2008). Similar differences in marine biota among closely situated reef communities have also been noted in French Polynesia (Gaither *et al*., 2010; Winters *et al*., 2010) and have been observed at microscales among intertidal limpets (Johnson, Black, 1982).

Although there was no defined genetic break in *S. psittacus*, a point which will be discussed later in this chapter, the highest nucleotide diversity for this species was recorded at the Cocos Keeling Islands, but only when the two Christmas Island samples were included (π = 1.09%). This suggests that more sampling from the central Indo–Pacific may reveal a stronger pattern of population structure between west Indo–Pacific and central/east Indo–Pacific populations. Similar lack of concordance in phylogeographic patterns has also been reported for two closely related, co-distributed gastropods *Nerita albicilla* and *N. plicata* (Crandall *et al*., 2007). Crandall *et al*. (2007) hypothesised that minute differences in ecology (e.g. pelagic larval duration) or adult habitat preference may have a large impact on patterns of population structure. Fauvelot *et al*. (2003) provided evidence for reduced genetic diversity in several lagoonal species of butterflyfish and damselfish compared to those inhabiting reef slopes; the authors attributed this lowered diversity to reduced shallow water habitat following Holocene low sea levels during glacial cycles. Arenas *et al* (2011) provided evidence through simulations that range contractions decreased genetic diversity, and that population structure following climate change was dependent on dispersal ability as well as the rate of that change. Given the slight differences in adult habitat preference between the four species in this study, the hypotheses put forth by Crandall *et al*. (2007) and Fauvelot *et al*. (2003) provide a possible explanation for the lack of concordance seen in *S. psittacus*. Of the four species, *S. psittacus* has the shortest pelagic larval duration (Choat pers. comm.), as well as the shallowest habitat preference, making it the most susceptible to sea level changes compared to the other three study species. In addition to congruent patterns of population structure on the scale of the entire Indo–Pacific, congruent patterns were also found on a smaller spatial scale, with peripheral populations yielding interesting patterns among the four study species.

6.3.2 Peripheral populations and isolated oceanic islands

Genetic diversity was consistently lower at the edges of the longitudinal distribution for all species, although which edge was affected in this manner varied between species. While *C. sordidus* exhibited low genetic diversity at the western periphery of its range at Oman, the three remaining species exhibited this pattern at the eastern edge of their ranges. *S. rubroviolaceus* showed lowest diversity at Hawaii, the Marquesas and the east Pacific; *S. ghobban* showed lowest diversity in the east Pacific; and *S. psittacus* exhibited the lowest diversity at Hawaii and the Marquesas. Interestingly, *S. ghobban* from the tropical east Pacific appears to be the most recently expanded population (1.02 mya), and while morphologically differentiated from its sister species *S. compressus*, its mitochondrial DNA is more similar to its tropical east Pacific sister than to other conspecifics from the main central Pacific Ocean stock (Chapter 5). Although this observation is based on only a small number of samples and a single marker, the taxonomy of *S. compressus* has been a subject of debate in the past, with some researchers believing it should be synonymous with *S. ghobban* due to their similar initial phase morphologies (Rosenblatt, Hobson, 1969; Schultz, 1958). The case of *S. compressus* and *S. ghobban* may be an example of either peripatric or parapatric speciation, in which populations have diverged, but the result is two sister species occupying two adjacent but different niches. While it is unlikely that the Eastern Pacific Barrier would have restricted dispersal long enough for speciation to occur in such a widely distributed fish, simulations by Gavrilets *et al*. (2000) found that rapid speciation could still take place with as many as 8 migrants per subpopulation per generation, which is consistent with the numbers of *S. ghobban* migrants between the east Pacific and other Pacific Ocean locations (Chapter 5). In addition, Lessios and Robertson (2006) found that the East Pacific Barrier significantly limited gene flow in a number of transpacific fishes, including *S. ghobban*, at different times in evolutionary history, and attributed such patterns to dispersal rather than vicariance. Malay and Paulay (2009) found a similar lack of sequence divergence despite distinctive colour patterns among hermit crabs, and proposed that either introgression was taking place or that the evolution of colour patterns may be more rapid than sequence divergence. The latter explanation is consistent with observations of different colour patterns between west Indo–Pacific and central/east Indo–Pacific individuals in both *C. sordidus* and *S. rubroviolaceus* (Bruce, Randall, 1983; Parenti, Randall, 2000).

Because peripheral populations are less stable than central populations (Hardie, Hutchings, 2010), greater selection pressures at these locations would likely result in adaptive divergence and ultimately, speciation (Losos, Glor, 2003). A study by Edmunds *et al*. (2010) revealed that populations of the tropical coastal fish *Lates calcarifer* from different thermal environments have different swimming abilities following acclimation to common thermal stresses. The peripheral areas in the present study that are associated with patterns of lowered genetic diversity are all situated near

seamounts. Thus, the cold water upwelling that is typically found at seamounts may provide the adaptive stress needed for such divergence that leads to speciation. Other explanations for the lack of neutral genetic diversity observed at peripheral locations include selective sweeps associated with thermal adaptation, or genetic drift acting on small founding populations. Another theory is that sexual selection may act to create a different genetic signal at peripheral locations, particularly in the case of *S. compressus*, which appears to be genetically indistinguishable from its sister species *S. ghobban* from the tropical east Pacific when mtDNA of populations of the two species are compared. Such a pattern has been observed in other reef fishes, including butterflyfishes (McMillan, Palumbi, 1995; Montanari *et al.*, 2012) and serranid fishes (McCartney et al., 2003; van Herweden et al., 2006), where introgressive hybridization has been identified between closely related species. It is therefore possible that the two sister species, *S. ghobban* and *S. compressus*, are hybridising in the tropical east Pacific. There is clearly some ambiguity surrounding these two sister taxa, and further work incorporating a more complete sampling scheme and additional markers is needed to resolve this.

Scarus rubroviolaceus also has recently diverged sister taxa, *S. ferrugineus* and *S. persicus*, both of which have narrow ranges that overlap with *S. rubroviolaceus* at the western periphery of its distribution. Interestingly, even though the split between *S. ghobban* and *S. compressus* occurred more recently (0.150–0.180 mya) (Choat et al., 2012) than the split between *S. rubroviolaceus* and its west Indo–Pacific sisters, the morphological differences between *S. ghobban* and *S. compressus* are more pronounced (Rosenblatt, Hobson, 1969). This reinforces the point that peripheral populations may be subject to greater selective pressures, or are more prone to the effects of genetic drift compared with central populations. Therefore, it is reasonable to assume that peripheral habitats are likely to be areas that generate biodiversity in widely distributed taxa and that the signals of genetic differentiation may be detected at peripheral populations earlier than in central populations. A recent study on reef corals suggested that species-edge zones may serve as cradles of evolution and should thus be considered a conservation priority (Budd, Pandolfi, 2010).

In three of the four species in this comparison, the coalescent age of the peripheral populations is older than that of central populations, which is consistent with the predictions for the Indo–Australian Archipelago as the ‘centre of overlap’ model (Barber, Bellwood, 2005), while coalescence analyses for *S. psittacus* indicate an Indo–Australian Archipelago ‘centre of origin’ model (Briggs, 1992; McManus, 1985). However, the large discrepancy between evolutionary and coalescent ages suggests that the central region served not as an origin but as a refuge from which the other populations were re-established after a severe bottleneck that effectively reset its evolutionary clock. Alternately, if *S. psittacus* maintained high levels of gene flow, we would expect a reduced coalescence compared with lineages that underwent a vicariance. Given the similar

natural histories and habitat preferences between *S. psittacus* and the other species in this comparison, however, it is unlikely that *S. psittacus* would have been able to maintain such high levels of gene flow when three other similar species could/did not, particularly when *S. psittacus* has the strongest degree of reef association compared to the other three species. A more likely alternative is that the loss of shallow water habitat following sea level changes would have meant a much more pronounced range contraction for *S. psittacus* than for the other species. Reduced genetic diversity resulting from a reduction in shallow water habitat post sea level change has been observed in other species of fish (Fauvelot *et al*., 2003). In addition, Arenas *et al* (2011) showed that range contractions decreased genetic diversity, and that slow contractions in particular resulted in smaller F-statistics, trees with shallow branches and thus shorter coalescence times. Avise (2000) states that small, unstructured ancestral populations can lead to shallower gene trees in daughter populations. Our findings are consistent with these observations, with the most specialised of the species, *S. psittacus*, exhibiting the shallowest coalescence, and the least specialised, *S ghobban*, exhibiting the greatest differentiation as well as the deepest coalescence. The juxtaposition of these contrasting patterns suggests that population structure goes through cycles over evolutionary time, with ongoing ebbing and flowing levels of genetic variability responding to the ever fluctuating climatic conditions. Additionally, the contrasting patterns observed here demonstrate that seemingly minute differences in life history and habitat preference will have dramatic impacts on how species respond to climate change.

6.3.3 Genealogical concordance

The first three species in this comparison, *C. sordidus*, *S. rubroviolaceus*, and *S. ghobban* demonstrated all four aspects of genealogical concordance. Each of these three species exhibits strong genetic partitioning into two distinct phylogroups, or clades. The concordance of these phylogenetic patterns is an indication that these "historical population units" are worth examining more closely in a biogeographic context. The lack of genealogical concordance in *S. psittacus* will be discussed later in this chapter. Similar patterns separating west Indian Ocean and east Indo-Pacific lineages have been recovered using nuclear markers in *S.rubroviolaceus* (Fitzpatrick et al., 2011), and *S. ghobban* (Bariche, Bernardi, 2009), as well as in preliminary findings by Robinson (unpublished) in *C. sordidus*. The fact that these three species are separated into two clades corresponding to the same geographic regions (west Indian Ocean and east Indo-Pacific) (Aspect III), and that between these two geographic regions is an important biodiversity hotspot (Aspect IV), provides evidence of shared historical evolutionary processes with strong biogeographic influences around Christmas Island and the Cocos Keeling Islands. Although these islands are not formally part of the biogeographic hotspot known as the Indo-Australian archipelago (IAA), its implication in the congruent patterns of population structure in three of the four species in this study provide strong

support that it deserves conservation attention, and perhaps, to be formally included in the region known as the IAA. These islands have recently been recognised as an important biodiversity hotspot (Spalding et al., 2007; Veron et al., 2009), and newly coined terms to describe them include the "marine hybrid hotspot at the Indo-Pacific biogeographic border" (Hobbs et al., 2009), and the "West Pacific diversity hotspot" (Fitzpatrick et al., 2011). Thus, their exclusion from the IAA is most likely due to the fact that they are less well known (Veron et al., 2009) rather than for their lack of species. Furthermore, the growing number of studies highlighting this area as an important contact/hybridization zone between West Indian Ocean and Pacific Ocean fauna (Hobbs *et al.*, 2009; Hobbs, Salmond, 2008; Montanari *et al.*, 2012) further suggests that this region should be included in the IAA.

6.3.4 Relationship between nodal age and the coalescent

The absence of strong population structure and genealogical concordance observed in *S. psittacus* at the broadest spatial scale initially seems to contradict the straightforward patterns observed in the other three species. However, a closer look at the coalescent history of *S. psittacus* reveals that only a small part of its evolutionary history has been retained (Winters *et al.*, 2010). Moreover, many of the stocks identified in *S. psittacus* are very similar to those found in the other three species, with the exception of the west Indo–Pacific, which was not sampled as thoroughly. In addition, the relative coalescence times of those stocks in *S. psittacus* are younger by nearly an order of magnitude than those of the other three species, despite all four species having diverged around the same time (2–4 mya) (Alfaro *et al.*, 2009). The genetic signature of *S. psittacus* is that of a much more recent species, with the oldest stock having a coalescence time of 0.14 million years (Winters *et al.*, 2010). This suggests that *S. psittacus* has undergone a much more recent bottleneck event and is representative of a species that is starting over. Thus, the pattern observed in *S. psittacus* is what we would predict for a widespread species in the formative stages of evolution, and is characterised by very little overall genetic diversity. What population structure does exist for this species appears to be at the same locations that generated the highest levels of population structure in the other three species—isolated oceanic islands, in this case, the Marquesas. This leads us to the reasonable conclusion that given enough time, those isolated and peripheral populations will become more and more differentiated, and the overall pattern will resemble that of the other three species. Alternatively, as mentioned in the previous sections, the range contractions and the rate at which they occurred may have affected the resulting population sizes and thus the resulting coalescence times. It is widely accepted that range contractions result in recent coalescent events due to small population sizes (e.g. Nei et al., 1975).

In contrast, *S. ghobban* exhibits much higher genetic as well as morphological differentiation between west Indo–Pacific and central/east Indo–Pacific individuals. In fact, a third clade consisting of 50% of all individuals from the Cocos Keeling Islands was also found, and individuals from this third clade were morphologically indistinguishable from Pacific Ocean congeners (Choat, pers. comm.). Interestingly, this third clade was more genetically differentiated than the other two clades, and is most likely an example of cryptic speciation involving a complex of three incipient species. Although only a small number of samples were obtained from the Cocos Keeling Islands for this species ($n < 10$), it is unlikely that more sampling would alter this finding. *S. ghobban* is one of the more resilient members of the family Scarinae, and has been known to use a number of atypical environments for a reef fish—it has been found at depths of up to 250 m in video surveillance (Rees *et al*., 1994), and has also invaded the Mediterranean Sea (Bariche, Saad, 2005). Its ability not only to colonise but thrive in nonreefal habitats may be one of the reasons for its distinctly different evolutionary history to *S. psittacus*, despite their similar evolutionary ages (Alfaro *et al*., 2009). The processes that effectively reset *S. psittacus* populations clearly did not have the same dire effect on the similarly distributed *S. ghobban*. However, the fact that all four species in this comparison have the same basic stock structure centred on the same biogeographic realms suggests that they are all affected by the same sets of processes across their distribution ranges over long periods of time. Furthermore, the contrasting levels of phylogenetic structure seen in this study are a reflection of the different species being at a different point in the cycle of population structure; *S. psittacus* is most likely at the beginning of the cycle while *S. ghobban* is at the end, with its newly diverged tropical east Pacific sister, *S. compressus*, starting anew.

Another congeneric phylogeographic comparison between three species of surgeonfishes with similar Indo–Pacific distributions found almost no evidence of geographic partitioning, even at the largest spatial scale; however, distinct temporal clades were observed in two of the three species (Horne *et al*., 2008). This may seem to contradict our results, but given the fact that the nasiid surgeonfishes are far less dependent on coral reef habitats relative to the scarine labrids in this study, along with their superior larval swimming ability, longer generation times, and ancient evolutionary and coalescent histories (with clades ranging from 1.2–15.4 my old) (Horne *et al.*, 2008; Klanten *et al.*, 2007), it is no surprise that the patterns are so different. Interestingly, despite the greater dispersive capacity of the surgeonfishes, the East Pacific Barrier is more of a significant barrier for *Acanthurus* species than it is for the parrotfishes in this study. A 20-species comparison of transpacific fishes on either side of the East Pacific Barrier demonstrated significant genetic differentiation between central Indo–Pacific and tropical east Pacific populations in 18 of the 20 species, and evidence of cryptic speciation in the two remaining species (Lessios, Robertson, 2006). Lessios and Robertson (2006) also concluded that dispersal, rather than a common historical event, was responsible for the transpacific patterns of distribution, and that for the species whose origins

could be determined, the direction of gene flow was opposite to the direction of original colonisation. For the three species whose migration patterns could be determined in Migrate (Beerli, Felsenstein, 1999), the majority of gene flow was from east to west, which is consistent with a Tethyan origin for this family (Bellwood, 1994; Streelman *et al*., 2002) and with observations regarding postcolonisation migration in other reef fish noted above (Lessios, Robertson, 2006). Kuhner (2009) states that caution should be exercised when interpreting results from what she calls "coalescent genealogy samplers", given the likelihood that key assumptions of programs such as MIGRATE will be violated due to the complex nature of biological data.

6.4 Conclusion

In conclusion, patterns of population structure in Indo–Pacific reef fish appear to be general within the broader spatial context among scarine labrids, but with some species-specific idiosyncrasies. Despite varying levels of population structure, each species comprised a main stock at the centre of its distribution and several smaller stocks at peripheral locations. Peripheral and isolated populations play an important role in the generation of marine biodiversity, evidenced by the presence of older peripheral lineages in all but *S. psittacus*, as well as newly diverged sister species that are confined to peripheral margins in two of the four species. The high degree of genealogical concordance in three of the four species examined in this thesis warrants several taxonomic reassessments, or assigning the different clades into ESUs or MUs. The older coalescent ages of peripheral lineages in three of the four species suggest that the Indo–Australian Archipelago is the centre of overlap. The absence of this pattern in *S. psittacus* can be explained by its very recent coalescent age relative to its evolutionary age, which in turn is most likely a function of its habitat specificity and moderate vagility compared with the other three species. Population structure fluctuates at regular intervals over evolutionary timescales. However, this study demonstrates that even closely related species are not fluctuating in synchrony; species at the beginning of the cycle will exhibit low genetic diversity, and those at the end will have higher diversity and possibly even newly diverged sister species. Phylogeographic studies often lack a phylogenetic aspect, but this study demonstrates the need for both in understanding patterns of population structure. Further studies of widely distributed sister taxa spanning multiple marine biogeographic realms would enable us to test the hypothesis of the cyclical nature of population structure in widespread marine taxa.

Chapter 7 General discussion

One of the major aims of this thesis was to test the hypothesis that peripheral habitats are responsible for the evolution of diversity in scarine labrids. The four species in this study have ranges that span much of the Indo–Pacific from the east coast of Africa to the central Pacific barrier (*C. sordidus* and *S. psittacus*) and further to the eastern periphery of the east Pacific Ocean (*S. ghobban* and *S. rubroviolaceus*). They are also found across a wide latitudinal range, with *S. ghobban* inhabiting numerous marginal habitats as far north as the Mediterranean Sea 35°N (Bariche, Saad, 2005) and as far south as Lord Howe Island off the New South Wales coast 31°33'S (Myers, 1999), making them ideal for a study of this nature. The subject of the origins of marine biodiversity is one of ongoing debate, however, results from this study provide four lines of evidence that support the theory that peripheral habitats are cradles of biodiversity in widely distributed scarine labrids.

First, all four species exhibited lowered genetic diversity and distinctive population structure at peripheral locations. Second, the coalescent ages of extant populations from peripheral locations were older than those of central populations for all but one species, *S. psittacus*. Coalescence estimates of extant populations of *S. psittacus* postdated the species nodal age by an order of magnitude, suggesting that an original source population may have come from elsewhere. Third, two of the four species in this study, *S. ghobban* and *S. rubroviolaceus*, have newly diverged sister species with narrowly sympatric distributions at the edges of their respective ranges. *S. compressus* is sister to *S. ghobban*, and is found only in the tropical east Pacific (Bruce, Randall, 1983). In the western Indian Ocean, *S. rubroviolaceus* has two recently diverged sister species: *S. persicus*, which is only found in northern Oman and the Persian Gulf (Bruce, Randall, 1983); and *S. ferrugineus*, which is confined to the Red Sea and the Gulf of Aden (Parenti, Randall, 2000). Finally, of these sister taxon pairs, *S. ghobban* from the tropical east Pacific showed greater similarity in mitochondrial DNA to its sister, *S. compressus*, than it did between congeners from other locations, despite obvious morphological differences between the terminal phases of these two species. In addition to these observations, there are many widely accepted perceptions regarding the differing evolutionary dynamics that exist at peripheral and marginal habitats compared with central habitats including:

a) increased selective pressures in marginal habitats at range edges where peripheral populations reside (Hardie, Hutchings, 2010; Johannesson, André, 2006)

b) increased rates of genetic divergence in small populations (such as those found at peripheral locations) under drift alone (Gaston, 1990)

c) more pronounced environmental fluctuations in peripheral/marginal habitats compared to central habitats (Lesica, Allendorf, 1995)

d) reduced dispersal to peripheral (and sometimes ecologically marginal) habitats result in lowered genetic diversity and reduced stability compared to central 'core' populations (Lesica, Allendorf, 1995).

My observations, in conjunction with these well-supported notions, led to the conclusion that diversification and extinction may occur more readily in peripheral than in central populations.

The importance of peripheral habitats in generating novel species has been documented in other reef fish (Rocha, Bowen, 2008), as well as reef corals (Budd, Pandolfi, 2010). In the present study, the absence of reciprocal monophyly at the mitochondrial DNA level between *S. ghobban* and its morphologically distinct tropical east Pacific sister *S. compressus* mirrors that found for some Indo–Pacific coral reef hermit crabs (Malay, Paulay, 2009). Malay and Paulay (2009) suggested that the evolution of colour pattern may, at times, exceed the rate of mitochondrial sequence divergence. However, in this study, introgression cannot be ruled out, and further studies on *S. ghobban* and *S. compressus* incorporating nuclear markers are warranted to resolve this relationship.

The lack of congruence between coalescent and nodal ages in *S. psittacus* (Winters et al., 2010) led to an important revelation about the differential efficacy of marine biogeographic barriers among widely distributed reef fish, as well as an alternative hypothesis for contrasting patterns of population structure in similar species. The same basic pattern of a main, centrally located stock accompanied by several smaller peripheral stocks could be seen across all four species. Further genetic differentiation was detected at the largest spatial scale, with distinctive populations on both sides of the Indo – West Pacific Barrier where it extends to the Cocos Keeling and Christmas islands in the east Indian Ocean. However, this feature was not evident in *S. psittacus*, which Winters et al. (2010) attributed to the recent coalescent history of extant populations (< 0.140 MY) compared to the evolutionary age of the species (4 MY, Alfaro et al., 2009).

I argue that the high level of apparent connectivity between *S. psittacus* individuals from the western Indian Ocean and the Pacific Ocean was most likely due to its habitat specificity which

would have facilitated smaller ancestral population sizes during times of low sea level stands. This would have created the genetic signal of a newer species, and given enough time and a large enough population size, similar patterns of population structure to the other three species is expected. Moreover, the absence of this defined break that was apparent in the other three species with much older coalescent ages suggests that the conditions or events (or both) that enabled genetic differentiation on either side of the Indo – West Pacific Barrier predate the more recent bottleneck event that effectively reset modern-day *S. psittacus* populations. The patterns of migration observed in *S. psittacus* are consistent with contemporary oceanographic currents, and are further supported by an expansion occurring over the last 120 000 years. My prediction was that processes, such as intensified climatic variations over large spatial and temporal (i.e. evolutionary) scales, would have generated genetic differentiation between parrotfish populations at range edges on both sides of the Indo – West Pacific Barrier, given that this has already been demonstrated for *S. psittacus* populations (Winters et al., 2010).

I hypothesised that widespread reef fish undergo cycles of population structuring characterised by ongoing fluctuating genetic variability over evolutionary timescales. Therefore, anomalous patterns of population structure in species with similar life-history traits and evolutionary ages may be the result of different species following this trajectory at different rates and, consequently, are at different points of that cycle at the time of comparison. I conclude that *S. psittacus* represents a species in the early stages of the most recent cycle, because it is characterised by overall low genetic diversity and comparatively little genetic differentiation between populations. In contrast, *S. ghobban* likely represents a species at a later stage of the cycle, because it is characterised by higher genetic diversity and possibly newly diverged sister taxa. The occurrence of discrete populations at peripheral locations for all species, including *S. psittacus*, suggests that despite being at a different point in the evolutionary cycle, all species in this comparison were affected by similar processes occurring at the edges of their distributional ranges. The effects of intrinsic and ecological differences between species cannot be disregarded, and that the degree of ecological specialisation may also play an important role in the patterns we see. For example, the process that caused the severe bottleneck in *S. psittacus* did not have the same effect on *S. ghobban* and this is likely due to its ability to colonise and use nonreefal environments, including temperate reefs and depths upwards of 250 m (Rees et al., 1994). Thus, the generalist parrotfish *S. ghobban* is less susceptible than the relatively specialist species *C. sordidus*, *S. rubroviolaceus* and *S. psittacus* during wide-scale climatic changes over evolutionary timescales. This was also suggested for Hawaiian butterflyfishes (Craig et al., 2010).

In addition to exhibiting strong population structure, as well as possessing genetically indistinguishable sister taxa based on mtDNA population genetics at the eastern periphery of its

range, I also found evidence of a third genetically distinct subclade in *S. ghobban* that was comprised of individuals exclusively from the Cocos (Keeling) Islands. No morphological distinction was detected between individuals from this third clade and those in the Pacific Ocean, but further, more detailed examination is called for. At the molecular level, however, there was a 15% (53 base pair) difference between haplotypes from the Cocos (Keeling) Islands subclade and those from the Pacific Ocean clade. This was similar to the difference between the Pacific Ocean clade haplotypes and the Indian Ocean clade haplotypes, which differed from one another by 14% (49 bp). The lack of distinguishable colour patterns in the three *S. ghobban* clades compared to its sister species, *S. compressus* from the east Pacific, suggests that sexual selection may be accelerating the evolution of colour pattern at peripheral locations and adds further support to peripheral habitats as cradles of biodiversity. In addition, the degree of genetic differentiation between the three clades in *S. ghobban* suggests that this putatively widespread species may actually comprise a series of cryptic species, as has been shown for the dascyllid species complex *Dascyllus trimaculatus* (Leray et al., 2010). Results from this study provide confirmation for the growing body of literature indicating that cryptic speciation may be far more common in marine taxa than previously acknowledged (Knowlton, 1993). Many management decisions are based on widespread species as the norm in the marine environment; however, increasing examples of cryptic species at range edges suggest that biodiversity may be significantly underestimated, and indicates that management policies should be updated to protect such edge populations.

Finally, I proposed *C. sordidus* to represent a species in the later stages of the evolutionary cycle. In the life of this doctoral thesis, *C. sordidus* has gone from a single, widely distributed Indo–Pacific species to being considered separate western Indian Ocean (*C. sordidus*) and Pacific Ocean (now *C. spilurus*) taxa (Randall, 2010). Based on the research presented here, as well as evidence from nuclear markers by Robinson et al. (unpublished) and previously noted morphological differences between western Indian Ocean and Pacific Ocean forms, I suggest that a formal taxonomic reappraisal of *C. sordidus* be undertaken. Furthermore, given the high degree of genealogical concordance across three of the four species, and considering that the base-pair differences between the three *S. ghobban* clades were three times greater than those between *C. sordidus* clades from the Indian Ocean and Pacific Ocean stocks (now considered to be separate species), more work is required to clarify the status of the *S. ghobban* lineages identified here. In conclusion, future comparative phylogeographic studies of closely related, widely distributed sister species are needed for more accurate estimates of marine biodiversity and to understand modes of speciation in this diverse assemblage more comprehensively. More work focusing on species from the Indo–Pacific, including the Cocos (Keeling) Islands and the west coast of Australia, as well as peripheral and isolated habitats, would provide valuable insight to this evolutionarily dynamic and important part of the world.

References

Alfaro ME, Brock CD, Banbury BL, Wainwright PC (2009) Does evolutionary innovation in pharyngeal jaws lead to rapid lineage diversification in labrid fishes? *BMC Evolutionary Biology* **9**, 255.

Allen GR (2007) Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems* **18**, 541–556.

Almany GR, Berumen ML, Thorrold S, Planes S, Jones G (2007) Local replenishment of coral reef fish populations in a marine reserve. *Science* **316**, 742-744.

Alvarado Bremer JR, Baker AJ, Mejuto J (1995) Mitochondrial DNA control region sequences indicate extensive mixing of swordfish (*Xiphias gladius*) populations in the Atlantic Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* **52**, 1720-1732.

Arenas M, Ray N, Currat M, Excoffier L (2011) Consequences of range contractions and range shifts on molecular diversity. *Molecular Biology and Evolution* **29**, 207-218.

Avise JC (2000) *Phylogeography: The history and formation of species.* Harvard University Press, Cambridge, MA.

Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, Neigel JE, Reeb CA, Saunders NC (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between populations genetics and systematics. *Annual Review of Ecology and Systematics* **18**, 489-522.

Avise JC, Ball J, R.M. (1990) Principles of genealogical concordance in species concepts and biological taxonomy. In: *Oxford Survyes in Evolutionary Biology* (eds. Futuyma D, Atonovics J), pp. 45-67. Oxford University Press, Oxford.

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

Barber PH, Bellwood DR (2005) Biodiversity hotspots: evolutionary origins of biodiversity in wrasses (*Halichoeres*: Labridae) in the Indo-Pacific and new world tropics. *Molecular Phylogenetics and Evolution* **35**, 235-253.

Barber PH, Erdmann MV, Palumbi SR (2006) Comparative phylogeography of three codistributed stomatopods: origins and timing of regional lineage diversification in the coral triangle. *Evolution* **60**, 1825-1839.

Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2000) A marine Wallace's line? *Nature* **406**, 692-693.

Bariche M, Bernardi G (2009) Lack of a genetic bottleneck in a recent Lessepsian bioinvader, the blue-barred parrotfish, *Scarus ghobban*. *Molecular Phylogenetics and Evolution* **53**, 592-595.

Bariche M, Saad M (2005) Settlement of the lessepsian blue-barred parrotfish *Scarus ghobban* (Teleostei: Scaridae) in the eastern Mediterranean. *Journal of the Marine Biological Association of the United Kingdom 2 - Biodiversity Records* **2005**, Published Online.

Bay L, Choat J, van Herwerden L, Robertson D (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.

Beerli P (2004) Effect of unsampled populations on the estimation population sizes and migration rates between sampled populations. *Molecular Ecology* **13**, 827 - 836.

Beerli P (2009) How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? . In: *Population Genetics for Animal Conservation* (eds. Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C), pp. 42-79. Cambridge University Press, Cambridge UK.

Beerli P, Felsenstein J (1999) Maximum likelihood estimation of a migration matrix and effective population numbers in two populations using a coalescent approach. *Genetics* **152**, 763-773.

Beerli P, Felsenstein J (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Science, USA* **98**, 4563-4568.

Bellwood DR (1994) A phylogenetic study of the Parrotfishes family Scaridae (Pisces: Labroidei), with a revision of genera. *Records of the Australian Museum* **Supplement 20**, 1-86.

Bellwood DR, Hughes TP (2001) Regional-scale assembly rules and biodiversity of coral reefs. *Science* **292**, 1532-1535.

Bellwood DR, Meyer CP (2009) Searching for heat in a marine biodiversity hotpost. *Journal of Biogeography* **36**, 569-576.

Bellwood DR, Wainwright PC (2002) The history and biogeography of fishes on coral reefs. In: *Coral reef fishes: dynamics and diversity in a complex ecosystem* (ed. Sale PF), pp. 5-32. Elsevier, San Diego, CA, USA.

Benzie JA (1999a) Genetic structure of coral reef organisms: ghosts of dispersal past. *American Zoologist* **39**, 131-145.

Benzie JA (1999b) Major genetic differences between crown-of-thorns starfish (*Acanthaster plancii*) populations in the Indian and Pacific Oceans. *Evolution* **53**, 1782-1795.

Bernardi G, Holbrook SJ, Schmitt RJ (2001) Gene flow at three spatial scales in a coral reef fish, the three-spot dascyllus, *Dascyllus trimaculatus*. *Marine Biology* **138**, 457-465.

Bernardi G, Holbrook SJ, Schmitt RJ (2003) Genetic evidence for two distinct clades in a French Polynesian population of the coral reef three-spot damselfish *Dascyllus trimaculatus*. *Marine Biology* **143**, 485-490.

Bernardi G, Holbrook SJ, Schmitt RJ, Crane NL, DeMartini E (2002) Species boundaries, populations and colour morphs in the coral reef three-spot damselfish (*Dascyllus trimaculatus*) species complex. *Proceedings of the Royal Society of London Series B: Biological Sciences* **269**, 599-605.

Bickford D, Lohman D, Sodhi N, Ng P, Meier R, Winker K, Ingram K, Das I (2007) Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22**, 148-155.

Brewer DT, Potter A, Skewes TD, Lyne V, Andersen J, Davies C, Taranto T, Heap AD, Murphy NE, Rochester WA, Fuller M, Donovan A (2009) Conservation values in Commonwealth waters of the Christmas and Cocos (Keeling) Islands remote Australian Territories. Report to Department of Environment and Water Resources. CSIRO, Cleveland, QLD.

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

Briggs JC (1999) Coincident biogeographic patterns: Indo-West Pacific Ocean. *Evolution* **53**, 326-335.

Briggs JC (2000) Centrifugal speciation and centres of origin. *Journal of Biogeography* **27**, 1183-1188.

Briggs JC (2005) The marine East Indies: diversity and speciation. *Journal of Biogeography* **32**, 1517-1522.

Brito PH, Edwards SV (2008) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* **135**, 439-455.

Bruce RW, Randall JE (1983) Scaridae. In: *FAO species identification sheets for fishery purposes. (Western Indian Ocean fishing area 51), Volume 3.* (eds. Fischer W, Bianchi G). FAO, Rome.

Budd AF, Pandolfi JM (2010) Coral Species Distributions Evolutionary. *Science* **328**, 1558-1561.

Carson HL, Templeton AR (1984) Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annual Review of Ecology and Systematics* **15**, 97-132.

Castelloe J, Templeton AR (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution* **3**, 102-113.

Chappell J, Shackleton N (1986) Oxygen isotopes and sea level. *Nature* **324**, 137-140.

Chen L (1999) *Ontogenetic development in post-settlement Scarids (Pisces: Scaridae): transition in feeding modes.* PhD Dissertation, James Cook University, Queensland.

Choat JH, Bellwood DR (1994) Wrasses and Parrotfishes. In: *Encyclopedia of fishes* (eds. Paxton JR, Eschmeyer WN), pp. 211-215. New South Wales Publishers, Sydney, NSW.

Choat JH, Klanten SO, van Herwerden L, Robertson DR, Clements KD (2012) Patterns and processes in the evolutionary history of parrotfishes (Family *Labridae*). *Biological Journal of the Linnean Society* **Article first published online : 5 SEP 2012**.

Clemens SC, Murray DW, Prell WL (1996) Nonstationary phase of the Plio-Pleistocene Asian Monsoon. *Science* **274**, 943-948.

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

Connolly SR, Bellwood DR, Hughes TP (2003) Indo-Pacific biodiversity of coral reefs: deviations from a mid-domain model. *Ecology* **84**, 2178-2190.

Craig MT, Eble JA, Bowen BW (2010) Origins, ages and population histories: comparative phylogeography of endemic Hawaiian butterflyfishes (genus *Chaetodon*). *Journal of Biogeography* **37**, 2125-2136.

Craig MT, Eble JA, Bowen BW, Robertson DR (2007) High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). *Marine Ecology Progress Series* **334**, 245-254.

Crandall ED, Frey MA, Grosberg RK, Barber PH (2007) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology* **17**, 611-626.

Cuvier GLCFD, Valeciennes A (1939) *Histoire naturelle des poissons.* Levrault, Paris.

Dawson MN, Hamner WM (2005) Rapid evolutionary radiation of marine zooplankton in peripheral environments. *Proceedings of the National Academy of Science, USA* **102**, 9235-9240.

Drew J, Allen GR, Kaufman L, Barber PH (2008) Endemism and regional color and genetic differences in five putatively cosmopolitan reef fishes. *Conservation Biology* **22**, 965-975.

Duda TF, Lee T (2009) Isolation and population divergence of a widespread Indo-West Pacific marine gastropod at Easter Island. *Marine Biology* **156**, 1193-1202.

Duda TF, Lessios HA (2009) Connectivity of populations within and between major biogeographic regions of the tropical Pacific in *Conus ebraeus*, a widespread marine gastropod. *Coral Reefs* **28**, 651-659.

Dudgeon CL, Gust N, Blair D (2000) No apparent genetic basis to demographic differences in scarid fishes across continental shelf of the Great Barrier Reef. *Marine Biology* **137**, 1059-1066.

Edmunds RC, van Herwerden L, Fulton CJ (2010) Population-specific locomotor phenotypes are displayed by barramundi, *Lates calcarifer*, in response to thermal stress. *Canadian Journal of Fisheries and Aquatic Sciences* **67**, 1068-1074.

Ekman S (1953) *Zoogeography of the sea.* Sidgwick and Jackson, London.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Excoffier L, Laval G, Schneider S (2006) Arlequin ver 3.01: An integrated software package for population genetics data analysis, Bern.

Fauvelot C, Bernardi G, Planes S (2003) Reductions in the mitochondrial DNA diversity of coral reef fish provide evidence of population bottlenecks resulting from Holocene sea-level change. *Evolution* **57**, 1571-1583.

Fitzpatrick J, Carlon DB, Lippe C, Robertson DR (2011) The West Pacific diversity hotspot as a source or sink for new species? Population genetic insights from the Indo-Pacific parrotfish *Scarus rubroviolaceus*. *Molecular Ecology* **20**, 219-234.

Fletcher C, Bochicchio C, Conger C, Engels M, Feirstein E, Frazer N, Glenn C, Grigg R, Grossman E, Harney J, Isoun E, Murray-Wallace C, Rooney J, Rubin K, Sherman C, Vitousek S (2008) Geology of Hawaii Reefs. In: *Coral Reefs of the USA* (eds. Riegl B, Dodge R), pp. 435-488. Springer, Berlin.

Forsskål P (1775) Descriptions animalium, avium, amphibiorum, piscium, insectorum, vermium; quae in itinere orientali observavit Petrus Forsskål, -Hauniae. Post mortem auctoris editdit Carsten Niebuhr.

Froukh T, Kochzius M (2008) Species boundaries and evolutionary lineages in the blue green damselfishes Chromis vitidis and Chromis atripectoralis (Pomacentridae). *Journal of Fish Biology* **72**, 451-457.

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

Funk D, Omland K (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology and Systematics* **34**, 397-423.

Gagliano M, McCormick M (2007) Maternal condition influences phenotypic selection on offspring. *Journal of Animal Ecology* **76**, 174-182.

Gaither MR, Bowen BW, Toonen RJ, Planes S, Messmer V, Earle J, Robertson DR (2010) Genetic consequences of introducing allopatric lineages of Bluestriped Snapper (*Lutjanus kasmira*) to Hawaii. *Molecular Ecology* **19**, 1107-1121.

Garcia LV (2003) Controlling the false discovery rate in ecological research. *Trends in Ecology & Evolution* **18**, 553-554.

Gaston KJ (1990) Patterns in the geographical ranges of species. *Biological Reviews Cambridge Philosophical Society* **65**, 105-129.

Gavrilets S, Li H, Vose MD (2000) Patterns of parapatric speciation. *Evolution* **54**, 1126-1134.

GeneCodes (2005) Sequencher v.4.5., Ann Arbor, MI.

Gopurenko ea (1999) Mitochondrial DNA evidence for rapid colnization of the Indo-West Pacific by the mudcrab *Scylla serata*. *Marine Biology* **134**, 227-233.

Goren M, Aronov A (2002) First record of the Indo-Pacific *Scarus ghobban* Forsskal, 1775, in the eastern Mediterranean. *Cybium* **27**, 239-240.

Grant WAS, Bowen BW (1998) Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity* **89**, 415-426.

Hardie DC, Hutchings A (2010) Evolutionary ecology at the extremes of species' ranges. *Environmental Review* **18**, 1-20.

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

Hayward B (2002) Late Pliocene to middle Pleistocene extinctions of deep-sea benthic foraminifera ("*Stilostomella* Extinction") in the Southwest Pacific. *Journal of Foraminiferal Research* **32**, 274–307.

Hickerson MJ, Cunningham CW (2005) Contrasting Quaternary Histories In An Ecologically Divergent Sister Pair Of Low-Dispersing Intertidal Fish (Xiphister) RevealedBy Multilocus Dna Analysis. *Evolution* **59**, 344-360.

Hobbs J-PA, Frisch AJ, Allen GR, van Herwerden L (2009) Marine hybrid hotspot at Indo-Pacific biogeographic border. *Biology Letters* **5**, 258-261.

Hobbs J-PA, Salmond JK (2008) Cohabitation of Indian and Pacific Ocean species at Christmas and Cocos (Keeling) Islands. *Coral Reefs* **27**, 933.

Hoeksema BW (2007) Delineation of the Indo-Malayan centre of maximum marine biodiversity: the coral triangle. In: *Biogeography, time, and place: distributions, barriers, and islands* (ed. Renema W), pp. 117-178. Springer, Dordrecht.

Horne JB, van Herwerden L, Choat JH, Robertson DR (2008) High population connectivity across the Indo-Pacific: Congruent lack of phylogeographic structure in three reef fish congeners. *Molecular Phylogenetics and Evolution* **49**, 629-638.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755.

Hughes TP, Bellwood DR, Connolly SR (2002) Biodiversity hotspots, centres of endemicity, and the conservation of coral reefs. . *Ecology Letters* **5**, 775–784.

Hyde J, Kimbress A, Budrick J, Lynne E, Vetter R (2008) Cryptic speciation in the vermilion rockfish (*Sebastes miniatus*) and the role of bathymetry in the speciation process. *Molecular Ecology* **17**, 1122-1136.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. In: *BMC Genetics*, p. 13.

Johannesson K, André C (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology* **15**, 2013-2029.

Johnson MS, Black R (1982) Chaotic genetic patchiness in an intertidal limpet, *Siphonaria* sp. *Marine Biology* **70**, 157-164.

Johnson MS, Watts RJ, Black R (1994) High levels of genetic subdivision in peripherally isolated populations of the atherinid fish Craterocephalus capreoli in the Houtman Abrolhos Islands, Western Australia. *Marine Biology* **119**, 179-184.

Jokiel P, Martinelli FJ (1992) The vortex model of coral reef biogeography. *Journal of Biogeography* **19**, 449-458.

Jones G, Planes S, Thorrold S (2005) Coral reef fish larvae settle close to home. *Current Biology* **15**, 1314-1318.

Jones GP, Milicich MJ, Emslie MJ, Lunow C (1999) Self-recruitment in a coral reef fish population. *Nature* **402**, 802-804.

Joshi MB, Rout PK, Mandal AK, Tyler-Smith C, Singh L, Thangaraj K (2004) Phylogeography and origin of Indian domestic goats. *Molecular Biology and Evolution* **21**, 454-462.

Kawagata S, Hayward B, Gupta A (2005) Benthic foraminiferal extinctions linked to late Pliocene–Pleistocene deep-sea circulation changes in the northern Indian Ocean (ODP Sites 722 and 758). *Marine Micropaleontology* **58**, 219-242.

Kingman JFC (1982) On the genealogy of large populations. *Journal of Applied Probability* **19**, 27-43.

Klanten SO, Choat JH, van Herwerden L (2007) Extreme genetic diversity and temporal rather than spatial partitioning in a widely distributed coral reef fish. *Marine Biology* **150**, 659-670.

Klanten SO, van Herwerden L, Choat JH, Blair D (2004) Patterns of lineage diversification in the genus Naso (Acanthuridae). *Molecular Phylogenetics and Evolution* **32**, 221-235.

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

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. *Hydrobiologia* **420**, 73-90.

Kon T, Yoshino T, Mukai T, Nishida M (2007) DNA sequences identify numerous cryptic species of the vertebrate: A lesson from the gobioid fish Schindleria. *Molecular Phylogenetics and Evolution* **44**, 53-62.

Kuhner MK (2009) Coalescent genealogy samplers: windows into population history. *Trends in Ecology & Evolution* **24**, 86-93.

Kuhnt W, Holbourn A (2004) Neogene History of the Indonesian Throughflow. *Geophysical Monograph* **149**, 299-320.

Ladd HS (1960) Origin of the Pacific island molluscan fauna. *American Journal of Science* **258**, 137-150.

Lavery S, Moritz C, Fielder DR (1996a) Genetic patterns suggest exponential population growth in a declining species. *Molecular Biology and Evolution* **13**, 1106-1113.

Lavery S, Moritz C, Fielder DR (1996b) Indo-Pacific population structure and evolutionary history of the coconut crab Birgus latro. *Molecular Ecology* **5**, 557-570.

Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G (2010) Allopatric divergence and speciation in coral reef fish: the three-spot dascyllus, Dascyllus trimaculatus, species complex. *Evolution* **64**, 1218-1230.

Lesica P, Allendorf FW (1995) When are peripheral populations valuable for conservation? *Conservation Biology* **9**, 753–760.

Lessios HA, Kessing BD, Pearse JS (2001) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution* **55**, 955-975.

Lessios HA, Kessing BD, Robertson DR (1998) Massive gene flow across the world's most potent marine biogeographic barrier. *Proceedings of the Royal Society of London Series B: Biological Sciences* **265**, 583-588.

Lessios HA, Robertson DR (2006) Crossing the impassable: genetic connections in 20 reef fishes across the eastern Pacific barrier. *Proceedings of the Royal Society of London Series B: Biological Sciences* **273**, 2201-2208.

Liu K, Pigram CJ, Paterson L, Kendall CG (1998) Computer simulation of a Cainozoic carbonate platform, Marion Plateau, north-east Australia. In: *Reefs and carbonate platforms in the Pacific and Indian Oceans. Special publication number 25 of the International Association of Sedimentologists.* (eds. Camoin GF, Davies C). Blackwell Science Pty Ltd, Victoria, Australia.

Losos JB, Glor RE (2003) Phylogenetic comparative methods and the geography of speciation. *Trends in Ecology & Evolution* **18**, 220-227.

Lourie SA, Green DM, Vincent ACJ (2005) Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Sygnathidae: *Hippocampus*). *Molecular Ecology* **14**, 1073-1094.

Malay MC, Paulay G (2009) Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: *Calcinus*). *Evolution* **64**, 634-662.

Marie AD, van Herwerden L, Choat JH, Hobbs J-PA (2007) Hybridization of reef fishes at the Indo-Pacific biogeographic barrier: a case study. *Coral Reefs* **26**, 841-850.

Mayr E (1963) *Animal species and evolution.* Belknap Press, Cambridge, MA.

McCartney MA, Acevedo JA, Heredia C, Rico C, Quenouille B, Bermingham E, McMillan WO (2003) Genetic mosaic in a marine species flock. *Molecular Ecology* **12**, 2963-2973.

McCoy ED, Heck KLJ (1976) Biogeography of corals, sea grasses and mangroves: an alternative to the centre of origin concept. *Systematic Zoology* **25**, 201-210.

McManus JW (1985) Marine speciation, tectonics and sea-level changes in Southeast Asia. **4**, 133–138.

McMillan WO, Palumbi SR (1995) Concordant evolutionary patterns among Indo-West Pacific Butterflyfishes. *Proceedings of the Royal Society of London Series B: Biological Sciences* **260**, 229-236.

McMillan WO, Palumbi SR (1997) Rapid rate of control-region evolution in Pacific butterflyfishes (Chaetodontidae). *Journal of Molecular Evolution* **45**, 473-484.

Messmer V, Jones G, van Herwerden L, Munday P (2005) Genetic and ecological characterisation of colour dimorphism in a coral reef fish. *Environmental Biology of Fishes* **74**, 175-183.

Meyer A, Morrissey JM, Schartl M (1994) Recurrent origin of a sexually selected trait in *Xiphophorus* fishes inferred from a molecular phylogeny. *Nature* **368**, 539-542.

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA*, 1-8.

Montanari SR, van Herwerden L, Pratchett MS, Hobbs J-PA, Fugedi A (2012) Reef fish hybridization: lessons learnt from butterflyfishes (genus *Chaetodon*). *Ecology and Evolution* **2**, 310-328.

Mora C, Chittaro PM, Sale PF, Kritzer JP, Ludsin SA (2003) Patterns and processes in reef fish diversity. *Nature* **421**, 933-936.

Moran MD (2003) Arguments for rejecting the sequential Bonferroni in ecological studies. *Oikos* **100**, 403-405.

Myers RF (1999) *Micronesian Reef Fishes: A Comprehensive Guide to the Coral Reef Fishes of Micronesia.* Coral Graphics, Guam.

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

Nei M, Maruyama T, Chakraborty R (1975) The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1-10.

Nunes F, Norris RD, Knowlton N (2009) Implications of isolation and low genetic diversity in peripheral populations of an amphi-Atlantic coral. *Molecular Ecology* **18**, 4283-4297.

Nylander JAA (2004) MrModeltest 2.2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Palumbi SR (1997) Molecular biogeography of the Pacific. *Coral Reefs* **16**, S47-S52.

Palumbi SR, Grabowsky G, Duda TF, Geyer L, Tachino N (1997) Speciation and population genetic structure in tropical pacific sea urchins. *Evolution* **51**, 1506–1517.

Pandolfi JM (1992) Successive isolation rather than evolutionary centres for the origination of Indo-Pacific reef corals. *Journal of Biogeography* **19**, 593-609.

Parenti P, Randall JE (2000) An annotated checklist of the species of the Labroid fish families Labridae and Scaridae. *Ictyological Bulletin of the J.L.B. Smith Institute of Icthyology* **68**, 1-97.

Pianka E (1978) *Evolutionary ecology.* Harper and Row, New York, NY.

Planes S, Fauvelot C (2002) Isolation by distance and vicariance drive genetic structure of a coral reef fish in the Pacific Ocean. *Evolution* **56**, 378-399.

Posada D (2006) ModelTest Server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucleic Acids Research* **34**, W700-W703.

Posada D, Crandall KA (2001) Intraspecific gene genealogies: trees grafting into networks. . *Trends in Ecology & Evolution* **16**, 37–45.

Randall JE (2010) *Shore Fishes of Hawai'i.* University of Hawaii Press, Honolulu, HI.

Randall JE, Allen GR, Steene RC (1997) *Fishes of the Great Barrier Reef and Coral Sea.* University of Hawaii Press, Honolulu, HI.

Randall JE, Bruce RW (1983) The parrotfishes of the subfamily Scarinae of the western Indian Ocean with a description of 3 new species. *Ictyological Bulletin of the J.L.B. Smith Institute of Icthyology* **47**, 1-39.

Ray N, Currat M, Excoffier L (2003) Intra-Deme Molecular Diversity in Spatially Expanding Populations. *Molecular Biology and Evolution* **20**, 76-86.

Reece JS, Bowen BW, Joshi K, Goz V, Larson A (2010) Phylogeography of Two Moray Eels Indicates High Dispersal Throughout the Indo-Pacific. *Journal of Heredity* **101**, 391-402.

Rees M, Heyward A, Cappo M, Speare P, Smith L (1994) Ningaloo Marine Park - Initial survey of seabed biodiversity in intermediate and deeper waters., p. 45. Produced for: Australian Government Department of the Environment and Heritage.

Reid DG, Lal K, Mackenzie-Dodds J, Kaligis F, Littlewood DTJ, Williams ST (2006) Comparative phylogeography and species boundaries in Echinolittorina snails in the central Indo-West Pacific. *Journal of Biogeography* **33**, 990-1006.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution* **43**, 223-225.

Roberts CM, McClean CJ, Veron JEN, Hawkins JP, Allen GR, McAllister DE, Mittermeier CG, Schueler FW, Spalding MD, Wells F, Vynne C, Werner TB (2002) Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science* **295**, 1280–1284.

Rocha LA, Bowen BW (2008) Speciation in coral-reef fishes. *Journal of Fish Biology* **72**, 1101-1121.

Rogers AR, Harpending HC (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-569.

Rosenberg NA, Feldman MW (2002) The relationship between coalescence times and population divergence times. In: *Modern Developents in Theoretical Population Genetics* (eds. Slatkin M, Veuille M). Oxford University Press, Oxford.

Rosenblatt R, Hobson ES (1969) Parrotfishes (Scaridae) of the eastern Pacific, with a generic rearrangement of the Scarinae. *Copeia* **1969**, 434-453.

Sáez AG, Lozano E (2005) Body doubles. Cryptic species: as we discover more examples of species that are morphologically indistinguishable, we need to ask why and how they exist. *Nature* **433**, 111.

Sale PF (1991) Reef fish communities: open nonequilibrial systems. In: *The ecology of fishes on coral reefs* (ed. Sale PF), pp. 564-598. Academic Press, San Diego, CA.

Sambrook J, Russell DW (2001) *Molecular cloning, a laboratory manual.* Cold Spring Harbour Laboratory Press., New York, NY.

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.

Scheltema RS (1986) On dispersal and planktonic larvae of benthic invertebrates: an eclectic overview and summary of problem. *Bulletin of Marine Science* **39**, 290-322.

Schneider S, Excoffier L (1999) Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. *Genetics* **152**, 1079-1089.

Schultz LP (1958) Review of the parrotfishes, family Scaridae. *Bulletin of the United States National Museum* **214**, 1-143.

Shanks A, Grantham B, Carr M (2003) Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications* **13**, S159–S169.

Sheppard C, Wilson S, Salm R, Dixon D (2000) Reefs and coral communities of the Arabian Gulf and Arabian Sea. In: *Coral reefs of the Indian Ocean: their ecology and conservation* (eds. McClanahan T, Sheppard C, Obura D), pp. 257-293. Oxford University Press, New York, NY.

Siegel D, Kinlan B, Gaylord B, Gaines S (2003) Lagrangian descriptions of marine larval dispersion. *Marine Ecology Progress Series* **260**, 83–96.

Slatkin M (2005) Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. *Molecular Ecology* **14**, 67-73.

Smith LL, Fessler JL, Alfaro ME, Streelman JT, Westneat MW (2008) Phylogenetic relationships and the evolution of regulatory gene sequences in the parrotfishes. *Molecular Phylogenetics and Evolution* **49**, 136-152.

Somero GN (2002) Thermal physiology and vertical zonation of intertidal animals: optima, limits, and costs of living. *Integrative and Comparative Biology* **42**, 780-789.

Spalding MD, Fox HE, Allen GR, Davidson N, Ferdaña Z, Finlayson M, Halpern B, Jorge M, Lombana A, Lourie S, Martin KD, McManus E, Molnar J, Recchia C, Robertson J (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience* **57**, 573-583.

Stehli FG, Wells JW (1971) Diversity and age patterns in hermatypic corals. . *Systematic Zoology* **20**, 115–126.

Streelman JT, Alfaro M, Westneat MW, Bellwood DR, Karl SA (2002) Evolutionary history of the Parrotfishes: Biogeography, ecomorphology, and comparative diversity. *Evolution* **56**, 961-971.

Swearer SE, Caselle JE, Lea DW, Warner RR (1999) Larval retention and recruitment in an island population of a coral-reef fish. *Nature* **402**, 799-802.

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony. Sinauer Associates, Sunderland, MA.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-460.

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

Tajima F (1996) The amount of DNA polymorphism maintained in a finite population when the neutral mutation rate varies among sites. . *Genetics* **143**, 1457-1465.

Templeton AR (1980) The theory of speciation via the founder principle. *Genetics* **94**, 1011–1038.

Templeton AR (2001) Using phylogeographic analyses of gene trees to test species status and processes. *Molecular Ecology* **10**, 779-791.

Teske PR, Hamilton H, Palsboll PJ, Choo CK, Gabr H, Lourie SA, Santos M, A. C, M.I. S, Matthee CA (2005) Molecular evidence for long-distance colonization in an Indo-Pacific seahorse lineage. *Marine Ecology Progress Series* **286**, 249-260.

Thompson J, Gibson T, Plewniak F, Higgins D (1997) The ClustalX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**, 4876-4882.

Timm J, Kochzius M (2008) Geological history and oceanography of the Indo-Malay Archipelago shape the genetic population structure in the false clown anemonefish (Amphiprion ocellaris). *Molecular Ecology* **17**, 3999-4014.

van Herweden L, Choat JH, Dudgeon CL, Carlos G, Newman S, Frisch AJ, van Oppen MJH (2006) Contrasting patterns of genetic structure in two species of the coral trout *Plectropomus* (Serranidae) from east and west Australia: Introgressive hybridisation or ancestral polymorphisms. . *Molecular Phylogenetics and Evolution* **41**, 420-435.

Veron JEN, Devantier LM, Turak E, Green A, Kininmonth S, Stafford-Smith M, Peterson N (2009) Delineating the Coral Triangle. *Galaxea, Journal of Coral Reef Studies* **11**, 91-100.

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

Vogler C, Benzie J, Lessios HA, Barber PH, Wörheide G (2008) A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. *Biology Letters* **4**, 696-699.

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

Werner FE, Cowen RK, Paris CB (2007) Coupled biological and physical models: Present capabilities and necessary developments for future studies of population connectivity. . *Oceanography* **20**, 54-69.

Werner R (2008) R/V Sonne Cruise SO199 CHRISP: New Insights into the Geodynamic History of northern Wharton Basin (South-East Indian Ocean) Eos Trans. In: *American Geophysical Union, 89 (53) Fall Meeting*.

Williams ST, Benzie JA (1996) Genetic uniformity of widely separated populations of the coral reef starfish *Linckia laevigata* from the West Pacific and East Indian Oceans, revealed by allozyme electrophoresis. *Marine Biology* **126**, 99-108.

Williams ST, Benzie JA (1997) Indo-West Pacific patterns of genetic differentiaiton in the high-dispersal starfish *Linckia laevigata*. *Molecular Ecology* **6**, 559-573.

Williams ST, Benzie JA (1998) Evidence of a phylogeographic break between populations of a high-dispersal starfish: congruent regions within the Indo-West Pacific defined by colour morphs, mtDNA and allozyme data. *Evolution* **52**, 87-99.

Williams ST, 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.

Winters K, van Herwerden L, Choat JH, Robertson DR (2010) Phylogeography of the Indo-Pacific parrotfish *Scarus psittacus*: isolation generates distinctive peripheral populations in two oceans. *Marine Biology* **157**, 1679-1691.

Woodland DJ (1983) Zoogeography of the Siganidae (Pisces): An interpretation of distribution and richness patterns. *Bulletin of Marine Science* **33**, 713-717.

Woodroffe CD, Berry PF (1994) Scientific studies in the Cocos (Keeling) Islands: An introduction. *Atoll Research Bulletin (Chapter 1)* **399**, 1-16.

Woodroffe CD, Veeh HH, Falkland A, McLean RF, Wallensky E (1991) Last interglacial reef and subsidence of the Cocos (Keeling) Islands, Indian Ocean. *Marine Geology* **96**, 137-143.

Zink RM, Barrowclough GF (2008) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* **17**, 2107-2121.

Zuvela-Aloise M (2005) *Modelling of the Indonesian throughflow on glacial-interglacial time-scales* PhD Dissertation, Christian Albrechts University.

Appendix

Manuscripts in preparation

Population structure of the widespread parrotfish *Chlorurus sordidus*: evidence of genetic partitioning at three marine biogeographic barriers. To be submitted to *Coral Reefs* (Chapter 3).

Strong population structure, zones of overlap and peripheral isolation in the widespread parrotfish, *Scarus rubroviolaceus* (Perciformes: Scarinae) (Chapter 4).

Population structure and cryptic speciation in the widespread Indo–Pacific scarine labrid, *Scarus ghobban* (Perciformes: Scarinae) (Chapter 5).

Comparative phylogeography of Indo–Pacific reef fish: a synthesis of four Indo–Pacific scarine labrids (Chapter 6).

Conference presentations

Beck E, van Herwerden L, Klanten OS, Choat JH, Robertson, DR (2006) Phylogeography, biogeography and the evolutionary history of a widespread parrotfish, *Scarus ghobban*: what does it tell us about IndoPacific reef fish evolution? Australian Coral Reef Society: Coral reef research: from the seas to the skies (2006), Mission Beach, Australia.

Beck E, vanHerwerden L, Klanten OS, Choat JH, Bay LK, Robertson, DR (2009) Population structure of the widespread parrotfish *Chlorurus sordidus*: evidence of genetic partitioning at three marine biogeographic barriers. Australian Society for Fish Biology 8th Indo Pacific Fish Conference: Biogeography & Biodiversity (2009), Fremantle, Australia.