Evolutionary and Ecological Genetic Patterns of Widespread Unicornfishes: Investigations of the Non-geographic Clades Phenomenon

Thesis by John Barton Horne BSc

For the degree of Doctor of Philosophy in Marine Biology

Submitted: April 2011

James Cook University

School of Marine and Tropical Biology

Publication list

Publications directly resulting from this thesis:

Chapter 2

Horne JB, McIlwain JL & van Herwerden (2010) Isolation of 15 new polymorphic microsatellite markers from the bluespine unicornfish *Naso unicornis*. *Conservation Genetics Resources*, **2**, 191-194.

Chapter 3 & 4

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.

Chapter 5

Horne JB, van Herwerden L & Choat JH (In prep) Horizontal gene flow between two Indo-Pacific unicornfish sister species *Naso hexacanthus* and *Naso caesius*.

Chapter 7

Horne JB, Abellana S, McIlwain JL, Choat JH & van Herwerden L (In prep) Recruitment patterns of the bluespine unicornfish in the Marianas archipelago.

Conceptual Paper

Horne JB, van Herwerden L & Choat JH (In prep) On the phylogeography of widespread Indo-Pacific coral reef dwelling species and perceived barriers to marine dispersal.

Publications that arose during the candidature of this thesis

Horne JB, Momigliano P, Welch DJ, Newman SJ & van Herwerden L (2011) Limited ecological population connectivity suggests low demands on self-recruitment in a tropical inshore marine fish (*Eleutheronema tetradactylum*: Polynemidae). *Molecular Ecology*, in press.

Horne JB, Momigliano P, Welch DJ, Newman SJ & van Herwerden L (in review) Searching for common threads in threadfins: comparative phylogeography of two Australian polynemid fisheries. *Marine Ecology Progress Series*.

Acknowledgements

I would like to thank the following people, who were either directly or indirectly instrumental to the completion of my research.

Crystal Horne: First and foremost I would like to thank my wife Crystal for her unwavering love and support. She is an amazing woman with many talents and qualities that I can only hope to emulate. She made many sacrifices so that I could pursue my doctorate degree here in Australia, not the least of which was working full time so that I could concentrate on school. Moreover, she sat by and watched many of her friends marry men with good jobs, watched them buy "cute" houses while we lived in less than desirable accommodation. She missed weddings and other important events. In particular she chose to stay in Australia and support me when her family back home was going through hard times. No man ever had a better, more loyal wife than mine. Crystal, I love you.

Tom and Susan Horne: I consider myself lucky to have had such good parents. All my life my parents have nurtured my interests and endured my obsession with fish. I remember my mother had to replace the carpets in our house because one of the filter hoses on my 55-gallon aquarium came loose. More importantly, however, my parents always taught me to value knowledge and to do what I wanted with my life. Without their direction I don't think I would have pursued the path that has lead me here today.

Lynne van Herwerden: I would like to thank my primary supervisor Lynne van Herwerden. Lynne is always very patient, very kind and I sometimes think it would be impossible to frustrate or make her angry. At the same time Lynne is remarkably dependable, always returns my manuscripts with revision in a timely manner and somehow manages to produce papers of her own while supervising 6-9 students. I know that she toils indefatigably behind the scenes though you'd never know how hard she works just from a superficial glance. Lynne has been a superlative example to me in so many ways. I used to worry how I would ever juggle a career as a scientist and a family but Lynne has taught me (by example) how to put my family first and still have a fruitful career.

Howard Choat: Oliver Wendell Holmes once likened men unto peaches or pears, saying that they tend to sweeten right before they decay. If that is true then Howard will not decay any time soon. Lynne was my nice supervisor, while Howard was my tough one. Lynne was good cop. Howard was bad cop. I still remember when he told me that my brain had been "fried by the green laser from Mars" and the time he told me that one of my manuscripts was a "self-indulgent ramble" and the time that he sharply asked me if I knew the difference between a noun and a verb. My memories of Howard Choat will forever arouse my sense of humility. I have benefited greatly from Howard's zero tolerance for mediocrity and his no-nonsense attitude. I thank him for having the gumption to tell me that my work was bad even when it very nearly broke my heart to hear it. He held me to a high standard, for which I will always be grateful.

Selma Klanten: Selma was my predecessor in Lynne's lab. In fact, her work laid the foundation for my research and paved the way for many of my papers. In many ways I have stood on her shoulders.

James Cook University Molecular Ecology and Evolution lab: I remember thinking what a dump the lab was when I first toured it. There were dead bugs everywhere and mold growing on the ceiling. Then the school built a new building to house the MEEL lab and now it is probably one of the nicest labs of its kind in the southern hemisphere. Too bad the move was after I had finished most of my lab work. Regardless, the MEEL lab has been like a home to me and has provided for all my needs. I will always remember sequestering my self there, sometimes for 15 hours at a time. I am especially grateful to the numerous lab managers we've had over the years: Adrian, Di, Stacy, Heather, Blanche. And also a warm thank you to Ainsley Calladine, the Megabace technician, who ran all of my microsatellite samples.

Other acknowledgements: I would like to thank the James Cook University international student center for helping and supporting me as an international student. I would like to thank the school of Marine and Tropical Biology and the post-graduate school of research. I would like to thank the University of Guam marine lab, where I spent several weeks, and Jenny McIlwain, who has been an important collaborator. I would like to thank all sources of funding: the James Cook University internal funding scheme and the U.S. Fish and Wildlife's Federal Assistance in Sportfish Restoration Program to the Guam Department of Agriculture. Thank you to all the people who did field work and supplied samples for this research: Sheena Abellana, Mark Priest, Brett Taylor, Alyssa Marshall, Andrew Halford, JP Hobbs, DR Robertson, W Robbins, M Meekan, BW Bowen, JA Eble and others.

Abstract

The tropical Indo-Pacific is a very large biogeographic region stretching from the Red Sea to the central Pacific, yet in spite of its size nearly 500 species of shallow water teleost fishes are mostly cosmopolitan throughout this vast region. To date only a small number of these species have been phylogeographically studied across their entire species ranges and most show some genetic population structure, usually between ocean basins. However, some Indo-Pacific coral reef fishes have genetically homogenous populations across the Indo-Pacific suggesting high levels of gene flow. Among those fishes that lack spatial population partitions some possess non-geographic or temporal population structures in the mtDNA. This means that their populations are composed of genetically differentiated lineages that are sympatrically distributed. The biological significance of this pattern is unclear and only a small number of reef fish taxa have been reported to have non-geographic clades. The first species in which this pattern was detected was the big nose unicornfish, *Naso vlamingii*, which is a widespread Indo-Pacific Acanthurid with a long pelagic larval duration (Klanten *et al.* 2007). The explanation given by Klanten *et al.* for the non-geographic clades of this species was temporary periods of isolation, mediated by low-sea-level barriers and followed by secondary contact and admixture of lineages. Yet, while the conclusions of Klanten *et al.* (2007) seem plausible, further investigation into non-geographic genetic patterns was warranted.

In this thesis four additional *Naso* species were phylogeographically surveyed at a broad scale across the Indo-Pacific: *Naso brevirostris*, *Naso unicornis*, *Naso hexacanthus* and *Naso caesius*. Like *N. vlamingii*, all species showed little in the way of spatial population structuring using conventional phylogeographic fixation indexes

(F_{st}). Model-based phylogenetic analysis, maximum parsimony and Bayesian inference revealed non-geographic population patterns in the mitochondrial control region with *N. brevirostris* being the most similar to *N. vlamingii*. Curiously, the sister species *N. hexacanthus* and *N. caesius* were not reciprocally monophyletic suggesting that some horizontal gene transfer has occurred between them. Expansion time based on mismatch distributions was also calculated for each species and each clade. Mismatch coalescence suggests mid-late Miocene expansion times for all taxa (considering *N. hexacanthus* and *N. caesisus* as a single taxon) and pre-Pleistocene expansion for non-geographic clades.

Coalescence age of species and clades was also assessed in the program BEAST using a fossil calibrated molecular dating and a relaxed molecular clock. For this analysis all *Naso* species previously mentioned, including *N. vlamingii*, were arranged into an interspecific phylogeny of cytochrome oxidase subunit one DNA sequences, with many representatives of each species and clade. Mean ages from this anaylsis suggest mid-miocene coalescence ages for each taxa and pre-Pleistocene coalescence for all non-geographic clades, indicating that Pleistocene processes probably do not account for non-geographic patterns in the genus *Naso*. Otherwise, little congruence could be found between the ages of clades, seemingly refuting the notion that a barrier was involved in the formation of these genetic features. Fossil calibrated coalescence ages were used to reconstruct demographic history of each species using Bayesian skyline plots. In all taxa, skyline plots suggest accelerated demographic expansion during the Pleistocene, rather than the bottlenecks or demographic decline expected based on the pretext that reef fishes suffered bottlenecks during the upheavals on coral reefs during the Pleistocene glacial cycles.

Complimentary to evolutionary genetic patterns, this thesis also includes a genetic survey of recently settled *N. unicornis* juveniles in the Marianas archipelago,

using 12 polymorphic microsatellite loci and sequences from the mitochondrial control region. This particular study reveals that post-settlement larval pools of *N. unicornis* have an elevated genetic diversity in both the nuDNA and mtDNA relative to the preexisting adult population. Recruitment events are genetically unstructured in space and time in this species within the Marianas and recruits seemed likely to have sibling relationships among sampling locations and between year classes. The genetic diversity of larval pools, as well as the likelihood of sibling pairs to be separated in space and time suggests demographically open populations. In light of the phylogeography of this species, *N. unicornis* and its congeners most likely experience high levels of gene flow on ecological as well as evolutionary time frames.

Non-geographic clades appear to be a genetic phenomenon observed in some widespread and highly dispersive coral reef fishes of the Indo-Pacific. There was little evidence to suggest that a low-sea-level barrier to gene flow was associated with these clades. Most likely non-geographic clades arise as a consequence of rapid range expansion across patchily distributed coral reef habitat in a large biogeographic region. The sheer size of the Indo-Pacific appears to be an important factor shaping the population patterns of widespread cosmopolitan coral reef fishes.

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction to Coral Reef Fish Studies

1.1 Speciation, Phylogeography and Marine Barriers

Fishes found on coral reefs have an impressive taxonomic diversity. Springer (1982) estimated there were 4,000 species of reef and shallow water fishes in the Indo-Pacific alone, though he suspected this number might be "much too low." Choat & Bellwood (1991) claimed that 4,000 species would comprise 18% of all fish species known at the time (roughly 22,000). In the past few decades, the number of described fish species has increased dramatically. According to Eschmeyer & Fricke (2009) the tally for total fish species has grown to over 31,000. If we assume that the number of described reef fishes has increased proportionally, we might, by way of crude extrapolation, approximate the number of reef fishes known currently to be about 5,600 species. Rocha & Bowen (2008 & references therein) also estimate the number of described reef fishes to be about one third of ~15,000 marine fishes. Considering that there are only about 5,000 living mammal species total (Wilson & Reeder 2005), the study of reef fishes is not only a massive scientific undertaking but also key to understanding the natural mechanisms that produce and sustain vertebrate biodiversity.

The fossil record reveals that the higher taxonomic groups of modern reef fishes first appeared and diversified early in the Cenozoic. The oldest benthic marine fish fossils that resemble modern day reef fish communities come from the lower middle Eocene deposits in Monte Bolca, Italy, where 247 fossil fish species from 82 families and 17 orders have been reported (Bellwood 1996). A large number of the

families present at Monte Bolca have extant representation on modern coral reefs and many fossils have skeletal morphologies that are highly similar to present-day species. Bellwood & Wainwright (2002) express that the Eocene Tethys Sea fish faunas in the Bolca region would have differed little in appearance and ecology from those seen in the tropics today.

Since the Eocene there has been a major shift in the biogeography of reef fishes. Obviously there are no longer coral reefs or reef fishes in northern Italy. Today the greatest biodiversity of fishes, corals and other reef denizens are found in the Indo-Australian Archipelago (IAA) but diverse assemblages of reef fishes are found in all tropical oceans with multiple centers of high endemicity (Randall 1998; Bellwood & Wainwright 2002; Allen 2008). The question asked by biogeographers is when did the present-day provinciality of reef fishes emerge and what were the factors that shaped it?

Many present-day reef fishes have long species histories that originate in the Miocene (23-5 Ma) (e.g. Bellwood *et al.* 2004; Bernardi *et al.* 2004; Klanten *et al.* 2004; Read *et al.* 2006; Dornburg *et al.* 2008; Cowman *et al.* 2009; Bellwood *et al.* 2010). If the initial great diversification of reef fishes happened in the Eocene, then the second great diversification was in the Miocene, at which time Renema *et al.* (2008) conclude that the new biodiversity hotspot in the IAA was established. But what were the evolutionary processes that promoted high levels of speciation starting in the Miocene and continuing on until the present day?

Speciation in its simplest form occurs across a geographic barrier, which prevents genetic exchange between two populations and allows them to diverge. This allopatric mode of speciation is presumed to have given rise to the majority of species on our planet, or has at least been the most documented (Coyne & Orr 2004).

Freshwater habitats, in particular, are intrinsically separated by allopatric barriers, which may account for the immense species diversity of freshwater fishes (Wiley 2002). The ocean, conversely, is a relatively homogenous environment and has few absolute barriers that would prevent migrant exchange in highly dispersive marine species. It is, therefore, difficult to invoke strict allopatry to explain marine biodiversity (Palumbi 1994), except in cases where it is unmistakable, such as across the Isthmus of Panama (see Lessios *et al.* 2008). Therefore, other models of speciation and more complex mechanisms of isolation must be considered to satisfactorily explain the species richness that exists on coral reefs and in other marine environments. Alternatives to strict allopatric speciation include ecological speciation (sympatry), partial geographic isolation (parapatry) and peripheral isolation (peripatry) (see Palumbi 1994; Choat 2006; Rocha & Bowen, 2008; Puebla 2009).

Models that predict speciation processes may be tested using a molecular approach, in which the genetic patterns that arise in selectively neutral gene regions are assessed. Neutral molecular markers may develop random polymorphic characters, which can be treated as a genetic signature of a population. Under the assumption of neutrality, two forces act upon such genetic signatures: gene flow and genetic drift (Slatkin 1987). Populations that regularly exchange genes through migration will eventually have their genetic signatures come to resemble each other. Alternatively, if populations are reproductively isolated, they will genetically drift apart and may eventually become different species. The structure and phylogeny of populations in a drift-equilibrium context indicates where barriers to gene flow have occurred, spatially or along an ecological gradient, and to what degree populations are genetically connected. This research method is often referred to as

phylogeography (Avise *et al.* 1987; Avise 1998) and is particularly useful for studying dispersive marine organisms, where population boundaries are often not obvious (Palumbi 1994; Hellberg 2007).

In recent years, genetic studies have shed considerable light on speciation process in the sea and its complexity. Several examples of ecological and sympatric speciation have been documented for marine fishes (Munday *et al.* 2004; Rocha *et al.* 2005a; Puebla *et al.* 2007; Wellenreuther *et al.* 2008; Hyde *et al.* 2008; Crow *et al.* 2010). However, speciation in sympatry is controversial (Coyne & Orr 2004) and difficult to demonstrate conclusively because the species ranges at the time of speciation are usually unknown (Bolnick & Fitzpatrick 2007). Even some good cases may not satisfy the strictest criteria prescribed to demonstrate sympatric speciation (Elmer & Meyer 2010). Peripatric isolation also appears to be an important driver of divergence in coral reef fishes (Gaither *et al.* 2010; Winters *et al.* 2010; Craig *et al.* 2010; Leray *et al.* 2010; DiBattista *et al.* 2011). Yet perhaps the best examples of peripatric speciation come from marine invertebrates (Budd & Pandolfi 2010; Malay & Paulay 2010). High levels of endemism of reef fishes at peripheral locations (Randall 1998; Allen 2008) might suggest peripatric speciation but endemism does not necessarily equate to speciation (Bellwood & Meyer 2009).

The potential for parapatric speciation in marine environments, in the form of soft or invisible oceanic barriers is becoming increasingly noted by marine researchers (Palumbi 1994; Rocha & Bowen 2008). Parapatry is difficult to distinguish from allopatry because the resulting biogeographic pattern is essentially the same (Coyne & Orr 2004). However, as Turelli *et al.* (2001) explain, any mechanism that can produce speciation in allopatry will do the same in parapatry, if the geographic scale is large enough. Considering the immense size of the marine

environment and the apparent paucity of absolute barriers, parapatry is likely to be common in the ocean. The abundance of soft barriers (i.e. stretches of water that have properties that restrict connectivity between separated populations) in the ocean is not readily apparent but there are a few soft barriers that are conspicuous. These include the East Pacific Barrier (EPB), the Benguela Barrier (BB), the Amazon Barrier (AB) and the Indo-Pacific or Sunda Shelf barrier (IPB) (reviewed by Rocha *et al.* 2007).

Soft barriers are, by nature, penetrable, to certain degrees, at certain times, or by certain species. The cold Benguela current that forms the BB, for example, normally segregates the fish faunas of the tropical Atlantic and West Indian Ocean. However, this barrier may have historically been subject to disturbances that have allowed colonizations of the Atlantic by Indian Ocean species (Rocha *et al.* 2005b; Bowen *et al.* 2006; Floeter *et al.* 2008). In another example, the EPB is a 5,000 km expanse of ocean that segregates the benthic faunas of two marine biogeographic provinces: the coastal Americas and the reef habitats of the Central Pacific (Bellwood & Wainwright 2002). Yet some reef fishes, such as the Acanthurid *Acanthurus triostegus*, have high genetic connectivity between the East and Central Pacific and are thought to regularly migrate across the EPB (Planes & Fauvelot 2002; Lessios & Robertson 2006). Therefore soft barriers may not be barriers at all to some species and may be more accurately described as filters. Rare dispersal across these filters and or fluctuations in their effectiveness appears to lead to speciation events in reef fishes (Floeter *et al.* 2008).

Considering that semi-allopatric soft barriers may create opportunities for speciation, it seems intuitive that there would be a soft barrier associated with the IAA biodiversity hotspot. In deed, a barrier does seem to operate between the

tropical Indian and Pacific, for although both oceans are faunistically similar, a large number of reef fishes are geographically confined to one ocean or the other (Springer & Williams 1990; Randall 1998) and there is growing data on a biogeographic suture zone between the two oceans (Hobbs *et al.* 2009). Phylogeographic studies also show that reef fishes found throughout the Indo-Pacific have deep genetic population partitions coinciding with a barrier somewhere in the IAA (Planes & Fauvelot, 2002; Bay *et al.* 2004; Craig *et al.* 2007; van Herwerden *et al.* 2009; Leray *et al.* 2010; Eble *et al.* 2010). The presence of such a barrier is puzzling because, unlike the BB and EPB, the mechanism that separates Indian and Pacific reef fish populations is not immediately obvious. One might expect the IAA, with its continuous warm water and abundant reef habitat to facilitate migrant exchange between oceans rather than suppress it. Nevertheless, for at least one widespread species (*Myripristis berndti*), the IPB appears to be a more effective barrier to dispersal than the EPB (Craig *et al.* 2007).

Given the geographic and oceanographic complexity of the IAA, multiple factors most likely act in concert to create the effect of the IPB. For some taxa, with lower dispersal abilities, such as many marine invertebrates, inter-oceanic dispersal is probably hindered by ocean circulations that operate in this area (Barber *et al.* 2006; Carpenter *et al.* 2011). But most authors have proposed that the major biogeographic effect of the IPB is caused by eustatic sea level changes in relation to the Indonesian Sunda Shelf (Springer & Williams 1990; Randall, 1998; Benzie 1999; Carpenter *et al.* 2011). This submerged continental shelf is the largest on Earth outside of the Polar Regions and comprises an area approximately 1.8×10^6 km^2 (Hanebuth *et al.* 2000). During low extremes in sea-level, most of this land mass would have been above water, reef habitat within the IAA would have been

drastically reduced and the coastline of the Sunda Shelf would have become a large obstacle for dispersing marine organisms. Still, even when sea-level falls 120 m below current levels, the tropical Indian and Pacific are connected by a large seaway (Voris 2000), the Indo-Pacific throughflow, which should permit some inter-oceanic migrant exchange. Indeed, some taxa show high genetic connectivity across the IPB (Lessios 2003; Klanten *et al.* 2007; Gaither *et al.* 2010; Reece *et al.* 2010). Therefore, the forces that create the IPB may be attenuated by dispersal ability, life history characteristics or chance migrations.

A recent development in the phylogeography of coral reef fishes is the presence of non-geographic or temporal clades. This genetic phenomenon was first observed in a widespread Indo-Pacific surgeonfish, *Naso vlamingii*, which had four well-defined clades that were sympatrically distributed rather than reciprocally monophyletic across a barrier (Klanten *et al.* 2007). The interpretation of this pattern was that anciently a barrier between the Indian and Pacific Oceans separated *N. vlamingii* populations. Subsequently, the barrier has disappeared and introgression, followed by high levels of inter-oceanic gene flow, has erased the ancient population boundary between oceans, though a genetic partition remains. Such an interpretation is consistent with a fluctuating sea-level barrier in the IAA and because there are more than two clades, isolation across this barrier would have occurred multiple times.

Non-geographic structuring of genetic lineages appears to be directly related to soft ocean barriers and therefore relevant to the parapatric processes that cause population differentiation and speciation in marine organisms. In fact, Klanten *et al.* (2007) speculated that non-geographic clades might be sympatrically distributed cryptic species. But what biological features set *N. vlamingii* apart from other marine

organisms such that the genetic patterns it exhibits would be so drastically different? Klanten *et al.* (2007) gave two reasons why this might be the case. 1) A lengthy pelagic larval duration and a larval ecology that would give *N. vlamingii* a high dispersal potential, even for a coral reef fish. Elevated dispersal would enable previously isolated populations to come into secondary contact after periods of isolation, thus resulting in widely overlapping lineages. 2) A semi-pelagic adult ecology, which would allow this species to utilize a variety of habitats and food sources. Such ecology would enable *N. vlamingii* to maintain large refuge populations in isolation, even during the suboptimal climatic conditions, such as glacial maxima. These populations would later expand to become geographically overlapping lineages. However, at the time this research thesis commenced *N. vlamingii* was the only known example of non-geographic clades and the suspicions of Klanten *et al.* lacked corroboration. Therefore much more work was required to explore these unusual genetic patterns.

1.2 Chapter outline

In the wake of Klanten *et al.* (2007), the broad-scale phylogeography of *N. vlamingii*, several questions remain unanswered:

1 Are the non-geographic genetic patterns present in *N. vlamingii* unique to this species?
2 Is a semi-pelagic ecology key to the formation of non-geographic clades?
3 Are non-geographic clades cryptic species?
4 Are non-geographic clades the result of temporal isolations across an ephemeral low-sea-level barrier? If so, how long ago did the barrier exist?

5 Does non-geographic structure imply elevated levels of dispersal, i.e., frequent long distance migrant exchange as opposed to evolutionary genetic connectivity?

The aim of this thesis was to address each of these questions by conducting additional phylogeographic surveys on other members of the genus *Naso* and to explore spatial and temporal genetic patterns in this genus using a variety of molecular approaches.

Chapter 2*: Methods*

The purpose of chapter to is to act as a general methods section for field and laboratory work. Included in this chapter is also a description of the protocol I used to develop microsatellite loci for *Naso unicornis.*

Chapter 3*: Broad-scale phylogeography of* **Naso brevirostris**

Are the non-geographic genetic patterns present in N. vlamingii *unique to this species?*

Because *N. vlamingii* was the only species known to possess non-geographic clades there is a possibility that the unusual patterns observed are stochastic in nature. The purpose of chapter three was to attempt to find another example of non-geographic structuring through a broad-scale phylogeographic study of a similar species *Naso brevirostris* using the methods of Klanten *et al* (2007). Both species are phylogenetically allied and have long species histories (Klanten *et al.* 2004), both have semi-pelagic habitat preferences and both forage on filamentous green algae

and gelatinous zooplankton (Choat *et al.* 2002). For these reasons *N. brevirostris* was chosen as the first species to be examined.

Chapter 4: *Broad-scale phylogeography of N. unicornis*

Is a semi-pelagic ecology key to the formation of non-geographic clades?

Chapter four is a broad-scale phylogeographic study of *N. unicornis*. Like *N. vlamingii* and *N. brevirostris* this species is widespread in the Indo-Pacific and has a long species history but is primarily benthic in its habitat preferences and has a diet dominated by brown macroalgae (Choat *et al.* 2002). The purpose of this chapter was to determine whether a *Naso* species without semi-pelagic ecology would also show non-geographic patterns.

Chapter 5: *Broad-scale phylogeography of widespread sympatric sister species:* Naso hexacanthus *and* Naso caesius

Are non-geographic clades cryptic species?

Chapter five is a broad-scale phylogeography of *Naso hexacanthus* and *Naso caesius,* which are two, ecologically very similar, sister species. They are so similar that they sometimes form heterospecific schools and forage together (Randall 2002). They are semi-pelagic, like *N. vlamingii* and *N. brevirostris,* but unlike the other species examined, which have long species histories, *N. hexacanthus* and *N. caesius* are purported to have diverged relatively recently (Klanten *et al.* 2004). In reality, the relationship between these two sister species may be analogous to the non-geographic clades found in other species. If so, the genetic differentiation between them should be similar to non-geographic clades and it might be concluded that non-geographic structures is grounds for cryptic speciation.

Chapter 6: *Molecular dating of non-geographic clades*

Are non-geographic clades the result of temporal isolations across an ephemeral low-sea-level barrier? If so, how long ago did the barrier exist?

The purpose of chapter six was to construct a phylogenetic tree containing all species and all clades, in which the nodes can be dated using a fossil calibration. In addition to the mismatch based expansion times of Klanten *et al.* (2007) and previous chapters, fossil calibrated molecular dating will illuminate the evolutionary history of non-geographic clades and the conditions under which they were formed.

Chapter 7: *Recruitment patterns of* Naso unicornis *on Guam*

Does non-geographic structure imply elevated levels of dispersal, i.e, frequent long distance migrant exchange as opposed to evolutionary genetic connectivity?

Chapter seven is an assessment in recruitment patterns in *N. unicornis* in the Marianas Archipelago, using 12 nuclear microsatellite loci. In connection with phylogeographic patterns, which indicate the scale of gene flow on an evolutionary time scale, recruitment patterns will indicate dispersal processes on an ecological time scale.

Chapter 8: *Discussion and Synthesis of research*

Chapter eight is a discussion on the main conclusions and contributions of this thesis and centers around the population processes that lead to the formation of non-geographic clades, the evolutionary history of the study taxa and the dispersal ability of *Naso.*

Chapter Two

Field and Laboratory Methods

2.1 Field Collections

The specimens used in this research come mostly from the tissue library of Howard Choat and were sampled from a broad range of geographic localities within the tropical Indo-Pacific between March 2000 and December 2009. The species used in this research were: *Naso brevirostris* (n = 102), *Naso unicornis* (n = 196), *Naso hexacanthus* (n = 92), *Naso caesius* (n =25) and *Naso vlamingii* (n = 30). Primarily these samples originated from the Seychelles Islands, Cocos Keeling, Christmas Island (Indian Ocean), Western Australia, the Australian Great Barrier Reef, Solomon Islands, New Guinea, the Philippines and French Polynesia. A small number of samples came from other locations, such as Oman, Taiwan, Rodriguez and Reunion Island. Samples were collected by spearing or purchased in fish markets and stored in either 80% ethanol or a 20% di-methyl-sulfoxide buffer (see Figures 2.1; 2.2; 2.3).

Additionally, some sampling was done independent of the Choat collection. *N. unicornis* was intensely sampled from the island of Guam, in cooperation with the Jennifer McIlwain lab of the University of Guam, between June 2008 and November 2009 (Figure 2.4). From Guam 176 adults and 394 newly settled recruits from five different sites were collected. Twenty-four *N. unicornis* (adults) from Hawaii were donated from the Brian Bowen lab at the Hawaii Institute of Marine Biology. Thirty more were purchased at fish markets in Nuku'alofa, Tonga. Eleven *N. hexacanthus* and 9 *N. caesius* were also purchased at markets in Tonga. All of the above samples

were placed in 80% ethanol, directly following sampling, for transport and storage. All sampling was conducted under James Cook University ethics approval numbers A503, A872.04 and A1539.

.

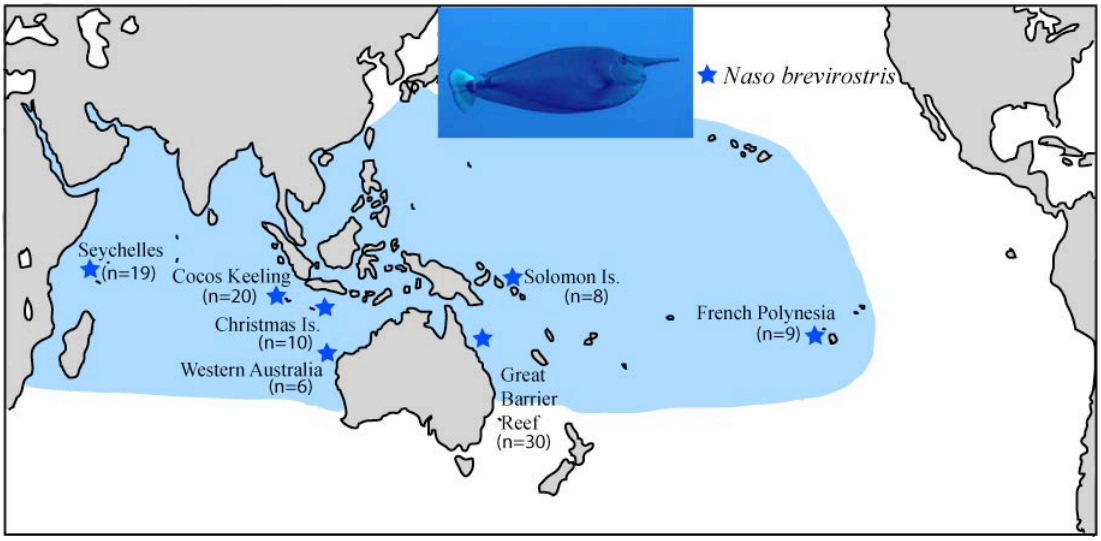


Figure 2.1: Sample collections of *N. brevirostris* at seven locations across the Indo-Pacific.

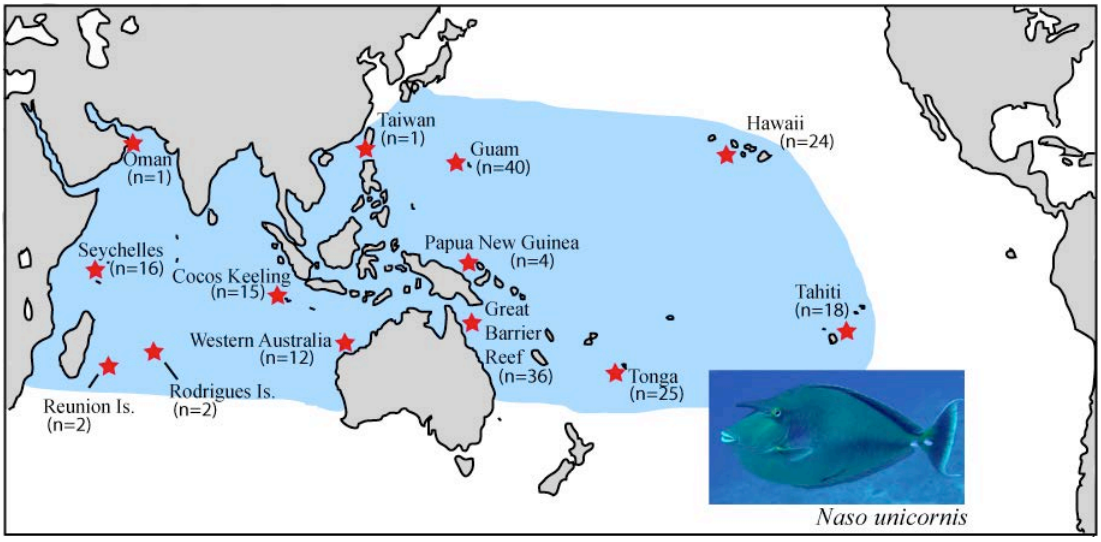


Figure 2.2: Sample collections of *N. unicornis* at 13 locations across the Indo-Pacific.

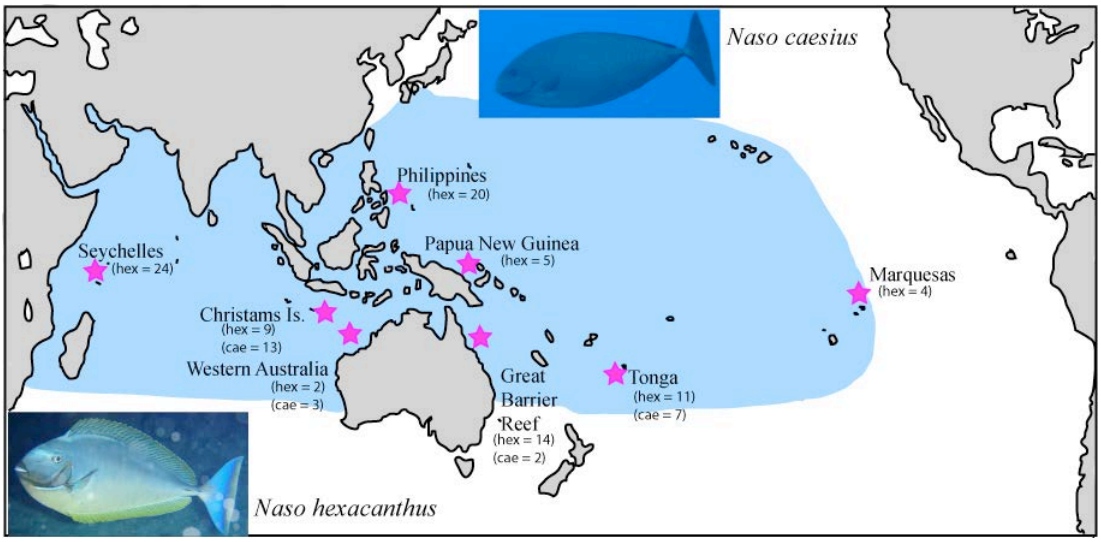


Figure 2.3: Sample collections of *N. hexacanthus* at eight locations and *N. caesius* at four locations across the Indo-Pacific.

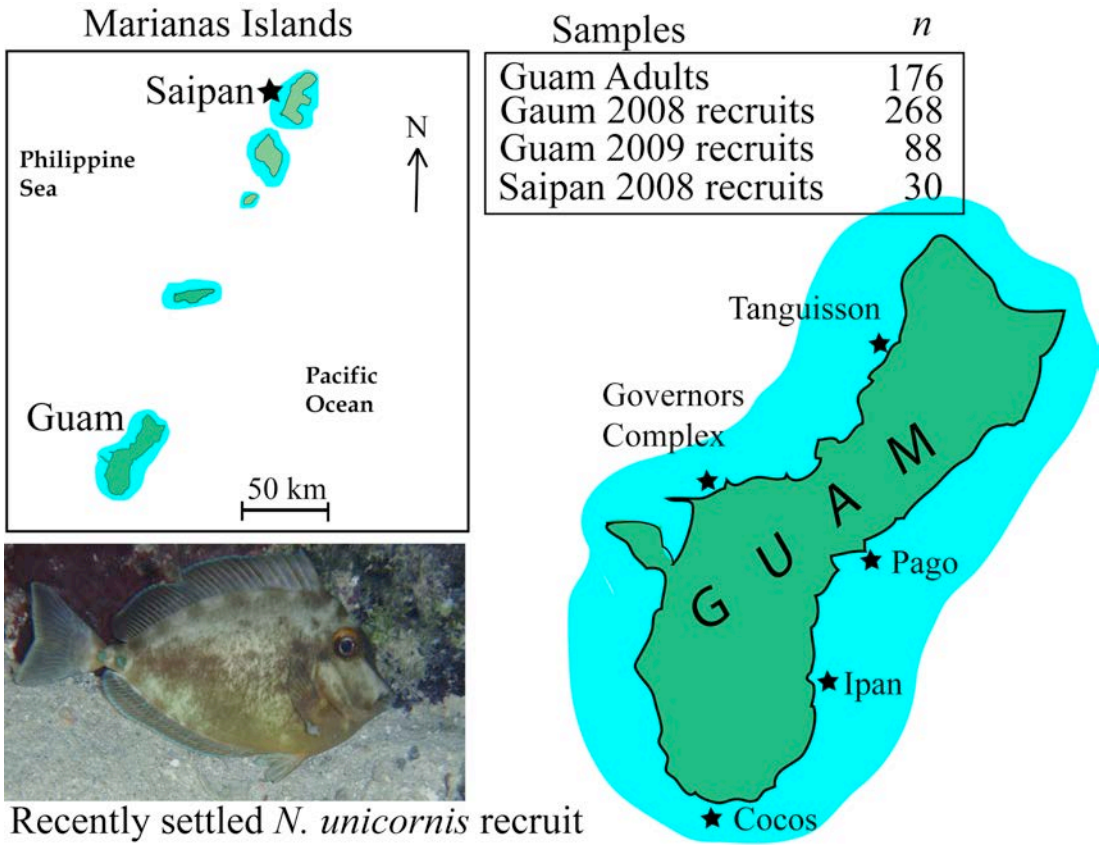


Figure 2.4: Sampling locations of recently settled *N. unicornis* recruits from Guam and Saipan.

2.2 DNA extraction

Total genomic DNA was extracted from fin clips using a salt chloroform protocol (Sambrook *et al.* 1989) or a Chelex extraction method (Walsh *et al.* 1991).

2.3 Mitochondrial Control Region

Klanten *et al.* (2007) designed genus specific mitochondrial control region (control region) primers for the broad-scale *N. vlamingii* phylogeography. These same NAI primers were employed to amplify this gene region in the other *Naso* species (Table 2.1). This gene is useful for phylogeography because it is non-coding and hypervariable (Mortiz *et al.* 1987) and should allow for the detection of subtle genetic signatures not found in other markers, particularly in population level analyses (AMOVA, mismatch distributions etc.). PCR parameters for these primers are as follows: DNA was amplified in 20 μL PCR reactions containing 2.5 mM Tris–Cl (pH 8.7), 5 mM KCL, 5 mM (NH4)2SO4, 200 μM each dNTP, 3.5 mM MgCl2, 10μM each primer, 1 U of Taq Polymerase (Qiagen Ltd.) and 1-10 ng of template DNA. Thermocycling was carried out with an initial denaturation of 94 ˚C for 2 min, 35 cycles of denaturation, annealing and extension (94 ˚C for 30 s, 50 ˚C for 30 s, 72 ˚C for 90 s) and a final extension of 72˚C for 10 min. PCR products were confirmed by gel electrophoresis on 1.5% agarose gels and purified by either a standard isopropanol purification or an ammonium acetate ethanol clean-up. PCR products were sequenced with the NA1 forward primer using ABI (Applied Biosystems Incorporated) technologies at Macrogen sequencing service Seoul, South Korea.

2.4 Cytochrome Oxidase Subunit 1

The relatively conserved, gene encoding cytochrome oxidase subunit 1 (COI) region

was also amplified for each species and a gene tree constructed for the purpose of divergence dating (chapter 6). The COI marker was chosen for this analysis for several reasons: 1) because it is more conserved than control region and therefore less prone to saturation. 2) COI has a greater range of phylogenetic signal than any other mitochondrial gene and is useful at both the species and population levels (Herbert *et al.* 2003; Hellberg 2009). 3) The universal COI primers work well, nearly always yielding strong PCR products. 4) As the designated barcode gene (Herbert *et al.* 2003) there are many genbank references to which comparisons may be made. COI was amplified using universal primers developed for fish by Ward et al. (2005) (Table 2.1). PCR parameters for these primers were as follows: 2 μL of 10x Pfu PCR buffer, 200 μM of each dNTP, 0.2 μM of each primer (Table 2.1), 0.2 U *Pfu* DNA polymerase (Promega, Madison, Wisconsin).

Table 2.1: Mitochondrial marker primers for the mitochondrial control region and cytochrome oxidase subunit 1 with the accompanying oligonucleotide sequences, annealing temperatures and magnesium concentration.

Primer Name	Oligo Sequence	T_a	$MgCl_2$
NAI **(Control Region)**	F 5'-AGCATTCTGAACTAAACTAC R 5'- TGTCCCTTGACTCTCAATA	50˚C	3.5mM
Fish 1 **(COI)**	F 5'- TCAACCAACCACAAAGACATT GGCAC R 5'-TAGACTTCTGGGTGGCCAAAGAATCA	55˚C	1.5-5.5mM

2.5 Microsatellite markers for *Naso unicornis*: development and PCR parameters

Phylogeographic patterns such as population-structure, genetic diversity indices and mismatch distributions reveal population dynamics primarily over evolutionary time. Genetic patterns that are relevant to demographic processes on a recent time scale are not easily detected in the mutations of nucleotide sequence data and in most cases these ecological signals appear as "background noise" in a drift-equilibrium context

(Hellberg 2007).

Microsatellite loci are hypervariable markers that may be employed in a number of highly sensitive population level analyses such as assignment tests, parentage analysis, mixture and admixture analysis (Manel *et al.* 2005). The merit of microsatellite loci for answering questions of demographic connectivity is found in their amalgamation. Individually, each locus is no more informative than any other type of marker but as suites of independently assorted loci the possibility of two populations having the same allelic frequencies by chance alone is very unlikely. Thus, the more unlinked loci used the more sensitive the analysis becomes.

In recent years these microsatellite loci have become increasingly sought after for studying coral reef fish species, which often have large geographic ranges and complex patterns of population connectivity at varying spatial scales (Purcell *et al.* 2006; Thacker *et al.* 2007; Hepburn *et al.* 2009; Planes *et al.* 2009). However, compared to the diversity of reef fishes, the number of published microsatellite loci available to coral reef fish biologists is few and at the time that this thesis began, no microsatellite loci had been developed for any Acanthurid. Therefore, it became necessary to develop novel microsatellite loci for *Naso* and the magnetic bead capture protocol of Glenn & Schable (2005) was used to this end.

We extracted total genomic DNA from fin clips of three *N. unicornis* adults collected from the Seychelles, Lizard Island (Great Barrier Reef) and Tonga. Tissue was digested with proteinase K and DNA extracted using a salt-chloroform method (Sambrook *et al.* 1989). Total genomic DNA was digested into 300-1000 bp fragments using the restriction enzymes Rsa I and Hae III separately. Double stranded SNX linkers were ligated to both sides of the blunt ended fragments using T4 DNA ligase, in accordance to the protocol of Glenn & Schable (2005).

Linker ligated DNA fragments were PCR amplified using the super SNX-24 primer and subsequently annealed to four different combinations of biotinylated oligo probes of di, tri and tetra-nucleotide motifs [$(AG)_{16}$ + $(AC)_{16}$, $(AAG)_8$, $(AAAC)_6$ + $(AATT)_8$, $(AAGG)_6$ + $(AGCG)_4$]. Reaction mixtures consisted of 25 μL 2x hyb solution, 10 μL mixed oligos (1 μM each), 10 μL linker ligated DNA, 5 μL nuclease free H_2O for a total volume of 50 μL. Thermocycler temperatures for the annealing of probes were as follows: 95˚C for five minutes, 70˚C for five seconds followed by 99 five-second incremental step downs of 0.2˚C and 50˚C for 10 minutes followed by 20 five-second incremental step down of 0.5˚C.

50 μL of Streptavidin bonded magnetic beads (Dynabeads, Invitrogen Dynal, Olso, Norway) were washed twice in 250 μL of TE buffer, twice in 1x hyb solution and suspended in a final volume of 150 μL 1x hyb solution. The hybridized DNA-oligo fragments were added to the magnetic bead solution and captured with a magnetic particle separator, while the supernatant and miscellaneous DNA was discarded.

Microsatellite enriched DNA was again PCR amplified using the super SNX-24 primer and ligated into the pCR®2.1-TOPO vector (Invitrogen) as per the manufacturer's instructions. Cloned inserts were sent to the Australian Genome Research Facility at the University of Queensland, Brisbane for bacterial transformation and sequencing with universal M13 primers.

The resulting DNA sequences were screened for microsatellite loci in the program MSATCOMMANDER (Faircloth 2008) and primers for each locus were designed in the same program. A total of 71 loci were detected and the 38 best (those with the longest number of repeats) were chosen for further screening. Loci were amplified in a minumum of five individuals using standard PCR parameters as

described above (see also Table 2.2). Polymorphism was tested by direct sequencing, which revealed high sequence variation, within loci and in the flanking regions. Additionally, length variation was not consistently uniform across loci, with many alleles differing in length by single base pairs due to small indels in the flanking sequences. Most of the microsatellite loci of *N. unicornis* could be described as imperfect or compound, as many had different repetitive elements adjacent to each other or were bisected by non-repeating base pairs. Though not optimal for microsatellite analysis, the excessive amount of polymorphism observed in the nuclear genome of *N. unicornis* is consistent with mitochondrial studies which show high genetic diversity in this and other *Naso* species (e.g. *Naso vlamingii* Klanten *et al.* 2007).

The 15 cleanest (those with fewest indels) and most polymorphic loci were screened against 90 adult *N. unicornis* (114-485mm FL) collected from Guam in the tropical West Pacific (Table 2.2). The genbank accession numbers for these loci are: GU189390-GU189404. Sample DNA was PCR amplified using forward primers that were flurolabeled with either HEX, TET or FAM dyes (Geneworks pty. Ltd. Hindmarsh, Southern Australia). PCR products of each of the three dyes were combined into 96-well plates, purified using a standard ethanol ammonium acetate clean-up and read using Amersham MegaBACE instrumentation at the James Cook University Genetics Analysis Facility. The program GENEPOP 4.0 (Rousset 2008) was used to assess allelic diversity, estimates of heterozygosity for each locus, to test for departures from Hardy-Weinberg equilibrium (HWE) and to test linkage disequilibrium. Mean observed and expected heterozgosities across all loci were 0.66 and 0.82 respectively. The program Micro-Checker (van Oosterhout *et al.* 2004)

indicated that Nuni02, Nuni06, Nuni13 and Nuni15 might be confounded by null alleles.

PCR amplifications of microsatellite loci were carried out in 10 μL reactions with the following reagents: 1 μL 10x Pfu buffer, 200 μM of each dNTP, 10 μM of each primer (table), 0.1 U *Pfu* DNA polymerase (Promega, Madison, Wisconsin) and 6.2 μL of H_2O. PCR products were purified using an ethanol and ammonium acetate precipitation and read using Amersham MegaBACE instrumentation at the James Cook University Genetics Analysis Facility.

Table 2.2: Description of 15 microsatellite loci for the blue-spine unicorn fish (*Naso unicornis*).

Locus	Repeat Motif	Primer Sequence	T_a (˚C)	N_a	Size range (bp)
Nuni01	$(CA)_{12}$	F 5' ACGCACAGTTGAGGGAGAG R 5' AAGGACAAAGTGTGAGGGG	60	10	167-185
Nuni02	$(CA)_{16}$	F 5' CTCTGGGTATGCTTAATGGGC R 5' CCAGCCTTGTTGTTCACCG	60	11	166-189
Nuni03	$(CA)_{20}$	F 5' GATGAGGCTACACACGCTG R 5' AGGAGTTTCACTTCCTCCCAC	60	21	151-196
Nuni04	$(CA)_{20}$	F 5' GAACACACGGGCTGCTG R 5' CCATGTATTTGGAGAGTAGTAGTGC	58	19	140-185
Nuni05	$(CA)_{23}$	F 5' CCCCTTCCTGTGGCTGTAG R 5' CCTGGTTTGCACTTGGAGC	60	17	181-214
Nuni06	$(CA)_{22}$	F 5' AGTGTCGCTCCTTCAGTGC R 5' CGCAGGTGAACGGCATATC	61	21	171-215
Nuni07	$(GA)_{34}$	F 5' GATTCAGGCACGCCACAC R 5' TGTTTGTGCAGCTTGGGAG	60	23	212-246
Nuni08	$(GTT)_{11}$	F 5' CGCATTTTGTTTCCACTGCC R 5' AGGATCCGCTGGTTACCTC	60	12	158-179
Nuni09	$(TAGA)_{13}$	F 5' TCCCAGTTATCACCGCCTG R 5' TCCAATACACCTGTTCTGCC	60	17	173-249
Nuni10	$(CATT)_{11}$	F 5' TGCTCCTACTCGACTCATTTCC R 5' CTGGAGTTTGTGAGTTGTCGG	60	21	205-248
Nuni11	$(CTT)_9$,$(GTT)_2$	F 5' TGCTAACTGCCAAGGACCC R 5' TGAACCTGAAAACGAGGAGC	60	5	161-174
Nuni12	$(CA)_{10}$,$(ACGC)_{13}$	F 5' TCACAGAGTGTGTATGATTGTCTG R 5' CCCTGCTGGTCATTGTGTTG	60	16	202-250
Nuni13	$(GGTT)_6$, TTT, $(GTT)_5$	F 5' TAGTTCCTCAGCACAGCCC R 5' TCCTAATCTCAATGCACTGGC	60	19	204-247
Nuni14	$(ATT)_6$, $(GTT)_{10}$	F 5' TGTTTCGCTGCCATCAGAG R 5' TCGACACAATGAAGTGCCAG	60	15	234-262
Nuni15	$(CTT)_2$, TTTCTC, $(CTT)_5$	F 5' TCCTCTCCACTGGCATCTG R 5' GCCTCCATGCAGACATTAGC	60	9	210-229

Screening results from 90 adult *N. unicornis* from Guam. (T_a) refers to PCR primer annealing temperature in degrees Celsius, (N_a) refers to number of alleles. Genbank accession numbers: GU189390-GU189404.

Chapter Three

Broad-scale Phylogeography of *Naso brevirostris*

3.1 Introduction

The broad-scale phylogeography of *Naso vlamingii*, by Klanten *et al.* (2007), revealed genetic patterns that were previously unobserved in a coral reef fish. To summarize the results of Klanten *et al.* these patterns were: 1) abnormally high genetic diversity in the mitochondrial gene region, 2) a general lack of population structure at all spatial scales, 3) the presence of clades, separated by deep population partitions, seemingly unrelated to geography (non-geographic clades). The interpretation of these patterns was that the non-geographic clades are remnant of historical episodes of isolation across an ancient barrier (presumably the IPB) that has since been lifted. Additionally, because these clades are sympatrically occurring across the sampled distribution, gene flow on a large spatial scale must be high in order to have so thoroughly mixed populations that were once allopatric.

An alternative interpretation is that the genetic patterns observed in *N. vlamingii* do not reflect biological processes but are stochastic in nature, having arisen randomly in the absence of a barrier. Based on computer simulations, de Aguiar *et al.* (2009) have suggested that population subdivision and even speciation can arise stochastically, without the influence of geography. Computer simulation studies should be accepted cautiously as they cannot account for all variables in complex natural populations but the possibility of stochastic speciation should be taken seriously, especially in the marine environment where biodiversity is

considered high and the evolutionary mechanisms that have produced it are poorly understood.

Equally as problematic as stochasticity is the potential for homoplasy in the *N. vlamingii* data. The genetic diversity in the mitochondrial control region of this species is extreme and, according to Klanten *et al.* (2007), is the highest of any fish studied to date. Such high genetic diversity may be evidence that the infinite sites model (Kimura 1969) has been violated. In other words, analysis of DNA sequence data assumes that there has been only a single mutation at each nucleotide position. If multiple mutations have occurred at some positions then there is a hidden signal in the data. In the case of *N. vlamingii*, violations of the infinite sites model may make individuals from separate populations appear to be in the same clade. This homoplasy may be so extensive as to destroy any geographic signal in the data.

A third conflict arises because there were no haplotypes shared among locations in the *N. vlamingii* data. Notwithstanding that there was no geographic population partitioning, there is no direct evidence that *N. vlamingii* populations are exchanging migrants without haplotypes shared among locations. AMOVA fixation indices, the primary metric used for assessing population structure, determine structure by dividing among-population variation by within-population variation. In a data set like that of *N. vlamingii*, where within population variation is extreme, there is little variation to be distributed among populations and fixation indices are unable to detect structure if it exists (Hellberg 2007). In other words, sample size is not great enough to capture the natural genetic signal and delimit genetic population boundaries. Therefore, in spite of the biological interest of the genetic patterns observed in *N. vlamingii*, substantiation is required to draw any concrete conclusions.

If a fluctuating barrier is responsible for the non-geographic clades of *N. vlamingii* then population patterns of similar species may have been likewise influenced by the same historical isolations. Of all *Naso* species, *Naso brevirostris* is perhaps the most similar to *N. vlamingii*, in terms of its ecology. Both species are pelagic foragers with a diet that is dominated by gelatinous zooplankton and filamentous green algae (Choat *et al.* 2002) and both species are known detritivores (Randall 2002). So alike are these two species in their habitat preferences and feeding modes that they are commonly seen in close proximity to each other (pers. observ). *N. brevirostris* might also be considered the sister species of *N. vlamingii* based on the phylogeny Klanten *et al.* (2004), though they are believed to have diverged over 20 million years ago. *A priori*, no other species seems as likely to share a parallel evolutionary history with *N. vlamingii* than *N. brevirostris.* Thus, the aims of this chapter were to assess the congruence between the phylogeographic patterns of these two species following the methods as Klanten *et al.* (2007). If the patterns found in *N. vlamingii* are also observed in *N. brevirostris*, it will suggest that biological rather than merely stochastic processes underpin non-geographic population patterns.

3.2 Methods

3.2.1 samples

102 *N. brevirostris* samples from seven locations (Seychelles, n = 19; Cocos Keeling, n = 20; Christmas Island, n = 10; Western Australia, n = 6; Great Barrier Reef, n = 30; Solomon Islands, n = 8; French Polynesia, n = 9) were included in this study (Fig. 2.1). All samples came from the tissue library of Howard Choat and many were collected at the same time as the *N. vlamingii* samples used by Klanten *et*

al. (2007), making comparisons between the two studies highly relevant. The mitochondrial control region was PCR amplified using the NA1 *Naso* specific primers designed by Klanten *et al.* (2007). Laboratory processing of these samples is outlined in chapter 2. PCR products were sequenced with the NA1 forward primer using ABI (Applied Biosystems Incorporated) technologies at Macrogen sequencing service Seoul, South Korea. Sequences from this study can be obtained from GenBank Accession Nos. FJ216727-FJ216828.

3.2.2 Phylogenetic analysis

Sequences were first aligned using a Clustal W alignment (Higgins *et al.* 1994) implemented in BioEdit version 7.0.9.0 (Hall 1999). Sequences were then further aligned and edited visually in BioEdit. The best substitution model for the data was evaluated in jMODELTEST 0.1.1 (Posada 2008) using a maximum likelihood approach and the Akaike information criterion. Phylogenetic trees were constructed using two different strategies. Maximum likelihood (ML) estimation of phylogenetic trees was performed in Garli 0.951 (Zwickl 2006) using the best substitution model with 100 bootstrap replicates. ML majority rule consensus trees were computed in PAUP* (Swofford 1999). Bayesian inference phylogenetic analysis (BI) was done in Mr. Bayes 3.1. (Huelsenbeck & Ronquist 2001). Bayesian analysis was run with two Markov chain Monte Carlo MCMC chains for 2,000,000 generations, sampling every 100 generations, and a 20% burnin. Convergence of the MCMC chains was assessed using the online service: Are we there yet (AWTY)(Wilgenbusch *et al.* 2004). *N. brevirostris* trees were rooted with *N. vlamingii* as the outgroup. Non-geographic clades were identified based on monophyly as indicated by bootstrap support values (ML) and posterior probability values (BI).

3.2.3 Population genetic analysis

Arlequin version 3.1 (Excoffier *et al.* 2005) was used to calculate molecular diversity for all population level analyses. Median joining haplotype networks (Bandelt et al. 1995) were drawn based on parsimony using Network version 4.5.0.1 (copyright 2004–2008, Fluxus Technologies Ltd.) for each species to visually illustrate haplotype variability and population geneology. Network calculations did not include gaps in the sequence alignment. Haplotype diversity (*h*) and nucleotide diversity (*%π*) were used as measures of genetic diversity (Nei 1987) for all sampled populations. Genetic distances between populations was calculated as a pairwise F_{st} matrix, plotted against geographic distance and tested for isolation by distance with a Mantel test correlation executed by IBD web service (Jensen *et al.* 2005). Negative F_{st} values were reset to zero. Genetic population structure, was further explored among populations and clades with analysis of molecular variance (AMOVA), also implemented in Arlequin 3.1 (Excoffier *et al*. 1992; Weir 1996). Non-geographic clades were assessed using a χ^2 goodness-of-fit test to determine if individuals from one ocean basin significantly dominated clades. The null hypothesis for this test was that clade membership of an individual was independent of the ocean basin where it came from.

3.2.4 Demographic Expansion

Time since expansion was calculated with the formula: $t = \tau/2u$ and $u = 2\mu k$ (Rogers & Harpending 1992), where u = the mutation rate for the entire sequence, t = generations since expansion began, μ is the mutation rate at each nucleotide position and k = the sequence length. Female generation times for *N. brevirostris* were

calculated using the formula T = (α + ω)/2, where α = the age at first reproduction and ω = the age at last reproduction (Pianka 1978). This is the best proxy for generation time available in the absence of data regarding age specific mortality and fecundity, which cannot be acquired by normal sampling programs for fish that have extended life spans and multiple spawning episodes within the annual cycle. We calculated α and ω using previously published otolith age estimates (Choat & Axe 1996), which give some indication of transitions between stages in life history. Therefore, with an approximate α and ω of 3 and 40 years, average female generation time is 21.5 years.

Mutation rate (μ) for *N. brevirostris* was calculated from divergence rate estimates of mitochondrial control region from swordfish (*Xiphias gladius*) (Alvarado Bremer *et al.* 1995) as per Klanten et al. (2007). The proportional combination of 12.9% divergence rate for variable sites and 1.1% for conserved sites per million years provided the overall sequence mutation rate. Mismatch distributions were produced in Arlequin (Li 1977; Rogers 1995; Harpending 1994; Schneider and Excoffier 1999). Mismatch distributions provided (τ), the final variable in the formula, which is a measure of time since expansion, expressed as $1/(2u)$ generations (Rogers & Harpending 1992). The mismatch distributions also provide θ_0 and θ_1, which are values that, respectively, represent effective female population size (N_{ef}) at the time of the last common ancestor and the current effective female population size.

Time since expansion for the entire *N. brevirostris* data set, as well as for individual non-geographic clades was calculated. Fu & Li's D and F tests of selective neutrality were used to evaluate neutrality in our markers (Fu & Li 1993). However, because neutrality tests are based on the assumption of a constant population size,

significantly negative values can also indicate an expanding population. Fu & Li's test are thought to be the most appropriate tests for detecting background selection but for estimating population growth we used a second, Fu's *F*s test, which is more appropriate (Fu 1997; Ramos-Onsins & Rozas 2002). All neutrality tests were executed in DnaSP version 4.50.3 (Rozas *et al.* 2003).

3.3 Results

3.3.1 Diversity

A total of 249 base pairs of the mitochondrial control region (HVR-1) were analyzed for *N. brevirostris*. There were 78 parsimony informative sites and 34 singleton sites. The ratio of transitions to transversions was approximately 5 to 1. Notwithstanding the high number of variable sites, sequences were aligned easily due to somewhat conserved regions at either end of the sequence. Overall haplotype and nucleotide diversities were high, $h = 0.98$, $\%\pi = 7.32$. but not as high as those observed for *N. vlamingii* ($h = 1.0$, $\%\pi = 13.6$) (Klanten *et al.* 2007).

3.3.2 Phylogenetic analysis

The best substitution model for *N. brevirostris* was SYM+G, $\alpha = 0.2930$. BI placed *N. brevirostris* samples into three distinct lineages that were not arranged according to geography (see Fig 3.1). χ^2 tests revealed that none of these lineages were significantly dominated by individuals from either ocean basin (χ^2 goodness-of-fit test, Yate's correction, df = 1, $\chi^2 = 0$; 0.21052; 3.5119, $p > 0.05$). Although similar, the three lineages differed slightly in support and overall structure. The lineage designated as clade 2 appears to be monophyletic and well supported by both ML and BI. Clade 1 also appears to be monophyletic but has less support. Clade 3 is

basal to the other two clades. All three lineages were treated as three distinct non-geographic clades in AMOVA and for mismatch distributions. However, in many of our phylogenetic trees, clade 1 and clade 3 formed a single clade and are treated as such for some of our analyses. There were also three individuals that did not belong to any clade and were the basal taxa in our phylogenetic tree. It is likely that some homoplasy is present in the data (see Fig. 3.1 for example) but the underlying phylogenetic pattern is probably not destroyed by it.

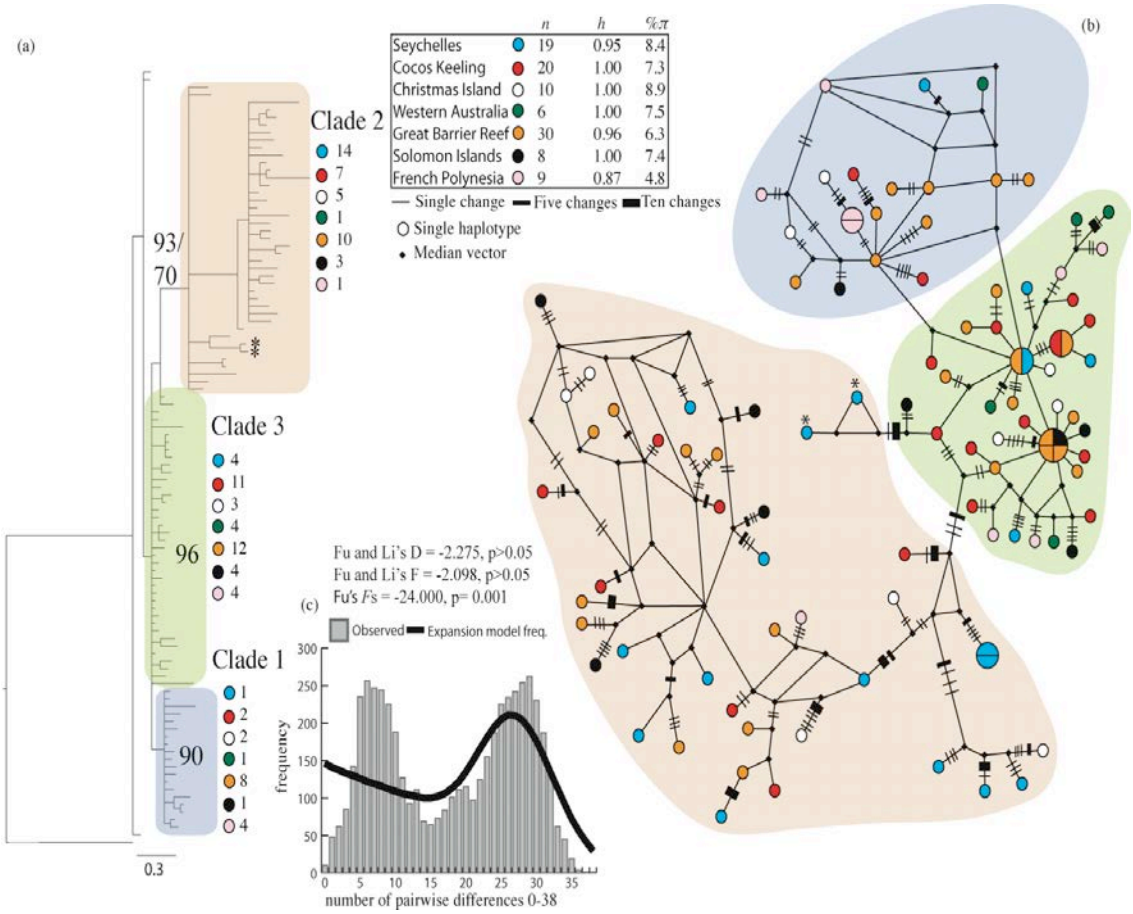


Figure 3.1: (a) Rooted phylogram of 102 *N. brevirostris* mitochondrial control region sequences from BI. Posterior probability values indicate three clades in which there is no geographic pattern. Clade 2 also has bootstrap support from ML analysis (b) Median joining network of *N. brevirostris* haplotypes. Sample size (n), haplotype diversity (*h*) and nucleotide diversity (*%π*) are given. (c) Mismatch distribution of *N. brevirostris* haplotypes Fu and Li's D and F tests of selective neutrality and Fu's *Fs* value. Haplotypes marked with asterisks indicate sequences that are peculiarly arranged in the network and may be affected by homoplasy.

3.3.3 Population structure

Pairwise F_{st} comparisons showed genetic differentiation between a subset of populations of *N. brevirostris* (Table 3.1). In particular, the Seychelles population was significantly different from Cocos Keeling, the GBR and French Polynesia. Pair wise F_{st} values increased gradually from west to east. Mantel test showed a significant and positive correlation between pairwise F_{st} values and pairwise geographic distance for *N. brevirostris*, indicating that isolation by distance is a factor in the gene flow of this species ($r = 0.4757$, $p = 0.046$). Log transformation of either variable did not significantly alter the outcome of the tests.

All AMOVA fixation indices for this study showed no evidence of population structure for *N. brevirostris* between the Indian and Pacific Oceans ($\Phi_{ct} = 0.008$, $p = 0.217$) (Table 3.2). For this analysis, the boundary between ocean basins was experimentally placed in three different locations: (i) between eastern and western Australia, (ii) between western Australia and Christmas Island and (iii) between Christmas Island and Cocos Keeling. Moving the boundary in this way did not significantly change the AMOVA result. Hence, probably no barrier to gene flow exists between the Indian and Pacific Ocean basins. Likewise, there was no evidence of population structure for *N. brevirostris* between the east and west Indian Ocean, or western and central Pacific Ocean ($\Phi_{ct} = 0.002$, $p = 0.453$). However, when samples were segregated into the seven sampling locations (i.e. GBR, Seychelles, etc.), shallow population structure was detected for *N. brevirostris* ($\Phi_{st} = 0.0773$, $p = 0.006$). In contrast to geographic tests of population structure, when the non-

geographic clades of *N. brevirostris* were tested, deep population structure was found (Φ_{st} = 0.4996, p = 0.001). AMOVA analysis of *N. vlamingii* populations revealed shallow structure between sampling locations and ocean basins (Φ_{st} = 0.065, p < 0.05) and (Φ_{st} = 0.076, p < 0.05), respectively (Klanten *et al.* 2007). Unfortunately Klanten *et al.* (2007) did not perform AMOVA on the non-geographic clades of *N. vlamingii* therefore no comparison can be made with *N. brevirostris.*

Table 3.1: Population pairwise F_{st} for sampled *N. brevirostris* populations (below) and corresponding p values (above). Significant values appear in bold. Only the Seychelles population showed structure with any other population. Structure becomes stronger as the distance grows greater suggesting that remote populations are isolated by distance.

	S	CK	CI	WA	GBR	SI	FP
Seychelles		**0.036**	0.441	0.189	**0.018**	0.207	**0.000**
Cocos	**0.042**		0.549	0.360	0.801	0.846	0.117
Christmas	-0.008	-0.019		0.585	0.270	0.531	0.144
W. Australia	0.063	-0.003	-0.034		0.162	0.459	0.702
GBR	**0.074**	-0.021	0.012	0.023		0.621	0.072
Solomon Is.	0.019	-0.043	-0.024	-0.023	-0.033		0.108
French Poly.	**0.162**	0.059	0.042	-0.044	0.068	0.071	

Table 3.2 AMOVA fixation indices (Φ_{st}) for *N. brevirostris*, percentage of genetic variation (%) and accompanying p values for four different comparisons: Overall (seven locations Seychelles-French Polynesia), Indian Ocean vs. Pacific Ocean, West Indian vs. East Indian+West Pacific vs. Central Pacific and clade vs. clade.

Comparison	Fixation index	%	p
Overall AMOVA	Φ_{st} = 0.0773	2.63	p = 0.006
Indian Ocean vs. Pacific Ocean	Φ_{st} = 0.022 Φ_{sc} = 0.021 Φ_{ct} = 0.008	97.01 2.11 0.88	p = 0.081 p = 0.128 p = 0.217
West Indian vs. East Indian+West Pacific vs. Central Pacific	Φ_{st} = 0.026 Φ_{sc} = 0.024 Φ_{ct} = 0.002	97.3 2.46 0.22	p = 0.083 p = 0.122 p = 0.453
Clade vs. Clade	Φ_{st} = 0.499	50.4	p = 0.001

3.3.4 Demographic Expansion

Unlike *N. vlamingii*, the mismatch distribution of the *N. brevirostris* data set was bimodal (Fig. 3.1) and had a reported τ value of 25.5 with upper and lower bounds of 10.9 and 35.3 respectively. Using the methodology of Klanten *et al.* (2007), *N. brevirostris* mean expansion time was 11.1 MY, compared to *N. vlamingii*, which had a mean expansion time of 8.1 MY. Mean expansion times for the clades of *N. brevirostris* ranged between 2 and 5 million years ago (Figure 3.2) and are comparably dated with the clades of *N. vlamingii*. Fu & Li's tests were not significant indicating that our marker is not under selection. Fu's *F*s was significantly negative (*F*s=-24.00, p= 0.001). Globally, *N. brevirostris* is probably an expanding population rather than one that has remained stable over prolonged periods. Comparisons of θ_0 and θ_1 are suggestive of long-term population growth for this species as they were for *N. vlamingii*.

Clade specific mismatch distributions of *N. brevirostris* (Fig. 3.2) revealed that clades one and three have unimodal mismatch distributions, suggestive of expansion, while clade two has an independently bimodal mismatch distribution and may not have evolved under the same conditions as the other clades. Fu's *F*s values are significantly negative for all clades. Fu & Li's tests are not significant for clades one and two but for clade three they were significant, suggesting that there may be background selection acting on this clade alone.

A combined clade 1 and 3 revealed a unimodal mismatch distribution and a significantly negative Fu's *F*s value (Fig. 3.2). Fu & Li's tests of selective neutrality differed for this arrangement: Fu & Li's D test was not significant, while Fu & Li's F test was significant. According to Fu & Li (1993), F is a more sensitive test than D, therefore, we must assume a selection signal here. Nevertheless, combining clades 1

and 3 appears only to have diluted the strong selection signal of clade 3 and may suggest that clades 1 and 3 are independent lineages that differ characteristically.

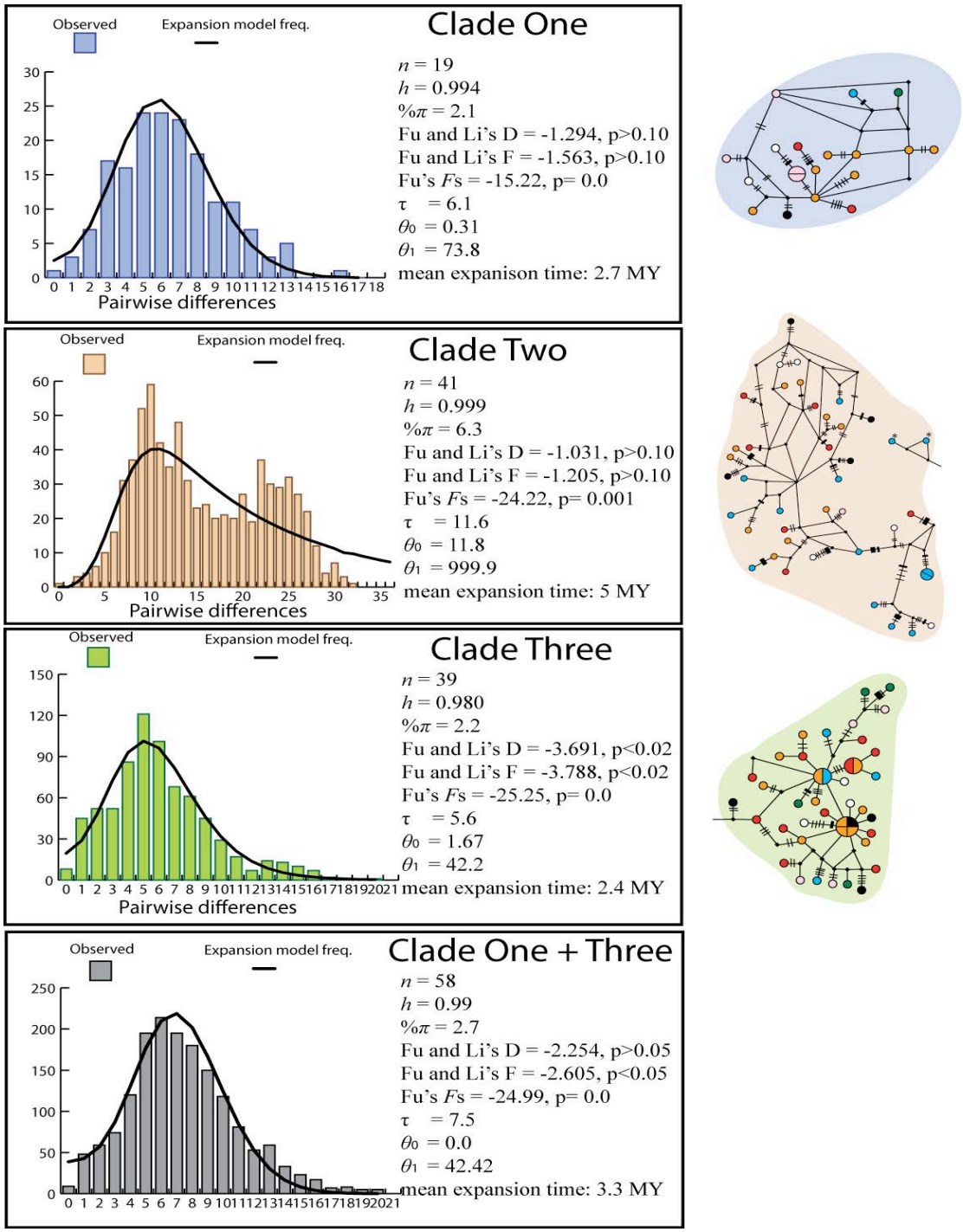


Figure 3.2: Clade Specific mismatch distributions with accompanying genetic diversity indices, neutrality tests and historical demographic values: τ, θ_0,θ_1, which are respectively the time since expansion measured as 1/(2u) generations (see methods), effective female population size (N_{ef}) at the time of expansion and N_{ef} of the current population. Fu and Li's D and F test assumptions of the neutrality of the genetic marker, while Fu's *F*s gives an indication of population expansion.

3.4 Discussion

3.4.1 Congruence with Naso vlamingii

The broad-scale phylogeography of *N. brevirostris* is generally congruent with that reported for *N. vlamingii* by Klanten et al. (2007). *N. brevirostris* exhibits: 1) high genetic diversity in the mitochondrial control region, 2) a lack of spatial population partitions at a large spatial scales, 3) non-geographic clades that appear to be analogous with those found in *N. vlamingii*. The presence of non-geographic clades in a second *Naso* species suggests that these are not stochastic patterns but residual population structures. Considering that the expansion ages of clades from both species are comparable, it seems unlikely that the similarities between the two data sets could be explained by chance alone. However, it should be noted that these dates are heuristic estimations and not precise ages.

3.4.2 Spatial patterns

Haplotypes shared among sampled areas in the *N. brevirostris* data compellingly argue that distant locations are exchanging genes and that the minimal population structure is not due to inflated genetic variation. Some of these haplotypes are shared between Indian and Pacific populations, suggesting that inter-oceanic gene flow occurs in this species. However, isolation-by-distance was significant in *N. brevirostris* implying that dispersal in this species is limited and that distant populations exchange genes in a stepping-stone fashion.

Overall AMOVA fixation indices showed no population structure on the largest geographic scales but on the smaller scale of our sampling locations (i.e. GBR, Solomon Islands, French Polynesia) shallow, yet significant, population structure was detected (Table 3.2). Population structure on a small scale but not on a

large scale can be due to a scaling effect. When populations are classed by ocean basin, the ocean-wide population automatically has a much larger within-population diversity and a genetic signature that is more vague. Large conglomerate groups of populations, therefore, may appear to be highly connected but this does not mean that migration rates to each individual population are high (Wakeley & Aliacar 2001). Also, it is thought that only a small amount of gene flow is necessary to genetically homogenize populations (Slatkin 1987). Thus, migration between Indian and Pacific Ocean populations need not be high in order to produce the patterns observed in *N. brevirostris*. Nevertheless, because many reef organisms do show population structure across the IPB, *N. brevirostris* probably experiences more gene flow than these.

We know that *N. brevirostris* is an impressive disperser because it has been reported vagrantly from the Galapagos islands (Randall 2002) nearly 5,500 km from the closest permanent populations in the central Pacific. The dispersal ability of this species appears to homogenize genetic populations between the Indian and Pacific Ocean basins across a potent biogeographic barrier. But the overall phylogeographic picture indicates that long-distance dispersal is a small percentage of the total gene flow. In reality, this species may employ several dispersal strategies on multiple spatial scales, as has been suggested for other reef fishes (Planes *et al.* 2009).

3.4.3 Non-geographic clades

The mismatch distribution of the entire *N. brevirostris* data set was distinctly bimodal (Fig. 3.1). Bimodal mismatch distributions are commonly observed in small terrestrial vertebrates such as rodents (Miller *et al.* 2006; Mora *et al.* 2007) and lizards (Strasburg *et al.* 2007), with limited dispersal ability. In these cases, bimodal

mismatch distributions are associated with population fragmentation, often involving a land-based barrier to dispersal. The two modes represent two different expansion events; the mode on the left, with the least amount of pairwise differences represents a relatively recent coalescence event, while the mode on the right represents an older coalescence event (Ray et al. 2003). Importantly, a bimodal mismatch distribution does not refute expansion, rather it suggests that the rate of expansion has been slow (Ray *et al.* 2003: Excoffier *et al.* 2009), possibly because expansion was hindered by a barrier. To find this pattern in *N. brevirostris* may validate the position of Klanten *et al.* (2007) that fluctuations in the Indo-Pacific Barrier have bisected widespread *Naso* populations in the past.

Clade 2 also had a bimodal mismatch distribution and the same interpretation explained above applies. Moreover, it seems that this clade is the source of the bimodal mismatch distribution in the entire data set because clades 1 and 3 have a combined mismatch distribution that is unimodal suggesting single expansion event for these two groups. In other words, non-geographic clades and a bimodal mismatch distribution are probably the result of two separate population expansions, one much slower than the other. The expansion of clade 2 is probably older than the expansion of the other clades as indicated by clade expansion times (Fig. 3.2), not withstanding it appears to be derived in the control region gene tree (see chapter 6).

Clade 3 contains all the haplotypes shared across locations (Fig. 3.1) and is characterized by small star-shaped phylogenies indicative of demographic expansions (Castelloe & Templeton, 1994; Posada & Crandall, 2001). If positive selection is present in clade 3, as is indicated by neutrality tests, this may explain the strong characteristics of expansion. It can only be speculated what selection pressures may be favoring clade 3. When Fu's *F*s and Fu and Li's D and F tests are

all significant it may indicate that the mitochondrial genome is genetic hitchhiking on top of a nuclear gene (Fu 1997). Or, because the pressure appears to be range-wide it may be acting on a molecular pathway such as oxidative phosphorylation, as in a mitochondrial selective sweep (Gillespie 2000; Meiklejohn *et al.* 2007). However, both genetic hitchhiking and mitochondrial selective sweeps are expected to reduce genetic diversity as a sort of genetic bottleneck (Gillespie 2000), which is not consistent with the observed genetic diversity of clade 3 relative to the other clades. Because clade 3 is not any less diverse than the others I am reluctant to conclude that selection is operating here.

The fact that the non-geographic clades of *N. brevirostris* have contrasting population histories, and indeed, sum up to an odd number of clades, is at variance with the notion of temporal isolation across a low-sea-level barrier proposed by Klanten *et al.* (2007). In light of the data, a broad-scale barrier may not be necessary to explain the formation of non-geographic clades. Here I would like to present two alternative explanations that may also account for the presence of non-geographic clades:

1) Allen & Erdmann (2009) have recently reported high endemism of reef fishes from Chenderawasih Bay in the West Papuan region. This endemism is purportedly the result of historical periods when Chenderawasih Bay was isolated from the surrounding ocean for much of the last five million years, due to a complex geological mechanism (Polhemus, 2007 as referenced by Allen and Erdmann, 2009). Carpenter and Springer (2005) also propose that similar vicariance events have occurred historically within the Philippine archipelago (see also Ravago-Gotanco & Junio-Meñez 2010). Thus a small population of *Naso*, which would require much less time for genetic drift to act upon than a large ocean basin population, trapped in

an area like Cenderawasih Bay, would be able to undergo intense expansion after it is reintroduced to the ocean at large. Therefore, clades, such as clade 1, that exhibit patterns of expansion may be the result of small-scale isolations that occurred over relatively short periods of time rather than large-scale isolations that occurred over long-periods of time.

2) Klanten *et al.* (2007) note that a high dispersal potential intuitively suggests long distance dispersal, which can lead to a patchily distributed gene frequencies (Nichols & Hewitt 1994; Ibrahim *et al.* 1996). It was the belief of Klanten *et al.* that long distance dispersal between periods of population contraction and isolation that resulted in non-geographic clades, but I argue that long distance dispersal and range expansion can explain non-geographic population patterns even without the help of a barrier.

In general, range expansion can take one of two forms: first, as a wave, when expansion is accomplished primarily by short distance dispersal. Expansion waves can sometimes lead to genetic differentiation of populations at the forefront of expansion through a process known as “genetic surfing” (Excoffier & Ray 2008; Excoffier *et al.* 2009). Essentially, genetic surfing is a form of founder effect in which rare genes can achieve elevated frequencies in populations at the edge of expansion.

The second form of range expansion is called pioneer expansion and occurs primarily when long distance dispersal is pervasive. Pioneer expansion leads to the establishment of patchy colonies ahead of a general expansion wave (Nichols & Hewitt 1994). Through founder effect, pioneer colonies are genetically differentiated from parent colonies, a pattern that can persist for many generations after these

colonies are introgressed into the population at large through further expansion (Ibrahim *et al.* 1996). Furthermore, pioneer colonies can sometimes produce a second expansion wave that radiates back in the opposite direction of expansion and blocks the original expansion in a process called “embolism” (Bialozyt *et al.* 2006; Excoffier *et al.* 2009; Fayard *et al.* 2009). Considering the enormous size of the Indo-Pacific and the highly dispersive nature of marine larvae, the potential for pioneer colonization and genetic embolisms in marine populations seems high. But genetic embolisms can be overcome through high gene flow, leading to admixed lineages.

In *N. brevirostris* both forms of expansion could have occurred simultaneously. As evidenced by its bimodal mismatch distribution, expansion in clade 2 could have occurred as a slow moving expansion wave. But long distance dispersal could have led to the establishment of pioneer colonies well ahead of the main expansion wave. Genetic embolisms may have ensued but were eventually introgressed, resulting in the sympatric lineages observed today. While not an absolute barrier, the IPB may have assisted in this process by acting as a narrow migration corridor, which would have slowed migration and exacerbated the effect of genetic embolism (Excoffier *et al.* 2009).

3.5 Conclusion

Based on the congruence between the results of this study and Klanten *et al.* (2007) I reject the possibility that stochasticity and homoplasy are creating chaotic genetic patterns in *Naso* species. The similarity of genetic patterns observed argues that both *N. vlamingii* and *N. brevirostris* have had similar evolutionary histories. The proposed scenario of episodic isolation between the Indian and Pacific Ocean acting

on the population structures in *Naso* does not fit the *N. brevirostris* data well. Other explanations for the formation of non-geographic clades, such as small-scale isolations that diverge from the population at large, and range expansion are worthy of consideration.

Chapter Four

Broad-scale phylogeography of *Naso unicornis*

4.1 Introduction

In this chapter I extend the approach of Klanten *et al.* (2007) to another *Naso* species for comparative purposes. *Naso unicornis* is one of the more common *Naso* species and is found in relative abundance throughout the tropical Indo-Pacific between the Red Sea and the Marquesas. In the Pacific, it is also found at a great range of latitudes: between southern Japan (Masuda *et al.* 1984) and New South Wales (Randall *et al.* 1990). This species is cosmopolitan in its habitat preferences and is a dominant herbivore in virtually all coral reef zones (Hoey & Bellwood 2009). According to my contacts in Tonga, at certain times of the year this fish increases its fat reserves, not only making it better to eat but also indicating that in some areas it may take advantage of seasonally available food sources.

As a widespread habitat generalist and opportunist, *N. unicornis* may exhibit historical patterns of isolation, introgression and expansion heretofore thought to be essential to the formation of non-geographic population structures. *N. unicornis* is also distinct from *N. vlamingii* and *N. brevirostris* in certain aspects of its ecology, namely its diet and benthic foraging mode. The purpose of this chapter is to determine if a species with a benthic foraging ecology exhibits the same type of genetic patterns observed in its pelagic foraging congeners.

4.2 Methods

196 *N. unicornis* from 14 locations were included in this study (Seychelles, n = 16;

Reunion Island, n = 2; Rodriguez, n = 2; Oman, n = 1; Cocos Keeling, n = 15; Western Australia, n = 12; Great Barrier Reef, n = 36; Papua New Guinea, n = 4; New Caledonia, n = 1; Taiwan, n = 1; Guam, n = 40; Tonga, n = 25; Tahiti, n = 18; Hawaii, n = 24) (Fig 2.2). Samples were collected as outlined in chapter 2. Laboratory processing is also outlined in chapter 2. Data analysis was performed as outlined in chapter 3. *N. unicornis* trees are rooted with *N. lituratus* as outgroup. Most of the sequences from this study are available on GenBank, Accession Nos. FJ216829- FJ216935. All remaining sequences are available upon request.

4.3 Results

4.3.1 Diversity

279 base pairs of the mitochondrial control region were analyzed for *N. unicornis*. There were 145 parsimony informative sites and 17 singleton sites. The ratio of transitions to transversions was approximately 4 to 1. As with other *Naso* species, overall haplotype and nucleotide diversities for the mitochondrial control region were high, ($h = 0.998$, $\%\pi = 7.8$).

4.3.2 Phylogenetic Analysis

The best substitution model for *N. unicornis* was GTR+G, with a gamma shaped distribution of $\alpha = 0.436$. ML and BI identified a number of well-supported clades as well as many other lineages (Fig. 4.1). The exact number of clades that exist in this species may be 13 or higher, based on posterior probability support, however, some of these clusters contained very few individuals in our sample set, therefore only nine clades (clade A – clade I), which had at least seven representative individuals, will be

discussed further. Only 102 out of 196 individuals belonged to these nine clades, the remaining sample units form a large polytomy.

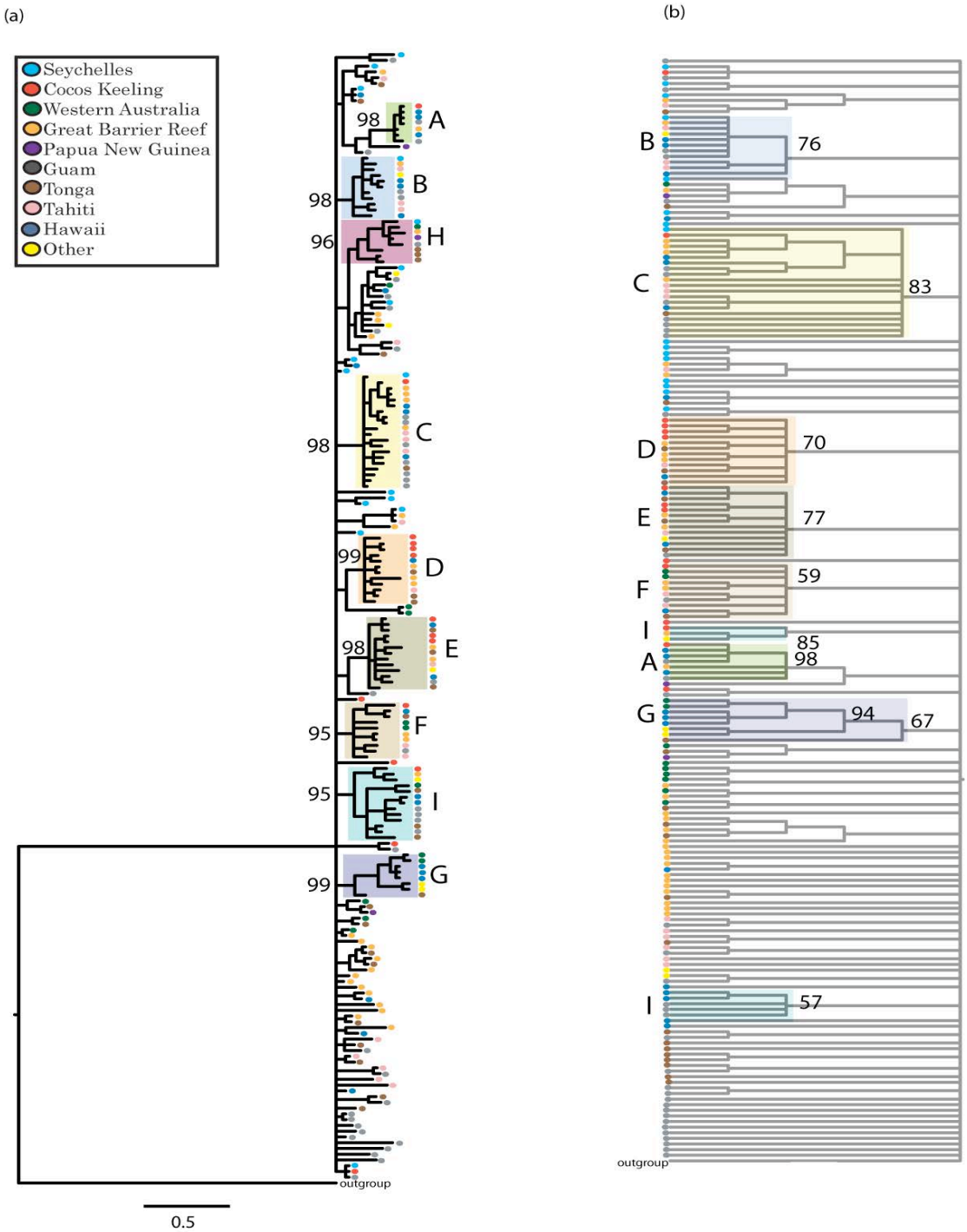


Figure 4.1 (a) BI consensus tree of 196 *N. unicornis* control region sequences, with posterior probability values for clades A-I. Values of >94 and membership >6 were the criteria used to designate clades. Color codes follow from figure 4.1. (b) ML majority rule consensus tree from 100 bootstrap replicates generated from the same data. Bootstrap support values provided for clades. Clade H is absent from this tree. Clade I is split in two, however, because the tree is largely a polytomy the separation does not represent genetic distance.

4.3.3 Population structure

Pair wise F_{st} values were small and not significant for any pair of populations (Table 4.1). Isolation-by-distance analysis conducted on these values show a negative relationship, albeit a weak and non-significant negative relationship, between genetic and geographic distances (r = -0.2716, p = 0.889). AMOVA fixation indices were small and not significant for all spatial comparisons (Table 4.2). There was no population structure detected between the Indian and Pacific oceans at the IPB (Φ_{ct} = -0.006, p = 1.0). There was no population structure detected within the Pacific Ocean between west (GBR, Guam, Tonga) and central (Tahiti, Hawaii) populations (Φ_{ct} = 0.001, p = 0.3). Likewise, there was no structure between north (Guam, Hawaii) and south (GBR, Tonga, Tahiti) Pacific populations (Φ_{ct} = 0.001, p = 0.9). AMOVA was also performed on our GBR samples, which collectively contained samples from the northern GBR (Lizard Island), the central GBR (Townsville and Orpheus Island) and southern GBR (One Tree Island). No structure was detected between the northern, central and southern GBR (Φ_{st} = 0.0256, p = 0.151).

Spatially, *N. unicornis* appears to have unstructured populations across its entire Indo-Pacific range in a pattern that resembles panmixia. Moreover, the haplotype network for this species boasts nine haplotypes from more than one location (Fig. 4.2). Five haplotypes were sampled from both the Indian and Pacific Oceans. Even remote parts of the Pacific, such as Hawaii, appeared to have high connectivity with the Indian Ocean. Thus, genetic exchange at a large spatial scale may be somewhat common *N. unicornis.* Nevertheless, when our data set was organized as clades, AMOVA fixation indices were deep and significant (Φ_{st} = 0.426, p = 0.0001).

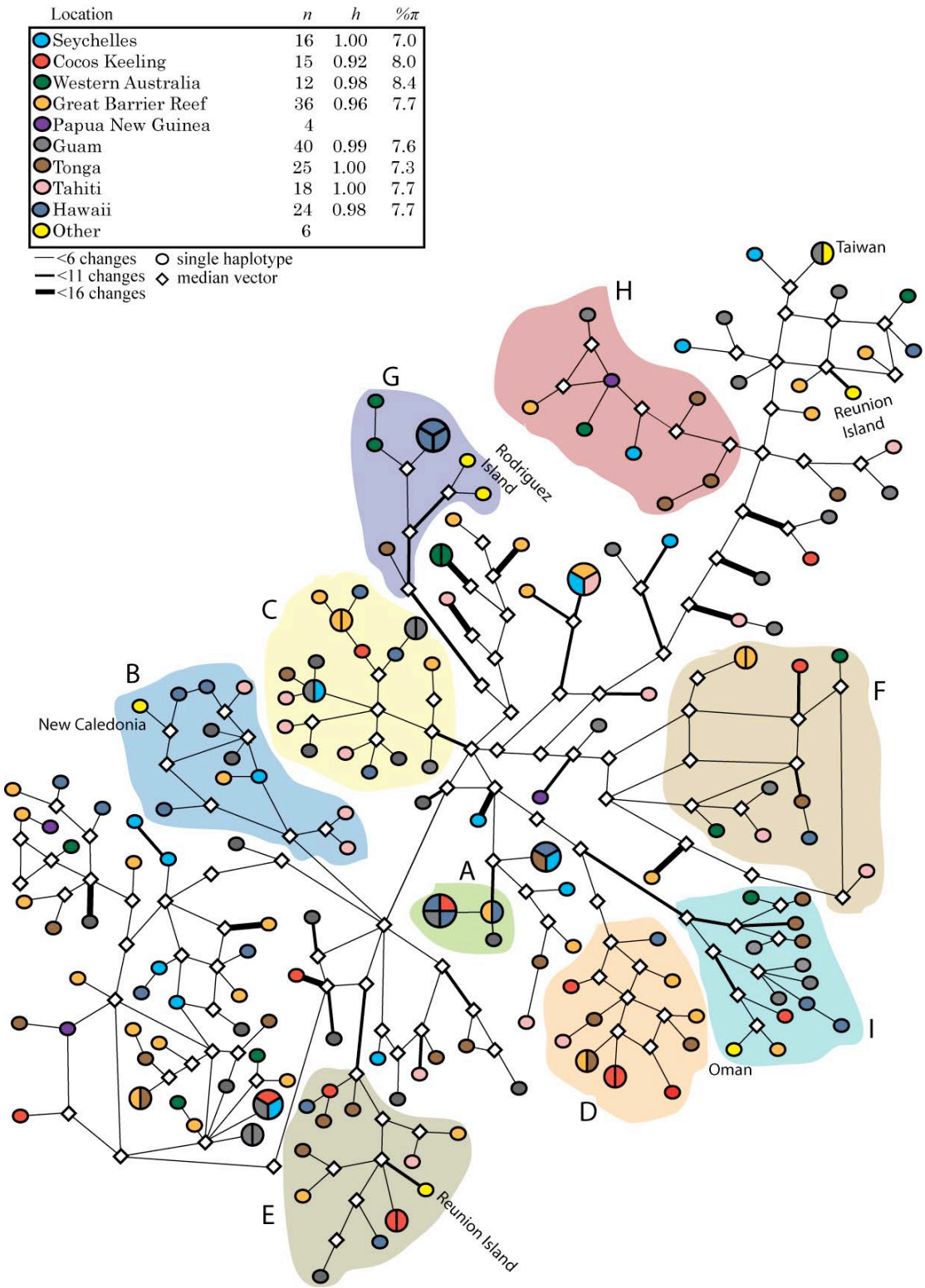


Figure 4.2 Median joining network of 196 *N. unicornis* haplotypes. For each population, sample size (*n*), haplotype diversity (*h*) and nucleotide diversity (*%π*) are provided.

Table 4.1 Pair wise population F_{st} for eight Indo-Pacific populations and p values on the above diagonal. All values are not significant.

	Seychelles	**Cocos**	**WA**	**GBR**	**Tahiti**	**Hawaii**	**Tonga**	**Guam**
Seychelles		0.08	0.16	0.85	0.36	0.25	0.45	0.79
Cocos	0.02363		0.07	0.27	0.26	0.21	0.41	0.11
W. Austr.	0.01827	0.03348		0.09	0.11	0.24	0.34	0.14
GBR	-0.01285	0.00638	0.02067		0.27	0.25	0.55	0.34
Tahiti	0.00257	0.00846	0.02123	0.00513		0.21	0.41	0.50
Hawaii	0.00859	0.01249	0.01184	0.00548	0.01004		0.23	0.13
Tonga	-0.00120	0.00050	0.00483	-0.0029	0.00062	0.00751		0.45
Guam	-0.01036	0.01532	0.01576	0.00150	-0.0019	0.01049	-0.0008	

Table 4.2 AMOVA fixation indices (Φ_{st}), percentage of genetic variation (%) and accompanying p values for population comparisons at multiple spatial scales. Overall = (eight locations Seychelles-Hawaii).

Comparison	Fixation index	%	p
Overall AMOVA	$\Phi_{st} = 0.0052$	0.52	p = 0.136
Indian Ocean vs. Pacific Ocean	$\Phi_{st} = 0.0009$ $\Phi_{sc} = 0.007$ $\Phi_{ct} = -0.006$	99.91 0.73 -0.64	p = 0.134 p = 0.078 p = 1.000
Central Pacific vs. West Pacific	$\Phi_{st} = 0.003$ $\Phi_{sc} = 0.002$ $\Phi_{ct} = 0.001$	99.61 0.21 0.18	p = 0.253 p = 0.201 p = 0.301
North Pacific vs. South Pacific	$\Phi_{st} = 0.002$ $\Phi_{sc} = 0.004$ $\Phi_{ct} = -0.001$	99.76 0.43 -0.19	p = 0.250 p = 0.221 p = 0.904
North GBR vs. Central GBR vs. South GBR	$\Phi_{st} = 0.0256$	2.5	p = 0.151
Clade vs. Clade	$\Phi_{st} = 0.426$	42.61	p < 0.001

4.3.4 Demographic History

Fu & Li's neutrality tests were not significant for the total *N. unicornis* data set, or for any of the non-geographic clades meaning that selection pressure acting on our marker is not likely. Surprisingly, however, clades E, F, G had neutrality test values that were positive, indicating that there is a disproportionate amount of polymorphism attributed to the internal branches (Fu & Li, 1993). Fu's *F*s was significantly negative for the entire data set (-23.694, $p = 0.007$) meaning that there is a strong expansion signal in the data. The mismatch distribution for the entire data set was unimodal and agreed nicely with a computer simulated expansion model frequency, further suggesting that *N. unicornis,* as a species, has experienced positive population growth for much of its species history (Fig. 4.3). The parameter τ was estimated at 23.95, with an upper and lower boundary of 18.3 and 26.3 respectively. Based on an average generation time of 21.5 years (the same generation time used for previous *Naso* species) the *N. unicornis* data set had a mean expansion time of 6.4 mya (Fig. 4.3).

With the exception of clade C, the non-geographic clades of *N. unicornis* do not show the characteristics of expanding populations and contrast starkly with the overall data set (Fig. 4.3). *N. unicornis* clades are also quite different from those of *N. brevirostris* for the same reason. While the clades of *N. brevirostris* appear to be expanding populations, *N. unicornis* clades have mismatch distributions that are exceptionally ragged and multimodal. What's more, Fu's *F*s values for clades G – H were not significant (a strong non-expansion signal).

In an expanding population, the value of τ is where the crest of the mismatch distribution falls over the x-axis (Rogers and Harpending, 1992). In those clades with highly ragged mismatch distributions, the shape of the mismatch distribution

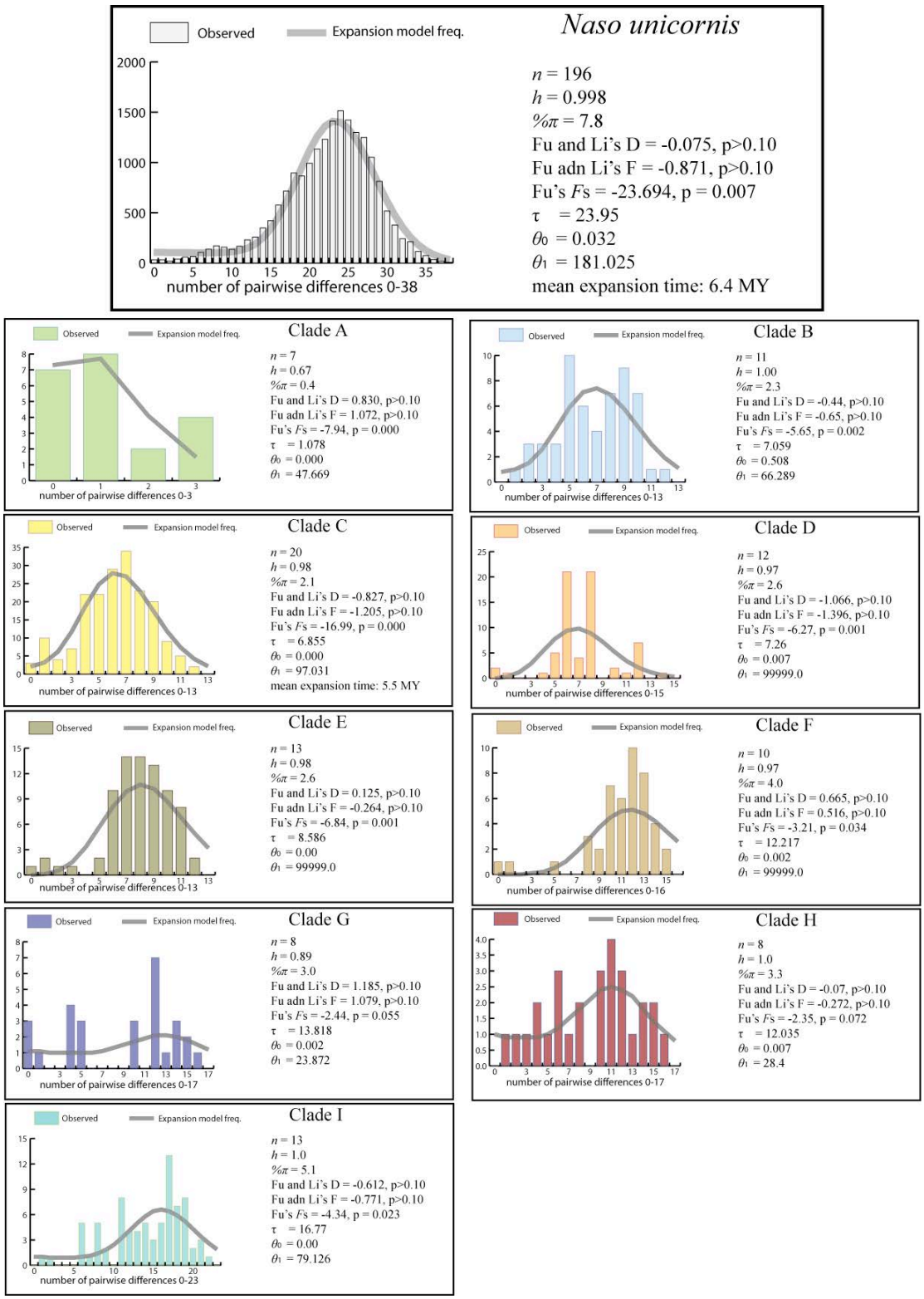


Figure 4.3 Mismatch distributions for the entire *N. unicornis* data set and for each clade individually. Also included: number (*n*), haplotype diversity (*h*), nucleotide diversity (*%π*), Fu and Li's D and F tests of selective neutrality, Fu's *F*s test of population expansion, the mutational expression τ, Population size before expansion (θ_0), present population size (θ_1) and mean expansion time in millions of years (MY) for the entire data set and clade C.

provides a poor estimate of τ because the frequency histogram actually has no crest. In general, τ is most informative when data sets exhibit a smooth and unimodal mismatch distribution, much less so where no expansion is evident (Rogers and Harpending, 1992). Therefore, of all the clade structures detected in *N. unicornis*, only for clade C was time since expansion calculated. At an age of 5.5 mya expansion in clade C appears nearly as old as the entire data set.

4.4 Discussion

In comparison with *N. brevirostris* and *N. vlamingii*, the data for *N. unicornis* is largely complementary. As with its congeners, the mitochondrial control region of *N. unicornis* had high genetic and haplotype diversity. Spatially the populations of *N. unicornis* are unstructured at all geographic scales and non-geographic clades were present. However, there are some key differences between *N. unicornis* and the other two species.

4.4.1 Spatial patterns

Although *N. brevirostris* and *N. vlamingii* lacked geographic population patterns in a general sense, both had a small number of populations that were genetically structured in pairwise comparisons. For *N. brevirostris*, specifically, the two most genetically differentiated populations were the Seychelles and French Polynesia, which are also the two most geographically distant populations. Not surprisingly, isolation-by-distance was significant for *N. brevirostris*. But *N. unicornis* populations absolutely lack pairwise differentiation (Tab. 4.1) and AMOVA failed to detect population partitions anywhere within the geographic range. So notwithstanding that

all three *Naso* species are deficient in phylogeographic structures, *N. unicornis* is the only that is free of any.

The *N. unicornis* data features a number of haplotypes that are shared across locations. The existence of haplotypes shared across vast distances can be interpreted in two different ways. Because the mitochondrial control region is one of the fastest evolving and hyper variable gene regions known (Moritz *et al.* 1987), the presence of shared haplotypes may indicate that gene flow between distant populations has occurred on a relatively recent evolutionary time scale. Alternatively, the most frequently occurring haplotypes, as well as those found at the largest geographic scales, can also be interpreted as being the oldest (Posada & Crandall 2001). In the latter scenario the shared haplotypes, separated by vast distances, may actually be much older than other haplotypes and have had more time to traverse long distances through many successive generations. In the *N. brevirostris* data set, the shared haplotypes are predominately found at or near the center of small star shaped phylogenies within the haplotype network (Fig. 3.1). This central position, with a high number of connections to other haplotypes is usually occupied by the oldest haplotypes (Castelloe & Templeton 1994). Contrastingly, in *N. unicornis* the majority of the most frequent haplotypes (those that occur more than once) are found on the periphery of the haplotype network and have only a single connection to the main body of haplotypes (Fig. 4.2). The small number of connections to these shared haplotypes indicates that they are much less likely to be the oldest (Castelloe & Templeton 1994). Thus, although both these *Naso* species show a general lack of geographic population structure, it is possible that they differ in the tempo of gene flow and expansion. Because of the positioning of shared haplotypes in the haplotype networks, distant *N. unicornis* populations may have been connected much more

recently than equivalent populations of *N. brevirostris.*

Three lines of evidence suggest that *N. unicornis* populations experience high genetic connectivity on a large spatial scale: 1) Total lack of significant fixation indices across its species range. 2) Distant populations that share haplotypes in common in a hypervariable gene region, in spite of high haplotype diversity. 3) High nucleotide diversity in populations on the periphery of the species range (Ray *et al.* 2003). However, the extent to which the genetic connectivity of a mitochondrial marker indicates demographic interdependence and migrant exchange between populations cannot be determined from this data (Hellberg 2007) and will be addressed in chapter 7.

4.4.2 Non-geographic clades

The non-geographic genetic patterns of *N. unicornis* appear different from those of its congeners. Both *N. vlamingii* and *N. brevirostris* had a small number of non-geographic clades that dominated the gene genealogy, with a few rare basal lineages that did not belong to clades. In contrast, only about half (52%) of the *N. unicornis* samples formed into about nine small non-geographic clades, with a large number of rare, basal lineages. If non-geographic clades are the result of low sea level barriers, why are there so many individuals that belong to basal lineages?

At first glance, the non-geographic patterns of *N. unicornis* appear so different that it seems they are unrelated to the patterns observed in the other species. Based on the simulations of Slatkin & Hudson (1991) we might conclude that the clades of *N. unicornis*, with ragged mismatch distributions, are no different than random samples from a panmictic population of stable size. Indeed, the branching observed here may not be more than would be expected from panmixia given our

sample. Furthermore, as Maynard-Smith (1989) once said "… it is fatally easy to read a pattern into stochastically generated data." Hence, it may be unwise to try and explain these patterns, as it would lead to over-interpretation.

Panmixia at a large geographic scale in *N. unicornis* is perhaps the best interpretation of this data and certainly cannot be ruled out, regardless of other interpretations, in light of the spatial genetic homogeneity observed in the mitochondrial control region. However, equally as conspicuous as the lack of spatial population structure in the data is a strong signal of expansion. Therefore, even if present-day *N. unicornis* meta-populations effectively comprise a genetic panmixia across the Indo-Pacific, the effects of range expansion, early in the species history, could have had a profound effect on the observed genetic patterns.

A gene genealogy that has undergone expansion displays a "comb" or "star" shape that is formed because most of the sampled lineages coalesce around the onset of expansion, at the base of the gene tree (Slatkin & Hudson 1991; Castelloe & Templeton 1994). As time progresses, coalescent events in the genealogy become less common (Harpending & Rogers 2000) resulting in a large number of unique haplotype lineages in a sample with long branch lengths (the teeth of the comb), especially if the migration rate between demes is high (Ray *et al.* 2003). If migration rates are low, or if there are physical constraints to diffusion, a greater number of recent coalescent events and shorter branches become evident in the genealogy because lineages would tend to coalesce before they migrate to another deme (Ray *et al.* 2003). At the spatial limits of expansion, for example, diffusion constraints would not have allowed new lineages to disperse before they coalesce (Ray *et al.* 2003; Excoffier 2004; Excoffier *et al.* 2009). Spatial diffusion constraints at peripheral habitat in the Indo-Pacific may have led to the formation of clades in the *N. unicornis*

mitochondrial genealogy. Under this scenario clade formation would not be due to the subdivision of pre-existing populations by arising barriers and the slow process of genetic drift or stochastic mutation but rather by a stochastic sorting of available polymorphisms into new habitat through colonization (Wakeley & Aliacar 2001). These spatially constrained lineages at the edge of tropical Indo-Pacific would have proliferated locally at first before undergoing an expansive scattering phase (Wakeley 1999; Wakeley & Aliacar 2001), thereby becoming widespread over time.

The process of clades forming as the result of spatial diffusion constraints at the limits or range expansion can probably be seen in other reef fishes, with shallower coalescence times. The parrotfish *Scarus psittacus,* for example, with a mean coalescence age of 115 kya, is much younger than *N. unicornis* and appears to have undergone large-scale range expansions much more recently (Winters *et al.* 2010). *S. psittacus* has a mitochondrial gene genealogy similarly shaped to that of *N. unicornis,* with many of the coalescent events occurring at the base of the tree and possessing a large number of rare basal lineages that stem from the original expansion. There are also many small clades that are well supported by ML and BI. Several of the clades of *S. psittacus* are widespread but some are endemic to single locations on the periphery of the species range (e.g. the Marquesas and Hawaii). Conceivably, this is what the phylogeography of *N. unicornis* may have looked like at some point in its species history, when expansion was recent, before all clade lineages eventually became widespread.

Another species with the potential for non-geographic clades is *Chromis multilineata* (Rocha *et al.* 2008). Like *N. unicornis* and *S. psittacus*, *C. multilineata* has an expanding mitochondrial genealogy, with many lineages appearing around the same time, near the onset of expansion. Presently, these lineages are mostly

segregated into different geographical areas of the tropical Atlantic but Rocha *et al.* (2008) were able to show that there is some movement of lineages between the Caribbean and the tropical southern Atlantic. Given enough time and sufficient migration, the geographic signal in *C. multilineata* populations could be lost, as it is in *N. unicornis.*

In summary, if the non-geographic clades of *N. unicornis* are not simply the result of stochastic mutation in the mtDNA, they may be lineages that colonized peripheral habitats in the Indo-Pacific as part of a rapid range expansion. These lineages then diversified on the fringes of the species range before undergoing expansions themselves. The principle difference between *N. unicornis* and some other reef fish species that undergo large scale range expansion is that *N. unicornis* is older, more dispersive, or both. In this regard it is noteworthy that the center of the haplotype network of *N. unicornis* is largely vacant with all bodies of haplotypes being appendages to a non-existent core. This is unusual because most expanding reef fish populations show very strong root haplotypes (Bay *et al.* 2004; Craig *et al.* 2007; Winters *et al.* 2010; Gaither *et al.* 2010; Eble *et al.* 2010; DiBattista *et al.* 2011). Most probably the root is no longer extant in *N. unicornis,* suggesting that the initial expansion of this species is ancient (see chapter 6).

4.4.3 A lost expansion signal

The fact that most of the clades of *N. unicornis* do not show a signal of expansion (Fig. 4.3), as the clades of *N. brevirostris* did, is perplexing because each lineage would have had to undergo expansion at some point in order to become widespread across two oceans. There are at least three possible explanations that could independently account for a lost expansion signal:

1. Inadequate sample numbers. Clade C may be the only clade with an expansion signal because it also had the most members. In other words, more sample numbers contain more data, which gives a clearer distribution. To confirm this, I constructed mismatch distributions for each sampling location (not shown) to see if these differed in shape from clade mismatch distributions. Generally, the more sample numbers a location contained the more normal the distribution became. What this means is that there is a large amount of genetic diversity even within clades that must be sampled in order to adequately represent the variation that exists in nature.

Sample numbers for most clades may be inadequate to fully capture the genetic signal of expansion that at one time took place but this does not mean that the observed raggedness is misleading. On the contrary, if the signal of expansion is difficult to detect it can only be because expansion was not recent. The many peaks at large values and general genetic diversity of each clade may be evidence that any expansion in the data is very old.

2. Demographic decline. Sharp demographic decline is a way in which a pattern of expansion can rapidly converge into one of equilibrium and population bottlenecks are thought to often generate ragged distributions with many peaks at large values (Rogers & Harpending 1992). Hence, *N. unicornis* may have experienced periodic population crashes after clades were already widespread. Such drops in population size may be related to the drying of the Sunda Shelf at times of low sea level and the loss of shallow water habitat in the IAA and other parts of the tropical Indo-Pacific (but see chapter 6).

3 Low dispersal. Range expansion is usually accompanied by demographic expansion. For this reason the signal of range expansion and demographic expansion are difficult to separate (Ray *et al.* 2003; Excoffier 2004). Yet, range expansion does not always result in a signal of demographic expansion, as in a unimodal mismatch distribution, only when migration between demes is high (Ray *et al.* 2003). Therefore, if clades became widespread through rare migration, with little genetic contact between populations thereafter, clades could have expanded across the Indo-Pacific without developing a unimodal distribution of pairwise nucleotide differences. This possibility seems questionable however, for if *N. unicornis* had subdued levels of gene flow, even if only at the largest of scales, one would expect *N. unicornis* to show a pattern of isolation by distance as was seen in *N. brevirostris*, which is not the case.

4.5 Conclusion

Like its semi-pelagic congeners, *N. unicornis* exhibited non-geographic population patterns suggesting that foraging ecology and adult habitat preference is not fundamental to the formation of non-geographic clades. The patterns of *N. unicornis* differed from those of other *Naso* both in the number of clades present and the greater number of basal lineages. Nevertheless, the explanation given for the non-geographic clades of *N. unicornis* is essentially similar to that given in chapter three, which is that clades are the result of extensive range expansions in the Indo-Pacific.

Chapter Five

Broad-scale phylogeography of sympatric sister species: *Naso hexacanthus* and *Naso caesius*

5.1 Introduction:

In the two previous chapters I have explored different ideas of how non-geographic genetic patterns might be formed. In chapter 3 geographic barriers were the focus. In chapter 4 the effect of rapid large-scale range expansion was discussed. In chapter 5 I turn my attention to another aspect of non-geographic clades: to what extent they can be treated as distinct evolutionary units. Klanten *et al.* (2007) posited that non-geographic clades might be cryptic species. If they are, some members of the genus *Naso* might deserve a closer scrutiny of taxonomy.

Cryptic speciation is becoming increasingly documented in demersal marine fishes (Lima *et al.* 2005; Hyde *et al.* 2008; Craig *et al.* 2009; Lin *et al.* 2009; Leray *et al.* 2010; DiBattista *et al.* 2011) and it may be that cryptic speciation represents a substantial portion of biodiversity in marine ichthyofaunas (Rocha & Bowen 2008). When selection pressures constrain morphological variation lineages may become cryptic through genetic drift and evolutionary divergence in the prolonged absence of population connectivity, also known as stabilizing selection. This process may be facilitated by geological barriers that bisect previously continuous populations, such as the Isthmus of Panama (Craig *et al.* 2009), occur due to a combination of geographical and ecological factors (Leray *et al.* 2010), or habitat partitioning (Hyde *et al.* 2008). Regardless, the issue of cryptic speciation in demersal fishes is typically a matter of spatial segregation of populations. However, one notable exception to this

is also one of the most well-known and extensively examined case of cryptic speciation in a marine fish.

Bonefishes of the genus *Albula* were previously regarded as a single circumtropically distributed species (*Albula vulpes*) until eight divergent lineages were detected indicating that this was actually a putative cryptic species complex (Colborn *et al.* 2001). Many of these lineages are now considered as species (Hidaka *et al.* 2008; Pheiler *et al.* 2008; Wallace & Tringali 2010). Moreover, some of the lineages uncovered by Colborn *et al.* 2001 were sympatric, not entirely dissimilar to the non-geographic clades of *Naso*.

Perhaps, like bonefishes, widespread *Naso* are complexes of cryptic species or widespread hybrid swarms that were once cryptic species. Testing this hypothesis would be very difficult considering that diversification most likely occurred long ago. However, one approach to testing whether non-geographic clades are in fact cryptic species is to look for analogs of non-geographic clades. For example, the sister species pair *Naso hexacanthus* and *Naso caesius* are two semi-pelagic unicornfishes with widely overlapping geographic distributions in the Indo-Pacific. Both species are very similar morphologically. Ecologically they are effectively identical and are commonly observed foraging in heterospecific schools (Randall 2002). Furthermore, these two species are purported to have diverged relatively recently. According to the molecular dating of Klanten *et al.* (2004), the split between *N. hexacanthus* and *N. caesius* is estimated between 3.2 and 2.7 mya, which would make them of similar age or younger than many of the non-geographic clades of their congeners. Therefore, the relationship between these two co-occurring sister taxa may in fact be analogous with the non-geographic clades observed in other *Naso* species.

In this chapter, samples of *N. hexacanthus* and *N. caesius* are submitted to the same phylogeographic methods used in previous chapters except that both are analyzed jointly in the same data set, as if they were the same species. Previous genetic studies have been able to distinguish these two species using allozyme loci (Dayton *et al.* 1994; Dayton 2001) and mtDNA and nuDNA (Klanten *et al.* 2004) and there are a number of morphological characters that are diagnostic to each (Randall & Bell 1992), therefore my *a priori* prediction is that *N. hexacanthus* and *N. caesius* will segregate into reciprocally monophyletic units non-geographically. If this prediction is correct and the relationship between these two species is similar to that of non-geographic clades found in other species, then the non-geographic clades described in previous chapters may deserve greater taxonomic scrutiny, as has been given to bonefishes.

5.2 Methods

92 *N. hexacanthus* from eight locations, 26 *N. caesius* from four locations were used in this study (Fig. 2.3). Phylogenetic trees were rooted with *N. brevirostris* as outgroup. Laboratory procedures are outlined in chapter 2 and analytical methods for this chapter are the same as in chapter 3.

5.3 Results

5.3.1 Diversity

244 base pairs of the mitochondrial control region were used for analysis. Excluding a small segment of each sequence, between 9 and12 base pairs, containing many small indels, which was deleted to improve the alignment. This may have been a mutation hotspot (Galtier *et al.* 2009), which would be saturated. However, identical

haplotypes were not shown as identical in the haplotype network unless the raw sequences (with the segment included) were the same. There were 136 polymorphic sites with 119 parsimony informative sites and 23 singletons. The best substitution model for the data was GTR+I+G, proportion of invariable sites = 0.2010 and a gamma shape parameter $\alpha = 0.594$ Haplotype and nucleotide diversities for *N. hexacanthus* and *N. caesius* were respectively $h = 1.0$; $\%\pi = 11.0$ and $h = 0.99$; $\%\pi = 10.4$. Combined haplotype and nucleotide diversities for both species were $h = 0.99$ and $\%\pi = 10.5$.

5.3.2 Phylogenetic analysis

Both ML and BI failed to resolve a reciprocally monophyletic relationship between *N. hexacanthus* and *N. caesius* based on mitochondrial control region sequences (Fig. 5.1). In some instances *N. hexacanthus* haplotypes and *N. caesius* haplotypes shared close identity to each other based on posterior probability support and bootstrap support but no identical haplotypes between species were observed. The combined data set possessed a single large non-geographic clade, similar to those observed in *N. brevirostris* and will hereafter be referred to as clade X. There were also a number of smaller non-geographic structures that were well supported, that were similar to the non-geographic clades of *N. unicornis* but most of these contained few individuals and will not be given attention hereafter.

5.3.3 Population structure

Pair wise F_{st} values are reported in table 5.1 and AMOVA fixation indices are reported in table 5.2. All analyses were performed on the combined data set from

both species. No pairwise population structures were observed. Isolation-by-distance was not significant, was slightly negative (r = -0.1344, p = 0.6).

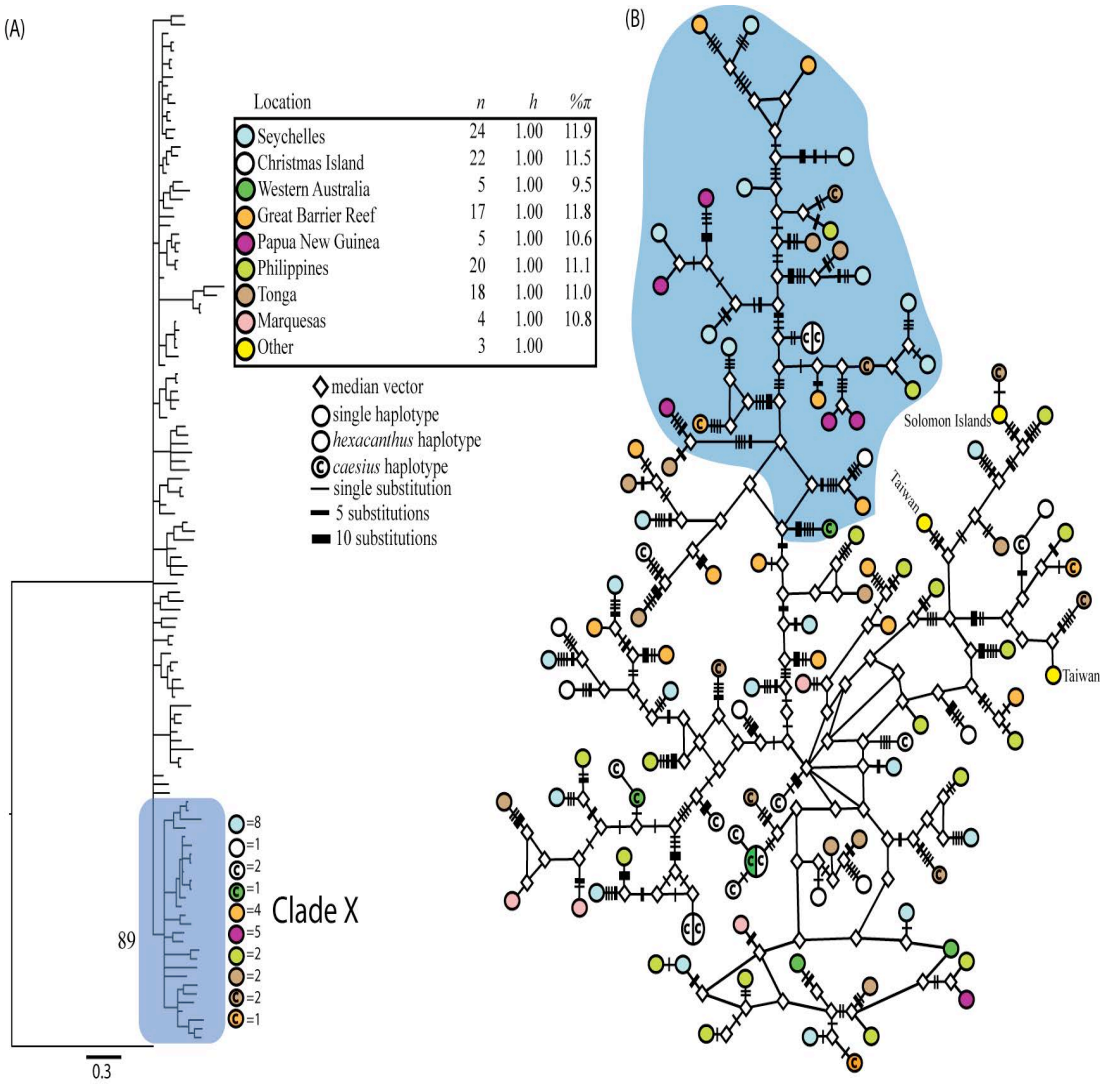


Figure 5.1: (a) Consensus tree of 19,350 phylogenetic trees generated from BI inference with selected posterior probability values. Shaded area is a cluster of haplotypes with reasonably high support, clade X. Colored circles indicate the composition of clade X. (b) minimum spanning network of 92 *N. hexacanthus* haplotypes, 26 *N. caesius* haplotypes, which contain and encircled "c". Shaded area corresponds to clade X.

Note, however, that some populations: the Marquesas, Western Australia and Papua New Guinea had very few representative samples ($n \leq 5$) and were excluded from the analysis. This is unfortunate but these two species are not as common as others, are

challenging to capture by spearing and therefore optimal sample numbers could not be obtained from every location.

AMOVA of spatial comparisons for the combined data set yielded small but significant fixation indices: among all populations (Φ_{st} = 0.0141, p = 0.023). Such small values probably do not reflect any substantial restrictions on gene flow because the genetic diversity (%π; Fig 5.1; Fig 5.3) is more or less uniform across all sampled populations and absolute diversity takes longer to equilibrate between populations than fixation index estimations of differentiation (Birky et al. 1989; Pannell and Charlesworth, 2000). In contrast, AMOVA found no significant population structure between ocean basins (Φ_{ct} = -0.099, p = 0.8). Also, there was no structure detected between species (*N. hexacanthus* vs. *N. caesius*; Φ_{st} = 0.00579, p = 0.15). Thus, there was greater genetic differentiation observed between locations than between species.

Table 5.1: Combined *N. hexacanthus* and *N. caesius* pairwise F_{st} comparisons for eight sampling locations: Seychelles, Christmas Island, Great Barrier Reef (GBR), Philippines and Tonga (see Fig. 1). Probability values for each F_{st} are given on the upper diagonal. All values are not significant.

	S	**CI**	**GBR**	**PH**	**TG**
Seychelles		0.073	0.633	0.082	0.852
Christmas	0.015		0.073	0.363	0.472
GBR	-0.005	0.020		0.075	0.766
Philippines	0.013	0.002	0.019		0.488
Tonga	-0.009	-0.001	-0.009	0.005	

Table 5.2: AMOVA fixation indices (Φ_{st}), percentage of genetic variation (%) and accompanying p values for three population analyses: Overall (eight locations Seychelles-Marquesas), Indian Ocean vs. Pacific Ocean and *N. hexacanthus* vs. *N. caesius.*

Comparison	Fixation index	%	p
Overall AMOVA	Φ_{st} = 0.0141	1.41	p = 0.023
Indian Ocean vs. Pacific Ocean	Φ_{st} = 0.00991 Φ_{sc} = 0.01963 Φ_{ct} = -0.099	99.01 1.98 -0.99	p = 0.024 p = 0.010 p = 0.815
N. hexacanthus vs. *N. caesius*	Φ_{st} = 0.00579	0.58	p = 0.151

5.3.4 Demographic History

For the combined data set, τ was estimated at 36.539, with an upper and lower bound of 38.627 and 29.383 respectively. Mean expansion time was estimated at 10.1 mya, a very old age considering the fossil calibrated dating of Klanten *et al.* (2004) placed the split between these two species around 3 mya. The mismatch distribution for the combined data set was unimodal and Fu's *F*s value was significantly negative (*F*s = -23.938, p = 0.002), all characteristics of population expansion (Fig. 5.2). Non-significant Fu & Li's D and F values suggest no signal of selection acting upon this marker in this taxon. A mismatch distribution was also constructed for the non-geographic clade X (Fig. 5.2). The characteristics of expansion were weaker in the non-geographic clade than in the entire combined data set (Fu's *F*s = -9.644, p = 0.007). Mean expansion time for the non-geographic clade was 12.01 MY. Clearly, this age is exaggerated due to raggedness in the mismatch distribution and a poor estimation of τ (see chapter 4).

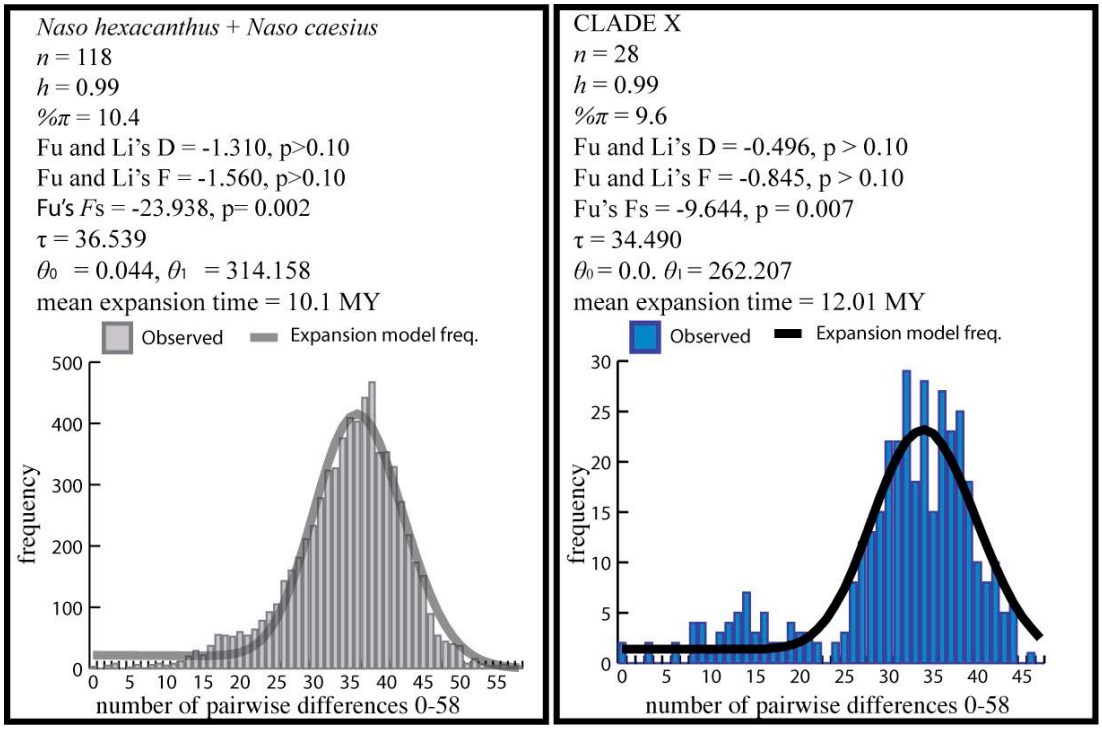


Figure 5.2: Mismatch distribution and expansion dating for 92 *N. hexacanthus* and 26 *N. caesius* haplotypes and that for the non-geographic clade X. Neutrality test results, coalescence values: τ, θ_0 and θ_1, which are respectively the evolutionary age in units of generations, a metric of the effective female population size at the time of expansion and a metric of the current female population size. Mismatch distribution and mean expansion time based on a female generation time of 21.5 years.

5.4 Discussion

5.4.1 Non-monophyly

The main conclusion of this chapter is that *N. hexacanthus* and *N. caesius* are not reciprocally monophyletic in this mtDNA marker. Possibly there has been horizontal gene transfer between these two taxa but whether hybridization is historical or ongoing is presently uncertain. Incomplete lineage sorting is also likely but again the time scale is not readily apparent. Nevertheless, two observations about the morphology and behavior are worth mentioning. First, *N. hexacanthus* and *N. caesius* were originally identified as different species because each has unique male spawning coloration (Randall & Bell 1992). Additionally, the caudal spines of each

species, which probably play a role in reproductive behavior, such as competition for reproductive status (pers. observ.), are differently shaped (Randall & Bell, 1992). These differences in mating morphology suggest that there are prezygotic barriers, in the form of behavioral isolation (Coyne & Orr 2004) that may safeguard against hybridization. However, and secondly, Randall & Bell (1992) originally noticed the distinct spawning colors because they were visible in the same spawning aggregation! This begs the question, how do broadcast spawning fishes avoid accidental hybrid fertilization of gametes when they form heterospecific spawning aggregations? Whether there are presently genetic postzygotic barriers that help maintain the species boundary cannot be determined from the present data and is beyond the scope of this thesis but it seems that *N. hexacanthus* and *N. caesius* are incompletely speciated (De Queiroz, 2007).

Given the phylogeography of *N. hexacanthus* and *N. caesius* it is unlikely that the non-geographic clades of other members of the genus are cryptic species. I do not wish to belabor this point but it suffices me to say that my expectations of monophyly for *N. hexacanthus* and *N. caesius* were wrong and that evolution in this group of fishes is more complex than I originally surmised.

5.4.2 Spatial patterns

As in previously studied *Naso* species, *N. hexacanthus* and *N. caesius* lacked spatial population structures and probably experience high levels of gene flow across large distances. Some of the pair wise F_{st} values were significant but most were only marginally so and there was no discernible pattern to the significant values. I suspect that, were additional samples available for some populations, such structures would not appear. Also included in this chapter is a figure (Fig. 5.3) depicting the

distribution of nucleotide diversity of all four *Naso* taxa across the Indo-Pacific. These measures of absolute diversity are important because locations that receive a greater number of migrants will generally have higher genetic diversities (Ray *et al.* 2003).

In Fig. 5.3, *N. vlamingii* visually shows a trend of decreasing diversity from west to east, except for a large peak in genetic diversity in the east Indian Ocean at Cocos Keeling and Christmas Island. Both of these islands are very small and have limited reef area, so it might seem surprising that they would have higher genetic diversities than large reef areas such as the Seychelles or the GBR. Logically, the elevated genetic diversity of these populations points towards a higher traffic of larval migrants. The same general pattern of nucleotide diversity is seen in *N. brevirostris,* only less dramatic. Conversely, *N. unicornis* and *N. hexacanthus + caesius* do not show visual longitudinal trends in nucleotide diversity. These species appear closer to achieving equilibrium of genetic diversity across their species ranges and probably experience greater gene flow between populations. Note: the nucleotide diversity of one mitochondrial marker is hardly representative of total genetic diversity, however, comparisons of genetic diversity in the mtDNA and nuclear microsatellites in chapter seven might suggest that this is not a bad proxy (see also Johannesson & Andre 2006 for correlations of mtDNA and nuDNA diversity).

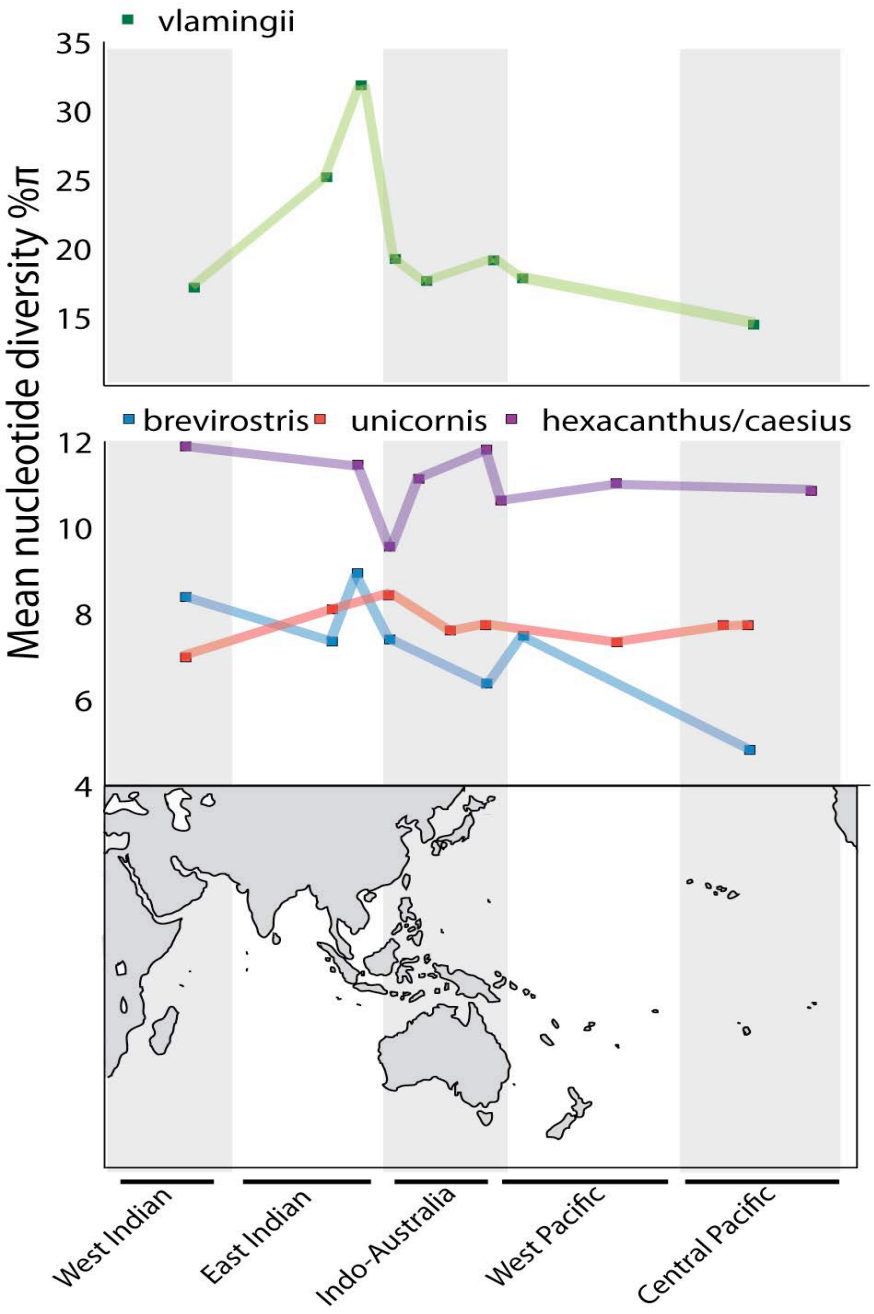


Figure 5.3: Trends in nucleotide diversity of the mitochondrial control region longitudinally across the Indo-Pacific.

If there is little genetic distinction between *N. hexacanthus* and *N. caesius* then it is a mystery why these two species have different geographic distributions. while *N. hexacanthus* is widespread between the Red Sea and Pitcairn, in the south central Pacific (Randall 2002), *N. caesius* has an irregular distribution and is comparatively rare. *N. caesius* is primarily a Pacific plate species, it is also in the east

Indian Ocean but is absent from the west Indian Ocean and though it is expected in the IAA, it has never been documented in Indonesia, New Guinea or the Philippines (Randall 2002; Allan & Erdmann, 2009; J.H. Choat, personal communication).

Ecologically there seems to be no reason why *N. hexacanthus* is more widespread or abundant than *N. caesius*. It is possible that there are underlying genetic reasons for differences in geographic distribution. For instance, if these species do currently hybridize the direction of hybridization could be asymmetrical as has been shown for greenlings (Crow *et al.* 2010) and sea urchins (Zigler *et al.* 2005; Geyer & Lessios 2009). Male *N. hexacanthus* could have more success producing hybrid offspring with the opposite species than *N. caesius* males. *N. hexacanthus* might also have more success back-crossing with hybrids, as has been shown for breams (Roberts *et al.* 2009), such that any *N. caesius* would be virtually unable to colonize areas already populated by *N. hexacanthus*. Nevertheless, this remains speculation without substantiating evidence. Moreover, no documented hybrids of these two species have ever been reported.

5.4.3 Non-geographic clades

There are many similarities between the phylogeographic patterns of *N. unicornis* and *N. hexacanthus* and *N. caesius*. For example, they possess many non-geographic clades and, in the same fashion as *N. unicornis*, these structures appear to radiate from an early expansion from which the root is noticeably absent. But the sample size for *N. hexacanthus* and *N. caesius* is limited and because most of these non-geographic clades are not well represented I overlook them, for now. The large clade X may be no different than the smaller clades and by some random sampling bias is

over-represented in the data. Apart from this possibility, clade X is a prominent genetic feature and comprises about 25% of all individuals.

5.5 Conclusion

The phenomenon of non-geographic clades, as observed in the mtDNA of the genus *Naso* does not appear to be grounds for a taxonomic delimitation of cryptic species, in contrast to the case of the bonefishes (*Albula*) (Colborn *et al.* 2001). Further, it seems inappropriate to treat the non-geographic clades of *Naso* as distinct evolutionary units. If anything, the data from this chapter has called into question the existing species boundary between *N. hexacanthus* and *N. caesius*. In reality, these could be two color variants of the same species. Or conversely, incomplete lineage sorting may be obscuring the species boundary but this would suggest very recent divergence if they are not sorted by the fast-evolving mitochondrial control region. If they are recently diverged species the possibility of contemporary hybridization between these two species is also possible but cannot adequately be confronted using the present data.

Chapter Six

Molecular dating of non-geographic clades

6.1 Introduction

Understanding the evolutionary significance of non-geographic population structuring in coral reef fishes rests heavily upon our ability to properly evaluate the age of these structures and identify the time periods and circumstances under which they arose. Up until this point we have used a substitution rate of 12.9% for variable sites for noncoding mtDNA following the methods of Klanten *et al.* (2007). This substitution rate has provided a heuristic approximation of time since expansion but must be interpreted with caution because this rate is somewhat arbitrary and is based on an unrelated taxon, swordfish (*Xiphias gladius*).

There are several approaches to estimating the age of taxa and several programs that perform molecular dating. Some of these methods are genealogy samplers based upon ML or BI, which mine much more data from a sample than can be expressed in a pairwise mismatch distribution (Kuhner 2009). It is possible, for example, to estimate molecular age using a coalescence approach, where time is counted from the present backwards. This can be done within a ML or BI framework in a program like LAMARC (Kuhner 2006) or BEAST (Drummond & Rambaut, 2007). However, this approach assumes a molecular clock and requires a user specified substitution rate (μ). Although there are many molecular clock calibrations for several mtDNA markers in marine fish (Lessios 2008), to date no calibration studies exist for the genus *Naso* or the family Acanthuridae.

An alternative to using a fixed substitution rate is to use a fossil calibration at the root of a phylogenetic tree. Not all organisms have a substantial fossil record but Acanthurid reef fishes have a considerable fossil fauna, particularly from the Eocene (Tyler 2000; Bellwood & Wainwright 2002). A previous study of molecular dating on the genus *Naso* (Klanten *et al.* 2004) also serves as a comparison by which to measure the results. Here I assess the relative ages of species and non-geographic clades of five *Naso* species: *N. brevirostris, N. caesius, N. hexacanthus, N. unicornis* and *N. vlamingii* in a comparative manner in a single data set (I also included a few sequences of *N. lituratus* to give a more complete taxon sampling for the ingroup) using fossil calibrated dating. These species were selected for their known non-geographic clades and comprise two out of five major lineages in the overall *Naso* phylogeny (Klanten *et al.* 2004). An additional species, *Naso lopezi*, which belongs to a third clade, was used as a secondary outgroup. Samples from all species were compiled into a single data set and a single gene tree on which nodes could be dated in the program BEAST 1.4.8 (Drummond & Rambaut, 2007). The advantage of using BEAST is the option of estimating age using BI with a relaxed molecular clock, which has very few assumptions, tolerates long branches, does not assume a particular tree topology and allows substitution rates to vary throughout the tree and can be used with a strong root prior (preferably a fossil calibration) at the basal node of the tree.

6.2 Methods

Laboratory methods for this section are described in chapter 2.

6.2.1 Phylogenetic comparisons of species and clades

First, an alignment of 142 sequences of cytochrome oxidase subunit 1 (COI) [*N. brevirostris* = 27, *N. caesius* = 7, *N. hexacanthus* = 24, *N. vlamingii* = 16, *N. unicornis* = 63, *N. lituratus* = 5] was created in BIOEDIT version 7.0.9.0 (Hall 1999). This was the only marker used for the phylogeny. Ideally at least one other unlinked nuclear marker would also be used, however, the utility of a nuclear marker for assessing non-geographic clade structure is questionable given recombination and typically less polymorphism than in mtDNA. Some initial experimentation was done with the ETS2 intron (Klanten *et al.* 2004) and the scDNA Enolase gene (Kelly & Palumbi 2009) with limited success. In the end, the time and resources required to develop a useful nuclear marker in this regard proved to be beyond the scope of this thesis.

Samples for the COI tree were not selected based on relationships observed in the control region sequences as this might bias results. Rather, because clades appear to be ubiquitous throughout all locations, samples from one or two locations were chosen for all species. *N. vlamingii* samples are from the GBR, *N. brevirostris* samples are from the Seychelles, *N. hexacanthus* samples are from the Philippines and PNG, *N. caesius* samples are from Tonga and *N. unicornis* samples are from Hawaii and Tonga. For this chapter I also constructed a control region phylogeny for *N. vlamingii* (not shown), from the GenBank sequences of Klanten *et al.* (2007). The purpose of this was to enable comparisons between the COI and control region topologies. And for the construction of Bayesian Skyline plots (see below).

As the outgroup for the tree, a *Paracanthurus hepatus* COI sequence was acquired from GenBank (accession: FJ583809). The higher acanthurines (*Paracanthurus* + *Zebrasoma* + *Acanthurus* + *Ctenochaetus*) all posses a shared

derived folding spine on the caudal peduncle, a character that is present in the putative primitive sister taxa, *Proacanthurus* of Eocene Monte Bolca, (Tyler 2000). Divergence between the higher and lower acanthurids based on caudal spine morphology therefore becomes a working minimum fossil age calibration for our tree (~ 50 mya) and is the same as used in Klanten *et al.* (2004). As a second outgroup we also acquired a single COI sequence of *N. lopezi* from genbank (accession: NC_009853), which belongs to an exterior clade in the *Naso* phylogeny (Klanten *et al.* 2004).

The best substitution model for the data was evaluated in jMODELTEST 0.1.1 (Posada 2008) using a ML approach and the Akaike information criterion. An initial starting tree for BEAST was first constructing using Bayesian inference in MRBAYES version 3.1 with 25 million iterations, sampling every 100 generations and a burnin of 20%. MRBAYES was run under several different conditions of data partitions and substitution parameters, which were assessed for appropriateness *post hoc* using log Bayes factor comparisons (computed in TRACER 1.4; Rambaut & Drummond 2007). A 50% majority rule MRBAYES consensus tree of the best 16,000 trees was converted into ultrametric format in R, 2.12.0 (http://www.Rproject.org; Ihaka & Gentleman,1996) using the APE package (commands written by E. Paradis & D. Lawson). For molecular dating, superfluous taxa were pruned from the tree (also using APE), final $n = 65$.

6.2.2 Molecular dating

BEAST analysis aimed to date the major nodes in the tree and was run under a number of different conditions including different tree priors, root priors and monophyly constraints that were assessed for appropriateness using Bayes factor

comparisons in TRACER. The final BEAST analysis was run for 20 million iterations with a sample taken every 1000 generations and a burnin of 10%. An exponential root prior was used for the tree with a zero offset of 50 million years and an exponential mean of five million years, which allowed the root calibration to wander back to the KT boundary ~65 million years ago (mya). The node at the base of the second outgroup, *N. lopezi*, was also constrained with a normal distribution prior with a mean age of 40 mya, in following with the dating of Klanten *et al.* (2004). A third node constraint of 30 mya and a normal distribution was used at the base of the ingroup taxa, also following the dating of Klanten *et al.* (2004). Both normally distributed node constraints were given a liberal standard deviation of five million years to allow flexibility in these dates.

A problem arises when trying to select a speciation model for the tree because this data has elements of both a species level phylogeny and an intraspecific phylogeny. There is a potentially large disparity between the nucleotide substitution rates within species and among species, the former being much higher. This phenomenon has been attributed to purifying selection (Ho *et al.* 2005; Ho *et al.* 2007) and ancestral polymorphism (Peterson & Masel 2009; Charlesworth 2010). Regardless of the cause, coalescent tree priors that estimate coalescence age as a function of effective population size are not suitable for species level phylogenies (BEAST manual, Rambaut & Drummond 2007), which leaves only two models: the Yule process (pure birth) model and the birth-death model provided by BEAST. Data was run under both models and the appropriateness of each were similar based on Bayes factor comparisons, however, because extinction rates cannot be accurately estimated from trees without the incorporation of ancient DNA (Rabosky 2009) here the simpler, Yule process model, was preferred.

6.2.3 Bayesian Skyline Plots

Following molecular dating with COI, data was reanalyzed using the Bayesian Skyline function in BEAST, which generates a posterior distribution of effective population size through time from sequence data (Drummond *et al.* 2005). For this analysis root age was constrained for taxa and clades based on the previously estimated node ages. Both COI sequences (from this chapter) and control region data (from previous chapters) sets were analyzed for each taxon, including *Naso vlamingii*. Both markers yielded similar results but control region skyline plots had greater resolution. Therefore, only control region skyline plots are reported hereafter (All available control region sequences were included for Bayesian Skyline analysis). Three replicate runs for each taxon were run with MCMC chains 10 million generations long, which were combined *post hoc* into a single file in LOGCOMBINER (Rambaut & Drummond 2007) and given a 10% burn-in after combination. Analyses were run using ten groupings in a piece-wise-constant model, which is suitable to a wide range of demographic scenarios (Drummond *et al.* 2005) but assumes that population size remains constant between coalescent events. The alternative to the piece-wise-constant model is the piece-wise-linear model, which allows population size to increase between coalescent events, but which requires a user defined substitution rate. Some experimentation with an arbitrary mutation rate of 3.6% per million years (Donaldson & Wilson 1999) was done for control region skyline plots in order to compare results with plots from another reef fish study that used this rate (Visram *et al.* 2010). The results of these explorations (not shown) confirm that demographic expansion in *Naso* is much older than the late Pleistocene expansion seen in parrotfishes (Visram *et al.* 2010; Winters *et al.* 2010).

6.3 Results

6.3.1 Phylogenetics results

I resolved 601 bp of the COI region for 142 *Naso* samples from six different species, excluding outgroups. Out of 601 sites 155 were polymorphic, 146 of these were parsimony informative and nine were singletons. The best substitution model for the data was TrN+G with a gamma shape parameter $\alpha = 0.1160$. The second best model was GTR+G with a gamma shape parameter $\alpha = 0.1140$. Bayes factor comparisons between runs of MRBAYES favored a more complex model, where all three codon positions were individually partitioned, each with 6 substitution rate parameters (Nst=6). Nevertheless, the tree topology was very consistent between runs and parameter explorations.

Resolving the phylogenetic relationships of the ingroup taxa was not the purpose of this chapter, nevertheless, resulting phylogenetic structure was complimentary to that of Klanten *et al.* (2004). *N. unicornis* and *N. lituratus* were well resolved sister taxa as were *N. brevirostris* and *N. vlamingii.* However, *N. hexacanthus* and *N. caesius* were not well resolved with respect to each other (Fig. 6.1). Non-geographic clades were visible in the COI sequences, although not to the same degree of resolution as observed for the control region in previous chapters. In *N. brevirostris,* clades 1 and 3 were recovered as a single clade, while clade 2 individuals were basal. In *N. unicornis,* five of the original nine clades were observed. In *N. vlamingii,* clades 1, 2 and 3, from Klanten *et al.* (2007) were observed but clade 4 sequences were basal and did not form a clade. In *N. hexacanthus* and *N. caesius,* clade X was not well resolved, being split near the base of this clade (Xa and Xb). Surprisingly, other supported structures, which were not

apparent in the control region appeared in COI in *N. hexacanthus* and *N. caesius*. While the nature of these new structures was enigmatic, I decided to name them clades Y and Z and date them along with the rest of the clades (Fig. 6.1).

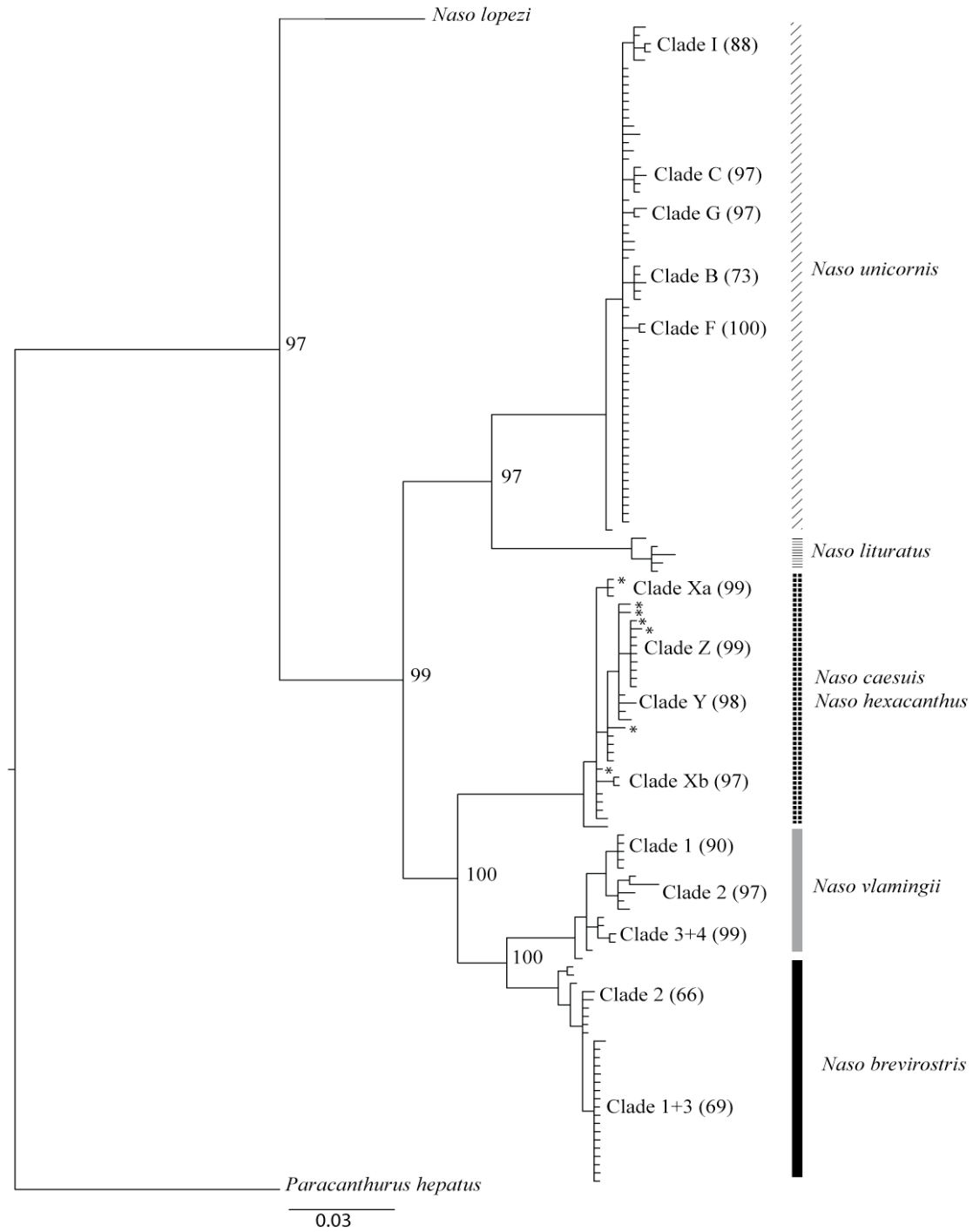


Figure 6.1: Phylogenetic consensus tree from16,000 trees computed using MrBayes for six ingroup *Naso* species. Posterior probability values are given at internal nodes

and in parentheses for each identified clade. *N. caesius* individuals are marked with *.

6.3.2 Molecular dating of deep phylogenetic nodes and species divergences

The mean rate of nucleotide substitutions averaged across all COI lineages was 0.0018 per million years and the speciation rate, or the rate at which new lineages emerged from parent lineages under the Yule process model, was 0.111 per million years. Node ages for the tree can be viewed in Figure 6.2 and Table 6.1. In general, the results were complimentary with the fossil dating of Klanten *et al.* (2004) and most estimations fell within, or just outside of the ranges of Klanten *et al.* Mean age for the root was resolved at 50.4 mya, with an upper 95% posterior density of 57.9 mya. The mean node age at the base of the second outgroup, *N. lopezi*, was estimated at 41.7 mya and the node at the base of the ingroup had a mean age of 28.3 mya. The mean divergence estimate between sister taxa *N. unicornis* and *N. lituratus* was 21.7 mya, compared to 18.5 mya from Klanten *et al.* (2004). The mean estimate for the node at the base of the *brevirostris, caesius, hexacanthus, vlamingii* complex was 26.3 mya, compared to 21.7 from Klanten *et al.* The split between *N. brev*irostris and *N. vlamingii* had a mean age of 20 mya.

6.3.3 Molecular dating of species ages

Molecular ages for root nodes of the intraspecific phylogenies (which are not the same as species divergence times noted above) of *N. brevirostris* and *N. vlamingii* are surprisingly similar to the previously estimated expansion times based on mismatch distributions, 11.5 and 7.2 mya respectively (chapter 3; Klanten *et al.* 2007). In contrast the molecular age for *N. unicornis* was nearly three times as old as previously estimated in chapter 4 (16.4 compared to 6.1 mya) and may indicate a

slower mutation rate in this species. This early-middle Miocene age is consistent with my earlier suspicion that this species is very old, because it lacks a visible root haplotype (Fig 4.2). The mean node age for *N. hexacanthus* and *N. caesius* taxon was estimated at 12.5 million years ago, much older than Klanten et al. (2004)'s divergence of the two species, which was estimated at 2.9 mya. Like *N. unicornis* the joint haplotype network for these two species also lacked a visible root haplotype (Fig 5.1) and suggests that this is a very old taxon.

6.3.4 Molecular dating of non-geographic clade ages

Clade ages were highly variable. For *N. unicornis*, the oldest was clade I with a mean age of 13.3 mya, followed by clade B (8.6 mya), clade C (6.9 mya), clade G (3.9 mya) and Clade F (0.3 mya). Ages for clades G and F should be interpreted cautiously because they only had two representatives in the COI data set and may not coalesce near the base of the clade. The two individuals from clade F in particular had a very shallow coalescence even in the control region phylogeny. Unlike the phylogeny from chapter 3, *N. brevirostris* clade 2 was basal in the tree. This COI topology is probably more accurate than the control region topology presented in chapter 3 because clade 2 is the oldest and the most genetically diverse. *N. brevirostris* clades 1 and 3 comprised a single clade in the COI phylogeny. Clade 1+3 had a coalescence age of 8.9 mya, much older than the previous clade expansion age of 3.3 mya. Clades 1 and 2 of *N. vlamingii* had coalescence ages of 4.7 and 2.8 mya respectively. The combined clades 3 and 4 of *N. vlamingii* had a mean age of 4.4 mya. Clades Y and Z of the *N. hexacanthus* and *N. caesius* complex had ages of 11.5 and 7 mya respectively.

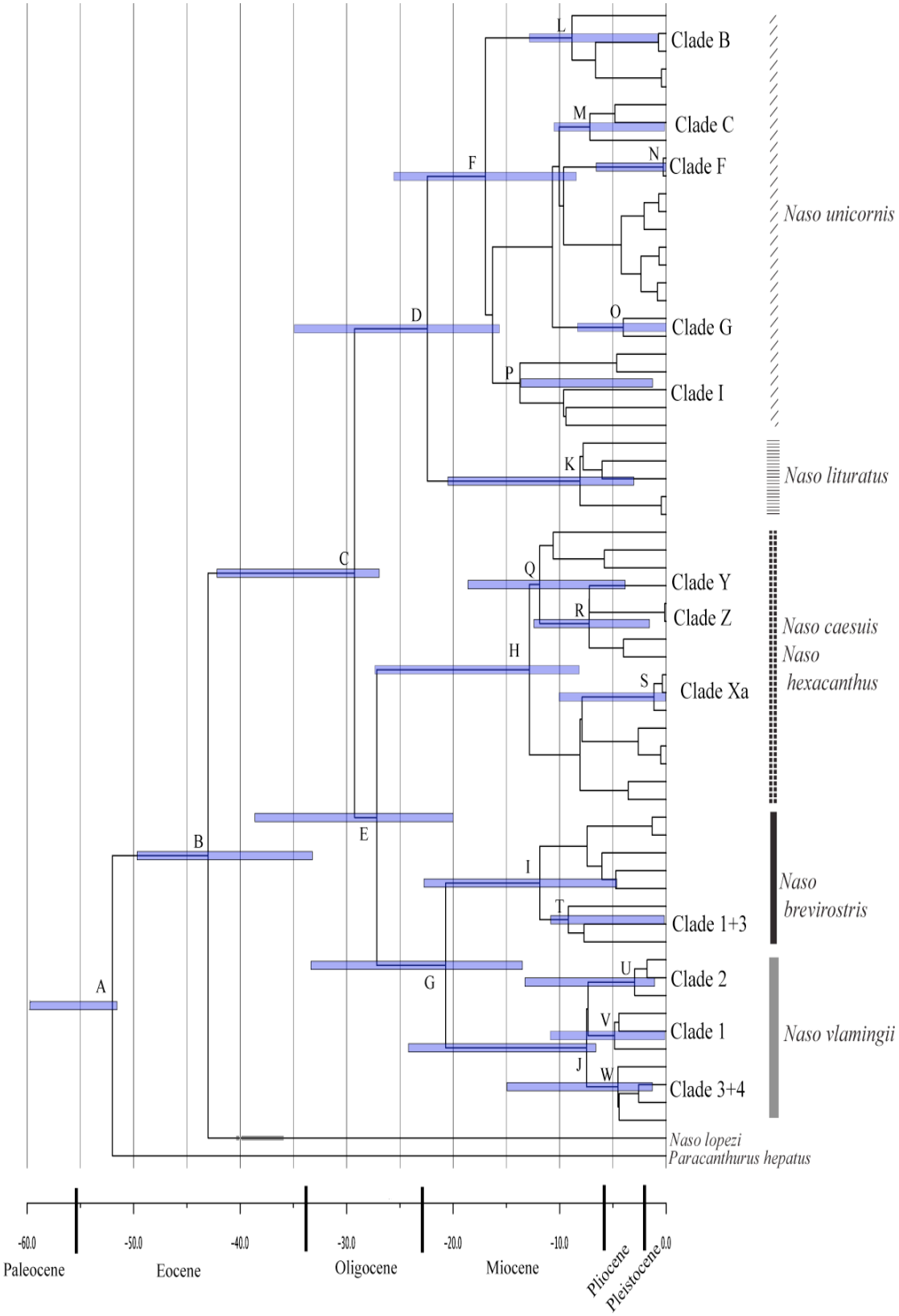


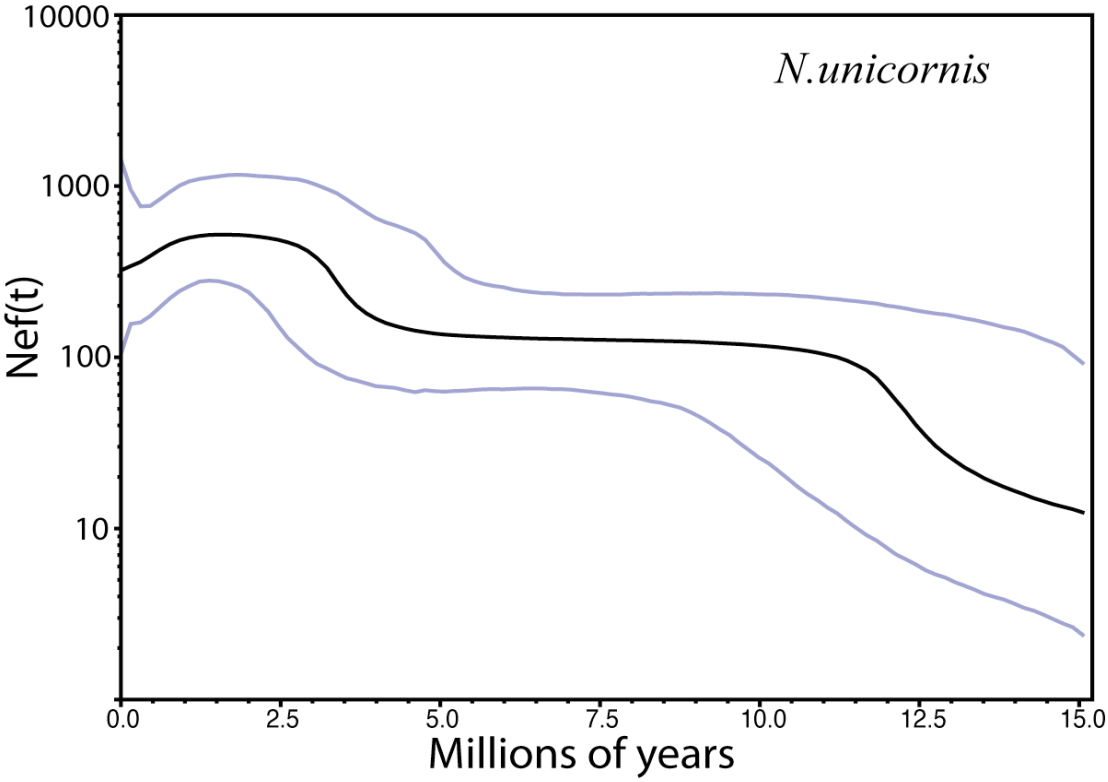
Figure 6.2: Chronogram depicting node ages in the *Naso* phylogeny. Ages are given in millions of years and error bars denote 95% posterior densities for age estimates.

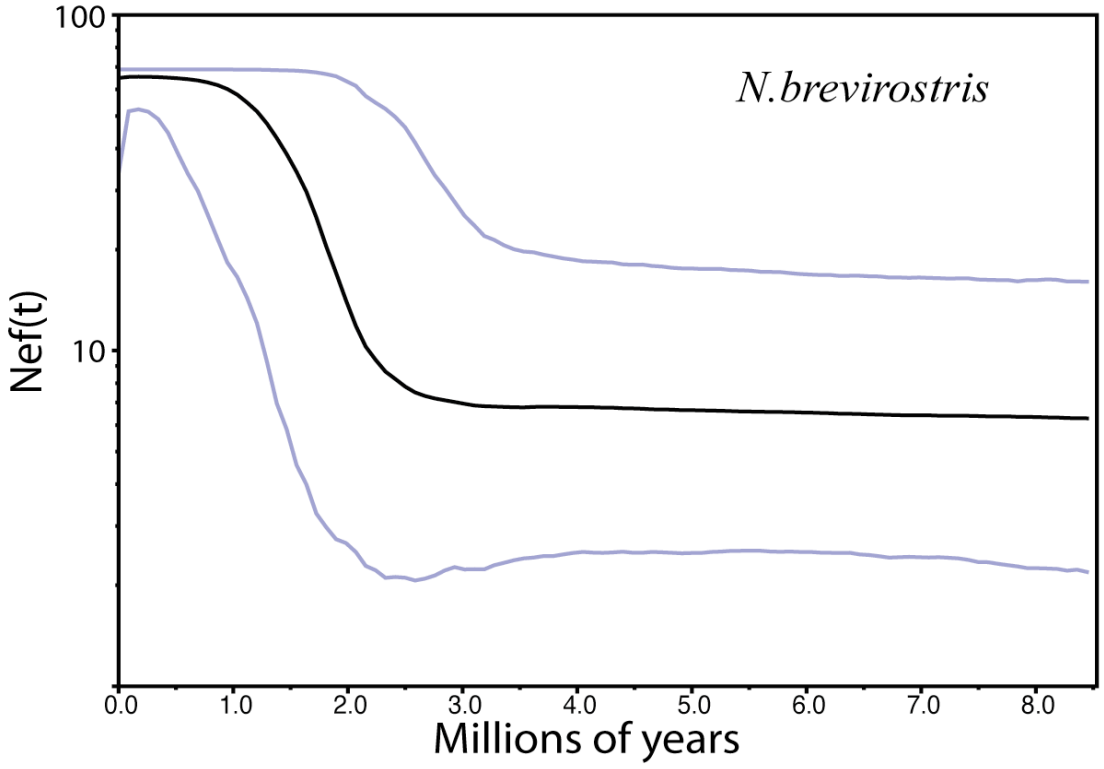
Table 6.1 Estimated node age or time to most recent common ancestor in millions of years for various taxa within the *Naso* phylogeny. Upper and lower 95% confidence interval of node heights given under Max age and Min age. Previous estimates of mean age are also given in parentheses. Mean expansion age based on mismatch distributions from previous chapters and from Klanten *et al.* (2007) are marked with †. Mean divergence time based on fossil calibrations Klanten *et al.* (2004) are marked with *.

Node	Max age	Mean Age	Min Age	Taxon
A	57.9	**50.4 (47.6)***	50.0	**Tree root**
B	48.2	**41.7 (40.1)***	32.2	***Naso***
C	40.9	**28.3 (31.3)***	26.1	**Ingroup**
D	33.9	**21.7 (18.5)***	15.2	***N. unicornis* + *N. lituratus***
E	37.4	**26.3 (21.7)***	19.4	***N. hexacanthus* + *N. caesius* + *N. brevirostris* + *N. vlamingii***
F	24.8	**16.4 (6.4)†**	8.2	***N. unicornis***
G	32.3	**20.0**	13.1	***N. brevirostris* + *N. vlamingii***
H	26.5	**12.5(3.0)*(10.1)†**	7.9	***N. hexacanthus* + *N. caesius***
I	22.0	**11.5 (11.1)†**	4.5	***N. brevirostris***
J	23.4	**7.2 (8.1)†**	6.4	***N. vlamingii***
K	19.9	**7.8**	2.9	***N. lituratus***
L	12.5	**8.6**	0.8	***unicornis* clade B**
M	10.2	**6.9**	0.1	***unicornis* clade C**
N	6.4	**0.3**	0.0	***unicornis* clade F**
O	8.1	**3.9**	0.0	***unicornis* clade G**
P	13.2	**13.3**	1.2	***unicornis* clade I**
Q	18.0	**11.5**	3.7	***hexacanthus-caesius* clade Y**
R	12.0	**7.0**	1.5	***hexacanthus-caesius* clade Z**
S	9.7	**1.1**	0.5	***hexacanthus-caesius* clade Xa**
T	10.5	**8.9 (3.3)†**	0.2	***brevirostris* clade 1+3**
U	12.9	**2.8 (5.5)†**	1.1	***vlamingii* 2**
V	10.5	**4.7 (3.4)†**	0.1	***vlamingii* 1**
W	14.5	**4.4 (5.0, 2.9)†**	1.3	***vlamingii* 3+4**

6.3.5 Bayesian Skyline Plots

Graphical representations of effective population size through time are given as Bayesian skyline (Figure 6.3). In all cases *Naso* species appear to have experienced positive population growth throughout much of their species histories as was previously indicated from mismatch distributions and tests of selective neutrality.


10000
1000
100
10
Nef(t)
N.unicornis
0.0
2.5
5.0
7.5
10.0
12.5
15.0
Millions of years


100
10
Nef(t)
N.brevirostris
0.0
1.0
2.0
3.0
4.0
5.0
6.0
7.0
8.0
Millions of years

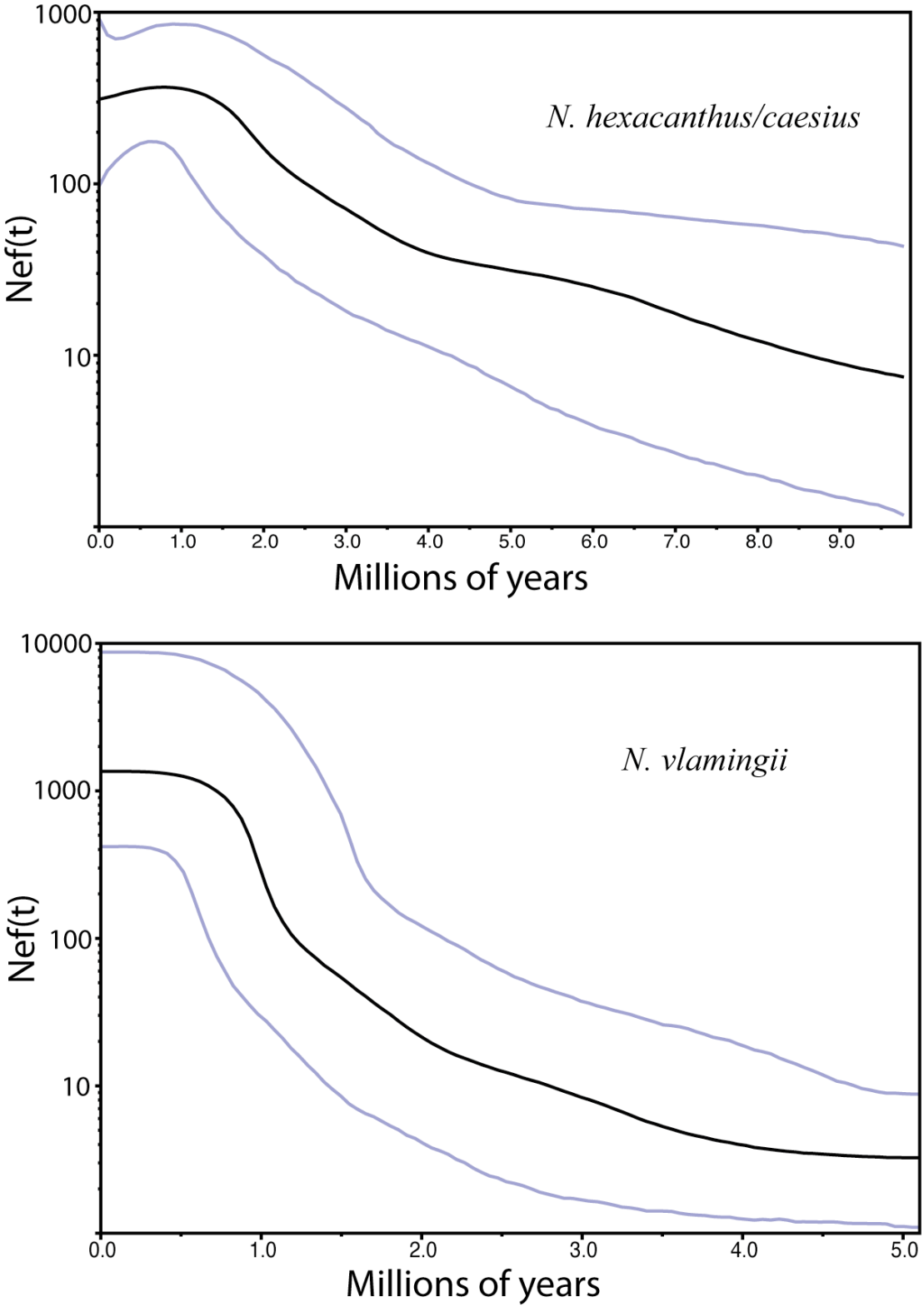


Figure 6.3: Bayesian skyline plots depicting demographic change through time. Time is given on the X axis in millions of years and the Y axis is the effective female population size multiplied by generation time (t) also in units of millions. The black line represents the median of the sampled posterior distribution and the blue lines represent the upper and lower 95% posterior densities.

Growth is also observed to increase sharply between four and one mya. This is unexpected because it roughly coincides with the dramatic climate fluctuations that commenced near the beginning of the Pleistocene 2.5 mya (Müller *et al.* 2008; Herbert *et al.* 2010), which are conventionally supposed to have caused widespread extinctions (Jackson & Johnson 2000; Renema *et al.* 2008) and population bottlenecks (Springer & Williams 1990; Fauvelot *et al.* 2003) in many tropical marine species. Estimates of present day effective female population size ranged from 2.5 million in *N. brevirostris* to over 48 million in *N. vlamingii,* given an average generation time of 21.5 years. Skyline plots of individual clades were also performed but these were remarkably similar to the plots generated for entire species and are not shown. Given a fossil calibrated coalescences ages substitution rates across the entire control region marker are: *N. brevirostris* 1.3%, *N. hexacanthus* 1.4%, *N. unicornis* 0.5%, *N. vlamingii* 2.5% per million years.

6.4 Discussion

6.4.1 Molecular evolution, mtDNA and comparative studies

Recently the sole use of mtDNA as a molecular marker for phylogenetic and phylogeographic research has come under intense criticism (Ballard & Whitlock 2004; Galtier *et al.* 2009) but this is an ongoing controversy (see Avise 2000; Zink & Barrowclough 2007; Barrowclough & Zink 2009). Here and in the previous three chapters I have dealt exclusively with mtDNA. In reality, this thesis has thus far been a study of evolution of a cellular organelle in widespread and long-lived marine species, rather than a study of evolution in unicornfishes themselves. However a comparative approach involving congruent mtDNA datasets from several species suggests that patterns are not merely a reflection of mtDNA stochasticity. We can

probably assume that the evolutionary history of these fishes is more or less parallel to the mitochondrial histories and, most importantly, we can interpret these data in the context of other mtDNA studies done on reef fishes, which commonly show much less nucleotide diversity and shallower coalescence than that observed in *Naso* (Grant & Bowen 1998; Bowen *et al.* 2006; Craig *et al.* 2007; Gaither *et al.* 2010; Visram *et al.* 2010; Winters *et al.* 2010; Reece *et al.* 2010; Ravago-Gotanco & Junio-Meñez 2010; Craig *et al.* 2010). Nevertheless, it must be acknowledged that non-geographic clades are a feature of the mitochondrial genome of widespread coral reef fishes. Understanding the biological significance of non-geographic population structuring is the main purpose of this chapter.

A note about the nucleotide substitution rates inferred herein. Compared to trans-isthmian calibrations molecular clocks in fish, a substitution rate of 0.18% per my, as reported here, is slow for the COI locus in a marine fish (Lessios 2008). However this rate is not homologous with trans-isthmian molecular clock calibrations because in this case 0.18% is the average substitution across the entire tree, including deeper internal branches. Given that there is a perceptual acceleration in substitution rate towards the terminal branches of the tree caused by ancestral polymorphism (Peterson & Masel 2009), short-term substitution rates in *Naso* are expected to be higher than the overall mean rate for the whole tree. Control region substitution rates for each species, estimated from Bayesian Skyline coalescence, were also low, however, between 1.3 and 2.5%. Only sharks have been reported to have such low control region substitution rates among fishes (Lessios 2008). But apart from the lowness of estimates, there is nothing to suggest that this rate is not accurate. A faster substitution rate, as reported in other teleosts would make coalescence ages proportionately younger and potentially alter how the species

history is interpreted. Nevertheless, the actual substitution rate would have to be approximately four times faster for expansion to be post Pleistocene.

6.4.2 The Oligocene and Miocene

The molecular ages estimated for *Naso* are quite old. The root age of the ingroup probably dates to sometime in the early to mid Oligocene and for all species pairs represented in this study (*N. caesius* excepted), time to most recent common ancestor is likely to have been in the late Oligocene or early Miocene. If modern *Naso* species have their origins in the Oligocene then the environmental conditions of this period may have influenced their biology.

The Oligocene is often viewed as a transitional stage in the evolution of Cenozoic organisms and is associated with a mass extinction of tropical marine faunas that occurred in three phases over a 10 million year period (Prothero 1994). The reasons for this faunal upheaval are most likely climate related. Around the commencement of the Oligocene the Tasmania-Antarctic passage and the Drake Passage between South America and Antarctica were opened, the circum-Antarctic current was established and the Antarctic ice sheet first appeared (Zachos *et al* 2001). In fine, the boundary between the Eocene and Oligocene marks a global change from a tropical "greenhouse" climate to a glaciated temperate climate (Lear *et al.* 2008).

During most of the Oligocene coral reefs were not abundant (Budd 2000; Keissling 2009) and may not have been common during the Paleogene at all (Keissling 2009). Fossil Nasine Acanthurids from the Oligocene have specializations that are highly suggestive of a pelagic lifestyle and were probably not reef associated (Tyler 2000). However, in the late Oligocene climate again warmed and the Australian plate began to collide with the Asian plate creating an abundance of

shallow water habitat in the tropics (Renema *et al.* 2008; Williams & Duda 2008). Coral reefs then became abundant and in the early Miocene, coral reefs can be described as modern (Bellwood & Wainwright 2002; Renema *et al.* 2008; Keissling 2009). The collision of the Australian and Eurasian plate also begins to define the Indian and Pacific basins and to constrict the seaway connecting these two oceans. Some of the oldest cases of vicariance among Indian-Pacific geminate fishes are inferred to be associated with this mid-Miocene partitioning of ocean basins (Read *et al.* 2006).

The proliferation of Cenozoic coral reef ecosystems at the end of the Oligocene and in the early Miocene coincides nicely with the early divergence dates of this study. Deductively, primitive *Naso* may have become reef associated around this same time and lineage diversification was promoted by transitions into open adaptive zones. Coral reefs in general are thought to be conducive to the generation of new taxa (Keissling *et al.* 2010). It has also been noted by Choat *et al.* (2002) that the dietary characteristics of different *Naso* species are poorly correlated with phylogenetic relationships, suggesting that at some point in time there was a departure from the ancestral foraging mode to trophic novelty. It also seems probable that modern *Naso* are descended from pelagic ancestors like *Arambourgthurus* (Tyler 2000) as they maintain several morphological features more characteristic of pelagic fishes than other coral reef species (Tyler *et al.* 1989). Also, out of all extant *Naso,* those with the strongest pelagic habitat preferences are the subgenus *Axinurus,* which are the basal clade in the *Naso* phylogeny (Klanten *et al.* 2004) and are most likely to represent the ancestral state.

During the mid-Miocene (~14 mya) there was a punctuated increase in Antarctic ice and the global climate cooled again, returning to a more Oligocene-like

condition (Zachos *et al.* 2001; Shevenell *et al.* 2008). Around this time there appears to be another pulse of new lineages for almost all of the extant *Naso* of our study have mean coalescence ages younger than this and actual ages that may not predate this climatic change (Table 6.1). *N. unicornis* underwent a period of elevated demographic expansion 14-12 mya (Fig 6.3), suggesting that this change in climate was positive for this species. All things considered, climate change in the early Neogene appears to have been an important driver of evolutionary processes in the genus *Naso*.

6.4.3 Non-geographic clades and the Pleistocene

Overall there appears to be no discernable pattern regarding the age of non-geographic clades. If one uses mean time to a most recent common ancestor as the most reliable estimation, most non-geographic clades appear to be of pre-Pleistocene origin (Table 6.1). Ancestral polymorphisms may be confounding the inference, causing these ages to be overestimated (Edwards & Beerli 2000). However, the effects of ancestral polymorphism are less of a concern when divergence time is great (Charlesworth 2010), as is the case here. Still, there is a considerable range of time within which these lineages might coalesce. Clade 1+3 of *N. brevirostris*, for instance, could be of mid Miocene origin at its maximum age or coalesce 0.2 mya at its minimum. Therefore, Pleistocene age processes cannot be definitively ruled out as contributing to these structures.

Based on Bayesian skyline analysis, the Pleistocene appears to be a time of population growth in all the *Naso* species examined here. This result is counterintuitive because this period is believed to be a time of extinctions in tropical marine faunas (Jackson & Johnson 2000; Renema *et al.* 2008). Habitat reductions

and climate fluctuations during the Pleistocene are believed to have caused population contractions in widespread Indo-Pacific taxa (Springer & Williams 1990; Randall 1998; Benzie 1999). Yet, while many coral reef fishes were probably in decline during most of the Pleistocene, these results suggests that many of the Nasine acanthurids thrived in these cooler conditions.

Why were *Naso* so successful during the Pleistocene? I can think of three reasons why this might have been the case. 1) *Naso* have a fermentive digestive system that enables them to take advantage of a wide range of dietary items including macro algae, gelatinous zooplankton and even detritus (Choat *et al.* 2002; Randall 2002). 2) Many *Naso* are also habitat generalists, associated with coral reefs but capable of surviving in various environments. During the course of this thesis a mature *N. unicornis* was reported from North Island New Zealand (Howard Choat, personal communication) suggesting that coral reefs and tropical water are not absolute habitat requirements for this species. Also, several species, such as *N. brevirostris, N. hexacanthus, N. caesius* and *N. vlamingii* have been described as semi-pelagic (Randall 2002), a possible inheritance from Oligocene forefathers. The caudal armature and venomous dorsal spines of this genus probably facilitate the occupation of pelagic environments as well. 3) Most *Naso* species are encountered at greater depths than most reef fishes: *N. unicornis* 80m, *N. brevirostris* 122m and *N. vlamingii* 50m, *N. hexacanthus* 229m (Lieski & Meyers 1994; Chave & Mundy 1994). In fact, it would probably be a mistake to consider *Naso* as shallow-water fishes. Many species such as *Naso annulatus* are rarely seen above 25m (Lieski & Meyers 1994) and some, such as *Naso maculatus* are typically found 76-120m and are amphitropical (Randall 2002). To exist at such depths shows a tolerance to colder water. This, in combination with a semi-pelagic and opportunistic ecology, is

consistent with Pleistocene demographic expansion. Ergo, *Naso* appears to be a reef fish taxon with few environmental inhibitions and which found the conditions of the Pleistocene favorable.

Another possibility is that coral reef habitat during the Pleistocene was not as scarce as previously believed. Recent oceanographic research has revealed a number of "drowned" coral reefs that would have been shallow during the Pleistocene but failed to grow at the same pace as rising sea levels. For instance, in the Hawaiian archipelago, there is evidence of extensive reef area in the so-called Maui-Nui complex (Webster *et al.* 2010; Faichney *et al.* 2011). Submerged Pleistocene reefs have been reported from New Guinea (Webster *et al.* 2004) and even in the Gulf of Carpentaria (Harris *et al.* 2008). The extensiveness of drowned Pleistocene reefs may currently be under appreciated. Therefore, ample Pleistocene refugia might have been available and buffered the effects of sea level change for some species that were not hypersensitive to disturbances (see Fauvelot *et al.* 2003; Craig *et al.* 2010).

Population growth during the Pleistocene might also explain why there is no phylogeographic structuring at the IPB and refute the involvement of this barrier in the formation of non-geographic clade structures. Genetic drift is weak when populations are expanding because it is hard for genes to become fixed when the birth rate exceeds the death rate (Slatkin & Hudson 1991). Therefore it would have been difficult for deep genetic partitions to arise during positive population growth by the force of genetic drift alone. Other processes such as small scale vicariance (see chapter 3) and range expansions (see chapter 4) could have been operating during the Pleistocene and may have contributed to non-geographic genetic structuring but isolation across the IPB due to Pleistocene sea level fluctuations and subsequent introgression presently seems like an improbability. Even if gene flow

was halted during the relatively brief extreme lows in sea level (Voris 2000), it is unlikely that genetic drift would have been able to differentiate large oceanwide populations (Slatkin 1987).

If non-geographic population structure is the result of isolation at all, this isolation probably occurred during the Miocene and Pliocene, when *Naso* populations were smaller, instead of in the Pleistocene. Where, when and how this isolation may have occurred is a matter of conjecture. Read *et al.* (2006) suggest that low-sea-level barriers, upwelling and oceanographic circulation associated with the tectonic formation of the IAA, starting in the mid-Miocene, would have separated Indian and Pacific populations, as it still does for many coral reef organisms today (Carpenter *et al.* 2011). However, if these same processes failed to differentiate *Naso* populations in the recent Pleistocene, I see no reason why they would have succeeded in the Miocene-Pliocene, when Australia was much further south and the Indonesian throughflow was wider than it is today. Regardless, while glacio-eustatic lows in sea level may explain vicariance and phylogeography in some marine organisms, this does not appear to be a good fit for *Naso*, which, at any rate, are atypical in the types of patterns they exhibit.

6.4.4 Naso hexacanthus *and* Naso caesius

The results of this study support the conclusions of chapter 5 that *N. hexacanthus* and *N. caesius* do not have a clearly defined species boundary. Also, the difference in non-geographic structuring between the control region and COI makes it hard to take these structures seriously in this taxon. Furthermore, given intense population growth during the last 3-4 million years the possibility of a recent divergence between these sister species, as proposed by Klanten *et al.* (2004) and Dayton *et al.* (1994), seems

remote unless some form of incomplete ecological speciation is invoked. Clearly more work is needed and warranted on these putative sister species to understand a very complex taxonomic relationship.

6.5 Conclusions

Overall, historical changes in climate seem to have been important to the diversification and demographic history of *Naso* species but for the opposite reason as most reef fishes. While episodes of cold climate and low sea levels brought about by glacial maxima have been detrimental to many coral reef taxa several *Naso* species appear to have benefited from these conditions. The ability of *Naso* species to survive during the frequent disturbances of habitat during the Pleistocene could be due to an ecological inheritance from an ancestry that evolved during the Oligocene, another period of great faunal turnover in the marine environment. The non-geographic clades that were the focus of this chapter show no discernible pattern of age that would indicate reciprocal monophyly across an ephemeral barrier to gene flow such as the IPB. It may be, however, that the demographic expansions of the Pleistocene, in connection with high levels of larval dispersal, brought about a rapid diversification of lineages leading to a non-geographic population structure.

Chapter Seven

Recruitment patterns of *Naso unicornis* in the Marianas Archipelago

7.1 Introduction

In previous chapters I have explored the broad-scale phylogeography of several members of the genus *Naso* and I have made inferences about the deep evolutionary history of these fishes based on genetic data. The results of these chapters indicate that *Naso* are highly dispersive marine fishes. The extent of the genetic homogeneity observed in these species at large geographic scales exceeds that of most other coral reef associated fishes studied to date, suggesting that the pelagic larvae of these fishes have a tendency to disperse and that genetic connectivity between regions has been great. Nevertheless, these patterns only point towards an overall trend of gene flow in evolutionary time over many generations (Palumbi 2003; Hellberg 2007; Jones *et al.* 2009). With respect to patterns of pelagic larval dispersal, demographic exchange of migrants and recruitment processes in recent ecological time, the mtDNA phylogeography only vaguely indicates what these patterns might be. Yet the idea that *Naso* are highly dispersive and populations regularly exchange migrants is integral to current understanding of non-geographic clades. In this final data chapter I hope to bring the discussion full circle by examining ecological patterns of larval movement in a *Naso* species.

The study of the dispersive pelagic larval phase of marine organisms continues to be one of the most challenging areas of marine research. The logistical difficulties of studying tiny larvae in the vast pelagic environment have earned this area of research nicknames like ‘the black box’ (Leis 1991) and ‘the holy grail of

marine biology' (Purcell *et al.* 2006). While there are several approaches to assessing larval dispersal, one method that is becoming increasingly popular is to sample cohorts of recently recruited individuals and survey a suite of highly polymorphic microsatellite loci to infer patterns of relationship between recruits and the preexisting adult populations and demographic exchange between neighboring populations (Selkoe *et al.* 2006; Buston *et al.* 2008; Hepburn *et al.* 2009; Planes *et al.* 2009; Christie *et al.* 2010). This type of genetic methodology can be approached in two different ways:

1) By using genetic markers as direct measures of larval movement, also known as parentage analysis (see Jones *et al.* 2009). This approach can produce direct evidence of dispersal or self-recruitment but requires large numbers of unlinked loci in order to confidently connect offspring to parents and demands extensive sampling in order to find parent offspring pairs amid large and demographically complex reef fish populations. Also, most parentage analysis requires estimates of demographic proportions of adults for obtaining confidence, which can be difficult to obtain in marine populations (Christie *et al.* 2009). While parentage analysis has been successfully applied to some reef fishes (Jones *et al.* 2005; Planes *et al.* 2009), this may not be a pragmatic approach for many marine species in which the amplitude and scale of dispersal may be intractable for most research initiatives.

2) Indirect appraisals of genetic boundaries, such as conventional population genetic inferences. Again, in order to deduce ecological patterns of dispersal multiple loci are needed and, in general, the more loci the better. These types of inferences are particularly useful for assessing recent migration when populations are structured. When adult populations group into specific genetic clusters recruiting larvae can be

assigned to the parent population of origin using a variety of "assignment" tests (Manel *et al*. 2005). Certain Bayesian based analyses also exist that are able to quantify recent migration, within a few generations, as the proportion of individuals that are recent migrants, or that are immediately descended from recent migrants (Wilson & Rannala 2003). These methods, however, are severely compromised when there is a lack of genetic structure between populations. For reef fishes that have genetically homogenous populations across large geographic ranges, indirect methods would probably reveal little. In taxa like *Naso*, the utility of such analyses to resolve ecological migrant exchange is questionable.

Recently Christie (2009; *et al.* 2010) have suggested that both indirect and direct measures of dispersal can be combined and implemented effectively even with a modest number of loci. For example, if recruits and adults belong to the same genetic cluster, the probability that two individuals share a parent offspring relationship may be determined based on estimated allele frequencies rather than demographic proportions. Christie *et al.* (2010) used this method to show self-recruitment within genetically connected populations of the damselfish *Stegastes partitus* in the Bahamas, with 437 adults and 314 recruits and using only seven loci. If such a method can detect self-recruitment amid high gene flow in an abundant species, like *Stegastes partitus*, at a small spatial scale, it might also detect recruitment in a population of *Naso*.

Here I present a genetic study of settling cohorts of *N. unicornis* larvae from the island of Guam. Apart from being a subject of previous chapters, *N. unicornis* has several attributes that make it a desirable study species for this type of research. For instance, it is known to settle in massive recruitment events (Doherty *et al.* 2004) and has a specific settlement habitat: shallow reef flats where it is easy to sample.

Settling recruits of *N. unicornis* are also relatively large compared to most reef fishes, attaining lengths of typically no less than 5 cm (often 6-8 cm) upon settlement and are easy to identify from the two blue spots on the caudal peduncle. Overall, the amplitude of recruitment, settlement habitat and large size of readily distinguishable post-larvae allow for robust sampling.

This is not the first time that *N. unicornis* has been the subject of such a study. Planes *et al.* (2002) surveyed 132 larvae and 124 recently settled juvenile *N. unicornis* from Moorea, French Polynesia, using 19 allozyme loci. Amongst other results, two conclusions that were drawn from this study are: 1) No evidence for self-recruitment. And, because otolith aging of the recruits revealed that larvae settling on the same night were of different ages, the recruits most likely originated from multiple spawning events. 2) Larvae generally shared closer overall genetic relatedness with each other than with juveniles or adults from the same reef, suggesting that siblings may have recruited together but that not all recruits in the same cohort were siblings. In light of the phylogeographic data of chapter four, which revealed genetic homogeneity across a broad-spatial scale, the chaotic genetic recruitment patterns observed by Planes *et al.* (2002) could be consistent with high amounts of external recruitment and demographically open populations.

The first objective of this chapter was to attempt to reproduce the results of Planes *et al.* (2002): do recruiting *N. unicornis* share a genetic identity against that of the preexisting adult population? Is there any evidence for sibling relationships within recruitment cohorts? The present study aims to elaborate on these issues and differs in design from that of Planes *et al.* (2002) in two main ways. 1) While Planes *et al.* used only one recruitment site for their study and only sampled recruits from one year, in this study recently recruited *N. unicornis* were sampled from five

different sites on Guam and one on the neighboring island of Saipan and samples were collected from two consecutive years permitting an assessment of the temporal stability of the genetic signal of recruits. 2) Rather than use allozyme loci, I used highly polymorphic microsatellite loci (see chapter two), which are considered to be much more sensitive to ecologically relevant genetic patterns (Waples & Gaggiotti 2006).

Secondly, is there any evidence for self-recruitment in *N. unicornis* populations on Guam? While this fish is numerous on most Indo-Pacific coral reefs, on Guam the population size is kept artificially low due to intense fishing pressure from local recreational and commercial fishermen. Considering the massive size of recruitment events in this species and the artificially low numbers of adults, if self-recruitment is prevalent on Guam it may be detectable using the methods of Christie *et al.* (2010).

7.2 Methods

7.2.1 Field collections on Guam.

Beginning in June 2008 a massive recruitment event of the blue-spine unicornfish (*N. unicornis*) occurred on the island of Guam. The exact scale of this mass recruitment is unknown but was large enough to deliver high numbers of settling larvae to a large portion of the shallow reef flat area fringing the island of Guam. During a similar event in French Polynesia, involving the same species, Doherty *et al.* (2004) reported that as many as 10,000 recruits per km^2 per night may settle on the reef at the height of settlement. Therefore, it may not be unreasonable to conservatively estimate the total number of these settlers in the hundreds of thousands. The first pulse of recruitment appears to have been the largest but based

on personal observation there was also a second distinct pulse sometime in July and possibly in September of that same year as well. It is also likely that small numbers of larvae continued to settle throughout this time. Recruitment was particularly intense on the western coast of Guam, most notably in the later months. By late August recruits on the east coast were seen only in very small numbers, while west coast sites continued to house numerous (and more recent) recruits.

Recruits were sampled at five main sites around the island of Guam (Fig. 2.4). Sampling was done mainly at night by snorkel at depths no greater than 1m. *N. unicornis* recruits were spotted with a dive light and collected by hand. Because of the greater abundance, most of our samples come from sites on the west side of the island. The recruits of other *Naso* species were also found alongside *N. unicornis* recruits but could be easily distinguished by the absence of two diagnostic blue spots on the caudal peduncle (see Fig. 2.4). *N. lituratus* recruits were also common on the fringing reef flats around Guam and appeared to have settled synchronously with *N. unicornis*. *N. vlamingii* juveniles were found in small numbers as well as an unidentified species that I suspect to be *N. tonganus* as the adults of this species were common on Guam.

Immediately following collection, samples were transported back to the University of Guam marine lab, where a small fin clip was taken and preserved in 80% EtOH for further genetic study. Samples were also weighed, measured for fork length and the otoliths removed during this process. [Note: It is my hope that the final published version of this data will use the daily otolith growth rings from each individual to be able to segregate genetics samples by age since settlement and pelagic larval duration, however, this data is not available at the present time.]

Adult *N. unicornis* were collected from the island of Guam in a variety of ways. Some of these were sampled directly by spearing and for these individuals I have collection site data. Yet other *N. unicornis* adults were purchased from commercial fishers on Guam or were confiscated fish taken from poachers. For these individuals there is no data on where the fish were caught. Hence, for most of the inferences here adult *N. unicornis* from Guam are treated as a single population. I justify grouping all adult samples together based on the fact that *N. unicornis* aggregate for spawning (Lieski & Meyers 1994). Therefore, while short-term observations reveal that this species has limited home range sizes (Meyer & Holland 2005), mating events probably draw individuals from large areas. Recent observations of occasional home range shifts in other *Naso* species (Meyer *et al.* 2010) might also indicate that individuals do not spend their entire lives in one area.

7.2.2 Laboratory procedures

Genetic processing of *N. unicornis* samples from Guam, for the mitochondrial control region and for microsatellite loci, was performed as outlined in chapter two. All individuals were genotyped with at least 12 of the original 15 loci (Nuni 1-12; table 2.2). Seventy-seven specimens were genotyped twice for quality control and returned consistent results.

7.2.3 Genetic analysis of mtDNA

MtDNA haplotype sequences were aligned as described in chapter 3. Number of haplotypes, haplotype diversity and nucleotide diversity were calculated in DNASP version 4.50.3 (Rozas *et al.* 2003). Pairwise *F*st and hierarchal AMOVA were performed between age cohorts and collection sites in ARLEQUIN version 3.1

(Excoffier *et al.* 2005) using 10,000 permutations. Twenty-five *N. unicornis* individuals from Tonga were also analyzed against all Guam samples as an external control and to help place the Guam samples in the broad scale genetic context of chapter 4.

7.2.4 Summary statistics, tests of Hardy-Weinberg equilibrium and linkage disequilibrium of microsatellite loci

Exact tests for departure from Hardy-Weinberg equilibrium (HWE) were conducted in GENEPOP version 4.0.10 (Rousset 2008) and also tests of linkage disequilibrium (LD) using the Markov chain algorithm, a dememorization of 10,000, with 20 batches and 5000 iterations per batch. Loci that did not conform to HWE may be under selection and were not used for downstream analyses but this was based solely on allele frequencies in the one adult population sampled in this study because recruitment cohorts are not reproductive populations, potentially contain kinship aggregations and may display allele frequencies that are misleading in this regard. The presence of null alleles, large allele drop out, stuttering and other genotyping errors were assessed in MICROCHECKER version 2.2.3 (van Oosterhout *et al.* 2004). Number of alleles, allelic richness, private alleles, observed and expected heterozygosities based on Hardy-Weinberg proportions were estimated in FSAT version 2.9.3 (Goudet 2001) and in GENALEX version 6.4 (Peakall & Smouse 2006). Pair wise *F*st were performed for each sample site and hierarchal AMOVA with various configurations to detect genetic structure in time and space were implemented in ARLEQUIN.

7.2.5 Discriminant analysis of principal components

Discriminant analysis of principal components (DAPC) is a multivariate analysis that may be used to extract information from large genetic data sets and assign individual genotypes to predefined groups (Jombart *et al.* 2010). First, data is transformed into uncorrelated components, containing most of the genetic information, using principal components analysis (PCA). These components are then subjected to a linear discriminant analysis (DA) that minimizes the genetic variance within groups while maximizing among group variation, thus providing the best discrimination of predefined genetic groups. When prior population information is unavailable, it is possible to identify and describe genetic clusters without defining groups *a priori* using a k-means algorithm, which is based on the same statistical model as discriminant analysis. This method has been shown to be as sensitive as Bayesian clustering programs (Jombart *et al.* 2010) but does not require large amounts of computational time. Furthermore, DAPC does not assume HWE or LD and has very few assumptions making it an extremely versatile methodology.

I used DAPC scatter plots, created in the R package adegenet (Jombart 2008; Jombart *et al.* 2010), to visually represent genetic patterns among *N. unicornis* age classes and cohorts in space and time in order to give the best genetic discrimination of these groups as they were collected in the field. However, due to a lack of spatial or temporal genetic structure (see Results) there was no justification for assigning individuals to predefined sample locations. Instead I used the “find.clusters” function from the R package adegenet (Jombart *et al.* 2010) to detect genetic clustering without prior group information. First, PCA was performed on the data, for this analysis data was scaled and all missing data was assigned to the mean of PCA. Next a number of PCs were retained as predictors for discriminant analysis. There are no

strict guidelines for determining how many PCs should be retained during this dimensions-reduction step but it is a compromise between the statistical power of more PCs and the stability of assignments (Jombart *et al.* 2010 & references therein). For the purposes of this chapter, the 100 PCs were retained containing 80% of the variation of the data. The correct number of demes (k) was selected based on likelihood score and the Bayesian information criterion with 10,000 iterations.

7.2.6 Parentage analysis and tests of relatedness

To investigate relationships between individuals, Queller & Goodnight's coefficient of relatedness (*r*) was calculated for all individuals in a pair wise fashion using the program RELATEDNESS v. 5.0.8 (Quller & Goodnight 1989). Mean *r* was also calculated for groups based on collection data. On average, $r = 0$ in unrelated individuals or when the relatedness of a group is random, $r = 0.25$ in half siblings and $r = 0.5$ in full siblings. This program calculates *r* based on a regression using allele frequencies. For this analysis all individuals were weighted equally and, because no genetic structure was detected, the allele frequencies of the entire data set were used (see Results).

To detect potential parent-offspring pairs in the data I used Mark Christie's method (Christie 2009; Christie *et al.* 2010). These analyses are presently in the form of script written for the program R (http://www.Rproject.org; Ihaka & Gentleman 1996) and are available from the website http://sites.google.com/site/parentagemethods/ along with detailed instructions. The main parameters generated are the putative parent-offspring pairs, the probability that any pair is false (Pr Φ) and the expected number of false pairs given the data set. Missing data was reset to the most common allele for each locus for the calculation

of allele frequencies (Christie 2009; Christie *et al.* 2010) but not when comparing the allele identities of putative parent-offspring pairs.

7.3 Results

7.3.1 Molecular diversity

I resolved approximately 250 bp of the mitochondrial control region and genotyped all samples with 12 microsatellite loci. Genetic diversity of the mtDNA is shown in table 7.1 and is similar between sample sites, cohorts and age classes. In other words, the genetic diversity of recruiting individuals was just as great as that of the preexisting adult population. In total, there were 373 haplotypes observed and approximately 80% of all individuals had unique haplotypes. There was no genetic structure observed in the mitochondrial control region between sites, cohorts or age classes (table 7.2), except for one pair wise test between the October sample of recruits from Governors and the adult population. Apart from this one exception, data from the mtDNA suggested that recruitment events of *N. unicornis* were unstructured in space and time. Neither was there any structure detected between the Micronesian samples and 25 adult *N. unicornis* from Tonga and there were four haplotypes shared between Tongan adults and recruits on Guam.

Four of the twelve microsatellites (Nuni 2,4,9,10) were found to have significant departures from HWE and were not used in subsequent analyses (table 7.3). Excluded loci also appeared to be the more erratic, with imperfect repeats, of the set. Out of 66 pairwise tests of LD only two were significant (loci Nuni 01 and Nuni 11, $p = 0.03$; Nu07 and Nu12, $p = 0.04$). However these were based only on one adult population and further investigations of LD in the recruitment samples did not repeat this linkage. Therefore these results should be treated cautiously.

According to MICROCHECKER, loci Nuni 02 and Nuni 09 are affected by null alleles but both of these loci were already excluded for HWE violations. It might also be mentioned that null alleles are expected to exaggerate genetic differentiation (Chapuis & Estoup 2007), which is clearly not a problem with this data set. Diversity indices for the eight remaining loci are presented in table 7.4.

	n	N_h	*h*	π
1. Adults	**94**	**89**	**0.99**	**0.0787**
2008 Recruits	**246**	**217**	**0.99**	**0.0852**
West Coast	**150**	**138**	**0.99**	**0.0847**
2. Governors June	27	26	0.99	0.0790
3. Governors Aug.	56	56	1.0	0.0857
4. Governors Oct.	32	31	0.99	0.0916
5. Tanguison June	39	38	0.99	0.0821
6. Tanguison Aug.	13	13	1.0	0.0802
East Coast	**78**	**76**	**0.99**	**0.0869**
7. Pago	24	24	1.0	0.0813
8. Ipan	35	35	1.0	0.0924
9. Cocos	20	20	1.0	0.0841
2009 Recruits	**79**	**74**	**0.99**	**0.0800**
West Coast	**44**	**43**	**0.99**	**0.0809**
10. Governors	33	32	0.99	0.0788
11. Tanguison	11	11	1.0	0.0859
East Coast	**35**	**33**	**0.99**	**0.0790**
12. Pago	11	9	0.96	0.0778
13. Ipan	13	13	1.0	0.0866
14. Cocos	11	11	1.0	0.0717
15. Saipan Recruits	**28**	**28**	**1.0**	**0.0817**
Total	**469**	**373**	**0.99**	**0.0828**

Table 7.1: Mitochondrial control region genetic diversity for adult *N. unicornis* from Guam and recent recruits from Guam and Saipan: number of samples (*n*), number of haplotypes (N_h), haplotype diversity (*h*) and nucleotide diversity (π).

Table 7.2: Pair wise *F*st values from mtDNA control region sequences for adult and recruit *N. unicornis* from Guam as numbered above. Also recruits from Saipan (15) and adults from Tonga (16). Values lower than 0.001 are displayed as 0. Uncorrected significant values ($\alpha = 0.05$) are in bold. No pair wise comparisons were significant after Bonferroni correction ($\alpha = 0.000416$).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0															
2	0.001	0														
3	0.004	0.003	0													
4	**0.012**	0	0	0												
5	0.008	0	0	0	0											
6	0	0	0	0	0	0										
7	0	0	0	0	0	0	0									
8	0.007	0.007	0	0	0.003	0.001	0.002	0								
9	0	0	0	0	0	0	0	0	0							
10	0.003	0.009	0.001	0.006	0	0	0	0	0	0						
11	0.015	0.027	0	0	0	0.011	0.004	0	0	0.009	0					
12	0.022	0.008	0	0	0	0.010	0.011	0	0	0.012	0	0				
13	0	0.014	0.002	0.007	0.006	0	0.002	0	0.009	0.004	0.021	0.036	0			
14	0	0	0	0	0	0	0	0	0	0	0	0.001	0	0		
15	0.009	0.006	0	0	0.003	0.003	0.004	0	0.002	0.007	0	0	0.022	0	0	
16	0.007	0	0	0	0	0	0	0	0	0	0	0	0	0	0.00	0

As with the mtDNA, genetic diversity in the nuclear microsatellites was highly similar in both recruits and adults. To the extent that my sample represents the actual genetic diversity, it would appear that the recruits are slightly more genetically diverse than the preexisting population on Guam, in both the mitochondrial and nuclear genome. The average inbreeding coefficient for all groupings is low and for some of the sites it is negative, indicating a lack of inbreeding and, by extension, a lack of relationship within some age cohorts from the same site. There were also a large number of private alleles (40 in total) most of which belonged to recruits. In this case a rare alleles in the recruits may indicate that the genetic diversity is greater than our sample would suggest. It might also be taken as evidence for external recruitment. Pairwise *F*st values across eight microsatellite loci are reported in table 7.5. As with the mtDNA there is a striking absence of structuring between sites,

cohorts and age classes, with only two exceptions. One significant test of population structure was between June and October samples of the 2008 recruits at Governors, suggesting temporal instability at this site. All hierarchal AMOVA analyses produced non-significant fixation indices of $\Phi_{st} = 0.0$, regardless of how populations were arranged.

Table 7.3: Exact tests for departure from Hardy-Weinberg equilibrium from 176 adult *N. unicornis* from Guam, for 12 polymorphic microsatellite loci and the total number of alleles (N_a) for each. The metric used is Weir and Cockerham's (1984) F_{IS} with corresponding probability. P values less than 0.05 are considered significant departures and indicate possible selection acting on these loci (highlighted with bold italics) and were not used in further analyses.

Locus name	**N_a**	***F_{IS}***	**p**
Nuni01	12	0.0050	0.5568
Nuni02	***17***	***0.2457***	***0.0000***
Nuni03	33	0.0416	0.0640
Nuni04	***26***	***0.0358***	***0.0382***
Nuni05	22	0.0658	0.2652
Nuni06	38	0.0364	0.2240
Nuni07	24	0.0293	0.1131
Nuni08	13	0.0425	0.0531
Nuni09	***25***	***0.2313***	***0.0000***
Nuni10	***25***	***0.0458***	***0.0053***
Nuni11	11	0.1174	0.1389
Nuni12	28	-0.0037	0.4875

Table 7.4: Genetic diversity indices for all *N. unicornis* adults from Guam and recruits from Guam and Saipan across eight microsatellite loci. Number of samples (*n*), average number of alleles across 8 loci (N_a), observed and expected heterozygosity (H_O, H_E), the population specific average inbreeding coefficient (F_{IS}), allelic richness (R_S) and number of private alleles (P_a). Bold type serves only to highlight larger groupings of samples.

	n	N_a	H_O	H_E	F_{IS}	R_S	P_a
1. Adults	**176**	**18**	**0.798**	**0.827**	**0.039**	**12**	**10**
2008 Recruits	**268**	**19.75**	**0.787**	**0.829**	**0.051**	**12.32**	**21**
West Coast	**176**	**18**	**0.783**	**0.828**	**0.057**	**13.49**	**11**
2. Governors June	30	12.25	0.798	0.808	0.016	7.42	1
3. Governors Aug.	58	15.25	0.770	0.824	0.066	7.78	5
4. Governors Oct.	33	12.625	0.776	0.811	0.047	7.77	2
5. Tanguison June	39	13.5	0.791	0.829	0.048	7.75	2
6. Tanguison Aug.	13	10.875	0.796	0.803	0.015	7.76	0
East Coast	**92**	**16.75**	**0.795**	**0.824**	**0.034**	**13.65**	**10**
7. Pago	25	12.625	0.784	0.813	0.059	7.87	2
8. Ipan	39	14.25	0.801	0.821	0.012	7.88	3
9. Cocos	28	12.75	0.795	0.810	0.006	7.65	2
2009 Recruits	**88**	**16.25**	**0.793**	**0.821**	**0.035**	**11.97**	**5**
West Coast	**47**	**14.375**	**0.798**	**0.821**	**0.025**	**13.31**	**3**
10. Governors	32	13.0	0.804	0.818	0.017	8.09	1
11. Tanguison	16	10.375	0.784	0.803	0.017	7.45	0
East Coast	**40**	**13.75**	**0.788**	**0.809**	**0.031**	**13.27**	**1**
12. Pago	11	8.875	0.783	0.791	0.013	7.57	2
13. Ipan	20	10.375	0.802	0.795	-0.003	7.66	1
14. Cocos	9	7.25	0.766	0.758	-0.007	7.65	0
15. Saipan Recruits	**30**	**13.25**	**0.801**	**0.823**	**0.026**	**12.64**	**4**
Total	**562**	**16.813**	**0.795**	**0.825**	**0.038**	**13.53**	**40**

7.3.2 Relatedness and Parentage analysis

Based on eight microsatellite loci, 28 putative parent-offspring pairs were identified, however, given the allele frequencies of the data set, the probability that any of these were false pairs was extremely high (Pr Φ = 1.0) and the expected number of false pairs in the data was 44.93. Furthermore, none of these relations could be confirmed from mtDNA or by adding more loci. Therefore, there is no evidence that any parent offspring-pairs were sampled in this study.

Pair wise relatedness between individuals and the mean relatedness coefficient for each group is given in figures 7.1a and 7.1b. Overall, recruits had a mean relatedness of 0.0 and the mean for adults was only slightly lower (*r* = -

0.0057). Mean relatedness within the sampling sites of recruits ranged between r = -0.0175 and -0.125. Notwithstanding the negative mean relatedness, some individuals from the same sites were genetically similar enough to suspect half or even full sibling relationships and may be evidence that siblings recruit together. For the purposes of this study, however, putative sibling pairs were defined as those individuals which shared a mtDNA haplotype and had a pair wise coefficient of relatedness in the microsatellites of $r > 0.19$. In reality this only indicates maternal siblings and individuals with $r < 0.19$ also have the potential for sibling relationship. However, notwithstanding that these criteria are very strict and may underestimate the true number of sibling relationships, using the diversity of mtDNA haplotypes (N_h = 373) in combination with the pair wise relatedness values is conservative and helps ensure that type II errors are avoided when inferring kinship between two individuals.

Table 7.5: Pair wise *F*st values across eight microsatellite loci for adult and recruit *N. unicornis* from Guam as numbered above. Plus recruits from Saipan (15). Values lower than 0.001 are displayed as 0. Uncorrected significant values are in bold (α = 0.05). No pair wise comparisons were significant after bonferroni correction (α = 0.000476).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1	0														
2	0.002	0													
3	0	0	0												
4	0.002	**0.009**	0.002	0											
5	0	0	0	0.003	0										
6	0	0	0	0	0	0									
7	0	0.002	0	0.004	0.002	0	0								
8	0	0	0	**0.006**	0.001	0	0	0							
9	0	0.002	0	0.006	0	0	0.003	0	0						
10	0	0.005	0	0	0.001	0	0.002	0.003	0	0					
11	0	0	0	0	0	0	0	0	0.002	0	0				
12	0	0.006	0.007	0.004	0	0	0.003	0.008	0.006	0	0	0			
13	0	0.002	0	0	0	0	0	0	0	0	0	0	0		
14	0.003	0	0	0.007	0	0	0.002	0.003	0.005	0.002	0	0.006	0	0	
15	0.007	0.002	0	0	0	0.001	0.013	0.013	0.010	0	0	0.002	0	0	0

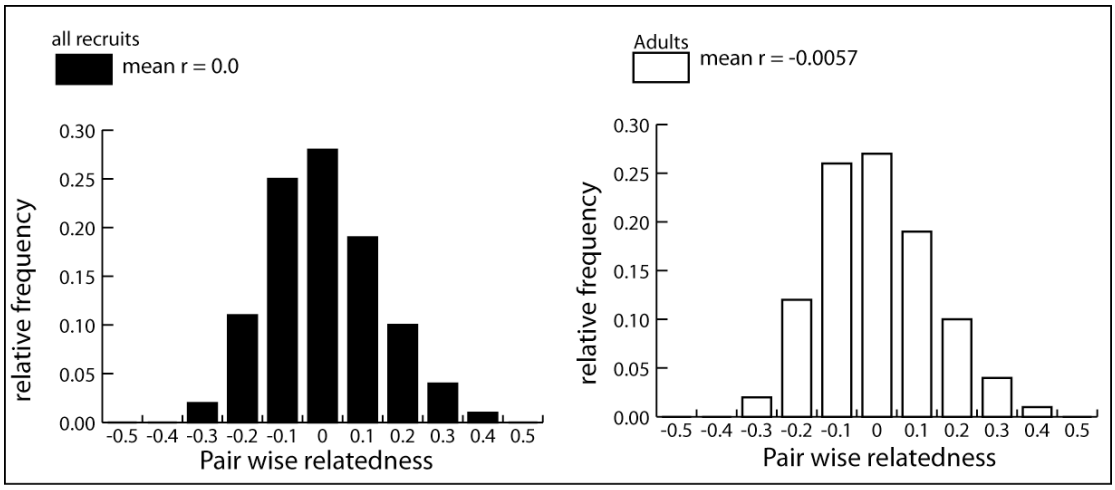
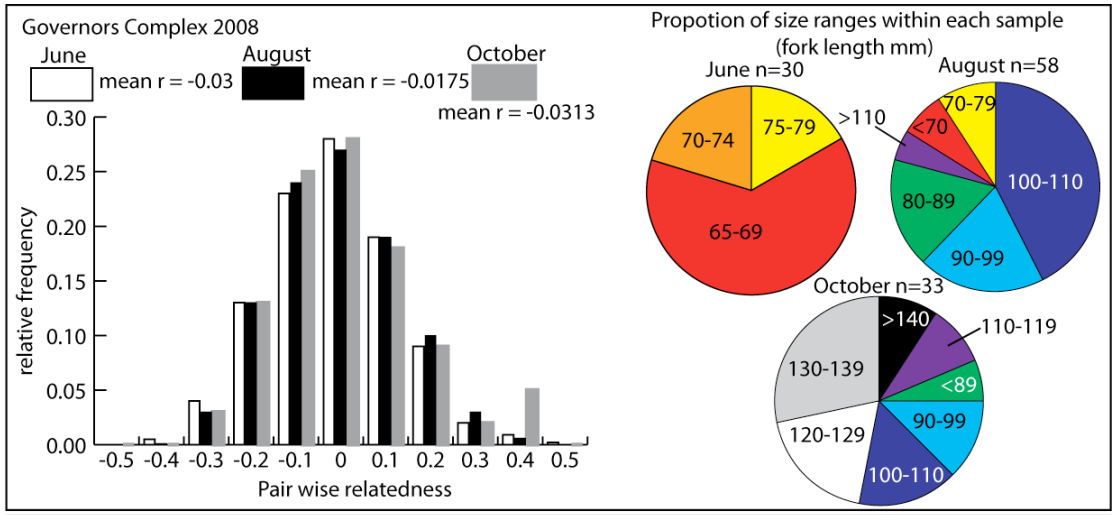
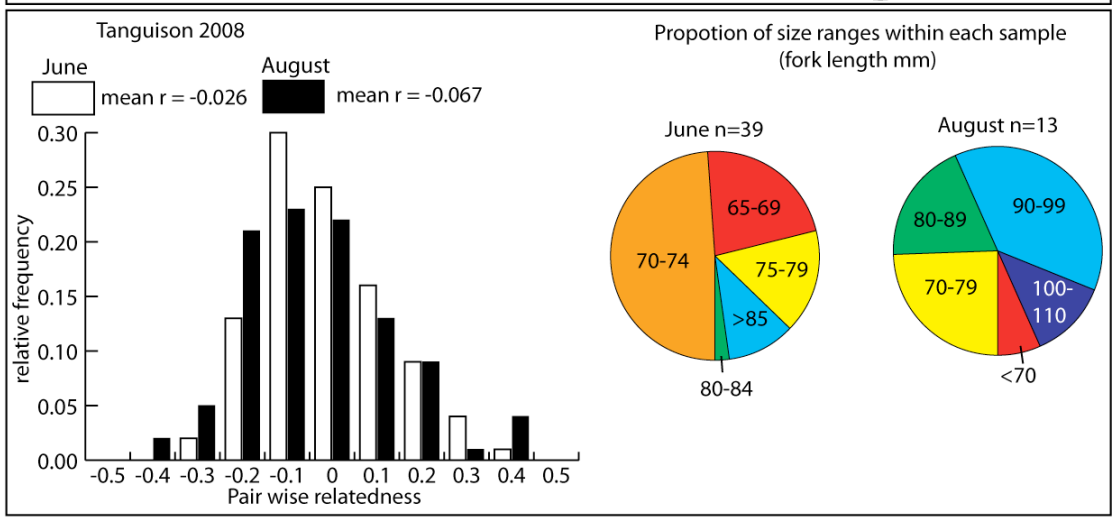
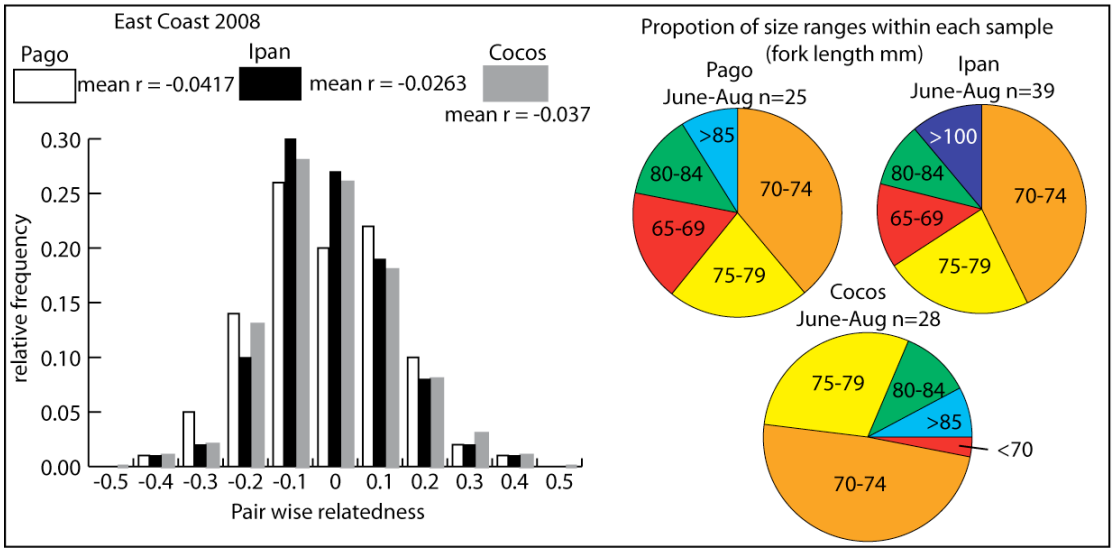


Figure 7.1a: Pair wise relatedness distributions for all recruit samples from both Guam and Saipan from 2008 and 2009, and all adult *N. unicornis* samples from Guam along with the average relatedness (mean *r*) from both groups.


Governors Complex 2008
June
August
October
mean r = -0.03
mean r = -0.0175
mean r = -0.0313
relative frequency
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.5 -0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5
Pair wise relatedness
Propotion of size ranges within each sample
(fork length mm)
June n=30
70-74
75-79
>110
65-69
August n=58
70-79
<70
80-89
100-110
90-99
October n=33
>140
110-119
130-139
<89
90-99
120-129
100-110


Tanguison 2008
June
August
mean r = -0.026
mean r = -0.067
relative frequency
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.5 -0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5
Pair wise relatedness
Propotion of size ranges within each sample
(fork length mm)
June n=39
65-69
70-74
75-79
>85
80-84
August n=13
80-89
90-99
70-79
100-
110
<70


East Coast 2008
Pago
Ipan
Cocos
mean r = -0.0417
mean r = -0.0263
mean r = -0.037
relative frequency
0.30
0.25
0.20
0.15
0.10
0.05
0.00
-0.5 -0.4 -0.3 -0.2 -0.1 0 0.1 0.2 0.3 0.4 0.5
Pair wise relatedness
Propotion of size ranges within each sample
(fork length mm)
Pago
June-Aug n=25
>85
80-84
70-74
65-69
75-79
Ipan
June-Aug n=39
>100
80-84
70-74
65-69
75-79
Cocos
June-Aug n=28
75-79
80-84
>85
<70
70-74

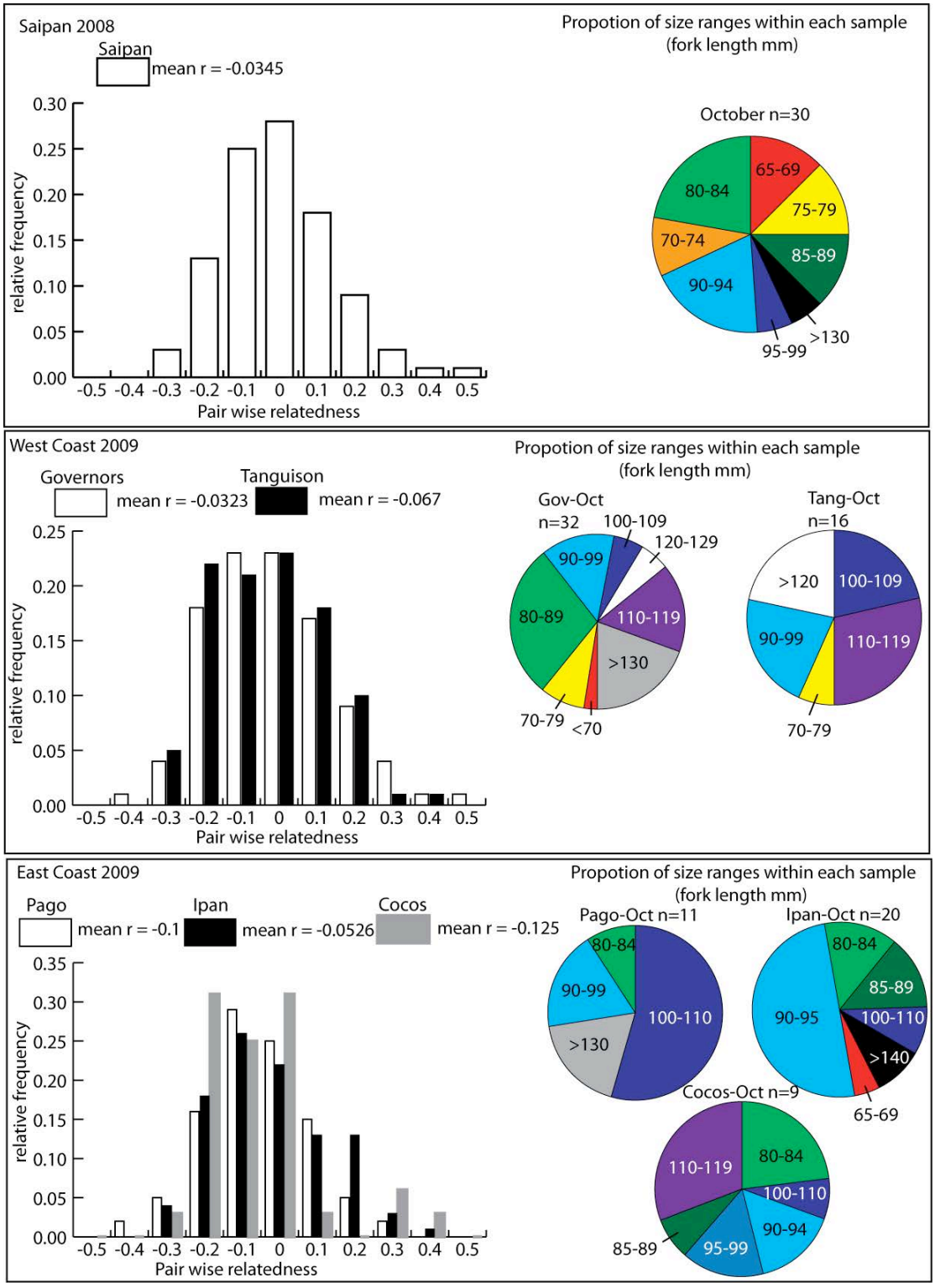


Figure 7.1b: [left] Pair wise relatedness distributions for all sample sites of *N. unicornis* recruits from Guam and Saipan during the 2008-2009 recruitment season, along with the average relatedness (mean *r*) from each group. **[right]** Pie graph depictions of the proportions of size ranges of recruits from each sample site and the number of samples collected.

Given the above criteria there were 13 putative sibling pairs identified among sampled recruits (table 7.6). Only a single pair, from among the Governors complex samples, came from the same site, all others were not sampled together. Three of these pairs were separated between Guam and Saipan. In one of these cases, a recruit from Saipan had a relationship of $r = 0.4$, enough to suspect a full kinship, with another recruit sampled a year later at Governors on Guam. There was no missing data in either of these recruits. Therefore, while there is some evidence that siblings may have settled in close proximity, recruiting *N. unicornis* were perhaps more likely to be related to individuals outside of their sample site than within. Note also that mean relatedness among recruits, as a whole, was higher than mean relatedness for any specific sample site. Six out of 13 putative sibling relationships were between year classes, so recruits were perhaps just as likely to be related to recruits of other generations as to their own cohorts.

Table 7.6: List of putative sibling pairs among sampled *N. unicornis* recruits from Guam and Saipan. All putative pairs share a mtDNA haplotype and have a pair wise coefficient of relatedness of $r > 0.19$.

Sibling 1	**Sibling 2**	**Pair wise coefficient of relatedness**
Governors, Oct, 2009	Saipan Oct, 2008	$r = 0.40$
Governors, Jun, 2008	Cocos, Oct, 2009	$r = 0.27$
Governors, Jun, 2008	Tanguison, Aug, 2008	$r = 0.32$
Governors, Jun, 2008	Governors, Aug, 2008	$r = 0.34$
Pago, Jun, 2008	Tanguison, Aug, 2008	$r = 0.25$
Pago, Jun, 2008	Governors, Aug, 2008	$r = 0.24$
Ipan, Jun, 2008	Governors, Aug, 2008	$r = 0.28$
Ipan, Jun, 2008	Saipan, Oct, 2008	$r = 0.24$
Governors, Oct, 2008	Cocos, Oct, 2009	$r = 0.26$
Governors, Aug, 2008	Tanguison, Aug, 2009	$r = 0.20$
Governors, Oct, 2008	Tanguison, Jun, 2008	$r = 0.20$
Saipan, Oct, 2008	Tanguison, Jun, 2008	$r = 0.26$
Ipan, Jun, 2008	Governors, Oct, 2009	$r = 0.20$

7.3.3 Discriminant analysis of principal components

Scatter plots of genetic variation in multivariate space revealed considerable overlap between Guam adults, 2008 and 2009 recruits on Guam and recruits from Saipan

(Fig. 7.2). A failure to discriminate between these groups strongly suggests genetic homogeneity in space and time. Each group also occupies all quadrants of the plot and possesses individuals that lie well beyond the 95% inertia elipses of the group. The same may be said of the recruits alone, when they are segregated according to sample site (Fig. 7.3), with the exception of Pago 2008 and Pago 2009, which do not overlap and mostly occupy separate quadrants.

The k-means algorithm identified three genetic clusters (Fig. 7.4) but one of them contained only a single individual. The remaining two clusters were composed of individuals from all age classes, year classes and both contained fishes from Saipan. When subjected to DAPC these two genetic clusters were clearly differentiated by a single discriminant function (Fig. 7.5). DAPC can sometimes create artifactual clusters (Jombart *et al.* 2010), therefore these "non-geographic" clusters might not be a biologically significant pattern. However, to find a non-geographic signal in the nuclear genome of a *Naso* species may be more than a coincidence. It should also be noted that AMOVA performed on the same microsatellite data detected significant genetic differentiation between these two clusters ($\Phi_{st} = 0.04$, $p = 0.000$). There is no visible concordance between the non-geographic partitioning observed here and non-geographic structures found in the mtDNA. If these two genetic clusters are a true pattern in the data, they are a signal that has arisen beyond the spatial and temporal scale of this study.

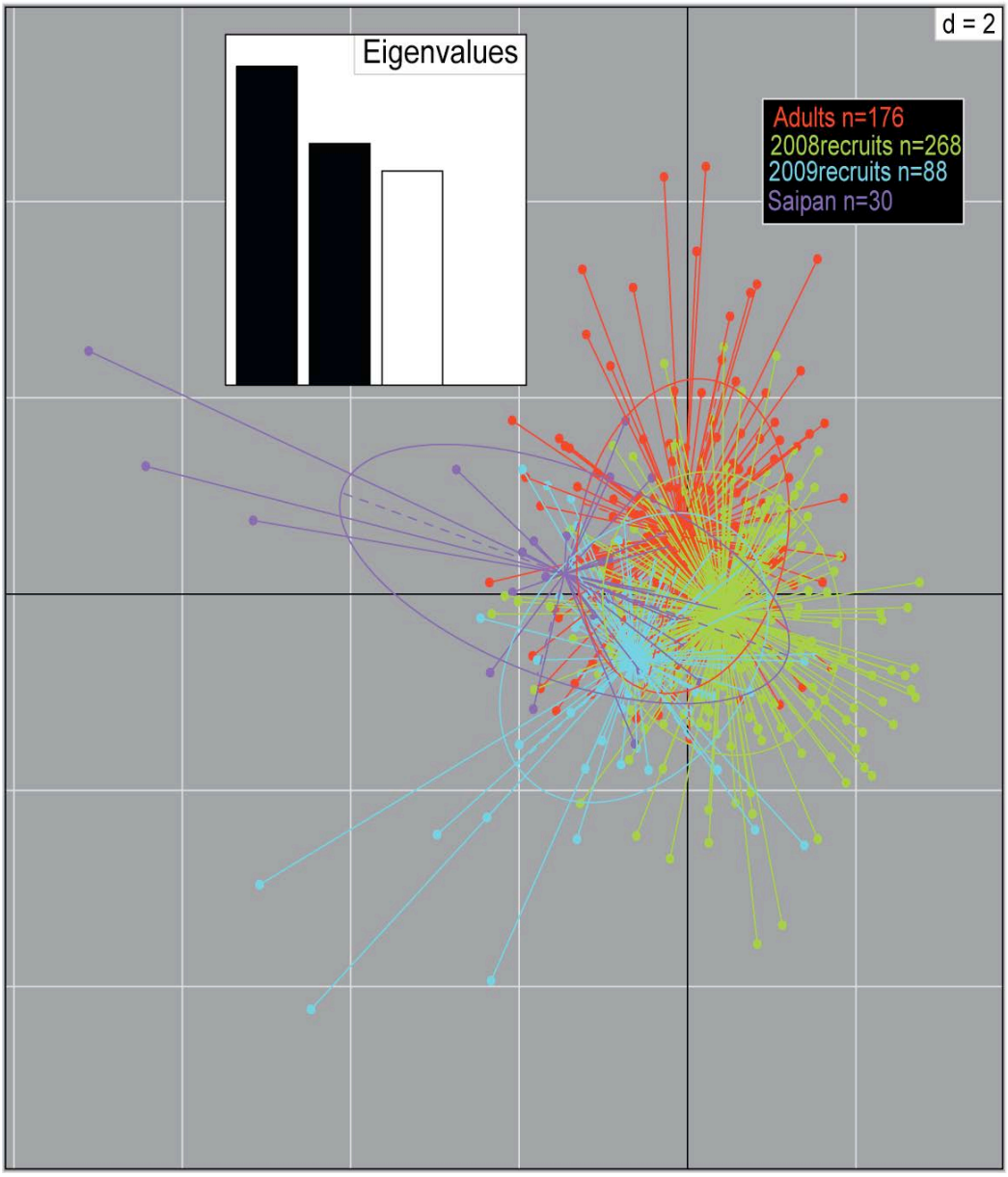


Figure 7.2: DAPC scatterplot of *N. unicornis* adults from the island of Guam and recruits from Guam and Saipan. Individual genotypes are represented by dots. Dots are grouped by location, age class or year class and represented by colors and 95% inertia elipses. Eigenvalues are displayed in the top left quadrant. The first two eigenvalues (black) show the amount of genetic information shown in the x and y-axes respectively.

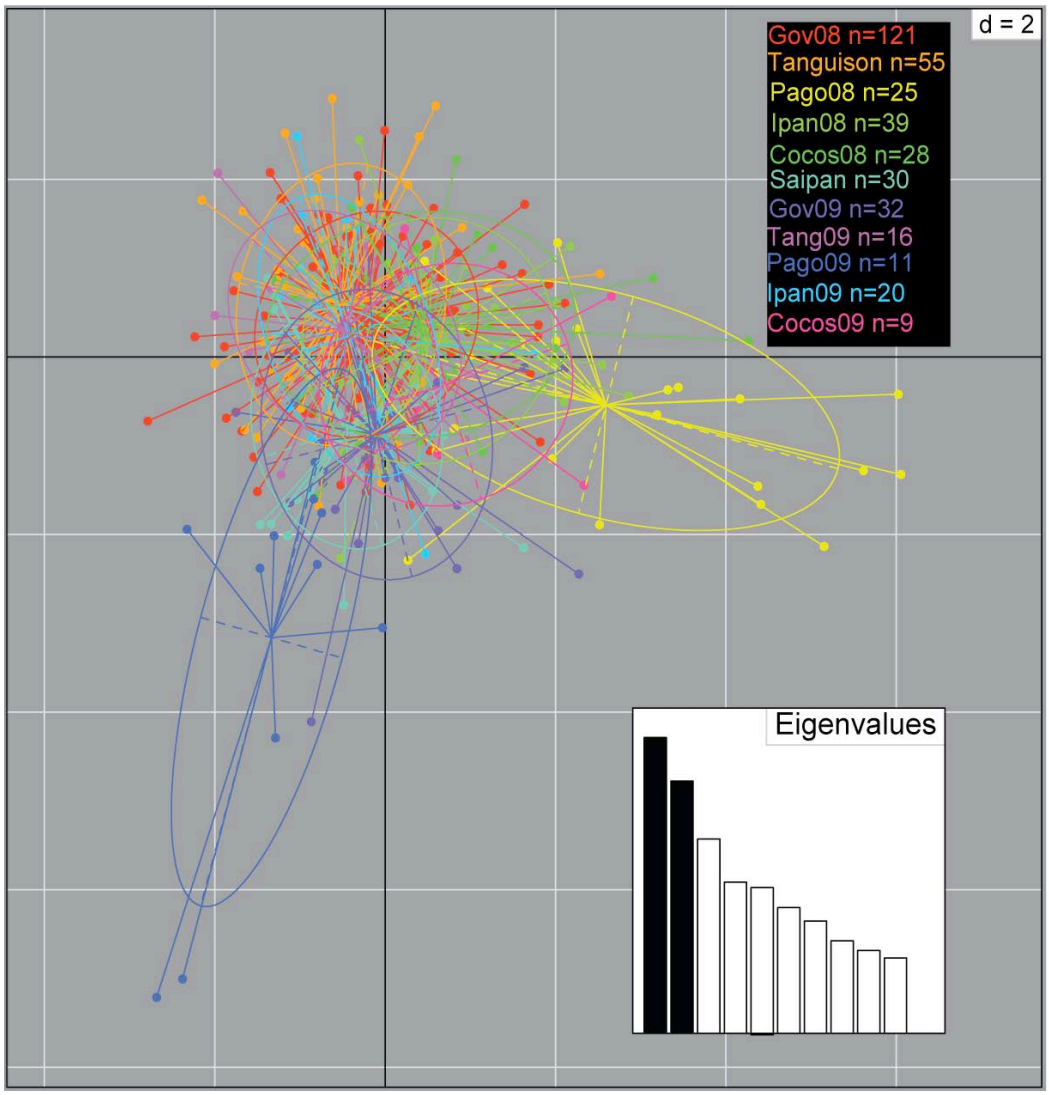


Figure 7.3: DAPC scatterplot of *N. unicornis* recruits from the island of Guam. Individual genotypes are represented by dots. Dots are grouped by sample site and years, represented by colors and 95% inertia elipses. Eigenvalues are displayed in the bottom right corner. The first two eigenvalues (black) show the amount of genetic information shown in the x and y-axes respectively.

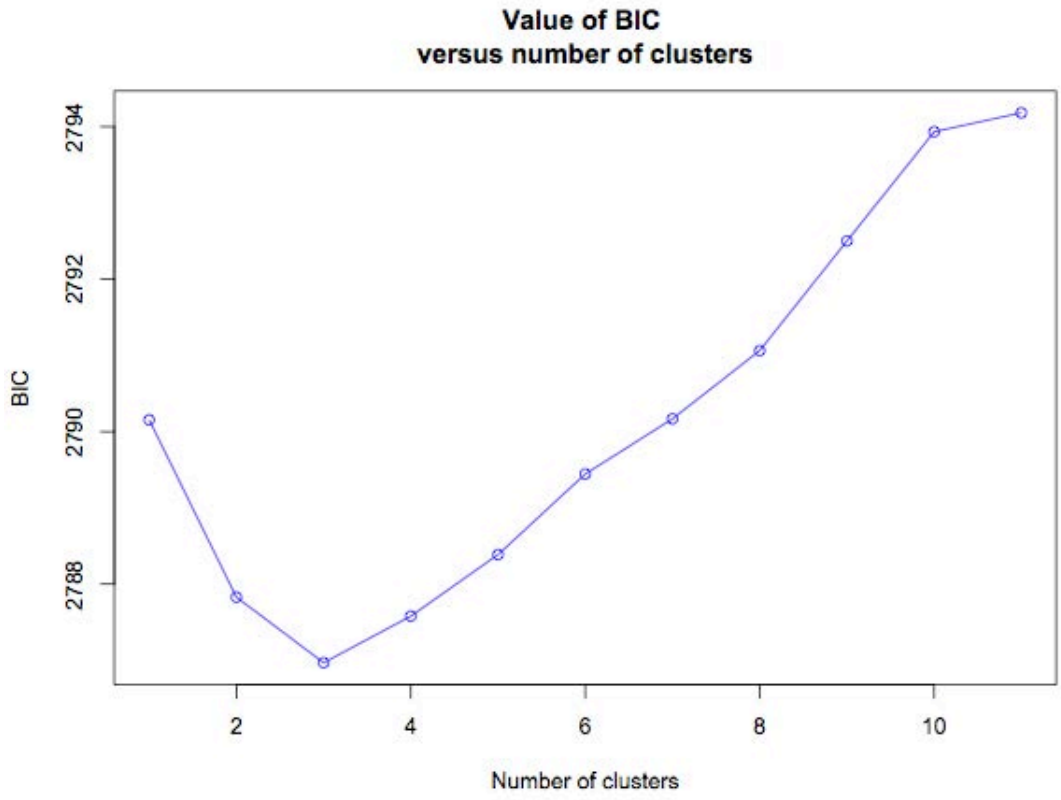


Figure 7.4: Likelihood of the number of genetic clusters based on the Bayesian information criterion (BIC), using the k-means algorithm (Jombart *et al.* 2010).

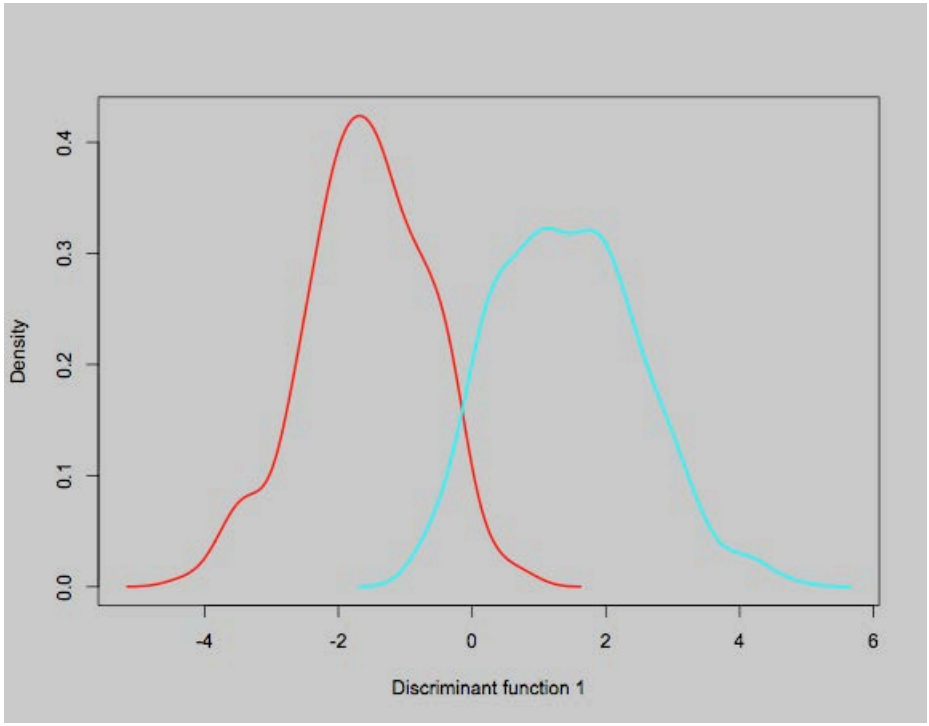


Figure 7.5: Discrimination of genetic cluster 1 (red) and genetic cluster 2 (blue). Here genetic variation between the two clusters can be represented on a single axis (discriminant function 1).

7.4 Discussion

7.4.1 Comparisons with previous studies

In many ways the results of this study and that of Planes *et al.* (2002) are complimentary. Planes *et al.* found that the mean relatedness of recruiting *N. unicornis* larvae in Moorea was $r = 0.002$ and was close enough to random to presume that larvae originated from different spawning events. Likewise, there was no evidence that recruits from Guam were any more closely related than random, and in many instances, mean relatedness in groups of larvae from the same reef flat was much less than zero. Yet, when Planes *et al.* grouped recruits based on otolith age they found that groups of recruits of the same age had a much higher relatedness ($r = 0.05 – 0.33$). In the present study recruits were not aged but at many locations recruits with pair wise relatedness values of 0.5 or higher were observed. It is unknown whether these individuals are the same age or not but they may likely be the progeny of single spawning events. However, unlike Planes *et al.* (2002) I surveyed multiple settlement sites around the island of Guam and observed equally high pair wise relatedness between recruits that settled many kilometers apart. Again these were not aged but the same logic applies. Even so, because I lacked demographic information from otolith analyses and I only used eight microsatellite loci, I opted for strict criteria, requiring $r > 0.19$ and shared mtDNA haplotypes, before I was willing to infer kinship. Only 13 pairs of recruits qualified as putative siblings under these criteria and of these only one pair was from the same settlement site (table 7.6). Planes *et al.* (2002) offered the first conclusive evidence that sibling reef fishes remained together throughout the pelagic larval phase and settled together (see also Buston *et al.* 2009). But from my data it also appears that few of the

surviving offspring of a single spawning choose to settle together, remain together for the duration of the pelagic larval phase or even settle synchronously.

Planes *et al.* (2002) also detected significant genetic structure between recruiting larval *N. unicornis* and juveniles on the same reef flat collected only ten days after larvae were captured in crest nets. Planes *et al.* explain this structure as the result of genetic drift due to the family-structure of pre-settlement larval pools and that mortality and mixing alter allele frequencies post settlement. Planes *et al.* further argue that the family-structure of larval pools and detectable genetic heterogeneity between cohorts of recruits and developmental stages is evidence of a stochastic sorting of genes during the pelagic larval phase also known as sweepstakes reproduction (Hedgecock *et al.* 2007). My samples were collected as post-settlement juveniles, so I am unable to comment on differences in allele frequency between the pre and post-settlement recruitment pools. However, based on the data of this study I question the notion that successive recruitment pulses of *N. unicornis* are structured in time by stochastic processes and sweepstakes reproduction for three reasons. First, even if pre-settlement larval pools contain disproportionate numbers of siblings, the extreme genetic diversity of the post-settlement recruits refutes that only a small number of progenitors contribute to the next generation's gene pool, regardless of whether recruits belong to the same larval pool or not. Allele frequencies in *N. unicornis* recruits may experience sorting following recruitment events due to high post-larval mortality (Doherty *et al.* 2004). Still, on Guam, old groups of large juveniles were just as genetically diverse as young groups of small juveniles (table 7.4; Fig. 7.1b). Second, if successive recruitment pulses are structured in time from stochastic processes one might expect more differentiation between entire year classes than was observed on Guam. Third, nearly half of the putative sibling pairs

were separated by year class. Clearly, many of the same progenitors are contributing to the gene pool year after year.

7.4.2 Self-recruitment and external recruitment

In spite of the intensity of *N. unicornis* recruitment and an adult population size kept unnaturally low by intense fishing pressure, there was no evidence that any parent-offspring pairs were sampled in this study. But the presence of self-recruitment cannot be rejected outright because of the lack of genetic differentiation between adults and recruits. Overlap of adult and juvenile genotype ordinates in reduced space (Fig. 7.2) might also suggest self-recruitment. Additionally, because some recruits from different year classes may be siblings I suspect that they may have been locally spawned, if not on Guam itself then elsewhere in the Marianas. Nevertheless, given the high genetic diversity of recruits, it seems unlikely that the massive recruitments of *N. unicornis* could be exclusively composed of locally spawned individuals. Most likely, there are a variety of sources for recruits, including self-recruitment. Multi-generational siblings could also be evidence of source-sink population dynamics, rather than local replenishment. Therefore, a significant portion of the recruits on Guam probably originated elsewhere.

The *N. unicornis* populations on Guam and Saipan, although separated by more than 200 km, appeared to be demographically connected, if not by direct migrant exchange then by drawing recruits from the same pools of migrants. Amazingly, three of the recruits sampled on Saipan (10%) were likely to have siblings that settled on Guam, suggesting that frequent migrant exchange on an ecological time scale occurs at least within the Marianas archipelago and probably further. From the perspective of mtDNA structure, Guam recruits were just as

genetically undifferentiated with adults from Tonga as they were with adults from Guam. Another line of evidence that may suggest migrant exchange occurs beyond the Marianas is the large number of private alleles observed (table 7.4). Some of these may have naturally low frequencies, or be underrepresented in the data due to sampling bias but it is also possible that some of these recruits are migrants originating in populations where these alleles were not so rare.

Recent studies propose that both self-recruitment and dispersal are important demographic processes for marine organisms (Planes *et al.* 2009; Jones *et al.* 2009). Concerning highly dispersive species, one modeling study, of Caribbean spiny lobsters, suggests a bimodal dispersal kernal, where ~60% of recruits settle < 450 km from where they were spawned and ~22% were predicted to disperse over 1000 km, with few at intermediate distances (Butler *et al.* 2011). A bimodal dispersal kernal for the *N. unicornis* population on Guam seems not only plausible but necessary, because habitat at a distance between 300 -1000 km is scarce. Hypothetically, dispersal of just over 1000 km in *N. unicornis* puts many areas in range of Guam: Iwo Jima and the Ogasawara islands to the north, Yap and Palau to the southwest, the Caroline islands and Pohnpei to the southeast. Dispersal on the order of 2000 km puts Guam within reach of the Philippines, New Guinea, Okinawa and the Marshall islands. And we know that the maximum dispersal potential for *Naso* species is much greater than this because some species have been reported as strays across the 5000 km expanse of the East Pacific Barrier (Randall 2002). Therefore, recruits sampled in this survey could have come from a large number of areas, especially those lying to the east from which the north equatorial current flows (see Eble *et al.* in press).

7.5 Conclusion

Massive recruitment events of the blue spine unicornfish, *N. unicornis* are genetically diverse assemblages of offspring from many spawning events and are not the progeny of a few fecund progenitors with skewed reproductive fitness due to stochastic survival of larvae and sweepstakes reproduction. Evidence suggests that some of these recruits maintain close proximity to siblings throughout the larval phase and settlement but that not all kin settle together. Moreover, kinship relationships between recruits can also exist between year classes suggesting that reproductive success from year to year is not purely stochastic. Evidence for self-recruitment on Guam was not strong but a portion of recruits are likely to have been spawned locally, suggesting that both self-recruitment and dispersal are demographically significant in this species, as has been proposed for other reef fishes (Planes *et al.* 2009). Isolated adult populations of *N. unicornis* in the Marianas archipelago almost certainly exchange migrants directly. The results of this study demonstrate that it is possible to observe kinship in highly dispersive coral reef fishes at a spatial scale of hundreds of km. Such results encouragingly suggest that close genetic relationships could be detected between individuals at an even greater scale.

Chapter Eight

Discussion and Synthesis of research

8.1 Introduction

The Klanten *et al.* (2007) paper on the broad-scale phylogeography of *N. vlamingii* was an intriguing study with unprecedented results. Klanten *et al.* (2007) served as a foundation for this research thesis and all of the questions asked were related to the unusual non-geographic patterns observed in *N. vlamingii*. Did other species exhibit non-geographic clades [chapter 3]? Are non-geographic clades only found in species with a semi-pelagic ecology [chapter 4]? Are non-geographic clades cryptic species [chapter 5]? How old are non-geographic clades [chapter 6]? Does non-geographic structure imply elevated levels of dispersal [chapter 7]? While many mysteries about the population dynamics of widespread coral reef fishes remain, the results of this thesis are informative, in the context of the *N. vlamingii* work and concerning coral reef fishes as a whole. In this concluding chapter I will briefly discuss the contributions of this research thesis to the greater field of coral reef fish biology and to draw overall conclusions about the main issues investigated.

8.2 Main contributions

1) The addition of four more widespread reef fishes to the growing corpus of species that have been genetically surveyed in both Indian and Pacific Oceans [chapters 3-5]. These data show that genetic homogeneity at large geographic scales is not just an artifact of extreme genetic variance in the mtDNA, rather, shared derived molecular characters found in populations from different oceans strongly argues that gene flow

does occur at this scale. Additionally, in 2010, two other papers were published, showing similar genetic uniformity across the Indo-Pacific in reef fishes from two other families, namely the Lutjanidae and Mureanidae (Gaither *et al.* 2010; Reece *et al.* 2010). Thus, with increasing amounts of evidence, the reality of high gene flow between the tropical Indian and Pacific Oceans by some tropical shallow water fishes now seems almost beyond question, though some authors might argue otherwise (see Carpenter *et al.* 2011).

2) More examples of non-geographic population patterns and a fossil calibrated molecular dating of non-geographic clades [chapter 6]. Therefore, the unique patterns observed in *N. vlamingii* may not be so unique after all. The underlying factors that cause these structures remain somewhat enigmatic but the molecular dating presented herein suggests that large biogeographic barriers are not required to explain non-geographic population patterns. Most likely, non-geographic clades are lineages that came into secondary contact following rapid range expansion and colonization of a very large biogeographic region [chapter 4]. Fundamentally, this amounts to founder effect in a peripatric context.

3) Evidence that, unlike many coral reef organisms, some *Naso* species could have experienced positive population growth during the Pleistocene [chapter 6]. This population growth could be due to the generalist ecology of *Naso* species. Alternatively, it is also possible that coral reef habitat was not as scarce as previously supposed as there is much evidence for drowned Pleistocene reefs in Hawaii (Webster *et al.* 2010; Faichney *et al.* 2011), Papua New Guinea (Webster *et al.* 2004) and even the Gulf of Carpentaria (Harris *et al.* 2008). Other papers suggest that

coalescent ages in reef fish are older than the Pilocene/Pleistocene (Haney *et al.* 2010), suggesting positive population growth throughout this period.

4) The non-monophyletic relationship of *Naso hexacanthus* and *N. caesius* in the mitochondrial genome [chapter 5]. This is an unusual result given that unlike many reef fish sister species that have a clear hybrid zone (Marie *et al.* 2007) *N. hexacanthus* and *N. caesius* are sympatric over a large geographic area. It is also unusual because sympatric sister species are often thought to have stronger reproductive barriers than allopatric ones (Coyne & Orr 2004). Yet, in spite of some evidence for prezygotic barriers in the form of assortative mating (Randall & Bell 1992), horizontal gene transfer between these two species may have occurred.

5) A comprehensive molecular study of *Naso unicornis,* from fossil calibrated molecular dating of the earliest node in a mtDNA gene tree [chapter 6], to broad scale phylogeography [chapter 4], to ecological patterns of recruitment [chapter 7]. Few marine fish species have been examined at so many different temporal and spatial scales as my treatment of *N. unicornis.* Also, in this thesis we provide the first microsatellite loci developed for *N. unicornis* [chapter 2]. This expansive approach has permitted insight into the patterns of connectivity at both evolutionary and ecological timescales and has illustrated that the lack of geographic structure observed for many *Naso* species, is due to both historic and ongoing connectivity in at least this first case study where it was possible to explore ecological connectivity as well.

6) Evidence for non-geographic clustering in the nuclear genome of *N. unicornis* [chapter 7]. While the validity of these findings remain uncertain, due to the highly sensitive nature of cluster analyses and the possibility of artificial patterns, non-geographic population structuring has only been known from mtDNA and I now report, for the first time, non-geographic patterns in the nuclear genome. Without corroboration from other species, the uncertainty surrounding this pattern in *N. unicornis* is tantamount to the initial interpretations of the Klanten *et al.* (2007) study of *N. vlamingii*. However, it does suggest that the issue of non-geographic patterning is something that will continue to appear in genetic studies of coral reef fishes.

8.3 Discussion and conclusions

What are non-geographic clades?

Since the commencement of this thesis several additional papers reporting non-geographic clades in a variety of coral reef fish taxa have been published (Evans *et al.* 2010; Visram *et al.* 2010; Winters *et al.* 2010; Reece *et al.* 2010). Therefore, in addition to the other *Naso* taxa surveyed in this thesis, the non-geographic population structure of *N. vlamingii* does not appear to be unique to this species, nor is it merely a stochastic signal due to elevated genetic diversity in a single mitochondrial gene region.

In order to understand the biological significance of this pattern, it becomes necessary to ask: what do all of the species that have non-geographic structure share in common? 1) All species with non-geographic structuring are Indo-Pacific species, no examples of Atlantic reef fishes with this type of pattern have been reported. 2) All species with non-geographic structuring are widespread, being found in both Indian and Pacific Oceans. 3) All species with non-geographic structuring are

presumed to have larvae that are highly dispersive (especially *Naso* and *Gymnothorax*; Reece *et al.* 2010). 4) All species with non-geographic clades show a strong signal of demographic, and by extension, range expansion. Curiously, there is not a lot of congruence in the age of non-geographic clades among the various taxa. While the clades of Scarids and Mureanids both appear to be of Pleistocene origin, *Naso* clades are probably much older, and span a very large time period [chapter 6].

If the ages of non-geographic clades among the various taxa give no great insight into the origins of these genetic structures then low-sea-level barriers to dispersal are probably not the most likely explanation. In fact, simple drift-migration population genetics may be insufficient for describing these structures. Genetic processes relating to range expansion (reviewed by Excoffier *et al.* 2009), may be more appropriate. As discussed in chapter three, long distance colonization can lead to the founding of pioneer populations, which are genetically differentiated through founder effect (Nichols & Hewitt 1994; Ibrahim *et al.* 1996) and these populations can independently expand creating a genetic embolism (Bialozyt *et al.* 2006; Excoffier *et al.* 2009; Fayard *et al.* 2009). Potentially, genetic embolisms could explain many population genetic breaks in marine organisms that are currently difficult to explain. However, gene flow will eventually overcome a genetic embolism given a high enough rate of migration or a sufficient amount of time (Excoffier *et al.* 2009; Fayard *et al.* 2009). Therefore, in highly dispersive species, such as *Naso*, with populations that may have at one time become genetically differentiated through rapid expansions, the geographic ranges of all lineages eventually come to overlap completely.

Non-geographic population structures are a feature of Indo-Pacific coral reef fishes. Perhaps it is the vastness of the Indo-Pacific that has influenced these genetic

patterns. Certainly, the enormous size of the Indo-Pacific would provide opportunity for populations to expand their ranges at an impressive geographic scale, if the dispersal ability of the organism allowed for it. For *Naso* this appears to be the case. Apart from a long PLD (Wilson & McCormick 1999), morphologically, Acanthurid larvae have all the prerequisites of a long-distance disperser, including high aspect ratio caudal fins, narrow caudal peduncle, large propulsive area and large anterior body depth (Fisher *et al.* 2005; Fisher & Hogan, 2007). These morphological adaptations hydromechanically maximize thrust while minimizing drag and are also characteristic of migratory pelagic fishes (Tyler et al. 1989; Blake, 2004). Also, the results of chapter 7 strongly suggest that the distances of over 200 km (between Guam and Saipan) are trivial in comparison to the dispersal range of these larvae. Therefore, physical barriers to gene flow are not required to explain non-geographic population structuring, which may result from rapid range expansion and long distance dispersal in a very large biogeographic region. Spatial bottlenecks, or narrow migration corridors, such as the IPB, that may have acted as barriers for other species could have exacerbated the effects of expansion (Excoffier *et al.* 2009), but the idea of genetic drift across an absolute barrier does not seem likely. Ergo, structuring is not due to the slow process of genetic drift but rather a stochastic sorting of available polymorphism into new habitats through colonization (Wakeley & Aliacar 2001).

The signal of expansion in gene genealogies is unmistakably a star or comb shaped phylogeny (Slatkin & Hudson 1991), which is evident in all taxa exhibiting non-geographic clades. This shape is formed because most of the coalescent events in the genealogy happened around the onset of expansion (Ray *et al.* 2003; Excoffier 2004). As time goes on coalescent events in the gene genealogy become increasingly

rare resulting in a large number of unique or singleton haplotypes present in a sample (Harpending & Rogers 2000; Excoffier *et al.* 2009). In the *Naso* phylogeographies the overwhelming majority of individuals possessed unique control region haplotypes but what is more, all reef fishes with non-geographic clades contain a number of basal individuals that do not belong to a non-geographic clade. In *N. vlamingii, N. brevirostris* and *Scarus ghobban* these basal individuals are few in number but in *N. unicornis* and *Scarus psittacus* a substantial portion of individuals do not group into clades. Branch lengths for these individuals are generally long, suggesting that they are rare lineages that diverged long ago, around the time of initial range expansion of the species. Why then are some lineages (i.e. the non-geographic clades) so much more abundant than others? As previously discussed, non-geographic clades could likely be pioneer populations that originated as small number of founders but later expanded across the Indo-Pacific. Other explanations are also possible, including small-scale geographic isolation or genetic hitchhiking [chapter 3].

The age of expansion in *Naso* species and clades appears to be old. Both mismatch based estimation of expansion age using a fixed mutation rate and fossil calibrated molecular dating using a relaxed clock placed much of the early demographic events within the Miocene. Perhaps more important than the estimated historical demography by complex analyses is the relative ages of *Naso* compared to other reef fishes. The phylogeography of *S. psittacus* (Winters *et al.* 2010) also employs the same coalescent methods as Klanten *et al.* (2007) for the mitochondrial control region. Yet expansion in *S. psittacus* appears to be much younger, more than two orders of magnitude younger than *N. unicornis.* Likewise, I performed some limited analyses on *Naso* using Visram *et al.* (2010) methods (see chapter 6). Again

expansion in *Naso* was much older than what is reported for *S. ghobban.* Presuming that the substitution rate in *Naso* is not dramatically higher than in *Scarus* and other reef fishes that have been studied using a range of substitution rates (e.g. Bowen *et al.* 2006; Craig et al. 2007; Eble *et al.* 2010), *Naso* appears to be older than many other reef fishes.

8.4 Final conclusions

Non-geographic clades are a genetic feature, primarily of mtDNA gene genealogies, of some widespread Indo-Pacific coral reef fish populations. Early papers reporting this genetic phenomenon considered it unusual but the results of this thesis and other recent publications suggest that non-geographic clades will continue to be manifested in reef fish phylogeography studies. Considering that but few cosmopolitan reef fish species, from only a handful of families have been genetically surveyed across the Indo-Pacific, much more work will be required to fully understand the prevalence of this pattern. In this regard certain widespread members of the family Balistidae may show promise as well as some reef associated Carangids. Deep water *Naso,* such as *Naso annulatus,* are of particular interest to me but to find non-geographic patterns in these species would not be revealing.

Why some Indo-Pacific reef fishes show non-geographic phylogeographic patterns is not entirely certain. Some authors (Klanten *et al.* 2007; Winters *et al.* 2010; Reece *et al.* 2010) have proposed historical barriers to gene flow but there is no direct evidence linking non-geographic genetic structures to a barrier. Based on the results of this thesis, colonization through long distance dispersal in a large biogeographic region and multiple subsequent expansions is the smoking gun. Certainly, *Naso* display strong genetic hallmarks of expansion, hence, at the present

time this appears to be the best explanation. Rapid expansions would have been made possible by highly dispersive pelagic larvae, but apart from this few ecological explanations seem related to non-geographic clades.

For *Naso*, the extent to which non-geographic clades can be treated as independent evolutionary units seems small and the possibility of these lineages being sympatrically distributed cryptic species seems extremely remote. The absence of reciprocal monophyly between *N. hexacanthus* and *N. caesius* makes the evolutionary independence of the non-geographic clades of other species dubious. However, further investigations of the intriguing genetic relationship between these two sister species, especially using nuclear markers, should be a research priority. Hence, non-geographic population patterns in *Naso* appear to be little more a signal of extensive population expansions facilitated by a highly dispersive pelagic larval phase.

References

Allen GR (2008) Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 541-556.

Allen GR & Erdmann MV (2009) Reef fishes of the Bird's Head Peninsula, West Papua, Indonesia. *Check List* 5,**3**, 587-628.

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 Science*, **52**, 1720–1732.

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

Avise JC (1998) The history and purview of phylogeography: a personal reflection. *Molecular Ecology*, **7**, 371-379.

Avise JC (2000) *Phylogeography: The History and Formation of Species.* Harvard University Press, Cambridge, Massachusetts.

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

Bandelt HJ, Forster P, Sykes BC, Richards MB (1995) Mitochondrial portraits of human populations using median networks. *Genetics*, **141**, 743-753.

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.

Barrowclough GF & Zink RM (2009) Funds enough and time: mtDNA, nuDNA and the discovery of divergence. *Molecular Ecology*, **18**, 2934-2936.

Bay LK, Choat JH, van Herwerden L & Robertson DR (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.

Bellwood DR (1996) The Eocene fishes of Monte Bolca: the earliest coral reef fish assemblage. *Coral Reefs*, **15**, 11-19.

Bellwood DR (1998) What are reef fishes? - Comment on the report by D. R. Robertson: Do coral-reef fish faunas have a distinctive taxonomic structure? *Coral Reefs*, **17**, 187-189.

Bellwood DR, Wainwright PC (2002) The history and biogeography of fishes on coral reefs. In Sale, P.F. (Ed.), *Coral reef fishes: dynamics and diversity in a complex ecosystem.* Academic press, San Diego, pp 5-32.

Bellwood DR, van Herwerden L & Konow N (2004) Evolution and biogeography of marine angelfishes (Pisces: Pomacanthidae). *Molecular Phylogenetics and* Evolution, **33**, 140-155.

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

Bellwood DR, Klantens S, Cowman PF, Pratchett MS, Konow N & van Herwerden L (2010) Evolutionary history of butterflyfishes (f: Chaetodontidae) and the rise of coral feeding fishes. *Journal of Evolutionary Biology*, **23**, 335-349.

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

Bernardi G, Bucciarelli G, Costagliola D, Robertson DR & Heiser JB (2004) Evolution of coral reef fish *Thalassoma* spp [Labridae]. 1. Molecular phylogeny and biogeography. *Marine Biology*, **144**, 369-375.

Bialozyt R, Zieganhagen B, Petit RJ (2006). Contrasting effects of long distance seed dispersal on genetic diversity during range expansion. *Journal of Evolutionary Biology*, **19**, 12-20.

Birky WC, Fuerst P & Maruyama T (1989) Organelle gene diversity under migration, mutation and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells and comparison to nuclear genes. *Genetics*, **121**, 613-627.

Bolnick DI & Fitzpatrick BM (2007) Sympatric speciation: models and empirical evidence. *Annual review of Ecology Evolution and Systematics*, **38**, 459-487.

Bowen, BW, Bass AL, Rocha LA, Grant WS & Robertson DR (2001) Phylogeography of the trumpetfishes (Aulostomus): ring species complex on a global scale. *Evolution*, **55**, 1029-1039.

Bowen BW, Muss A, Rocha LA & Grant WS (2006) Shallow mtDNA coalescence in Atlantic pygmy angelfishes (Genus *Centropyge*) indicates a recent invasion from the Indian Ocean. *Journal of Heredity*, **97**, 1-12.

Budd AF (2000) Diversity and extinction in the Cenezoic history of Caribbean coral reefs. *Coral Reefs*, **19**, 25-35.

Budd AF & Pandolfi JM (2010) Evolutionary novelty is concentrated at the edge of coral species distributions. *Science*, **328**, 1558-1560.

Buston PM, Fauvelot C, Wong MYL & Planes S (2009) Genetic relatedness in groups of the humbug damselfish *Dascyllus aruanus*: small similar sized individuals may be close kin. *Molecular ecology*, **18**, 4707-4715.

Butler MJ, Paris CB, Goldstein JS, Matsuda H & Cowen RK (2011) Behavior constrains the dispersal of long-lived spiny lobster larvae. *Marine Ecology Progess Series*, **422**, 223-237.

Carpenter KE & Springer VG (2005) The center of the center of marine shore fish biodiversity: the Philippine islands. *Environmental biology of fishes*, **72**, 467-480.

Carpenter KE, Barber PH, Crandall ED, Ablan-Lagman MCA, Ambariyanto, Mahardika GN, Manjaji-Matsumoto BM, Juinio-Menez MA, Santos MD, Starger CJ & Toha AHA (2011) Comparative phylogeography of the coral triangle and implications for marine management. *Journal of Marine Biology*, Article ID 396982, 14 pages, doi:10.1155/2011/396982.

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

Chapuis MP & Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, **24**, 621-631.

Charlesworth D (2010) Don't forget the ancestral polymorphisms. *Heredity*, **104**, 1-2.

Chave EH & Mundy BC (1994) Deep sea benthic fish of the Hawaiian archipelago, Cross Seamount and Johnston Atoll. *Pacific Science*, **48**, 367-409.

Choat JH & Bellwood DR (1991) Reef fishes: their history and evolution. In Sale, P.F. (Ed.), *The ecology of fishes on coral reefs.* Academic press, San Diego, pp. 39-61.

Choat JH & Axe LM (1996) Growth and longevity in acanthurid fishes; an analysis of otolith increments. *Marine Ecology Progress Series*, **134**, 15-26.

Choat JH, Clements KD & Robbins WD (2002) The trophic status of herbivorous fishes on coral reefs. *Marine Biology*, **140**, 613-623.

Choat JH (2006) Phylogeography and reef fishes: bringing ecology back into the argument. *Journal of Biogeography*, **33**, 967-968.

Christie MR (2009) Parentage in natural populations: novel methods to detect parent offspring pairs in large data sets. *Molecular Ecology Resources*, **10**, 115-128.

Christie MR, Johnson DW, Stallings CD & Hixon MA (2010) Self-recruitment and sweepstakes reproduction amid extensive gene flow in a coral reef fish. *Molecular Ecology*, **19**, 1042-1057.

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

Cowman PF, Bellwood DR & van Herwerden L (2009) Dating the evolutionary origins of wrasse lineages (Labridae) and the rise of trophic novelty on coral reefs. *Molecular phylogenetics and evolution*, **52**, 621-631.

Coyne JA & Orr HA (2004) *Speciation*. Sinauer Associates Inc., Sunderland, Massachusetts.

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.

Craig MT, Graham RT, Torres RA, Hyde JR, Freitas MO, Ferriera BP, Hostim-Silva M, Gerhardinger LC, Bertoncini AA & Robertson DR (2009) How many species of Goliath Grouper are there? Cryptic genetic divergence in a threatened marine fish and the resurrection of a geopolitical species. *Endangered Species Research*, **7**, 167-174.

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

Crow KD, Munehara H & Bernardi G (2010) Sympatric speciation in a genus of marine fishes. *Molecular Ecology*, **19**, 2089-2105.

Dayton C, Santayana AC & Lacson JM (1994) Genetic evidence for reproductive isolation of the recently described unicornfish *Naso caesius* and its sibling *N. hexacanthus*. *Marine Biology*, **118**, 551-554.

Dayton CL (2001) Genetic evolution among selected members of the genus *Naso* (Nasinae), "unicornfishes" from Guam. *Marine Biology*, **139**, 771-776.

De Aguiar MAM, Baranger M, Baptestini EM, Kaufman L & Bar-Yam, Y., 2009. Global patterns of speciation and diversity. *Nature*, **460**, 384-387.

De Queiroz K (2007) Species concepts and species delimitation. *Systematics Biology*, **56**, 879-886.

DiBattista JD, Wilcox C, Craig MT, Rocha LA & Bowen B (2011) Phylogeography of the Pacific blueline surgeonfish, *Acanthurus nigroris*, reveals high genetic

connectivity and a cryptic endemic species in the Hawaiian archipelago. *Journal of Marine Biology*, Article ID 839134, 17 pages, doi:10.1155/2011/839134.

Doherty PJ, V. Dufour, R. Galzin, M. A. Hixon, M. G. Meekan, S. Planes (2004) High mortality during settlement is a population bottleneck for a tropical surgeonfish. *Ecology*, **85**, 2422-2428.

Donaldson KA & Wilson RR (1999) Amphi-panamic geminates of snook (Percoidei: Centropomidae) provide a calibration of the divergence rate in the mitochondrial DNA control region of fishes. *Molecular Phylogenetics and Evolution*, **13**, 208-213.

Dornburg A, Santini F & Alfaro ME (2008) The influence of model averaging on clade posteriors: an example using the triggerfishes (Family Balistidae). *Systematics Biology*, **57**, 905-919.

Drummond AJ, Rambaut A, Shapiro B, & Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, **22**, 1185-1192.

Drummond AJ & Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**, 214.

Eble JA, Rocha LA, Craig MT & Bowen BW (2011) Not all larvae stay close to home: insights into marine population connectivity with a focus on the brown surgeonfish, (*Acanthurus nigrofuscus*). *Journal of marine biology*, Article ID 518516, 12 pages, doi:10.1155/2011/518516.

Eble JA, Toonen RJ, Sorenson L, Basch LV, Papastamatiou YP & Bowen BW (in press) Escaping paradise: larval export from Hawaii in an Indo-Pacific reef fish, the Yellow Tang Zebrasoma flavescens. *Marine Ecology Progress Series*, doi: 10.3354/meps09083.

Edwards SV & Beerli P (2000) Gene divergence, population diverence and the variance in coalescent time in phylogeographic studies. *Evolution*, **54**, 1839-1854.

Elmer KR & Meyer A (2010) Sympatric speciation without borders? *Molecular Ecology*, **19**, 1991-1993.

Eschmeyer WN & Fricke R (eds.) Catalog of Fishes electronic version (updated 2 July 2009). http://research.calacademy.org/ichthyology/catalog/fishcatsearch.html.

Evans RD, van Herwerden L, Russ GR & Frisch AJ (2010) Strong genetic but not spatial subdivision of two reef fish species targeted by fishers on the Great Barrier Reef. *Fisheries Research*, **102**, 16-25.

Excoffier L, Smouse PE & Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Excoffier L (2004) Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Molecular Ecology, **13**, 853-864.

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

Excoffier L & Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology and Evolution*, **23**, 347-351.

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. *Annual Reviews of Ecology, Evolution and Systematics*, **40**, 481-501.

Faichney IDE, Webster JM, Clague DA, Braga JC, Renema W & Potts DC (2011) The impact of mid-Pleistocene transition on the composition of submerged reefs of the Maui-Nui complex, Hawaii. *Paleogeography, Paleoclimatology and Paleoecology*, **299**, 493-506.

Faircloth BC (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus specific primer design. *Molecular Ecology Resources*, **8**, 92-94.

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.

Fayard J, Klein EK & Lefevre F (2009) Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary Biology*, **22**, 2171-2182.

Fisher R, Leis JM, Clark DL & Wilson SK (2005) Critical swimming speeds of late-stage coral reef fish larvae: variation within species, among species and between locations. *Marine Biology*, **147**, 1201-1212.

Fisher R & Hogan JD (2007) Morphological predictors of swimming speed: a case study of pre-settlement juvenile coral reef fishes. *Journal of Experimental Biology*, **210**, 2436-2443.

Floeter S, Rocha LA, Robertson DR, Joyeux JC, Smith-Vaniz W, Wirtz P, Edwards AJ, Barreiros JP, Ferriera CEL, Gasparini JL, Brito A, Falcon M, Bowen BW & Bernandi G (2008) Atlantic reef fish biogeography and evolution. *Journal of Biogeography*, **35**, 22-47.

Fu YX & Li WH (1993) Statistical tests of neutrality of mutations. *Genetics*, **133**, 693-709.

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

Gaither MR, Toonen RJ, Robertson DR, Planes S & Bowen BW (2010) Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulva*). *Journal of Biogeography*, **37**, 133-147.

Galtier N, Nabholz B, Glemin S & Hurst GDD (2009) Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, **18**, 4541-4550.

Geyer LB & Lessios HA (2009) Lack of character displacement in the male recognition molecule, bindin, in Atlantic sea urchins of the genus *Echinometra*. *Molecular Biology and Evolution*, **26**, 2135-2146.

Gillespie JH (2000) Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics*, **155**, 909-919.

Glenn TC & Schable NA (2005) Isolating microsatellite DNA Loci. *Methods in Enzymology*, **395**, 202-222.

Goudet J (2001) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www.unil.ch/izea/softwares/fstat.html.

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

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95-98.

Hanebuth T, Stattegger K & Grootes PM (2000) Rapid flooding of the Sunda Shelf: a late-glacial sea-level record. *Science*, **288**, 1033-1035.

Haney RA, Silliman BR & Rand DM (2010) Effects of selection and mutation on mitochondrial variation and inferences of historical population expansion in a Caribbean reef fish. *Molecular Phylogenetics and Evolution*, **57**, 821-828.

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

Harpending HC & Rogers A (2000) Genetic perspectives on human origins and differentiation. *Annual Reviews of Genomics and Human Genetics*, **1**, 361-385.

Harris PT, Heap AD, Marshall JF & McCulloch M (2008) A new coral reef province in the Gulf of Carpentaria: colonization growth and submergence during the early Holocene. *Marine Geology*, **251**, 85-97.

Hedgecock D, Barber PH & Edmands S (2007) Genetic approaches to measuring marine connectivity. *Oceanography*, **20**, 70-79.

Hellberg ME (2007) Footprints on the water: the genetic wake of dispersal among reefs. *Coral Reefs*, **26**, 463-473.

Hellberg ME 2009. Gene flow and isolation among populations of marine animals. *Annual Review of Ecology Evolution and Systematics*, **40**, 291–310.

Hepburn RI, Sale PF, Dixon B & Heath DD (2009) Genetic structure of juvenile cohorts of bicolor damselfish (*Stegastes partitus*) along the Meso-American barrier reef: chaos through time. *Coral* Reefs, **28**, 277-288.

Herbert PDN, Ratnasignham S & deWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closesly related species. *Proceedings of the Royal Society of London series B*, **270**, S96-S99.

Herbert TD, Peterson LC, Lawrence KT & Liu Z (2010) Tropical ocean temperatures over the past 3.5 million years. *Science*, **328**, 1530-1534.

Hidaka K, Iwatsuki Y & Randall JE (2003) A review of the Indo-Pacific bonefishes of the *Albula argentea* complex, with a description of a new species. *Ichthyological research*, **55**, 53-64.

Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG & Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673-4680.

Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution*, **22**, 1561-1568.

Ho SYW, Shapiro B, Phillips MJ, Cooper A, Drummond AJ (2007) Evidence for time dependency of molecular rate estimates. *Systematics Biology*, **56**, 515-522.

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

Hoey AS & Bellwood DR (2009) Limited functional redundancy in a high diversity ecosystem: single species dominates key ecological process on coral reefs. *Ecosystems*, **12**, 1316-1328.

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

Hyde JR, Kimbrell A, Budrick JE, Lynne EA & Vetter RD (2008) Cryptic speciation in the vermillion rockfish (*Sebastes miniatus*) and the role of bathymetry in the speciation process. *Molecular Ecology*, **17**, 1122-1136.

Ibrahim KM, Nichols RA & Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282-291.

Ihaka R & Gentleman R (1996) R: a language for data analysis and graphics. *Journal of Computer Graphics and Statistics*, **5**, 299-314.

Jackson JBC & Johnson KG (2000) Life in the last few million years. *Paleobiology*, **26**, 221-235.

Jensen JL, Bohonak AJ, Kelley ST (2005) Isolation by distance, web service. *BMC Genetics*, **6**, 13. v.3.14. Available from: <http://ibdws.sdsu.edu/>.

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

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403-1405.

Jombart T, Devillard S & Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC genetics, **11**, 94-109.

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

Jones GP, Almany GR, Russ GR, Sale PF, Steneck RS, van Oppen MJH & Willis BL (2009) Larval retention and connectivity among populations of corals and reef fishes: history, advances and challenges. *Coral Reefs*, **28**, 307-325.

Keissling W (2009) Geologic and biologic controls on the evolution of reefs. *Annual Reviews of Ecology Evolution and Systematics*, **40**, 173-192.

Keissling W, Simpson C, Foote M (2010) Reefs as cradles of evolution and sources of biodiversity in the Phanerozoic. *Science*, **37**, 196-198.

Kelly RP, Palumbi SR 2009. General-use polymerase chain reaction primers for amplification and direct sequencing of enolase, a single copy nuclear gene, from different animal phyla. *Molecular Ecology Resources*, **9**, 144-147.

Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutation. *Genetics*, **61**, 893-903.

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.

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

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of

population parameters. *Bioinformatics*, **22**, 768–770.

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

Lear CH, Bailey TR, Pearson PN, Coxall HK & Rosenthal Y (2008) Cooling and ice growth across the Eocene-Oligocene transition. *Geology*, **36**, 251-254.

Leis JM (1991) The pelagic stage of reef fishes: the larval biology of coral reef fishes. In: Sale, P.F. (Ed.), *The Ecology of Fishes on Coral Reefs*. Academic Press, San Diego, pp. 183–230.

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

Lessios HA, Kane J & Robertson DR (2003) Phylogeography of the pantropical sea urchin Tripneustes: contrasting patterns of population structure between oceans. *Evolution*, **57**, 2026–2036.

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

Lessios HA (2008) The great American Schism: divergence of marine organisms after the rise of the Central American Isthmus. *Annual Review of Ecology, Evolution and Systematics*, **39**, 63-91.

Li WH (1977) Distribution of nucleotide differences between two randomly chosen cistrons in a finite population. *Genetics*, **85**, 331-337.

Lieske E & Myers R (1994) Collins Pocket Guide. Coral reef fishes. *Indo-Pacific & Caribbean including the Red Sea*. Haper Collins Publishers, p. 400.

Lima D, Freitas JEP, Araujo ME & Sole-Cava AM (2005) Genetic detection of cryptic species in the frillfin goby *Bathygobius soporator*. *Journal of Experimental Marine Biology and Ecology*, **320**, 211-223.

Lin HC, Sanchez-Ortiz C & Hastings PA (2009) Color variation is incongruent with mitochondrial lineages: Crypric speciation and subsequent diversification in a Gulf of California reef fish (Teleostei: Blenioidei) *Molecular Ecology*, **18**, 2476-2488.

Lowe WH & Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038-3051.

Malay MG & Paulay G (2010) Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: Calcinus). *Evolution*, **64**, 643-672.

Manel S, Gaggiotti OE & Waples, RS (2005) Assignement methods: matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, **20**, 136-142.

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

Masuda H, Amaoka K, Araga C, Uyeno T & Yoshino T (1984) *The Fishes of the Japanese Archipelago*. Vol. 1. Tokai University Press, Tokyo, Japan. p 437.

Maynard-Smith J (1989) The causes of extinction. *Philosophical Transactions of the Royal Society of London Series B*, **325**, 241-52.

Meiklejohn CD, Montooth KL & Rand DM (2007) Positive and negative selection on the mitochondrial genome. *Trends in Genetics*, **23**, 259-263.

Meyer CG & Holland KN (2005) Movement patterns, home range size and habitat utilization of the bluespine unicornfish, *Naso unicornis* (Acanthuridae) in a Hawaiian marine reserve. *Environmental Biology of Fishes*, **73**, 201-210.

Meyer CG, Papastamatiou YP & Clark TB (2010) Differential movement patterns and site fidelity among trophic groups of reef fishes in a Hawaiian marine protected area. *Marine Biology*, **157**, 1499-1511.

Miller MP, Bellinger RM, Forsman ED & Haig SM (2006) Effects of historical climate change, habitat connectivity and vicariance on the genetic structure and diversity across the range of the red tree vole (*Phenacomys longicaudus*). Molecular Ecology, **15**, 145-159.

Mora MS, Lessa EP, Cutrera AP, Kittlein MJ & Vasallo AI (2007) Phylogeographical structure in the subterranean tuco-tuco *Ctenomys talarum* (Rodentia: Ctenomyidae): contrasting the demographic consequences of regional and habitat specific histories. *Molecular Ecology*, **16**, 3453-3465.

Moritz C, Dowling TE & Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Reviews of Ecology Evololution and Systematics*,**18**, 269-292.

Muller RD, Sdrolias M, Gaina C, Steinberger B & Heine C (2008) Long-term sea-level fluctuations driven by ocean basin dynamics. *Science*, **319**, 1357-1362.

Munday PL, van Herwerden L & Dudgeon CL (2004) Evidence for sympatric speciation by host shift in the sea. *Current Biology*, **14**, 1-20.

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

Nichols RA & Hewitt GM (1994) The genetic consequences of long distance dispersal during colonization. *Heredity*, **72**, 312-317.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547-72.

Pannell JR & Charlesworth B (2000) Effects of metapopulation processes on measures of genetic diversity. Philosophical Transactions of the Royal Society of London B, **355**, 1851-1864.

Peakall R & Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288-295.

Peterson GI & Masel J (2009) Quantitative prediction of molecular clock and K_a/K_s at short timescales. *Molecular Biology and Evolution*, **26**, 2595-2603.

Pheiler E, Bitler BG, Ulloa R, van der Heiden AM & Hastings PA (2008) Molecular identification of the bonefish *Albula esuncula* (Albuliformes: Albulidae) from the tropical eastern Pacific, with comments on distribution and morphology. *Copeia*, **2008**, 763-770.

Pianka ER (1978) *Evolutionary Ecology.* Harper & Row Publishers, New York.

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.

Planes S, Lencaillon G, Lenfant P & Meekan M (2002) Genetic and demographic variation in new recruits of *Naso unicornis*. *Journal of Fish Biology*, **61**, 1033-1049.

Planes S, Jones GP & Thorrold SR (2009) Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Science* USA, **106**, 5693-5697.

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

Posada D (2008) jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**, 1253-1256.

Prothero DR (1994) The late Eocene-Oligocene extinctions. *Annual Reviews of Earth and Planetary Science*, **22**, 145-165.

Puebla O, Bermingham E, Guichard F & Whiteman E (2007) Color pattern as a single trait driving speciation in *Hypoplectrus* coral reef fishes? *Proceedings of the Royal Society of London B*, **274**, 1265-1271.

Puebla O (2009) Ecological speciation in marine v. freshwater fishes. *Journal of Fish Biology*, **75**, 960-996.

Purcell JFH, Cowen RK, Hughes CR & Williams DA (2006) Weak genetic structure indicates strong dispersal limits: a tale of two coral reef fish. *Proceedings of the Royal Society of London B*, **273**, 1483-1490.

Rabosky DL (2009) Extinction rates should not be estimated from molecular phylogenies. *Evolution*, **64**, 1816-1824.

Ramos-Onsins SE, Rozas J (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092-2100.

Randall JE, GR Allen & RC Steene (1990) *Fishes of the Great Barrier Reef and Coral Sea*. University of Hawaii Press, Honolulu, Hawaii. p. 506.

Randall JE & Bell LJ (1992) *Naso caesius* a new acanthurid fish from the central Pacific. *Pacific Science*, **46**, 344-352.

Randall JE (1998) Zoogeography of shore fishes of the Indo-Pacific region. *Zoological Studies*, **37**, 227-268.

Randall JE (2002) *Surgeonfishes of Hawaii and the World*. Mutual Publishing and Bishop Museum Press, Honolulu.

Rambaut A & Drummond AJ (2007) Tracer v1.4. URL http://beast.bio.ed.ac.uk/Tracer.

Ravago-Gotanco RG & Juinio-Meñez MA (2010) Phylogeography of the mottled spinefoot *Siganus fuscescens*: Pleistocene divergence and limited genetic connectivity across the Philippine archipelago. *Molecular Ecology*, **19**, 4520-4534.

Ray N, Currat M & Excoffier L (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, **20**, 76-86.

Read CI, Bellwood DR & van Herwerden L (2006) Ancient origins of Indo-Pacific coral reef fish biodiversity: A case study of the leopard wrasses (Labridae: *Macropharyngodon*). *Molecular Phylogenetics and Evolution*, **38**, 808-819.

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.

Renema W, Bellwood DR, Braga JC, Bromfield K, Hall R, Johnson KG, Lunt P, Meyer CP, McMonagle LB, Morley RJ, O'Dea A, Todd JA, Wesselingh FP, Wilson MEJ & Pandolfi JM (2008) Hopping hotspots: global shifts in marine biodiversity. *Science*, **321**, 654-657.

Roberts DG, Gray CA, West RJ & Ayre DJ (2009) Evolutionary impacts of hybridization and interspecific gene flow on an obligately estuarine fish. *Journal of Evolutionary Biology*, **22**, 27-35.

Rocha LA, Robertson DR, Roman J & Bowen BW (2005a). Ecological

speciation in tropical reef fishes. *Proceedings of the Royal Society of London Series B*, **272**, 573–579.

Rocha LA, Robertson DR, Rocha CR, van Tassell JL, Craig MT & Bowen BW (2005b) Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. *Molecular Ecology*, **14**, 3921-3928.

Rocha LA, Craig MT & Bowen BW (2007) Phylogeography and the conservation of coral reef fishes. *Coral Reefs*, 26, 501–512.

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

Rocha LA, Rocha CR, Robertson DR & Bowen BW (2008) Comparative phylogeography of Atlantic reef fishes indicates both origin and accumulation of diversity in the Caribbean. *BMC Evolutionary Biology*, **8**, 157.

Rogers AR (1995) Genetic evidence for Pleistocene population explosion. *Evolution*, **49**, 608–615.

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

Rousset F (2008) Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular Ecology Resources, **8**, 103-106.

Rozas J, Sanchez-DelBarrio JC, Messeguer X & Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496-2497.

Sambrook J, Fritsh EF & Maniatis T (1989) *Molecular Cloning, a Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

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

Shevenell AE (2008) Middle Miocene ice sheet dynamics, deep-sea temperatures, and carbon cycling: a Southern Ocean perspective. *Geochemistry Geophysics and Geosystems*, **9**, 1-14.

Slatkin M (1985) Rare alleles as indicators of gene flow. *Evolution*, **39**, 53-65.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787-792.

Slatkin M & Hudson RR (1991) Pair wise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555-562.

Springer VG (1982) Pacific plate biogeography, with special reference to shore fishes. *Smithsonian Contributions to Zoology*, **367**, 1-182.

Springer VG & Williams JT (1990) Widely distributed Pacific plate endemics and lowered sea level. *Bulletin of Marine Science*, **47**, 631-640.

Strasburg JL, Kearny M, Moritz C & Templeton AR (2007) Combining phylogeography with distribution modeling: Pleistocene range expansions in a parthenogenic gecko from the Australian arid zone. *PLoS ONE*, **2**, e760.

Swearer SE, Shima JS, Hellberg ME, Thorrold SR, Jones GP, Robertson DR, Morgan SG, Selkoe KA, Ruiz GM & Warner RR (2002) Evidence of self-recruitment in demersal marine populations. *Bulletin of Marine Science*, **70**, S251-271.

Swofford DL (1999) PAUP*: Phylogenetic Analysis Using Parsimony. Sinauer Associates, Champaign.

Thacker CE, Thompson AR, Roje DM & Shaw EY (2007) New expansions in old clades: population genetics and phylogeny in a *Gnatholepis* species (Teleostei: Gobioidei) in the Pacific. *Marine Biology*, **153**, 375-385.

Turelli M, Barton NH & Coyne JA (2001) Theory and speciation. *Trends in Ecology and Evolution*, **16**, 330-343.

Tyler JC, Johnson GD, Nakamura I & Collette BB (1989) Morphology of *Luvarus imperialis* (Luvaridae), with a phylogenetic analysis of the Acanthuroidei (Pisces). Smithsonian Contributions to Zoology, **485**, 1-78.

Tyler JC (2000) *Arambourgthurus*, a new genus of hypurostegic surgeonfish (Acanthuridae) from the Oligocene of Iran, with a phylogeny of the Nasinae. *Geodiversitas*, **22**, 525-537.

van Herwerden L, Choat JH, Newman SJ, Leray M & Hillersoy G (2009) Complex patterns of population structure and recruitment of *Plectropomus leopardis* (Pisces: Epinephelidae) in the Indo-West Pacific: implications for fisheries management. *Marine Biology*, **156**, 1595-1607.

van Oosterhout C, Hutchinson WF, Wills DPM & Shipley P (2004) Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535-538.

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.

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

Wakeley J (1999) Non-equilibrium migration in human history. *Genetics*, **153**, 1863-1871.

Wakeley J & Aliacar N (2001) Gene genealogies in a metapopulation. *Genetics*, **159**, 893-905.

Wallace EM & Tringali MD (2010) Identification of a novel member in the family Albulidae (bonefish). *Journal of Fish Biology*, **76**, 1972-1983.

Walsh PS, Metzger DA & Higuchi R (1991) Chelex® 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, **10**, 506.

Waples RS & Gaggiotti O (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, **15**, 1419-1439.

Ward RD, Zemlak TS, Ines BH, Last PR & Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society, London Series B*, **360**, 1847–1857.

Webster JM, Wallace L, Silver E, Potts D, Braga JC, Renema W, Riker-Coleman K & Gallup C (2004) Coralgal composition of drowned carbonate platforms in the Huon Gulf, Papua New Guinea: implications for lowstand reef development and drowning. *Marine Geology*, **204**, 59-89.

Webster JM, Clague DA, Faichney IDE, Fullagar PD, Hein JR, Moore JG & Paull CK (2010) Early Pleistocene origins of reef around Lanai. *Earth and Planetary science letters*, **290**, 331-339.

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

Weir BS (1996) *Genetic Data Analysis II: Methods for Discrete Genetic Data.* Sinauer Associates, Inc., Sunderland, MA, USA.

Wellenreuther M, Syms C, Clements KD (2008) Body size and ecological diversification in a sister species pair of triplefin fishes. *Evolutionary Ecology*, **22**, 575-592.

Wiley EO (2002) On species and speciation with reference to fishes. *Fish and Fisheries*, **3**, 161-170.

Wilgenbusch JC, Warren DL & Swofford DL (2004) AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. http://ceb.csit.fsu.edu/awty.

Williams ST, Duda TF (2008) Did tectonic activity stimulate Oligo-Miocene speciation in the Indo-West Pacific? *Evolution*, **62**, 1618-1634.

Wilson DT & McCormick MI (1999) Microstructure of settlement marks in the otoliths of tropical reef fishes. *Marine Biology*, **134**, 29-41.

Wilson GA & Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, **163**, 1177-1191.

Wilson DE & Reeder DAM (editors) (2005) *Mammal Species of the World. A Taxonomic and Geographic Reference* (3rd ed), Johns Hopkins University Press, 2,142 pp.

Winters KL, 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.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 ma to present. *Science*, **292**, 686-693.

Zigler KS, McCarteney MA, Levitan DR & Lessios HA (2005) Sea urchin bindin divergence predicts gamete compatibility. *Evolution*, **59**, 2399-2404.

Zink RM & Barrowclough GF (2007) Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, **17**, 2107-2121.

Zwickl DJ (2006) GARLI. Program distributed by the author. Section of Integrative Biology, The Univ. of Texas at Austin. available at www.zo.utexas.edu/faculty/antisense/Garli.html.