UNIVERSITY OF CALIFORNIA
SANTA CRUZ

# INTERACTIONS AND EVOLUTION OF THE ANEMONEFISHES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

**James L. O'Donnell**

June 2014

The Dissertation of James L. O'Donnell
is approved:

Professor Giacomo Bernardi, Chair

Professor Peter T. Raimondi

Professor John N. Thompson

Dean Tyrus Miller
Vice Provost and Dean of Graduate Studies

# Table of Contents

# List of Figures

# List of Tables

**Abstract**

Interactions and evolution of the anemonefishes

by

James L. O'Donnell

The interactions among organisms shape the earth's biodiversity in striking ways, and these interactions are known to vary substantially over both time and space. Identifying the scale of variation in species interactions allows for a better understanding of the origins and organization of biodiversity, and informs expectations of the dynamics of these systems. In order to study the scale of variation in species interactions, I focused on the anemonefishes, a group of coral reef damselfishes engaged in a mutualism with sea anemones. I used three complementary approaches. First, I used population genetic methods to show that populations of anemonefishes are genetically isolated over relatively small spatial scales in the Mozambique Channel. Second, I employed data from social media to reveal spatial variation in the interactions between anemonefishes and their host sea anemones. Third, I reconstructed the phylogenetic history of the anemonefishes to examine their pattern of diversification in the context of both ecological and geographic processes. The relatively low dispersal potential of anemonefishes likely contributes to both the spatial structure among populations within species, and the strong geographic pattern of diversification among species. In turn, the spatial variation in diversity within the anemonefishes has led to variation in their interactions with sea anemones.

To my wife, Sally,

whose love, support, and laughter kept me going through graduate school. I could not have done this without you.

And to my mother, Jan,

who did not merely tolerate my fascination with scaly and slimy creatures, but embraced it. I cannot thank you enough for your tireless love and encouragement.

## Acknowledgments

While at times writing a dissertation can feel isolating, I think of this as a collaborative process for which I have many people to thank.

First, I am grateful to the organizations that provided the funding for this research, including the UCSC Ecology and Evolutionary Biology Department, the Society for the Study of Evolution, Sigma Xi, the American Museum of Natural History Lerner-Gray Grant for Marine Research, the American Society of Ichthyologists and Herpetologists Raney Award, the UCSC Graduate Student Association, the UCSC Doctoral Sabbatical Fellowship, the Earl and Ethel Myers Oceanographic and Marine Biology Trust, the Friends of Long Marine Lab, and the INEE-CNRS. In particular, I wish to acknowledge the support of the UCSC Chancellor's Dissertation Year Fellowship, which permitted a much needed reprieve from 17 quarters as a teaching assistant.

This work would not have been possible without my committee members, who contributed their time not just willingly but enthusiastically. While they generously gave input on my work, any oversights contained here are my own.

Pete Raimondi is responsible for instilling in me a logic-based approach to hypothesis testing. His ardor for ecology and commitment to mentorship is embodied by his open office door. Pete was quick to set aside what he was doing to make time for questions, and provided answers in a way that made me feel more like a colleague than a subordinate. I am lucky to have spent time teaching with Pete in Corsica and Moorea, where he was always the first out helping students during the day, and the last

one helping students analyze data at night. I will always admire his attitude and work ethic.

John Thompson always helped put my work in a bigger perspective, which has played a significant role in shaping the way I think about my own research program. John shared and indulged my broad interests in natural history, and I always enjoyed they tendency of our chats to include talk of some interesting species or behavior we'd recently encountered. I am glad to have been able to participate in the Species Interactions Workshop, and will look back on his guidance with great admiration.

There is too much to say about Giacomo Bernardi to include in a single dissertation. Giacomo had seemingly never-ending patience with me as I learned the ropes of a molecular wet lab and in the analysis of genetic data. He was never bothered by impromptu questions in the office, and always made time to talk through questions and analyses over a chocolate croissant. I have very fond memories of our time in the Philippines, Madagascar, Corsica, and Moorea, and I look forward to the next round of "name that fish".

I'd also like to thank Krikor Andonian, Kathleen Kay, Greg Gilbert, and Ingrid Parker, who were willing to take a chance on letting a fish person branch out into the worlds of plants and insects. Tim Miller was particularly patient with me as I bumbled through botany. I also owe a great deal of gratitude to Gustav Paulay and the staff at FLMNH who were instrumental in setting me on this path, and to Jada-Simone White for her mentorship and bringing me on my first field season.

I am thankful for the opportunity to be immersed in a large community of

people interested in better understanding the natural world, and there is no way I can name all of the people with whom I shared invigorating discussions and experiences. Several friends provided not only moral support, but reminded me of how cool this stuff really is, especially Adrian Stier, Justin Yeakel, David Hembry, Sea McKeon, Abe Borker, Tim Miller, Eric Crandall, and Sharifa Crandall. My cohort of Justin Cummings, Justin Yeakel, Melinda Fowler, Rodrigo Beas, Erin McCreless, Cyril Michel, and Alicia Abadia helped make this a fun place right from the beginning.

Because graduate school is not always easy, it made a big difference to be able to come into the lab every day knowing I'd be laughing with my lab mates at some point. The Bernardi lab had great barbecues, and even if it got a little hot and crowded, I'll miss our tiny office.

The hardest thing about finishing school will be leaving behind a great group of friends in Santa Cruz, and I'm fortunate to say the list is far too long to be included here. The gatherings of Los Clupars will be memories I shall cherish forever. Thanks so much to all of you who made this a great journey.

My family has been wonderfully supportive of me throughout this process, making sure to keep in touch in spite the physical distance. Beyond moral support, my Aunt Maureen spared me the cost of expensive eyeglasses with the gift of a beautiful monitor on which to analyze sequences and write computer code. My brothers Johnny, Danny, and Timmy all provided great conversation and feedback in both general and specific ways, and I consider myself incredibly lucky to have a built in group of friends for life. Tim also provided assistance with the assembly of some of the figures.

I would not have even gotten on the road to grad school if it wasn't for the tireless love and support of my mom. When I was eleven years old, making lists of chameleon species while listening to Tom Waits, she never tried to convince me to collect baseball cards. Instead, she embraced my interests, and selflessly supported my pursuits. The things that my mom has taught me have been the most important lessons in my life, and if I am ever the curator of a museum, I hope you'll visit often.

Finally, I am most indebted to my wife Sally for her constant love, support, friendship, and patience. Without her there to consistently remind me of the bigger picture, I would have surely let the "number checking" drive me crazy. I cannot tell you enough how much your support has meant to me.

# Introduction

All living things are connected by invisible links that form an enormous network connecting every organism that has ever existed. These arbitrary human constructs allow us to make sense of the natural world around us. We organize our own social groups around families, connected by reproduction that has resulted in shared genetic material among individuals. The same concept of relatedness is used to classify the breathtaking diversity of life into hierarchical groups evolved over billions of years. Similarly, we refer to organisms living in the same place at the same time as a community or ecosystem, some of which interact directly with one another. Interactions among organisms affect the reproductive output of the participants. For example, a sperm whale that kills and eats a giant squid acquires energy that increases its ability to find mates and produce offspring, while reducing the squid's prospect of reproduction to nil.

The interactions among species have so deeply shaped the diversity of life that it would be unrecognizable without them. In the absence of the selective pressure imposed by species interactions, cheetahs would not be fast, flowers would not have pleasant fragrances, and grass would not be green. For about as long as humans have

pondered the natural world, they have recognized the importance of antagonistic interactions – those in which at least one of the participants experiences a net cost. But in the past several decades, mutualistic interactions that benefit both organisms have garnered increasing attention from ecologists (Howe, 1984; Bronstein, 2001). Research has enhanced our knowledge of the conditions under which mutualism might evolve, the evolutionary outcomes of mutualisms, and the organization of communities of mutualistic species, using both theoretical and empirical approaches (Thompson, 2005). Theoretical research on mutualisms has been exceptionally productive, but the answers that theory has provided to these questions can only be confirmed by empirical evidence, ideally from a diverse range of systems. One area of study on mutualisms that has gained substantial attention is whether mutualistic interactions tend to be more or less specialized than antagonistic interactions (Thompson, 2005).

Ecological specialization – the breadth of resources on which an organism depends – has captivated biologists for many years (Darwin, 1862). Though conceptually intuitive, it is difficult to quantify because it is inherently relative. Clearly, a sea otter that consumes only one type of prey over its lifetime is more specialized than a neighboring otter consuming five prey types. But are these otters more or less specialized than those consuming the same number of prey types in an area with different availability of prey? Such information is not always available, and thus no consistent metric is appropriate for all purposes.

Despite the difficulties in quantifying specialization, theory provides expectations of how specialization should affect the ecology and evolution of species, and these

expectations have been confirmed by empirical evidence. For example, specialization is commonly invoked as the mechanism driving adaptive radiation, the evolution of ecological and phenotypic diversity within a rapidly multiplying lineage (Schluter, 2000). Yet, it remains to be known which types of specialization drive the population divergence leading to speciation. Quantifying the links between specialization and the divergence of lineages is critical to our understanding of the role of species interactions on speciation mechanisms.

While evolutionary biologists have long speculated about the dynamics and outcomes of specialization for speciation and community organization, it is only in the last decade that the tools necessary to assess complex ecological networks within an evolutionary context have become available. Scientists have begun to address this primarily using three approaches: (1) Analyzing interaction network structure in the context of a known phylogenetic framework to test for phylogenetic constraints (Cattin et al., 2004; Rezende et al., 2007); (2) inferring the rate of transitions to and from specialization using character mapping on a phylogenetic tree (Kelley and Farrell, 1998; Nosil and Mooers, 2005; Stireman, 2005; Yotoko et al., 2005); and (3) comparing the amount of genetic structure or distance between sympatric species of closely related specialists and generalists (Dobler and Farrell, 1999; Kelley et al., 2000; Brouat et al., 2003). Each of these approaches has its own strengths and weaknesses, and thus the combination of these three approaches within a single system would yield more robust results.

Here, I focus on the mutualism between anemonefishes (Perciformes: Pomacen-

tridae: Amphiprioninae) and their host sea anemones. This mutualism provides the fish with shelter from predators, while the fish guards the anemone from predators and increases anemone growth and reproduction by supplementing nutrients to the anemone's photosynthetic zooxanthellae via excrement (Porat and Chadwick-Furman, 2004; Holbrook and Schmitt, 2005; Roopin and Chadwick, 2009). The interaction is obligatory for the fish, as they are never found outside of anemones, though in some areas host anemones may be found without fish (Fautin and Allen, 1997). The 30 described species of anemonefish can be found on Indo-Pacific coral reefs from East Africa to Polynesia (Fautin and Allen, 1997). The species are grouped into two genera, *Amphiprion* and *Premnas*, though *Premnas* has repeatedly been found to be nested within *Amphiprion* (Santini and Polacco, 2006; Cooper et al., 2009; Frédérich et al., 2013). Ten species of anemone hosts are recognized (Fautin and Allen, 1997), and individual fish associate with a single host for the entirety of their post-settlement life. Some fish species are only known to associate with a single host species, while others can be found with any of the ten hosts. All species are protandrous hermaphrodites and are usually found in small monospecific groups. Mated pairs guard nests of demersal eggs until hatching, at which point larvae begin a pelagic phase lasting 7-22 days (Wellington and Victor, 1989; Thresher et al., 1989). Here, I study the interactions and evolution of the anemonefishes using three approaches.

First, I investigate the degree of population genetic structure of an anemonefish species in the Mozambique Channel. The extent of connectivity among populations is of fundamental importance to ecology and evolutionary biology, and plays an important

role in the planning and management of natural reserves; however, it is poorly understood in most marine organisms. Studies of demersal coral reef fishes have been key to enhancing our understanding of marine population connectivity, but few have examined gene flow in the Western Indian Ocean, a marine biodiversity hotspot. We studied the population genetic structure of *Amphiprion akallopisos* among four island sites in the Mozambique Channel using highly variable microsatellite loci. We employed two Bayesian clustering algorithms along with classical population genetic approaches, and found evidence for subtle population structure among island sites. While bathymetric and oceanographic features may have predicted genetic homogeneity in this region, we argue that the complex system of seasonal eddies may serve to increase the likelihood of larvae recruiting close to their natal habitat, thereby increasing population structure.

Second, I characterized the interaction frequency and spatial variation between anemonefish and their host anemone species, using a novel data source. Understanding the dynamics of species interactions is a central goal of ecology, but comprehensive assessments of interactions across space and taxonomic groups are difficult to obtain. I collected and analyzed 10,167 georeferenced photos from the entire geographic range of the interaction to compile a data set of 11,029 occurrence records of all species of anemonefish and sea anemones, including 4,830 records explicitly depicting an interaction. I used this data set along with a detailed habitat map to model species distribution and diversity, and reconstruct both local and global quantitative interaction matrices. The nestedness of these interaction matrices is not spatially variable at the regional scale, and the structure of regional networks does not differ from that of the global

network. Within suites of species, there is no correlation between the spatial overlap of species and their pattern of partner interaction. These results suggest that similar processes structure the interaction network at multiple scales, and that species with overlapping distributions partition host resources in order to coexist.

Third, I reconstructed the evolutionary history of the diversification of the anemonefishes using a phylogenetic approach. While much is known about their behavior and ecology, their evolutionary history remains poorly resolved. We generated DNA sequence data from 24 of the 29 species for three mitochondrial loci and four nuclear loci, and include samples from multiple geographic variants of wide-ranging species. We reconstructed phylogenies at both the gene and species level using Bayesian and maximum likelihood methods to infer patterns of the evolution of the group. Gene phylogenies revealed reciprocal monophyly among species in most cases, though substantial divergence between populations indicates the presence of cryptic diversity in at least one wide-ranging taxon, *A. clarkii*. We recovered exceptionally strong support for most nodes in the species phylogeny, which upheld the taxonomic grouping of species based on morphology. Strongly supported yet shallowly divergent sister relationships between morphologically disparate taxa support the putative reports of interspecific hybridization by other authors. The earliest branching taxa are restricted to the center of the spatial distribution of the Amphiprioninae, the Coral Triangle, while more recently diverging species are found in the periphery of this region and in the Western Indian Ocean. This pattern is indicative of an origin within the Coral Triangle, followed by diversification driven by spatial isolation outside of the core of diversity.

# Chapter 1

# Population structure of *Amphiprion akallopisos* in the Mozambique Channel: Connectivity, management, and conservation in the Scattered Islands

## Introduction

The Western Indian Ocean (WIO) is one of the most biodiverse yet understudied regions of the sea (Roberts et al., 2002). In addition to high levels of local diversity and endemism, there is substantial turnover in species composition between this region and the epicenter of marine biodiversity, the Coral Triangle (Briggs and Bowen, 2012; Allen, 2008; Obura, 2012). These attributes make the WIO an important priority for conservation, yet its coral reefs are afforded little protection from exploitation (Mora

[1] Additional coauthors on this manuscript: Ricardo Beldade, Hannah Williams, Suzanne C. Mills, and Giacomo Bernardi

et al., 2006). Human populations in this region are growing rapidly, and these populations rely on marine organisms as a food source and as an economic resource. The combined threat of global climate change (McClanahan et al., 2014) and anthropogenic disturbances has generated growing interest in creating reserves that will help sustain populations of marine organisms both for their economic and intrinsic values.

Optimization of the size, spatial arrangement, and regulation of ecological reserves so that they effectively serve their intended purpose requires an understanding of the movement of individuals among populations in the area of interest (Halpern and Warner, 2003; Palumbi, 2003). The optimal size, arrangement, and regulation of reserves depends on whether populations in proposed areas can be sustained under harvested conditions, which in turn depends on the source of new individuals recruited to local populations (Halpern and Warner, 2003). If population self-recruitment is high, then local harvesting pressure should have strong effects on population size. Conversely, if immigration is high, local pressures should have little effect on local population size. Along with factors like natural mortality and emigration rates, the relationship between local harvesting pressure and population size is expected to be mediated by the influx of individuals from outside areas (Carr and Reed, 1993; Shanks et al., 2003).

Population connectivity (the movement of individuals among spatially isolated populations) of marine organisms remains one of the most elusive problems in ecology. The majority of marine organisms begin life as tiny larvae that are suspended in the water column for days to months (Leis, 1991), during which time they are currently impossible to track using traditional methods. Both intrinsic and extrinsic factors have

been implicated as major contributors to realized larval dispersal, though their relative contributions are unknown. It is intuitive to suspect a relationship between the duration of the pelagic larval phase and population connectivity: If the pelagic larval duration (PLD) is long, larvae could be carried over long distances by even moderate currents. Conversely, the larvae of species with a very short PLD are expected to settle near their natal habitat. Due to the difficulty in tracking actual larvae in the water column, a number of studies have tried to use genetic tags and relate PLD with gene flow (Doherty et al., 1995; Riginos and Victor, 2001; Shulman and Bermingham, 1995; Waples, 1987). Yet, no clear picture has emerged from these studies, suggesting that simple oceanographic explanations may not be forthcoming.

Simplistic dispersal models take place in the context of extremely complex and dynamic oceanographic features. Currents differ in direction and speed at different depths, and drastic changes take place over timescales from seasonal to decadal. While results to date implicate currents as playing a role in larval dispersal, evidence for strong and consistent patterns has been obscured by problems incorporating the complexities of oceanographic forcing at relevant spatial scales. Furthermore, the influence of these factors may be mediated by larval behavior. Simulations have shown that even modest abilities for larvae to modulate their position in the water column can obscure the relationship between PLD, oceanographic features, and larval dispersal (Cowen et al., 2006; Leis, 2007).

At large scales discussed above, patterns of population connectivity are not always consistent across species, in part because barriers to gene flow are relative to

dispersal ability. Thus, attention has focused on smaller scales that are more relevant to ecological processes. Early work on Labrids and Pomacentrids showed that self-recruitment was more common than previously expected (Jones et al., 1999; Swearer et al., 1999). This was followed by parentage analysis on anemonefishes using microsatellites, which showed the precise point of origin and settlement for individual fishes(Jones et al., 2005; Planes et al., 2009). These methods were also used to assess the effectiveness of marine reserves and how larvae travel in and out of such areas (Almany et al., 2007; Planes et al., 2009; Harrison et al., 2012).

We brought together the necessity for understanding the level of connectivity and the use of genetic techniques on anemonefishes to address connectivity issues in the Mozambique Channel. Very few studies have examined population connectivity in the Mozambique Channel, despite much interest in the establishment of reserves for both human use and conservation in the region. The channel is approximately 500 km wide and 1200km long, with a string of islands running along its middle, The Scattered Islands. This provides a landscape of habitats which are close enough for potential connectivity, but distant enough to allow populations to diverge genetically. These islands, Europa, Bassas da India, Juan de Nova, and Iles Glorieuses, are relatively pristine (Fricke et al., 2013) and a primary candidate for marine protected areas.

To assess the connectivity among isolated areas in the Mozambique Channel, we studied the spatial genetic structure and differentiation of the skunk anemonefish, *Amphiprion akallopisos*. Congeneric species are known to have a brief PLD of 7-22 days (Thresher et al., 1989; Wellington and Victor, 1989), simplifying the relationship

between intrinsic factors, extrinsic factors and population connectivity (Shanks, 2009). They are symbiotic mutualists of sessile sea anemones, and adults rarely move further than a few meters from their host (Fautin and Allen, 1997). While this species is not a target of harvesting in this region, its life history makes it likely to show fine-scale levels of population structure should it exist. Significant levels of population structure have been shown in other members of this genus, even at very fine spatial scales (Beldade et al., 2012; Jones et al., 2005; Timm and Kochzius, 2008; Pinsky et al., 2010).

We set out to test whether significant genetic differentiation exists among populations of *Amphiprion akallopisos* at four sites in the Mozambique Channel.

# Methods

## Sample Collection

We collected samples at four island sites in the Mozambique Channel: Glorieuses (GLO), Juan de Nova (JDN), Bassas da India (BAS), and Europa (EUR) (Figure 1.1). These islands, along with the island of Tromelin, make up the Îles Éparses, an overseas administrative division of France. Individuals were captured using hand nets, and a small piece of fin tissue was collected before returning fish to their host anemone. Tissue samples were placed in 95% ethanol as soon as possible, and stored at -20C.

## Laboratory Methods

Genomic DNA was extracted using a chloroform protocol (Sambrook et al., 1989). Microsatellite loci were originally isolated and developed for *A. chrysopterus* and *A. polymnus* (Beldade et al., 2009; Quenouille et al., 2004). From this existing library we chose 20 microsatellites to test for variability on 32 individuals. Nine polymorphic loci were selected based on their variability and amplification success rate (Table 1.1). The polymerase chain reaction (PCR) was carried out in an Applied Biosystems GeneAmp PCR system 9700 using fluorescently labeled forward primers (Table 1.1). Each reaction contained 5$\mu$L of Qiagen Multiplex PCR mastermix, 0.2 pmol of each primer, and 0.8 $\mu$L of DNA, diluted with RNAase-free water to a total reaction volume of 10 $\mu$L. The following temperature profile was used: 15 min at 95 C , followed by 40 cycles of 30 seconds at 94C, 1 min and 30 s at 57 C, and 1 min at 72 C with a final extension of 7 min at 72C. PCR product was diluted with 50$\mu$L of water, and 0.5$\mu$L of this diluted PCR product was added to 9.74$\mu$L of HiDi formamide and 0.26$\mu$L ROX 500 size standard (Qiagen, Valencia, California). Fragment size was analyzed on an ABI 96 capillary 3730XL DNA Analyzer (Applied Biosystems, Darmstadt, Germany).

## Data Analysis

We genotyped individuals using GeneMapper version 3.7 (Applied Biosystems), and convert v1.31 (Glaubitz, 2004) was used to generate input files for various analysis programs. We checked for null alleles and scoring errors using micro-checker v.2.2.3 (Van Oosterhout et al., 2004). Expected and observed heterozygosities, pairwise

$F_{ST}$ values, an exact test of deviation from Hardy-Weinberg Equilibrium (HWE) using a Markov Chain of length 100000, as well as an analysis of molecular variance (AMOVA), were calculated in the software Arlequin v3.1 (Excoffier et al., 2005). One locus was omitted from the genetic structure analysis due to too much missing data (0.109). Significance of $F_{ST}$ values was assessed using 100 permutations.

We conducted a Bayesian clustering analysis in STRUCTURE v2.3 (Pritchard et al., 2000). STRUCTURE implements a Markov chain Monte Carlo (MCMC) simulation approach to estimate the posterior probability that a sampled individual belongs to each of K clusters based on its multi-locus genotype. We conducted analyses for values of K from 1 to 10, each consisting of 20 independent runs of MCMC length 1,000,000. The first 10,000 iterations of each chain were discarded as burn-in. We chose an optimum value of K using the method described by Evanno et al. (2005), and combined runs using the software CLUMPP (Jakobsson and Rosenberg, 2007), and visualized the results in DISTRUCT (Rosenberg, 2004).

As an independent line of evidence, we carried out a similar Bayesian clustering approach that explicitly accounts for the spatial arrangement of sampled individuals. This algorithm is implemented in Geneland (Guillot et al., 2008), a package developed for the R computing framework (R Core Team, 2013). Geneland uses MCMC to simultaneously estimate the most likely number of clusters to which genotypes belong, the posterior probability of individuals' membership to those clusters, and the most likely distribution of those clusters in space. After a preliminary run of 1 million MCMC generations using the full dataset, we calculated the frequency of null alleles at each locus,

and found the two loci with the highest proportion of missing data to show evidence of high proportions of null alleles (0.22 and 0.38), while the remainder had very low frequencies of null alleles (0.05 or less). We ran a second analysis using only individuals with no missing data under identical conditions.

# Results

## Microsatellite Analysis

We successfully amplified nine microsatellite loci from 110 individuals from four sites in the Mozambique Channel (Table 1.1, Figure 1.1). While microsatellites were designed for congeneric species, these nine loci showed sufficient variability for a population study (2-35 alleles, mean = 16.1). Within sites, there was evidence of significant deviation from HWE in seven cases (Table 1.2), but no site or microsatellite showed consistent deviation from HWE, indicating that all sites and microsatellite loci were usable for this analysis.

## Population Structure

Classical $F_{ST}$ values as well as Bayesian approaches were used in this study. In marine populations, where effective population sizes are very large, $F_{ST}$ values are expected to be very low (Bird et al., 2011). As expected, $F_{ST}$ values were very low (less than 0.06), but were statistically significant in all pairwise comparisons except between Bassas da India and Glorieuses.

Both of the Bayesian clustering approaches were consistent with the $F_{ST}$ analysis presented above. Indeed, both indicated genetic structure among sites (Figure 1.2,1.3). Notably, the same numbers of clusters (three) were identified by both the Evanno et al. (2005) method and by Geneland using the dataset trimmed of missing data (Figure 1.4,1.5). However, the two methods differed in the assignment of individuals and sites among clusters. STRUCTURE assigned most individuals from Glorieuses and Bassas da India to two clusters with roughly equal probability, while the majority of the individuals from Juan de Nova and Europa were assigned to a third cluster with very high probability. Geneland's spatially explicit algorithm also identified three clusters, but specifically assigned those clusters to the three regions sampled here (north, central and south Mozambique Channel), with the geographically close Bassas da India and Europa grouped together (Figure 1.3).

## Discussion

For about a decade, anemonefish have been used to assess population structure and ecological patterns of larval dispersal in several systems (Berumen et al., 2012; Saenz-Agudelo et al., 2009). Here, we capitalize on a unique geographic system, the Mozambique Channel, and a proven biological system, anemonefish, to estimate the level of connectivity between an island chain in the Mozambique Channel. Using variable microsatellite markers, our data show low but significant levels of genetic structure based on $F_{ST}$ values, as well as distinct patterns of structure based on two indepen-

dent Bayesian clustering algorithms. The spatially explicit approach implemented by Geneland separated geographic regions into genetic clusters. Interestingly, both pairwise $F_{ST}$ values and the algorithm employed by STRUCTURE indicate a close relationship between Glorieuses, the northernmost site, and Bassas da India in the far south. This pattern is contrary to expectations based on an isolation by distance model.

One explanation of this pattern is the complex oceanographic features that exist in the region (see Figure 1 in Schouten et al., 2003). Water at the surface of the central Indian Ocean is carried westward by the South Equatorial Current before being interrupted by Madagascar. Some of that flow is diverted northwards, forming the East African Coastal Current. The remainder flows southwards along the coast of Madagascar before arriving at the coast of continental Africa, where it forms the Agulhas Current. The Mozambique Channel lies in the shadow of this powerful current, and while previously thought to be dominated by the southward flowing Mozambique Current (Sætre, 1985), it has now been shown that strong eddies form approximately four times per year (Schouten et al., 2003). Though the sites included in our study are small and isolated, these eddies may serve to enhance self recruitment and thereby increase genetic structure among sites.

Biogeographic patterns corroborate the important role of oceanographic features in the Southwestern Indian Ocean and Mozambique Channel. For example, within *Amphiprion* alone, another four species can be found in this region. *Amphiprion allardi* is common on the reefs of Africa on the western side of the Mozambique Channel, but is not found on Madagascar. Madagascar is home to its own endemic anemonefish, *A.*

*latifasciatus*, while two other endemics with extremely restricted ranges can be found in the Seychelles (*A. fuscocaudatus*) and Reunion (*A. chrysogaster*). While these species' small ranges may be mediated by other factors, such as competition for hosts, the complex circulation patterns are likely to play a role.

We present robust evidence for genetic population structure in a demersal reef fish in the Mozambique Channel, an understudied hotspot of marine biodiversity (Roberts et al., 2002). While bathymetric and oceanographic features may have predicted genetic homogeneity in this region, we argue that the complex system of seasonal eddies may serve to increase the likelihood of larvae recruiting close to their natal habitat, thereby increasing genetic structure. Future work should focus on the inclusion of samples from sites on both the eastern and western sides of the Mozambique Channel, as well as the comparison of these results to other species.

## Acknowledgements

We wish to thank the TAAF and the crew of the Marion Dufresne for logistical support, and the Friends of Long Marine Lab, Sigma Xi, The Lerner-Gray Fund for Marine Research, The American Society of Ichthyologists and Herpetologists, the UCSC Ecology and Evolutionary Biology Department, and INEE-CNRS for financial support.

## Tables and Figures

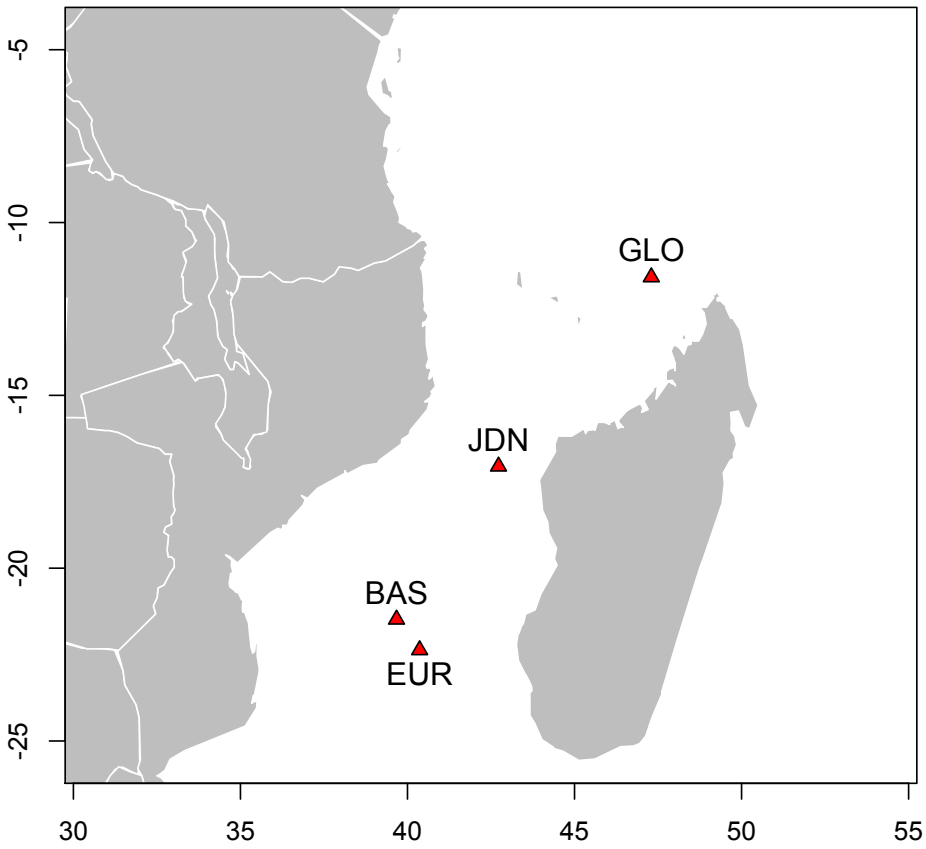



Figure 1.1: Location of collection sites in the Mozambique Channel (GLO: Glorieuses, JDN: Juan da Nova, BAS: Bassas da India, EUR: Europa).

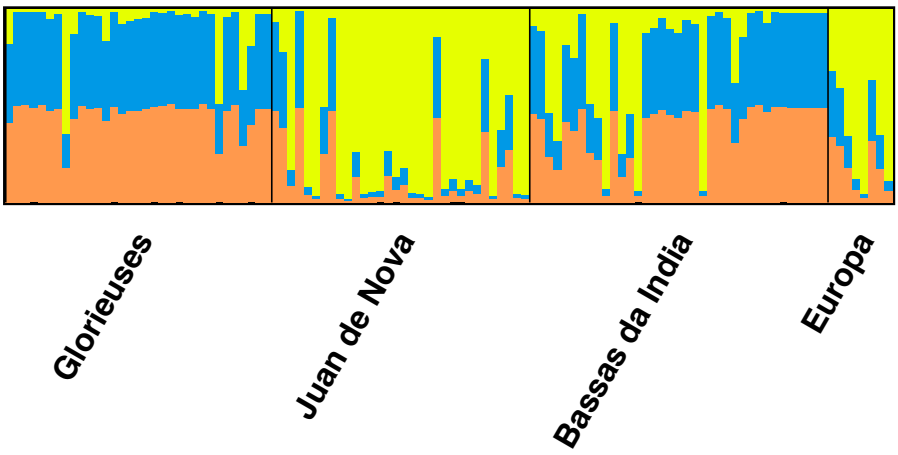



Figure 1.2: Plot representing output of Bayesian clustering algorithm implemented by STRUCTURE (Pritchard et al., 2000). Each vertical bar represents an individual, and the colors represent the posterior probability of its membership to each of three clusters. The number of clusters (three) was chosen using the method outlined by Evanno et al. (2005).

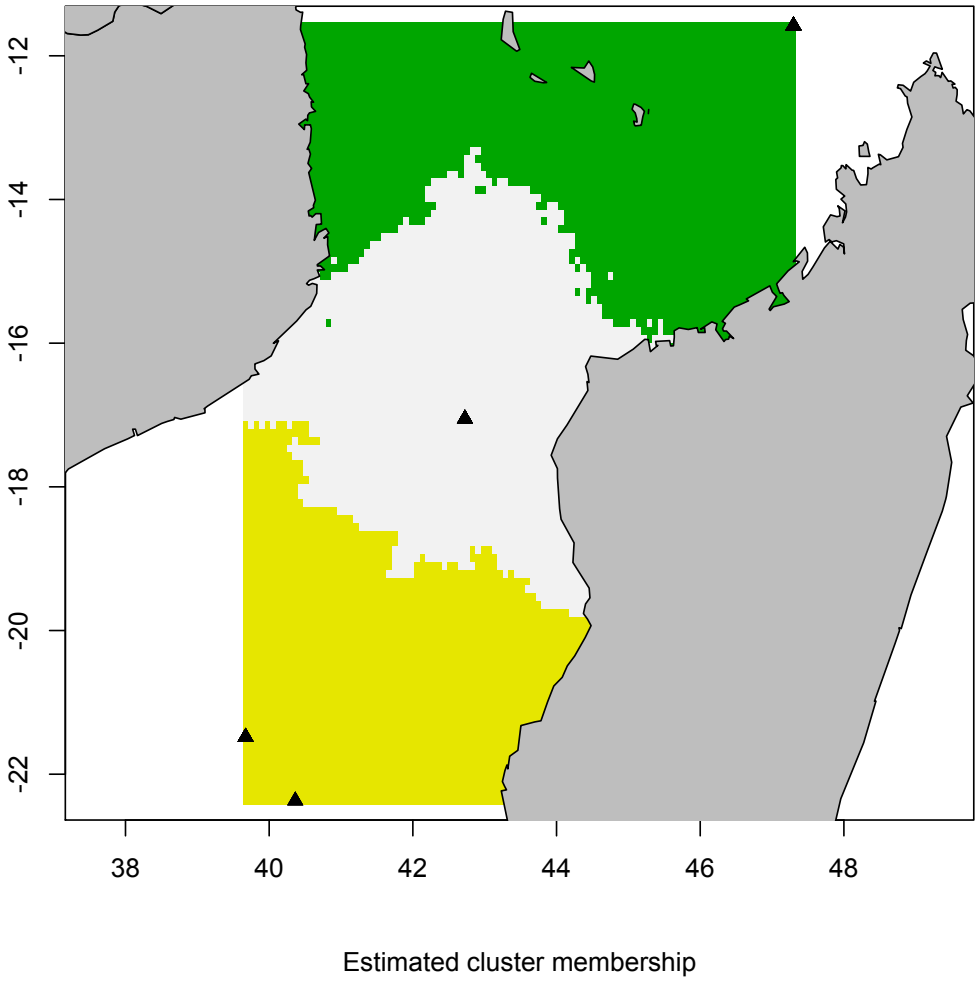



Figure 1.3: Estimated cluster membership for each pixel of a 100 by 100 grid based on posterior probability. Pixels are colored by the modal cluster membership of 1 million iterations. This represents the dataset after removal of individuals for which no data was available for at least one locus. A preliminary analysis using individuals with missing data (not shown) yielded qualitatively similar results, but did not appropriately sample the cluster parameter or establish convergence after multiple runs.

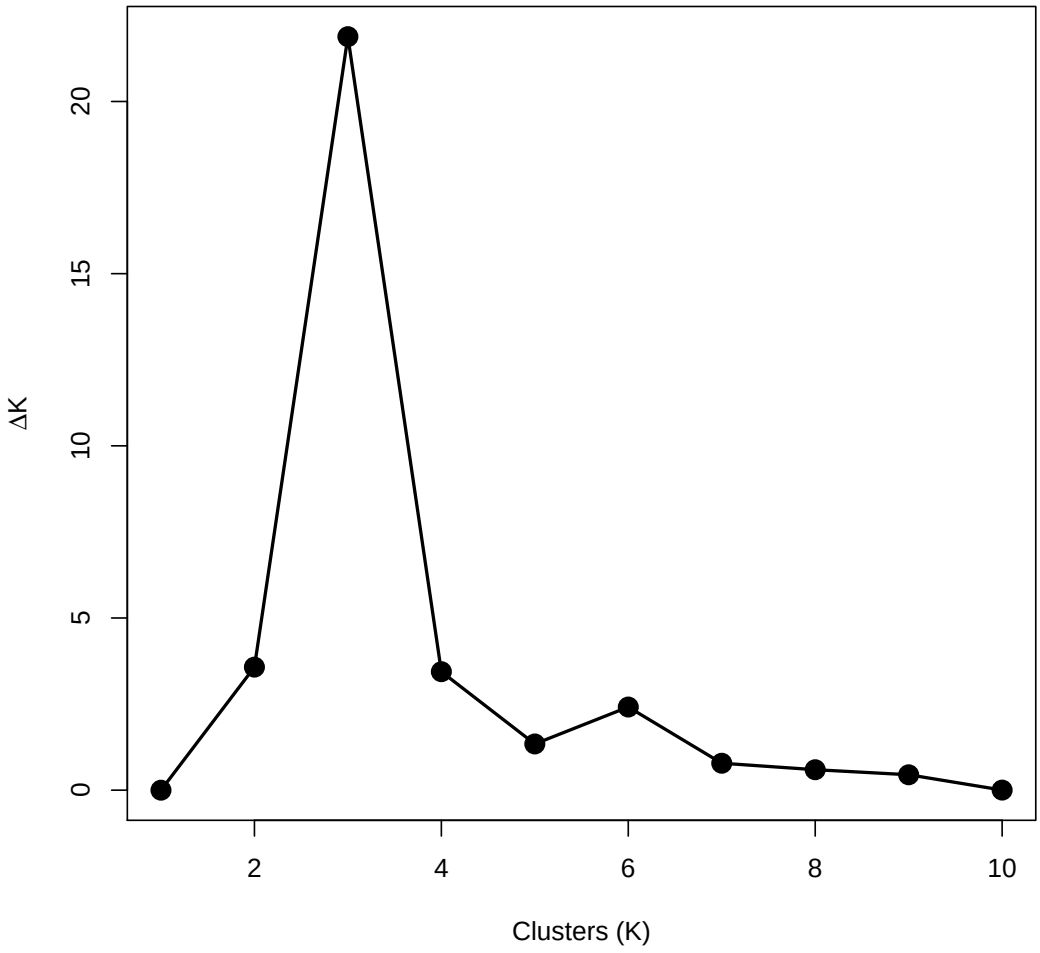



Figure 1.4: Plot of ΔK for each number of clusters (K) assessed over 20 runs in STRUCTURE.

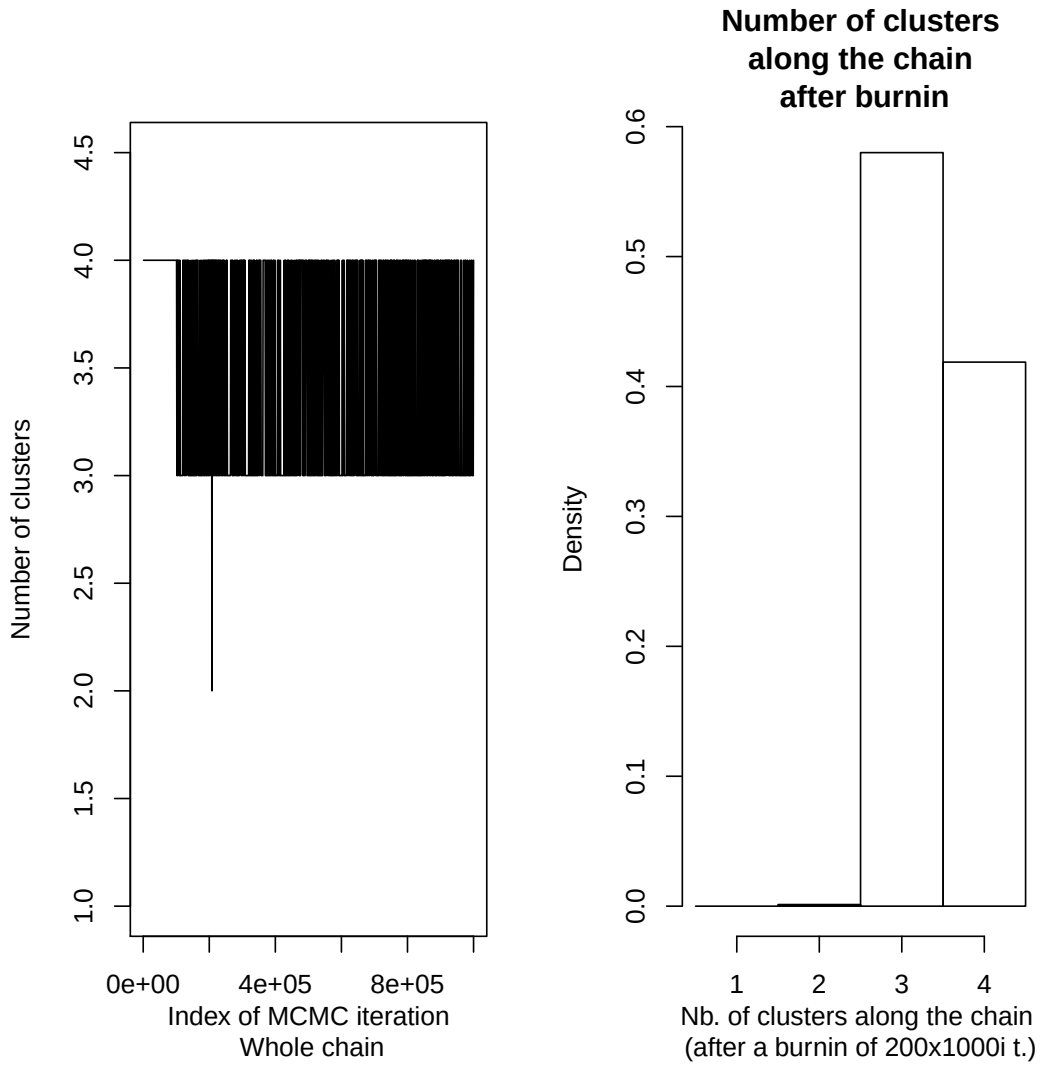



Figure 1.5: Sampling of cluster number parameter and histogram of posterior probability density of clusters sampled by the MCMC implemented by Geneland.

Table 1.1: Primer sequences, repeat motifs, fluorescent label name, and original reference for the nine microsatellite loci used in this study.

| Locus | Primer Sequence (5' to 3') | Repeat | Label | Reference |
|---|---|---|---|---|
| 10TCTA | F: GGGACGTATCTGTTGGAAATGAT<br>R: TTAAGGTACTGTGAGATGAGACT | (TCTA)26 | HEX | Quenouille et al., 2004 |
| 44 | F: TTGGAGCAGCGTACTTAGCT<br>R: AGATGTGTTTACGCACGCTT | (GT)13 | TAMRA | Quenouille et al., 2004 |
| 61 | F: TGAACACATAAACGCTCACTCAC<br>R: AAGACAATGCCTCCACATATCTA | (GT)49 | FLUO | Quenouille et al., 2004 |
| 120 | F: TCGATGACATAACACGACGCAGT<br>R: GACGGCCTCGATCTGCAAGCTGA | (GT)18N20(GT)14 | FLUO | Quenouille et al., 2004 |
| A130 | F: GCACTCAACACAAAGACCTTA<br>R: ACCCAAACAACATCCAGTC | (CA)24 | FLUO | Beldade et al., 2009 |
| B6 | F: TGTCTTCTCCCAAGTCAG<br>R: ACGAGGCTCAACATACCTG | (CATC)14 | FLUO | Beldade et al., 2009 |
| CF11 | F: GCTGGTTACAACACCTTG<br>R: GTAATTGCTGCAAGACAG | (CT)15(CA)16 | FLUO | Beldade et al., 2009 |
| D103 | F: GTTGGCTAATGGTGCTGTG<br>R: GATTCTGTGGTGGCATCAG | (GATA)13 | FLUO | Beldade et al., 2009 |
| D114 | F: TGTTCCAGCTCTGATATTTGAC<br>R: TTGGCAGTGTTTTATACCTGTC | (GATA)19 | TAMRA | Beldade et al., 2009 |

Table 1.2: Number of alleles (Na) and fragment lengths (RR, given in base pairs) for nine microsatellite loci used in this study, as well as observed and expected heterozygosities for each locus at each sampled site. Asterisks indicate significant deviations from Hardy-Weinberg Equilibrium ($p < 0.05$) based on 100,000 generations of Markov chain simulation. Sample sizes are given next to each site name: Glorieuses (GLO), Juan de Nova (JDN), Bassas da India (BAS), and Europa (EUR).

| | | | GLO (n = 32) | | JDN (n = 37) | | BAS (n = 33) | | EUR (n = 8) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Locus | Na | RR (bp) | $H_o$ | $H_e$ | $H_o$ | $H_e$ | $H_o$ | $H_e$ | $H_o$ | $H_e$ |
| 10TCTA | 35 | 500-558 | 1 | 0.937 | 0.91429 | 0.8911 | 0.87097 | 0.91909 | 0.75 | 0.75 |
| 44 | 11 | 241-256 | 0.54839 | 0.728 | 0.5 | 0.67332* | 0.6087 | 0.63961* | 0.375 | 0.525 |
| 61 | 34 | 296-359 | 0.84375 | 0.88492 | 0.7027 | 0.87079 | 0.81818 | 0.88112* | 0.875 | 0.83333 |
| 120 | 2 | 454-456 | 0.375 | 0.34722 | 0.21622 | 0.19548 | 0.18182 | 0.26107 | + | + |
| A130 | 15 | 171-194 | 0.08 | 0.81143* | 0.19444 | 0.48983* | 0.09375 | 0.65278* | 0.33333 | 0.54545 |
| B6 | 8 | 141-159 | 0.78125 | 0.69147 | 0.75676 | 0.74639 | 0.78788 | 0.68951 | 1 | 0.66667* |
| CF11 | 19 | 184-214 | 0.77419 | 0.71444 | 0.83784 | 0.79193 | 0.78788 | 0.69697 | 0.75 | 0.73333 |
| D103 | 15 | 250-306 | 0.96875 | 0.88492 | 0.81081 | 0.88227 | 0.81818 | 0.88392 | 0.875 | 0.91667 |
| D114 | 6 | 202-222 | 0.6129 | 0.63194 | 0.62162 | 0.63865 | 0.72727 | 0.67552 | 0.75 | 0.71667 |

Table 1.3: Pairwise $F_{ST}$ values for sites sampled in this study. Asterisks indicate significance ($p < 0.05$) based on 100 permutations computed in Arlequin v3.1 (Excoffier et al., 2005).

| | GLO | JDN | BAS | EUR |
|---|---|---|---|---|
| GLO | 0 | - | - | - |
| JDN | 0.04746* | 0 | - | - |
| BAS | 0.00909 | 0.03373* | 0 | - |
| EUR | 0.06064* | 0.05622* | 0.0353* | 0 |

# Chapter 2

# Social media reveals properties of a mutualistic network at multiple spatial scales

## Introduction

The interactions among species are major drivers of ecological and evolutionary dynamics. They are responsible for both short and long-term evolutionary change, and can cause sudden and drastic ecosystem shifts (Moreau et al., 2006; Thompson, 1998; Estes et al., 2011; Myers and Worm, 2003). The accumulation of interaction data for a range of systems has permitted ecologists to draw inference about the dynamics of groups of interacting species. These interactions can be analyzed as a network, where species and the interactions among them are represented by nodes and edges, respectively. All organisms are part of a single global network that describes the interactions between all living things, but ecologists can gain insight by analyzing a discrete subset of interactions, the organization and spatial and taxonomic bounds of which are chosen

arbitrarily depending on the question of interest. Food webs, perhaps the most familiar type of interaction network, may be organized hierarchically, with species grouped into discrete trophic levels (Estes et al., 1998), while plant-pollinator systems can be thought of as a bipartite network in which species are categorized into one of two suites of species (Bascompte et al., 2003). Ecologists are eager to understand the rules that govern the structure of interaction networks because they represent the manifestation of processes operating at the levels of the the individual, population, and species. The study of the structure of interaction networks has revealed food web dynamics such as trophic cascades (Estes et al., 1998) and non-random extinction potential in mutualistic networks (Rezende et al., 2007).

The spatial arrangement of individuals and populations is an important predictor of the distribution of genetic and phenotypic variation within and among species (Hanski, 1998; Turelli et al., 2001; Futuyma and Mayer, 1980), and it has been shown that selection on interspecific interactions varies among populations (Thompson, 2009, 2005), even over small spatial scales (Thompson and Cunningham, 2002). Because interactions among species are constrained by their spatio-temporal co-occurrence, a robust analysis of the structure of the interaction network between suites of species should incorporate information concerning the spatial scale of the interaction (i.e. the rate of incidence at the level of individual, population, and species). Though many studies have examined the structure of interaction networks (Bascompte et al., 2003), none explicitly consider the spatial component of a network inclusive of all potentially interacting species. Thus, it is not known whether the structural properties of interaction networks

are consistent across spatial scales. An understanding of the critical spatial scale at which patterns arise informs the search for plausible processes structuring interaction networks.

The size of the spatial distributions of interacting species may play an important role in structuring interaction networks. Among species that vary in their degree of ecological specialization, generalists are often expected to have larger ranges, though it is unclear whether range size is a driver or product of specialization (Gaston et al., 1997; Bell, 2001). Generalists may establish and sustain larger range sizes because their fitness is not tied to the presence of a single partner, or species with larger ranges may be generalists simply because they encounter more partner species. Both explanations of this relationship implicitly assume that the spatial distribution of diversity of partnering species is uneven. If partner diversity is even in space, the likelihood of encounter is equal for all species regardless of range size, and specialization is therefore merely not a consequence of small range size.

Because interactions vary in their identity, nature, and strength over small spatial scales (Thompson, 2005), coarse-scale data may not provide the resolution necessary to capture the true organization of interaction networks. Additionally, indices of specialization and network structure can be biased by small sample sizes; for example, rarely observed species are likely to be incorrectly categorized as specialists (Blüthgen et al., 2008; Nielsen and Bascompte, 2007) Taken together, it is critical that ecologists who wish to better understand broad-scale patterns in species interactions compile data sets that are as spatially and taxonomically inclusive as possible, and comprise a large

number of observations. However, acquiring comprehensive datasets spanning large geographic areas is rarely feasible due to the time and cost associated with gathering such data. Coincidentally, the extinction rate of species and populations is growing at an unprecedented rate (Estes et al., 2011; Wake and Vredenburg, 2008), making the need for such large-scale datasets more urgent than ever.

One potential source of such data is photographs shared publicly online. Though not intended to be used by scientists, these photos constitute a wealth of potential data, but their use in this capacity has been limited to date (though see Stafford et al. (2010)). For example, Flickr, a popular photo sharing website, hosts more than six billion user-uploaded photos, nearly three million of which are associated with specific geographic coordinates ("geotagged"/georeferenced?). In order to test the efficacy of such repositories in resolving large scale problems in ecology, I focused on the anemonefishes, a conspicuous and uniquely photogenic clade of coral reef fishes engaging in an obligate mutualism with sea anemones. Specifically, I used these data to characterize the structure and specialization of the anemone-fish interaction on a global scale (the broadest spatial scale in the context of the interaction: the scale at which all participants in the system are included), and to answer the following questions:

1. Does network structure differ among regional networks? Does this differ from that of the global network?

2. Is there a relationship between the degree of spatial overlap and the degree of interaction overlap among species within suites?

If network structure at local scales is not significantly different from that at global or regional scales, similar processes and mechanisms are likely to generate and maintain the structure of the interaction network across spatial scales. Conversely, significant differences in network structure across spatial scales suggest that local and global factors operate and independently influence network structure. A positive relationship between spatial overlap and host use suggests species coexist without competing, while a negative relationship implies that species partition space, resources, or both in order to coexist. These outcomes inform the expectation of the spatial scale at which coevolution acts among partners in this system.

# Methods

## Study System

The anemonefishes (Perciformes: Pomacentridae: Amphiprioninae) are a monophyletic subfamily of damselfishes found on coral reefs from East Africa to French Polynesia. Thirty species are recognized, though one of these (*Amphiprion thielli* Burgess 1981) is known only from two aquarium specimens of dubious geographic origin, and I disregard it here (see also Fautin and Allen (1997)). All species are obligate mutualistic symbionts of a polyphyletic group of ten species of sea anemones (Anthozoa: Actiniaria) from three families. Individual fish settle to and remain with a single host for the (span/duration) of their lives. Allen (Allen, 1975; Fautin and Allen, 1997) is the primary source for host association data, but the frequency of associations has only been

explicitly examined in two studies of limited geographic scope (Elliott and Mariscal, 2001; Ricciardi et al., 2010).

## Data Collection

I queried Flickr for geotagged photos matching any of a variety of keywords across the entire spatial range of the anemone-fish interaction (Supplemental Material). For each photo containing an anemonefish, host anemone, or both, I identified the fish and anemone to the species level following Allen (1975) and Fautin (1981). I considered each photo to represent one interaction regardless of the number of individual fish or anemones portrayed, and mixed species groups were included as a single event for each species. Duplicate records were omitted from the analysis (Supplemental Material). The density of photographs on Flickr was low in some locations, and likewise for endemics of these regions (e.g. *Amphiprion chagosensis* from the Chagos Archipelago). In order to ensure distributions and interactions were estimated from as much information as possible, I incorporated data from the primary literature, solicited unpublished data from colleagues, and extended the search to include photos that were not geotagged by coupling the search terms with geocodable place names, both on Flickr and Google (Supplemental Material).

## Data Analysis

Data were processed and analyzed in R 3.0.2 (R Core Team, 2013) using packages cited individually below.

I estimated the distribution of each species using a simple presence/absence geographic distance model informed by the data (Supplemental Material), and overlaid these distributions onto a high resolution digitized map of the world's coral reefs (UNEP-WCMC et al., 2010). Despite the simplicity of this approach, the modeled distributions of both fish and anemones were nearly identical to those published by experts in the field (Fautin and Allen, 1997). I used these to quantify the range size (sea surface area encompassed by the modeled distribution) and habitat area (area of coral reef encompassed by the modeled distribution) of each species, and to provide the expected presence/absence of species at a given location.

I quantified specialization at the network level using the index $H_2'$ proposed by Blüthgen et al. (2006) and at the species level using the Shannon entropy ($H'$) (Shannon, 1948). Both indices are derived from Kullback-Leibler divergence: $H_2'$ is standardized to range from 0 for extreme generalization to 1 for extreme specialization, while Shannon's Entropy H$'$ ranges from 0 for a species which interacts with only one partner and increases with the number and evenness of partners (formulas given in Supplemental Material). Because Shannon entropy is unbounded and blind to the pool of available resources (i.e. a species using only one resource has a Shannon entropy of zero whether there is one resource available or 50), I also used a corrected measure of Shannon entropy, $H_c$ which is scaled from 0 for generalization to 1 for specialization (Supplemental Material).

An interaction network is said to be nested if species with fewer partners interact with a subset of those with more partners, for both suites of species (Patterson and

Atmar, 1986). The prevalence of this pattern not only in mutualistic interaction networks but in species incidence among habitat patches suggests it may be a consequence of variation in species abundance. I quantified the nestedness of the anemone-fish network using a weighted metric based on overlap and decreasing fill (WNODF), which increases with nestedness from 0 to 100 (Almeida-Neto and Ulrich, 2011). I also calculated its binary counterpart (NODF) to facilitate comparison to the published system-wide host associations which do not account for interaction frequency (Almeida-Neto et al., 2008). Indices were computed using the R packages bipartite and vegan (Dormann et al., 2008; Oksanen et al., 2013). I regard a network as nested if its index is greater than half the maximum index value.

I generated null matrices on which the various network indices could be computed for comparison to that of empirical matrices according to a null model in which the relative species abundance of both suites is held constant to the empirical matrix, and resulting interactions are filled randomly within this constraints (Algorithm AS159 (Patefield, 1981); implemented in R by function "r2dtable"). I also generated resampled null matrices by sampling interaction observations from the original matrix with replacement, with the number of samples either equal to the original matrix sum, or half of the matrix sum.

Observations were not distributed in space randomly, evenly, or uniformly with respect to habitat; therefore, in order to assess the spatial variation of various measures, I generated 10,000 randomly distributed points over the potential habitat identified by the species distribution models (i.e. the intersection of all species distribution models

and all coral reef habitat) (Figure 2.1). At each of these points, I inferred the species incidence from the species distribution models, and estimated local specialization for each suite by calculating the mean corrected Shannon Entropy ($\overline{H}_c$) of the interactions within the more conservative of the buffers used for the species distributions models (Supplemental Material).

The occurrence of species at each of these points was also used to compute pairwise Bray-Curtis dissimilarity of spatial overlap among all species within a suite (fish or anemones). Likewise, pairwise Bray-Curtis dissimilarities were calculated from the quantitative and binary interaction matrices, for each suite of species. I tested for correlation between distribution and partner dissimilarity among species by conducting Mantel tests on these distance matrices, and employed a permutational test of significance by permuting matrix rows and recalculating correlation coefficient (10,000 replicates).

# Results

## Search Method

The automated Flickr query returned a total of 4881 images containing at least one anemonefish or host from 3344 unique locations and comprising 3864 anemonefish pairings (Table 2.1). Some images contained more than one species of fish or anemone, thus the total number of observations of a fish, anemone, or both was 5290. An additional 1084 records were gleaned from auxiliary sources, bringing the total number

of observations in the data set to 6374, representing all species of anemonefishes and host anemones.

The number of observations varied by species, but was consistent with expectations based on the amount of habitat available to each species. I calculated the expected number of observations for each species by multiplying the proportion of 10,000 random points that occur within it's modeled distribution by the total number of observations for all fish or anemone species (Figure 2.6).

## Species Distributions

Three distinct regions were apparent with respect to fish species richness: the Indian ("West"), the Indo-Australian Archipelago ("Central"/"IAA"), and Pacific ("East") (Figure 2.1). Twenty-two (79.3%) fish species were restricted to one of these three regions. Only one species (3.4%; *A. clarkii*) occurred in all three regions, one species (3.4%; *A. akallopisos*) occurred in both the West and Central regions, and four species (13.8%; *A. akindynos*, *A. chrysopterus*, *A. melanopus*, and *A. perideraion*) were found in the Central and Eastern regions. The Western and Central regions are comparable in size and species richness of fish and anemones, but share only two fish species. Diversity varies substantially between them: The fauna of the Central region is both rich and spatially even, with high local diversity, while the Western region is rich in total number of species but with low local diversity.

Conversely, local anemone richness was relatively high in nearly every part of the geographic range of the interaction. All species of anemone occur in the Indo-

Australian Archipelago; all but one (*S. gigantea)* occur in the Indian Ocean basin, and all species' ranges extend at least partially into the Pacific.

At the randomly generated sites ($N$ = 10,000 ), anemone diversity was nested ($N_{col}$ = 94; $N_{row}$ = 74; NODF = 74; Matrix fill = 0.76), while fish diversity was not nested ($N_{col}$ = 20; $N_{row}$ = 30; NODF = 30; Matrix fill = 0.16) (Figure 2.3).

## Species Interactions

Of the 290 possible anemone-fish species pairings, 46 were determined to be impossible due to spatial non-overlap of the species' ranges, and 98 species pairs were observed. The overall pattern of binary host association did not differ from that reported in Ollerton et al. (2007) (Pearson's chi-squared test; $\chi^2$ = 130, df = 225, $p$ = 1). However, the identity of interactions differed: 14 interactions (4.8% of 290) reported by Ollerton were not observed in the present study, 32 observed in the present study were previously unreported (11%), and the remaining 244 were consistent between studies, comprising 66 interactions (23%) and 178 empty cells (61.4%) (Figure 2.4). Previously unpublished host association data were obtained for four fish species (*A. chagosensis*, *A. omanensis*, *A. barberi*, and *A. pacificus*). The two matrices were similarly slightly nested ($NODF_{Ollerton}$ = 56; $NODF_{present}$ = 58).

## Regional Network Structure

The global quantitative matrix ($N_{obs}$ = 4830) is not nested (WNODF = 33), and significantly less so than null matrices generated with equivalent marginal sums

and matrix fill ($WNODF_{mean} = 73$). The Western and Central regional networks had a value of WNODF less than 100% of 10,000 null matrices; the value of WNODF was less than 98.7% of 10,000 null matrices (Figure 2.5A). Resampled values of $H_2'$ also did not differ significantly by region, and all networks were much more specialized than expected based on the null hypothesis (Figure 2.5B).

### Spatial Overlap and Interaction Overlap

Neither fish nor anemones showed any relationship between their spatial distribution and pattern of interactions, as indicated by the Mantel test on pairwise dissimilarities of those properties (fish: $r = 0.003$; significance = 0.45; anemones: $r = 0.177$; significance = 0.13) Additionally, binary host associations did not show any relationship with spatial distribution for the fish ($r = 0.051$; significance: 0.257), though they did result in a positive and significant correlation for the anemones ($r = 0.535$; significance: 0.002) Thus, fish species with similar spatial distributions do not have similar host use patterns, but anemones with similar spatial distributions do tend to host similar fish assemblages.

## Discussion

These analyses demonstrate that the structure of this mutualistic network dissolves once the spatial scale of focus becomes smaller than the ocean basin level. Outside of the central core of diversity in the Coral Triangle, local species richness of fish is so low (1-3), that network indices are no longer meaningful. Nevertheless, these

areas retain high species richness of anemones (8-10), indicating that a nested network structure is not required to maintain diversity at local scales.

Differences in the distribution of diversity between fish and anemones may reflect different scales of dispersal. It is known that the dispersal of anemonefish is extremely localized, more so than other marine fishes (Jones et al., 2005). The scale of dispersal in anemones is unknown, but the lack of fine-scale endemism and large range size shown here predicts that dispersal scales should be broader in this group.

Few studies have examined network structure on large spatial scales, but we know from studies of pollination and seed dispersal that these networks tend to be nested at small spatial scales (Bascompte et al., 2003). Given that the anemone-fish network is only slightly more nested than random at the broadest spatial scales, and that local networks are even less nested, it appears the anemone-fish interaction differs fundamentally from the more intensely studied networks from which many of the generalizations about interaction networks have been made. Coevolutionary theory predicts that nestedness and specialization should vary as a function of interaction intimacy (the degree of biological association between partners) (Thompson, 2005), and empirical evidence for this prediction is accumulating (Guimarães et al., 2007; Hembry, 2012; Thompson et al., 2013). Additionally, the spatial scale of dispersal of participating species may influence selection and thus network structure.

In considering only local patterns, it appears the anemone-fish mutualism deviates substantially from expectations, though these expectations are based largely on pollination and seed dispersal networks. The results presented here provide further ev-

idence that patterns of network structure are not consistent across all mutualisms, and instead, that specific properties such as interaction intimacy may be the best predictors of network structure.

The role of interaction intimacy reflects the importance of the balance between costs and benefits to each species, balanced by selection. While the benefits of mutualisms often gain substantial attention, costs can be less obvious but equally important drivers of selection (Addicott, 1986; Bronstein, 2001). Investigating the costs of the anemone-fish interaction may provide fruitful insights into the processes that structure it on broad levels. In particular, the production of nematocysts (stinging cells) may be energetically costly to anemones, but mediated by hosting resident fish.

The lack of a relationship between spatial distributions and interactions is indicative of, but not evidence for, the presence of competition for hosts among fish species. While a negative relationship may be expected, such relationships are uncommon (Dutilleul et al., 2000; Legendre and Fortin, 2010), and the non-overlapping spatial distributions of many fish species may preclude such a result. Nevertheless, competition has been implicated in behavioral studies of these species, and likely structures large-scale patterns of the interaction (Elliott and Mariscal, 1995). On the anemone side, the positive relationship between spatial and binary interactions with fish species likely reflects the anemones' generally large spatial distributions.

Broad-scale studies of interaction networks have been hampered by a lack of taxonomically and spatially comprehensive data sets. I present a novel method for obtaining such data, and demonstrate the utility of those data to addressing questions

regarding the structure of the interaction network of the anemone-fish system, at both the network and species level. While this study exploits the large amount of data available in a photogenic species, the increased use of cameras, mobile phones, and social media suggests this method will become a widespread utility for other organisms. The economy of this method will also increase rapidly with the development of the field of computer vision. Incorporating technology to discriminate among subjects in photos even on coarse levels will reduce the workload of taxonomists, and facilitate ecologists' access to global-scale datasets. While this method cannot replace field surveys with respect to the sampling intensity at small scales, it is instrumental in gathering data for many species over broad geographic areas.

## Acknowledgements

I wish to thank Amy McDermott, Walter Sessions, and Julie Erickson for help with data collection and writing, as well as John N. Thompson, Giacomo Bernardi, and Peter Raimondi for their feedback during conceptualization, analysis, and writing. David Hembry's input on an earlier draft was indispensable. I was supported by a Dissertation Year Fellowship from the University of California Santa Cruz while conducting this research.

Table 2.1: Query results.

| Category | Number of Observations |
|---|---|
| Georeferenced photos on Flickr matching query terms | 9083 |
| Photos documenting $\geq 1$ species of interest | 4880 |
| Site records from Flickr query (fish OR anemone) | 5290 |
| Interactions from Flickr search | 3863 |
| Accessory records | 1084 |
| Total unique site records (Flickr + accessory) | 11029 |
| Total unique site records (anemones) | 5844 |
| Total unique site records (fish) | 5185 |
| Total photos analyzed (Flickr + accessory) | 10167 |
| Total interactions (Flickr + accessory) | 4830 |

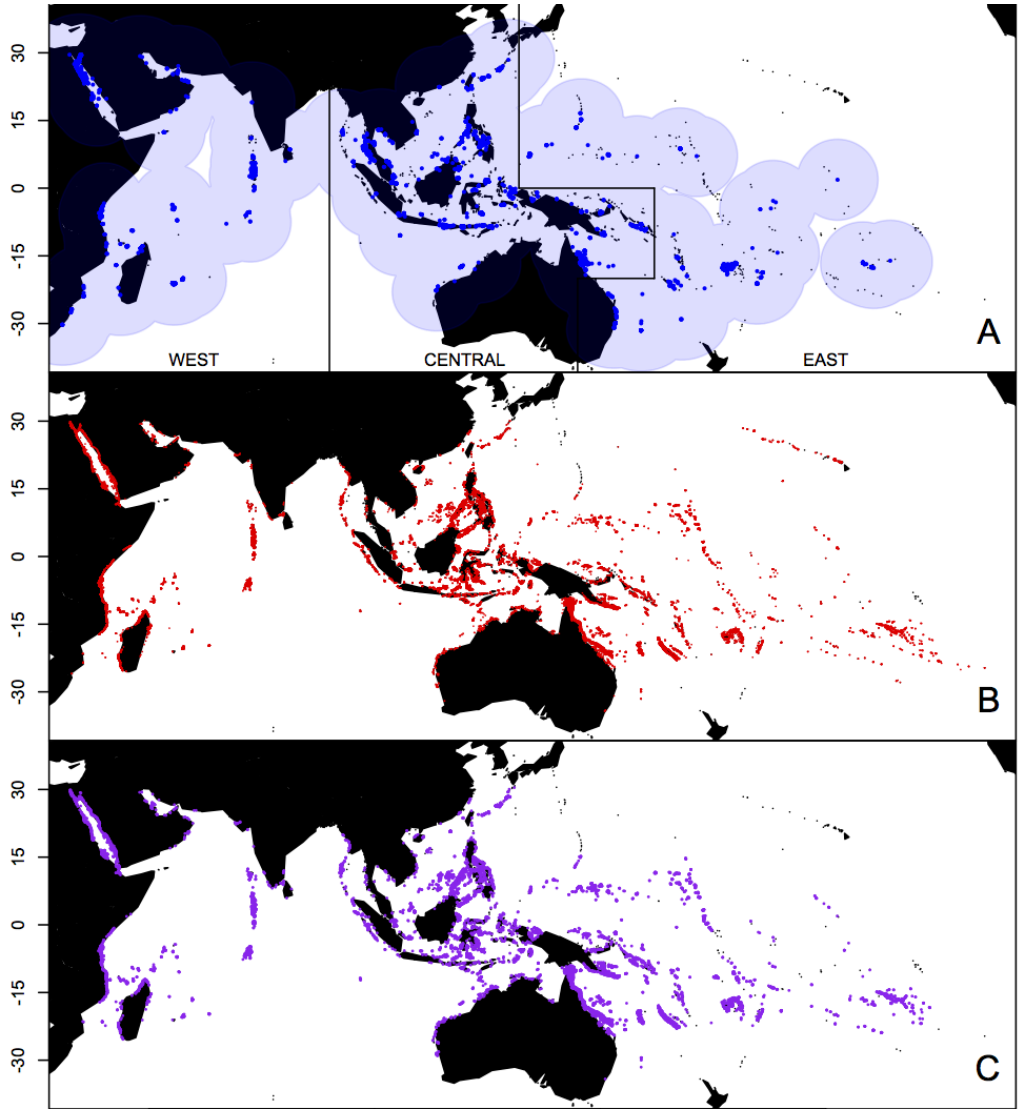



Figure 2.1: Maps illustrating spatial approach. Blue dots (A) represent observations from the image search for all species; shaded areas represent the buffer around those points used for the distribution model. The system-wide distribution model was then mapped onto polygons of a digitized map of coral reefs (B; red). Random points (purple dots in C; $n$ = 10,000) were generated on the intersection of the distribution model (shaded blue areas in A) and coral reef habitat (B). At each of these points, I inferred species occurrence from individual species distribution models. Note the absence of anemonefishes in the Hawaiian archipelago (A) and the presence of coral reefs there (B) result in the exclusion of that region from the random point set (C). Distinct regions of diversity are outlined in A.

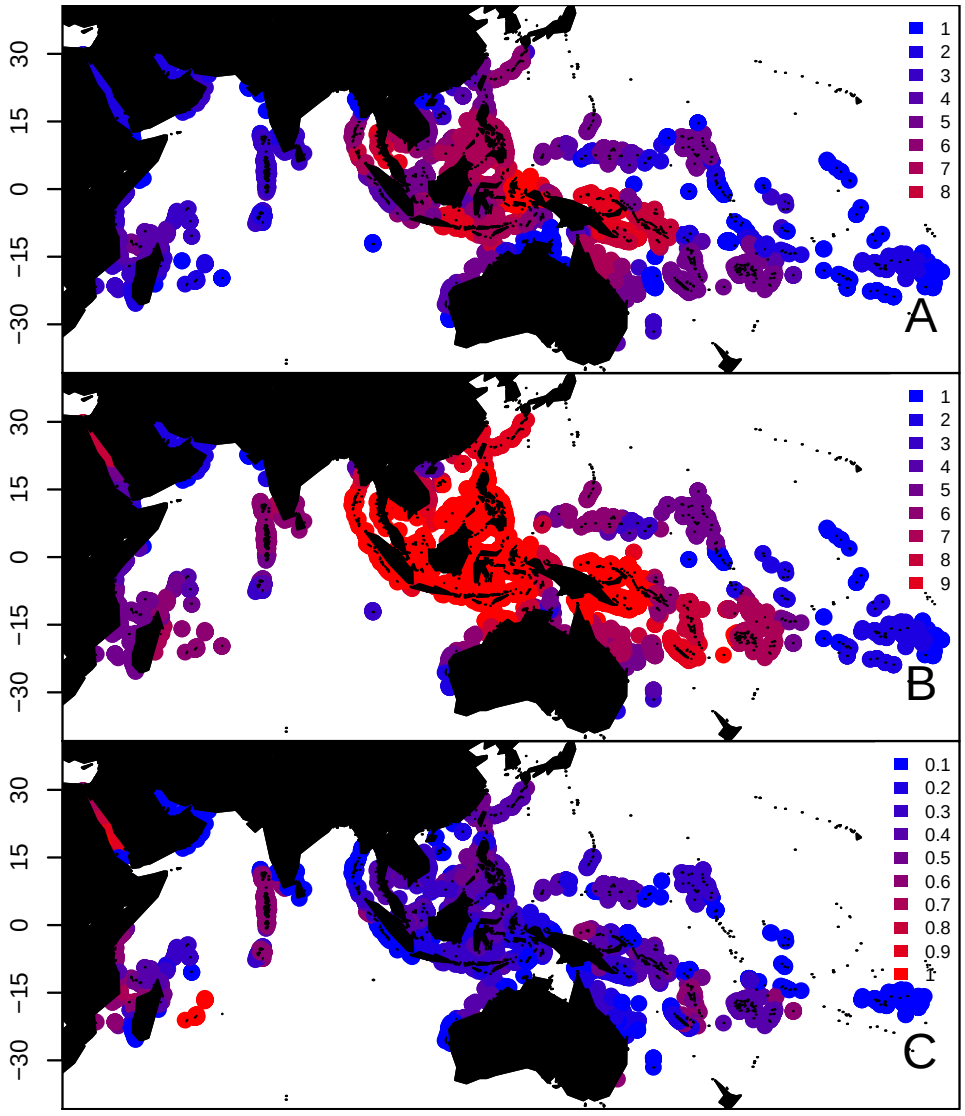



Figure 2.2: Random points colored by (A) fish richness, (B) anemone richness, and (C) local fish specialization ($\overline{H}_c$). Local specialization is not displayed for random points which fall further than 528720 m from an observed interaction.

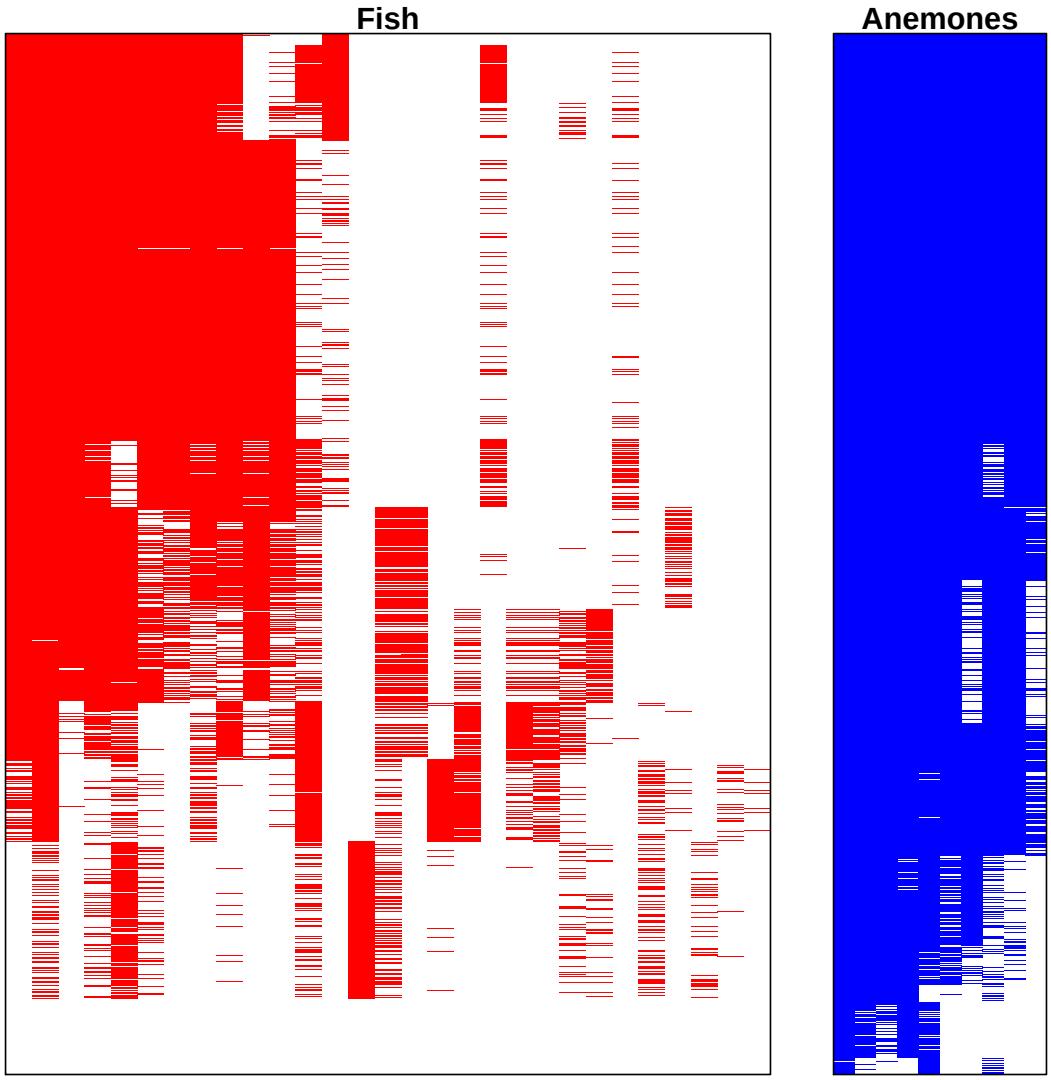
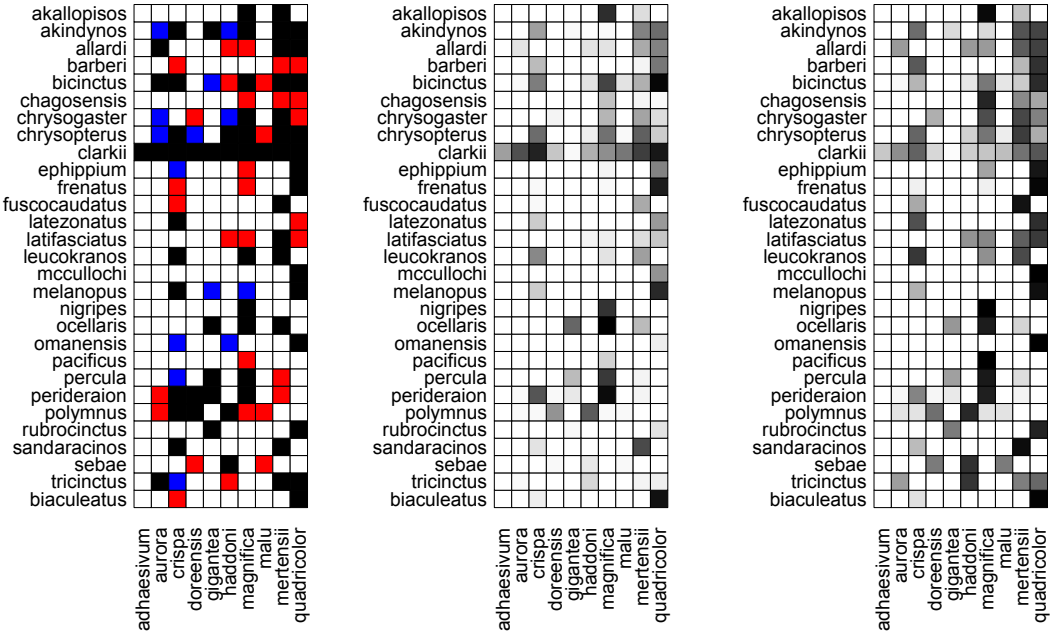



Figure 2.3: Geographic nestedness of fish (left; red) and anemones (right; blue). Rows represent 10,000 randomly generated sites across all potential habitat; columns represent species. Fill indicates a species' presence at a given location, and rows and columns are arranged by decreasing frequency from top left to bottom right.

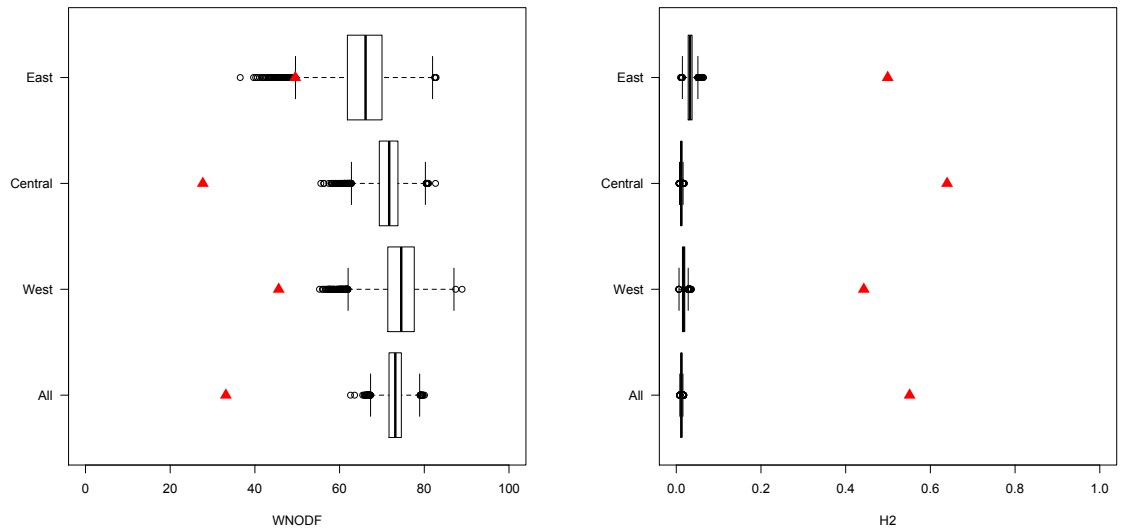



(a) Interactions colored by their occurrence in each of two studies. Blue: Ollerton et al. (2007); red: present study, black: both. White boxes represent interactions not observed by either study.

(b) Interactions shaded by their frequency relative to the entire data set. N = 4831

(c) Interactions shaded by their frequency relative to each fish species. N = 4831

Figure 2.4: Interactions between anemones and fish. Note that rows and columns are arranged alphabetically to facilitate visual comparison among matrices, rather than being arranged by fill, as is done to calculate nestedness.




(a) (b)

Figure 2.5: Variation in nestedness (WNODF; a) and specialization ($H_2'$; b) of networks by region. Box plots illustrate median value of resampled or null matrices. Boxes encompass the interquartile range (IQR), lines extend to 1.5IQR, and dots represent outliers. Empirical values of the index are represented by red triangles.

# Supplemental Material

## Search Method

Images were queried and downloaded from Flickr[1] using the Python interface to the Flickr API[2]. For a variety of bounding boxes encompassing the range of all anemonefish species, I queried and downloaded images tagged with any of seven relevant character strings ("amphiprion", "anemone", "anemonefish", "clownfish", "clown fish", "nemo", and "premnas"). As an informal test for the occurrence of false positives, I also conducted the query around the Hawaiian islands, a heavily touristed, dived, and photographed area where anemonefishes are known to be absent. No relevant photos were returned from this query.

Records containing specimens of uncertain taxonomic identity were omitted from the analysis, as were multiple photos of the same subject as determined by identical location, interaction identity, and photo similarity. Cases in which more than one fish species shared a single host or vice versa were recorded as a single event for each species. Records that constituted a significant range expansion were reviewed for evidence such as album titles, tags, and captions to confirm the location. If it was not possible to confirm the location based on secondary lines of evidence, the photo was omitted from the analysis. The identity of the individuals in each photo was assessed at least three times: first when both species names were recorded, and a second and third time after the photos were parsed by either fish or anemone species for error checking.

[1] original code at http://graphics.cs.cmu.edu/projects/im2gps/flickr_code.html
[2] available from: http://pypi.python.org/pypi/flickrapi

Both fish and host species can be identified visually, making them uniquely suited to this application. For the morphologically cryptic species *A. ocellaris* and *A. percula*, I designated individuals to one of these species based on their location with respect to the genetic break defined in Timm et al. (2008) For records taken from sources besides the standardized Flickr search, place names were manually geocoded using a variety of data sources ranging from Wikipedia's GeoHack Tool to personal blogs and scuba diving websites. Coordinates were only extrapolated for place names that were sufficiently specific (e.g. for "Guam", but not "Papua New Guinea").

### Search Efficacy and Sampling Evenness

In order to account for sampling bias, I generated 10,000 random points over the intersection of coral reef habitat and the union of all species distributions, and calculated which species are present at each of these points according to the species distribution models. This provides estimates of species level variables that are spatially random with respect to habitat available to all species, as well as providing an expected number of observations if samples had been randomly placed. The number of photos returned varied by species. I compared these values to the expected number of observations for each species, which I defined as the proportion of 10,000 spatially random points that lie within the species' range, multiplied by the total number of observations in the dataset (Figure 2.6).

I also used these points to calculate pairwise spatial overlap for all within-suite (fish and anemones) species pairs using Bray-Curtis dissimilarity values. I conducted a

Mantel test to compare these values to pairwise niche overlap, defined as the Bray-Curtis dissimilarity of two species' resource use.

### Species Distribution Models

Both anemonefishes and their host anemones have small home ranges (fish do not leave their sessile hosts), and dispersal estimates over evolutionary time scales (i.e. 100 generations) are for fishes uninformatively broad at best, and nonexistent for anemones. Therefore, I used the location of observations to inform simple presence-absence geographic distance models of species' distributions whereby a species is present within a given radius of each observation, and absent beyond. The maximum nearest neighbor distance of a set of observations is informative to this end because it defines the greatest distance over which we can be certain that individuals of a given species can travel. I calculated the maximum nearest neighbor distances of the observations of each species (Magnusson, 2012), and used the median of these as the radius of the buffer around each observation. For fish, the median of the maximum nearest neighbors was 528.72 km. For anemones, this value was so huge as to be uninformative (1772.688 km), due to great distances between observations of anemone species which have broad ranges but are very rare. Anemonefish pelagic larval duration is among the shortest of any marine fish with pelagic larvae, while the anemone species that have been investigated had high levels of population connectivity over broader scales. Under the assumption that interaction dynamics are likely constrained by the dispersal potential of the least-dispersive suite of species, I used the fish estimate for both sets of species

distribution models. I assumed each species' distribution to be the union of the buffers around observations of that species. General spatial operations were performed using the package sp (Pebesma and Bivand, 2005) and rgeos (Bivand and Rundel, 2013) and area was calculated using the package geosphere (Hijmans et al., 2012).

### Null Models

The choice of null model can affect the outcome of network analyses, and therefore warrants careful consideration. Null models for ecological networks have been discussed at length (Ulrich et al., 2009), and the appropriate model may vary depending on both the biological system and question of interest. Compared to biogeographic species-site matrices, interaction networks in particular suffer from a lack of consensus regarding null model choice. Moreover, the anemone-fish mutualism differs from the more commonly studied pollination and seed dispersal mutualisms in several fundamental ways, thus making null model selection less than straightforward.

Anemones are sessile, while mobile fish larvae recruit to a host during settlement from the larval stage, presumably with some potential for selectivity during this time. Fish do not leave the host after settlement, thus, only one interaction occurs per individual fish. Mixed species groups are rare; it is therefore likely that either choice or post-settlement survival is affected by the presence and absence of conspecifics and heterospecifics. However, no studies have been published regarding the competitive hierarchies among species, or the effect of heterospecifics on larval settlement or post-settlement mortality. Because the presence of fish increases anemone survival,

growth, and reproduction (Holbrook and Schmitt, 2005), there is also likely to be positive feedback between relative abundance of both the fish and anemones, mediated by fish selectivity. Given these considerations, I regard the most biologically realistic null model as one in which anemone relative abundance (column totals) is held constant, fish relative abundance (row totals) is held constant, and resulting interactions are filled randomly within these constraints (Algorithm AS159 (Patefield, 1981); implemented in R by function "r2dtable"). Nevertheless, I also evaluated network indices for various other null models, including unconstrained equiprobable models, and in no case did this affect the direction of effects reported here.

## Quantifying Specialization

I calculated the $H_2'$ index of network in the R package bipartite, the value of which is given by

$$H_2' = -\sum_{i=1}^{r}\sum_{j=1}^{c}(p_{ij} \cdot \ln p_{ij}) \tag{2.1}$$

where $i$ and $j$ represent rows and columns of an interaction matrix, the numbers of which are represented by $r$ and $c$, respectively, and the proportion of each interaction is denoted by $p_{ij}$. Shannon's Entropy is a measure of the evenness and richness of community of species. Its value is given by

$$H' = -\sum_{i=1}^{S}(p_i \cdot \ln p_i) \tag{2.2}$$

where $p_i$ is the proportional occurrence of species $i$, for a given number of species $S$.

In order to make comparisons across species and account for differences in partner availability, I used a Shannon index scaled to the minimum and maximum values possible given the local richness of partners. This correction, similar to Blüthgen et al. (2006) and Yeakel et al. (2013), is given as follows:

$$H'_c = \frac{H' - H'_{even}}{H'_{uneven} - H'_{even}} \tag{2.3}$$

where $H'_{uneven}$ is the Shannon entropy if only one partner is used to the exclusion of others (and is thus always equal to zero), $H'_{even}$ is the Shannon entropy if all available partners are used with equal proportions, and $H'$ is the uncorrected Shannon entropy. The potential number of local partners is computed from species distribution models. While this index does not take into account differences in relative abundance of resources, it does account for "forbidden interactions", those which cannot occur due to non-overlap of species' distributions. Thus, species that interact with all resources in equal proportions have identical corrected indices, regardless of the number of resources available. Because $H'_{even}$ is 0 for a system with only a single resource, the corrected entropy in these systems is undefined. In order to include these networks in analyses, I assigned them a value of zero, indicating maximum generalization, because the fish is using all resources available to it.

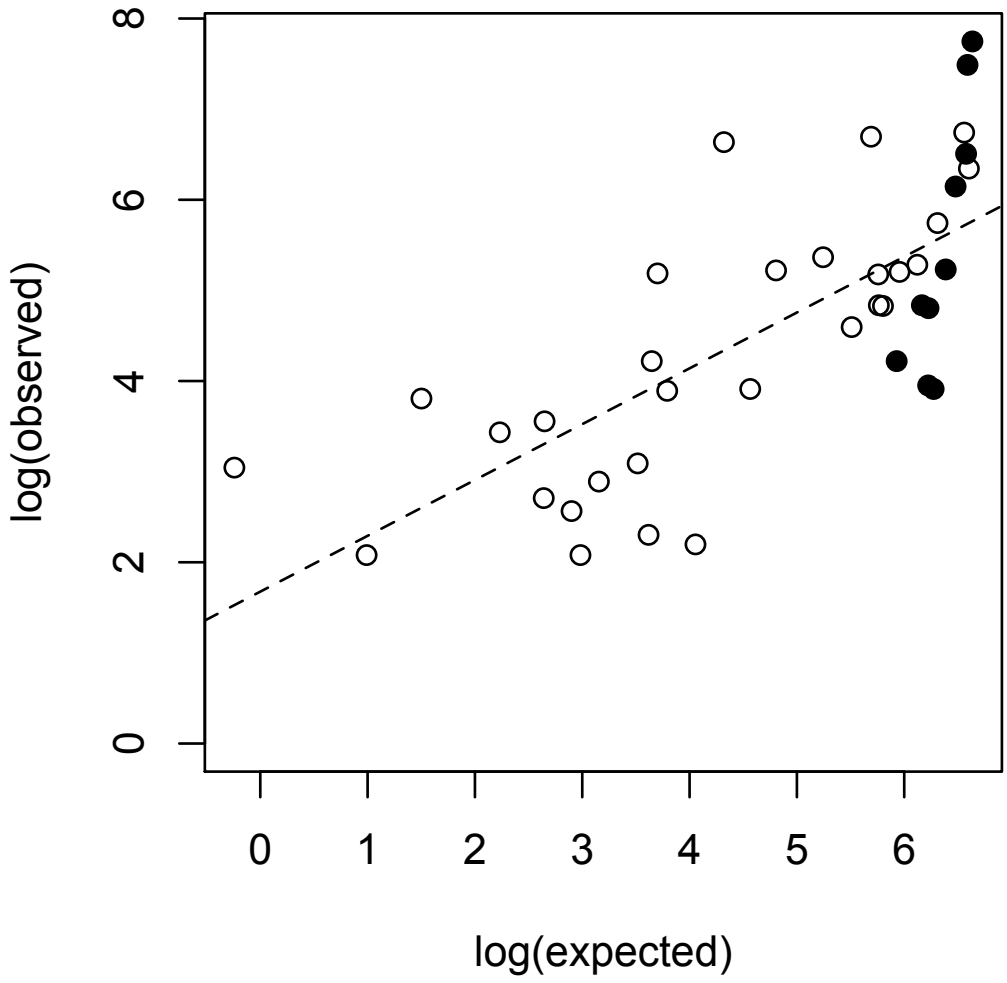



Figure 2.6: Relationship between number of expected and observed observations. Open circles: fish species; filled circles: anemones; dashed line: line of best fit. F-statistic = 41.29 on 1 and 37 df; $p$ = 1.669e-07; $R^2_{adj}$ = 0.5146. [$\log(N_{obs}) = 0.616\log(N_{exp}) + 1.67589$]

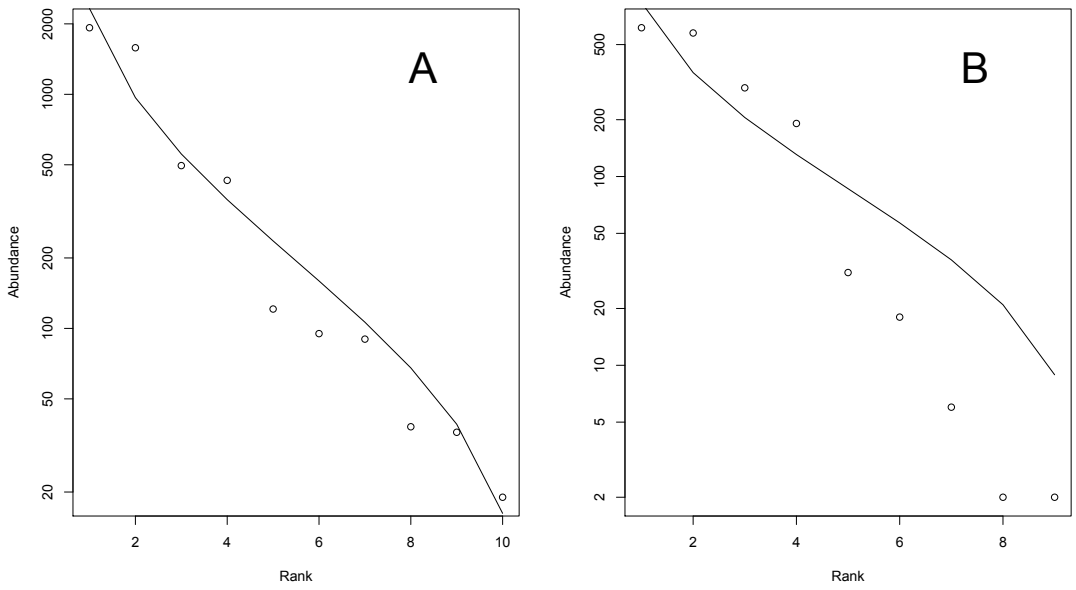



Figure 2.7: Abundance of anemone species in the present data set (A) and in Elliott and Mariscal (2001). Lines represent fitted values to a lognormal distribution with mean and standard deviation equal to (A) 5.268963, 1.605332; and (B) 4.456957, 1.519102, respectively.

# Chapter 3

# A Multi-locus Phylogeny of the Anemonefishes

## Introduction

More than one-third of all extant vertebrate species are ray-finned fishes (class Actinopterygii), and coral reef associated families of fish are among the largest (Froese and Pauly, 2014). Elucidating the evolutionary history of these groups is an important step towards understanding the diversity of vertebrates. Their diversification is likely the result of a combination of biogeographic and ecological processes, and the study of large and ecologically diverse families, such as damselfishes, may shed light on such processes.

The damselfishes (Perciformes: Pomacentridae) are a large and conspicuous family of reef fishes whose evolutionary history is of central interest to ecologists and evolutionary biologists. The group originated in the Eocene (Frédérich et al., 2013) and currently has 375 recognized extant species, which are distributed primarily on

[1]Additional coauthors on this manuscript: Paul H. Barber and Giacomo Bernardi

coral reefs (Froese and Pauly, 2014). Due to their abundance, relatively small size, and amenability to experimental manipulation, Pomacentrids have contributed immensely to our understanding of marine population dynamics (Carr et al., 2002), the evolution of reproductive strategies (Fricke and Fricke, 1977), and general patterns in the structure of biodiversity (Hixon and Brostoff, 1983). Recent work has shown that within the Pomacentridae, the subfamily Amphiprioninae has undergone extremely rapid diversification over the past 10-22 million years (Frédérich et al., 2013).

The Amphiprioninae (anemonefishes) are small damselfishes that live amongst the tentacles of sea anemones on Indo-Pacific coral reefs from East Africa to French Polynesia. This obligate mutualism provides the fish with shelter from predators, while the fish guards the anemone from predators and increases anemone growth and reproduction by supplementing nutrients to the anemone's photosynthetic zooxanthellae via excrement (Porat and Chadwick-Furman, 2004; Holbrook and Schmitt, 2005; Roopin and Chadwick, 2009).

Thirty species of anemonefish are currently recognized (Froese and Pauly, 2014), though one of these (*Amphiprion thielli* Burgess 1981) is known only from two aquarium specimens of dubious geographic origin, and we disregard it here (see also Fautin and Allen, 1997). The species are grouped into two genera, *Amphiprion* and *Premnas*, though *Premnas* has repeatedly been shown to be nested within *Amphiprion*. Ten species of anemone hosts are recognized (Fautin and Allen, 1997), and individual fish associate with a single host for the entirety of their post-settlement life. Some fish species are only known to associate with a single host species, while others can be found

with any of the ten hosts. All species are protandrous hermaphrodites and are usually found in small monospecific groups. Pairs guard nests of demersal eggs until hatching, at which point larvae begin a pelagic phase lasting 7-22 days (Wellington and Victor, 1989; Thresher et al., 1989).

Species richness of anemones and anemonefishes parallels larger diversity patterns within the family and is highest in the Coral Triangle (the reefs of the Indo-Australian Archipelago), where all 10 species of anemone and up to nine fish species can be found in a single locale (Fautin and Allen, 1997; Elliott and Mariscal, 2001). Anemone diversity declines steadily outward from this area, while fish diversity outside of this area is lower and composed largely of regional endemics (Fautin and Allen, 1997).

The unique symbiotic lifestyle and unusual mating system of Amphiprioninae make their evolutionary history of considerable interest to a wide range of biologists, but no robust phylogeny has been attained. Multiple, conflicting hypotheses have been put forth regarding the diversification of this group, beginning with a morphological analysis by Allen (1972; Figure 3.1). Morphological characters differentiating these species are mostly distinct, but can be quite subtle (Allen, 1975), and conspicuous characters like color pattern can be variable within species (Fautin and Allen, 1997). Several molecular phylogenies have since been presented (Elliott et al., 1999; Santini and Polacco, 2006), using 1-3 mitochondrial loci and a subset of the species. Family-level studies have added better marker coverage, but for fewer taxa (Cooper et al., 2009; Tang, 2001). Two recent studies reanalyzed data from previous work (Frédérich et al., 2013; Litsios et al., 2012), and these phylogenies were largely consistent with respect to the sister relationships

of recently diverged species. However, substantial inconsistencies with respect to the identity of the earliest diverging species and of the arrangement of the morphological species complexes preclude accurate estimates of character evolution within the group.

The most recent phylogeny was constructed using sequence data from three loci from the nuclear genome (nDNA) and six loci from the mitochondrial genome (nDNA) (Litsios et al., 2012). A total of 106 sequences from 26 species were obtained from GenBank (45% of a complete data matrix). Seven of the species were represented by sequences for 8-9 loci, while the remaining 19 species by 4 or fewer loci. No locus was sampled for all species, and six of the loci were sampled for 12 or fewer species. It included sequence data from an individual that is verifiably misidentified (*A. sebae*, GenBank accession number FJ582825), as well as other sequences from that study (see Appendix).

Hence, despite several efforts to reconstruct the evolutionary history of the group from DNA sequence data, important nodes in the phylogeny remained unresolved by recent studies. Here we present a more robust molecular phylogeny of the Amphiprioninae, constructed using a multi-locus analysis of the most taxonomically and genetically exhaustive data set to date.

## Methods

### Taxon Sampling

We obtained samples from 24 species of Amphiprioninae, as well as samples from multiple geographic locales of wide-ranging species (Table 3.4), and confirmed the identity of the species either via photograph, voucher, or taxonomic expert. Wherever possible, we used sequences of the different loci from a single individual. For the species *A. nigripes* and *A. tricinctus*, we were only able to obtain DNA material; however, these species are unmistakable both in morphology and geographic origin, and samples were provided by an expert on the group (J. K. Elliott). For outgroup-rooted trees, we chose *Pomacentrus pavo* based on its placement in published phylogenies (Cooper et al., 2009; Frédérich et al., 2013) and availability of sequence data.

### Laboratory Methods

Genomic DNA was extracted using a chloroform protocol (Sambrook et al., 1989). We amplified three mitochondrial loci: 16S ribosomal DNA (16S), cytochrome b (CYTB), cytochrome c oxidase subunit I (CO1); and four nuclear loci: recombination activating gene 2 (RAG2), rhodopsin (RHOD), the first intron of 40S ribosomal protein S7 (S71), and its second intron (S72). Target loci were amplified using the polymerase chain reaction (PCR) primers and parameters outlined in Tables 3.1 and 3.2. Reactions took place in 13$\mu$L volumes containing 11.25$\mu$L ThermoPrime ReddyMix (Thermo Scientific), 0.625$\mu$L each primer, and 0.5$\mu$L extracted DNA. PCR products were sequenced

on an Applied Biosystems capillary 3730xl DNA Analyzer.

## Alignment

We resolved nucleotide ambiguities and removed primer sequences manually in Geneious Pro 5.6.5 (Biomatters Ltd.). The S7 intron sequences were marked by several regions of A-T repeats of variable length. Some individuals were heterozygous for this sequence length polymorphism. In these cases, we sequenced PCR product using the reverse primer, and concatenated the two sequences. The region of repeats was removed from aligned sequences in order to reduce potential noise due to intraspecific variation in this region.

Sequences were aligned with the MAFFT alignment program v7.017 (Katoh et al., 2002) using the L-INS-i iterative refinement algorithm with a 200PAM / k=2 scoring matrix, and set a gap open and offset penalty of 1.53 and 0.123, respectively. Variation in sequence length due to insertion or deletion events (indels) can cause gaps in an alignment for which theoretical models of evolution are poorly developed. The alignment process introduced no gaps into CO1, CYTB, RAG2, or RHOD. Gaps from 1-12 bp were inserted into 16S, S71, and S72. We analyzed the full data set with and without the gaps removed and found no qualitative differences, so further analyses were conducted with gaps included. We did not make any additional adjustments to the alignments.

We inferred the models and parameters most likely to have generated each locus's alignment using jModelTest (v2.1.4) (Darriba et al., 2012), and chose models

and parameters based on Bayesian information criterion (BIC) scores (Table 3.3).

## Phylogenetic Reconstruction

### Gene Trees

We confirmed individuals within a species were more similar to one another than to heterospecifics by reconstructing phylogenies of individual loci using PhyML 3.0 (Guindon et al., 2010) on a server hosted by LIRMM (Montpellier) at http://www.atgc-montpellier.fr/phyml/. These alignments contained multiple sequences per species wherever possible (Table 3.4). We employed the models of nucleotide substitution specified by jModelTest for each gene, except for CYTB and S72, whose specified models were not available in PhyML. In these cases, we used the next more complex model, the general time reversible model (GTR). Start trees were constructed using BIONJ, the SPR algorithm was used to search tree topologies, and both topology and branch lengths were optimized. Node support was evaluated on the maximum likelihood (ML) tree by analyzing 100 non-parametric bootstrap replicates of this tree. For comparison, also we concatenated the species tree alignments and analyzed them under these conditions.

### Species Trees

We conducted a multi-locus phylogenetic analysis of the species using a Metropolis-coupled Markov chain Monte Carlo (MCMCMC) approach implemented in MrBayes v3.2.2 (Ronquist et al., 2012). Alignments were concatenated and partitioned by locus.

We assigned a GTR+I+G substitution model to all loci because Bayesian phylogenetic inference has been demonstrated to be robust to model overspecification (Huelsenbeck and Rannala, 2004). That is, even if the true process of nucleotide substitution is simple (e.g. JC or HKY), a more complex model will recover posterior probabilities that accurately reflect the probability that the tree is correct. As confirmation, We also ran an analysis with models specified by jModelTest. If a selected model was not available in MrBayes (e.g. TrN or TPM), we chose the next best model according to BIC score (e.g. HKY). In all analyses, the proportion of invariable sites, the shape of the gamma parameter, the character state frequencies, and the substitution rate were allowed to vary among partitions.

We performed two independent runs of the MCMCMC analysis, each with four parallel chains with a temperature parameter of 0.1. Chains were run for 10 million generations, and sampled every 1000 generations. We ensured the chains had adequately sampled the parameters by confirming the runs had established a minimum effective sample size (ESS) greater than 1500 and a potential scale reduction factor (PSRF) of 1 for all parameters.

## Results and Discussion

We generated DNA sequence data from three mitochondrial loci and four nuclear loci, from a total of 145 individuals, comprising 24 species. For species tree reconstruction, our data matrix was 84% full, all but three species were represented by

more than half the loci, and 15 species had no missing data. In total, 4547 base pairs were included in the analysis. We provide the first published molecular sequence data for *Amphiprion tricinctus*, the first multi-locus phylogenetic treatment of four species (*A. barberi*, *A. ephippium*, *A. latezonatus*, and *A. latifasciatus*), and the first nDNA data for nine species (*A. akallopisos*, *A. bicinctus*, *A. chrysopterus*, *A. ephippium*, *A. leucokranos*, *A. mccullochi*, *A. nigripes*, *A. percula*, *A. polymnus*). We recovered substantially higher support values for internal nodes than has previously been obtained for this group (Figure 3.4).

Gene phylogenies revealed reciprocal monophyly among species in most cases. There were some notable exceptions. *Amphiprion leucokranos* was indistinguishable from *A. chrysopterus* in mitochondrial gene trees, but rooted within the *A. sandaracinos* clade for nuclear gene trees (Figures 3.6-3.12). In analyses that included multiple individuals of *A. clarkii* from multiple geographic locations, *A. tricinctus* rendered *A. clarkii* paraphyletic, suggesting that *A. clarkii* may be composed of multiple spatially isolated cryptic species. For species spanning the Coral Triangle region, the analysis of multiple individuals from across this region often exhibit substantial divergence between locales, emphasizing the importance of barriers to gene flow there. Most relationships among closely related species were recovered in each gene tree (e.g. species of the "Tomato" complex including *A. melanopus* and *A. frenatus*). The RHOD analysis resulted in the least well-resolved tree with respect to node support; however, conspecifics continued to cluster together, and many clades found in other gene trees were recovered. Nearly all gene trees supported the validity of the recently described *A. barberi*.

The majority rule consensus species trees produced using both GTR+I+G and simpler models were nearly identical in both topology and branch lengths, except that the node leading to *A. bicinctus* was reduced to a polytomy in the tree generated using a simpler model. Support for internal nodes on the consensus tree was generally high (Figure 3.2). The majority of nodes had posterior probability values of 1. Only three nodes had posterior probability values below 0.7: the nodes placing *A. rubrocinctus*, *A. barberi*, and the clade *A. akindynos* + *A. mccullochi*. *Amphiprion rubrocinctus*, *A. akindynos*, and *A. mccullochi* were represented by two, four, and five sequences respectively, and data for *A. rubrocinctus* came only from mtDNA.

We present robust phylogenies of both individual genes and species, and we attribute their high levels of support to increased marker coverage and careful taxon sampling. Nevertheless, support for the nodes within *Amphiprion* is relatively low compared to phylogenies of similar groups, such as the Stegastinae (Frédérich et al., 2013). The Amphiprioninae are both a recently diverged and rapidly diversifying group (Frédérich et al., 2013), phenomena that are known to contribute to phylogenetic uncertainty. Resolution may be improved through the incorporation of data from additional markers, particularly from the nuclear genome. Data from massively parallel sequencing techniques such as sequence capture of ultra-conserved elements (Faircloth et al., 2012), would be ideally suited to this problem.

The genus *Premnas* is embedded within the genus *Amphiprion*, thus rendering *Amphiprion* polyphyletic. *Premnas biaculeatus* was first described as *Chaetodon biaculeatus* (Bloch 1790), Cuvier erected the genus *Premnas* in 1816, while *Amphiprion*

was designated by Bloch and Schneider in 1801 (Eschmeyer, 2014). The International Code on Zoological Nomenclature grants priority to the oldest available name given to a taxon (Chapter 6, Article 23.1), in this case *Amphiprion*. In Allen's (1972) definitive treatment of the Amphiprioninae, he made this change, but later resurrected the genus *Premnas* in the second edition of that work (Allen, 1975). These changes were made in the absence of molecular evidence; however, all subsequent molecular treatments of the group have included *Premnas* within *Amphiprion*. Because of the historical precedence of the name, along with irrefutable evidence that the continued use of *Premnas* results in a polyphyletic *Amphiprion*, we propose the synonymy of *Premnas* (Cuvier 1816) with *Amphiprion* (Bloch and Schneider 1801).

Immediately following the well-supported *A. chrysopterus* split, there is extremely rapid diversification of several groups of species, which are for the most part morphologically similar within groups and distinct among groups. These include the "Skunk" clade (*A. akallopisos*, *A. perideraion*, and *A. sandaracinos*), the "Tomato" clade (*A. barberi*, *A. ephippium*, *A. frenatus*, *A. melanopus*, and *A. rubrocinctus*), an Australian clade (*A. akindynos* and *A. mccullochi*), and a Western Indian Ocean clade (*A. allardi*, *A. latifasciatus*, *A. bicinctus*, *A. chagosensis*, and *A. nigripes*). *Amphiprion polymnus* also arose from this rapid diversification event, a species that is associated with anemones usually found in muddy or sandy substrates.

Perhaps the most striking pattern in the tree is the clade containing only species restricted to the Western Indian Ocean basin west of 90 degrees longitude (*A. bicinctus*, *A. allardi*, *A. latifasciatus*, *A. chagosensis*, and *A. nigripes*), which form the

bulk of the species found in that area. Outside of this clade, the only other species found in the region are *A. clarkii* and *A. akallopisos*, both which are also found in the Coral Triangle. None of these species co-occur, suggesting an ancestral invasion of the region followed by subsequent diversification driven by spatial isolation of populations. While we did not include data from *A. chrysogaster* (endemic to Reunion and Mauritius), *A. fuscocaudatus* (endemic to the Seychelles), or *A. omanensis* (endemic to Oman) in our analysis, other analyses have these species in this clade, strengthening this notion.

The relationship between *A. clarkii* and *A. tricinctus* is an intriguing recent speciation event. *Amphiprion clarkii* has the largest range of any Amphiprioninae (from the Persian Gulf to Fiji), while *A. tricinctus* has among the smallest, occupying only the Marshall Islands archipelago. This suggests that populations of *A. clarkii* have been able to maintain sufficient genetic connectivity to prevent speciation across biogeographic barriers known to have promoted speciation in other lineages within this group (Timm 2008). Yet the amount of gene flow in the ancestral species between the Marshall Islands and elsewhere was restricted. Two other species can be found in the Marshall Islands (*A. melanopus* and *A. chrysopterus*), though *A. clarkii* is absent (Fautin and Allen, 1997). Endemism resulting from peripheral isolation is known to occur across lineages in other archipelagos (Drew et al., 2008; Hodge et al., 2012). Populations of *A. melanopus* and *A. chrysopterus* in the Marshall Islands may be similarly divergent, and warrant further investigation.

The sister relationship between *A. chrysopterus* and *A. leucokranos* is inconsistent with morphology, but unsurprising given the hypothesized hybrid origin of *A.*

*leucokranos*. *Amphiprion sandaracinos* and *A. chrysopterus* are one of the very few instances of species that regularly form mixed-species groups on the same host individual (Elliott 2001). This phenomenon occurs where their ranges overlap along the northern coast of New Guinea. Because family groups on anemones are dominated by the largest female, and because *A. chrysopterus* is much larger than the diminutive *A. sandaracinos*, *A. chrysopterus* is expected to be the maternal species of interspecific hybrids. The assignment of our *A. leucokranos* mitochondrial sequences to *A. chrysopterus* is consistent with that hypothesis.

The evolutionary history of the species endemic to the seas off of eastern Australia (*A. akindynos*, *A. mccullochi*, and *A. latezonatus*) has been difficult to resolve. *Amphiprion mccullochi* is restricted to the isolated Lord Howe Island and nearby reefs, while *A. akindynos* is found in New Caledonia and the Great Barrier Reef. The two species are morphologically distinct, much more so than other more distantly related species (e.g. the species of the "Tomato" complex), making it unlikely the two species merely represent the result of an allopatric speciation event. Two individuals of *A. akindynos* have been reported from Lord Howe Island, a presumably extremely rare event (Crean et al., 2010). Further, gene flow between the species occurs, suggesting multiple hybridization events (van der Meer et al., 2012). Similarly puzzling is the role of *A. latezonatus*, found both on the eastern coast of Australia and at Lord Howe Island, sympatric with both *A. akindynos* and *A. mccullochi*. An in depth consideration of the relationship between these three species using a broad suite of molecular data is warranted.

In addition to the molecular evidence for hybridization provided by (van der Meer et al., 2012), *Amphiprion* species are known to interbreed regularly in captivity, and implicated in the creation of aberrant morphs in the wild (Carlson, 1996; Fautin and Allen, 1997), including crosses of species that are not closely related (e.g. *A. frenatus* x *P. biaculeatus*, collected in the Philippines https://reefbuilders.com/2012/11/28/wild-tomato-maroon-clownfish/). It is likely that such events have occurred throughout the history of the group, obfuscating the true evolutionary history of the group using traditional phylogenetic approaches. Approaches that explicitly accommodate such reticulate evolutionary patterns may help to resolve a more accurate evolutionary history.

The species most directly descended from the most recent common ancestor (*A. ocellaris*, *A. percula*, and *P. biaculeatus*) are all restricted to the Coral Triangle. Though species distributions are known to not be stationary over long periods of time, changes are unlikely to be both fast and drastic in species with small adult ranges and short dispersal phases. Given the anemonefishes originated approximately 10-22 million years ago, have among the shortest dispersal phases of any reef fish, and have limited realized dispersal (Jones et al., 2005), it is unlikely these species' ranges could have moved substantially in this time. Instead, our results support the notion that the transition to a symbiotic lifestyle was brought about in the exceptionally diverse Coral Triangle, perhaps a product of high levels of competition for niche space. Subsequently, lineages diverged as the group expanded outwards from the Coral Triangle. The invasion into the Western Indian Ocean led to the origin of eight new species, nearly one third of the family diversity. Likewise, expansion of lineages from the Coral Triangle into the

Pacific basin may have led to the origin of *A. chrysopterus* and *A. tricinctus*, along with the recently described, morphologically cryptic *A. barberi*, and *A. pacificus* (Allen et al., 2010, 2008). The only other clade of similar diversity to the Western Indian Ocean clade is the "Tomato" complex containing *A. ephippium*. As a group, these species span a range smaller than that of *A. clarkii*, most of which is contained in the center of Amphiprioninae diversity.

Overall, our findings support a pattern of diversity originating in the Coral Triangle with speciation events occurring at the periphery of this region. This pattern is evident with respect to the entire group, but also to more recent events such as the radiation of the Western Indian Ocean clade and the origin of *A. tricinctus* and *A. barberi*. Phylogenetic analyses of other groups of Indo-Pacific coral reef organisms have found similar patterns (Hodge et al., 2012; Malay and Paulay, 2010; Drew et al., 2008). As suggested by previous authors, following the divergence of populations at the margins of the Coral Triangle, peripatric species may subsequently expand their range into the region, further enhancing the high diversity found there.

## Acknowledgements

We are extremely grateful to the following individuals for contributing samples: Peter J. Hundt, Michelle Gaither, Bruce Carlson, Philippe Borsa, Serge Planes, Elizabeth Jones Sbrocco, Kathryn Scott, Joel Elliott, Thierry Mulochau, Sue Morrison, Laurie Sorenson, and especially Martin van der Meer. We acknowledge the finan-

cial support of the Lerner-Gray Fund for Marine Research, the Earl and Ethel Myers Oceanographic and Marine Biology Trust, the UCSC Ecology and Evolutionary Biology Department, and the American Society of Ichthyologists and Herpetologists.

# Appendix

There exists at least at least one instance where vouchers were misidentified and resultant sequences deposited to GenBank under the incorrect taxon name. Photographs of the vouchers (found at http://www.marinebarcoding.org/species/region/1/id/10537) for GenBank sequences FJ582817-FJ582830 confirm that these individuals are *Amphiprion clarkii*, not *A. sebae* as labeled in GenBank. This error has been reported but not corrected. These sequences have been used in phylogenetic studies (Frédérich et al., 2013; Litsios et al., 2012), where they are concatenated with sequences that may or may not come from the same species. Such errors can contribute to poor resolution of phylogenetic reconstruction.

## Tables and Figures

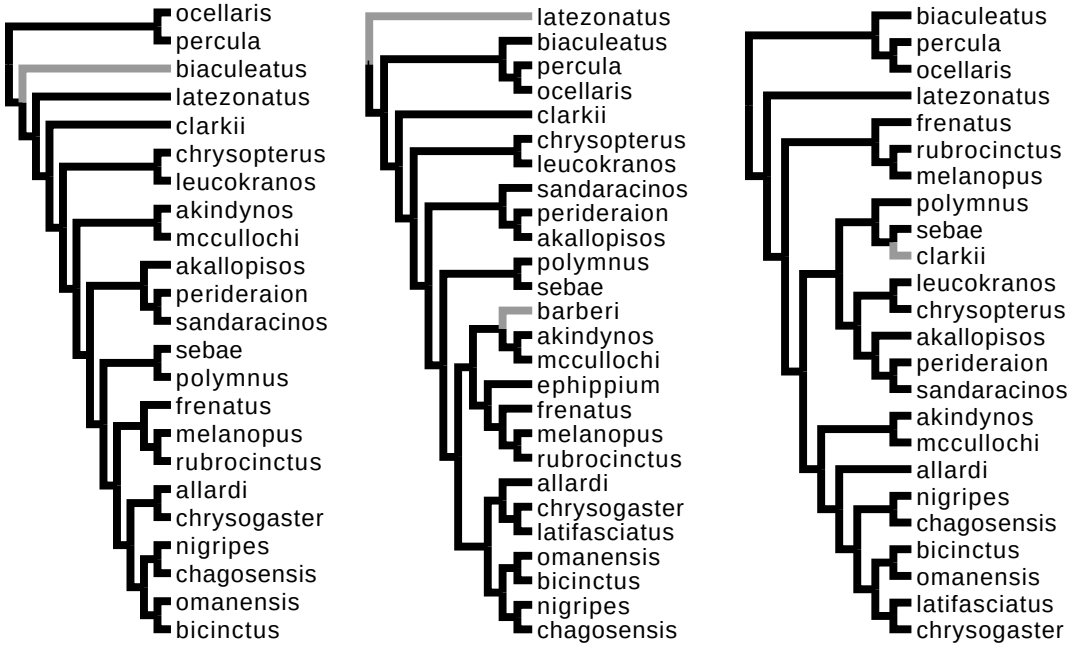



Figure 3.1: Trees representing topology supported by previous studies. (A) Santini and Polacco 2006, (B) Frederich et al 2013, (C) Litsios et al 2013. Tips with particularly problematic placement are highlighted in grey.

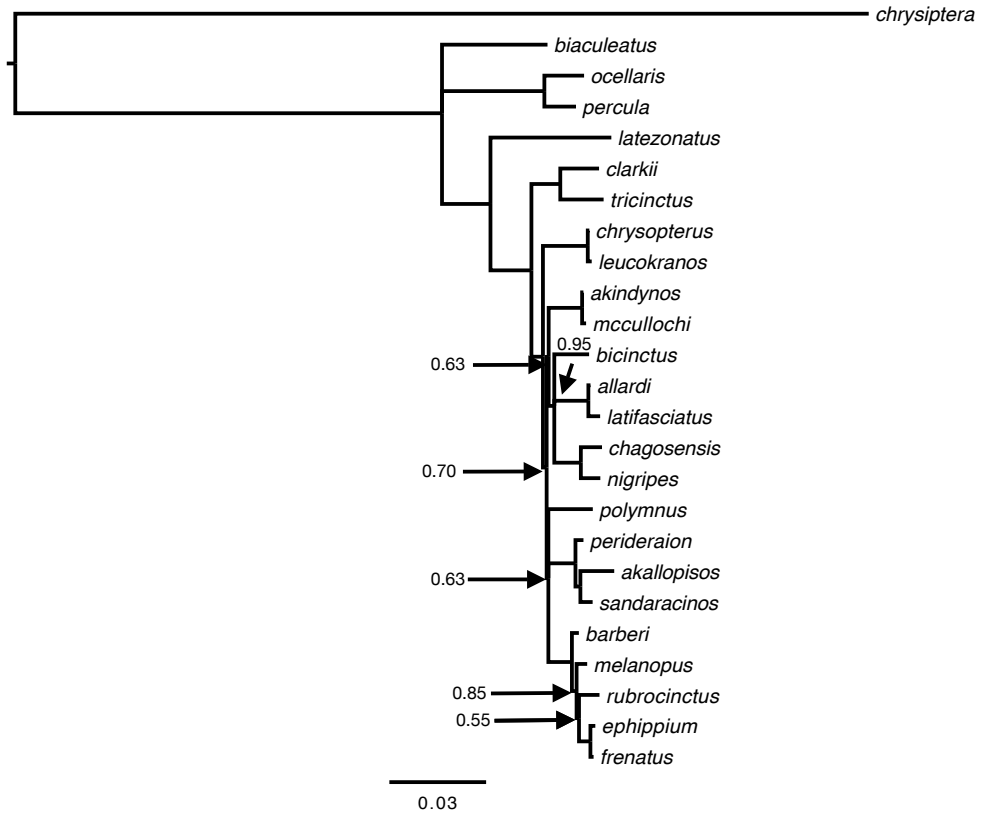



Figure 3.2: 50% majority rule consensus tree of *Amphiprion* species inferred from Bayesian analysis. Branch lengths are expected substitutions per site, and the values at the nodes represent approximations of the posterior probability of the bipartitions. Unlabeled nodes have support $\geq 0.99$.

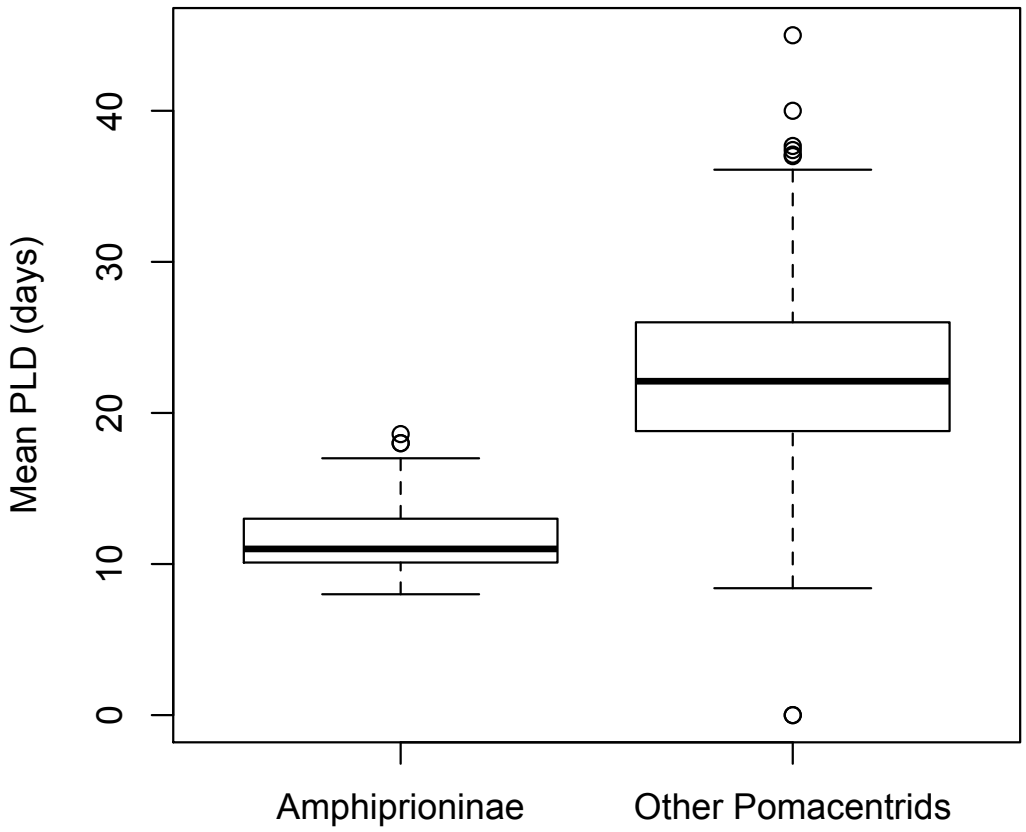



Figure 3.3: Plot of the mean pelagic larval duration (PLD) in days for species of Amphiprioninae (n = 21) and species of other Pomacentrids (n = 309). Data from Luiz et al. (2013). Boxes encompass the interquartile range (IQR), lines extend to 1.5 IQR, and circles represent outliers.

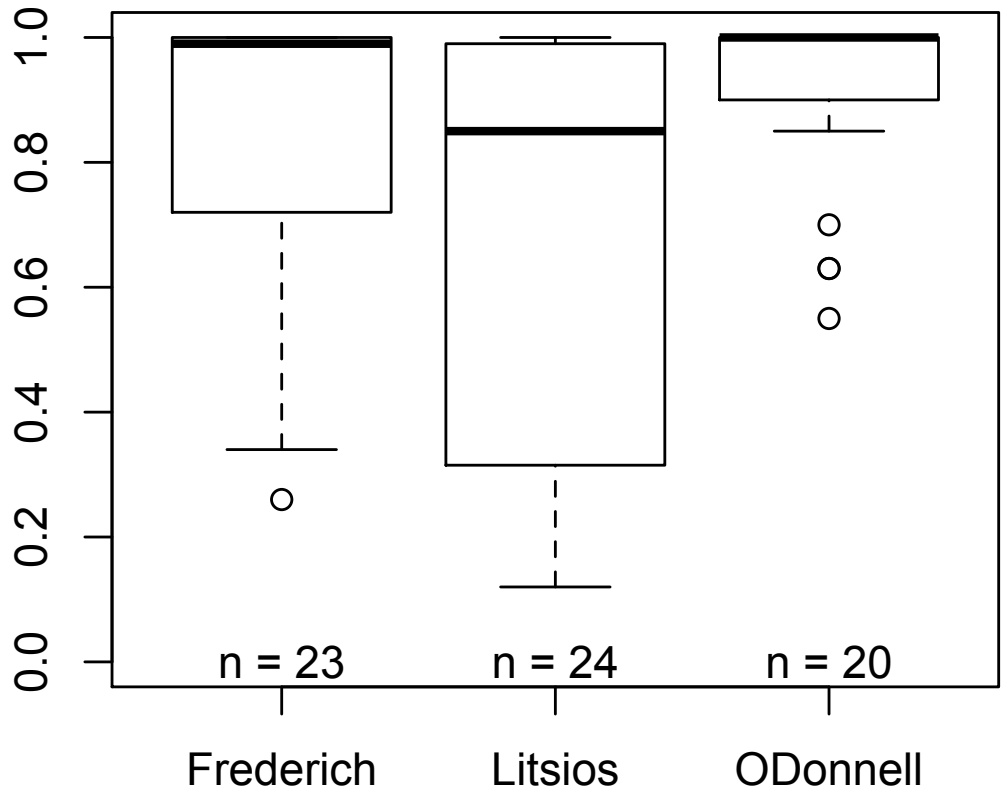



Figure 3.4: Comparison of support values (posterior probability) for internal nodes recovered by two previous species level phylogenies (Frédérich et al., 2013; Litsios et al., 2012) and the present study. Boxes encompass the interquartile range (IQR), lines extend to 1.5 IQR, and circles represent outliers.

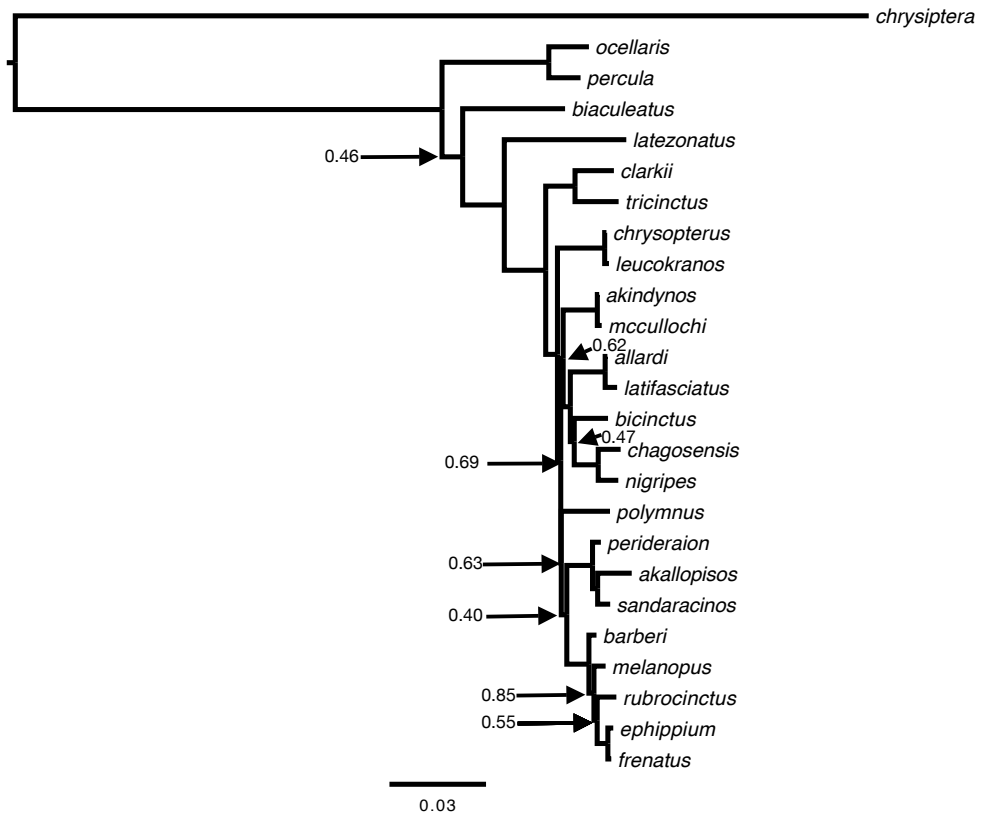



Figure 3.5: Maximum clade credibility tree of *Amphiprion* species inferred from Bayesian analysis.

Table 3.1: PCR primers used in this study.

| Genome | Locus | Primer Name | Direction | Primer Sequence | Reference |
|---|---|---|---|---|---|
| mtDNA | 16S | 16SAR | Forward | CGCCTGTTTATCAAAAACAT | Palumbi 2002 |
| | | 16SBR | Reverse | CCGGTCTGAACTCAGATCACGT | |
| mtDNA | CO1 | CO1_Vr1dT1 | Forward | TAGACTTCTGGGTGGCCRAARAAYCA | Ivanova 2007 |
| | | CO1_VF2_T1 | Reverse | TCAACCAACCACAAAGACATTGGCAC | |
| mtDNA | CYTB | GLUDG-L | Forward | TGACTTGAARAACCAYCGTTG | Palumbi 2002 |
| | | CB3-H | Reverse | GGCAAATAGGAARTATCATTC | |
| nDNA | RAG2 | RAG2_F1 | Forward | GAGGGCCATCTCCTTCTCCAA | Cooper 2009 |
| | | RAG2_R2 | Reverse | GTCTGTAGAGTCTCACAGGAGAGCA | |
| nDNA | RHOD | Rod-F2X | Forward | AGCAACTTCCGCTTCGGCGAGAA | Sevilla 2007 |
| | | Rod-R4n | Reverse | GGAACTGCTTGTTCATGCAGATGTAGAT | |
| nDNA | S71 | S7-F | Forward | TGGCCTCTTCCTTGGCCGTC | Chow 1999 |
| | | S7-R | Reverse | AACTCGTCTGGCTTTTCGCC | |
| nDNA | S72 | S7RPEX2F | Forward | AGCGCCAAAATAGTGAAGCC | Chow 1999 |
| | | S7RPEX3R | Reverse | GCCTTCAGGTCAGAGTTCAT | |

Table 3.2: PCR conditions used in this study. Temperatures are given in degrees Celsius.

| | Initial | | Denaturation | | Annealing | | Extension | | | Final Extension | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Locus | Temp | Time | Temp | Time | Temp | Time | Temp | Time | N cycles | Temp | Time |
| 16S | 94 | 5:00 | 94 | 0:30 | 54 | 0:30 | 72 | 0:30 | 35 | 72 | 7:00 |
| CO1 | 94 | 4:00 | 94 | 0:45 | 54 | 0:45 | 72 | 1:00 | 35 | 72 | 1:00 |
| CYTB | 94 | 5:00 | 94 | 0:45 | 45 | 0:45 | 72 | 0:45 | 35 | 72 | 7:00 |
| RAG2 | 94 | 3:00 | 94 | 0:45 | 54 | 0:45 | 72 | 0:45 | 35 | 72 | 5:00 |
| RHOD | 95 | 7:00 | 94 | 0:30 | 56 | 0:30 | 72 | 0:30 | 40 | 72 | 7:00 |
| S71 | 94 | 3:00 | 94 | 0:45 | 54 | 0:45 | 72 | 0:45 | 35 | 72 | 5:00 |
| S72 | 94 | 3:00 | 94 | 0:45 | 54 | 0:45 | 72 | 0:45 | 35 | 72 | 5:00 |

Table 3.3: Substitution models selected for each locus based on BIC scores in jModelTest2. N Taxa = number of taxa for individual gene tree construction; -lnL = negative log likelihood; K = number of estimated parameters.

| Genome | Locus | Length | N Taxa | Model | -lnL | K |
|---|---|---|---|---|---|---|
| mtDNA | 16S | 538 | 101 | K80+I+G | 1583.15 | 203 |
| mtDNA | CO1 | 618 | 80 | TrN+I+G | 2466.03 | 165 |
| mtDNA | CYTB | 764 | 86 | TPM2uf+I+G | 3323.24 | 177 |
| nDNA | RAG2 | 755 | 86 | K80+I | 1486.58 | 172 |
| nDNA | RHOD | 410 | 95 | K80 | 716.66 | 189 |
| nDNA | S71 | 715 | 70 | JC | 1792.15 | 138 |
| nDNA | S72 | 721 | 72 | TPM3uf+G | 1836.96 | 148 |

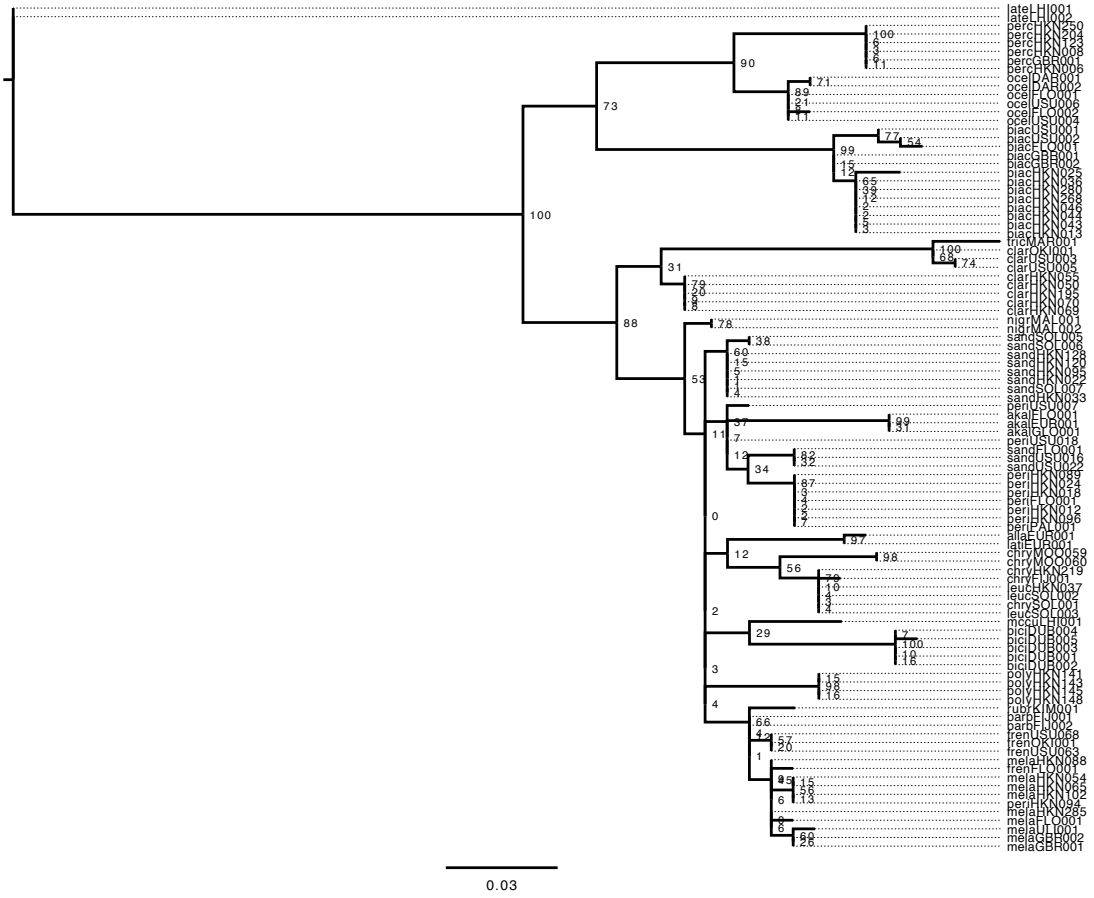



Figure 3.6: Maximum likelihood tree for 16S. Numbers at nodes indicate the number of bootstrap replicates out of 100 in which nodes were recovered, and branch lengths are expected substitutions per site.

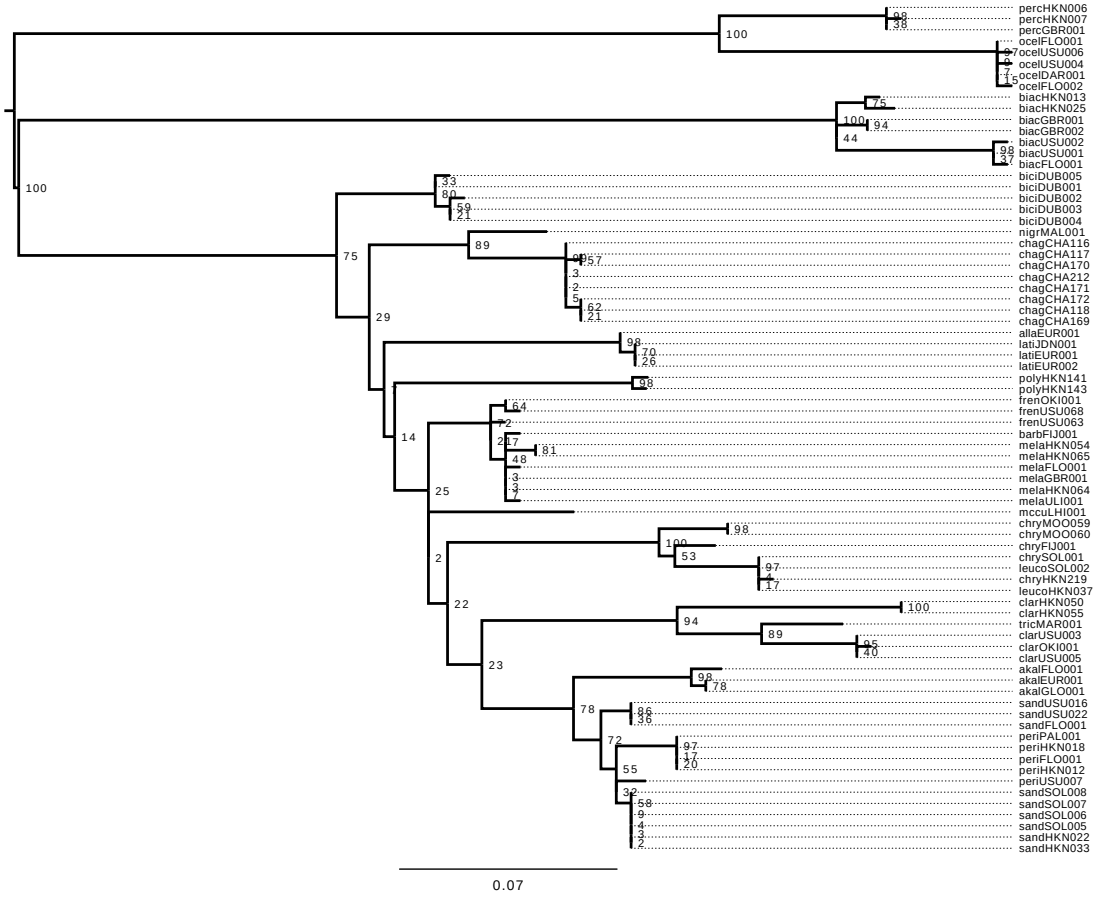



Figure 3.7: Maximum likelihood tree for CO1. Numbers at nodes indicate the number of bootstrap replicates out of 100 in which nodes were recovered, and branch lengths are expected substitutions per site.


Figure 3.8: Maximum likelihood tree for CYTB. Numbers at nodes indicate the number of bootstrap replicates out of 100 in which nodes were recovered, and branch lengths are expected substitutions per site.

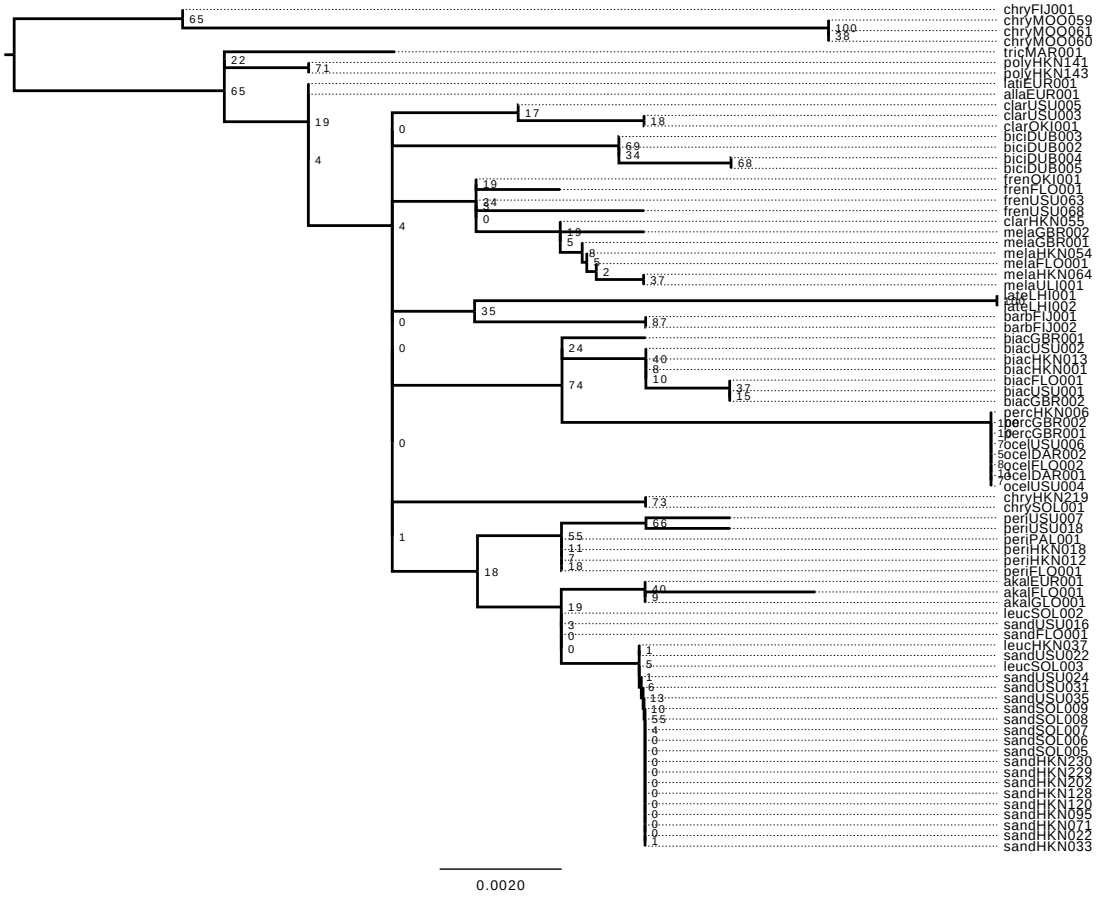



Figure 3.9: Maximum likelihood tree for RAG2. Numbers at nodes indicate the number of bootstrap replicates out of 100 in which nodes were recovered, and branch lengths are expected substitutions per site.

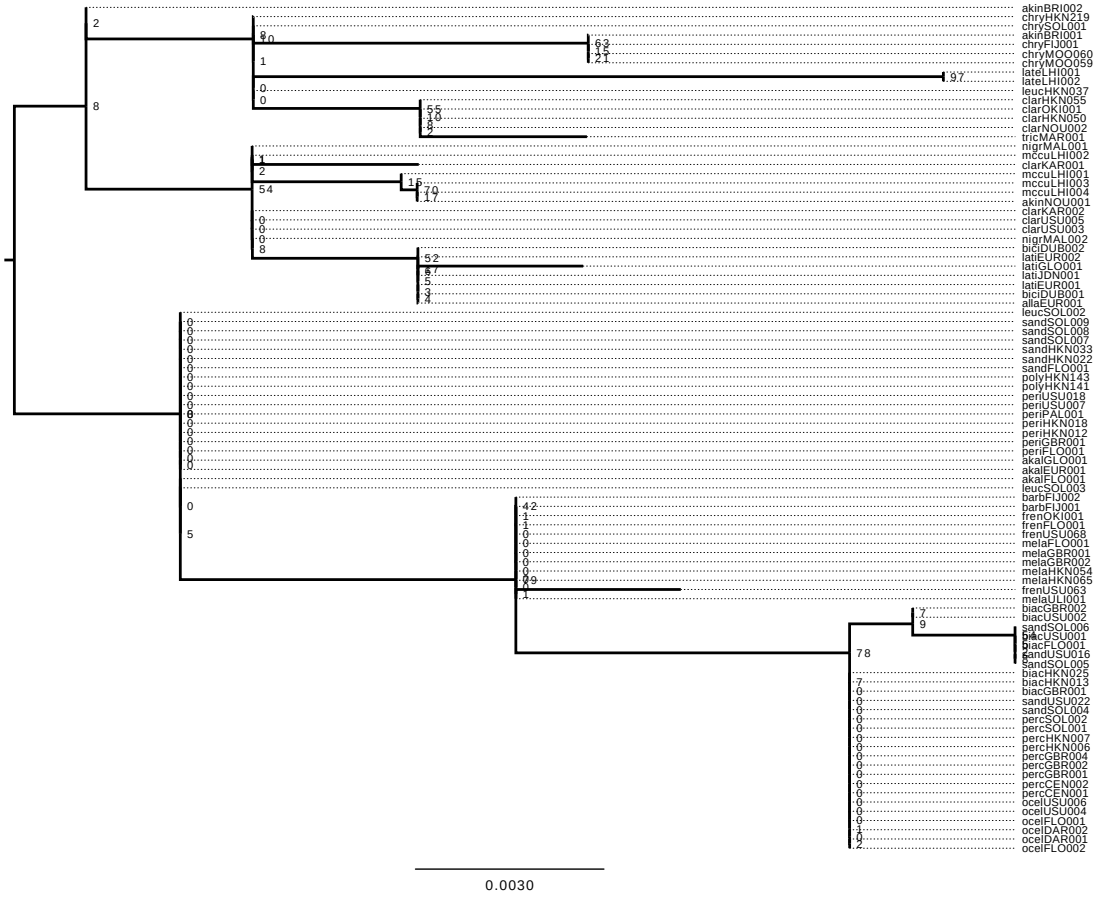



Figure 3.10: Maximum likelihood tree for RHOD. Numbers at nodes indicate the number of bootstrap replicates out of 100 in which nodes were recovered, and branch lengths are expected substitutions per site.

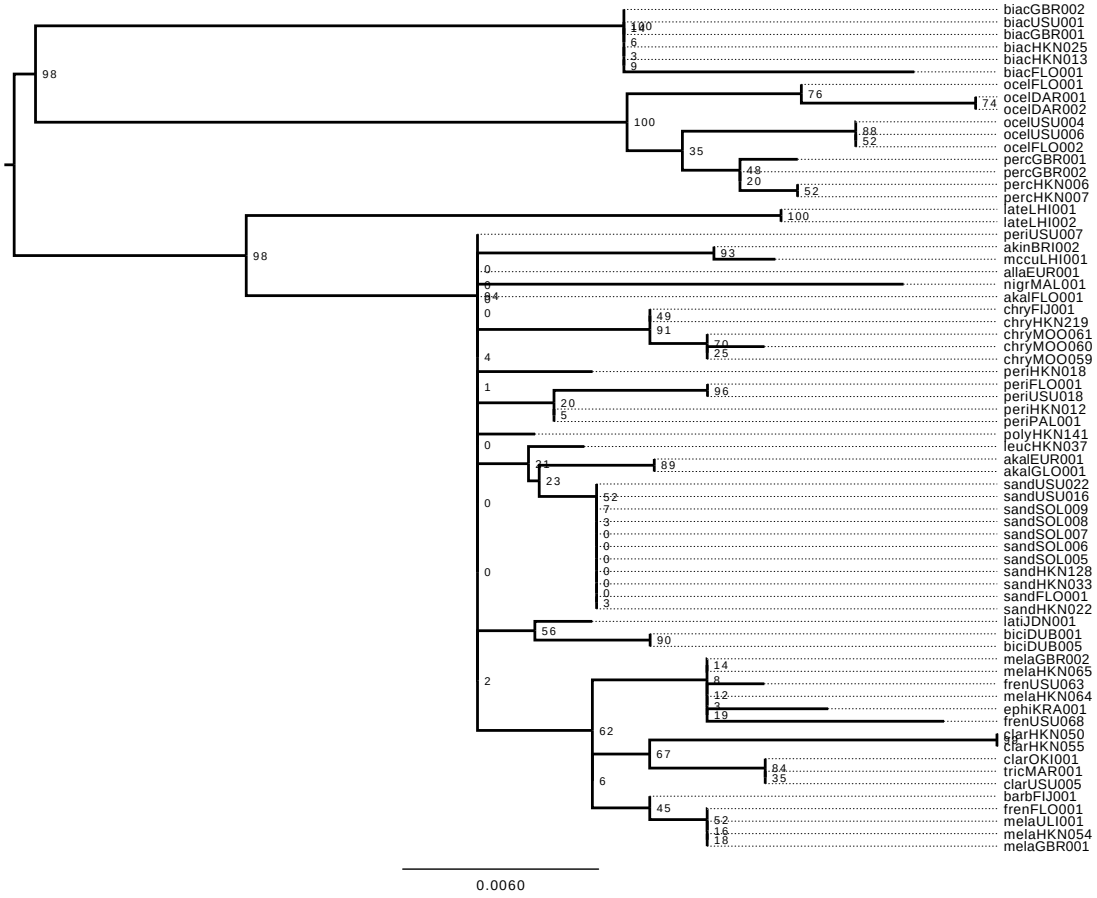



Figure 3.11: Maximum likelihood tree for S7-1. Numbers at nodes indicate the number of bootstrap replicates out of 100 in which nodes were recovered, and branch lengths are expected substitutions per site.

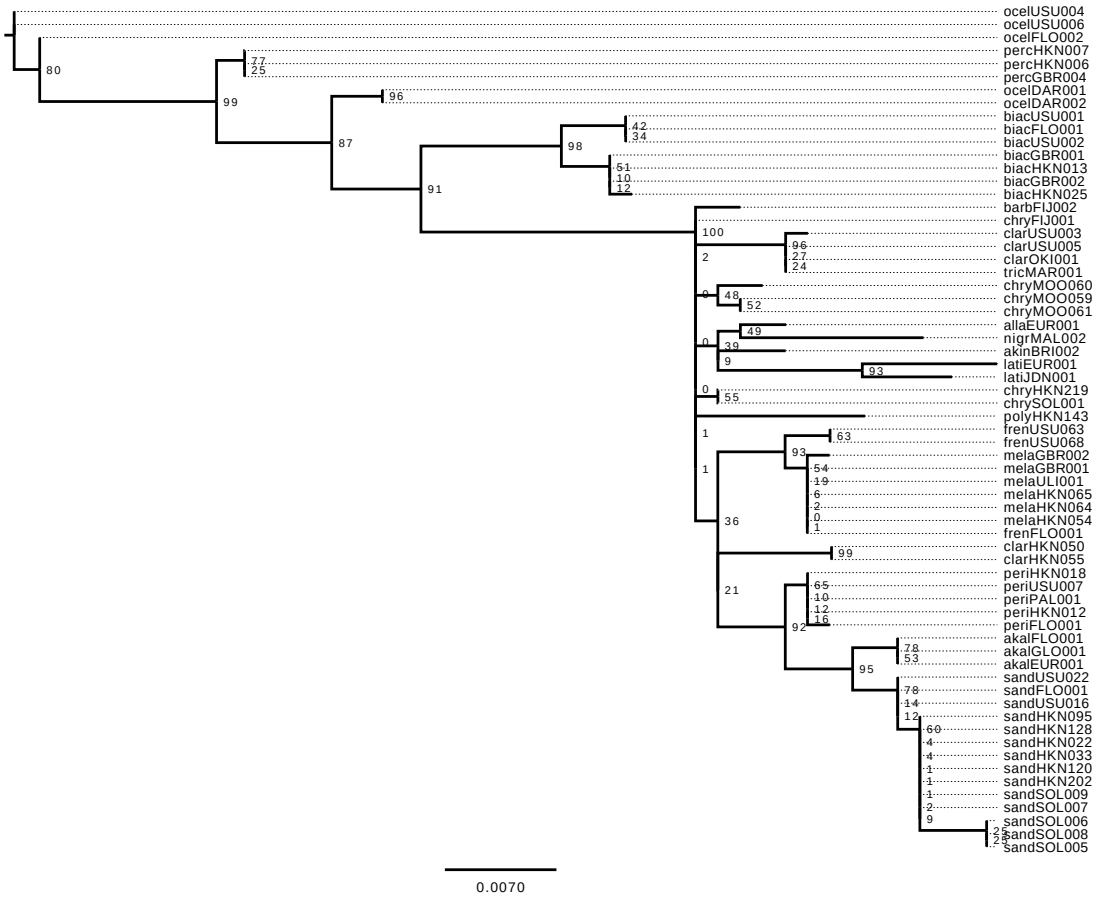



Figure 3.12: Maximum likelihood tree for S7-2. Numbers at nodes indicate the number of bootstrap replicates out of 100 in which nodes were recovered, and branch lengths are expected substitutions per site.

Table 3.4: Species and sampling locales for taxa used in this study. Sequences used in gene tree reconstruction are indicated by a 1, an asterisk indicates a sequence used in species tree reconstruction, and a 0 indicates no sequence.

| Species | Origin | Sample | 16S | CO1 | CYTB | RAG2 | RHOD | S71 | S72 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| *A. akindynos* | Great Barrier Reef | AMAIBRI001 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. akindynos* | Great Barrier Reef | AMAIBRI002 | 0 | 0 | * | 0 | * | * | * | 4 |
| *A. akindynos* | New Caledonia | AMAINOU001 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. akallopisos* | Mozambique Channel | AMAKEUR001 | * | * | 0 | * | * | * | * | 6 |
| *A. akallopisos* | Flores | AMAKFLO001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. akallopisos* | Mozambique Channel | AMAKGLO001 | 1 | 1 | * | 1 | 1 | 1 | 1 | 7 |
| *A. allardi* | Mozambique Channel | AMALEUR001 | * | * | * | * | * | * | * | 7 |
| *A. barberi* | Fiji | AMBAFIJ001 | * | * | * | * | * | * | 0 | 6 |
| *A. barberi* | Fiji | AMBAFIJ002 | 1 | 0 | 1 | 1 | 1 | 0 | * | 5 |
| *A. bicinctus* | Red Sea | AMBIDUB001 | * | * | 0 | 0 | * | * | 0 | 4 |
| *A. bicinctus* | Red Sea | AMBIDUB002 | 1 | 1 | 0 | * | 1 | 0 | 0 | 4 |
| *A. bicinctus* | Red Sea | AMBIDUB003 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 3 |
| *A. bicinctus* | Red Sea | AMBIDUB004 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 3 |
| *A. bicinctus* | Red Sea | AMBIDUB005 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 4 |
| *A. chagosensis* | Chagos | AMCHCHA116 | 0 | * | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. chagosensis* | Chagos | AMCHCHA117 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. chagosensis* | Chagos | AMCHCHA118 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. chagosensis* | Chagos | AMCHCHA169 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. chagosensis* | Chagos | AMCHCHA170 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. chagosensis* | Chagos | AMCHCHA171 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. chagosensis* | Chagos | AMCHCHA172 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. chagosensis* | Chagos | AMCHCHA212 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. clarkii* | New Guinea | AMCLHKN042 | 0 | 0 | * | 0 | 0 | 0 | 0 | 1 |

Continued on next page

Table 3.4 – Continued from previous page

| Species | Origin | Sample | 16S | CO1 | CYTB | RAG2 | RHOD | S71 | S72 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| *A. clarkii* | New Guinea | AMCLHKN050 | * | * | 1 | 0 | * | * | * | 6 |
| *A. clarkii* | New Guinea | AMCLHKN052 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| *A. clarkii* | New Guinea | AMCLHKN055 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. clarkii* | New Guinea | AMCLHKN069 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *A. clarkii* | New Guinea | AMCLHKN070 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *A. clarkii* | New Guinea | AMCLHKN195 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. clarkii* | New Guinea | AMCLHKN223 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| *A. clarkii* | Java | AMCLKAR001 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. clarkii* | Java | AMCLKAR002 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. clarkii* | New Caledonia | AMCLNOU002 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. clarkii* | Japan | AMCLOKI001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. clarkii* | Philippines | AMCLUSU003 | 1 | 1 | 1 | * | 1 | 0 | 1 | 6 |
| *A. clarkii* | Philippines | AMCLUSU005 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. chrysopterus* | Fiji | AMCPFIJ001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. chrysopterus* | New Guinea | AMCPHKN219 | * | * | * | * | * | * | * | 7 |
| *A. chrysopterus* | French Polynesia | AMCPMOO059 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. chrysopterus* | French Polynesia | AMCPMOO060 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 6 |
| *A. chrysopterus* | French Polynesia | AMCPMOO061 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 3 |
| *A. chrysopterus* | Solomon Islands | AMCPSOL001 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 6 |
| *A. ephippium* | Sumatra | AMEPKRA001 | 0 | 0 | * | 0 | 0 | * | 0 | 2 |
| *A. frenatus* | Flores | AMFRFLO001 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 6 |
| *A. frenatus* | Japan | AMFROKI001 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 5 |
| *A. frenatus* | Philippines | AMFRUSU063 | * | 1 | * | * | * | * | * | 7 |
| *A. frenatus* | Philippines | AMFRUSU068 | 1 | * | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. latifasciatus* | Mozambique Channel | AMLFEUR001 | * | * | * | * | * | 0 | * | 6 |
| *A. latifasciatus* | Mozambique Channel | AMLFEUR002 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 2 |
| *A. latifasciatus* | Mozambique Channel | AMLFGLO001 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |

Continued on next page

Table 3.4 – Continued from previous page

| Species | Origin | Sample | 16S | CO1 | CYTB | RAG2 | RHOD | S71 | S72 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| *A. latifasciatus* | Mozambique Channel | AMLFJDN001 | 0 | 1 | 1 | 0 | 1 | * | 1 | 5 |
| *A. leucokranos* | New Guinea | AMLUHKN037 | * | * | * | 1 | * | * | 0 | 6 |
| *A. leucokranos* | Solomon Islands | AMLUSOL002 | 1 | 1 | 1 | * | 1 | 0 | 0 | 5 |
| *A. leucokranos* | Solomon Islands | AMLUSOL003 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 4 |
| *A. latezonatus* | Lord Howe Island | AMLZLHI001 | * | 0 | * | * | * | * | 0 | 5 |
| *A. latezonatus* | Lord Howe Island | AMLZLHI002 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 5 |
| *A. mccullochi* | Lord Howe Island | AMMCLHI001 | * | * | * | 0 | 1 | * | 0 | 5 |
| *A. mccullochi* | Lord Howe Island | AMMCLHI002 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. mccullochi* | Lord Howe Island | AMMCLHI003 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. mccullochi* | Lord Howe Island | AMMCLHI004 | 0 | 0 | 0 | 0 | * | 0 | 0 | 1 |
| *A. melanopus* | Flores | AMMEFLO001 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 5 |
| *A. melanopus* | Great Barrier Reef | AMMEGBR001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. melanopus* | Great Barrier Reef | AMMEGBR002 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 6 |
| *A. melanopus* | New Guinea | AMMEHKN054 | * | * | * | * | * | * | * | 7 |
| *A. melanopus* | New Guinea | AMMEHKN064 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 4 |
| *A. melanopus* | New Guinea | AMMEHKN065 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 6 |
| *A. melanopus* | New Guinea | AMMEHKN075 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| *A. melanopus* | New Guinea | AMMEHKN088 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *A. melanopus* | New Guinea | AMMEHKN102 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. melanopus* | New Guinea | AMMEHKN285 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. melanopus* | Micronesia | AMMEULI001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. nigripes* | Maldives | AMNIMAL001 | * | * | * | 0 | * | * | 0 | 5 |
| *A. nigripes* | Maldives | AMNIMAL002 | 1 | 0 | 0 | 0 | 1 | 0 | * | 3 |
| *A. ocellaris* | Darwin, Australia | AMOCDAR001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. ocellaris* | Darwin, Australia | AMOCDAR002 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 6 |
| *A. ocellaris* | Flores | AMOCFLO001 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 5 |
| *A. ocellaris* | Flores | AMOCFLO002 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |

Continued on next page

Table 3.4 – Continued from previous page

| Species | Origin | Sample | 16S | CO1 | CYTB | RAG2 | RHOD | S71 | S72 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| *A. ocellaris* | Philippines | AMOCUSU004 | * | * | * | * | * | * | * | 7 |
| *A. ocellaris* | Philippines | AMOCUSU006 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. percula* | Cenderawasih Bay | AMPCCEN001 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. percula* | Cenderawasih Bay | AMPCCEN002 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. percula* | Great Barrier Reef | AMPCGBR001 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 6 |
| *A. percula* | Great Barrier Reef | AMPCGBR002 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 3 |
| *A. percula* | Great Barrier Reef | AMPCGBR004 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 3 |
| *A. percula* | New Guinea | AMPCHKN006 | * | * | * | * | * | * | * | 7 |
| *A. percula* | New Guinea | AMPCHKN007 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 5 |
| *A. percula* | New Guinea | AMPCHKN008 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *A. percula* | New Guinea | AMPCHKN011 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| *A. percula* | New Guinea | AMPCHKN123 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. percula* | New Guinea | AMPCHKN204 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *A. percula* | New Guinea | AMPCHKN250 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. percula* | Solomon Islands | AMPCSOL001 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. percula* | Solomon Islands | AMPCSOL002 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. perideraion* | Flores | AMPIFLO001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. perideraion* | Great Barrier Reef | AMPIGBR001 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. perideraion* | New Guinea | AMPIHKN012 | * | * | * | * | * | * | * | 7 |
| *A. perideraion* | New Guinea | AMPIHKN018 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. perideraion* | New Guinea | AMPIHKN024 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *A. perideraion* | New Guinea | AMPIHKN089 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *A. perideraion* | New Guinea | AMPIHKN096 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. perideraion* | Palau | AMPIPAL001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. perideraion* | Philippines | AMPIUSU007 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. perideraion* | Philippines | AMPIUSU018 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 5 |
| *A. polymnus* | New Guinea | AMPOHKN141 | * | * | * | * | * | * | 0 | 6 |

Continued on next page

Table 3.4 – Continued from previous page

| Species | Origin | Sample | 16S | CO1 | CYTB | RAG2 | RHOD | S71 | S72 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| *A. polymnus* | New Guinea | AMPOHKN142 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| *A. polymnus* | New Guinea | AMPOHKN143 | 1 | 1 | 1 | 1 | 1 | 0 | * | 6 |
| *A. polymnus* | New Guinea | AMPOHKN145 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *A. polymnus* | New Guinea | AMPOHKN148 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *A. rubrocinctus* | Kimberley, Australia | AMRUKIM001 | * | 0 | * | 0 | 0 | 0 | 0 | 2 |
| *A. sandaracinos* | Flores | AMSAFLO001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. sandaracinos* | New Guinea | AMSAHKN022 | 1 | * | 0 | * | * | * | * | 6 |
| *A. sandaracinos* | New Guinea | AMSAHKN033 | 1 | 1 | * | 1 | 1 | 1 | 1 | 7 |
| *A. sandaracinos* | New Guinea | AMSAHKN071 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *A. sandaracinos* | New Guinea | AMSAHKN095 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 3 |
| *A. sandaracinos* | New Guinea | AMSAHKN120 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 3 |
| *A. sandaracinos* | New Guinea | AMSAHKN128 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 4 |
| *A. sandaracinos* | New Guinea | AMSAHKN202 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 2 |
| *A. sandaracinos* | New Guinea | AMSAHKN229 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *A. sandaracinos* | New Guinea | AMSAHKN230 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *A. sandaracinos* | Solomon Islands | AMSASOL004 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| *A. sandaracinos* | Solomon Islands | AMSASOL005 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 6 |
| *A. sandaracinos* | Solomon Islands | AMSASOL006 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 6 |
| *A. sandaracinos* | Solomon Islands | AMSASOL007 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. sandaracinos* | Solomon Islands | AMSASOL008 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 5 |
| *A. sandaracinos* | Solomon Islands | AMSASOL009 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 4 |
| *A. sandaracinos* | Philippines | AMSAUSU016 | * | 1 | 0 | 1 | 1 | 1 | 1 | 6 |
| *A. sandaracinos* | Philippines | AMSAUSU022 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *A. sandaracinos* | Philippines | AMSAUSU024 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *A. sandaracinos* | Philippines | AMSAUSU031 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *A. sandaracinos* | Philippines | AMSAUSU035 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *A. tricinctus* | Marshall Islands | AMTRMAR001 | * | * | * | * | * | * | * | 7 |

Continued on next page

Table 3.4 – Continued from previous page

| Species | Origin | Sample | 16S | CO1 | CYTB | RAG2 | RHOD | S71 | S72 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| *P. biaculeatus* | Flores | PRBIFLO001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *P. biaculeatus* | Great Barrier Reef | PRBIGBR001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *P. biaculeatus* | Great Barrier Reef | PRBIGBR002 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *P. biaculeatus* | New Guinea | PRBIHKN001 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| *P. biaculeatus* | New Guinea | PRBIHKN013 | * | * | * | * | * | * | * | 7 |
| *P. biaculeatus* | New Guinea | PRBIHKN025 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 6 |
| *P. biaculeatus* | New Guinea | PRBIHKN036 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *P. biaculeatus* | New Guinea | PRBIHKN043 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *P. biaculeatus* | New Guinea | PRBIHKN044 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *P. biaculeatus* | New Guinea | PRBIHKN046 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 2 |
| *P. biaculeatus* | New Guinea | PRBIHKN268 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *P. biaculeatus* | New Guinea | PRBIHKN280 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| *P. biaculeatus* | Philippines | PRBIUSU001 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| *P. biaculeatus* | Philippines | PRBIUSU002 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 6 |

# Bibliography

Addicott, J. F. (1986). Variation in the costs and benefits of mutualism: the interaction between yuccas and yucca moths. *Oecologia*, 70(4):486–494.

Allen, G. R. (1972). *The Anemonefishes, their classification and biology*. Tropical Fish Hobbyist Publications, Neptune City, New Jersey, first edition.

Allen, G. R. (1975). *The Anemonefishes: Their Classification and Biology*. Tropical Fish Hobbyist Publications, Hong Kong, second edition.

Allen, G. R. (2008). Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 18(5):541–556.

Allen, G. R., Drew, J. A., and Fenner, D. (2010). Amphiprion pacificus, a new species of anemonefish (Pomacentridae) from Fiji, Tonga, Samoa, and Wallis Island. *Aqua, International Journal of Ichthyology*, 16(3):129–138.

Allen, G. R., Drew, J. A., and Kaufman, L. (2008). Amphiprion barberi, a new species

of anemonefish (Pomacentridae) from Fiji, Tonga, and Samoa. *Aqua, International Journal of Ichthyology*, 14(3):105–114.

Almany, G. R., Berumen, M. L., Thorrold, S. R., Planes, S., and Jones, G. P. (2007). Local replenishment of coral reef fish populations in a marine reserve. *Science*, 316(5825):742–744.

Almeida-Neto, M., Guimarães, P. R., Loyola, R. D., and Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems : reconciling concept and measurement. *Oikos*, 117(March):1227–1239.

Almeida-Neto, M. and Ulrich, W. (2011). A straightforward computational approach for measuring nestedness using quantitative matrices. *Environmental Modelling & Software*, 26(2):173–178.

Bascompte, J., Jordano, P., Melian, C. J., and Olesen, J. M. (2003). The nested assembly of plant-animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387.

Beldade, R., Holbrook, S. J., Schmitt, R. J., Planes, S., and Bernardi, G. (2009). Isolation and characterization of eight polymorphic microsatellite markers from the orange-n anemonesh, Amphiprion chrysopterus. *Conservation Genetics Resources*, 1(1):333–335.

Beldade, R., Holbrook, S. J., Schmitt, R. J., Planes, S., Malone, D., and Bernardi, G.

(2012). Larger female fish contribute disproportionately more to self- replenishment. *Proceedings of the Royal Society B: Biological Sciences*, 279(1736):2116–2121.

Bell, G. (2001). Neutral Macroecology. *Science*, 293(5539):2413–2418.

Berumen, M. L., Almany, G. R., Planes, S., Jones, G. P., Saenz-Agudelo, P., and Thorrold, S. R. (2012). Persistence of self-recruitment and patterns of larval connectivity in a marine protected area network. *Ecology and Evolution*, 2(2):444–452.

Bird, C. E., Karl, S. A., Smouse, P. E., and Toonen, R. J. (2011). Detecting and measuring genetic differentiation. In Held, C., Koenemann, S., and Schubart, C. D., editors, *Phylogeography and Population Genetics in Crustacea*, pages 31–55. CRC Press.

Bivand, R. S. and Rundel, C. (2013). *rgeos: Interface to Geometry Engine - Open Source (GEOS).*

Blüthgen, N., Fründ, J., Vázquez, D. P., and Menzel, F. (2008). What do interaction network metrics tell us about specialization and biological traits? *Ecology*, 89(12):3387–99.

Blüthgen, N., Menzel, F., and Blüthgen, N. (2006). Measuring specialization in species interaction networks. *BMC Ecology*, 6(9):9.

Briggs, J. C. and Bowen, B. W. (2012). A realignment of marine biogeographic provinces with particular reference to fish distributions. *Journal of Biogeography*, 39(1):12–30.

Bronstein, J. L. (2001). The Costs of Mutualism. *American Zoologist*, 41(4):825–839.

Brouat, C., Sennedot, F., Audiot, P., Leblois, R., and Rasplus, J. Y. (2003). Fine-scale genetic structure of two carabid species with contrasted levels of habitat specialization. *Molecular Ecology*, 12(7):1731–1745.

Carlson, B. (1996). The Amphiprion leucokranos Mystery. *Aquarium Frontiers*, 3(3):34–37.

Carr, M. H., Anderson, T. W., and Hixon, M. A. (2002). Biodiversity, population regulation, and the stability of coral-reef fish communities. *Proceedings of the National Academy of Sciences*, 99(17):11241–11245.

Carr, M. H. and Reed, D. C. (1993). Conceptual Issues Relevant to Marine Harvest Refuges: Examples from Temperate Reef Fishes. *Canadian Journal of Fisheries and Aquatic Sciences*, 50(9):2019–2028.

Cattin, M. F., Bersier, L. F., Banasek-Richter, C., Baltensperger, R., and Gabriel, J. P. (2004). Phylogenetic constraints and adaptation explain food-web structure. *Nature*, 427(6977):835–839.

Chow, S. and Hazama, K. (1998). Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology*, 7(9):1255–1256.

Cooper, W. J., Smith, L. L., and Westneat, M. W. (2009). Exploring the radiation of a diverse reef fish family: Phylogenetics of the damselfishes (Pomacentridae), with new classifications based on molecular analyses of all genera. *Molecular Phylogenetics and Evolution*, 52(1):1–16.

Cowen, R. K., Paris, C. B., and Srinivasan, A. (2006). Scaling of connectivity in marine populations. *Science*, 311(5760):522–527.

Crean, A. J., Swearer, S. E., and Patterson, H. M. (2010). Larval supply is a good predictor of recruitment in endemic but not non-endemic fish populations at a high latitude coral reef. *Coral Reefs*, 29(1):137–143.

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8):772.

Darwin, C. R. (1862). *On the various contrivances by which British and foreign orchids are fertilised by insects.* John Murray, London.

Dobler, S. and Farrell, B. D. (1999). Host use evolution in Chrysochus milkweed beetles: evidence from behaviour, population genetics and phylogeny. *Molecular Ecology*, 8(8):1297–1307.

Doherty, P. J., Planes, S., and Mather, P. (1995). Gene flow and larval duration in seven species of fish from the Great Barrier Reef. *Ecology*, 76(8):2373–2391.

Dormann, C. F., Gruber, B., and Fruend, J. (2008). Introducing the bipartite Package: Analysing Ecological Networks. *R News*, 8(2):8–11.

Drew, J. A., Almany, G. R., Kaufman, L., and Barber, P. H. (2008). Endemism and regional color and genetic differences in five putatively cosmopolitan reef fishes. *Conservation Biology*, 22(4):965–975.

Dutilleul, P., Stockwell, J. D., Frigon, D., and Legendre, P. (2000). The Mantel test versus Pearson's correlation analysis : Assessment of the differences for biological and environmental studies. *Journal of agricultural, biological, and environmental statistics*, 5(2):131–150.

Elliott, J. K., Lougheed, S. C., Bateman, B., McPhee, L. K., and Boag, P. T. (1999). Molecular phylogenetic evidence for the evolution of specialization in anemonefishes. *Proceedings of the Royal Society B*, 266(1420):677–685.

Elliott, J. K. and Mariscal, R. N. (1995). Host selection, location, and association behaviors of anemonefishes in field settlement experiments. *Marine Biology*, 122(3):377–389.

Elliott, J. K. and Mariscal, R. N. (2001). Coexistence of nine anemonefish species: differential host and habitat utilization, size and recruitment. *Marine Biology*, 138(1):23–36.

Eschmeyer, W. N. (2014). Catalog of Fishes: Genera, Species, References.

Estes, J., Tinker, M., Williams, T., and Doak, D. (1998). Killer whale predation on sea otters linking oceanic and nearshore ecosystems. *Science*, 282:473–476.

Estes, J. A., Terborgh, J., Brashares, J. S., Power, M. E., Berger, J., Bond, W. J., Carpenter, S. R., Essington, T. E., Holt, R. D., Jackson, J. B. C., Marquis, R. J., Oksanen, L., and Oksanen, T. (2011). Trophic Downgrading of Planet Earth. *Science*, 333:301–306.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of

individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8):2611–2620.

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

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., and Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5):717–26.

Fautin, D. G. (1981). The Clownfish Sea Anemones: Stichodactylidae (Coelenterata: Actiniaria) and Other Sea Anemones Symbiotic with Pomacentrid Fishes. *Transactions of the American Philosophical Society*, 71(1):3–115.

Fautin, D. G. and Allen, G. R. (1997). *Anemonefishes and their host sea anemones*. Western Australian Museum, Perth.

Frédérich, B., Sorenson, L., Santini, F., Slater, G. J., and Alfaro, M. E. (2013). Iterative ecological radiation and convergence during the evolutionary history of damselfishes (Pomacentridae). *The American Naturalist*, 181(1):94–113.

Fricke, H. W. and Fricke, S. (1977). Monogamy and sex change by aggressive dominance in coral reef fish. *Nature*, 266(5605):830–832.

Fricke, R., Durville, P., Bernardi, G., Borsa, P., Mou-Tham, G., and Chabanet, P. (2013). Checklist of the shore fishes of Europa Island, Mozambique Channel, south-

western Indian Ocean, including 302 new records. *Stuttgarter Beitrage zur Naturkunde A*, 6(April 30).

Froese, R. and Pauly, D. (2014). FishBase.

Futuyma, D. J. and Mayer, G. C. (1980). Non-allopatric speciation in animals. *Systematic Biology*, 29(3):254–271.

Gaston, K. J., Blackburn, T. M., and Lawton, J. H. (1997). Interspecific Abundance-Range Size Relationships: An Appraisal of Mechanisms. *Journal of Animal Ecology*, 66(4):pp. 579–601.

Glaubitz, J. C. (2004). convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes*, 4(2):309–310.

Guillot, G., Santos, F., and Estoup, A. (2008). Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics*, 24(11):1406–1407.

Guimarães, P. R., Rico-Gray, V., Oliveira, P. S., Izzo, T. J., dos Reis, S. F., and Thompson, J. N. (2007). Interaction intimacy affects structure and coevolutionary dynamics in mutualistic networks. *Current Biology*, 17(20):1797–1803.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, 59(3):307–321.

Halpern, B. S. and Warner, R. R. (2003). Matching marine reserve design to reserve objectives. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1527):1871–1878.

Hanski, I. (1998). Metapopulation dynamics. *Nature*, 396(6706):41–49.

Harrison, H. B., Williamson, D. H., Evans, R. D., Almany, G. R., Thorrold, S. R., Russ, G. R., Feldheim, K. A., van Herwerden, L., Planes, S., Srinivasan, M., Berumen, M. L., and Jones, G. P. (2012). Larval Export from Marine Reserves and the Recruitment Benefit for Fish and Fisheries. *Current Biology*, 22(11):1023–1028.

Hembry, D. H. (2012). *Coevolutionary diversification of leafflower moths (Lepidoptera: Gracillariidae: Epicephala) and leafflower trees (Phyllanthaceae: Phyllanthus sensu lato [Glochidion]) in southeastern Polynesia*. Ph.d. dissertation, University of California, Berkeley.

Hijmans, R. J., Williams, E., and Vennes, C. (2012). *geosphere: Spherical Trigonometry*.

Hixon, M. A. and Brostoff, W. N. (1983). Damselfish as keystone species in reverse: Intermediate disturbance and diversity of reef algae. *Science*, 220(4596):511–513.

Hodge, J. R., Read, C. I., van Herwerden, L., and Bellwood, D. R. (2012). The role of peripheral endemism in species diversification: evidence from the coral reef fish genus Anampses (Family: Labridae). *Molecular phylogenetics and evolution*, 62(2):653–63.

Holbrook, S. J. and Schmitt, R. J. (2005). Growth, reproduction and survival of a

tropical sea anemone (Actiniaria): benefits of hosting anemonefish. *Coral Reefs*, 24(1):67–73.

Howe, H. F. (1984). Constraints on the evolution of mutualisms. *The American Naturalist*, 123(6):764–777.

Huelsenbeck, J. P. and Rannala, B. (2004). Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. *Systematic biology*, 53(6):904–13.

Ivanova, N. V., Zemlak, T. S., Hanner, R. H., and Hebert, P. D. N. (2007). Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes*, 7(4):544–548.

Jakobsson, M. and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14):1801–1806.

Jones, G. P., Milicich, M., Emslie, M., and Lunow, C. (1999). Self-recruitment in a coral reef fish population. *Nature*, 402(6763):802–804.

Jones, G. P., Planes, S., and Thorrold, S. R. (2005). Coral reef fish larvae settle close to home. *Current Biology*, 15(14):1314–8.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059–3066.

Kelley, S. T. and Farrell, B. D. (1998). Is specialization a dead end? The phylogeny of host use in Dendroctonus bark beetles (Scolytidae). *Evolution*, 52(6):1731–1743.

Kelley, S. T., Farrell, B. D., and Mitton, J. B. (2000). Effects of specialization on genetic differentiation in sister species of bark beetles. *Heredity*, 84(2):218–227.

Legendre, P. and Fortin, M.-J. (2010). Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, 10(5):831–844.

Leis, J. M. (1991). The pelagic stage of reef fishes: The larval biology of coral reef fishes. In *The Ecology of Fishes on Coral Reefs*, pages 183–230. Academic Press.

Leis, J. M. (2007). Behaviour as input for modelling dispersal of fish larvae: behaviour, biogeography, hydrodynamics, ontogeny, physiology and phylogeny meet hydrography. *Marine Ecology Progress Series*, 347:185–193.

Litsios, G., Sims, C. a., Wüest, R. O., Pearman, P. B., Zimmermann, N. E., and Salamin, N. (2012). Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, 12:212.

Magnusson, A. (2012). *gmt: Interface between GMT Map-Making Software and R.*

Malay, M. C. M. D. and Paulay, G. (2010). Peripatric speciation drives diversification and distributional pattern of reef hermit crabs (Decapoda: Diogenidae: Calcinus). *Evolution*, 64(3):634–662.

McClanahan, T. R., Ateweberhan, M., Darling, E. S., Graham, N. A. J., and Muthiga, N. A. (2014). Biogeography and Change among Regional Coral Communities across the Western Indian Ocean. *PLoS ONE*, 9(4):e93385.

Mora, C., Andrèfouët, S., Costello, M. J., Kranenburg, C., Rollo, A., Veron, J., Gaston, K. J., and Myers, R. A. (2006). Coral Reefs and the Global Network of Marine Protected Areas. *Science*, 312(5781):1750–1751.

Moreau, C. S., Bell, C. D., Vila, R., Archibald, S. B., and Pierce, N. E. (2006). Phylogeny of the ants: Diversification in the age of angiosperms. *Science*, 312(5770):101–104.

Myers, R. A. and Worm, B. (2003). Rapid worldwide depletion of predatory fish communities. *Nature*, 423(6937):280–283.

Nielsen, A. and Bascompte, J. (2007). Ecological networks, nestedness and sampling effort. *Journal of Ecology*, 95(5):1134–1141.

Nosil, P. and Mooers, A. O. (2005). Testing hypotheses about ecological specialization using phylogenetic trees. *Evolution*, 59(10):2256–2263.

Obura, D. (2012). The diversity and biogeography of Western Indian ocean reef-building corals. *PLoS One*, 7(9):e45013.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., and Wagner, H. (2013). *vegan: Community Ecology Package.*

Ollerton, J., McCollin, D., Fautin, D. G., and Allen, G. R. (2007). Finding NEMO: nestedness engendered by mutualistic organization in anemonefish and their hosts. *Proceedings of the Royal Society B*, 274(1609):591–598.

Palumbi, S. R. (2003). Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications*, 13(1):S146–S158.

Palumbi, S. R., Martin, A., Romano, S., McMillan, W. O., Stice, L., and Grabowski, G. (2002). The simple fool's guide to PCR. Technical report, University of Hawaii.

Patefield, A. (1981). An efficient method of generating r  c tables with given row and column totals. *Applied Statistics*, 30:91–97.

Patterson, B. D. and Atmar, W. (1986). Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological Journal of the Linnean Society*, 28(1-2):65–82.

Pebesma, E. J. and Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2).

Pinsky, M. L., Montes, H. R., and Palumbi, S. R. (2010). Using isolation by distance and effective density to estimate dispersal scales in anemonefish. *Evolution*, 64(9):2688–700.

Planes, S., Jones, G. P., and Thorrold, S. R. (2009). Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences*, 106(14):5693–7.

Porat, D. and Chadwick-Furman, N. E. (2004). Effects of anemonefish on giant sea anemones: expansion behavior, growth, and survival. *Hydrobiologia*, 530:513–520.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959.

Quenouille, B., Bouchenak-Khelladi, Y., Hervet, C., and Planes, S. (2004). Eleven microsatellite loci for the saddleback clownfish Amphiprion polymnus. *Molecular Ecology Notes*, 4(2):291–293.

R Core Team (2013). R: A language and environment for statistical computing.

Rezende, E. L., Jordano, P., and Bascompte, J. (2007). Effects of phenotypic complementarity and phylogeny on the nested structure of mutualistic networks. *Oikos*, 116(11):1919–1929.

Ricciardi, F., Boyer, M., and Ollerton, J. (2010). Assemblage and interaction structure of the anemonefish-anemone mutualism across the Manado region of Sulawesi, Indonesia. *Environmental Biology of Fishes*, 87(4):333–347.

Riginos, C. and Victor, B. C. (2001). Larval spatial distributions and other early lifehistory characteristics predict genetic differentiation in eastern Pacific blennioid fishes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1479):1931–1936.

Roberts, C. M., McClean, C. J., Veron, J. E. N., Hawkins, J. P., Almany, G. R., McAllister, D. E., Mittermeier, C. G., Schueler, F. W., Spalding, M., Wells, F.,

Vynne, C., and Werner, T. B. (2002). Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science*, 295(5558):1280–1284.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., and Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology*, 61(3):539–542.

Roopin, M. and Chadwick, N. E. (2009). Benefits to host sea anemones from ammonia contributions of resident anemonefish. *Journal of Experimental Marine Biology and Ecology*, 370(1-2):27–34.

Rosenberg, N. A. (2004). distruct: a program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1):137–138.

Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R., and Planes, S. (2009). Estimating connectivity in marine populations: an empirical evaluation of assignment tests and parentage analysis under different gene flow scenarios. *Molecular Ecology*, 18(8):1765–1776.

Sætre, R. (1985). Surface currents in the Mozambique channel. *Deep Sea Research Part A. Oceanographic Research Papers*, 32(12):1457–1467.

Sambrook, J., Fritsch, E. F., Maniatis, T., and Fritschi, E. F. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, second edition.

Santini, S. and Polacco, G. (2006). Finding Nemo: Molecular phylogeny and evolution of the unusual life style of anemonefish. *Gene*, 385:19–27.

Schluter, D. (2000). *The Ecology of Adaptive Radiation*. Oxford University Press, Oxford.

Schouten, M. W., Ruijter, W. P. M. D., Jan van Leuwen, P., and Ridderinkhof, H. (2003). Eddies and variability in the Mozambique Channel. *Deep Sea Research II*, 50:1987–2003.

Sevilla, R. G., Diez, A., Noren, M., Mouchel, O., Jerome, M., Verrez-Bagnis, V., van Pelt, H., Favre-Krey, L., Krey, G., and Bautista, J. M. (2007). Primers and polymerase chain reaction conditions for DNA barcoding teleost fish based on the mitochondrial cytochrome b and nuclear rhodopsin genes. *Molecular Ecology Notes*, 7(5):730–734.

Shanks, A. L. (2009). Pelagic Larval Duration and Dispersal Distance Revisited. *The Biological Bulletin*, 216(3):373–385.

Shanks, A. L., Grantham, B. A., and Carr, M. H. (2003). Propagule dispersal distance and the size and spacing of marine reserves. *Ecological Applications*, 13(sp1):159–169.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423; 623–656.

Shulman, M. J. and Bermingham, E. (1995). Early life histories, ocean currents, and the population-genetics of Caribbean reef fishes. *Evolution*, 49(5):897–910.

Stafford, R., Hart, A. G., Collins, L., Kirkhope, C. L., Williams, R. L., Rees, S. G., Lloyd, J. R., and Goodenough, A. E. (2010). Eu-social science: the role of internet social networks in the collection of bee biodiversity data. *PLoS One*, 5(12):e14381.

Stireman, J. O. (2005). The evolution of generalization? Parasitoid flies and the perils of inferring host range evolution from phylogenies. *Journal of Evolutionary Biology*, 18(2):325–336.

Swearer, S. E., Caselle, J. E., Lea, D. W., and Warner, R. R. (1999). Larval retention and recruitment in an island population of a coral-reef fish. *Nature*, 402(6763):799–802.

Tang, K. L. (2001). Phylogenetic Relationships among Damselfishes (Teleostei: Pomacentridae) as Determined by Mitochondrial DNA Data. *Copeia*, 2001(3):591–601.

Thompson, A. R., Adam, T. C., Hultgren, K. M., and Thacker, C. E. (2013). Ecology and evolution affect network structure in an intimate marine mutualism. *The American naturalist*, 182(2):E58–72.

Thompson, J. N. (1998). Rapid evolution as an ecological process. *Trends in Ecology and Evolution*, 5347(1977):329–332.

Thompson, J. N. (2005). *The Geographic Mosaic of Coevolution*. University of Chicago Press, Chicago.

Thompson, J. N. (2009). The Coevolving Web of Life. *American Naturalist*, 173(2):125–140.

Thompson, J. N. and Cunningham, B. M. (2002). Geographic structure and dynamics of coevolutionary selection. *Nature*, 417(6890):735–8.

Thresher, R. E., Colin, P. L., and Bell, L. J. (1989). Planktonic duration, distribution and population structure of western and central Pacific damselfishes (Pomacentridae). *Copeia*, 1989(2):420–434.

Timm, J., Figiel, M., and Kochzius, M. (2008). Contrasting patterns in species boundaries and evolution of anemonefishes (Amphiprioninae, Pomacentridae) in the centre of marine biodiversity. *Molecular Phylogenetics and Evolution*, 49(1):268–276.

Timm, J. and Kochzius, M. (2008). Geological history and oceanography of the Indo-Malay Archipelago shape the genetic population structure in the false clown anemonefish (Amphiprion ocellaris). *Molecular Ecology*, 17(18):3999–4014.

Turelli, M., Barton, N. H., and Coyne, J. A. (2001). Theory and speciation. *Trends in Ecology and Evolution*, 16(7):330–343.

Ulrich, W., Almeida-Neto, M., and Gotelli, N. J. (2009). A consumer's guide to nestedness analysis. *Oikos*, 118(1):3–17.

UNEP-WCMC, Centre, W., WRI, and TNC (2010). Global distribution of warm-water coral reefs, compiled from multiple sources, including the Millennium Coral Reef Mapping Project.

van der Meer, M. H., Jones, G. P., Hobbs, J.-P. a., and van Herwerden, L. (2012). Historic hybridization and introgression between two iconic Australian anemonefish and

contemporary patterns of population connectivity. *Ecology and evolution*, 2(7):1592–604.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M., and Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4(3):535–538.

Wake, D. and Vredenburg, V. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences*, 105:11466–11473.

Waples, R. S. (1987). A multispecies approach to the analysis of gene flow in marine shore fishes. *Evolution*, 41(2):385–400.

Wellington, G. M. and Victor, B. C. (1989). Planktonic larval duration of one hundred species of Pacific and Atlantic damselfishes (Pomacentridae). *Marine Biology*, 101(4):557–567.

Yeakel, J. D., Guimarães, P. R., Bocherens, H., and Koch, P. L. (2013). The impact of climate change on the structure of Pleistocene food webs across the mammoth steppe. *Proceedings of the Royal Society B: Biological Sciences*, 280(1762):20130239–20130239.

Yotoko, K. S. C., Prado, P. I., Russo, C. A. M., and Solferini, V. N. (2005). Testing the trend towards specialization in herbivore-host plant associations using a molec-

ular phylogeny of Tomoplagia (Diptera : Tephritidae). *Molecular Phylogenetics and Evolution*, 35(3):701–711.