UNIVERSITY OF CALIFORNIA

SANTA CRUZ

MOLECULAR EVOLUTION OF TROPICAL AMERICAN SHOREFISH

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

DOCTOR OF PHILOSOPHY

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

Yvette R. Alva-Campbell

June 2011

The Dissertation of
Yvette R. Alva-Campbell is approved:

_____
Professor Giacomo Bernardi, Chair

_____
Professor Peter T. Raimondi

_____
Professor Don Croll

_____
Professor Eric J. Routman

_____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Molecular Evolution of Tropical American Shorefish

by

Yvette R. Alva-Campbell

The goal of my dissertation is to understand the mechanisms that may inhibit gene flow among Tropical American shorefish. The first chapter of my dissertation involves the study of speciation, the outcome of limited gene flow between populations. To study speciation in Tropical America, I reconstructed the phylogeny of *Holacanthus* angelfish. I calibrated a molecular clock based on the Isthmus of Panama and observed many cases of recent speciation events. Five of the seven species of angelfish diverged from one another less than 3.5 myr. Recent speciation events resulted in geminate clades instead of geminate species found on either side of the Isthmus of Panama. The basal species of *Holacanthus* is from Africa, suggesting that *Holacanthus* may have originated from the Indian Ocean and migrated westward to the Pacific. For the next two chapters of my dissertation I studied the processes that limit gene flow among populations. I performed a multidisciplinary study combining molecular evolution, larval ecology, spawning behavior, biogeography, and geology in order to measure the amount, direction, and timing of past migration. The combination of these fields serves as a valuable tool in explaining the evolutionary history of a species. Although understanding the

evolutionary history of one species is important, it is of more interest to determine whether there is concordance across taxa that occur in the same area (Avise 2000). For the second and third chapter of my dissertation I tested for genetic concordance among shorefish at two different spatial scales the Tropical America (Tropical Atlantic and Tropical Eastern Pacific) and the Tropical Eastern Pacific. Each comparative phylogeographic study resulted in different genetic signatures for each species. Oceanic currents and habitat fragmentation limited population connectivity for some species while having no affect in other species. Species that form spawning aggregations and have positively buoyant eggs exhibited more gene flow than species that are demersal spawners and have negatively buoyant eggs. The results of this study indicate that the ecology and biogeography of a species is important in order to infer the evolutionary history of a species.

Acknowledgements

A special thank you to my husband and best friend, Benjamin C. Campbell, who was such a motivating force and support system through my dissertation. A humble thanks to my parents and my brother who instilled in me the values and skills I needed to succeed in my academic career. A special thank you to my son, Isaac, who kept me sane during this process and helped me realize that there is more in life than academia. To my extended family and friends who kept me laughing through this process. I am indebted to the people above, they are inspirational beings in my life and without them none of this would be possible, I hope I made you all proud. I would like to thank my committee (Eric Routman, Don Croll, Pete Raimondi, Giacomo Bernardi) for their insightful comments and suggestions for improving my dissertation research. An additional thank you to Giacomo Bernardi who was such an inspirational advisor. The advice he gave me in both academia and life will always be treasured. I am so grateful to my labmates for their continuing support, motivation, and laughter.

The laboratory and fieldwork was accomplished with the aid of several individuals, too many to mention, but I would like to thank all of you and I will forever be grateful to each and every one of you. This work would not have been completed without the financial assistance of the MBRS program.

Introduction

How do environmental factors along with the physical and ecological constraints of the organism influence the movement of individuals between populations? Studying how these factors affect population connectivity is essential for understanding how populations adapt to changing environments (Avise 2000). If barriers to dispersal persist, populations will evolve independently into genetically distinct populations, and possibly undergo speciation or worse extinction (Avise 2000). In the marine realm, addressing these issues may prove difficult due to the lack of oceanographic and ecological information. The advent of molecular techniques has become essential in inferring past migratory movements in marine systems (Palumbi 1995). In 1987, Avise and colleagues coined the term "intraspecific phylogeography" for the use of molecular data to reconstruct population histories in relation to geography (Wiley and Hagen 1997, Avise et al. 1987). Coupling past migratory movements along with the geology of the region, scientists are now able to reconstruct the evolutionary history of a species to better understand how past and current climatic changes may have affected the distribution of a species. Although understanding the phylogeography of one species is important, the overall goal is to obtain concordance across multiple taxa that occupy the same region (Avise 2000). Comparative phylogeographic studies are less common than single-species phylogeographic studies, but have become

extremely useful in revealing common genetic patterns as well as possible causes of genetic heterogeneities (Avise 2000, 2004).

Genetic heterogeneities found among species can result from ecological life history variation. Marine fish exhibit an array of life history traits, specifically differences in mating and larval behavior that may facilitate or limit migration. Most reef fish have two main life stages: a pelagic larval stage and a sedentary reef-associated adult stage. The pelagic larval phase of a reef fish is thought to provide the only opportunity for dispersal (Leis 1991). Knowledge of larval ecology is limited; it is still unclear whether there is a positive correlation between pelagic larval duration (PLD) and gene flow (Riginos and Victor 2001), few studies comparing fish with differing PLD have found any relationship between these two parameters (Shulman and Bermingham 1995; Victor 2000; Riginos and Victor 2001). Because PLD is an inconsistent indicator of gene flow, other life history traits, such as egg buoyancy and spawning behavior, are showing up in the literature as possible factors affecting dispersal for marine taxa (Shulman and Bermingham 1995). For instance, an organism that forms spawning aggregations, produces pelagic eggs, and has a long pelagic larval duration may have more opportunities for dispersal compared to a demersal spawner with negatively buoyant eggs that does not form spawning aggregations (Shulman and Bermingham 1995). Differences in mating and larval behavior may lead to differences in genetic signatures. Therefore, multiple

disciplines, including ecology, need to be combined to reconstruct the evolutionary history of a species.

For my dissertation research I attempted to understand the underlining mechanisms involved in migration among populations of shorefish that reside in Tropical America. The main goal of my dissertation is to understand how environmental and life history factors play a role in dispersal? To provide insight on dispersal capabilities among marine fish I divided my dissertation according to three different spatial scales, the first chapter being the broadest and the last chapter being the smallest spatial scale.

The first chapter of my dissertation involves studying speciation within the genus *Holacanthus*. *Holacanthus* comprises of 7 recognized species of angelfish that are found in the Tropical Eastern Pacific and Atlantic. A reconstruction of the phylogeny of angelfish will provide information on the evolutionary history of this genus. In addition the implementation of a molecular clock, based on the Isthmus of Panama, will reveal the timing of speciation events. A taxonomic comparison will allow me to address whether certain speciation events are linked to documented climatic events that may have prohibited dispersal. A study of this magnitude provides information on how barriers to dispersal can lead to speciation events.

To determine whether environmental factors play a significant role in dispersal capabilities, for the second chapter of my dissertation, I compared the

population genetics of two geminate pairs (*Abudefduf troschelii vs. A. saxatilis* and *Lutjanus argentiventris vs. L. apodus*) that reside in two contrasting oceans, the Tropical Eastern Pacific (TEP) and the Tropical Western Atlantic (TWA). The TEP and TWA, separated by the rise of the Isthmus approximately 3 million years ago, differ drastically in the amount and type of habitat available for settlement as well as oceanographic features such as the stability of currents, water temperature, and tidal changes. By choosing geminate species that are found on either side of the Isthmus and share similar ecological requirements, I minimized phylogenetic variation to test whether differences in population structure are caused by differing oceanographic environments. To determine whether ecological factors play a significant role in affecting dispersal capabilities, I performed a comparison among two taxonomic families, *Abudefduf sp.* and *Lutjanus sp.* that occupy the same region, but differ in life history characteristics. *Abudefduf sp.* are benthic spawners and are expected to have fewer opportunities for dispersal than *Lutjanus sp.* that are broadcast spawners, have pelagic eggs, and form spawning aggregations. By choosing species that have overlapping distributions, I minimized environmental variation in order to test whether ecological traits, for example differences in spawning behavior, influence levels of gene flow.

Although the Tropical Eastern Pacific (TEP) has received less attention compared to its sister ocean, the Tropical Western Atlantic, the TEP has a unique

landscape that makes it a perfect location to study dispersal capabilities of marine species. For the last chapter of my dissertation I performed a fine-scale population genetic study among three endemic species found in the Tropical Eastern Pacific (TEP). The TEP coastline is divided into four recognized biogeographic regions (Pacific Baja, Cortez, Mexican, and Panamic) each containing distinct thermal gradients and oceanic currents. Each province is separated from one another by gaps of habitat not suitable for reef fish settlement (the Sinaloan Gap, 370km; and the Central American Gap, 1,200km). To determine whether long stretches of sandy beaches act as true barriers to dispersal, I chose to reconstruct the phylogeography for the following species of marine fish: *Lutjanus argentiventris, Anisotremus interruptus,* and *Abudefduf troschelii.* All fish are abundant, associate with similar rock or reef habitat, have overlapping distributions in the TEP, and have sister species that reside in the Atlantic to enable the calibration of a molecular clock. These species were further chosen because they differ in life history characteristics. Differences in genetic signatures among fish that are assumed to have the same evolutionary history, may indicate that life history differences are influencing dispersal capabilities more so than geological processes.

Overall my dissertation incorporates multiple disciplines, such as ecology, oceanography, and molecular evolution, in order to identify the environmental and ecological forces responsible for dispersal in marine fish with the goal of

understanding how marine organisms will adapt to their environment. A comparative approach among fish that reside in Tropical America will further provide regional information that can be used by other scientists to provide insight on the genetic composition of other marine organisms that inhabit the same region. Understanding the forces that govern population structure in marine organisms is crucial to addressing fundamental questions about evolutionary and ecological processes affecting life in the oceans.

Chapter 1: Molecular phylogenetics and evolution of *Holacanthus* angelfishes (Pomacanthidae)

Abstract

The angelfish genus *Holacanthus* includes seven species in the Tropical Eastern Pacific and Atlantic. In this study we performed an analysis on all species, the closely related regal angelfish, and the Cortez angelfish, using four mitochondrial and one nuclear marker. Our results support a monophyletic *Holacanthus*. The close relationship between *P. diacanthus* and *Holacanthus* suggests a potential Indian Ocean origin of *Holacanthus*. We found a split into two clades with divergences that were consistent with the rise of the Isthmus of Panama. An internally calibrated molecular clock thus placed the origin of *Holacanthus* to ~10.2 - 7.6 million years ago.

Introduction

The angelfish genus *Holacanthus* (family Pomacanthidae) is relatively small with only 7 recognized species. *Holacanthus* resides in the tropical portion of three discrete biogeographic regions: The Eastern Atlantic, the Western Atlantic, and the Tropical Eastern Pacific (Figure 1.1). The West African angelfish, *Holacanthus africanus*, is the only species found along the tropical eastern Atlantic coastline, in the Gulf of Guinea (São Tomé) and the Cape Verde Islands. Within the western

Atlantic, three species occur on coral and rocky reefs, *H. tricolor, H. bermudensis, and H. ciliaris*. The rock beauty, H. tricolor, and the queen angelfish, *H. ciliaris* have the widest distributions, from Bermuda to Brazil, where different color morphs of *H. ciliaris* can be found on the NE coast and at isolated offshore islands (Luiz Jr 2003, Feeley et al. 2009). The blue angelfish, *H. bermudensis* has a more restricted distribution, from Bermuda, along the entire Atlantic coast of the US from North Carolina south, and throughout the Gulf of Mexico (Smith-Vaniz et al 1999). Atlantic *Holacanthus* have traditionally been divided into two separate subgenera, *Angelichthys*, which comprises *H. africanus, H. bermudensis,* and *H. ciliaris,* and *Holacanthus* which comprises only *H. tricolor* (Allen et al. 1998, Debelius et al. 2003).

Within the Tropical Eastern Pacific, *H. passer, H. clarionensis, and H. limbaughi* occur on rocky and coral reefs, (Robertson and Allen, 2008) (Figure 1.1). The king angelfish, *H. passer*, has the widest distribution, from central Baja California, Mexico, through the Gulf of California south to northern Peru, plus the Revillagigedo (its abundance there is unclear), Cocos, Malpelo, and the Galapagos Islands. Two insular endemic species, the Clipperton angelfish *H. limbaughi*, and the Clarion angelfish, *H. clarionensis,* are found at Clipperton Atoll and the Revillagigedo Islands, respectively. Small numbers, most likely vagrants, of *H. clarionensis* are found at the southern tip of Baja California, while there is a single record of that species at Clipperton Island (Allen and Robertson 1997). There are no records of *H. limbaughi* from any site

other than Clipperton, or of *H. passer* at that island. Traditionally, these three species have been grouped in a single subgenus, *Plitops* (Allen et al. 1998, Debelius et al. 2003).

The rise of the Central American Isthmus separated many populations of marine organisms, with the final closure of the Isthmus of Panama producing geminate pairs of similar-looking species (Jordan 1908). *Holacanthus* has been described as a potential candidate, with *H. passer* and *H. bermudensis* being the likely geminates (Thomson et al. 2000). In a molecular (mtDNA) study of angelfishes that included four *Holacanthus* species, Bellwood et al. (2004) proposed a phylogenetic hypothesis for the group. That analysis indicated that *H. bermudensis* and *H. passer* were sister species (thus potentially breaking up the subgenera *Plitops* and *Angelichthys*). In addition, that study, which included a thorough representation of most angelfish genera, indicated that the regal angelfish, *Pygoplites diacanthus*, the sole member of its genus, is the closest relative of the genus *Holacanthus*, raising the possibility of it being a basal *Holacanthus* (Bellwood et al. 2004), or potentially well within *Holacanthus*, thus disrupting its monophyletic status. Although *Pygoplites diacanthus* was originally described as *Holacanthus diacanthus* (Bleeker 1857), an early morphological phylogenetic analysis by Shen and Liu (1978) indicated that *Pygoplites* is sister to *Apolemichthys*. A later allozyme comparison done by Chung and Woo (1998) then placed *Pygoplites* closer to *Pomacanthus* than to *Holacanthus*

(see Fig. 1 from Bellwood et al. 2004). However, neither of the two molecular studies incorporated sampling of all seven members of *Holacanthus*.

Pygoplites diacanthus is widely distributed over the entire Indo-central Pacific. This distribution does not indicate whether *Holacanthus* originated either from the ancestor of *Pygoplites* in the TEP after migrating from the central Pacific, or in West Africa, after colonization from the Indian Ocean. Thus, a complete phylogeny of the genus *Holacanthus* is necessary to elucidate the evolutionary history of this genus.

To achieve this goal, we sampled all 7 known *Holacanthus* species (including different color morphs), *P. diacanthus* and the Cortez angelfish, *Pomacanthus zonipectus*, a TEP species, as an outgroup, and used four mitochondrial and one nuclear molecular markers. We first established a complete phylogeny of the genus, and then further investigated the status of geminate groups using the more variable mitochondrial control region marker.

Materials and Methods

Sampling sizes and locations are listed in Table 1.1. Preservation and DNA extraction protocols followed Bernardi et al. (2008). Amplifications of 12S rRNA, 16S rRNA, cytochrome b, and control region segments were performed using the universal primers 12SAL-12SBH, 16SAR-16SBR, GLUDGL-CB3H, and CRA-CRE, respectively (Palumbi et al. 1991, Lee et al. 1995) with 35 cycles at a denaturation

temperature of 94^{0}C for 30s, an annealing temperature of 52-54^{0}C, and an extension of 30s. at 72^{0}C. Amplification of the 1st intron of the nuclear S7 ribosomal protein used the primers S7RPEX1F and S7RPEX2R (Chow and Hazama 1998) followed similar cycles but with an annealing temperature of 56^{0}C.

After purification of the PCR products, following the manufacturer's protocol (Applied Biosystems, Foster City, CA), sequencing was performed in both directions with the primers used in the PCR amplification on an ABI 3100 automated sequencer (Applied Biosystems, Foster City, CA). In the case of the nuclear marker, heterozygous individuals were found to be very rare, and when present, only one allele was scored per individual (we did not find individuals with more than one heterozygotic site, thus making the calling of the two alleles possible).

Phylogenetic analyses

We used the computer program Clustal V implemented by Sequence Navigator (Applied Biosystems) to align the DNA sequences. Overall, few insertions and deletions (indels) were observed. For 12SrRNA, 3 indels of one base pair (bp) each were observed, for 16S rRNA 6 indels accounting for a total of 9bp, and for S7 fragments, 4 indels accounting for a total of 14bp were observed. These indels were removed from the subsequent phylogenetic analyses, yet their inclusion did not change the results. Phylogenetic relationships were assessed by Maximum Likelihood (ML, GARLI software, Zwickl 2006), Maximum Parsimony (MP, PAUP*

software, Swofford 2003), and Neighbor-Joining (NJ, PAUP software), methods. For Maximum Likelihood topologies, we conducted 10 independent runs in GARLI, using default settings and the automated stopping criterion, terminating the search when the ln score remained constant for 20,000 consecutive generations. The tree with the highest likelihood was retained and is presented here. MP searches included 100 random addition replicates and TBR branch swapping with the Multrees option. NJ reconstructions used distances based on substitution models obtained with Modeltest (Posada and Crandall 1998) (HKY + G). Statistical confidence in nodes was evaluated using 2000 non-parametric bootstrap replicates (Felsenstein 1985) (100 replicates for Maximum Likelihood in GARLI, using the automated stopping criterion set at 10,000 generations). Topological differences were tested using a Shimodaira and Hasegawa test (Shimodaira and Hasegawa 1999) implemented by PAUP, based on resampling of estimated log-likelihoods tests (RELL, 1000 replicates). Topological differences with or without an enforced molecular clock were tested using a Shimodaira and Hasegawa test (Shimodaira and Hasegawa 1999) implemented by PAUP. Genetic divergence was estimated using distances based on substitution models obtained with Modeltest (HKY + G).

Results

We used two sets of data, one broad set for a phylogenetic analysis and one more restricted set for a detailed analysis of Trans-Isthmian species. The dataset for

the broad analysis comprised 2283 aligned base pairs (bp), which included 346bp, 547bp, 699bp and 692bp for 12S rRNA, 16S rRNA, cytochrome b, and 1st intron of the ribosomal protein S7 fragments respectively. The dataset for the restricted analysis, based on the mitochondrial control region, was a set of 21 individuals that were representative of Trans-Isthmian geminate species (see Figure 1.2B). This dataset comprised 410 aligned base pairs.

Phylogenetic relationships

Phylogenetic relationships based on each molecular marker were found not to be statistically different from each other, thus we decided to concatenate all markers (three mitochondrial and one nuclear marker). The three reconstruction methods (ML, MP, NJ) resulted in identical topologies (Figure 1.2A shows the maximum likelihood topology). Using *P. zonipectus* as an outgroup, we found that the regal angelfish, *P. diacanthus* was not a member of *Holacanthus*, a result that is consistent with previous findings (Shen and Liu 1978, Allen 1981, Chung and Woo 1988, Bellwood et al. 2004). Forcing *Pygoplites* inside *Holacanthus* always resulted in a topology that was significantly worse than what is presented in Figure 1.2A (SH test, $p<0.05$).

The West Africa angelfish, *H. africanus,* from São Tomé were not distinct from Cape Verde Islands individuals, and were found to be basal to the other members of the genus (Figure 1.2A). The second species branching off in that tree was the rock

beauty, *H. tricolor*. Two clades developed after *H tricolor* branched off, one that included the Caribbean species *H. ciliaris* and *H. bermudensis,* the other included the TEP species, *H. passer, H. clarionensis,* and *H. limbaughi.*

Very little genetic differentiation was found among the extreme localities. For example, individuals from Brazil and the Caribbean of *H. tricolor* were genetically indistinguishable (Figure 1.2A).

Trans-Isthmian phylogeny

In order to obtain a fine-scale phylogeny for the group of species that occur across the Isthmus of Panama, we decided to use more individuals and the faster mitochondrial control region as a molecular marker. Results obtained with the control region are consistent with results of the wider analysis presented above (Figure 1.2B). In the TEP, one clade included *H. passer* and *H. clarionensis*. *Holacanthus clarionensis* did not separate in its own lineage. No coloration or morphological characters indicated that the sequenced individual was a hybrid, yet further work is necessary to determine the genetics of this species. The *passer-clarionensis* clade was sister to all Clipperton angelfish, *H. limbaughi*, which formed a monophyletic group. Considering its very restricted geographic range (Clipperton Atoll) and its known small population size (Robertson 2001 estimated that there were ~60,000 adults there in 1998), *Holacanthus limbaughi* exhibited a remarkable genetic diversity (four haplotypes out of six individuals).

The West Atlantic clade included *H. bermudensis* and *H. ciliaris,* which formed two reciprocally monophyletic clades. Here again little intra-specific diversity was found. Different *H. ciliaris* color morphs from NE Brazil (namely 'blue face' and 'yellow face' morphs collected along Ceará State coast) and the Caribbean are very similar (sequences divergence was less than 1%).

Genetic divergence and temporal divergence

A molecular clock could not be rejected for either datasets (Shimodaira and Hasegawa test, p=1.00 for the first dataset and p=1.00 for the second dataset). In the case of the mitochondrial control region, average pairwise sequence divergence between the Western Atlantic and TEP clades was 18.7.%. A thorough review performed by Lessios (2008) on multiple species of geminate pairs of marine fish found that the species that most likely separated due to the rise of the Isthmus of Panama exhibited a divergence range of 9.7 to 22%. These values are consistent with the idea that the rise of the isthmus may have coincided with the divergence of the two Trans-isthmian *Holacanthus* clades.

When analyzing cytochrome b sequences, which are widely used for timing divergences in fishes, the average sequence divergence between the Western Atlantic and the TEP clade was approximately 4.5%. This value again falls within the range of other geminate fish pairs (using the same gene region) that are most likely

to have been separated at the closure of the Isthmus of Panama (3.3 – 4.8%, Lessios 2008).

The average cytochrome b sequence divergence between *H. africanus* and the rest of the *Holacanthus* clade was approximately 12.1%, the divergence between *Pygoplites diacanthus* and *Holacanthus* species was 14.7%. These two values provide the boundaries to estimate the age of the genus *Holacanthus.* Using the calculated rate of divergence based on the Trans-Isthmian geminate clades, this suggests that *Holacanthus* appeared approximately 10.2 to 7.6 million years ago.

Discussion

Phylogenetic relationships, evolutionary history and biogeographic patterns

The genus *Holacanthus* comprises only 7 species, yet it offers unique possibilities in testing biogeographic, ecological and evolutionary hypotheses. The overall monophyly of the genus is confirmed here. Considering the thorough investigation presented by Bellwood et al. (2004), it is unlikely that other unsampled angelfish species would fall within *Holacanthus. Pygoplites diacanthus* was found to be very close and basal to *Holacanthus*. Keeping *Pygoplites* in its own monospecific genus or considering it a basal *Holacanthus* is more of a semantic issue at this stage.

Our data indicate that the genus *Holacanthus* likely originated approximately 10.2 - 7.6 mya. Since the basal species is found in western Africa, *Holacanthus* may have originated from an Indian Ocean invasion, a scenario that has been observed in

other groups, including pygmy angelfishes (Bowen et al. 2006, Rocha et al. 2005b). These dates, however, are different from the proposed timeframe of the split between *P. diacanthus* and *Holacanthus* by Bellwood et al. (2004): 13 Mya (minimum age) or even 23 Mya, which would lead to a Tethyan separation hypothesis (see proposed scenarios in Floeter et al. 2008). In any case, the first split (or invasion with further speciation) within *Holacanthus* is likely to have occurred between the Eastern Atlantic and the Indian Ocean. In contrast, an invasion via the Central American Seaway prior to the closure of the Isthmus of Panama, such as proposed for the wrasse genus *Halichoeres* (Barber and Bellwood 2005; Floeter et al. 2008) is a less likely scenario for *Holacanthus* as the TEP species (subgenus *Plitops*) are relatively recent and not the immediate sisters to *Pygoplites*.

The genus *Holacanthus* has previously been divided into four subgenera. Our data do not support the subgenus *Angelichthys* as it were originally proposed (*H. africanus, H. bermudensis, H. ciliaris*). Our data, however, support a separation of species by geographic regions: the eastern Atlantic species from the western Atlantic clade, and the Atlantic clade from the TEP clade.

Modes of speciation

Holacanthus affords hypothesis testing due to its presence on both sides of known biogeographic barriers. While *H. bermudensis* and *H. passer* were thought to be trans-Isthmian geminate species (Bellwood et al. 2004), our study shows that

they actually belong to multispecific clades that were separated by the closure of the Isthmus of Panama. The genetic divergence between these clades is similar to other known cases of geminate species (e.g. Domingues et al. 2005, Bernardi et al. 2008, Lessios 2008 comments about this issue). Thus the rise of the Isthmus of Panama likely preceded the diversification of those clades in both the W Atlantic and TEP. Vicariant speciation is likely to be responsible for the initial divergence of angelfish among these two neotropical clades.

The close relationship of the species within the Caribbean, and within the TEP, and the presence of hybrids within these species complexes (*H. bermudensis x ciliaris, H. passer x clarionensis*) underscore their evolutionary proximity (Sala et al. 1999). The broad distribution of *H. passer* in the TEP relative to the other two insular species suggests that *H. passer* may have been the ancestral species that gave rise first to *H. limbaughi* and then to *H. clarionensis*. The fact that *H. clarionensis* is almost entirely limited to the Revillagigedo islands, and that *H. limbaughi* only occurs at Clipperton, may suggest that these species arose via peripatric speciation, particularly in the case of the latter, with its extremely isolated population. However, more information is needed on the abundance of *H. passer* at the Revillagigedo Islands; if it is common there then the case for isolation leading to a peripatric origin of *H. clarionensis* weakens. The origin of *H. bermudensis* is more problematic than that of the two insular Caribbean species. Although *H.*

bermudensis is effectively isolated at Bermuda (*H. ciliaris* is very rare there, Smith-Vaniz et al 1999), its geographic range and population is mostly continental and its sister *H. ciliaris* is common throughout the continental range of *H. bermudensis*. A peripatric origination scenario would have to have *H. bermudensis* originating in Bermuda, whose reef fish fauna has been subject to major recent extinctions during ice ages (Smith-Vaniz et al 1999). Parapatric speciation, speciation with geographic segregation but ongoing gene flow (see Rocha and Bowen 2008) provides an alternative mechanism that could account for the origins of *H. bermudensis* as a northern, warm-temperate form separating off from a southern, tropical form (*H.ciliaris*). The distributions of *H .bermudensis* and *H. ciliaris* are similar to those of a labrid for which such parapatric speciation has been proposed (see Rocha et al 2005a.) Parapatric speciation in response to differences between insular and continental environmental conditions could also account for the development of *H. clarionensis* in the Revillagigedo Islands despite ongoing immigration by *H. passer*. Two of the three oldest divergences in *Holacanthus* (the split of *H. africanus* from an apparent Indian Ocean ancestor, the trans-Atlantic split of *H. tricolor* from *H. africanus* and the division of the two transisthmian clades) are also consistent with allopatric speciation scenarios. The geographic distributions of the neotropical species that evolved after the closure of the isthmus of Panama are most consistent with modes of speciation that involve partial or complete allopatry. Therefore at

present there seems to be no reason to invoke purely sympatric speciation as a contributor to the evolution of *Holacanthus*.

Conclusion

Holacanthus, although a small and geographically confined genus (compared to other angelfish genera), joins other angelfishes in providing a very useful model for the study of marine reef fish evolution. The results of our phylogenetic analyses indicate when and how the genus originated, that its 7 species include trans-Isthmian geminate species complexes, and that speciation likely occurred well after the closure of the Central American Isthmus in both the Caribbean and the Tropical Eastern Pacific.

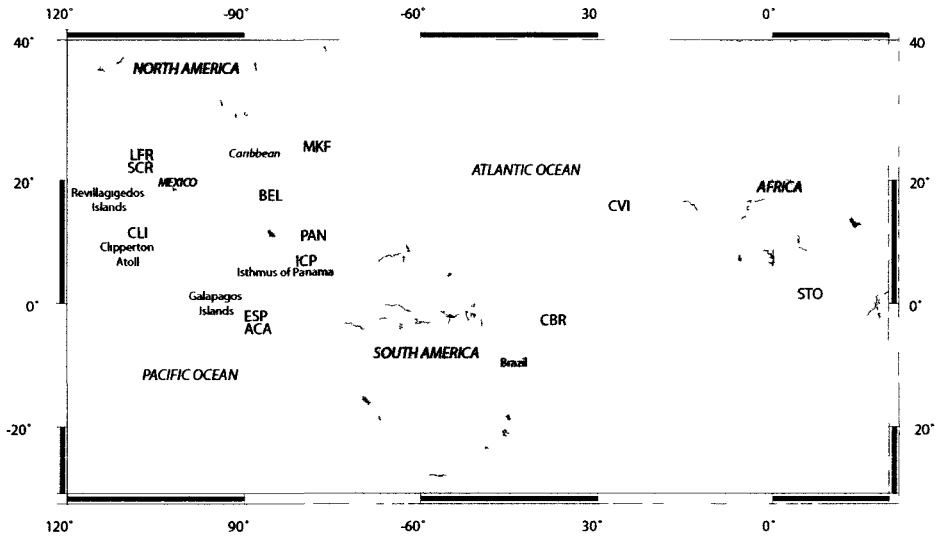

Figure 1.1. Distribution map and sampling locations of *Holacanthus*. Labels refer to sampling locations described in Table 1 (map created by OMC, Martin Weinelt).

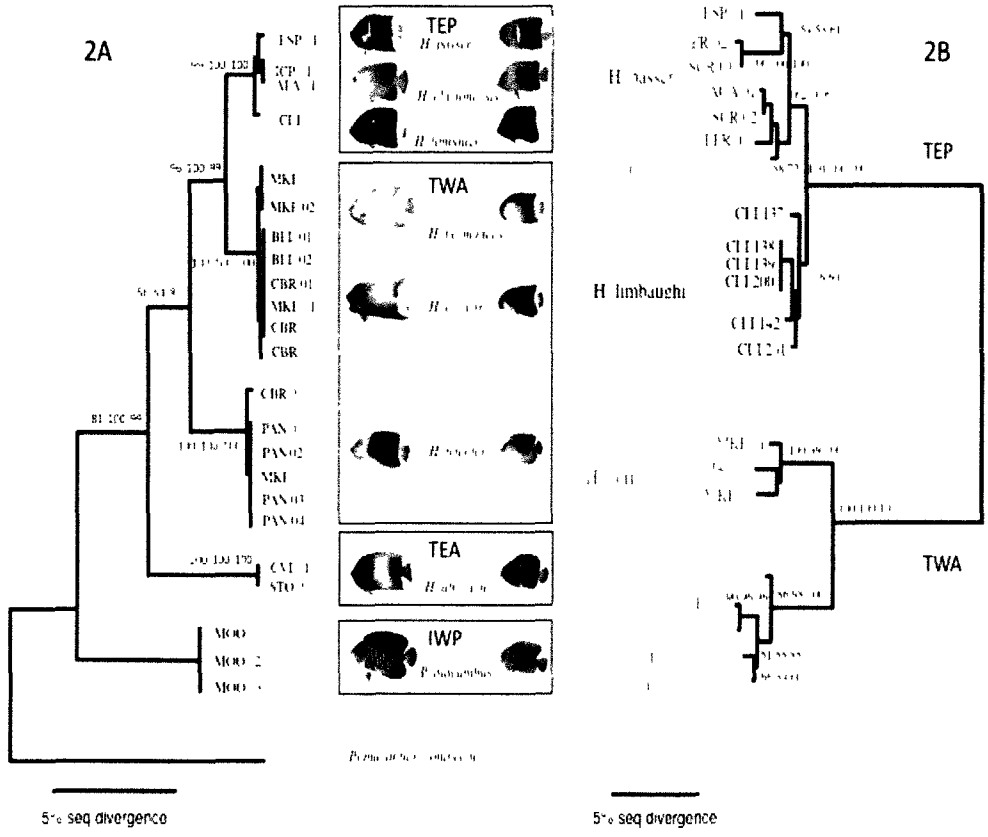

Figure 1.2. Panel 2A. Molecular phylogeny of *Holacanthus* spp. and *Pygoplites diacanthus* based on mitochondrial and nuclear molecular markers (12SrRNA, 16S rRNA, cytochrome b, 1st intron of ribosomal protein S7) using Maximum Likelihood (Maximum Parsimony, and Neighbor-Joining reconstructions resulted in the same topology). *Pomacanthus zonipectus* was used as an outgroup. Labels are described in Table 1.1. Bootstrap support (or consensus value) is shown when above 50%, for the three methods used, ML, NJ, and MP, in that order. Percent sequence divergence is represented on the scale bar. Panel 2B. Molecular phylogeny of 5 species of *Holacanthus* that reside in the TEP and TWA. based on 410 bp of the mitochondrial control region using Maximum Likelihood. Labels are described in Table 1.1 Bootstrap support (or consensus value) is shown when above 50%, for the three methods used, ML, NJ, and MP, in that order. Percent sequence divergence is represented on the scale bar.

Species	Sampling site	n	Label
Holacanthus africanus	Cape Verde Islands	1	CVI
	São Tomé	1	STO
Holacanthus bermudensis	Marathon Key, Florida	3	MKF
Holacanthus ciliaris	Turneffe Atoll, Belize	2	BEL
	Regular morph, Ceará State, Brazil	1	CBR1
	Blue morph, Ceará State, Brazil	1	CBR2
	Yellow morph, Ceará State, Brazil	1	CBR3
	Marathon Key, Florida	1	MKF
Holacanthus clarionensis	Aquarium of the Pacific	1	APA
Holacanthus limbaughi	Clipperton Atoll	6	CLI
Holacanthus tricolor	Ceará State, Brazil	1	CBR
	San Blas, Panama	4	PAN
	Marathon Key, Florida	1	MKF
Holacanthus passer	Isla Contradora, Panama	1	ICP
	Santa Cruz, Mexico	2	SCR
	Los Frailes, Mexico	2	LFR
	Espanola, Galapagos, Ecuador	1	ESP
	Santa Cruz, Galapagos, Ecuador	1	ACA
Outgroups			
Pomacanthus zonipectus	Isla Contradora, Panama	1	RCP
Pygoplites diacanthus	Moorea, French Polynesia	3	MOO

Table 1.1. Collection localities for *Holacanthus spp.*, *Pygoplites diacanthus*, and the outgroup, *Pomacanthus zonipectus*. Columns represent the number of individuals included in the study, and the abbreviations used in Figure 1.1 and 1.2.

Chapter 2: A Comparative Phylogeographic of Tropical Eastern Pacific and Western Atlantic Shore Fish

Abstract

In this study I examine how environmental factors beyond physical and ecological constraints influence the movement of individuals between populations. While lack of specific oceanographic and ecological information makes this difficult, understanding how these factors affect population connectivity is essential for understanding how populations adapt to changing environments. I conducted genetic comparison of marine fish across two taxonomic levels to address two key questions: 1) whether environmental factors play a significant role in dispersal capabilities in two sister species residing in two contrasting ocean environments - the Tropical Eastern Pacific (TEP) and the Tropical Western Atlantic (TWA). These oceans differ drastically in the amount and type of habitat available for settlement as well as oceanographic features (e.g. current stability, water temperature, and tidal changes). By choosing sister species that are found on either side of the Isthmus and share similar ecological requirements, I minimized phylogenetic variation to test whether differences in population structure are caused by differing oceanographic environments. Here I performed a comparison between two taxonomic families that occupy the same region, but differ in life history

characteristics. *Abudefduf sp.* are benthic spawners and are expected to have fewer opportunities for dispersal than *Lutjanus sp.* that are broadcast spawners, have pelagic eggs, and form spawning aggregations. By choosing species that have overlapping distributions, I minimized environmental variation in order to test whether ecological traits (e.g. spawning behavior) influence levels of gene flow. For both comparisons I examined differences in two mitochondrial markers: ATPase subunit 8 and the Dloop. In the first question I found the TEP species exhibited more genetic differentiation than those in the TWA, suggesting that oceanographic differences affect population connectivity. In the second question I found that *Abudefduf* showed more genetic differentiation than *Lutjanus*, suggesting life history characteristics (spawning behavior and egg buoyancy) play a significant role in dispersal. Together, these findings suggest that both environmental processes and life history characteristics influence gene flow.

Introduction

Population connectivity via dispersal (gene flow) is essential for maintaining genetic variation within populations (Avise 2004). When climatic events inhibit dispersal, populations can become isolated, leading to genetic drift that ultimately may inhibit the genetic variation necessary to adapt to a changing environment (Palumbi 1992, 1994). As isolates go extinct the survival of the entire species is

threatened. Therefore, it is essential to understand the underlying factors responsible for dispersal.

In the marine realm, understanding population structure poses unique challenges (Palumbi 1992, 1994). With few known physical barriers and limited oceanographic information, it is difficult to determine how physical and oceanic processes may influence marine dispersal (Bermingham 1996; Hastings 2000; Lessios 2008; Riginos 2005; Shulman and Bermingham 1995). Studies that have been successful in addressing the population structure of marine fish have used well-known vicariant events (Riginos 2005, Waters et al. 2004, Bernardi et al. 2003) or regions where biogeographic breaks have been documented based on marine composition (Hastings 2000). Other studies have used oceanographic features, such as currents, to predict dispersal movements (Muss et al 2001). Generally, an understanding of the geological history and oceanic features of the study region is necessary to assess the dispersal capabilities of marine organisms.

Another challenge in studying the population dynamics of marine species lies in the complex life histories of many marine organisms (Warner and Palumbi 2003). Most marine fish have a pelagic larval phase, which is thought to provide the primary opportunity for dispersal (Leis 1991). It is still unclear whether pelagic larval duration (PLD), which is the time spent in the open water column, is the most important factor determining dispersal capabilities (Shulman and Bermingham

1995). Many studies have found no relationship between PLD and gene flow (Shulman and Bermingham 1995; Warner and Palumbi 2003). However, there are other life history traits such as spawning behavior and egg buoyancy that may influence dispersal capabilities (Shulman and Bermingham 1995). Fish that are known to travel to aggregation sites to spawn are more likely to come in contact with individuals from other populations compared to fish that mate within their home range. This ecological behavior may facilitate gene flow (Sala et al 2003). Marine organisms that are broadcast spawners, organisms that release egg and sperm into the water column, have positively buoyant eggs that are pelagic which may assist with dispersal, compared to benthic spawners that have negatively buoyant eggs that settle on surrounding substrate (Munro 1973; Sala et al 2003; Victor and Wellington 2000).

The goal of this study is to understand how environmental and life history factors may affect dispersal capabilities. The aims of the research are two fold: 1) How do environmental factors affect dispersal capabilities of marine fish? 2) How do ecological requirements affect dispersal capabilities of marine fish? I will answer these two questions by comparing two different taxonomic levels. To address the first question I will compare sister species that reside in two contrasting environments. A comparison of sister taxa will allow me to compare species with minimal evolutionary differences (compared to distantly related taxa) in order to

determine whether differing oceanic environments create differences in levels of gene flow. To test the second question I will compare species from different taxonomic families with different ecological requirements that reside in the same region. By comparing species that have overlapping distributions, I will be able to minimize environmental variation in order to determine whether life history differences affect gene flow differently among taxonomic families. To address these two questions, a combination of phylogenetic and population genetic methods will be used to compare the population history and genetic structure for each species.

Background

Study Region

The region encompassing the Tropical Eastern Pacific (TEP) and Tropical Western Atlantic (TWA) serves as a perfect platform to study evolutionary processes of marine biota, due to its unique landscape and rich geological history. These oceanic regions were separated by the closure of the Central American Isthmus approximately 3 million years ago (Coates and Obando 1996, Jackson et al 1996). In middle Miocene (16-15mya), the tropical Atlantic and eastern Pacific were separated by a few small islands. During the late Pliocene, parts of Panama began to rise, and by 3 Mya, small remaining water canals were filled with sediment which completed the formation of the Isthmus (Coates and Obando 1996; Coates 1996; Jackson et al 1996).

The rise of the Isthmus had dramatic effects on the physical environment of both newly formed ocean basins. The Atlantic North Equatorial Current that flowed in a westward direction between the two oceans was deflected northward increasing the movement of the Gulf Stream (Shulman and Bermingham 1995). Compared to the TEP, TWA is noted for having stable currents. Unstable currents are a defining feature in the TEP due to the periodic occurrence of El Nino Southern Oscillation events, the movement of warm currents into cold water, which affects marine distributions and disrupts ongoing oceanic processes (Robertson and Allen 2002). In addition, the tidal range on the Tropical Eastern Pacific can range up to 6 meters, while the tropical western Atlantic generally has less than 0.5 meter tidal fluctuations. As a result the Caribbean water is warmer and saltier in comparison to the TEP (Shulman and Bermingham 1995). Another phenomenon that has led to the dramatic differences observed between the two oceans is the occurrence of upwelling events in the TEP (Muss et al 2003; Robertson and Allen 2002). In the Eastern Pacific, Northern trade winds continuously push surface water away from the coast initiating upwelling events which replace surface water with nutrient rich water (Robertson and Allen 2002). Upwelling events have led to the scarcity of coral reefs compared to the TWA, but has led to an overall high amount of primary productivity in the TEP that can support a diverse biota (Bermingham et al 1997; Coates and Obando 1996; Coates 1997; Muss et al 2003). In addition to climatic

instability, the TEP also has fewer islands and long stretches of uninterrupted sandy shores (Hastings 2000). The rise of the Isthmus has reshaped the physical environment of TEP and TWA oceans, the Atlantic Ocean is a closed basin with stable currents, while the Pacific is an open basin with unstable currents, variable tide oscillations, and upwelling (Figure 2.1). The environmental differences found between the TEP and TWA provides an excellent opportunity to study how environmental factors may influence dispersal capabilities of marine biota.

Along with the creation of two differing oceans, the rise of the Isthmus severed preexisting marine distributions into two populations starting them on independent evolutionary trajectories (Bermingham et al 1997; Jordan 1908; Lessios 2008). There are many pairs of sister species, known as geminate species, found on either side of the Isthmus that share similar morphological features, life histories, and ecological niches. The term geminate species was coined by Jordan (1908), when he observed morphological similarities between marine species that reside in both the TEP and TWA. The list of morphologically identified geminate species presently exceeds 100 taxa (Bermingham et al 1997, Lessios 2008). Ongoing research on putative geminate pairs based on molecular phylogenies, has revealed a few cases of misidentification (Alva-Campbell et al 2010, Rocha et al 2008). Species that share morphological traits have been found to be more related to taxa from the opposite ocean that shares fewer traits (Rocha et al 2008). Misidentification can

also occur when geminates diverge from one another and then undergo another round of speciation, this will lead to "geminate clades" on either side of the Isthmus (Alva-Campbell et al 2010). The use of molecular techniques proves to be crucial when categorizing pairs of species as geminates.

The rise of the Isthmus of Panama is one of a few marine vicariant events that are well documented and which provide a unique opportunity to study molecular clocks within a recent time frame of earth history (Bermingham 1997). Molecular clocks are useful for estimating dates of divergence between similar taxa in order to time speciation events and within taxa to make phylogeographic inferences (Avise 2004; Zukerkandl and Pauling 1965). Most molecular clocks are calibrated using fossil records, but some studies have been successful with using vicariant events estimated using geological information to calculate timing of divergences (Avise 2004; Banford et al 2004; Bermingham 1997; Knowlton 1993; Lessios 2008; Zukerkandl and Pauling 1965). The Isthmus of Panama serves as a perfect natural experiment to test the molecular clock because its geology is well known (Bermingham 1997; Knowlton 1993; Lessios 2008; Zukerkandl and Pauling 1965).

Study organism:

For this study, I chose two pairs of proposed geminate species from two different families: snappers (*Lutjanus argentiventris vs. L. apodus)* and damselfish

(*Abudefduf troschelii vs. A. saxatilis*). These fish are relatively abundant, endemic to the study region, and are associated with rock or reef habitat (Robertson and Allen 2002). The taxa found in the TEP, *L. argentiventris* and *A. troschelii*, have completely overlapping distributions, as well as those found in the TWA, *L. apodus* and *A. saxatilis*. In addition *Lutjanus sp.* are also valuable to local and commercial fisheries.

I chose to study these two taxonomic families because although PLD is similar, the two taxonomic groups differ in other related life history traits (Table 2.1). Damselfish are benthic spawners and have negatively buoyant eggs that fall down to the surrounding substrate (Munro 1973; Victor and Wellington 2000). Snappers are known to travel to spawning aggregations where spawning takes place in the open water column (Sala et al 2003; Zapata 2002).

This behavior is called broadcast spawning; fish that exhibit this behavior commonly have pelagic eggs that float in the water column for a certain amount of time (Munro 1973; Victor and Wellington 2000). This stage of egg movement may facilitate gene flow.

Using species that are closely related can minimize life history variation and specifically test whether differences in oceanic processes in the TEP and TWA lead to differences in genetic signatures. I performed a comparison among taxonomic families that share geographic ranges and are assumed to have undergone the same historical pressures, to test whether differences in ecological requirements influence

dispersal capabilities differently across taxonomic families. I will reveal whether differences in life history traits, such as spawning behavior and egg buoyancy, influence dispersal capabilities differently among taxonomic families (Munro 1973; Sala et al 2003; Victor and Wellington 2000; Zapata 2002).

This study will estimate the evolutionary history for each species, which includes: an assessment of the levels of gene flow, directionality of gene flow, levels of inter- and intra-population genetic variation, mutation rate, and timing of divergences. Compiling these results will allow me to reconstruct the phylogeography of each species in order to address the two main questions for this study: *How do environmental factors play a significant role in dispersal?* and *How do life history factors play a role in dispersal?*

Methods

Sampling and DNA preparation

Sampling locations and sample sizes are shown in Figure 2.2.1and listed on Table 2.1. Fish were collected by hand nets, spear fishing, and at local fish markets. Tissue samples were cut from the caudal fin and placed in 95% ethanol for preservation. In the laboratory, DNA was prepared from 75 to 150 mg of muscle by proteinase K digestion in lysis buffer (10mM Tris, 400 mM NaCl, 2 mM EDTA, 1% SDS overnight at 55C. This was followed by purification using chloroform extractions and alcohol precipitation (Sambrook et al. 1989). Amplification of the 5' hypervariable

mitochondrial D loop region was performed with primers: ProL (CTACCTCCAACTCCCAAAGC) and tPhe (GTGTTATGCTTAGTTTAAGC) for *Abudefduf* sp.; ProL and STDloop (CTGGAYAGAYRGCACGGC) for *Lutjanus* sp. In addition to the Dloop region, mitochondrial ATPase8, tRNA, and a partial segment of cytochrome oxidase subunit II was sequenced using the following primers: ATPase 8 (AAAGCRTYRGCCTTTTAAGC) and Co3 (GTTAGTGGTGACCGGCTTGGRTC). Both regions used a cycling profile of 95 C for 1 minute, 52 C for 1minute and 72 for 1 minute for 35 cycles. After purification following the manufacturer's protocol (ABI, PerkinElmer), sequencing was performed on an ABI PRISM 3700 DNA automated sequencer (Applied Biosystems).

Phylogenetic and population genetic methods

Sequences were aligned using Sequencher 4.1 (Gene Codes Corporation), and checked manually. A model of evolution for each dataset was constructed using the program MODELTEST version 3.6 (Posada 1998). Approximately 56 models of nucleotide sequence evolution were evaluated using MODELTEST, the models selected for each data set are described in Figure 2.2. Shimodaira and Hasegawa test implemented by PAUP* (Phylogenetic Analysis Using Parsimony, version 4.0; Swafford 1998) was used to test whether topological differences existed between both the markers used in this study as well as to determine if there were differences in topologies with or without an enforced molecular clock (Shimodaira and

Hasegawa 1999). Maximum likelihood estimates per gene were used to obtain a maximum likelihood ratio without the input of MODELTEST parameters. Because no significant differences were found between the Dloop and ATPase8 topologies, both loci were combined for further analysis. Using the model of evolution suggested from MODELTEST, phylogenetic relationships were measured using Maximum Likelihood (ML), Maximum Parsimony (MP) and Neighbor-Joining methods (NJ). Both MP and NJ topologies were constructed with the program PAUP. Topological confidence was evaluated for MP and NJ with 1000 bootstrap replicates. The program GARLI (Zwickl 2006) was used to construct a Maximum Likelihood tree making the necessary adjustments to implement the model of evolution proposed by MODELTEST. The search was set to terminate when the likelihood score remained constant for 20,000 consecutive generations. The tree with the highest likelihood score out of 10 independent runs was retained. Bootstrap values were added to the topology based on 100 replicates performed in GARLI. Outgroups within the same genus were included in the topology. Average pairwise sequence divergence was calculated among individuals within the same species and between geminate pairs using PAUP for each marker and then for both markers combined.

For each population the number of haplotypes (H_n), nucleotide diversity (π), and haplotype diversity (h) was estimated using the software package DnaSP (Rozas et al 2003). Minimum spanning networks based on haplotype differences were

constructed using the program HAPSTAR. To detect genetic structure for each dataset, pairwise comparisons between populations (ϕst values) were calculated in ARLEQUIN 3.5.1.2 (Schneider et al 2000). A hierarchical analysis of molecular variance (AMOVA) was performed using ARLEQUIN to test for significant regional subdivision. The program MIGRATE (Beerli 2006) uses coalescent theory to estimate the number of migrants and the direction of migratory movements. To estimate the number of migrants (Nm), I used a Bayesian method implemented in MIGRATE consisting of 1 long chain, a sampling increment of 100, the recorded steps in a chain was set at 100,000, and the option of heating was turned on. Analysis of each dataset was performed 10 times to generate a mean and standard deviation.

Historical demography was evaluated with the program FLUCUATE 1.4 (Kuhner et al. 1998). To determine population expansion or contraction, growth (g) was estimate. To describe effective population size the parameter θ was calculated with two models, one with growth kept constant (θ_c) and the other with option for population growth (θ_v). All three parameters were estimated by running 10 replicates which generated a mean value and standard deviation to ensure convergence of parameters. Analysis for each run was performed with 10 short Monte Carlo chains of length 20000 followed by 2 long chains, with a sampling increment of 20. Coalescence time was estimated under the assumption that coalescence is reached when the population size was reduced to 1% of its present-

day value (Wares and Cummingham 2001). Two parameters need to be estimated to obtain coalescence time: mutation rate (μ) and generation time (g). Under the assumption that these gene are evolving in a neutral manner, estimates of substitution rate will be used to infer mutation rate. To calculate mutation rate average pairwise divergence between geminate species was subtracted from the average pairwise divergence within a species to account for intraspecific polymorphism and then divided by generation time (g). The time at reproductive age, for each species was acquired from the literature (*Abudefduf species* =1yr; *L. apodus* = 2yrs; *L. argentiventris*= 4yrs) this measurement was divided by 3.5myr to obtain the number of generations since the proposed closure of the Isthmus (Robertson 1989; Martinez-Andadre 2010).

Life history information

Life history characteristics of egg buoyancy (buoyant or non-buoyant) and spawning behavior (benthic or broadcast) (aggregations or not) were acquired from previous studies. Abudefduf species are benthic spawners that do not form aggregations and have negatively buoyant eggs that settle to the bottom (Munro 1973; Sala et al 2003; Victor and Wellington 2000; Zapata 2002). In contrast *Lutjanus* species form spawning aggregations and are broadcast spawners with positively buoyant eggs. *Lutjanus* species have a slightly longer PLD of 23 days than *Abudefduf* species (PLD= 18.2days) (Victor and Wellington 2000; Zapata 2002). A

comparison between these life history traits and gene flow among all species will be performed to determine if there is a relationship between any of the life history components and gene flow.

Results

DNA Sequence Analysis

Five populations of *Abudefduf sp.* and four populations of *Lutjanus sp.* were included in this study. Both the 5'-end portion of the mitochondrial Dloop and ATPase8 was sequenced for each individual, fragment sizes can be found in Table 2.4. A Shimodaira and Hasegawa test found no significant differences between the mtDloop and ATPase8 topologies (P=1.000); therefore further analysis will include both markers combined unless otherwise stated. The number of parsimony informative characters per dataset is given in Table 2.4.

Phylogenetic inference

The three methods of tree reconstruction (MP, NJ, ML) resulted in identical topologies for each major taxon. A neighbor-joining tree was constructed based on the Kimura 2 Parameter genetic distance method, which weighs tranversions and transitions. Maximum likelihood topologies along with the bootstrap values are presented for each data set in Figure 2.2. Tree topologies between geminate pairs were combined to compare genetic divergences. The topology belonging to *Abudefduf sp.* included *Abudefduf troschelii* from the Tropical Eastern Pacific and its

proposed geminate, *A. saxatilis,* from the Tropical Western Atlantic (Figure 2.2). Concordant with the results of previous studies (Bermingham et al 1997; Jordan 1908; Shulman and Bermingham 1996), bootstrap values support the grouping of *Abudefduf troschelii* as sister to *A. saxatilis* (Figure 2.5). All pairwise differences between the two species were calculated in PAUP and this term was subtracted from the average pairwise differences found within species to correct for intraspecific polymorphism. The corrected average pairwise difference between *A. troschelii* and *A. saxatilis* was estimated to be 4.4% (Table 2.5).

The topology belonging to *Lutjanus sp.* included *L. argentiventris* from the Tropical Eastern Pacific and its proposed geminate, *L. apodus* (Figure 2.2). Another proposed geminate pair belonging to the family Lutjanidae was included in the analysis for comparison, *L. novemfaciatus* from the Tropical Eastern Pacific, and *L. cynopterus* from the Tropical Western Atlantic. *L. jocu* was included as an outgroup. The topology of *Lutjanus species* differed from results of previous studies (Bermingham 1997, Lessios 2008). Based on the Maximum Likelihood topology for *Lutjanus* species presented in this paper, *L. argentiventris* is sister to *L. jocu* not *L. apodus,* its proposed geminate (Figure 2.2 and 5). Unlike *Abudefduf sp.*, no extensive phylogeny has been constructed for the family Lutjanidae. *L. argentiventris* and *L. apodus* were thought to be geminate species based morphological and ecological synapomorphies, but they genetically differ by 13.2%

sequence divergence. The corrected average pairwise difference between *L. argentiventris* and *L. jocu* is 3.8%, lower than the estimated genetic divergence for *Abudefduf* geminate sp. (4.4%) and between *L. novemfaciatus* and *L. cynopterus* (5.3%)(Table 2.5). *Lutjanus jocu* is found in the Tropical Western Atlantic. Based on these confounding results I will consider *L. argentiventris* and *L. jocu* as geminate species. However, I will continue to compare the population structure of *L. apodus* with *L. argentiventris* due to insufficient sample numbers for *L. jocu*.

Population Genetic Inferences

Haplotype diversity was high among all datasets, making a minimum spanning network difficult to reconstruct. To simplify the network I collapsed the dataset by only using parsimony informative characters of one mitochondrial marker (ATPase8). This process allowed me to build a network with a manageable number of haplotypes in (Figure 2.3). The number of haplotypes based on parsimony informative characters in the ATPase gene, differed among taxa (*L. argentiventris* H_n=22; *L. apodus* H_n=8; *A. troschelii* H_n=35; *A. saxatilis* H_n=27). Two trends were evident- *Lutjanus sp.* had a lower number of haplotypes than *Abudefduf sp.* and species found in the TWA had fewer haplotypes than geminate species found in the TEP. Haplotype networks were also built to display haplotype frequency and the number of mutational differences found between haplotypes (Figure 2.3). In all data sets, two to five haplotypes dominated the network, with many singletons that

differ from the dominant haplotype by one single-step mutation. This suggests either high gene flow or a recent population expansion. Geographic distributions of haplotypes for each population were constructed using pie charts to show haplotype numbers and frequencies per population (Figure 2.4a,b). Private alleles, alleles found in only one population, are shown in black and ranged from 3 to 11 private alleles per dataset. The lowest number of private alleles (H_n =3) was found in *Lutjanus apodus* populations which may be caused by high gene flow. The highest number of private alleles were found in *Abudefduf species*, which may be an indicator of low gene flow. Overall fewer private alleles were found in *Lutjanus* species than *Abudefduf* species.

Gene flow analysis comparing genetic differentiation among populations identified various degrees of structure among taxa (Table 2.2). *Lutjanus* species showed no significant genetic structure with pairwise population comparisons (Table 2.2 c,d). In comparison, *Abudefduf saxatilis* detected one instance of population subdivision (Table 2.2b), while *A. troschelii* showed many instances of population subdivision (Table 2.2a). AMOVA was used to detect genetic differentiation between groupings. Prior information using the geographic history of the region was used to make informed groupings to test with AMOVA. *Lutjanus* species showed no genetic structure among all possible groups, all variation found was attributed to within population differentiation. Differences between groups were detected for

Abudefduf species. *A. troschelii* demonstrated a substantial amount of variation between Mexico and non-Mexican groups (3.8%), but this difference was not significant (Table 2.3). *A. saxatilis* revealed a significant amount of genetic structure between groups consisting of populations from BRA and NPR versus MKF, BEL, and SPB (Table 2.3). Group divisions that showed a substantial amount of among group variation were implemented into the program MIGRATE to further analyze the number of migrants and direction of migratory movements.

Patterns of gene flow, including the number of migrants and direction of migration events, varied among taxa. A north clockwise trend dominated all other migratory events in *L. apodus* (Figure 2.4). *A. saxatilis,* which has an overlapping distribution with *L. apodus* displayed a panmictic form of migration. The number of migrants and patterns of migratory movements were more frequent in the Tropical Eastern Pacific species. Both *L. argentiventris* and *A. troschelii*, showed a larger number of migrants (Nm=950; Nm=500, respectively) from the Mexican populations to non-Mexican populations than from the other direction (Nm=36; Nm=100)(Figure 2.4 a,b).

Molecular Clock Estimates

No significant topological differences with or without an enforced molecular clock were found using a Shimodaira and Hasegawa test (P=1.000); therefore a molecular clock can be used with all the data sets. To calculate the divergence

rates, I used the average pairwise differences between the geminate pairs along with the timing of the closure of the Isthmus, which was suggested to occur 3.1-3.5 myr ago. The average pairwise sequence divergence between TEP and TWA geminate pairs slightly varied (Table 2.5). Dividing the average pairwise difference by the proposed closure of the Isthmus of Panama (3.1-3.5 myr), the calculated rate of divergence for the combined markers between species pairs varied between 1.1-1.7% per million years (Table 2.5). A UPGMA tree based on a subset of samples included in the analysis was constructed along with a time scale based on the divergence of the two geminate species and under the assumption that these geminate species diverged at the time of closure (3.1-3.5myr) (Figure 2.5). The topologies for *Abudefduf* and *Lutjanus* are scaled accordingly for visual comparisons. Based on the topology it is estimated that *L. apodus* diverged from the other *Lutjanus sp.* used in this study approximately 7 mya.

Historical Demography

Growth rate (*g*) for each dataset was estimated using the program FLUCTUATE (Table 2.4). Growth rate was positive (suggesting population expansion) for each species except for the species *L. apodus* which was estimated to have 0 growth rate. A value of 0 for growth rate implies that the population is stable, with no sudden expansions or contractions. The study of geminate species across the Isthmus of Panama provided a unique opportunity to calculate mutation rate (μ)

that can be used to infer historical demography. To obtain mutation rate (μ), I estimated generation time (g) which was the time of divergence (3.5 myr) divided by the age at reproduction, acquired from previous studies. The average pairwise divergence was divided by two to estimate the substitutions on a single branch and then divided by generation time to obtain μ. Under the assumption that coalescence is reached when the population size was reduced to 1% of its present-day value, the formula $N_t=\theta e^{-(gt\mu)}$ was further simplified to $t= 1/(g\mu \ln 100)$, to solve for time of coalescence (t). The time of coalescence for each data set varied between 24,120- 20,355,633 years ago (Table 2.4). The time of coalescence for three of the four species was less than 3.5myr (the approximate closure of the Isthmus). Because *L. apodus* had such a small growth value (*g*), the time of coalescence was inflated (~20mya). Unlike the other species used in this study, *L. apodus* was found not to have a geminate, suggesting that this species may not be likely to expand or diversify as quickly as the other species.

Two values of θ were calculated, first under a model that forced the population to stay at a constant growth (θ_c) and then under a model that allowed the population to expand or contract (θ_v). These parameters are valuable in comparing effective population sizes, given approximately equal mutation rates among lineages. According to the results listed on Table 2.4, when growth is constant (θ_c), both species in the tropical Eastern Pacific have a higher θ_c value than

those species that reside in the TWA. *Lutjanus apodus* has a lower θ value than all the other species. When the population is allowed to grow (θ_v), the same pattern is found, *L. apodus* has a lower θ value, suggesting that its effective population size is smaller than the other species.

Discussion

Genetic Structure

I found high genetic variation within populations for both the Tropical Western Atlantic species and Tropical Eastern Pacific species examined for environmental effects. For species examined for ecological effects I found low but significantly nonzero levels of genetic differentiation between populations of *Abudefduf* species, and no genetic differentiation between populations of *Lutjanus* species. In addition, haplotype diversity was high among all taxa; however nucleotide diversity was low. These observations are highly characteristic of marine biota that have high dispersal rates (Bowen et al 2006).

Shallow genealogies were found for all four species (Figure 2.2). The haplotype networks displayed closely related haplotypes, along with few haplotypes with high frequency spawning many single haplotypes. *Abudefduf* species had a larger number of haplotypes and private alleles indicating lower levels of gene flow concordant with the AMOVA results. Shallow genealogies with many singletons along with widely distributed haplotypes are characteristic of high levels of gene

flow, recent expansion events, or frequent historic population fluctuations.

The genetic partitions in *A. troschelii* correspond exactly to the biogeographic provinces of the Pacific described by Briggs (1974). *Abudefduf troschelii* exhibited the largest amount of genetic differentiation (Fct = 0.037, p=0.11) between Mexico populations and populations from Panama and Peru (Table 2.3b). This biogeographic break may be a result of a long stretch of sandy coastline (~1200km) separating two biogeographic regions with rocky habitat, the Mexican and Panamic Province (Briggs 1974). Therefore in the case of *A. troschelii*, the expansive stretch of unsuitable coastal habitat, may be an effective barrier to dispersal. No such pattern was found for *Lutjanus argentiventris* which shares the same distribution and habitat requirements as *A. troschelii*. Thus, while important for *A. troschelii*, habitat fragmentation along the coastline does not appear to be an effective barrier to dispersal among *L. argentiventris* individuals.

I found significant differences between populations of *Abudefduf saxatilis* from Brazil and Puerto Rico and populations belonging to Panama, Belize, and Florida (Table 2.3), suggesting low gene flow between these two regions. Although there is no evidence of biogeographic breaks in this region, this genetic break may best be explained as a result of oceanographic currents in the TWA. The Caribbean current, which flows Northward from Venezuela into the Caribbean and back around to Florida may explain migratory movements between Panama, Belize, and Florida

(Figure 2.1). I also found evidence of migratory movements between Brazil and Puerto Rico populations using Migrate analysis, although the movement of individuals was bidirectional with the northward trajectory dominant (Figure 2.5), again consistent with currents found in the TWA where the North Equatorial Current from Africa is deflected northward from Brazil towards the outer Caribbean (Figure 2.1). In MIGRATE results for *Lutjanus apodus,* I found evidence of panmixia or random gene flow.

Do environmental factors play a significant role in dispersal?

By comparing species that are closely related I was able to reduce the variation attributed to life history differences and test for regional effects on population structure. Because the TWA is a closed basin with stable oceanic features, species that inhabit this region should exhibit less gene flow than their sister counterpart that resides in the TEP where the instability of currents and tides may facilitate dispersal movements. I found the opposite in *Abudefduf sp. A. troschelii,* resident of the TEP, exhibited a larger degree of genetic structure than *A. saxatilis from the TWA,* suggesting *A. saxatilis* has a higher level of exchange between populations than *A. troschelii.* This pattern could have been the result of favorable conditions in the Caribbean facilitating dispersal in comparison to unstable conditions found in the TEP that may hinder dispersal.

Lutjanus sp. showed no difference in population structure between the two regions. This result suggests relatively high levels of gene flow for both species, suggesting that there are no effective barriers to dispersal in both the TEP and TWA for these species. High levels of gene flow in *Lutjanus sp.* compared to *Abudefduf sp.* may be a result of differences in life history traits between the two families (see below).

Do life history factors play a role in dispersal?

A comparison among taxonomic families that reside in the same oceanic environment and are assumed to have been exposed to similar past and present environmental events, enabled me to test whether intrinsic differences in life history factors influence dispersal movements. Both taxonomic groups show differing levels of gene flow - *Lutjanus sp.* demonstrate high levels of gene flow, while *Abudefduf sp.* have lower levels. These differences in levels of gene flow may be a result of differences in life history characteristics. *Lutjanus sp.* travel to aggregation sites to spawn, while *Abudefduf* individuals spawn within the proximity of their home range (Sala et al 2003; Zapata 2002). Once spawning occurs, *Abudefduf sp.* eggs settle to the floor where the male protects them until hatching takes place (Munro 1973). In *Lutjanus sp.* there is no parental investment after spawning events; eggs are pelagic and vulnerable to predation. Along with a slightly longer PLD, the life history traits possessed by *Lutjanus sp.* provides many more opportunities for dispersal than

Abudefduf individuals. Based on the results of this study, differences in life history traits contribute to differences in genetic signatures.

Historical Demography

Past population expansions were found for all species except for *L. apodus*. This result implies that *Abudefduf species* and *L. argentiventris* populations are unstable and likely to expand quickly (Kuhner et al 1998). *A. troschelii* had the highest rate of population growth (N=443) as well as the highest value of theta indicating a high effective number (Table 2.4). For populations belonging to *Lutjanus apodus,* zero growth was found. This result suggests that *L. apodus* has a stable population size (Kuhner et al 1998). *L. apodus* may be less likely to diversify, an observation that may support the lack of a geminate counterpart for *L. apodus* opposite of the results found for the other species used in this study that revealed a positive growth value and a geminate counterpart.

Time of coalescence for *L. apodus* was inflated due to a low growth parameter. Time of coalescence ranged from 1myr-20myr (Table 2.4). This value is well above the projected date of the closure of the Isthmus (3.5 mya). This result implies that L. apodus most likely originated well before the final closure of the Isthmus of Panama. *L. argentiventris* and *Abudefduf* species had significantly lower coalescence times (16,529- 79,783 years); with all species diverging after the closure of the Isthmus.

Molecular Clock Estimates

Three pairs of trans-Isthmian geminate pairs were used to calibrate a molecular clock. The genetic divergences between the three geminate pairs were surprisingly similar to one another, suggesting that all geminate pairs diverged from one another roughly around the same time (Figure 2.5, Table 2.5). Of interest is that *L. apodus* is not sister to *L. argentiventris*. Based on morphological similarities it is understandable why these two species were thought to be geminate species (Bermingham et al 1997; Jordan 1908). However, the topology suggests that *L. apodus* and *L. argentiventris* are 13% divergent, four times more divergent than *L. argentiventris* and *L. jocu* (3.8%). Based on similar mutation rates found with the other geminates used in this study, *L. argentiventris* and *L. jocu* are geminate species. A detailed phylogeny of the family Lutjanidae needs to be performed to determine whether *Lutjanus apodus* has a geminate counterpart.

Conclusion

This study incorporated ecology, oceanography, and molecular tools, to identify the environmental and ecological forces responsible for dispersal in marine. My comparative approach among fish that reside in the TEP and TWA provides regional information that can be applied to the genetic composition of other marine organisms that inhabit the same region. In particular, my study demonstrates that insight into the environmental and life history context of the organism proves useful

when examining patterns in genetic composition and gene flow. Habitat fragmentation and oceanic currents likely affects dispersal and flow among benthic spawners more than broadcast spawners that form spawning aggregations. Including these factors in our understanding of the factors governing population structure in marine organisms is crucial to addressing fundamental questions about evolutionary and ecological processes affecting life in the oceans. From a conservation perspective, this study helps inform our understanding of how marine organisms will genetically adapt to a changing environment.

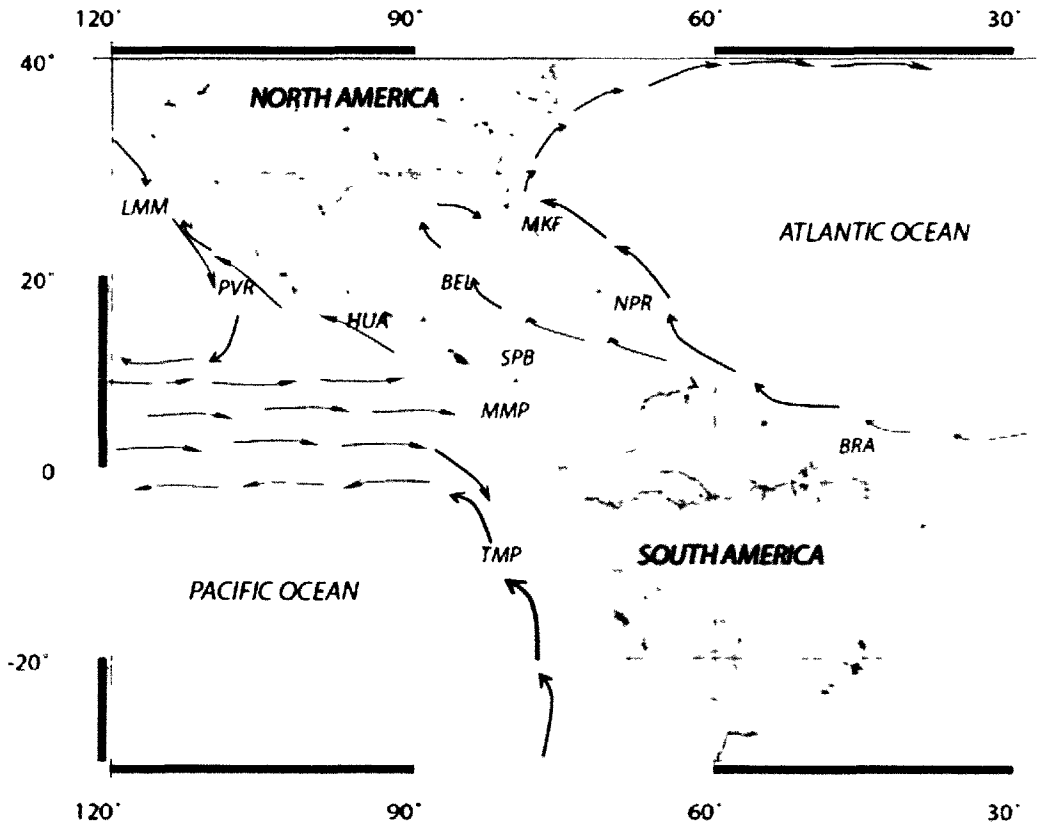

Figure 2.1. Map of the Tropical Eastern Pacific and Western Atlantic. Sampling locations for species that reside in the Tropical Western Atlantic along with abbreviations are the following: Marathon Key, Florida (MKF); Caye Caulker, Belize (BEL); Farjado, Puerto Rico (NPR); San Blas, Panama (SPB); Sao Luis, Brazil (BRA). The following are the populations sampled in the Tropical Eastern Pacific: Lopez Mateos, Mexico (LMM); Puerto Vallarta, Mexico (PVR); Huatulco, Mexico (HUA); Panama City, Panama (MMP); Tumbes, Peru (TMP). Currents are also displayed for both regions. Blue arrows represent cold currents while red arrows represent warm currents.

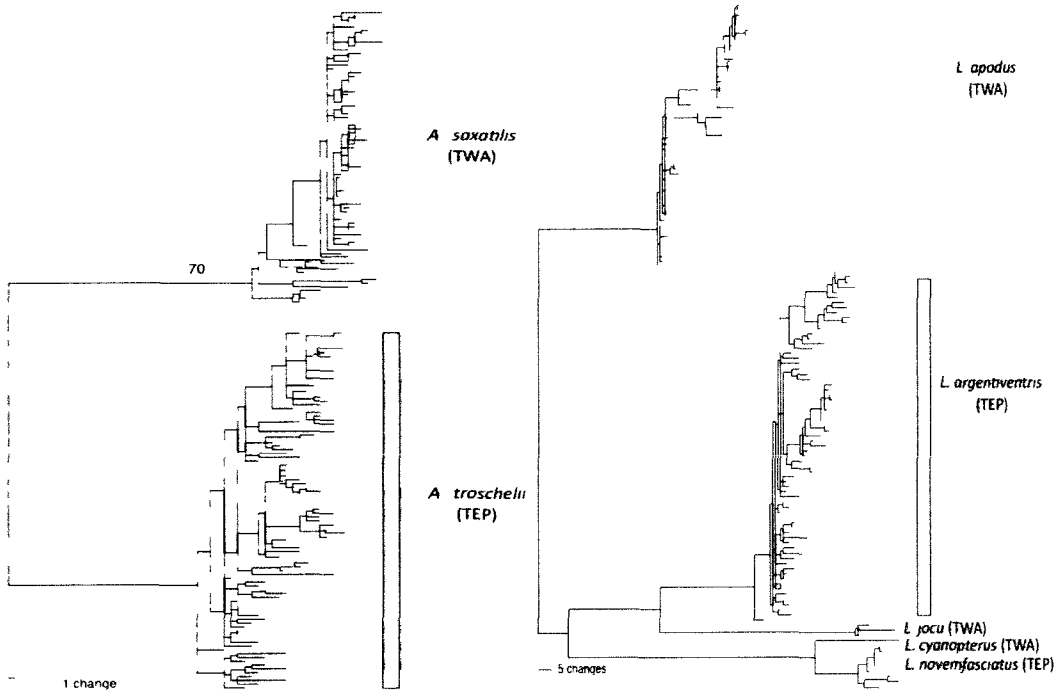

Figure 2.2. Phylogeny of *Abudefduf species and Lutjanus species* based on two mitochondrial markers (ATPase8 and Dloop) using the Maximum Likelihood method. Bootstrap values are shown when above 50%. *Lutjanus* ML based on a TIM+I+G model with the following likelihood settings: base frequencies set at 0.3300, 0.2724, 0.1197; Rate matrix=(1, 44.2131, 2.4535, 2.4535, 31.9002); Gamma shape distribution= 0.3810; Parsimony invariable characters= 0.4637. *Abudefduf* ML based on a TrN+I+G model with the following settings: Base frequencies=0.3115, 0.2788, 0.123; Rate matrix=(1, 35.2399,1, 1, 12.9571); Gamma shape distribution= 0.4081; Parsimony invariable character = 0.5588.

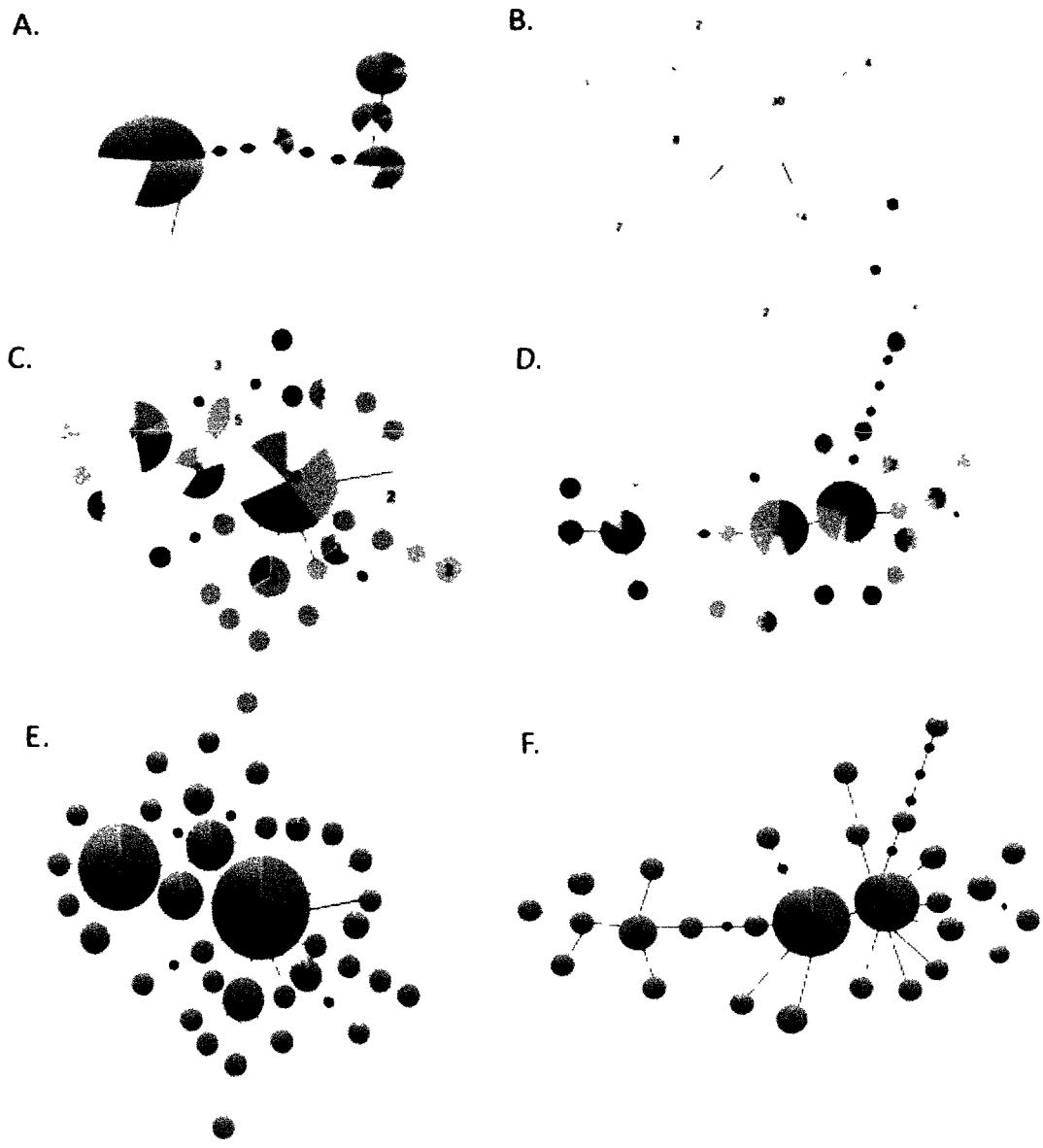

Figure 2.3. Minimum spanning networks based on parsimony informative characters of ATPase8. Each circle represents a single haplotype. Each line represents a single step mutation, black circles refer to a haplotype not found in the dataset. Designated numbers refer to the number of individuals that contain the haplotype. Colors refer to populations (given in the accompanied tables) where the haplotype is found. A) *L. apodus;* B) *L. argentiventris*; C) *A. troschelii;* D) *A. saxatilis*. The last two networks show regional difference of haplotypes for E) *A. troschelii* (Mexico-blue vs. Non-Mexican Populations-red); F) *A. saxatilis* (BRA/CPR-red vs. SPB/BEL/MKF-blue).

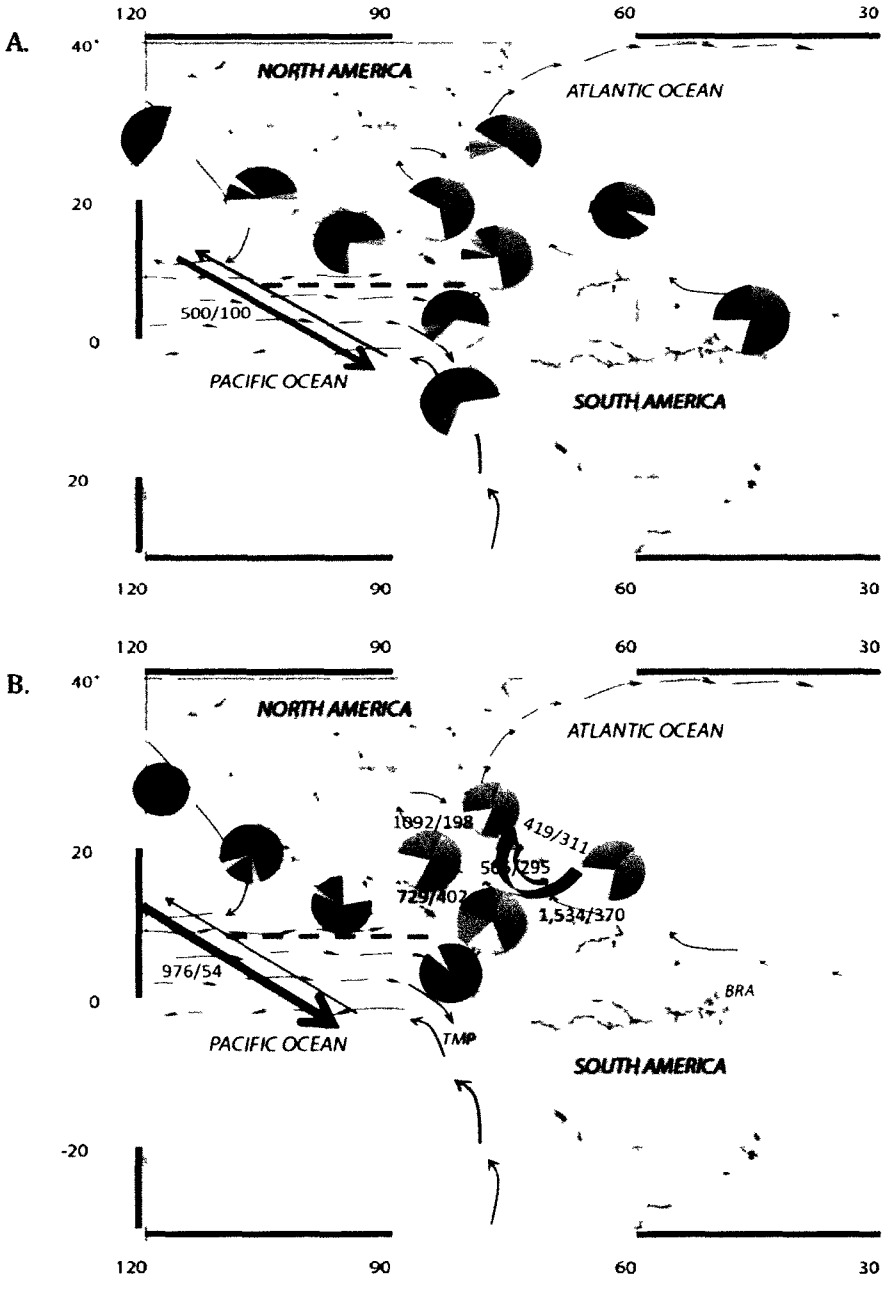

Figure 2.4. Geographic distribution of haplotypes. Each pie chart represents a population and haplotypes frequencies found within each population. Color with each pie chart represents a different haplotype; black refers to the frequency of private alleles. Brown arrows denote migratory movements estimated using MIGRATE. The width of the arrow represents the strength of migration. Purple arrows represent AMOVA results. A) *Lutjanus* species; B) *Abudefduf* species.

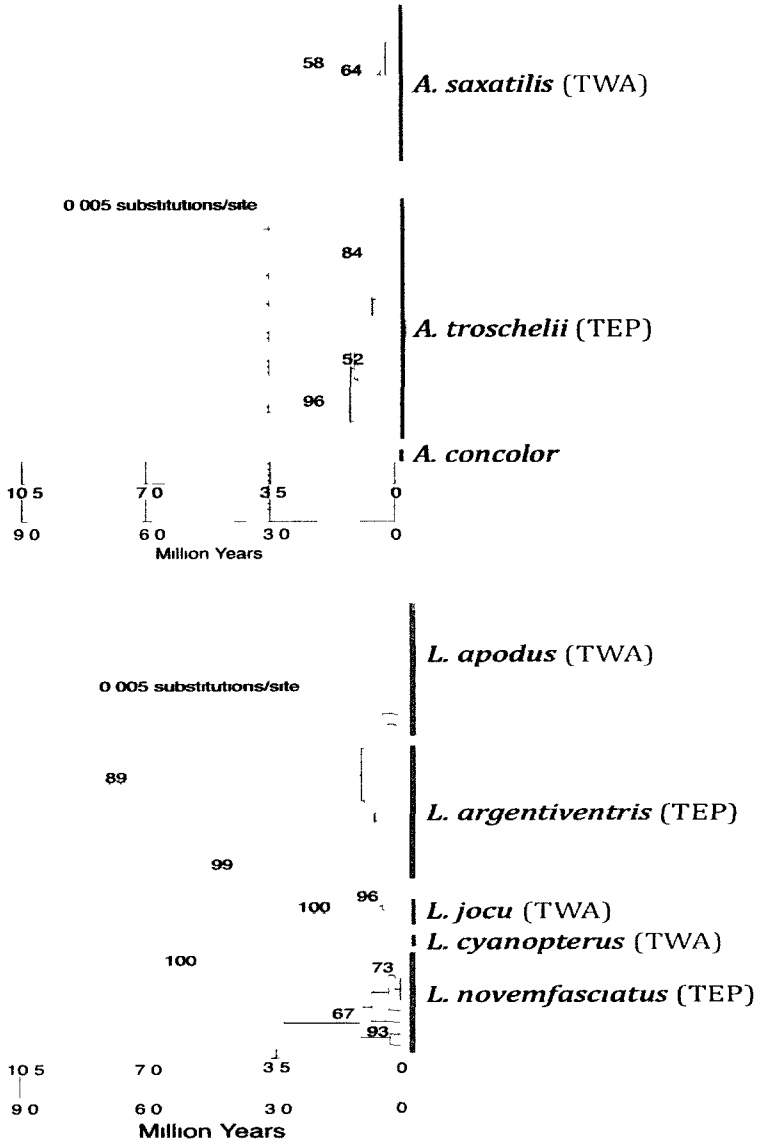

Figure 2.5. UPGMA method of A) *Abudefduf sp.* and B) *Lutjanus sp.* to display the implementation of a molecular clock based on the closure of the Isthmus of Panama (3.1-3.5myr) denoted with a red line. Both topologies are scaled with one another.

Species	N	H_n	H_d	π
Abudefduf saxtatilis				
All Populations	98	95	0.9993	0.0140
MKF	19	19	1	0.0105
BEL	21	21	1	0.0206
NPR	17	17	1	0.0232
SPB	19	17	0.9883	0.0063
BRA	22	22	1	0.0088
Lutjanus apodus				
All Populations	65	53	0.9896	0.0168
MKF	19	17	0.9883	0.0185
BEL	14	12	0.9780	0.0148
NPR	18	16	0.9869	0.0192
SPB	16	15	0.9917	0.0148
Abudefduf troschelii				
All Populations	106	126	0.9974	0.0153
LMM	20	19	1	0.0176
PVR	23	23	1	0.0182
HUA	22	22	1	0.0165
MMP	21	21	1	0.0165
TMP	20	21	1	0.0196
Lutjanus argentiventris				
All Populations	80	80	1	
LMM	17	17	1	0.0271
PVR	21	21	1	0.0161
HUA	27	27	1	0.0187
MMP	15	15	1	0.0187

Table 2.1. Species are listed along with the number of individuals sampled per population; the number of haplotypes per population (Hn); haplotype diversity (Hd); nucleotide diversity (π) are listed.

A)

Location	HUA	IMM	PVR	RCP	TMP
HUA	*	0.306	0.847	**0.036**	**0.018**
IMM	0.007	*	0.496	**0.000**	**0.000**
PVR	-0.015	-0.007	*	**0.027**	**0.018**
RCP	**0.060**	**0.090**	**0.047**	*	**0.018**
TMP	**0.050**	**0.076**	**0.038**	**0.063**	*

B)

Location	BEL	BRA	CPR	MKF	SPB
BEL	*	0.991	0.991	0.991	0.477
BRA	0.000	*	0.991	0.991	**0.045**
CPR	0.000	0.000	*	0.991	0.315
MKF	0.000	0.000	0 000	*	0 216
SPB	0.003	**0.006**	0.006	0.006	*

C)

Location	BEL	MKF	SPB	NPR
BEL	*	0.955	0.811	0.640
MKF	-0.043	*	0.423	0.937
SPB	-0.039	-0.004	*	0.198
TPR	-0.030	-0.042	0.024	*

D)

Location	HUA	LMM	MMP	PVR
HUA	*	0.550	0.514	0.126
LMM	-0.006	*	0.991	0.820
MMP	-0.006	-0.033	*	0.919
PVR	0.015	-0.014	-0.023	*

Table 2.2. Population pairwise (ϕst) comparisons (bottom diagonal) and associated significance values for A) *A. troschelii*; B) *A. saxatilis*; C) *L. argentiventris*; D) *L. apodus*. Bold values are significant.

A)

Source of variation	SS	Variance components	Variation	F-Statistics (P-value)
Among groups (Fct)	12.714	0.141	2.41%	0.024 (0.00)
Among populations within groups (Fsc)	19.111	0.038	0.64%	0.007 (0.00)
Within populations (fst)	504.653	5.670	96.94%	0.031 (0.00)
Total	536.479	5.849		

B)

Source of variation	SS	Variance components	Variation	F-Statistics (P-value)
Among groups (Fct)	26.85	0.314	3.73%	0.037 (0.11)
Among populations within groups (Fsc)	33.34	0.149	1.77%	0.019 (0.00)
Within populations (Fst)	504.653	7.952	94.50%	0.055(0.00)
Total	536.479	8.415		

Table 2.3. Analysis of molecular variance (AMOVA) based on partitioning of populations into two groups A) *A. saxatilis*: BRA/CPR vs. SPB/NPR/BEL; B) *A. troschelii*: MMP/TMP vs. LMM/HUA/PVR.

Species	Dloop (bp)	ATPase (bp)	P_{ic}	θ_c	θ_v	g	Tc(years)
Abudefduf saxatilis	485	735	120	0.153 (±0.007)	0.283 (±0.050)	209.2 (±64.4)	165292.6
Lutjanus apodus	474	774	96	0.052 (±0.001)	0.051 (±0.002)	0.993 (±5.5)	20355663.67
Abudefduf troschelii	451	635	79	0.240 (±0.011)	1.126 (±0.084)	443.5 (±25.6)	79783.3
Lutjanus aregentiventris	397	520	78	0.237 (±0.013)	1.273 (±0.147)	419.5 (±34.7)	24120.7

Table 2.4. List of population statistics for each dataset. Number of base pairs (bp) used for each gene are listed. Parameters include: parsimony informative characters (P_{ic}); theta without(θ_c) and with growth (θ_v); growth (g), and coalescence time (Tc). Standard deviations are shown within the parentheses.

Species pair (Pacific vs. Atlantic)	Average pairwise differences			Average divergence rates					
				Dloop		ATPase		Both genes	
	Dloop	ATPase	Both genes	3.1myr	3.5myr	3.1myr	3.5myr	3.1myr	3.5myr
A. troschelii vs. A. saxatilis	0.062	0.043	0.044	1.990	1.762	1.382	1.224	1.415	1.253
L. argentiventris vs. L. jocu	0.064	0.021	0.038	2.079	1.841	0.668	0.592	1.213	1.074
L. novemfasciatus vs. L. cyanopterus	0.105	0.000	0.053	3.403	3.014	0.000	0.000	1.725	1.528

Table 2.5. List of average pairwise distances and rates of divergences for each species pair listed. Rates of divergences were estimated for each gene and for the two genes combined by dividing the average pairwise distance by 3.1-3.5myr, the estimated dates of the closure of the Isthmus of Panama.

Chapter 3: A Phylogeographic Study of Tropical Eastern Pacific Shorefish

Abstract

In this chapter I examine the underlining mechanisms involved in migration among populations of reef fish that reside in the Tropical Eastern Pacific (TEP). The Tropical Eastern Pacific (TEP) coastline is divided into four recognized biogeographic regions with distinct thermal gradients and oceanic currents (Pacific Baja, Cortez, Mexican, and Panamic). Each province is separated by gaps of habitat (sandy beach) not suitable for reef fish settlement (the Sinaloan Gap, 370km; and the Central American Gap, 1,200km). To determine whether long stretches of sandy beaches act as true barriers to dispersal, I reconstructed the phylogeography of 3 abundant rocky reef fish with overlapping distributions in the TEP (*Lutjanus argentiventris,* *Anisotremus interruptus,* and *Abudefduf troschelii*). All three species, have sister species that reside in the Atlantic, enabling the calibration of a molecular clock. These species differ in life history characteristics, allowing me to examine whether differences in genetic signatures among fish with the same evolutionary history may be driven by life history differences that influence gene flow more than geological processes, such as habitat fragmentation. To assess levels of gene flow I sequenced approximately 1000bp of mitochondrial DNA using two markers (ATPase and DLoop).

All three species had distinctly different genetic signatures with *Lutjanus argentiventris* exhibiting no genetic structure (suggesting high gene flow among populations while *Abudefduf troschelii* and *Anisotremus interruptus* had regional differentiation corresponding exactly to the biogeographic provinces found in the TEP. However these two species differed in the location of the greatest amount of regional subdivision with the greatest differentiation for *Abudefduf troschelii* in the southern TEP (between the Mexican and Panamic Province) compared to greatest differentiation in *Anisotremus interruptus* in the Northern TEP (between Baja California and the rest of the TEP). These differences in genetic signatures among species with shared evolutionary history are likely caused by differences in life history: *Abudefduf sp.* are benthic spawners that do not have pelagic eggs while *Lutjanus argentiventris* are broadcast spawners and form spawning aggregations. *Anisotremus interruptus* has an intermediate strategy as a broadcast spawner but does not form spawning aggregations. The signatures I found are concordant with what is expected based on these life history traits with *Abudefduf troschelii* showing the greatest genetic structure, *L. argentiventris* the least, and *A. interruptus* intermediate to the two. These results indicate that life history characteristics may lead to different effects of habitat fragmentation on population gene structure.

Introduction

The Tropical Eastern Pacific is unique because it has undergone radical geological transformations that are well documented, making it a perfect platform to study evolutionary and ecological processes, particularly the factor that may affect dispersal among populations. The Tropical Eastern Pacific (TEP) consists of the Pacific Ocean coastline from Mexico to Peru including oceanic islands: Revillagigedos, Malpelo, Galapagos, Clipperton Atoll, and Coco Islands (Figure 3.1). The TEP harbors little biodiversity compared to other tropical marine environments, with only 1,089 TEP shorefish species being described (Robertson and Allen 2002). However, the TEP is unique because it contains the highest rate of endemism (79.3%) of any tropical region its size (Robertson and Allen 2002). The lack of diversity and a high rate of endemism may be a result of effective barriers to dispersal surrounding the TEP along with past and ongoing geological processes affecting connectivity within the TEP.

Effective barriers surrounding the TEP restrict the movement of fish in and out of this region. West of the TEP lies the largest open ocean barrier to dispersal for shorefish, the Eastern Pacific Barrier which spans from the TEP to Central Pacific, a distance of approximately 4000-6000km (Briggs 1961). Since its formation 65mya, very few coastal species have been known to reach the Central Pacific due to the unavailability of shallow habitat to settle on (Briggs 1961; Lessios et al 1998). North and South of the TEP are regions separated by steep thermal gradients due to

coastal upwelling and basin current patterns (Hastings 2000; Robertson and Allen 2002). The northern region is undated with cold water from the California current which reaches North Baja, before continuing to the Eastern Pacific. Cold water in the Southern region is a result of the Humboldt current which reaches central Peru before being directed out to the Eastern Pacific. Cold water found North and South of TEP acts as an effective barrier to dispersal for tropical fish found in the TEP (Robertson and Allen 2002).

To the east of the TEP lies the Isthmus of Panama, an effective land barrier separating the Tropical Western Atlantic from the TEP. During the Miocene (16-15mya), the tropical Atlantic and eastern Pacific were one body of water with few small islands separating the two regions (Coates and Obando 1996; Jackson et al 1996). During the late Pliocene, parts of Panama began to rise, and ultimately severed the marine connection between the eastern Pacific Ocean and the Caribbean Sea. The rise of the Isthmus produced two extremely different oceans (Coates and Obando 1996; Jackson et al 1996). The Atlantic Ocean is a closed basin with stable currents, while the Pacific is an open basin with unstable currents, variable tide oscillations, and upwelling, which accounts for the scarcity of coral reefs (Case et al 2002; Coates and Obando 1996; Muss et al 2002)(Figure 3.1). Along with the creation of two differing oceans, the rise of the Isthmus severed preexisting marine distributions into two populations starting them on a path of independent

evolutionary trajectories (Bermingham 1996; Jordan 1908; Lessios 2008). Because of this recent separation there are many pairs of sister species, referred as geminate species, between the Atlantic and Pacific that share similar morphological features, life histories, and ecological niches (Jordan 1908). The Isthmus of Panama creates an opportunity for evolutionary biologists to study fish molecular clocks and historical biogeography within a recent time frame of earth history (Bermingham et al 1997; Lessios 2008).

Aside from intrinsic barriers restricting the movement of shore fish in and out of the TEP, the lack of biodiversity and high rate of endemism in the TEP may be in part due to past and ongoing geological processes occurring within the TEP. One geological event that continues to influence dispersal capabilities of marine taxa is the interruption of suitable reef and rock coastline habitat by long stretches of sandy beaches (Briggs 1974; Rosenblatt 1974; Robertson and Allen 2002). This form of habitat fragmentation has broken the coastline into three recognized biogeographic provinces (Cortez, Mexican, Panamic Province) each containing distinct thermal gradients and oceanic currents (Walker 1966; Muss et al 2001)(Figure 3.1). Each province is separated from one another by gaps of habitat not suitable for reef fish settlement (the Sinaloan Gap, 370km; and the Central American Gap, 1,200km) (Figure 3.1). Presently, there has been no extensive genetic study performed to determine whether these expansive gaps act as barriers to dispersal; however, one

study, based on species diversity of the family Chaenopsidae in the TEP, has found many cases of allopatric speciation as a result of these biogeographic breaks (Hastings 2000).

Another geological event that has influenced marine distributions in the TEP is the rise of offshore volcanic islands. Oceanic islands such as the Galapagos Islands (1000km from mainland), house a majority of the marine biodiversity we see today in the TEP (Walker 1966). Approximately 60% of the species found on these islands are from the mainland TEP (Robertson and Allen 2002). The islands mentioned above were never a part of the mainland, in some cases they were much farther away from the mainland then they are today (Walker 1960); therefore, it is still unclear how and when inshore fish reached these destinations, whether it was random or continuous gene flow.

The separation of Baja Peninsula from present day Mexico has also had a profound affect on the distribution of fish in the TEP. Before the Miocene, Baja California was attached to North America (Case et al. 200). Paleontologic and stratigraphic data indicate that initial rifting and drifting began in the middle Miocene (14mya), which led to a northern marine opening between the Pacific and the gulf ~5-7mya. Once this marine connection closed, the Gulf continued to open and finally ceased to move approximately four million years ago (Case et al. 2002; Jackson et al 1996). The opening of the gulf created available habitat for tropical

shore fish to occupy. Approximately 90% (785 species) of the fish found in the Gulf are from the Panamic province, the rest are either endemic or from the Pacific (Case et al. 2002). Within the gulf, Walker (1966) has classified four biogeographic provinces based on faunal diversity and climatic characteristics, which he named the upper gulf, the central gulf, the Cabo San Lucas area, and the southeastern gulf. The two most variable regions are the upper and central gulf. The upper gulf is shallow, has extreme tidal oscillations, and experiences fluctuating sea surface temperatures, making it difficult for tropical fish to survive in this region (Case et al. 2002). The central gulf, on the other hand, has much less tide and temperature variation as well as deep oceanic channels and canyons. Although fish assemblages differ between the two regions, there are some fish that are found throughout both regions. A study performed on five species of near shore fish in the gulf has found significant genetic differences between these two regions are most likely attributed to oceanic differences (Riginos 2005).

In this study I examine how ongoing geological processes in the TEP affect dispersal among resident marine fish with distinctly different life histories. I combine molecular evolution, ecology, biogeography, and geology to estimate the amount, direction, and timing of past migration events among tropical shore fish. The combination of these fields serves as a valuable tool in explaining the evolutionary history of a species. Although understanding the evolutionary history

of one species is important, it is of more interest to determine whether there is concordance across taxa that occur in the same region (Avise 2000). Comparative phylogeographic studies are less common than single-species phylogeographic studies, but have become extremely useful in revealing common genetic patterns as well as possible causes of genetic heterogeneities (Avise 2000, 2004). To test for geographic concordance, I performed a multi taxa comparison among marine fish species that reside in the Tropical Eastern Pacific to determine whether taxa exhibit similar genetic patterns or whether life history differences may be important in driving patterns in genetic signatures.

Specifically, I reconstruct the phylogeography of 3 abundant rocky reef fish with overlapping distributions in the TEP (*Lutjanus argentiventris, Anisotremus interruptus,* and *Abudefduf troschelii*). All three species, have sister species that reside in the Atlantic, enabling the calibration of a molecular clock, but differ in life history characteristics, allowing me to examine whether differences in genetic signatures among fish with the same evolutionary history may be driven by life history differences that influence gene flow more than geological processes, such as habitat fragmentation. *Lutjanus argentiventris* is a broadcast spawner forming large spawning aggregations were positively buoyant eggs are released (Munro 1974; Sala et al 2003; Victor and Wellington 2000). The time the larvae spend in the open water column (pelagic larval duration) before settlement is approximately 22 days

(Zapata and Herron 2002). *Anisotremus interruptus*, is also a broadcast spawner, but does not form large spawning aggregations *as L. argentiventris* (Cailliet et al 2000). The pelagic larval duration (PLD) of *A. interruptus* is approximately 40-50 days (Watson and Walker 1992). *Abudefduf troschelii* individuals do not form spawning aggregations nor have positively buoyant eggs. *Abudefduf species* are benthic spawners and eggs settle down to the floor where they are guarded and defended by the male (Munro 1974; Victor and Wellington 2000). More opportunities for dispersal are expected among species that form spawning aggregations, longer PLD, and have positively buoyant eggs. *Lutjanus argentiventris* and *Anisotremus interruptus* are also valuable to local and commercial fisheries. An assessment of gene flow and the degree of genetic variation within each population will be useful to determine the sustainability of these fisheries.

Methods

Sampling and DNA preparation

Sampling locations and sample sizes are shown in Figure 3.2.1and listed on Table 3.1. Fish were collected by hand nets, spear fishing, and at local fish markets. Tissue samples were cut from the caudal fin and placed in 95% ethanol for preservation. In the laboratory, DNA was prepared from 75 to 150 mg of muscle by proteinase K digestion in lysis buffer (10mM Tris, 400 mM NaCl, 2 mM EDTA, 1% SDS

overnight at 55°C. This was followed by purification using chloroform extractions and alcohol precipitation (Sambrook et al. 1989). Amplification of the 5' hypervariable mitochondrial D loop region was performed with primers: ProL (CTACCTCCAACTCCCAAAGC) and tPhe (GTGTTATGCTTAGTTTAAGC) for *Abudefduf* sp.; ProL and STDloop (CTGGAYAGAYRGCACGGC) for *Lutjanus* sp.; L15725 (AARCCYGARTGRTAYTTYYTNTTYGC) and DloopH (CCTGAAGTAGGAACCAGATG) for *Anisotremus sp*. In addition to the Dloop region, mitochondrial ATPase8, tRNA, and a partial segment of cytochrome oxidase subunit II was sequenced using the following primers for all species: ATPase 8 (AAAGCRTYRGCCTTTTAAGC) and Co3 (GTTAGTGGTGACCGGCTTGGRTC). Both regions used a cycling profile of 95°C for 1 minute, 52°C for 1 minute and 72°C for 1 minute for 35 cycles. After purification following the manufacturer's protocol (ABI, PerkinElmer), sequencing was performed on an ABI PRISM 3700 DNA automated sequencer (Applied Biosystems).

Phylogenetic and population genetic methods

Sequences were aligned using Sequencher 4.1 (Gene Codes Corporation), and checked manually. A model of evolution for each dataset was constructed using the program MODELTEST version 3.6 (Posada 1998). Approximately 56 models of nucleotide sequence evolution were evaluated using MODELTEST, the models selected for each data set are described in Figure 3.2. Shimodaira and Hasegawa test implemented by PAUP (Phylogenetic Analysis Using Parsimony, version 4.0; Swafford

1998), was used to test whether topological differences existed between both the markers used in this study as well as to determine if there were differences with or without an enforced molecular clock (Shimodaira and Hasegawa 1999). Because no significant differences were found between the Dloop and ATPase8 topologies, both loci were combined for further analysis. Phylogenetic relationships were measured using Maximum Likelihood (ML), Maximum Parsimony (MP) and Neighbor-Joining methods (NJ). Both MP and NJ topologies were constructed with the program PAUP. Topological confidence was evaluated for MP and NJ with 1000 bootstrap replicates. The program GARLI 0.951 (Zwickl 2006) was used to construct a Maximum Likelihood tree making the necessary adjustments to implement the model of evolution proposed by MODELTEST. The search was set to terminate when the ln score remained constant for 20,000 consecutive generations. The tree with the highest likelihood score out of 10 independent runs was retained. Bootstrap values were added to the topology based on 100 replicates performed in GARLI. Average pairwise sequence divergence was calculated between individuals within the same species and between geminate pairs using PAUP for each marker and then for both markers combined. A UPGMA was constructed to apply a timeline to the phylogeny.

For each population the number of haplotypes (H_n) and nucleotide (π) and haplotype (*h*) diversity was estimated using the software package DnaSP (Rozas et al

2003). Minimum spanning networks based on haplotype differences were constructed using the program HAPSTAR (Griffiths 2010). To detect genetic structure for each dataset, fixation indices were calculated using ARLEQUIN 2.0 (Schneider et al 2000). Pairwise comparisons between populations were calculated to test for population subdivision (ϕst). A hierarchical analysis of molecular variance (AMOVA) was performed using ARLEQUIN to test for significant regional subdivision (ϕct). A significant amount of variation found between populations or regions implies low levels of gene flow. If results are not significant, low levels of variation between populations and regions may be a result of the homogenizing effect of high gene flow. The program MIGRATE implements coalescent theory to determine the number of migrants and the direction of migratory movements (Beerli 2006). To determine the number of migrants (Nm), I used a Bayesian method implemented in MIGRATE consisting of 1 long chain, a sampling increment of 100, the recorded steps in a chain was set at 100,000, and the option of heating was turned on. Analysis of each dataset was performed 10 times to generate a mean and standard deviation.

Historical demography was evaluated with the program FLUCUATE 1.4 (Kuhner et al. 1998). To determine population expansion or contraction, growth (g) was estimated. To describe effective population size the parameter θ was calculated with two models, one with growth kept constant (θ_c) and the other with the option for population growth (θ_v). All three parameters were estimated by running 10

replicates which generated a mean value and standard deviation to ensure convergence of parameters. Analysis for each run was performed with 10 short Monte Carlo chains of length 20000 followed by 2 long chains, with a sampling increment of 20. Coalescence time was estimated under the assumption that coalescence is reached when the population size was reduced to 1% of its present-day value (Wares and Cummingham 2001). Two parameters need to be estimated to obtain coalescence time: mutation rate (μ) and generation time (g). To calculate mutation rate average pairwise divergence between geminate species was subtracted from the average pairwise divergence within a species to account for intraspecific polymorphism and then divided by generation time (g). The time at reproductive age, for each species was acquired from the literature: *Abudefduf troschelii* =1yr; *L. argentiventris*= 4yrs; *A. interruptus* = 3yrs (because this parameter is not available for *A. interruptus*, generation time of *A. davidsonii* was used instead) this measurement was divided by 3.5myr to obtain time since divergence in generations (Robertson 1989; Martinez Andadre 2010; Calliet 1999).

Life history information

Life history characteristics egg buoyancy (buoyant or non-buoyant) and spawning behavior (benthic or broadcast) (aggregations or not) were acquired from previous studies. *Abudefduf species* are benthic spawners that do not form aggregations and have negatively buoyant eggs that settle to the bottom (Munro

1974; Sala et al 2003; Victor and Wellington 2000). In contrast *Lutjanus* species form spawning aggregations and are broadcast spawners with positively buoyant eggs. *Lutjanus argentiventris* have a slightly larger larval duration (PLD=23 days) than *Abudefduf troschelii* (PLD= 18.2 days)(Munro 1974; Sala et al 2003; Victor and Wellington 2000; Zapata and Herron). *A. interruptus* is a broadcast spawner and produces pelagic eggs, but does not form spawning aggregations (Cailliet et al 2000). Due to the lack of information of larval duration for *Anisotremus interruptus,* the larval duration of *A. davidsonii* (PLD=40-50 days) will be used (Watson and Walker 1992). A longer PLD found in *Anisotremus* species may provide more opportunity for dispersal than the other species. A comparison between these life history traits and gene flow among all species will be performed to determine if there is a relationship between any of the life history components and gene flow.

Results

DNA Sequence Analysis and Phylogenetic inference

Sixteen populations of *Abudefduf troschelii*, twelve populations of *Anisotremus interruptus,* and fourteen populations of *Lutjanus argentiventris* were included for comparison. Included in the *Abudefduf sp.* sample was a population from the Galapagos (Figure 3.2.1and Table 3.1). Both the 5'-end portion of the mitochondrial Dloop and ATPase8 were sequenced for each individual. A

Shimodaira and Hasegawa test found no significant differences between the mtDloop and ATPase8 topologies (P=1.000); therefore further analysis will include both markers combined unless otherwise stated. Total base pair length and the number of parsimony informative characters for *Abudefduf troschelii, Anisotremus interruptus,* and *Lutjanus argentiventris* are 1169bp, 291 characters; 1341bp, 148 characters; 1151, 163 characters; respectively.

For each species, the three methods of tree reconstruction (MP, NJ, ML) resulted in identical topologies; maximum likelihood topologies along with bootstrap values are presented for each dataset (Figure 3.2). The models of evolution used to infer the ML topology are given in Figure 3.2. Shallow genealogies were found for all datasets with very little branch support. The average pairwise differences between individuals for *Abudefduf troschelii, Anisotremus interruptus, Lutjanus argentiventris* were 1.4%, 1.0%, and 1.2% respectively.

Population Genetic Inferences

Haplotype diversity was high for all species, making a minimum spanning network difficult to reconstruct. To simplify the network I collapsed the dataset by only using parsimony informative characters of mitochondrial marker ATPase8. This process allowed me to build a network with a manageable amount of haplotypes in order to infer relationships between haplotypes (Figure 3.3). The number of haplotypes based on parsimony informative characters in the ATPase gene differed

among species (*A. troschelii* H_n=75; *A. interruptus* H_n=36; *L. argentiventris* H_n=39). *A. troschelii* had nearly double the number of haplotypes than the other two species. Haplotype networks were also built to display haplotype frequency and the number of mutational differences found between regions (Figure 3.3. For each species, populations were clustered into four groups (Pacific Baja, Cortez, Mexico, Panamic) according to biogeographic provinces proposed by Briggs (1974). In all species, two to five haplotypes dominated the network, with many singletons that differ from the dominant haplotype(s) by one single-step mutation. This pattern indicates either high gene flow or a recent population expansion. Haplotypes were color-coded based on geographic placement. Geographic distributions of haplotypes for each province were constructed with pie charts to show haplotype numbers and frequencies per province (Figure 3.4a,b). Private alleles, haplotypes found in only one population, are shown in black and ranged from 1 to 23 private alleles per province per dataset, the lowest number of private alleles (H_n =1) was found in the *Anisotremus interruptus* populations belonging to the Cortez province. The highest number of private alleles (H_n =23) was found in the *Abudefduf troschelii* dataset in the Mexican province.

Gene flow analysis comparing genetic differentiation among populations identified different degrees of structure among species (Table 3.2). *Lutjanus argentiventris* showed six cases of significant genetic structure among pairwise

population comparisons, all involving population HUA (Table 3.2 c,d). In comparison, I detected 20 instances of population subdivision in *A. interruptus* (Table 3.2b), while *Abudefduf troschelii* had 39 significant cases of population subdivision (Table 3.2a). This is consistent with the hypothesis that *L. argentiventris* has high levels of gene flow among populations while *A. troschelii* has lower gene flow among populations.

An AMOVA was used to detect genetic differentiation between population clusters. All possible group comparisons for *Lutjanus argentiventris* resulted in no significant between region genetic variance. *Abudefduf troschelii* and *Anisotremus interruptus* both showed cases of genetic structure between assigned groupings. The first group comparison separated populations according to the proposed provinces (Pacific Baja, Cortez, Mexican, Panamic) explained by Briggs (1974). Both *Abudefduf troschelii* and *Anisotremus interruptus* showed a statistically significant amount of genetic differentiation between the four regions (Table 3.3). This result suggests lower levels of gene flow between the four biogeographic provinces for *A. troschelii and A. interruptus*. Comparisons made between the North and South Baja regions including and excluding Pacific Baja populations, were not significant. An AMOVA performed between the Galapagos and mainland TEP, also showed no significant amount of variation between the two regions. However, other assigned groupings using AMOVA found significant results for both *A. troschelii and A.*

interruptus. To detect the highest ϕct value (variation between regions), all combinations of regional groupings, following a southward stepping stone model of migration, were performed and plotted against ϕct in order to find the highest amount of regional subdivision (Figure 3.5). *A. troschelii and A. interruptus* showed differences in the location of genetic subdivision. The highest ϕct values were found when comparing the most Northern population (Juanico, MX) to the rest of the *Anisotremus* dataset; and the most Southern population (Tumbes, Peru) to the rest of the *Abudefduf* dataset. A comparison of one population to many may lead an inflated ϕct value, therefore I only recognized comparisons involving two or more populations per region. By these criteria *Anisotremus interruptus* exhibited the highest amount of genetic subdivision between Baja (including Pacific Baja and Cortez Province) and the populations belonging to the Mexican and Panamic Province (Figure 3.5). In contrast the highest amount of subdivision for *A. troschelii* was found between all Mexican populations and populations found in the Panamic province (Figure 3.5).

Using the regional subdivisions above, patterns of gene flow including the number of migrants and direction of migration events were estimated with MIGRATE. A similar southward pattern of migration was found for both *A. troschelii* and *A. interruptus*, but the strength of migration differed slightly. Southward migration (Nm=645) was stronger than in the northward direction (Nm= 477) for *A.*

troschelii. Although the number of migrants in the southern direction was smaller in *A. interruptus* (Nm=461) compared to *A. troschelii*, the strength of the southward migration in *A. interruptus* was stronger by more than 2 fold than Northward migration (Nm=150) in *A. interruptus* (Figure 3.4 a,b). No concordant results were obtained for *Lutjanus argentiventris.*

Molecular Clock Estimates

No significant differences with or without an enforced molecular clock were found using a Shimodaira and Hasegawa test (P=1.000); therefore a molecular clock was used for all the datasets. Previous studies have found that *Abudefduf troschelii,* from the Tropical Eastern Pacific, is sister to *A. saxatilis,* from the Tropical Western Atlantic; high bootstrap values found in this study confirms that they are true geminates (Bermingham 1996; Shulman and Bermingham 1995)(Figure 3.6). A comprehensive phylogeny was also constructed for *Anisotremus* sp., confirming two geminate pairs (*A. interruptus vs. A. surinamensis,* and *A. taeniatus vs. A. virginicus*)(Bernardi et al 2008). Both geminate pairs were included in this study to compare rates of divergence and test for concordance. At this time no phylogenetic work has been performed for *Lutjanus* species. It has been proposed that *Lutjanus apodus* is the geminate of *L. argentiventris* (Bermingham 1996; Shulman and Bermingham 1995). Another proposed geminate pair belonging to the family Lutjanidae was included in the analysis for comparison, *L. novemfaciatus* from the

Tropical Eastern Pacific, and *L. cynopterus* from the Tropical Western Atlantic (Bermingham 1996; Lessios 2008; Jordan 1908). *L. jocu* was included as an outgroup. ML, NJ, and MP topologies were constructed for each dataset. A UPGMA tree for each dataset is presented to show relationships and rates of divergences (Figure 3.6). The topology for *Lutjanus* species presented in this paper supports *L. argentiventris* as sister to *L. jocu* not *L. apodus,* its proposed geminate. *L. argentiventris* and *L. apodus* were thought to be geminate species based morphological and ecological synapomorphies, but they genetically differ by 13.2%. The corrected average pairwise difference between *L. argentiventris* and *L. jocu* is 3.8%, similar to the estimated genetic divergence above for *Abudefduf* geminate sp. (4.4%) and between *L. novemfaciatus* and *L. cynopterus* (5.3%)(Table 3.4). *Lutjanus jocu* is found in the Tropical Western Atlantic. Based on these confounding results I will address *L. argentiventris* and *L. jocu* as geminate species.

The average pairwise differences between geminate pairs slightly varied (Table 3.4). To calculate rate of divergence the average pairwise difference was divided by the times proposed for closure of the Isthmus of Panama (3.1, 3.5 myr). The divergence rate for the combined markers between species pairs varied between 0.7-1.5% per million years. *Anisotremus interruptus* and *A. surinamensis* had the lowest rate of divergences compared to the other geminates (Table 3.4). A UPGMA based on a subset of samples included in the analysis was constructed along

with a time scale based on the divergence of the two geminate species and under the assumption that the these geminate species diverged at the time of closure (3.5myr) (Figure 3.4). The topologies for *Abudefduf sp.*, *Anisotremus sp.*, and *Lutjanus sp.* are scaled accordingly for visual comparisons.

Historical Demography

Growth (*g*) was for each dataset was estimated using the program FLUCTUATE (Table 3.5). Growth was positive for each dataset suggesting population expansion for each species. To obtain mutation rate (μ), I estimated number of generations (g), which was the time of divergence (for this calculation I used the proposed closure of 3.5 myr) divided by the age at first reproduction, acquired from previous studies. The average pairwise divergence was divided by two to account for one branch length and then further divided by number of generations to obtain μ per generation. Under the assumption that coalescence is reached when the population size was reduced to 1% of its present-day value, the formula $N_t=\theta e^{-(gt\mu)}$ was further simplified to $t= 1/(g\mu \ln 100)$, to solve for time of coalescence (t) (Wares and Cunningham). The time of coalescence for each data set varied between 84,344- 402,239 years ago (Table 3.5). The time of coalescence for all three species was below the 3.5myr (the approximate closure of the Isthmus). This result implies that branches diverged from one another after the final closure of the Isthmus for each dataset.

Two parameters for θ were calculated under a model that forced the population to stay at a constant growth (θ_c) and then under a model that allowed the population to expand or contract (θ_v). These parameters are valuable in understanding effective population sizes. Lutjanus argentiventris has a smaller θ value than the other two species, suggesting a smaller historic population size. *Anisotremus interruptus* had the highest value of θ, suggesting a larger effective population size than the other species (Table 3.5).

Discussion

Genetic Structure

Different degrees of genetic structure were found among *Abudefduf troschelii, Anisotremus interruptus,* and *Lutjanus argentiventris,* emphasizing the importance of comparative phylogeographic studies to inferring the evolutionary history of a region. Populations of *A. troschelii, A. interruptus,* and *L. argentiventris* were all found to have a large number of shared haplotypes, a large amount of genetic variation, and shallow genealogies. These results suggest high gene flow for all three species. The haplotype networks displayed closely related haplotypes and with a few haplotypes with high frequency spawning many single haplotypes. Most haplotypes that were shared were shared among multiple biogeographic regions. Of the three species studied, *A. troschelii* had a larger number of haplotypes and

private alleles. This result indicates lower levels of gene flow, which is concordant with the high number of ϕst comparisons that were significantly different from 0. Overall the pattern shown in the networks, a common haplotype giving rise to numerous rare haplotypes, is a common outcome when studying marine biota, but what is unique to this study is that for each dataset two or more haplotypes are common. Shallow networks with one major haplotype with many singletons along with widely distributed haplotypes are characteristic of high levels of gene flow, recent expansion events, or frequent historic population fluctuations. However, when two or more haplotypes are common this pattern can be explained by two scenarios: either two populations were isolated and diverged from one another fixating for a certain allele by random genetic drift and are now in contact with one another, spreading the two common alleles to the other populations; or two populations are just beginning to diverge from one another forming two common haplotypes. Because the haplotypes are widely distributed it is difficult to conclude which scenario is correct. An understanding of historical demography and migratory movements may resolve this issue.

The genetic partitions found in *Abudefduf troschelii* and *Anisotremus interruptus* correspond exactly to the biogeographic provinces of the TEP described by Briggs 1974. Both *Abudefduf troschelii* and *Anisotremus interruptus* exhibited a small but statistically significant amount of genetic differentiation between Pacific

Baja, Cortez, Mexican, and Panamic provinces as proposed by Briggs (1974). Limited gene flow between biogeographic regions may be a result of expansive stretches of unsuitable coastal habitat, which may act as an effective barrier to dispersal for *Abudefduf troschelii* and *Anisotremus interruptus.* Although, a southward movement was detected by using the program MIGRATE for both *Abudefduf troschelii* and *Anisotremus interruptus,* migration occurs at relatively low rates (Nm= 645; Nm=461, respectively).

Surprisingly, *Abudefduf troschelii* and *Anisotremus interruptus* share similar characteristics in genetic structure statistics, but they differ in geographic location of lowered gene flow. *Abudefduf troschelii* demonstrated a significant subdivision between Mexico populations and populations belonging to the Panamic Province. *Abudefduf* populations sampled in Panama and Peru showed significant pairwise ϕst values when compared to other populations, this could be the driving force for the regional subdivision observed in *A. troschelii* (Table 3.2). In contrast, *Anisotremus interruptus* showed high regional division between Baja populations and the rest of the populations (Cortez and Panamic populations). Genetic differentiation between these two regions can be largely attributed to the population found in the Northern most distribution of *Anisotremus interruptus*, Juanico, MX. Juanico populations differed from almost all other populations (Table 3.2). Because it is located in the most northern distribution of *A. interruptus,* Juanico may be a newly colonized

population that is beginning to diverge from the other populations. However, MIGRATE was not able to detect any migration or recent colonization into Juanico. *Lutjanus argentiventris* showed no genetic partitions. Habitat fragmentation along the coastline may not prove to be an effective barrier to dispersal among *L. argentiventris* individuals.

Historical demographic analysis found that all species exhibited past population expansions. This result implies that *A. troschelii, A. interruptus,* and *L. argentiventris* increased in numbers more than expected, suggesting that populations are unstable and likely to expand quickly. *A. interruptus* had the highest rate of population growth (g=436 individuals) as well as the highest value of theta indicating a high effective population size (Table 3.5). Time of coalescence for all three datasets ranged from 84,344- 402,239 years. These values are below the projected date of the closure of the Isthmus (3.1, 3.5 mya). *L. argentiventris* had a very low coalescence time possibly due to high levels of gene flow.

Molecular Clock Estimates

Five pairs of trans-Isthmian geminate pairs were used to calibrate a molecular clock. The genetic divergences between the five geminate pairs were surprisingly similar to one another, suggesting that all geminate pairs diverged from one another roughly around the same time (Figure 3.5, Table 3.5). *Anisotremus interruptus* and *A. surinamensis* had the lowest rate of divergence. Prior studies

have found similar results using different markers for this geminate pair and suggest that delayed divergence may be a result of a habitat preference. Near the closure of the Isthmus, small shallow water canals were all that connected TEP with TWA. Due to a greater affinity for coastal shallow waters, *Anisotremus interruptus* individuals that can tolerate such environments would be the last to diverge and therefore show the lowest divergence (Rocha et al 2008).

One confounding result from the topologies was that *L. apodus* is not sister to *L. argentiventris*. Based on morphological similarities it is understandable why these two species were thought to be geminate species. However, the topology suggests that *L. apodus* and *L. argentiventris* are 13% divergent, four times more divergent than *L. argentiventris* and *L. jocu* (3.8%). Based on similar rates of divergence found with the other geminates used in this study, *L. argentiventris* and *L. jocu* are geminate species. A detailed phylogeny of the family Lutjanidae needs to be estimated to determine whether *Lutjanus apodus* has a geminate counterpart.

Are there effective barriers to dispersal within the TEP?

A comparative phylogeography of species that reside in the TEP reveals differences in genetic signature among species that are assumed to have shared similar evolutionary histories. Genetic concordance was found among *Abudefduf troschelii* and *Anisotremus interruptus* between the four provinces (Pacific Baja, Cortez, Mexican, and Panamic), but the location of the highest regional subdivision

was different between the two species, resulting in different genetic signatures. Barriers to dispersal are effective between the provinces, but more so in the south of TEP for *A. troschelii* and in the north of TEP for *A. interruptus*. Surprisingly, genetic structure in the Sea of Cortez was absent in these species suggesting different oceanic processes within the gulf do not affect dispersal. In addition, no significant genetic differences were found between the Galapagos and the mainland for Abudefduf suggesting open water is not a barrier to dispersal. Unlike the other two species, *Lutjanus argentiventris* showed no population structure. This result indicates relatively high levels of gene flow, suggesting that there are no effective barriers to dispersal in the TEP for *Lutjanus argentiventris*. High levels of gene flow exemplified for *Lutjanus argentiventris* compared to the other two species, may be a result of differences in life history.

Do life history factors play a role in dispersal?

A comparison among taxonomic families that reside in the same oceanic environment and are assumed to have been exposed to similar past and present environmental events, enables me to propose that differences in life history factors influence dispersal movements. All three taxonomic groups show differing levels of gene flow, with *Lutjanus argentiventris* showing higher levels of gene flow than the other two species. These differences in levels of gene flow may be driven by differences in life history characteristics. For instance, *Lutjanus sp.* travel to

aggregation sites to spawn, while *Abudefduf* individuals spawn within the proximity of their home range. Once spawning occurs, *Abudefduf sp.* eggs settle to the floor where the male protects them until hatching takes place. In *Lutjanus sp.* there is no parental investment after spawning events; eggs are pelagic and vulnerable to predation (Zapata and Herron 2002). Along with a slightly longer PLD, the life history traits possessed by *Lutjanus sp.* provide many more opportunities for dispersal than *Abudefduf* individuals. *Anisotremus interruptus* has life history characteristics that fall intermediate between two species above. Similar to *Abudefduf sp.*, *Anisotremus* individuals do not form spawning aggregations, but they are broadcast spawners and have positively buoyant eggs similar to *Lutjanus sp.* Thus, my results showing that life history traits corresponding to genetic signatures underline the importance that life history characteristics play in population genetic structure. Based on life history characteristics, *Lutjanus argentiventris* is expected to have high levels of gene flow, which is concordant with the lack of genetic structure found for this species. Alternatively, the ecology of *Abudefduf troschelii* would limit the amount of migratory movements which is supported by the fact that *A. troschelii* exhibited the highest amount of genetic structure of the three species. Although *Anisotremus interruptus* has the longest PLD of 40-50 days, populations were found to exhibit an intermediate amount of subdivision similar to what's expected from a broadcast spawner that does not form spawning aggregations. Similar to previous

phylogeographic studies, this result may indicate that factors such as the formation of spawning aggregations or egg buoyancy may have a stronger influence on dispersal capabilities than pelagic larval duration (Victor and Rignios 2005; Shulman and Bermingham).

Conclusion

This study demonstrates the importance of comparative phylogeographic studies to infer the evolutionary history of a region. The phylogeography of all three species differed even though they have overlapping distributions and are assumed to share the same evolutionary history. For *A. troschelii and A. interruptus*, long stretches of sandy coastline in the TEP acts as a barrier to dispersal, for *L. argentiventris* no barriers seem to prohibit dispersal in the TEP. My study demonstrates that in addition to past evolutionary and geographic information, knowledge of the life history of the organism is important to infer evolutionary history. I found distinct differences in gene flow based upon life history characteristics: in benthic spawners unsuitable habitat separating the biogeographic provinces in the TEP proved to be an effective barrier to dispersal. In contrast, for pelagic spawners barriers to dispersal were not effective in limiting gene flow. Barriers to dispersal found within the TEP may reveal why the TEP is plagued with a high rate of endemism and a low rate of biodiversity.

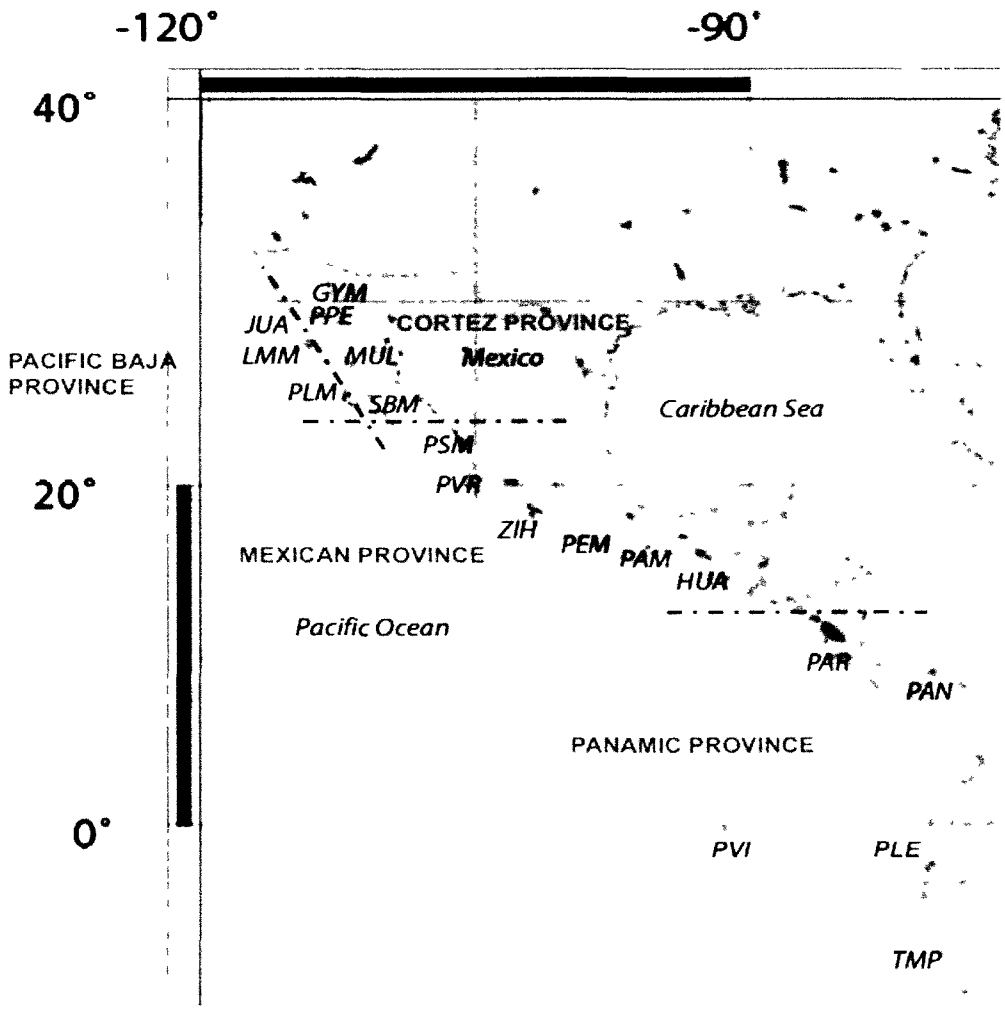

Figure 3.1. Map of the Tropical Eastern Pacific along with sample locations. Population abbreviations are explained in Table 3.1. Red typing designate Provinces. Dashed lines represent the proposed geographic breaks.

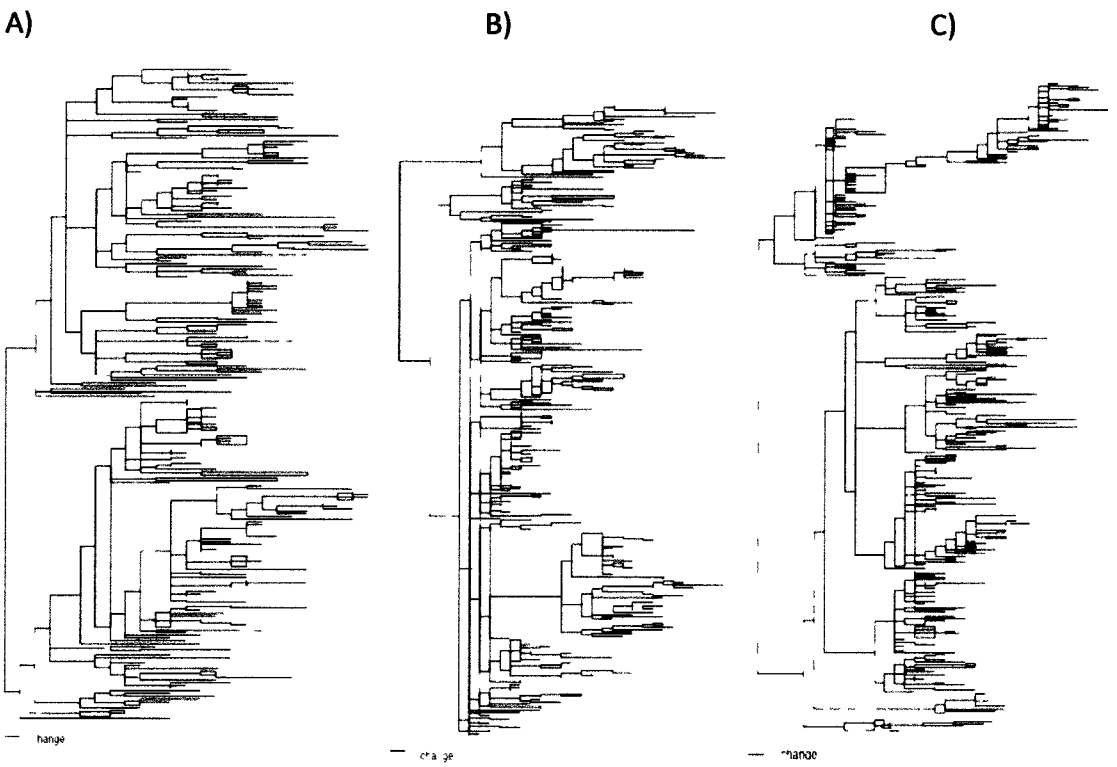

Figure 3.2. Maximum likelihood (ML) topologies for A) *A. troschelii,* B) *A. interruptus,* C) *L. argentiventris* based on two mitochondrial markers (ATPase8 and Dloop). *Abudefduf* ML is based on a TrN+I+G model; Anisotremus interruptus ML is based on a SYM+I+G model; *Lutjanus* ML is based on a HKY+I+G model.

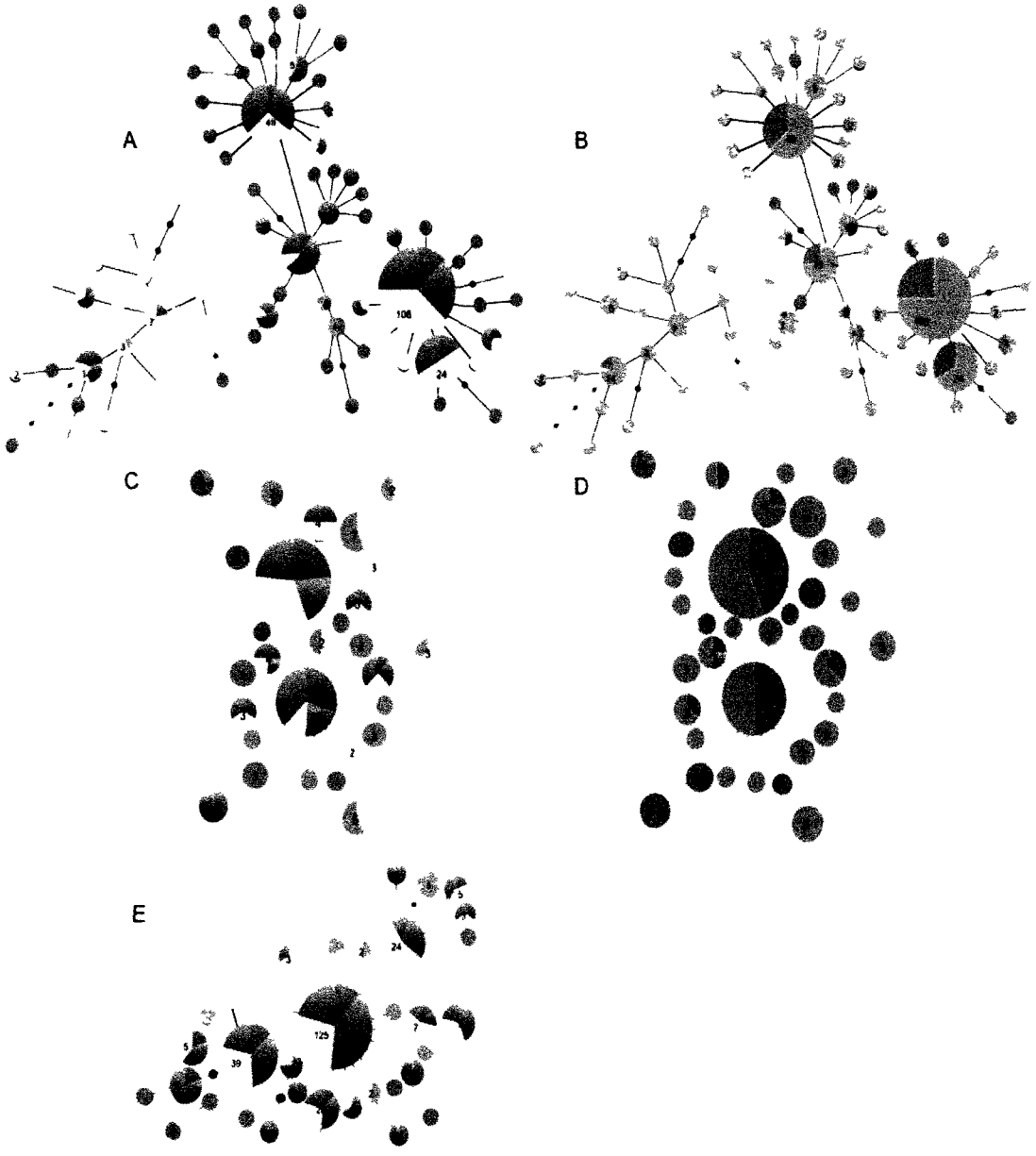

Figure 3.3. Minimum spanning networks based on parsimony informative characters found in ATPase8. Each circle represents a single haplotype. Each line represents a single step mutation, black circles refer to a haplotype not found in the dataset. Designated numbers refer to the number of individuals that contain the haplotype. Colors refer to provinces where the haplotype is found (Blue- Pacific Baja; Red-Cortez; Green-Mexican; Purple-Panamic). A) *A. troschelii*; B) *A. troschelii* (Mexican-Blue and Panamic-Red); C) *A. interruptus;* D) *A. interruptus* (Baja-red vs. Mexican and Panamic-blue); E) *L. argentiventris*.

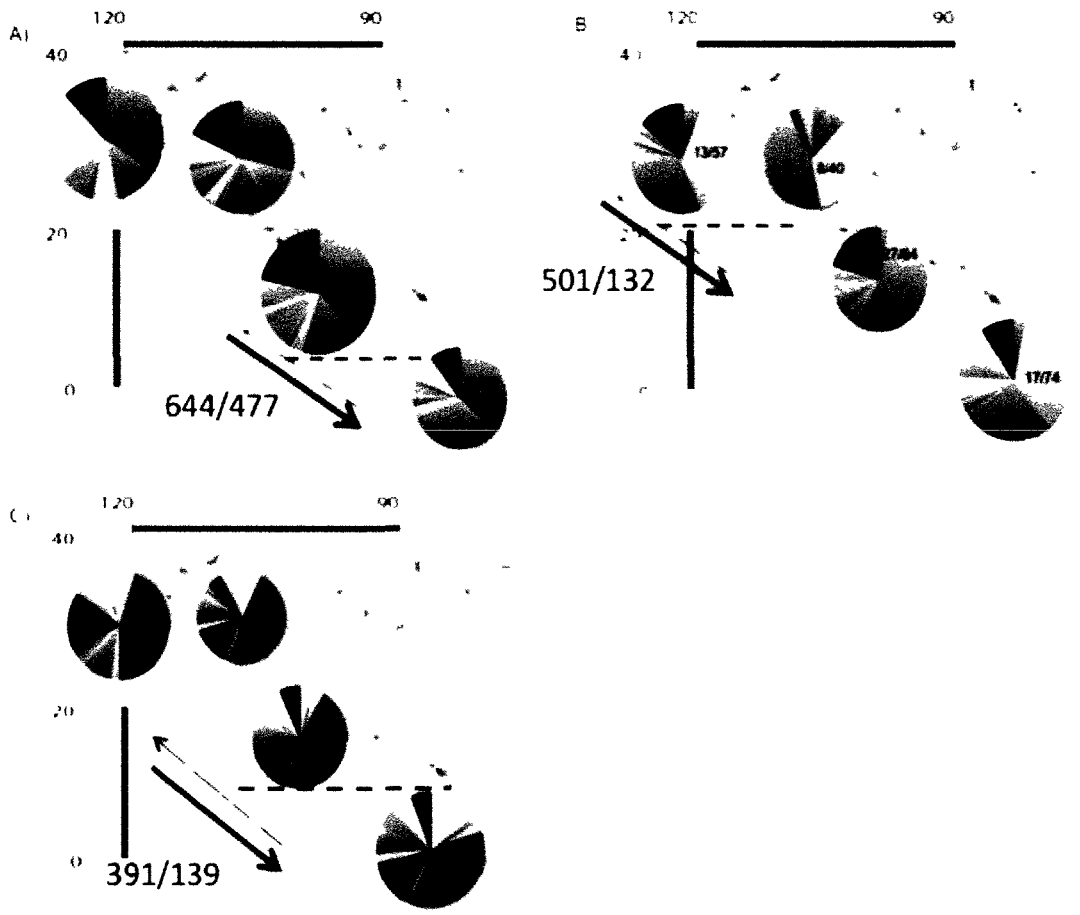

Figure 3.4. Geographic distribution of haplotypes. Each pie chart represents a population and haplotypes and frequencies found within each population. Brown arrows denote migratory movements between regions (see text) estimated using MIGRATE. Number of migrants given (South/North direction). The width of the arrow represents the strength of migration. A) *Abudefduf troschelii;* B*) Anisotremus interruptus*; C) *Lutjanus argentiventris*.

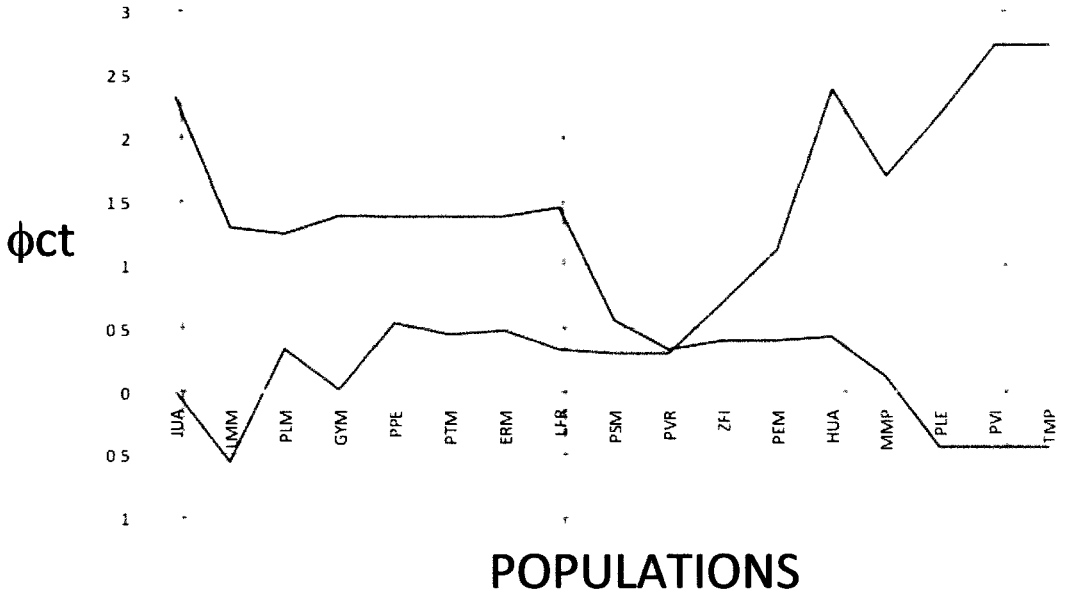

Figure 3.5. Plot based on populations subdivisions vs. ϕct values. Blue line represents *A. troschelii* populations; red line represents *A. interruptus;* and green line represents *L. argentiventris*. Dashed lines indicate the two highest ϕct values for each dataset.

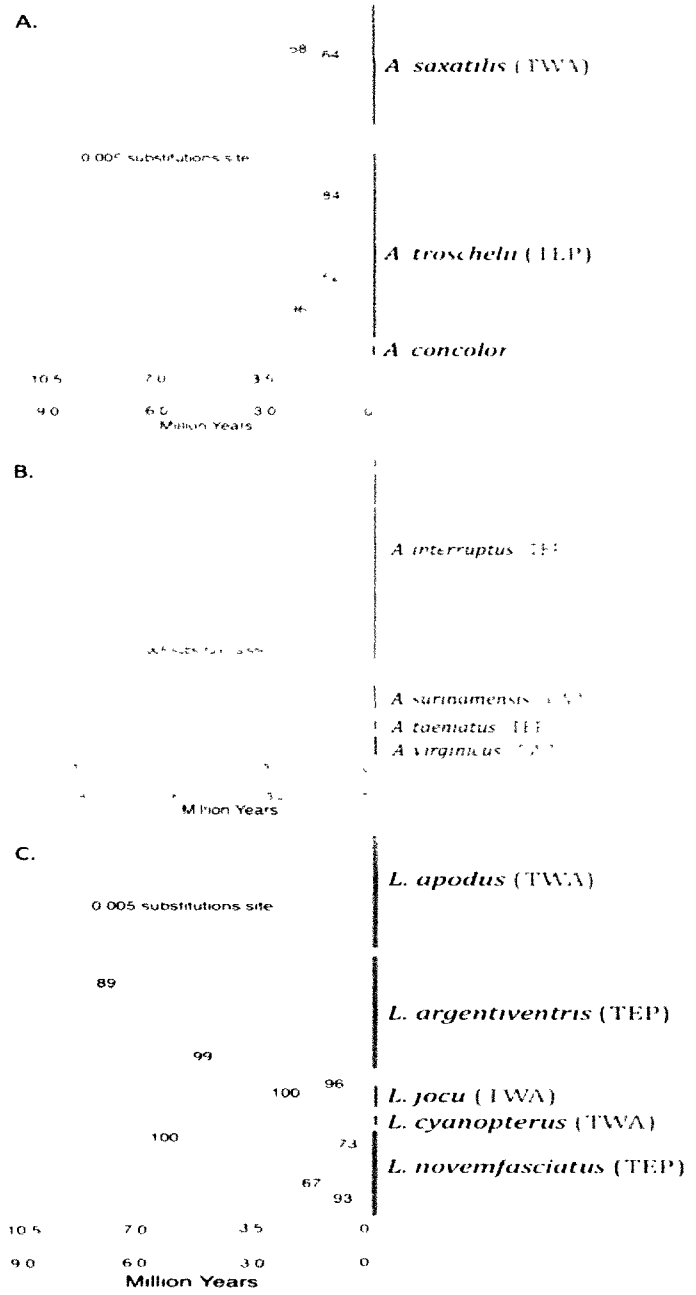

Figure 3.6. Molecular phylogeny of a subset of the dataset for A) *Abudefduf sp.*, B) *Anisotremus sp.*, C) *Lutjanus sp*. UPGMA method to display the implementation of a molecular clock based on the closure of the Isthmus of Panama (3.1; 3.5myr) denoted with a red line. All topologies are scaled with one another.

		Abudefduf troschelii				*Anisotremus interruptus*				*Lutjanus agentiventris*			
Location	Code	Ntaxa	H_n	H_d	π	Ntaxa	H_n	H_d	π	Ntaxa	H_n	H_d	π
Juanico, MX	JUA					19	17	0 9883	0 0145				
Lopez Mateos, MX	LMM	20	19	0 9974	0 0153	19	19	0 9942	0 0141	17	17	1 0000	0 0271
Puerto Lobos, MX	PLM	21	21	1 0000	0 0178	20	20	1	0 0134	21	20	0 9952	0 0182
Guaymas, MX	GYM	17	17	1 0000	0 0186	20	18	0 98947	0 0145	22	22	1 0000	0 0163
Puerto Penasco, MX	PPE	23	23	1 0000	0 0188					34	32	0 9964	0 0170
Puertocitos, MX	PTM	15	15	1 0000	0 0197					-	-	-	-
Mulege, MX	MUL	20	20	1 0000	0 0164					29	28	0 9975	0 0153
South Baja, MX	SBM	19	19	1 0000	0 0194	20	20	1	0 0135	13	13	1 0000	0 0153
Mazatlan, MX	PSM	16	16	1 0000	0 0193	16	16	1	0 0138	18	17	0 9935	0 0166
Puerto Vallarte, MX	PVR	23	23	1 0000	0 0176	16	16	1	0 0131	21	21	1 0000	0 0161
Zihuatanejo, MX	ZFI	26	26	1 0000	0 0177	19	19	1	0 0147	22	22	1 0000	0 0137
Puerto Escondido, MX	PEM	20	20	1 0000	0 0171	-	-	-	-	-	-	-	-
Huatulco, MX	HUA	22	22	1 0000	0 0182	13	12	0 9872	0 0134	27	27	1 0000	0 0187
Punta Arenas, Costa Rica	PAR	-	-	-	-	-	-	-	-	16	16	1 0000	0 0122
Panama City, Panama	MMP	21	21	1 0000	0 0165	25	25	1	0 0137	15	15	1 0000	0 0187
Puerto Lopez, Ecuador	PLE	19	19	1 0000	0 0186	22	22	1	0 0140	20	20	1 0000	0 0174
Galapagos, Ecuador	PVI	16	16	1 0000	0 0174	-	-	-	-	-	-	-	-
Tumbes, Peru	TMP	20	20	1 0000	0 0196	26	26	1	0 0142	22	21	0 9957	0 0128
	TOTAL	318	300	0 9995	0 0182	235	219	0 9993	0 0142	297	261	0 9988	0 0163

Table 3.1. Locations and abbreviations for each population. For each population the number of individuals sampled per population; number of base pairs sequenced per marker; the number of haplotypes per population (Hn); haplotype frequency (Hd); nucleotide diversity (π) are listed.

A)

Location	ERM	GYM	HUA	IMM	LFR	PCM	PEM	PLE	PPE	PSM	PTM	PVI	PVR	RCP	TMP	ZFI
ERM	*	0 234	**0.036**	0 108	**0.018**	0 081	**0.027**	**0.027**	0 108	0 081	**0.018**	0 162	**0.045**	0 180	**0.027**	**0.027**
GYM	0 012	*	0 523	0 468	0 144	0 360	0 288	0 270	0 928	0 847	0 351	0 324	0 964	0 081	0 459	0 631
HUA	**0.035**	0 005	*	0 351	**0.009**	0 126	0 820	**0.036**	0 306	0 667	0 387	0 108	0 820	**0.045**	**0.045**	0 910
IMM	0 034	0 003	0 003	*	**0.027**	0 153	0 486	**0.027**	0 676	0 441	0 189	**0.027**	0 685	**0.036**	**0.018**	0 649
LFR	**0.049**	0 018	**0.048**	**0.039**	*	**0 009**	**0.027**	0 315	0 099	0 108	**0.045**	0 550	0 072	0 189	0 063	**0.036**
PCM	0 030	0 004	0 015	0 022	**0.049**	*	0 063	**0.000**	0 577	0 180	**0.018**	0 054	0 477	**0.018**	**0.000**	0 297
PEM	**0.051**	0 005	0 015	0 004	**0.054**	0 030	*	**0.027**	0 198	0 459	0 486	0 054	0 405	**0.018**	**0.018**	0 613
PLE	**0.040**	0 012	**0.030**	**0.052**	0 007	**0.056**	**0.037**	*	**0.036**	0 477	**0.018**	0 288	0 072	0 856	0 063	**0.045**
PPE	0 015	-0 023	0 004	0 009	0 026	-0 007	0 018	**0.034**	*	0 586	0 234	0 099	0 982	**0.018**	0 054	0 595
PSM	0 022	0 020	-0 012	0 000	0 019	0 011	0 001	0 004	0 008	*	0 243	0 198	0 784	0 216	**0.036**	0 748
PTM	**0.044**	0 002	0 001	0 018	**0.044**	**0.044**	0 003	**0.047**	0 009	0 016	*	0 189	0 468	**0.018**	0 099	0 162
PVI	0 023	0 009	0 031	**0.058**	0 006	0 053	0 043	0 004	0 025	0 016	0 014	*	0 162	0 414	0 279	**0.027**
PVR	**0.027**	0 023	0 015	0 010	0 031	0 004	-0 001	0 027	0 020	0 016	0 002	0 016	*	**0.000**	0 063	0 775
RCP	0 016	0 025	**0.050**	**0.064**	0 010	**0.049**	**0.063**	-0 019	**0.038**	0 020	**0.056**	0 001	**0.038**	*	**0.018**	**0.009**
TMP	**0.049**	-0 002	**0.041**	**0.059**	0 038	**0.075**	**0.050**	0 042	0 032	**0.031**	0 040	0 008	0 028	**0.050**	*	**0.009**
ZFI	**0.032**	0 010	-0 017	0 009	**0.054**	0 005	0 006	**0.035**	-0 004	-0 014	0 020	**0.050**	-0 013	**0.051**	**0.049**	*

B)

Location	GYM	JUA	LMM	LPA	MMP	PAM	PLE	PLM	PTM	PVR	TMP	ZFI
GYM	*	0 559	0 874	0 982	0 541	0 261	0 234	0 505	0 117	0 261	0 532	**0.018**
JUA	-0 010	*	0 369	0 279	0 072	**0.027**	0 054	**0.045**	**0.018**	**0.027**	**0.018**	**0.000**
LMM	-0 019	0 001	*	0 865	0 505	0 108	0 207	0 459	**0.036**	**0.027**	0 541	**0.018**
LPA	-0 025	0 008	-0 018	*	0 604	0 468	0 387	0 153	0 234	0 081	0 270	0 054
MMP	0 004	0 027	-0 002	-0 005	*	0 189	0 441	**0.045**	0 054	0 117	0 595	**0.009**
PAM	0 006	**0.061**	0 027	0 002	0 017	*	**0.018**	**0.027**	0 712	0 135	**0.090**	0 126
PLE	0 007	0 031	0 007	0 001	0 001	**0.036**	*	0 090	**0.036**	0 153	0 423	0 261
PLM	0 001	**0.028**	0 003	0 011	**0.018**	**0.046**	0 013	*	0 009	**0.045**	0 324	**0.009**
PTM	0 019	**0.068**	**0.041**	0 014	0 025	-0 015	**0.026**	**0.055**	*	0 505	**0.045**	0 405
PVR	0 013	**0.059**	**0.037**	0 021	0 019	0 024	0 017	**0.033**	0 005	*	0 072	0 180
TMP	-0 003	**0.037**	-0 004	0 007	0 003	0 025	-0 002	0 003	**0.035**	0 027	*	**0.027**
ZFI	**0.042**	**0.075**	**0.044**	0 028	**0.038**	0 021	0 011	**0.047**	0 000	0 015	**0.040**	*

C)

Location	HUA	LAG	LMM	LMP	MMP	MUL	PAR	PLE	PLM	PSM	PVR	SBM	TAJ	ZFI
HUA	*	0 225	0 559	**0.009**	0 595	**0.027**	**0.009**	**0.018**	0 577	0 180	0 108	0 117	**0.036**	**0.000**
LAG	0 010	*	0 883	0 396	0 631	0 324	0 559	0 198	0 631	0 937	0 703	0 712	0 514	0 477
LMM	-0 006	-0 014	*	0 279	0 991	0 730	0 288	0 919	0 964	0 928	0 793	0 676	0 874	0 162
LMP	**0.033**	0 000	0 003	*	0 180	0 216	0 874	0 135	0 216	0 297	0 306	0 586	0 685	0 766
MMP	-0 006	-0 008	-0 033	0 013	*	0 793	0 126	0 496	0 964	0 559	0 946	0 225	0 577	0 234
MUL	**0.026**	0 004	-0 007	0 004	0 014	*	0 144	0 333	0 441	0 288	0 847	0 072	0 676	0 207
PAR	**0.055**	0 005	0 006	-0 016	0 022	0 009	*	0 225	0 144	0 667	0 387	0 532	0 928	0 919
PLE	**0.035**	0 011	-0 018	0 014	-0 008	0 002	0 015	*	0 342	0 577	0 559	0 126	0 802	0 279
PLM	-0 002	0 007	0 022	0 009	-0 027	0 000	0 014	0 003	*	0 928	0 703	0 577	0 369	0 324
PSM	0 012	-0 018	0 023	0 003	-0 012	0 007	-0 009	0 007	0 020	*	0 784	0 847	0 892	0 586
PVR	0 015	-0 006	0 014	0 002	-0 023	-0 009	0 002	0 004	-0 013	0 012	*	0 261	0 838	0 532
SBM	0 025	-0 012	-0 010	0 008	0 009	0 016	-0 004	0 021	-0 009	0 020	0 009	*	0 667	0 730
TAJ	**0.042**	-0 002	-0 014	0 007	-0 003	-0 008	-0 018	-0 012	0 000	0 018	0 012	-0 007	*	0 973
ZFI	**0.054**	-0 002	0 009	-0 009	0 011	0 006	-0 019	0 009	0 006	-0 007	0 003	-0 007	-0 022	*

Table 3.2. Population pairwise (ϕst) comparisons (bottom diagonal) and associated P values for A) *A. troschelii*; B) *Anisotremus interruptus;* C) *L. argentiventris.* Bold values are significant.

A)

Source of variation	SS	Variance components	Variation	Statistics (P-value)
Among groups (Fct)	53.473	0.15153	1.58%	0.01579 (0.002)
Among populations within groups (Fsc)	69.382	0.02558	0.27%	0.0027 (0.005)
Within populations (Fst)	1902.586	9.41874	98.15%	0.01846 (0.318)
Total	2025.441	9.59585		

B)

Source of variation	SS	Variance components	Variation	Statistics (P-value)
Among groups (Fct)	75.804	0.1672	1.56%	0.01563 (0.000)
Among populations within groups (Fsc)	146.543	0.08959	0.84%	0.00851 (0.000)
Within populations (Fst)	3143.297	0.44285	97 60%	0.024 (0.000)
Total	2025.441	9.59585		

C)

Source of variation	SS	Variance components	Variation	Statistics (P-value)
Among groups (Fct)	44.278	0.17968	1.87%	0.01868
Among populations within groups (Fsc)	78.577	0.021	0.22%	0 00222
Within populations (Fst)	1902.586	9.41874	97.91%	0.02086
Total	2025.441	9.59585		

D)

Source of variation	SS	Variance components	Variation	Statistics (P-value)
Among groups (Fct)	42.596	0.25781	2.38%	0.0238 (0.000)
Among populations within groups (Fsc)	179.751	0.121	1.12%	0.0114 (0.000)
Within populations (Fst)	3143.297	10.44285	96.55	0.035 (0.000)
Total	3365.644	10.82166		

Table 3.3. Analysis of molecular variance (AMOVA) based on partitioning of populations into multiple groups A) *A. troschelii*: Pacific Baja, Baja Peninsula, Cortez, Panamic; B) *A. troschelii*: Mexico vs. Panamic; C) *A. interruptus*: Pacific Baja, sea of Cortez, Cortez, Panamic; D) *A. interruptus*: Baja, Cortez vs. Panamic.

Species pair (Pacific vs. Atlantic)	Average pairwise differences			Average divergence rates					
				Dloop		ATPase		Both genes	
	Dloop	ATPase	Both genes	3.1myr	3 5myr	3 1myr	3.5myr	3 1myr	3 5myr
A. interruptus vs. A. surinemensis	0.037	0.013	0.026	1.204	1.066	0.413	0.366	0 825	0.731
A. taniatus vs. A.virginicus	0.044	0.035	0.040	1.422	1.260	1.135	1.005	1.279	1.133
A. troschelii vs. A. saxatilis	0.062	0.043	0.044	1.990	1.762	1 382	1.224	1.415	1.253
L. argentiventris vs. L. jocu	0.064	0.021	0.038	2.079	1.841	0.668	0.592	1.213	1.074
L. novemfasciatus vs. L. cyanopterus	0.105	0.022	0.053	3.403	3.014	0.718	0.636	1.725	1.528

Table 3.4. List of average pairwise distances and divergence rates for each species pair listed. Rates of divergence were estimated for each gene and for the two genes combined by dividing the average pairwise distance by 3.1 and 3.5myr, the estimated date of the closure of the Isthmus of Panama.

Species	Regional Groups	θ_c	θ_v	g	Tc(years)
A. interruptus	All populations	0.612	4.180	435.9	159,007.2
A. troschelii	All populations	0.792	2.790	301.6	402,239.2
L. argentiventris	All populations	0.237	1.273	419.5	84,344.4

Table 3.5. Listed statistics for each data set. Parameters include: theta without(θ_c) and with growth (θ_v); growth (g), and coalescence time (Tc).

Conclusion

The major goal of my dissertation was to study the processes that affect population connectivity in Tropical American shorefish. Assessing the limitations of dispersal among marine organisms proves to be difficult due to lack of ecological and oceanographic information. With the advent of molecular techniques, we can estimate the geographic distribution of genealogies in order infer whether genetic patterns correspond to past and ongoing climatic events. For my dissertation I used molecular genetics to study evolutionary processes among shore fish found in Tropical America by comparing different taxonomic levels at different spatial scales.

My dissertation began with a broad overview of speciation in the genus *Holacanthus*. Speciation results when two populations become reproductively isolated for a significant amount of time. Understanding the forces that govern speciation is important in addressing the mechanisms that limit dispersal. An assessment of the evolutionary history of the genus *Holacanthus* found many cases of recent speciation events. The calibration of a molecular clock based on the rise of the Isthmus of Panama allowed me to place a timeline onto the phylogeny of *Holacanthus*. Surprisingly the clock was based on geminate clades instead of geminate species, which is more commonly found. Each clade consisted of 2-3 species that diverged from one another after the closure of the Isthmus 3.5 myr ago. These findings suggest that the rate of speciation in *Holacanthus* angelfish is high

compare to other species that reside in the same ocean, suggesting ecological differences may inhibit dispersal in angelfish.

For the last two chapters of my dissertation I performed comparative phylogeographic studies in order to test for genetic concordance across taxa to better understand the evolutionary history of a region and the factors that may limit dispersal. Interestingly, each comparative phylogeographic study performed for my dissertation did not find genetic concordance among taxa that are assumed to have shared the same evolutionary history. Some taxa showed no evidence of genetic breaks, while other taxa displayed genetic breaks at different geographic locations. Discrepancies in genetic signatures may be a result of 1) an environmental force acting as an effective barriers for some species but not for others or 2) ecological life history variation. For each chapter I sought to identify the factors that limit dispersal.

For the second chapter of my dissertation, I performed a comparison among fish that reside in two contrasting oceans, the TEP and TWA. The results of this study found that TEP species exhibited a higher level of genetic differentiation than the sister species found in the TWA. Because these are closely related species that have similar life history traits, this finding suggests oceanographic differences do have an affect on population connectivity. Secondly, *Abudefduf sp.* showed more genetic differentiation than *Lutjanus* species. Since *Abudefduf sp.* is a demersal spawner

with negatively buoyant eggs and does not form aggregations like *Lutjanus* individuals, this result suggests that spawning behavior and egg buoyancy may facilitate dispersal.

For the final chapter of my thesis, I performed a fine scale comparative phylogeographic study on three shorefish species, *Lutjanus argentiventris, Anisotremus interruptus,* and *Abudefduf troschelii*, that have overlapping distributions in the TEP. The goal of this project was to determine whether I find genetic concordance among species that have differing ecological requirements. The second goal of this study was to determine whether long stretches of unsuitable habitat act as true barriers to dispersal for these species. The results of this displayed different genetic signatures for all three species. *Lutjanus argentiventris* showed no genetic structure, suggesting high gene flow among populations. Both *Abudefduf troschelii* and *Anisotremus interruptus* showed regional division corresponding exactly to the biogeographic provinces found in the TEP, but differed in the location of the highest amount of regional subdivision. The highest amount of regional differentiation for *Abudefduf troschelii* was found in the southern TEP between the Mexican and Panamic Province. For *Anisotremus interruptus* regional subdivision was found in the Northern TEP between Baja and the rest of the TEP. Differences in genetic signatures among species that are assumed to have shared the same evolutionary history may be caused by differences in ecology. *Abudefduf*

sp. are benthic spawners that do not have pelagic eggs, opposite are *Lutjanus argentiventris* individuals which are broadcast spawners and form spawning aggregations. Intermediate to these two life histories lies *Anisotremus interruptus* which is a broadcast spawner but does not form spawning aggregations. The signatures I found are concordant with what is expected based on life history traits. Out of the three species, *L. argentiventris*, a broadcast spawner, showed less structure suggesting more gene flow, followed by *A. interruptus. Abudefduf troschelii* showed the most genetic structure than the other two. These results indicate that habitat fragmentation along the TEP may affect benthic spawners more than broadcast spawners.

Based on the results of my dissertation, the ecology of a species and the geological information of its distribution is necessary when inferring evolutionary history. This study exemplifies that intrinsic barriers such as oceanic currents and habitat fragmentation inhibit dispersal capabilities of species that are demersal spawners and have negatively buoyant eggs, more so than species that migrate to spawning aggregations, perform broadcast spawning, and have positively buoyant eggs. Based on the conclusions of this study, scientists may be able to use this information to predict the rate of migration in other species that reside in the Tropical America, only if the ecological requirements of the study species are known. Further species need to be added to this study to conclude that spawning behavior

and egg buoyancy do have an affect in dispersal. Also, the addition of genetic data, especially nuclear data, will provide more statistical support for the genetic patterns found in this study.

Bibliography

Allen, G.R., Robertson, D.R., 1997. An annotated checklist of the fishes of Clipperton Island, tropical eastern Pacific. *Rev. Biol. Trop.* 45, 813-843.

Allen, G.R., Steene, R., Allen, M., 1998. A Guide to Angelfishes and Butterflyfishes. Odyssey Publishing, Perth.

Alva-Campbell Y, Floeter S, Robertson D, Bellwood D, Bernardi G (2010) Molecular phylogenetics and evolution of Holacanthus angelfishes (Pomacanthidae). Molecular Phylogenetics and Evolution 56, 456-461.

Avise J (2004) Molecular Markers, Natural History, and Evolution Sinauer Associates, Inc. Sunderland, MA.

Barber, P.H., Bellwood, D.R., 2005. Biodiversity hotspots, evolutionary origins of biodiversity in wrasses (*Halichoeres*, Labridae) in the Indo-Pacific and New World tropics. Mol. Phyl. Evol. 35, 235-253.

Banford HM, Bermingham E, Collett B (2004) Molecular phylogenetics and biogeography of transisthmian and amphi-Atlantic needlefishes (Belonidae:

Strongylura and Tylosurus): perspectives on New World marine speciation. Molecular Phylogenetics and Evolution 31, 833-851.

Beerli P (2006) Comparison of Baysian and Maximum likelihood inference of population genetic parameters. Bioinformatics 22, 341-345.

Bellwood, D.R., van Herwerden, L., Konow, N., 2004. Evolution and biogeography of marine angelfishes (Pisces, Pomacanthidae) Mol. Phyl. Evol. 33, 140-155.

Bermingham E, McCafferty SS, Martin AP (1997) Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: Molecular systematics of fishes (eds. Kocher TD, Stepien CA), pp. 113-128. Academic Press, San Diego.

Bernardi G, Alva-Campbell Y, Gaparini J, Floeter S (2008) Molecular ecology, speciation, and evolution of the reef fish genus Anisotremus. Molecular Phylogenetics and Evolution 48.

Bleeker, P. 1857. Achtste bijdrage tot de kennis der vischfauna van Amboina. Acta Societatis Regiae Scientiarum Indo-Neêrlandicae 2, 1-102.

Bowen B, Bass A, Muss A, Carlin J, Robertson D (2006) Phylogeography of two Atlantic squirrelfishes (Family Holocentridae): exploring links between pelagic laval duration and population connectivity. Marine Biology 149, 899-913.

Briggs J (1961) The East Pacific Barrier and the distribution of marine shore fishes. Evolution 15, 545-554.

Briggs J (1974) Marine Zoogeography McGraw-Hill.

Cailliet GB, EJ, Cope J, Kerr L (2000) Biological Characteristic of Nearshore Fishes of California: A review of the Exising Knowledge.

Case TJ, Cody ML, Ezcurra E (2002) A new island biogeography of the Sea of Cortes Oxford, New York.

Coates A, G. (1997) Central America: A Natural and Cultural History Yale University Press, New Haven.

Chow, S., Hazama, K., 1998. Universal PCR primers for S7 ribosomal protein gene introns in fish. Mol Ecol. 7, 1255-1256.

Chung, K.C., Woo, N.Y.S., 1998. Phylogenetic relationships of the Pomacanthidae (Pisces, Teleostei) inferred from allozyme variation. J. Zool. , 246, 215-231.

Debelius, H., Tanaka, H., Kuiter, R.H., 2003. Angelfishes, a comprehensive guide to Pomacanthidae. TMC Publishing, Chorley,UK.

Domingues, V., Bucciarelli, G., Almada, V.C., Bernardi, G., 2005. Historical colonization and demography of the Mediterranean damselfish, *Chromis chromis*. Mol. Ecol. 14, 4051-4063.

Feder, H.M., 1966. Cleaning symbiosis in the marine environment. Symbiosis (ed. S.M. Henry), pp. 327–380. Academic Press, New York.

Feeley, M.W., Luiz Jr, O.J. and Zurcher, N., 2009. Colour morph of a probable queen angelfish *Holacanthus ciliaris*, from Dry Tortugas, Florida. J. Fish. Biol. 74, 2415-2421.

Felsenstein, J., 1985. Confidence limits on phylogenies, an approach using the bootstrap. Evolution 39, 783-791

Floeter, S.R., Rocha, L.A., Robertson, D.R., Joyeux, J.C., Smith-Vaniz, W., Wirtz, P., Edwards, A.J., Barreiros, J.P., Ferreira, C.E.L., Gasparini, J.L., Brito, A., Falcon, J.M., Bowen, B.W., Bernardi, G., 2008. Atlantic reef fish biogeography and evolution. J. Biogeog. 35, 22-47.

Hastings P (2000) Biogeography of the Tropical Eastern Pacific: distribution and phylogeny of chaenopsid fishes. Zoological Journal of the Linnean Society 128, 319-335.

Jackson JBC, Budd AF, Coates A, G. (1996) Evolution and Environment in Tropical America The University of Chicago Press, Chicago, IL.

Jordan DS (1908) The law of geminate species. American Naturalist XLII, 73-80.

Knowlton N, Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. Proceedings Royal Society of London B 265, 2257-2263.

Kuhner M, Yamato J, Felsenstein J (1998) Maximum likelihood estimation of population growth rates based on the coalescent. Genetics 149, 429-434.

Lee, W.J., Conroy, J., Howell, W.H., Kocher, T.D., 1995. Structure and evolution of teleost mitochondrial control regions. J. Mol. Evol. 41, 54–66.

Leis JM (1991) The pelagic stage of reef fishes. In: The Ecology of Fishes on Coral Reefs (ed. Sale PF), pp. 183-230. Academic Press, San Diego.

Lessios HA (2008) The Great American Schism: Divergence of marine organisms after the rise of the Central American Isthmus. Annual Rev. Ecol. Evol. Syst. 39, 63-91.

Martinez-Andrade F (2003) A comparison of life histories and ecological aspects among snappers (Pisces: Lutjanidae), Louisiana State University.

Luiz Jr., O.J., 2003. Colour morphs in a queen angelfish *Holacanthus ciliaris* (Perciformes, Pomacanthidae) population of St. Paul's Rocks, NE Brazil. Trop. Fish Hob. 51, 82–90.

Munro J, Gaut V (1973) The spawning seasons of Caribbean reef fishes. Journal of Fish Biology: 567-68.

Muss A, Robertson D, Stepien C, al. e (2001) Phylogeography of Ophioblennius: the role of ocean currents and geography in reef fish evolution. Evolution 55, 561-572.

Palumbi, S., Martin, A., Romano, S., McMillian, W.O., Stice, L., Grabowski, G., 1991. The Simple Fool's Guide to PCR. University of Hawaii, Honolulu. 46 pp.

Palumbi SR (1992) Marine speciation in a small planet. TREE 7, 114-117.

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. ARES 25, 547-572.

Posada D (1998) Modeltest: testing the model of DNA substitution Bioinformatics 14, 817-818.

Riginos C, Victor BC (2001) Larval spatial distributions and other early life-history characteristics predict genetic differentiation in eastern Pacific blennioid fishes. Proc. R. Soc. Lond. B 268, 1931-1936.

Robertson D (1988) Settlement and population dynamics of Abudefduf saxatilis on patch reefs in Caribbean Panama 2.

Robertson D, Allen G (2002) Shorefishes of the tropical eastern Pacific: an information system. Tropical Research Institute, Balboa, Panama.

Rocha, L.A., Robertson, D. R. Rocha, C. R. Van Tassell, J. L. Craig M. and Bowen, B. W. 2005a. Recent invasion of the tropical Atlantic by an Indo-Pacific coral reef fish. Molecular Ecology 14, 3921-3928.

Rocha L.A., Robertson D.R., Roman J., and Bowen B.W., 2005b. Ecological speciation in tropical reef fishes. Proc. Roy. Soc. London, Ser. B. 272, 573-579.

Rocha L, Lindeman K, Rocha C, Laessios H (2008) Historical biogeography and speciation in the reef fish genus Haemulon (Teleostei: Haemulidae. Molecular Phylogenetics and Evolution 48, 918-928.

Rocha L.A., Bowen B.W., 2008. Speciation in coral-reef fishes. J. Fish Biol. 72, 1101-1121.

Rosenblatt R (1974) Faunal areas. Science 186, 1028–1029.

Rozas J, Sanchez-DelBarrio J, Messeguer X, Rozas R (2003) DnaSP polymorphism analyses by coalescent and other methods. Bioinformatics 19, 2496-1497.

Sambrook J, Fritsch E, Maniatis T (1989) Molecular Cloning: A laboratory Manual, 2nd edn. Cold Spring Harbor, New York.

Sala, E., Aburto-Oropeza, O., Arreola-Robles, J.L., 1999. Observations of a probable hybrid angelfish of the genus *Holacanthus* from the Sea of Cortez, México. Pac. Sci., 53, 181-184.

Schneider S, Roessli D, Excoffier L (2000) Arlequin, a software package for population genetics data analysis. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva.

Shen, S.C., Liu, C.H., 1978. Clarification of the genera of the angelfishes (family Pomacanthidae). *Proc. Stud. Essays Commemoration Golden Jubilee Acad. Sin.*, pp. 57-77.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Smith-Vaniz WF, Collette BB and Luckhurst BE 1999. Fishes of Bermuda: history, zoogeography, annotated checklist, and identification keys. American Society of Ichthyologists and Herpetologists. Special Publication No. 4.

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods), 4 edn. Sinauer, Sunderland, MA.

Taylor MS, Hellberg ME (2003) Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. Science 299, 107-109.

Thomson, D.A., Findley, L.T., Kersitch, A.N., 2000. Reef fishes of the Sea of Cortez. The rocky shore fishes of the Gulf of California. The University of Texas Press, Austin.

Victor BE, Wellington G (2000) Endemism and the pelagic larval duration of reef fish in the eastern Pacific Ocean. Marine Ecology Progress Series 205, 241-248.

Walker BW (1960) The distribution and affinities of the marine fish fauna of the Gulf of California. Systematic Zoology Systematic Zool., 9(3-4):123-133, 123-133.

Wares J, Cunningham C (2001) Phylogreography and historical ecology of the north Atlantic intertidal. Evolution 55, 2455-2469.

Warner R, Palumbi S Larvae retention: genes or oceanography? Science 300, 1657-1658.

Waters JM, Roy M (2004) Phylogeography of a high-dispersal New Zealand sea-star: does upwelling block gene-flow? Molecular Ecology 13, 2797-2806.

Watson W, Walker H (1992) Larval development of sargo (Anisotremus davidsonii) and salema (Xenistius californiensis). Bulletin of Marine Science 51, 360-406.

Zukerkandl E, Pauling L (1967) Evolutionary divergence and convergence in proteins. Evolving genes and proteins Academic Press, New York.

Zwickl D (2006) Genetic Algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.