PATTERNS OF CONNECTIVITY IN CORAL REEF FISHES ACROSS THREE SPATIAL SCALES

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ZOOLOGY (ECOLOGY, EVOLUTION AND CONSERVATION BIOLOGY)

DECEMBER 2019

By

Richard Ronald Coleman

Dissertation Committee:

Brian Bowen, Chairperson
Mark Hixon
Robert Thomson
Robert Toonen
Margaret McManus

ACKNOWLEDGEMENTS

There are many people who have helped me along my personal journey to complete this dissertation and I am extremely thankful for your assistance, support, and guidance. Apologies if I have forgotten to add you to this list, I sincerely am grateful for every one of you.

The completion of this dissertation could not have occurred without the incredible support of many individuals. First and foremost, I would like to thank my advisor, Brian Bowen for taking the chance on me despite having received a C in my undergraduate genetics course. Without a doubt, I would not be as strong of a scientist if I did not receive your mentorship. You have provided me with many professional opportunities by supporting my interests even though they were oftentimes not immediately related to my dissertation. You have trusted me to participate in multiple expeditions around the world and for that I am truly grateful. On top of everything you have done for me professional you are also an awesome individual. Not many students have shared the personal experiences we have with their advisor and it is that personal dynamic that truly sets you apart as a mentor. I have had the dream graduate experience – I have become knowledgeable about many faucets in the field of marine science, I received excellent training and guidance during my time in graduate school, I got to travel to many exotic locations around the globe, and the end of it all, I can call my advisor a friend. I look forward to maintaining a working relationship and exciting about futures projects we will collaborate on.

I also need to dedicate a special section for the “To” in the ToBo Lab, Robert Toonen. Being in the ToBo lab is a special environment as you get a two for one in terms of mentorship. Despite being a “Bo” student you never made me feel like my time was not as valuable as any of your

other students. You have been instrumental to my success as a scientist and have been a great mentor.

I'd like to thank the other members of my committee (in alphabetical order) for their time, advice and support. Mark Hixon has provided invaluable insight into my work and every chapter was substantially improved by his input. Margaret McManus always made it a point to recognize that my knowledge was on par with other experts in the field, despite my hesitation as a graduate student. In a world where imposter syndrome regularly creeps in this form or validation was a well needed. Bob Thompson continues to inspire me with his dedication to research and his invaluable critiques throughout my dissertation.

I offer heartfelt thanks to the ToBo lab members, past and present. You have all been instrumental in my growth as a scientist and as a person. Joey DiBattista, Iria Fernandez Silva, Michelle Gaither, and Matthew Iacchei – you have all moved onto outstanding careers but I am fortunate to have maintain strong relationships with your both personally and professionally. Your mentorship throughout my academic career has been instrumental to my success. All of the past and present members have come through in various ways during my time including, Kim Andrews, Evan Barba, Chris Bird, Iain Caldwell, Sean Canfield, Emily Conklin, Josh Copus, Annick Cros, ʻAleʻalani Dudoit, Zac Forsman. Mykle Hoban, JP Hobbs, Garrett Johnson, Erika Johnston, Cassie Kaʻapu-Lyons, Ingrid Knapp, Derek Kraft, Eileen Nalley, Jon Puritz, Derek Skillings, Zoltan Szabo, Molly Timmers, Eric Tong, Lindsey Veazey, Jan Vicente, Ellen Waldrop, Charley Westbrook, Jon Whitney, Christie Wilcox, and Johanna Wren.

Each of my chapters had some people that helped a lot on that particular part of this work: collecting samples, helping with lab work or data analysis, or just influential discussions.

Chapter 2: I want to thank Eric Mason and the crew at Dream Divers in Saudi Arabia, Nicolas Prévot at Dolphin Divers and the crew of the M/V Deli in Djibouti, the KAUST Coastal and Marine Resources Core Lab, the KAUST Reef Ecology Lab, Amr Gusti, the Administration of the British Indian Ocean Territory and Chagos Conservation Trust, as well as the University of Milano-Bicocca Marine Research and High Education Centre in Magoodhoo, the Ministry of Fisheries and Agriculture, Republic of Maldives, and the community of Maghoodhoo, Faafu Atoll. For assistance in collection efforts I thank Alfonso Alexander, Senifa Annandale, Howard Choat, Pat Colin, Lori Colin, Joshua Copus, Matthew Craig, Michelle Gaither, Brian Greene, Jean-Paul Hobbs, Garrett Johnson, Stephen Karl, Randall Kosaki, Cassie Lyons, David Pence, Mark Priest, Joshua Reece, D. Ross Robertson, Tane Sinclair-Taylor, Robert Thorne, and Robert Whitton. I also thank Michelle Gaither, Brent Snelgrove, Robert Toonen, and members of the ToBo lab for their assistance, logistical support, and feedback throughout this project. Thanks to Tane Sinclair-Taylor for the graphical abstract images

Chapters 3 and 4: I thank Matt Craig, Joshua Copus, Joseph DiBattista, Jeff Eble, Kimberly Tenggardjaja, Steve Karl, Randall Kosaki, Jo-Ann Leong, Cassie Lyons, Carl Meyers, Yannis Papastamatiou, David Pence, Zoltan Szabo, ʻAulani Wilhelm, Jill Zamzow, and the crew of the R/V Hiʻialakai for their logistical support and/or assistance in the field for collections in the Papāhanaumokuākea Marine National Monument (PMNM). Field sampling was made possible by the PMNM. For field assistance for Oʻahu collections I thank Anela Akiona, Keisha Bahr,

Evan Barba, Lindsay Bull, Gregory Burgess, Katia Capel, Bailey Carlson, Daniel Coffey, Delany Coleman, Joshua Copus, Chelsie Counsell, ‘Ale‘alani Dudoit, Kelle Freel, Campbell Gunnell, Griffin Holms, Kaleonani Hurley, Conor Jerolmon, Sarah Leitner, Cory Lescher, Claire Lewis, Dan Levy, Joshua Levy, Kaimana Lundquist, Sean Mahaffey, Eileen Nalley, Edna Diaz Negron, Julien Paillusseau, Kai Steuer, Kaliko Thomas, Sarah Tucker, Leon Weaver, Van Wishingrad, Julie Zill, and many others. For logistic support, I thank the University of Hawai’i’s Dive Safety Office, David Pence, Jason Jones, Daniel Coffey, Charley Westbrook, Ivor Williams, and the Hawai’i’s Sea Grant College Program. For RADseq library preparation and sequencing, I thank Chris Bird, Sharron Magnuson, and the Genomics Core Lab at Texas A & M Corpus Christie. Special thanks to members of the Fish Flow working group including Mark Hixon, Chelsie Counsell, Megan Donahue, Erik Franklin, Conor Jerolmon, Steve Karl, Jack Kittenger, Shilpa Lai, Margaret McManus, Anna Neuheimer, Brian Powell, and Eva Schemmel, for consultation and discussion.

For all chapters, many thanks to Shaobin Hou and the University of Hawaii’s Advanced Studies of Genomics, Proteomics, and Bioinformatics sequencing facility (ASGPB) for their assistance with DNA sequencing, as well as Mindy Mizobe, and Amy Eggers and the HIMB EPSCoR Core Genetics Facility for sequencing at HIMB. Thanks to the University of Hawai‘i Dive Safety Program (Dave Pence, Kevin Flanagan, Jason Jones Keoki Stender), the Hawai‘i Institute of Marine Biology office and fiscal staff.

I am extremely grateful to those individuals and agencies that have generously funded the research for this dissertation in particular the including the National Science Foundation (NSF)

Graduate Research Fellowship Program (GRFP) under award no. DGE-1329626 and the Dr. Nancy Foster Scholarship program under Award no. NA15NOS4290067. Both have provided invaluable guidance and training. Addition support was provided by NSF under awards no OCE-0929031 and OCE-0623678, the Seaver Institute, the Howard K.L. Castle Foundation grant #3846, Project# R/SB-17, sponsored by the University of Hawaii Sea Grant College Program, SOEST, under Institutional Grant No. NA14OAR4170071 from NOAA Office of Sea Grant, NOAA National Marine Sanctuaries Program MOA grant No. 2005-008/66882), King Abdullah University of Science and Technology (KAUST) Office of Competitive Research Funds under Award no. CRG-1-2012-BER-002 and baseline research funds from the National Geographic Society Grant 9024-11.

Finally, I must thank my family and friends, for without them, there is no way that I would be where I am today. I want to thank my parents Lawrence Coleman and Milagros Quintana Coleman, my brothers Michael Coleman, Terry Coleman, and Nyle Coleman, my brothers and sisters from another mother, Paul Afnan, Kennaya Lambert, Alfredo Vera, Julie Vera, and all my cousins, nieces and nephews for their lifetime of guidance and support. It has been nearly 15 years since I started this journey and you have all been there for me every step of the way. I love you guys. My Hawaii ohana here and on the mainland - I could not have done this without you. Your friendship and love help ground me. We also figured out ways to have a lot of fun! Though not an exhaustive list I have to give a special shout out to those individuals who have been instrumental for my mental well-being throughout this process and helping to shape to be the best version of myself - Keisha Bahr, Iain Caldwell, Jamie Caldwell, Danny Coffey, Adrienne Copeland, Rachel Dacks, Angela Richard Dona, Alessandro Dona, ʻAleʻalani Dudoit, Tim

Frankland. Caitlyn Genovese, Kaleo Hurley, Matt Iacchei, Ily Iglesias, Mel Mariano, Luke Nasegai, Jonathan Saupe, Nyssa Silbiger, and Kelly Yuka Walton, I love you all! Most of all I thank Maya Walton. We started graduate school together and you have been there for me since day one. You and your family have become my extended family and I cannot thank you enough for the emotional support you have given me throughout these years. Cheers! Cheers! Cheers!

Mahalo nui loa

ABSTRACT

The vast majority of reef fish have a life history consisting of a pelagic larval phase of typically 20 to 60 days and followed by larval settlement where they remain through their juvenile and adult phase. It is during the pelagic larval phase that nearly all dispersal across great distances is accomplished. Understanding connectivity and dispersal pathways, as well as identifying the underlying mechanisms influencing these patterns are essential to properly understand how biodiversity is generated and maintained in the sea. The scale in which these patterns can be identified can also illuminate evolutionary processes, and can inform conservation strategies. Since direct observation of larvae is impractical, a variety of methods have been developed to characterize connectivity and dispersal patterns in marine organisms. Here, I incorporated several different genetic based approaches to assess connectivity across a suite of spatial scales: across ocean basins (Indian and Pacific Oceans), across an isolated archipelago (Hawaiian Archipelago), and at the island scale (O‘ahu). From an ocean basin scale, the results of this work identified historic barriers to dispersal, refugia during the Pleistocene, and recovered cryptic diversity. At the archipelago and island scale, this work shows how biogeographic distribution can be predictive of dispersal potential. recovered previous unknown management units and showed the complex system of dispersal pathways and the role these systems play in regards to informing management strategies. By evaluating connectivity across different spatial scales, this work highlights the different processes facilitating evolution as well as enhancing our ability to inform conservation and management goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 INTRODUCTION

The vast majority of reef fish have a life history consisting of a pelagic larval phase of typically 20 to 60 days and followed by larval settlement where they remain through their juvenile and adult phase. It is during the pelagic larval phase that nearly all dispersal across great distances is accomplished (Leis & McCormick 2002). Previous studies that have investigated marine organisms with pelagic larvae often can often show little genetic differentiation across thousands of kilometers and are associated with high levels of gene flow and wide range distributions (Hellberg *et al.* 2002; Palumbi 2003). However, studies have also demonstrated high levels of self-recruitment and local larval retention in reef fishes illustrating that not all marine organisms exhibit broad-scale dispersal (Planes *et al.* 1998a; Swearer *et al.* 2002; Planes *et al.* 2009; Berumen *et al.* 2012; Jones 2015).

Since direct observation of larvae is impractical, a variety of methods have been developed to characterize connectivity and dispersal patterns in marine organisms. Common techniques include chemical tagging, hydrodynamic and biophysical models, and genetic based methods (Jones 2015). Genetic based approaches are well-suited for studying connectivity since identifying genetically differentiated populations only require minor differences in gene flow (Hellberg *et al.* 2002). Additionally, genetic signatures can be used to assign parents and offspring (e.g. parentage analysis) and therefore a direct measure of connectivity by identifying dispersal pathways from the natal spawning location to where the larvae ultimately settled.

Understanding connectivity and dispersal pathways, as well as identifying the underlying mechanisms influencing these patterns are essential to properly understand how biodiversity is generated and maintained in the sea. The scale in which these patterns can be identified can also

illuminate evolutionary processes, and can inform conservation and management strategies. For the studies presented here, I assessed connectivity across a suite of spatial scales: across ocean basins (i.e. Indian and Pacific Oceans), across an isolated archipelago, and at the island scale. By evaluating connectivity across different spatial scales, I was able to add to our understanding of the different processes facilitating evolution as well as inform conservation and management goals.

A phylogeographic approach was implemented to assess connectivity in the Regal Angelfish, *Pygoplites diacanthus*, a species whose distribution extends across the Red Sea, Indian and Pacific Oceans. Phylogeography is the study of the historical processes that are responsible for the contemporary distribution of genetic lineages within a species. These studies have been essential in identifying areas of endemism (DiBattista *et al.* 2015; DiBattista *et al.* 2016), cryptic species (Rocha *et al.* 2007; DiBattista *et al.* 2016), and the locations of biogeographic barriers (Rocha *et al.* 2007; Toonen *et al.* 2016) which often fall at the edge of biogeographic provinces (Briggs & Bowen 2013; Toonen *et al.* 2016). Often, after identifying cryptic genetic lineages, previously unobserved morphological characters are identified. For this study, two *P. diacanthus* morphotypes exist: one distributed in the Indian Ocean and Red Sea, the other one in the Pacific Ocean. By undertaking a phylogeographic approach we aimed to resolve the taxonomic distinction between the different morphotypes and identify historic geological processes that have promoted or restricted diversification within this species.

Assessing connectivity across the Hawaiian Archipelago provided an opportunity to compare connectivity against an evolutionary and contemporary framework. Hawaii is the most isolated archipelago in the world, with the closest land mass nearly 1000 km away, and its geological history has been extensively studied. Uniquely, every organism that is found there

originated elsewhere in the Pacific. Over the past few decades the archipelago has acted as natural laboratory to investigate connectivity patterns of more than 40 marine species representing a variety of organisms that are ecological and taxonomically distinct (see Toonen et al. (2011); Selkoe et al. (2016), and references therein). The capacity to assess connectivity across a broad spectrum of taxa, including endemic and wide-ranging species, has provided insight into the origins of biodiversity (Bird 2011; Eble *et al.* 2011b; Skillings *et al.* 2011), the evolutionary consequences of invasive species (Gaither *et al.* 2010a; Coleman *et al.* 2014), the ability to identify the location of ecological barriers to dispersal (Toonen *et al.* 2011), fishery related impacts (Iacchei *et al.* 2014), and the role of life history on influencing genetic structure and diversity (Selkoe *et al.* 2014; Selkoe *et al.* 2016b).

Phylogeographic and traditional population genetic approaches were typically used to characterize connectivity patterns for Hawaiian organisms. A common trait among these past studies was the use of targeted loci, including single markers, to describe connectivity within each species. Although, utilizing targeted loci methods are useful for describing patterns of connectivity across an evolutionary timescale, they have limited ability to identify contemporary connectivity patterns. By incorporating high-throughput sequencing, which generates thousands of loci in a single run, we now have the power to describe contemporary patterns that could not be accomplished in a practical manner using a targeted loci approach.

For this portion of my research, I identified two species of surgeonfishes, *Acanthurus triostegus* (*manini*) and *Ctenocheatus strigosus* (*kole*), that exhibit similar life history strategies and occupy similiar habitat. However, they differ greatly in their biogeographic distribution, thus providing an excellent framework to compare patterns of connectivity across Hawaii. Additionally, connectivity across the Hawaiian Archipelago has been previously described for

kole, a Hawaiian endemic, using a single mtDNA (Eble *et al.* 2009). *Manini* has a Indo-Pacific distribution and although connectivity descriptions in Hawaii are limited to between Oahu and Hawaii Island, there has been studies in other parts of the range based on allozymes and mtDNA (Planes *et al.* 1998a; Planes & Fauvelot 2002; Mirams *et al.* 2011). These previous descriptions provide a framework to compare connectivity between an evolutionary and contemporary timescale and will be the first archipelago-wide study describing population strucutre for reef fishes based on genomic-based sequencing. Here, I hypothesized that RADseq data will illuminate levels of population structure not yet described in each of these species as well as confirming the presence of ecological breaks found with other species.

Manini is a heavily targeted by recreational fisheries on the island of O‘ahu (e.g. sport, leisure or subsistence) and has been described as an exploited fish by Hawai‘i’s Division of Aquatic Resources (Longenecker *et al.* 2008). Current management strategies for *manini* are ineffective as the majority of females enter the fishery before reaching maturity (Longenecker *et al.* 2008; Schemmel & Friedlander 2017). In response, Native Hawaiian community leaders identified *manini* as a species of concern and an effort began to implement a community-based management plan for the eastern side of O‘ahu. To advance this goal, one of the first steps is to determine the source of fish populations.

A traditional population genetics approach is an effective means to characterize connectivity across ocean basins and archipelagos; however, it is ineffective at smaller spatial scales, such as individual islands, where it is difficult to detect signals of isolation within the existing pool of genetic diversity (Saenz-Agudelo *et al.* 2009). As most coral reefs fishes have limited home ranges as adults, the potential to disperse occurs at the larval planktonic stage (Leis & McCormick 2002). In understanding this dynamic, we can identify direct routes of dispersal

by assigning individuals back to the parents. Genetic parentage analyses has proven to be powerful tool at identifying this form of fine scale connectivity in a variety of taxa including butterflyfishes (Abesamis *et al.* 2017), clownfish (Jones *et al.* 1999; Saenz-Agudelo *et al.* 2012), gobies (D'aloia *et al.* 2013), groupers (Almany *et al.* 2013), snappers (Harrison *et al.* 2012), and surgeonfish (Christie *et al.* 2010).

For the island scale study, I used SNPs generated from RADseq to conduct a parentage analysis of *manini* to describe dispersal patterns around the island of O‘ahu. A concerted effort was made to describe dispersal pathways along the windward coast as part of an interdisciplinary project, involving ecologists, oceanographers and social scientists, titled *Fish Flow*. Within this working group the goal is to describe the route from spawning, including the use of biophysical models, the community level interactions after settlement, and identifying the route of consumption by local fishers. Although *Fish Flow* is focused on the windward coast, an attempt was made to describe connectivity across the entire island of O‘ahu. Based on the biophysical models and circulation patterns, I predicted that there would be high rates of local retention along the windward coast, particularly within Kāne‘ohe Bay. The results of this research will describe the flow of fish from reproduction to reef residency, which then can be used to identify key areas of larval productivity and recruitment. In turn, we can inform community-based management efforts by identifying propagule sources and sinks, and areas that are vulnerable to fishing pressure.

As the field of genetics continues to evolve and move towards genomics based efforts, the ability to identify evolutionary and contemporary processes will become easier to accomplish. By assessing connectivity across the spatial scales using a variety of genetic methods, I intend to showcase how different scales can be used to inform different processes.

Some techniques can lend themselves better to understanding evolutionary processes such as evaluating connectivity across a species range using targeted loci. Whereas, a parentage analysis provides more refinement for identifying connectivity on an island scale which can be valuable for informing management strategies, as well as for describing dispersal potential which can contribute to evolutionary processes. I anticipate that this research will enhance our growing knowledge of dispersal and connectivity in coral reef fishes and how these patterns promote biodiversity in both an evolutionary and a contemporary framework.

CHAPTER 2 REGAL PHYLOGEOGRAPHY: RANGE-WIDE SURVEY OF THE MARINE ANGELFISH *PYGOPLITES DIACANTHUS* REVEALS EVOLUTOINARY PARTITION BETWEEN THE RED SEA, INDIAN OCEAN AND PACIFIC OCEAN

ABSTRACT

The regal angelfish (*Pygoplites diacanthus*; family Pomacanthidae) occupies reefs from the Red Sea to the central Pacific, with an Indian Ocean/Rea Sea color morph distinct from a Pacific Ocean morph. To assess population differentiation and evaluate the possibility of cryptic evolutionary partitions in this monotypic genus, we surveyed mtDNA cytochrome *b* and two nuclear introns (S7 and RAG2) in 547 individuals from 15 locations. Phylogeographic analyses revealed four mtDNA lineages (d = 0.006 – 0.015) corresponding to the Pacific Ocean, the Red Sea, and two admixed lineages in the Indian Ocean, a pattern consistent with known biogeographical barriers. Christmas Island in the eastern Indian Ocean had both Indian and Pacific lineages. Both S7 and RAG2 showed strong population-level differentiation between the Red Sea, Indian Ocean, and Pacific Ocean (Φ_{ST} = 0.066 – 0.512). The only consistent population sub-structure within these three regions was at the Society Islands (French Polynesia), where surrounding oceanographic conditions may reinforce isolation. Coalescence analyses indicate the Pacific (1.7 Ma) as the oldest extant lineage followed by the Red Sea lineage (1.4 Ma). Results from a median-joining network suggest radiations of two lineages from the Red Sea that currently occupy the Indian Ocean (0.7 – 0.9 Ma). Persistence of a Red Sea lineage through Pleistocene glacial cycles suggests a long-term refuge in this region. The affiliation of Pacific and Red Sea populations, apparent in cytochrome *b* and S7 (but equivocal in RAG2) raises the hypothesis that the Indian Ocean was recolonized from the Red Sea, possibly more than once.

Assessing the genetic architecture of this widespread monotypic genus reveals cryptic evolutionary diversity that merits subspecific recognition

INTRODUCTION

The majority of reef fishes have a pelagic larval phase typically lasting 20 to 60 days, followed by settlement at a location where they remain through juvenile and adult phases. It is during the pelagic larval phase that nearly all dispersal is accomplished, sometimes across great distances (Leis & McCormick 2002; Hellberg 2009). However, even closely related species with similar life histories can show markedly different genetic structure across their respective ranges (Rocha *et al.* 2002; DiBattista *et al.* 2012). Despite these differences in realized dispersal, genetic partitions frequently align with boundaries between biogeographic provinces, which mark abrupt changes in species composition accompanied by obvious geological or oceanographic barriers (Kulbicki *et al.* 2013; Bowen *et al.* 2016). However, phylogeographic reef surveys usually examine genetic partitions both within and between congeneric species (e.g. Robertson et al., 2006; Leray et al., 2010; DiBattista et al., 2013; Gaither et al., 2014; Ahti et al . 2016; Waldrop et al., 2016). Less attention has been paid to monotypic genera, and it is unknown whether these species have evolutionary or ecological traits that promote species cohesion across time.

The family Pomacanthidae (marine angelfishes) is comprised of more than 85 species across seven genera. All of the genera have at least eight species (*Centropyge* has more than 30) except for the monotypic genus *Pygoplites*. The regal angelfish, *Pygoplites diacanthus* (Boddaert 1772), has a wide distribution from East Africa and the Red Sea to the Tuamotu Archipelago in the central Pacific. This distribution encompasses four biogeographic provinces (Fig. 2 in Briggs and Bowen, 2013): the Indo-Polynesian Province (IPP), the Sino-Japanese Province, the Western

Indian Ocean Province, and the Red Sea Province (which includes the Gulf of Aden; see Briggs and Bowen, 2012). Additionally, the range of *P. diacanthus* spans the Indo-Pacific Barrier, an episodic land bridge separating Pacific and Indian Ocean fauna during low sea levels associated with glaciation (Randall 1998; Rocha *et al.* 2007). *Pygoplites* diverged from the sister genus *Holacanthus* about 7.6 – 10.2 Ma (Alva-Campbell *et al.* 2010), and is monotypic despite occupying a very broad range and a variety of ecological conditions.

Randall (2005b) noted coloration differences between an Indian Ocean morph, with a yellow chest, and a Pacific Ocean morph with a gray chest and less yellow coloring on the head (Fig. 2.1), invoking the possibility of nomenclatural recognition of the two morphotypes. Historically color has been used for species delineation in reef fishes, however, coloration alone can be a deceptive foundation for taxonomical classification; molecular tools have been useful for identifying cryptic genetic partitions and resolving taxonomic uncertainty over color morphs (McMillan *et al.* 1999; Schultz *et al.* 2006; Drew *et al.* 2008; Drew *et al.* 2010; DiBattista *et al.* 2012; Gaither *et al.* 2014; Ahti *et al.* 2016; Andrews *et al.* 2016)

Here we obtained samples from across the range of *P. diacanthus* to assess genetic connectivity with mitochondrial (mtDNA) and nuclear (nDNA) markers. Our sampling allowed us to test for cryptic evolutionary partitions and evaluate the hypothesis of taxonomic distinction between Indian and Pacific morphotypes. We were further motivated to resolve the ecological and evolutionary conditions that restrict diversification within the genus *Pygoplites*, the sole monotypic genus in an otherwise speciose family of fishes.

MATERIAL AND METHODS

Sample Collections

Between 2004 and 2014, 547 tissue samples (primarily fin clips) of *P. diacanthus* were collected from 15 locations across the species distribution (Fig. 2.1), using nets and pole-spears while scuba diving or snorkeling. Tissues were preserved in salt-saturated DMSO buffer (Amos & Hoelzel 1991) and stored at room temperature. Total genomic DNA was isolated from preserved tissue following the "HotSHOT" method of Meeker *et al.* (2007) and stored at -20°C. Due to variation in DNA amplification and sequence resolution, not all specimens were resolved at all three loci outlined below, hence sample sizes in Fig. 2.1 do not match samples sizes provided in the tables.

MtDNA Analyses

A 568-base pair (bp) fragment of the mtDNA cytochrome *b* (cyt *b*) gene was resolved to identify the maternal lineage of each individual using the forward primer (5'-GTGACTTGAAAAACCACCGTTG-3') (Song *et al.* 1998) and reverse primer (H15573; 5'-AATAGGAAGTATCATTCGGGTTTGAT-3') (Taberlet *et al.* 1992). PCR was performed in 10 µl reactions containing 10-15 ng of DNA, 5 µl of premixed BioMixRed^{TM} (Bioline, Inc., Springfield, NJ, USA), 0.2 µM primer for each primer, and nanopure water (Thermo Scientific Barnstead, Dubuque, IA, USA) to volume and using the following conditions: 4 min at 94°C, 35 cycles of denaturing for 30 s at 94°C, annealing for 30 s at 50°C, extension for 45 s at 72°C, and a final extension for 10 min at 72°C.

PCR products were visualized using a 1.5% agarose gel with GelStar^{TM} (Cambrex Bio Science Rockland, Rockland MA, USA) and then purified by incubating with 0.75 units of Exonuclease and 0.5 units of Shrimp Alkaline Phosphatase (ExoSAP; USB, Cleveland, OH, USA) per 7.5 µl of PCR product for 30 min at 37°C, followed by 15 min at 85°C. DNA

sequencing was performed using fluorescently-labeled dideoxy terminators on an ABI 3730XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) at the University of Hawai‘i Advanced Studies of Genomics, Proteomics and Bioinformatics sequencing facility.

Sequences were aligned and edited using GENEIOUS v.8.0.3 (Gene Codes, Ann Arbor, MI, USA) and unique sequences were deposited into GenBank (Accession numbers: RAG2, KU885737 - KU885756; S7, KU885757 - KU885843; cyt *b*, KU885844 - KU885892). A model for DNA sequence evolution was selected using the program JMODELTEST v.2.1 (Guindon & Gascuel 2003; Darriba *et al.* 2012). The best-fit model of TIM1+G (gamma=0.0760) was identified by the Akaike Information Criterion (*AIC*) and the closest matched used for subsequent analyses. Mean genetic distance between lineages was calculated in DNASP v.5.10 (Librado & Rozas 2009). A haplotype network was constructed for each locus with NETWORK v.4.6.1.1 (http://www.fluxus-engineering.com/network_terms.htm) using a median-joining algorithm (Bandelt *et al.* 1999) and default settings.

To estimate the time to most recent common ancestor (TMRCA), we formatted the data with BEAUTI v.1.4.7 and used a Bayesian MCMC approach in BEAST v.2.2.0 (Drummond & Rambaut, 2007). We conducted our analysis with a strict clock of 2% per million years between lineages (Bowen et al., 2001; Reece et al., 2010a) and used a coalescent tree prior assuming exponential growth. We used default priors under the HKY+G+I model of mutation, the closest available model, and ran simulations for 10 million generations with sampling every 1000 generations. Ten independent runs were computed to ensure convergence, and log files were combined and ages averaged across runs using TRACER v.1.6 (http://tree.bio.ed.ac.uk/software/tracer/).

ARLEQUIN v.4.11 was used to generate haplotype and nucleotide diversity, as well as to

test for population structure (Excoffier *et al.* 2005). Genetic structure among and between regions was estimated by performing an analysis of molecular variance (AMOVA). Deviations from null distributions were tested with non-parametric permutation procedures ($N = 9999$). Pairwise Φ_{ST} statistics, an analog of Wright's F_{ST} that incorporates sequence evolution and divergence, were generated to assess structure and identify phylogeographic partitions. Locations where samples sizes were < 8 were excluded from population genetic analyses but included in overall diversity estimates. False discovery rates were controlled for and maintained at $\alpha = 0.05$ among all pairwise tests (Benjamini & Yekutieli 2001; Narum 2006).

Time since the most recent population expansion was estimated for each location using the equation $\tau = 2\mu t$, where t is the age of the population in generations and μ is the mutation rate per generation for the entire sequence (μ = number of bp x divergence rate within a lineage x generation time in years). We used a sequence divergence estimate within lineages of 1-2% per million years (Bowen *et al.* 2001; Reece *et al.* 2010a) to estimate population age. While generation time is unknown for *P. diacanthus*, we conditionally used the equation $T = (\alpha + \omega)/2$, where α is the age at first reproduction and ω is the age of last reproduction (or lifespan; Pianka, 1978). We obtained a generation time of 8.5 years based on an estimated reproductive age of 2 years and longevity of more than 15 years (Hinton 1962). Due to the tentative nature of generation time and mutation rates estimates, population age should be interpreted with caution, however rank-order comparisons among populations are robust to such approximations. Fu's F_S (Fu 1997) was calculated to test for evidence of selection or (more likely) population expansion using 10,000 permutations with significance determined at $P < 0.02$. A significant negative value of Fu's F_S is evidence for an excess number of alleles, as would be expected from a recent

population expansion, whereas, a significant positive value is evidence for a deficiency of alleles, as would be expected from a recent population bottleneck.

Nuclear DNA Analysis

We sequenced two nuclear loci: the recombination-activating gene 2 (RAG2) and intron 1 of the S7 ribosomal protein (S7). We resolved 431-bp of RAG2 using modified primers from Lovejoy (1999); the forward primer is 5'-SACCTTGTGCTGCAAAGAGA-3' and reverse primer is 5'-AGTGGATCCCCTTBTCATCCAGA-3'. We resolved 510-bp of S7 using primers S7RPEX1F and S7RPEX2R from Chowand Hazama (1998). For each intron, PCR was performed using the same reaction as described for cyt *b* but using the following temperature conditions: 5 min at 94°C, 35 cycles of denaturing for 30 s at 94°C, annealing for 30 s at 58°C, extension for 45 s at 72°C, and a final extension for 10 min at 72°C.

Allelic states with more than one heterozygous site were estimated using PHASE v.2.1 (Stephens & Donnelly 2003) as implemented in DNASP. Unique sequences were deposited in GenBank (Accession numbers: XXX - XXX). Three separate runs, each of 100,000 repetitions after a 10,000 iteration burn-in, were conducted for each locus; all runs returned consistent allele identities. Median-joining networks were created for each nuclear dataset as outlined above. To minimize circularity between closely related alleles, singletons were removed from the S7 network. However, this did not alter our overall interpretation of the results. Pairwise Φ_{ST} statistics were calculated for each nuclear dataset. The best-fit model of K80 and TPM1uf+I (proportion of invariable sites = 0.89) were identified for RAG2 and S7, respectively as determined by JMODELTEST. Observed heterozygosity (H_O) and expected heterozygosity (H_E)

were calculated for each locus and an exact test of Hardy-Weinberg Equilibrium (HWE) using 100,000 steps in a Markov chain was performed in ARLEQUIN.

Phylogenetic reconstruction

Phylogenetic reconstruction based on cyt *b* was rooted with *Holacanthus africanus* (family Pomacanthidae; GenBank accession number KC845351 and KC845352), as this genus is sister to *Pygoplites* (Bellwood et al., 2004; Alva-Campbell et al., 2010). Bayesian inference was conducted using MRBAYES v.4.1.2 (Huelsenbeck *et al.* 2001; Ronquist 2004) running a pair of independent searches for 1 million generations, with trees saved every 1000 generations and the first 250 sampled trees of each search discarded as burn-in. Due to high divergence between *P. diacanthus* and *H. africanus* (14.7% at cyt *b*) we were unable to resolve phylogenetic relationships within the genus *Pygoplites* using an outgroup, therefore an unrooted tree was also constructed with MRBAYES based on the concatenated dataset of all loci. A maximum likelihood tree was created using PHYML v.3.0.1 (Guindon *et al.* 2010) as implemented in GENEIOUS with clade support assessed with 1000 non-parametric bootstrap replicates. A neighbor-joining tree was created using GENEIOUS with clade support assessed after 1000 non-parametric bootstrap replicates.

RESULTS

Phylogenetic and coalescence analyses

All tree-building methods yielded identical topologies. The unrooted phylogenetic analysis recovered four lineages: a Pacific lineage that extends to Christmas Island in the eastern Indian Ocean (henceforth referred to as “Pacific lineage”), a lineage detected around Saudi

Arabia and Djibouti (henceforth referred to as "Red Sea lineage"), and two lineages with overlapping ranges in the Maldives and Diego Garcia (henceforth referred to as "Indian lineage 1" and "Indian lineage 2") (Fig. 2.2). The phylogenetic analyses were unable to resolve branch order among these lineages using an outgroup (Fig 2.2a), in part because the sister genus (*Holacanthus*) is deeply divergent at cyt *b* (Alva-Campbell *et al.* 2010). The Pacific lineage is 0.6% divergent from the Red Sea lineage and 1.2% and 1.5% from Indian lineage 1 and 2, respectively. The Red Sea lineage is 0.6% divergent from Indian lineage 1 and 1.0% from Indian lineage 2, and the two Indian lineages are distinguished by 1.5% divergence. Coalescence analysis based on cyt *b* yielded a TMRCA of 1.7 Ma for the Pacific lineage, and identified the Pacific as the oldest extant lineage (Table 2.1). The Red Sea lineage dates to 1.4 Ma and the two Indian lineages were the youngest: Indian lineage 1 at 0.7 Ma; Indian lineage 2 at 0.9 Ma.

MtDNA Sequences

Mitochondrial DNA molecular diversity indices are summarized for lineages in Table 2.1 and among populations in Table 2.2. Total haplotype diversity was $h = 0.817$ with 49 unique haplotypes. Among lineages, the Red Sea had the highest haplotype diversity ($h = 0.701$) with the lowest being observed in Indian Ocean lineage 2 ($h = 0.284$). Within populations, Indonesia had the highest haplotype diversity ($h = 1.00$) followed by Okinawa ($h = 0.867$) and the Maldives ($h = 0.808$). The lowest haplotype diversity was observed at Fiji ($h = 0.427$) and Mo'orea ($h = 0.483$). Total nucleotide diversity was $\pi = 0.005$. Among lineages, the Pacific Ocean and Red Sea had the higher nucleotide diversity ($\pi = 0.002$) with the lowest nucleotide diversity observed in both Indian Ocean lineages ($\pi = 0.001$). Among populations, the Maldives

and Diego Garcia had the highest nucleotide diversity for all locations, each at $\pi = 0.009$, whereas the lowest nucleotide diversity is observed at Fiji ($\pi = 0.0008$).

The median-joining haplotype network illustrates the low level of divergence between the four evolutionary lineages recovered from the phylogenetic analysis (Fig. 2.3a). However, the network also reveals that Red Sea haplotypes lie between the Pacific and Indian haplotypes. The presence of two Indian lineages radiating from the most common Red Sea haplotype provides evidence for two independent colonization events. The two Indonesia specimens are associated with Indian Ocean lineage 1; however, low samples size precludes any interpretation about lineage distribution. Christmas Island, located at the edge of the IPP, a region where Pacific and Indian Ocean fauna come into contact (Gaither & Rocha 2013), had both Pacific and Indian lineages. In subsequent comparisons between ocean basins, Christmas Island specimens grouped with Indian and Pacific cohorts based on mtDNA identity.

Population pairwise Φ_{ST} values for cyt *b* results are summarized in Table 2.3. Significance was determined after controlling for false discovery rates (corrected $\alpha = 0.009$). Φ_{ST} values show congruence with the haplotype network further supporting the Pacific, Indian, and Red Sea groups. There was little or no population structure detected within these groups, with two exceptions: Mo'orea (French Polynesia) shows significant genetic differentiation from all Pacific locations, with pairwise Φ_{ST} values ranging from 0.123 with Pohnpei to 0.229 with Fiji. Elsewhere in the Pacific, significant genetic structure was detected between the Marshall Islands and Fiji ($\Phi_{ST} = 0.061$, $P < 0.001$). Although population level data is not reported for the single location in the Sino-Japanese Province (Okinawa) due to low sample size ($N = 6$), preliminary runs show no significant population structure between the Sino-Japanese Province and the Pacific samples of *P. diacanthus*. The Red Sea lineage shows high levels of population

differentiation from all other samples (pairwise Φ_{ST}: 0.284 – 0.837). Likewise, the Indian lineages show significant population differentiation from all other samples (pairwise Φ_{ST}: 0.284 – 0.753). The AMOVA analysis supports the Pacific, Indian, and Red Sea geographic groupings based on mtDNA (Table 2.4) with the majority of the variation ($\Phi_{CT} = 0.66$, $P < 0.001$) existing among the groups.

The demographic results for cyt *b* show indications of population expansion at every Pacific location with the exception of Okinawa, the Marshall Islands, and Mo‘orea (Table 2.2). Estimates of population expansion indicate that the youngest dates are in the Pacific: Fiji and Christmas Island, with estimates of 39,000 and 49,000 years, respectively. The oldest Pacific expansion dates are in Okinawa, Pohnpei, and American Samoa, at 271,000, 230,000, and 212,000 years, respectively. Within the Red Sea Province, Saudi Arabia shows evidence for a population expansion (Fu’s F_S: - 4.73, $P < 0.01$) at 65,000 – 130,000 years, whereas Djibouti shows evidence for a neutral population (Fu’s F_S, $P = 0.35$) aged at 105,000 – 209,000 years. Locations in the Indian Ocean singularly show no evidence of population expansion (Fu’s F_S, $P > 0.02$) and have the oldest population expansions dates at 807,000 – 1,742,000 years. However, these estimates are shaped by the presence of two lineages that are not monophyletic (Fig. 2.3a). When considered individually, Indian lineage 1 has a population expansion date at 48,000 – 97,000 years (Fu’s F_S: -3.70, $P < 0.001$), and Indian lineage 2 has a population expansion date at 264,000 – 528,000 years (Fu’s F_S: -2.75, $P = 0.01$).

Nuclear DNA Sequences

A total of 10 variable sites yielded 12 alleles at the RAG2 locus and 31 variable sites yielded 46 alleles at the S7 locus. Samples from Palau and Tokelau were out of Hardy-Weinberg

equilibrium (Palau, $P < 0.001$; Tokelau, $P = 0.04$) with excess homozygotes at the S7 locus (Table 2.5). Overall expected heterozygosity (H_E) was 0.43 and 0.86 for RAG2 and S7, respectively. Across all samples $H_E = 0.06 – 0.64$ for RAG2 and $H_E = 0.41 – 1.00$ for the S7 intron. The median-joining networks based on intron sequences do not show distinct lineages in the Red Sea, Indian Ocean, and Pacific Ocean (Fig. 2.3b, c). However, both RAG2 and S7 networks include common alleles that are observed only in the Pacific, or only in the Indian Ocean locations. For S7, an Indian Ocean specific allele is also shared with a single individual from Christmas Island.

The population genetic results for the nuclear dataset are strongly concordant with mtDNA analyses for *P. diacanthus*, although they differ by degree. Genetic structure was absent within the Red Sea and within the Indian Ocean. The only significant differentiation in the Pacific was in 7 of 8 comparisons to Mo'orea (Society Islands, French Polynesia) with RAG2 ($\Phi_{ST} = 0.111 – 0.271$; Table 2.6). Curiously, none of the same pairwise comparisons for Mo'orea were significant with S7, however Mo'orea showed the highest differentiation from Red Sea populations.

Both nuclear markers show high levels of genetic structure that correspond to a Pacific, Indian, and Red Sea lineage. RAG2 was significant in 17 of 18 Pacific versus Indian comparisons ($\Phi_{ST} = 0.137 – 0.343$), significant in all Indian versus Red Sea comparisons ($\Phi_{ST} = 0.091 – 0.258$), and significant in 15 of 18 Pacific versus Red Sea comparisons ($\Phi_{ST} = 0.066 – 0.359$). The S7 differences were significant in all Pacific versus Indian comparisons ($\Phi_{ST} = 0.073 – 0.188$), all Indian versus Red Sea comparisons ($\Phi_{ST} = 0.253 – 0.512$), and all Pacific versus Red Sea comparisons ($\Phi_{ST} = 0.159 – 0.443$). The exceptions to these patterns were comparisons between the Red Sea lineage and Tokelau, as well as between Saudi Arabia and Pohnpei. For S7

the highest genetic structure was observed between the Indian and Red Sea populations. This contrasts with the RAG2 and cyt *b* comparisons, where the highest genetic structure differentiated the Pacific from both Indian and Red Sea regions.

DISCUSSION

Summary of results

Our data demonstrates that cryptic diversity exists within the monotypic genus *Pygoplites* as evidenced by significant levels of genetic structure among three regions: the Pacific Ocean (which includes a cohort at Christmas Island), the Indian Ocean (with two sympatric mtDNA lineages), and the Red Sea (Table 2.4). This pattern of genetic structure corresponds to known biogeographic provinces and phylogeographic barriers observed in other reef fishes (Rocha *et al.* 2007; Briggs & Bowen 2013; DiBattista *et al.* 2013; Eble *et al.* 2015; Gaither *et al.* 2015). The Red Sea biogeographic province is distinguished by a faunal break at the Gulf of Aden, and the Indo-Pacific Barrier is an intermittent terrestrial bridge between Australia and SE Asia that impedes water movement between Pacific and Indian Oceans during glacial low-sea levels (see Gaither & Rocha, 2013). The Sino-Japanese Province shows no genetic differentiation from the Pacific population (based on $N = 6$), but Mo'orea is highly isolated, a finding we attribute to prevailing oceanographic conditions (see Gaither et al., 2010). Below we discuss the phylogenetic implications of cryptic lineages and examine each of these regions in light of biogeographic theory

Phylogenetic considerations

Differences in coloration reviewed by Randall (2005) suggested that cryptic linages of *P. diacanthus* might exist in the Pacific and Indian Oceans. The three loci evaluated here support this Indian-Pacific distinction with diagnostic (albeit shallow) mtDNA differences and strong population genetic separations at two nuclear loci. A rooted phylogeny was unable to resolve relationships within the genus *Pygoplites* due to shallow separations and the deep divergence from the outgroup, *H. africanus,* (*d* = 15.5% at cyt *b,* this study), despite being the most closely related species to *P. diacanthus* (Alva-Campbell *et al.* 2010). Therefore, we were unable to determine the basal lineage from among the four lineages recovered. The oldest TMRCA in *P. diacanthus* is the Pacific lineage at 1.7 Ma, but the divergence between *Pygoplites* and *Holocanthus* is much older, estimated at 7.6 – 10.2 Ma (Alva-Campbell et al., 2010). Hence much of the evolutionary history of *Pygoplites* has been erased, at least for the loci examined here.

There are two possible explanations for the lack of diversity within the genus. First, there has been no evolutionary or selective pressure for *P. diacanthus* to diversify, a feature that may be attributed to the species ability to occupy a variety of ecological niches. *P. diacanthus* can be considered a generalist in that its range occupies more than half the globe in subtropical and tropical environments, its diet consists of sessile invertebrate, such as sponges and tunicates, and it appears to be a reef-habitat generalist where its range extends from the surface to depths greater than 60 m, a zone where shallow coral reef habitat is replaced by mesophotic ecosystems (Puglise *et al.* 2009). An alternative explanation is that other species within the genus went extinct while *P. diacanthus* persisted. However, with a poor fossil record, the evolutionary history of the marine angelfishes is poorly understood and limited to extant species. Therefore, we know of no species that may have existed during the 10 million year separation between

Holacanthus and *Pygoplites*. Nonetheless, the phylogeographic record for *Pygoplites* begins with a radiation in the last 2 MY. Although phylogenetic reconstruction was unable to determine branch order among the four lineages, the median-joining network indicates that the Red Sea lineage is basal to the two mtDNA lineages in the Indian Ocean. Coloration differences distinguish the Pacific lineage from both Indian and Red Sea lineages (Fig. 2.1); however, a preliminary morphological examination revealed no additional morphological characters that discriminate between Indian and Red Sea lineages (pers. comm. Luiz Rocha).

The geographical delineation between the Pacific and Indian lineages correspond with the exposure of the Sunda Shelf, which separates the Pacific and Indian Oceans during low sea level. The Red Sea lineage corresponds to the Red Sea biogeographic province, which encompasses the adjacent Gulf of Aden (Briggs & Bowen 2012) and whose populations have a disjunct distribution with the remainder of the range (see below). During glacial maxima the Red Sea is effectively cut off from the Indian Ocean by closure of the Strait of Bab al Mandab, the only natural gateway into the Red Sea, allowing sufficient time for populations to diverge into distinct evolutionary lineages (DiBattista et al., 2013, 2016a).

The mechanisms facilitating two sympatric mtDNA lineages in the Indian Ocean are less clear. Coalescence estimates indicate that lineages arose independently during roughly the same period (0.72 – 0.93 Ma). As there are no known phenotypic differences within this region, the unexpected recovery of two distinct lineages requires further investigation. Indian Ocean samples contained similar number of each lineage (Maldives: Lineage 1, $N = 8$; Lineage 2, $N = 8$; Diego Garcia: Lineage 1, $N = 17$; Lineage 2, $N = 11$) indicating that the two lineages are approximately equally represented.

The recovery of multiple evolutionary partitions within the monotypic genus *Pygoplites* may not be indicative of other monotypic genera. Cryptic evolutionary partitions are routinely discovered within species of marine fishes (Colborn *et al.* 2001; Rocha *et al.* 2008; DiBattista *et al.* 2012; Fernandez Silva *et al.* 2015), and in this regard *P. diacanthus* is similar to the more speciose inhabitants of Indo-Pacific reefs. The factors that produce a deep, monotypic lineage are therefore not reflected in an unusual phylogeographic architecture. However, part of the explanation for this monotype may be that five to eight million years after the divergence of *Pygoplites* and *Holocanthus*, the ancestor of all modern *Pygoplites* likely radiated out of the West Pacific Ocean, an extensive source of Indo-Pacific diversity (Cowman & Bellwood 2013).

In considering the phylogenetic results through a taxonomic lens, there are several issues. First, the Pacific and Indian morphs are distinguished by diagnostic differences, but they are not monophyletic. The Indian Ocean contains two mtDNA lineages, each more closely related to the Red Sea lineage than to each other. Second, the coloration difference between Pacific and Indian forms, now matched by $d = 0.006$ divergence, could be a platform to describe them as separate species. Third, the genetic divergence observed at all three loci is low in comparison to typical divergences for fish species ($d = 0.03 – 0.12$; Grant and Bowen, 1998; Johns and Avise, 1998). Fourth, the two morphs form mixed groups where they co-occur at Christmas Island (Hobbs and Allen, 2014). Since we lack diagnostic nDNA alleles for the two morphs, we do not have the power to test for hybrids between the lineages, but this is certainly a possibility. Given these considerations, we believe that it is problematic to invoke species status for these three regional forms and we endorse subspecies recognition distinguishing the Pacific lineage from the Indian and Red Sea lineages based on shallow but diagnostic distinctions in genetics and morphology. We propose the name *P. diacanthus flavescens* for the Indian Ocean and Red Sea lineages to give

recognition to the yellow chest coloration, a character not found in individuals from the Pacific lineage (*P. d. diacanthus*).

Red Sea isolation and refugia

The Red Sea Province is distinguished from the Indian Ocean by high levels of endemism found across a suite of taxa (Randall 1994; Cox & Moore 2000) as well as many fish species whose ranges extend from the Red Sea into the Gulf of Aden (Briggs & Bowen 2012; DiBattista *et al.* 2016). This distinction is supported by our findings that show the Djibouti population of *P. diacanthus* forms a genetically homogenous population with the Red Sea, coupled with a population break separating these two locations from adjacent populations in the Indian Ocean.

Population breaks between the Red Sea and Indian Ocean have previously been documented in *P. diacanthus*, in addition to other species (Vogler *et al.* 2008; DiBattista *et al.* 2013; Fernandez Silva *et al.* 2015). One possible explanation for breaks across multiple species in this region is the presence of an ecological barrier. Based on differences in fish assemblages, Kemp (1998) proposed that such a barrier separated the Red Sea and western Gulf of Aden from the eastern Gulf of Aden and Indian Ocean. Furthermore, the upwelling that occurs along the Arabian coast of southern Yemen, Oman and the Indian Ocean coast of Somalia impedes the formation of continuous reefs from Djibouti to Oman and southern Somalia, limiting opportunities for dispersal from the Gulf of Aden (for review see DiBattista et al., 2016a). Notably, we did not detect *P. diacanthus* during collection efforts in the Socotra Archipelago, Oman, and Somolia, which are located at the periphery of the Gulf of Aden and the Arabian Sea. This observation coincides with previous surveys conducted in the region indicating a gap in the distribution of *P. diacanthus* between the Gulf of Aden and the western Indian Ocean, a

phenomenon found in other wide-ranging species (Kemp 1998).

The parsimonious conclusion that a population of *P. diacanthus* has been in the Red Sea Province (including western Gulf of Aden) for over a million years implies that this population has been subjected to and survived Pleistocene glacial conditions. The only natural connection to the Indian Ocean is through the narrow (18 km) and shallow (137 m) Strait of Bab al Mandab at the southern end of the Red Sea. During periods of low sea level associated with glaciation, the connection from the Indian Ocean through the strait is reduced, and the Red Sea experiences extreme fluctuations in temperature and salinity (Bailey 2009). During the last 400,000 years in particular, the Red Sea has undergone at least two periods of hypersalinity (c. 19,000 and 30,000 years ago) that caused an aplanktonic environment in which larvae of many marine organisms presumably could not survive (Siddal et al., 2003; DiBattista et al., 2016a). Coalescence analysis dates the Red Sea lineage to 1.44 Ma (95% HPD = 0.51 – 2.53 Ma), which coupled with the Saudi Arabian population expansion (65,000 – 130,000 years) indicates that *P. diacanthus* likely survived the temperature and salinity crises that occurred during these periods, a conclusion that is corroborated by other species (DiBattista *et al.* 2013). Our neutrality tests show no evidence for changes in population size (Fu's Fs = -3.60, P = 0.035) providing evidence that refugia may have existed in the Red Sea Province (possibly in the Gulf of Aden) to support a large stable population of *P. diacanthus* despite the extreme environmental conditions.

Biogeographic inferences in the Indian Ocean

Christmas Island is located in the eastern Indian Ocean, a region (which includes Cocos-Keeling Island) of secondary contact between Indian and Pacific species that diverged in allopatry during Pleistocene glacial cycles (Gaither & Rocha 2013). Indian and Pacific Ocean

phenotypes of *P. diacanthus* have both been recorded in the eastern Indian Ocean region, and both Pacific and Indian Ocean mtDNA haplotypes are present at Christmas Island, indicating an area of overlap (Hobbs and Allen, 2014 , Fig 2.3a). This region is recognized as a hybridization hotspot (suture zone) with interbreeding documented between at least 27 reef fish species-pairs from across eight families, and it has been suggested that Indian and Pacific *P. diacanthus* lineages hybridize in this region (Hobbs & Allen 2014). However, additional molecular work will be needed to evaluate this hypothesis.

Genetic differences between Indian and Pacific Ocean populations are consistent with Pleistocene closures of the Indo-Pacific Barrier. Despite being located in the Indian Ocean basin and the presence of haplotypes that are associated with Indian Ocean lineages, our results indicate that Christmas Island is genetically differentiated from other locations in the Indian Ocean and instead has a stronger affiliation with the Pacific Ocean. A barrier to dispersal has been previously proposed to exist west of the Cocos-Keeling Islands and east of the Chagos-Laccadive ridge based on the presence of many Pacific species with distributions that extend no further west than Christmas and the Cocos-Keeling Islands (Blum 1989; Hodge & Bellwood 2016).

Elsewhere in the Indian Ocean, the Maldives and Diego Garcia (Chagos Archipelago) are genetically differentiated from the Pacific and Red Sea, but not from each other. Both archipelagos are located in the central Indian Ocean, which is the western extent of the IPP, although they also share faunal affinities with the Western Indian Ocean Province (Winterbottom & Anderson 1997; Gaither *et al.* 2010b; Eble *et al.* 2011a; Briggs & Bowen 2012). The grouping of Diego Garcia and the Maldives within the IPP is further evidenced by Pacific Ocean mtDNA being found at Diego Garcia (Fig. 2.3a), which provides a signal that some degree of gene flow

occurs between the Indian and Pacific Ocean. Coalescence estimates of the two Indian Ocean *P. diacanthus* lineages indicate they arose from an ancestor affiliated with the Red Sea.

The ability of *P. diacanthus* to persist throughout major geological and climatic shifts is demonstrated by the age of expansion for all populations of *P. diacanthus* which predate the Last Glacial Maximum, peaking at 26.5 – 19 ka (Clark *et al.* 2009) when global sea level dropped 130 m below present levels (Voris 2000). During this period, habitable shelf in the Pacific was reduced by as much as 92% from present day values and this reduction in habitat area has been linked to population bottlenecks (Ludt and Rocha, 2014), a feature not observed in *P. diacanthus*. As previously discussed, *P. diacanthus* can be considered an ecological generalist with a vertical range that extends to mesophotic depths. Thus, a reduction of shallow reef habitat due to sea level change may not have substantially reduced suitable ecological niches for this species.

Gene flow within the Pacific

Despite the wide expanse of the central and western Pacific Ocean, many species exhibit a high degree of genetic connectivity across the region (Schultz et al., 2006; Reece et al., 2010b; Gaither et al., 2011). However, populations breaks have been associated with isolated regions such as the Hawaiian Archipelago and the Marquesas, which are also known for high levels of endemism (Randall 2005b; Briggs & Bowen 2012). Here we found population genetic differentiation of Mo'orea (Table 2.3, Table 2.6), a pattern observed in other widely distributed Pacific species (Planes 1993; Bernardi *et al.* 2001; DiBattista *et al.* 2012; Timmers *et al.* 2012; Lemer & Planes 2014).

The isolation of Mo'orea may be attributed to ocean circulation patterns. The westward flow of the Southern Equatorial Current (SEC) and eddies created in the wake of Tahiti, located approximately 17 km east of Mo'orea, contribute to a strong counterclockwise flow around the island promoting the local retention of larvae (Leichter *et al.* 2013). _ENREF_42Plankton tows conducted in this region revealed that fish larvae were not recovered more than 300 km from the nearest reef (Lo-Yat *et al.* 2006). Additionally, Bernardi *et al.* (2012) found that 14% of juvenile damselfish (*Dascyllus trimaculatus*) recruiting to reefs around Mo'orea were very close relatives, including full siblings, indicating that the larvae traveled and settled together despite a PLD of several weeks.

Although the counterclockwise flow surrounding Mo'orea may explain local retention of larvae, it does not explain how larvae produced elsewhere in the Pacific are restricted from emigrating and settling onto Moorean reefs. One possible explanation may be that the westward flowing SEC restricts larvae from dispersing in an easterly direction. The SEC, located between 4°N and 17°S (Wyrtki & Kilonsky 1984; Bonjean & Lagerloef 2002), has been implicated in limiting connectivity between the Marquesas, located 1300 km northeast of Mo'orea, and other Pacific locales (Gaither et al., 2010; Szabo et al., 2014). Populations of *P. diacanthus* west of Mo'orea, located at the southern extent of the SEC, may be restricted in easterly dispersal by the strong current; however, the SEC may facilitate a western dispersal. American Samoa is the closest sample location downstream from Mo'orea; it is the only sample location that is not significantly differentiated from Mo'orea at RAG2 (Φ_{ST} = 0.039, P=0.054) and has one of the lowest levels of differentiation from Mo'orea at cyt *b* (Φ_{ST} = 0.128). Fine-scale sampling across French Polynesia would be required to determine the extent of genetic isolation. Additionally, further sampling from neighboring localities east and west of Mo'orea are needed to test our

hypothesis regarding the SEC. It is likely that a number of physical processes surrounding Mo'orea promote local retention of larvae and prevent the recruitment of larvae from elsewhere in the Pacific.

CONCLUSION

Pygoplites diacanthus is the first large angelfish to be surveyed across the Indo-Pacific. It appears to be highly dispersive, joining the ranks of smaller Pomacanthids such as the pygmy angelfish in showing little structure across ocean basins (Schultz *et al.* 2006; DiBattista *et al.* 2012). Pelagic larval duration tends to be shorter in the large angelfishes (~25 days in *Pygoplites* compared to 30 days or more in pygmy angelfishes; Thresher and Brothers, 1985), but this does not seem to restrict dispersal among the closely associated islands of the West and Central Pacific. However, this monotypic genus exhibits deep population genetic partitions between ocean basins. In every case, historical barriers existed at the junctions between observed populations, and in at least two cases (Red Sea and Mo'orea) oceanographic conditions may contribute to contemporary isolation. On the genetic continuum between isolated populations and evolutionary distinctions (Wright 1978; Frankham *et al.* 2002), the deep divergences between oceans indicate that the monotypic *Pygoplites* may be on the pathway to three emerging species. The genetic and morphological divergences are certainly sufficient to recognize subspecific evolutionary (and taxonomic) partitions.

Figure 2.1. Map of collection locations, sample sizes (in parentheses), and the two recognized morphotypes of *Pygoplites diacanthus*. (*left*) Indian Ocean and Red Sea individuals are characterized by a yellow chest and head, whereas the (*right*) Pacific Ocean morph is characterized by a gray chest and head. *Photos by L. Rocha (Djibouti; Great Barrier Reef, Australia)*

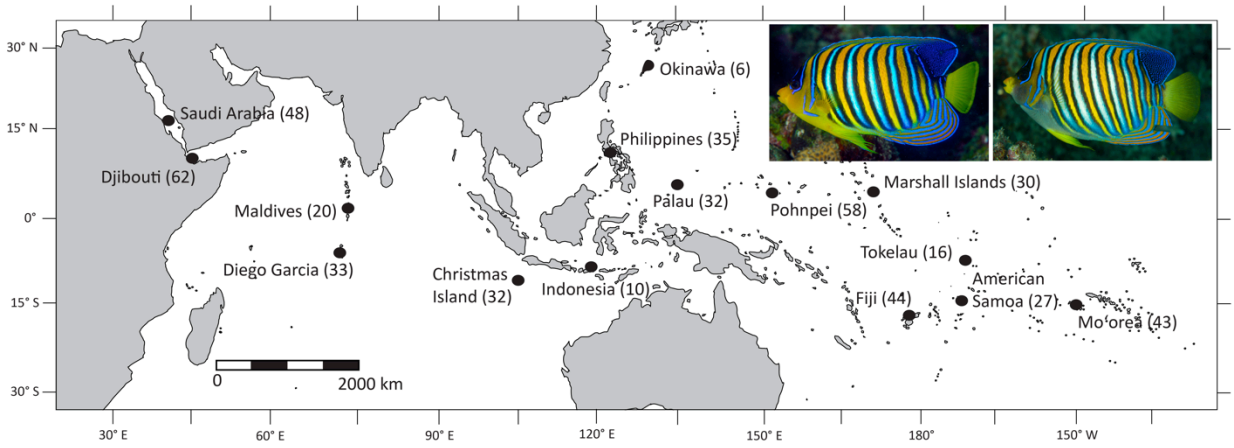

Figure 2.2. Molecular phylogenetic reconstruction of *Pygoplites diacanthus.* A) Rooted Bayesian tree based on mitochondrial cytochrome *b* with posterior probabilities, B) an unrooted maximum-likelihood tree based on mitochondrial and nuclear markers (cytochrome *b*, intron 1 of the S7 ribosomal protein, and the recombination-activating gene 2) with consensus values based on posterior probabilities from Bayesian inference (BI), maximum-likelihood bootstrap support (ML), and neighbor-joining bootstrap support (NJ). Percent sequence divergence is represented on the scale bar. The sizes of black triangles are proportional to the number of individuals within the lineage. Abbreviations: Red Sea Province, RS; Indian Ocean, IO.

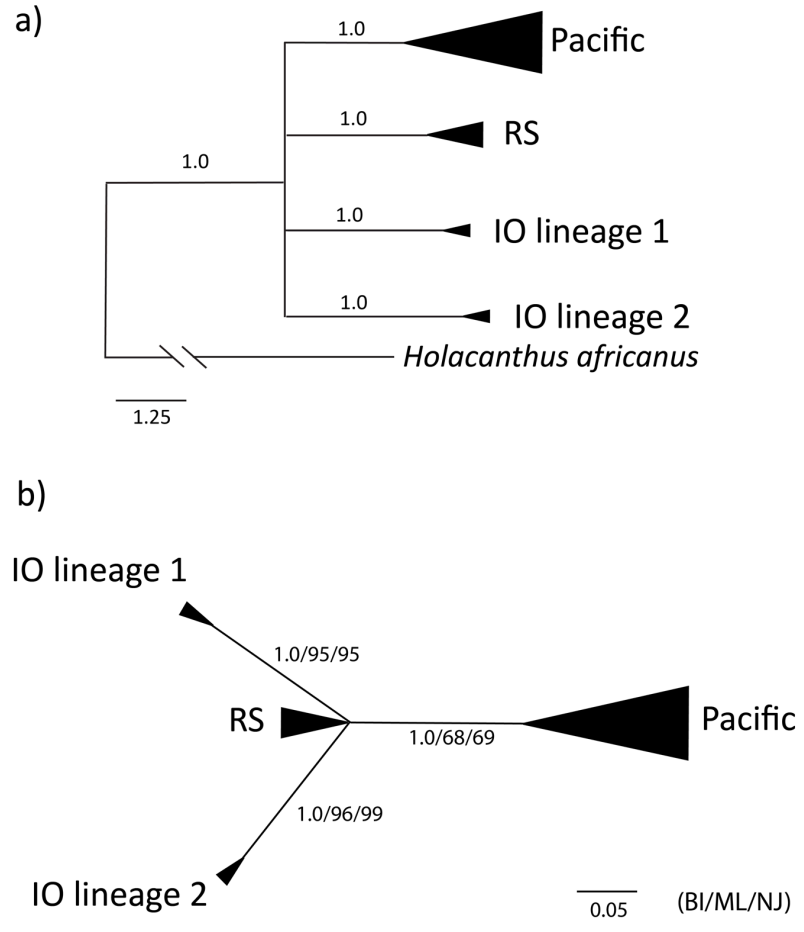

Figure 2.3. Median-joining network for *Pygoplites diacanthus* constructed using NETWORK for A) cytochrome *b* sequences (568 bp) from 386 individuals, B) alleles for RAG2 (431 bp) from 366 individuals, and c) alleles for the S7 intron (510 bp) from 288 individuals. Each circle represents a unique mitochondrial haplotype or nuclear allele, with the size being proportional to the total frequency. Open circles represent unsampled alleles, branches and crossbars represent a single nucleotide change, and color represents collection location (see key). All singleton alleles ($N = 22$) were removed from the S7 analysis to minimize circularity between closely related alleles.

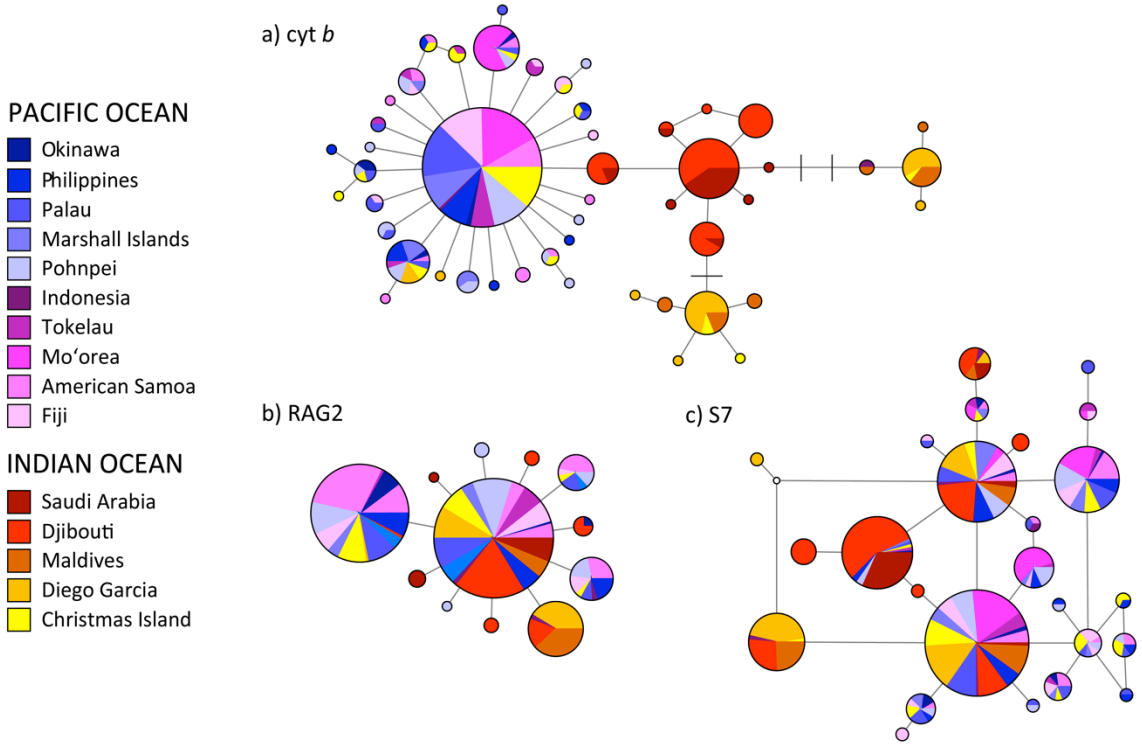

Table 2.1. Molecular diversity indices for lineages of *Pygoplites diacanthus* based on mitochondrial DNA (cytochrome *b,* 568 bp). Number of individuals sequenced (n), number of haplotypes (N_h), number of segregating (polymorphic) sites (S), haplotype diversity (h), and nucleotide diversity (π) are presented. Times to most recent common ancestor (TMRCA) are presented as million years. Bolded numbers denote significance at $P < 0.02$.

Lineage	n	N_h	S	$h \pm$ SD	$\pi \pm$ SD	TMRCA (95% HPD)	Fu's F_S	Fu's F_S P-value
Pacific Ocean[a]	257	33	37	0.628 ± 0.034	0.002 ± 0.010	1.71 (0.91 - 2.65)	-29.51	<0.001
Red Sea[b]	81	9	8	0.701 ± 0.042	0.002 ± 0.003	1.44 (0.51 - 2.53)	-3.602	0.035
Indian Ocean Lineage 1	28	6	5	0.439 ± 0.114	0.001± 0.001	0.72 (0.14 - 1.52)	-3.695	<0.001
Indian Ocean Lineage 2	20	4	3	0.284 ± 0.128	0.001 ± 0.001	0.92 (0.27 - 1.75)	-2.749	0.001
All Locations	386	49	45	0.817 ± 0.018	0.005 ± 0.003	--	-25.90	<0.001

[a]Pacific includes all Pacific Ocean populations plus Christmas Island
[b]Red Sea includes Saudi Arabia and Djibouti

Table 2.2. Molecular diversity indices for populations of *Pygoplites diacanthus* based on mitochondrial DNA (cytochrome *b,* 568 bp) divided into phylogeographical groupings. Number of individuals sequenced (n), number of haplotypes (N_h), number of segregating (polymorphic) sites (S), haplotype diversity (h), and nucleotide diversity (π) are presented. τ is used to estimate the age of most recent population expansion (population age) using the equation $\tau = 2\mu t$ (see Material and Methods). ∞ denotes values that could not be resolved. Bolded numbers denote significance at $P < 0.02$.

Sample Location	n	N_h	S	$h \pm$ SD	$\pi \pm$ SD	τ	Population Age (years)	Fu's F_S	Fu's F_S P-value
Pacific Ocean									
Okinawa	6	4	3	0.867 ± 0.129	0.002 ± 0.002	1.54	135,000 - 271,000	-1.454	0.052
Philippines	21	7	8	0.657 ± 0.104	0.001 ± 0.001	1.04	92,000 - 183,000	-3.473	0.003
Palau	32	9	8	0.488 ± 0.109	0.001 ± 0.001	0.67	59,000 - 118,000	-7.928	< 0.001
Marshall islands	23	4	3	0.549 ± 0.105	0.001 ± 0.001	0.78	69,000 -138,000	-0.936	0.208
Pohnpei	33	12	13	0.760 ± 0.076	0.002 ± 0.001	1.30	115,000 - 230,000	-8.754	< 0.001
Indonesia	2	2	7	1.000 ± 0.500	0.012 ± 0.013	∞	∞	1.946	0.519
Tokelau	16	6	5	0.617 ± 0.135	0.001 ± 0.001	0.92	81,000 - 162,000	-3.692	< 0.001
Mo'orea	42	2	1	0.483 ± 0.039	0.001 ± 0.001	0.73	64,000 - 128,000	1.766	0.738
American Samoa	25	10	9	0.730 ± 0.094	0.002 ± 0.001	1.20	106,000 - 212,000	-7.128	< 0.001
Fiji	25	6	5	0.427 ± 0.122	0.001 ± 0.001	0.56	49,000 - 98,000	-4.423	< 0.001
Christmas Island	32	13	19	0.720 ± 0.087	0.004 ± 0.003	0.44	39,000 - 77,000	-5.445	0.006
Red Sea Province									
Saudi Arabia	23	7	7	0.522 ± 0.124	0.001 ± 0.001	0.74	65,000 - 130,000	-4.731	< 0.001
Djibouti	58	6	4	0.738 ± 0.034	0.002 ± 0.001	1.19	105,000 - 209,000	-0.879	0.349
Indian Ocean									
Maldives	16	6	11	0.808 ± 0.069	0.009 ± 0.005	9.89	871,000 - 1,742,000	1.858	0.804
Diego Garcia	32	7	17	0.692 ± 0.059	0.009 ± 0.005	9.17	807,000 - 1,614,000	4.177	0.898
Pacific Ocean	257	33	37	0.628 ± 0.034	0.002 ± 0.009	0.94	83,000 - 165,000	-29.511	<0.001
Red Sea Province	81	9	8	0.701 ± 0.042	0.002 ± 0.003	1.09	96,000 - 192,000	-3.602	0.035
Indian Ocean	48	11	19	0.738 +/- 0.045	0.009 ± 0.008	9.37	825,000 - 1,649,000	1.013	0.700
All Locations	386	49	45	0.817 ± 0.018	0.005 ± 0.003	3.61	318,000 - 636,000	-25.897	<0.001

Table 2.3. Matrix of pairwise Φ_{ST} statistics for 13 populations of *Pygoplites diacanthus* based on mitochondrial DNA (cytochrome *b,* 568 bp) sequences. Bolded numbers indicate significance after controlling for false discovery rates at $\alpha = 0.05$ (as per Narum, 2006). The corrected $\alpha = 0.009$. Owing to low sample size, Okinawa and Indonesia have been excluded. Abbreviations: Red Sea Province, RS; Indian Ocean, IO.

		Pacific Ocean									RS		IO
	Sample location	1	2	3	4	5	6	7	8	9	10	11	12
Pacific Ocean	1. Philippines	--											
	2. Palau	0.02286	--										
	3. Marshall Is.	-0.00690	0.04436	--									
	4. Pohnpei	-0.00113	-0.00003	-0.00602	--								
	5. Tokelau	-0.00088	0.00970	0.02593	-0.00273	--							
	6. Mo'orea	**0.21137**	**0.15115**	**0.24879**	**0.12293**	**0.21718**	--						
	7. American Samoa	0.00241	0.01337	0.01878	-0.00796	-0.00602	**0.12806**	--					
	8. Fiji	0.04077	0.00310	**0.06141**	0.00599	-0.00058	**0.22924**	0.01738	--				
	9. Christmas Is.	0.02264	0.04064	0.02876	0.02685	0.01493	**0.13152**	0.02805	0.03767	--			
RS	10. Saudi Arabia	**0.76918**	**0.80926**	**0.80834**	**0.73676**	**0.79578**	**0.83717**	**0.75536**	**0.82668**	**0.55505**	--		
	11. Djibouti	**0.73859**	**0.76222**	**0.75577**	**0.72292**	**0.74532**	**0.78731**	**0.73266**	**0.76756**	**0.58986**	0.05789	--	
IO	12. Maldives	**0.64673**	**0.71028**	**0.67118**	**0.67292**	**0.63123**	**0.75342**	**0.65734**	**0.69346**	**0.53679**	**0.42875**	**0.50323**	--
	13. Diego Garcia	**0.51036**	**0.56474**	**0.52244**	**0.53912**	**0.49451**	**0.61025**	**0.52284**	**0.54533**	**0.41311**	**0.28493**	**0.35104**	-0.00728

Table 2.4. Results of the analysis of molecular variance (AMOVA) based on mitochondrial DNA (cytochrome *b*) sequence data for *Pygoplites diacanthus*. Bolded values denote significance at $P < 0.05$.

	Among groups			Among populations (within groups)			Within populations		
Regions	Φ_{CT}	*P*-value	% variation	Φ_{SC}	*P*-value	% variation	Φ_{ST}	*P*-value	% variation
Pacific Ocean vs. Indian Ocean	0.60	0.058	59.91	**0.19**	< 0.001	**7.46**	**0.67**	< 0.001	**32.63**
Pacific[a] vs. Indian[b] vs. Red Sea[c]	**0.66**	< 0.001	**65.53**	**0.04**	0.017	**1.44**	**0.67**	< 0.001	**33.03**
Indian[b] vs. Red Sea[c] vs. Christmas Is.	0.44	0.078	44.09	**0.02**	< 0.001	**0.92**	0.45	0.269	54.99
Pacific[a] vs. Mo'orea	0.08	0.184	7.92	**0.05**	< 0.001	**4.82**	**0.13**	< 0.001	**87.26**

[a]Pacific includes all Pacific Ocean populations plus Christmas Island.
[b]Indian includes the Maldives and Diego Garcia.
[c]Red Sea includes Saudi Arabia and Djibouti.

Table 2.5. Molecular diversity indices for populations of *Pygoplites diacanthus* based on nuclear DNA (introns RAG2 and S7) for all populations. Number of individuals sequenced (n), number of alleles (N_a), number of segregating (polymorphic) sites (S), observed heterozygosity (H_O), expected heterozygosity (H_E), and the corresponding P-value

	RAG2						S7					
Sample Location	n	N_a	S	H_O	H_E	P-value	n	N_a	S	H_O	H_E	P-value
Pacific Ocean												
Okinawa	6	3	1	0.50	0.59	1.00	5	8	8	0.80	0.93	0.37
Philippines	21	3	2	0.38	0.46	0.32	15	11	9	0.73	0.85	0.19
Palau	30	4	3	0.47	0.41	0.67	22	14	12	0.88	0.82	< 0.001
Marshall islands	27	3	2	0.26	0.29	0.55	14	9	8	0.71	0.82	0.16
Pohnpei	39	5	4	0.38	0.43	0.20	21	15	16	0.76	0.87	0.24
Indonesia	4	3	2	0.50	0.46	1.00	3	6	7	1.00	1.00	1.00
Tokelau	16	2	1	0.06	0.06	1.00	8	8	8	0.63	0.81	0.04
Mo'orea	31	4	3	0.61	0.64	0.83	30	7	7	0.80	0.71	0.33
American Samoa	18	3	2	0.44	0.54	0.40	16	10	10	0.75	0.85	0.14
Fiji	21	4	3	0.43	0.43	0.83	16	12	10	0.88	0.85	0.74
Christmas Island	25	4	3	0.28	0.39	0.14	18	12	11	0.78	0.84	0.58
Red Sea Province												
Saudi Arabia	19	3	2	0.21	0.20	1.00	15	5	7	0.47	0.41	1.00
Djibouti	59	6	5	0.22	0.22	0.61	52	8	7	0.85	0.80	0.43
Indian Ocean												
Maldives	19	2	1	0.37	0.46	0.61	18	5	5	0.61	0.70	0.24
Diego Garcia	31	3	2	0.39	0.38	0.25	31	8	9	0.84	0.72	0.93
All Locations	366	12	10	0.35	0.43	<0.001	284	44	31	0.77	0.86	< 0.001

Table 2.6. Matrix of pairwise *F*-statistics for 13 populations of *Pygoplites diacanthus*. Φ_{ST} values for RAG2 (below diagonal) and S7 (above diagonal). Bolded numbers indicate significance after controlling for false discovery rates at α = 0.05 (as per Narum, 2006). The corrected α = 0.009. Owing to low sample size, Okinawa and Indonesia have been excluded. Abbreviations: Red Sea Province, RS; Indian Ocean, IO.

		Pacific Ocean									RS		IO	
	Sample location	1	2	3	4	5	6	7	8	9	10	11	12	13
Pacific Ocean	1. Philippines	--	-0.00920	-0.01362	-0.01360	-0.01687	0.01687	0.01128	-0.00514	0.01515	**0.34364**	**0.17142**	**0.10393**	**0.09571**
	2. Palau	-0.00538	--	0.00508	-0.00296	-0.02460	0.02700	0.00922	-0.01550	-0.00726	**0.38275**	**0.22126**	**0.09033**	**0.08852**
	3. Marshall Is.	0.00818	-0.01120	--	-0.00031	-0.00704	0.05734	0.00961	-0.00460	0.04995	**0.34476**	**0.15924**	**0.14964**	**0.13677**
	4. Pohnpei	-0.00668	-0.01082	-0.00789	--	-0.02035	0.01691	0.00989	-0.00473	0.01182	**0.30090**	**0.18403**	**0.07945**	**0.08287**
	5. Tokelau	0.07767	0.04393	0.04048	0.02592	--	0.00217	-0.01313	-0.02998	-0.01183	**0.37543**	**0.20290**	**0.09332**	**0.09858**
	6. Mo'orea	**0.11389**	**0.13061**	**0.16703**	**0.16493**	**0.27102**	--	0.07372	0.03949	0.03840	**0.44338**	**0.25787**	**0.07336**	**0.07570**
	7. American Samoa	-0.00096	0.02568	0.05152	0.03906	**0.17279**	0.03921	--	-0.01129	0.02085	**0.36682**	**0.24862**	**0.18011**	**0.18764**
	8. Fiji	-0.02068	-0.01577	-0.00644	-0.01574	0.05125	**0.12489**	0.01209	--	0.00554	**0.40033**	**0.23471**	**0.13818**	**0.13340**
	9. Christmas Is.	-0.00815	-0.01338	-0.01051	-0.00520	0.07526	**0.11113**	0.00810	-0.01339	--	**0.43065**	**0.26948**	**0.09838**	**0.10584**
RS	10. Saudi Arabia	**0.10905**	**0.08062**	**0.08407**	0.06289	0.02759	**0.29673**	**0.19761**	**0.08516**	**0.11256**	--	0.07234	**0.51195**	**0.50068**
	11. Djibouti	**0.12103**	**0.08643**	**0.07487**	**0.06583**	0.00215	**0.3587**	**0.22863**	**0.09234**	**0.11642**	0.02992	--	**0.25737**	**0.25281**
IO	12. Maldives	**0.23970**	**0.23544**	**0.26566**	**0.23347**	**0.28203**	**0.34276**	**0.27897**	**0.23017**	**0.25607**	**0.25779**	**0.22058**	--	-0.00921
	13. Diego Garcia	**0.16033**	**0.14521**	**0.15459**	**0.1368**	0.13833	**0.31590**	**0.22167**	**0.14479**	**0.16565**	**0.13958**	**0.09067**	0.01181	--

CHAPTER 3 CONTEMPORARY CONNECTIVITY ACROSS THE HAWAIIAN ARCHIPELAGO IN TWO SPECIES OF SURGEONFISHES, *ACANTHURUS TRIOSTEGUS* AND *CTENOCHEATUS STRIGOSUS*, REVEAL FINESCALE STRUCTURE

ABSTRACT

The Hawaiian Archipelago has served as a natural lab to assess genetic connectivity patterns across a variety of organisms that are taxonomically and ecologically diverse. The ability to assess connectivity across a broad spectrum of taxa has provided insight into the location of ecological breaks, and the role of life history in influencing genetic structure and diversity. One common factor among these studies is the use of targeted loci, which illuminates connectivity over evolutionary timescales but is limited in explaining contemporary patterns. To evaluate contemporary connectivity patterns, we conducted a genomics-based analysis using SNPs generated from individual libraries as well as pooled DNA. The two species of surgeonfishes used in this study, *Acanthurus triostegus* (*manini*) and *Ctenochaetus strigosus* (*kole*), exhibit similar life history strategies but differ greatly in their biogeographic distribution. *Kole*, a Hawaiian endemic, showed island-by-island population structure, a pattern not previouosly exhibited amoug Hawaii reef fishes and countering previous results based on a single mtDNA marker. *Manini*, which has an Indo-Pacific distribution, showed highly structured population in the main Hawaiian Islands, but genetic homogeneity across the the northwestern extent of the archipelago. These results highlight the efficacy of genomic sequencing to characterize contemporary connectivity and invokes a mandate to revist past connetivity studies using targeted loci and reassessing them in a genomics framework.

INTRODUCTION

The linear Hawaiian Archipelago has hosted extensive research into genetic connectivity patterns across a variety of organisms that are taxonomically and ecologically diverse (e.g. Eble et al., 2009; (Andrews *et al.* 2010; Gaither *et al.* 2010a; Skillings *et al.* 2011; Timmers *et al.* 2011; Coleman *et al.* 2014; Iacchei *et al.* 2014; Tenggardjaja *et al.* 2016). This research has cumulated in several meta-analysis studies that aimed to identify common barriers to dispersal across the archipelago (Toonen *et al.* 2011), assess how life history traits influence population genetic structure (Selkoe *et al.* 2014), and to show how high coral cover harbors the greatest genetic diveristy (Selkoe *et al.* 2016b). Common amongst the data that was the foundation for these meta-analysis studies was the use of a targeted locus approach, which in some cases was conducted using a single mtDNA marker.

Population genetic studies have historically used a targeted loci approach to characterize genetic connectivity. However, during the past decade the field of genetics has steadily shifted to using high-throughput sequencing due in part to a reduction in cost (Wetterstrand 2019) and the ability to generate thousands of loci in a single run. Furthermore, when studies have incorporated a genomic component they are able to describe trends and patterns that could not be accomplished using a targeted loci approach (Keller *et al.* 2013; Gaither *et al.* 2015). With an increased number of loci from across the genome, we now have a greater ability to relate genomic trends to the ecology of the organism being investigated.

For this study, we used restriction site associated DNA sequencing (RADseq), a reduce genomic approach, to characterize connectivity across the Hawaiian Archipelago for two species of surgeonfish. By using RADseq, we were able to resolve contemporary pattern of connectivity as opposed to a targeted loci approach (i.e. mtDNA) which, in essence, uses loci that resolve

connectivity averaged across thousands to millions of years. Additionally, we can revist and compare our RADseq results against past studies.

The two species of surgeonfishes (family Acanthuridae) in this study exhibit similar life history strategies and occupy similiar habitat. However, they differ greatly in their biogeographic distribution, thus providing an excellent framework to compare patterns of connectivity across Hawaii. *Ctenocheatus strigosus*, locally known as *kole*, and *Acanthurus triostegus*, locally known as *manini*, are both distributed throughout the Hawaiian Archipelago as well as Johnston Atoll. *Kole* is endemic to Hawai‘i and Johnston Atoll. *Manini* has a broad Indo-Pacific distribution, although the Hawai‘i population is recognized as a sub-species (*A. triostegus sandvicensis*) based on diagnostic differences in coloration and morphology (Randall 1961; Randall 2007).

Genetic patterns of connectivity have previously been described for each species. Genetic connectivity across Hawai‘i was characterized for *kole* and found no population structure across the majority of the island chain except between Maro Reef and Pearl and Hermes Atoll located in the northwestern part of the archipelago (Eble *et al.* 2009). The single study that investigated connectivity of *manini* within Hawai‘i found isolation between Hawai‘i Island and O‘ahu (Planes & Fauvelot 2002). However, studies in other parts of the range have provided conflicting patterns of dispersal. In Northern Australia, *manini* had no significant (mtDNA) population structure across the Torres Strait (Mirams *et al.* 2011), a known biogeographic barrier (Voris 2000). However, by using allozymes, genetic structure was observed across various spatial scales from adjacent islands to a range-wide scale (Planes 1993; Planes *et al.* 1998a; Planes & Fauvelot 2002).

This will be the first analysis using a genomic approach to assess connectivity across the Hawaiian Archipelago for a marine fish species. We anticipate that RADseq data will uncover patterns that are not found using a targeted loci approach. The existence of previous single-locus surveys provides a valuable foundation for measuring the increased resolution expected with RADseq data.

MATERIAL AND METHODS

Taxon sampling and DNA extraction

Between 2003 and 2006, a total of 461 tissue samples (primarily fin clips) of *manini* from 10 locations and 790 samples of *kole* from 16 locations were collected from across the Hawaiian Archipelago and Johnston Atoll using pole spears with SCUBA or snorkeling (Tables 1, 2; Fig. 1). Tissues were preserved in salt-saturated DMSO buffer (Amos & Hoelzel 1991) and stored at room temperature. For *manini*, genomic DNA was extracted using Omega Bio-Tek E-Z 96® Tissue DNA Kit (Norcross, GA, USA) following the manufacturers protocol, and resuspended in nanopure water. High molecular weight was confirmed by visualizing on a 1.5% agarose gel with GelRed® (Biotium, Inc. Fremont, CA, USA). For *kole* when samples sizes were > 30, a random subsample of 30 individuals were selected to be included in the library pool.

Manini analysis

Library preparation and sequencing

RADseq library preparation and sequencing was conducted by the core lab at Texas A&M Corpus Christi, starting with 150 ng of high-molecular weight genomic DNA per sample and following the double-digest RAD (ddRAD) protocol (Peterson *et al.* 2012). Briefly, this

process included digesting each sample with *Msp*I and *Eco*RI (New England Biolabs, Ipswich, MA, USA) followed by cleaning each sample with PEG solution using retained beads. Samples were then normalized to the same concentration followed by ligation of adapters. After digestion and ligation, a PCR was performed using dual-indexed primers. Fragments of between 325 bp and 400 bp were selected using BluePippin (Sage Science, Beverley, MA, USA). Following size selection, a Fragment Analyzer was run to visualize library size range followed by a qPCR to determine molarity of libraries. The resulting libraries were sequenced on a Illumina HiSeq® 4000 (150 paired-end reads, perfromed by NYU Langone Health Genome Technology Center). Sequence data for the samples were demultiplexed based on the barcodes from the adapters using *process_radtags*. Each individual library was sequenced across two or three independent runs for the purpose of increasing the number of sequence reads for each sample and to ensure congruence in nucleotide assignments.

Genotyping and de novo assembly of RADseq libraries

Raw reads obtained from Illumina runs were assessed for sequence quality, AT/GC content and overrepresented and duplicate sequences using *FastQC* v.0.10.1. As a reference genome is not available for *A. triostegus*, after initial quality assessment a *de novo* reference genome was assembled using *Rainbow* v. 2.0.4 as performed in the dDocent pipeline (Puritz *et al.* 2014a; Puritz *et al.* 2014b). After generating the reference genome and read mapping, single nucleotide polymorphisms (SNPs) detection was performed using *FreeBayes* v.1.10.54. A second filtering step was performed where variants were excluded if they were not genotyped in 50% of individuals, had a minimum quality score of 30, and minor allele count of 3. Individuals

with 10% missing genotypes were also excluded. Additionally, SNPs that did not meet expectations for Hardy-Weinberg Equilibrium were excluded.

Population genetic analyses

GENODIVE v.2.0b27 (Meirmans & Van Tienderen 2004) was used to generate genetic diversity indices, as well as to test for population structure. Genetic structure among sample locations was evaluated with an analysis of molecular variance (AMOVA; (Excoffier *et al.* 2005). Deviations from null distributions were tested with non-parametric permutation procedures (N = 9999). Pairwise F_{ST} statistics were generated to assess genetic structure between locations. False discovery rates were controlled for and maintained at $\alpha = 0.05$ among all pairwise tests (Benjamini & Yekutieli 2001; Narum 2006). Populations for the east and west side of Hawai‘i Island, Maui, and O‘ahu were analyzed as separate populations but no genetic differentiation was identified. The results being presented combined all the samples from both sides and are presented as a single island population.

Genetic partitioning was assessed using STRUCTURE v.2.3.2 (Pritchard *et al.* 2000), a Bayesian method that estimates ancestry and categorizes individuals into discrete populations. The simulation was run for 1 million generations with the first 100,000 discarded as burn-in. Five replicates of each simulation from K=1 to 10 genetic clusters were run. We determined the most likely number of genetic clusters (K) using STRUCTURE HARVESTER v.0.6.93 (Earl & von Holdt 2012). STRUCTURE results were analyzed using the on-line tool CLUMPAK (http://clumpak.tau.ac.il/index.html) (Kopelman *et al.* 2015) which integrates the program CLUMPP v.1.1.2 (Jakobsson & Rosenberg 2007), which minimizes the variance across all iterations. CLUMPAK then creates the final visualized output.

Kole analysis

Pooled library preparation

To save on sequencing costs, samples of *kole* were pooled into a single library. Pooling multiple individuals from the same species and sequencing the homogenized DNA (Pool-seq) is a cost-effective method to estimate allele frequencies within populations that has been utilized to evaluate population structure across of suite of diverse taxa (Kozak *et al.* 2014; Guo *et al.* 2016; Dennenmoser *et al.* 2017; Fischer *et al.* 2017). Despite the controversy regarding the use of Pool-seq versus individual sequencing (Cutler & Jensen 2010), it has been demonstrated that estimations of allele frequencies for high-frequency alleles are robust when individuals DNA are pooled in equal molar concentrations, even when taking sequencing error into account (Futschik & Schlötterer 2010), making Pool-seq an economical approach for population genomic analyses.

To confirm the quantity and purity, DNA was quantified using an AccuBlue assay (Biotium, Hayward, CA, USA) and assessed with the software SoftMax Pro 4.8. DNA samples were precipitated, dried using a speedvac, and resuspended in nanopure water (Thermo Scientific*, Barnstead, Dubuque, IA, USA). Equal molar concentrations from individual samples were then pooled to a total volume of 25 µl for a final number of 16 pooled DNA samples representing each island population.

DNA digestion and library preparation followed the ezRAD protocol by Toonen *et al.* (2013). Briefly, we first used the Kapa Hyper prep kit (Kapa Biosystems, Wilmington, MA, USA) and the Truseq PCR-free kit (Illumina, Inc., San Diego, CA, USA). This process included digesting each sample with *Dpn*II (New England Biolabs, Ipswich, MA, USA), which cleaves sequences at GATC cut sites, followed by cleaning each sample with AMPure XP beads

(Beckman Coulter Life Sciences, Indianapolis, IN, USA). Samples were then normalized to the same concentration followed by end repairs, A-tailing, and ligation of adapters. Fragments of between 350 bp and 700 bp were selected using SPRI beads (Applied Biological Materials, Inc., Richmond, BC, Canada). Afterwards, libraries were amplified by conducting PCR. The resulting libraries were sequenced on a Illumina MiSeq® (paired-end 150 bp; perfromed by Hawai'i Institute of Marine Biologys Genetic Core Facility). Raw reads were subsequently processed using the dDocent bioinformatics pipeline as described for *manini* with the exception that it was adjusted for analysing pooled sequence data.

Population level analyses

Genetic differentation between populations was calculated using assessPool (https://github.com/ToBoDev/assessPool). *AssessPool* is a R pipeline designed to analyze population structure from pooled data. It incorporates *PoPoolation2* v.1.201 (Kofler *et al.* 2011b) to generate pairwise F_{ST} values and the associated Fishers T-test. General population molecular indices were calculated using *Popoolation* v1.2.2 (Kofler *et al.* 2011a). *AssessPool* conducts a pre-analysis filtering step based on population pool size, miniumum total coverage per variable site, and maximum insertion/deletion length. This was followed by an additional filtering step conducted by *PoPoolation2*. The *PoPoolation2* parameters were set as: minimum depth threshold, 4; = maximum indel length, 5; minimum count, 2; minimum coverage, 22; pool size, 30. All other parameters remained at their default settings.

RESULTS

Population structure of manini

After the initial trimming, filtering and demultiplexing, we retained 80,955 loci. Following the second filtering step which accounted for coverage, minimum allele frequency and presence among the individuals included in the dataset, we identified 716 loci that met all the criteria to be used in downstream analyses.

Molecular diversity indices are summarized in Table 1. The number of alleles present in all populations ranged from 1.17 to 1.46 at Kure and Kaua‘i. The effective number of alleles were similar across locations and ranged from 1.046 to 1.067. Total heterozygosity ranged from 0.031 at Maui to 0.045 at Kure. A review of the inbreeding coefficient found that the influence of inbreeding is negligible across all population.

Population pairwise F_{ST} values are summarized in Table 3. Significance was determined after controlling for false discovery rates (corrected $\alpha = 0.009$). In the main Hawaiian Islands (MHI), population structure was found to be significant between all islands. The highest differentiation found was between Hawai‘i Island and O‘ahu (F_{ST} =0.071). After correcting for false discovery rates Hawai‘i Island no longer was differentiated from Kaua‘i. In the Northwestern Hawaiian Islands (NWHI), all island expect Kure grouped with Johnston Atoll to form one panmictic population. Kure was differentiated from the other NWHI with the greatest differentiation between Kure and Pearl and Hermes (F_{ST} =0.070). When comparing between the MHI and the NWHI, O‘ahu was not significantly differentiated from Johnston Atoll and the islands in the NWHI with the exception of Kure (F_{ST} =0.073). Kure was also found to be significantly differentiated from Maui (F_{ST} =0.057) but was not differentiated from Hawai‘i Island or Kaua‘i. Johnston Atoll also showed significant differentiation from Hawai‘i Island, Maui and Kaua‘i. The AMOVA analysis found significant differences among populations overall (F_{ST} =0.033, $P < 0.001$; Table 4).

The STRUCTURE analysis recovered two population clusters (k=2) (Fig. 2, Fig. S1). The analysis found one population consisting of Hawai‘i, Maui, Kaua‘i, and Kure, and a second population consisting of the remaining islands in the NWHI, O‘ahu and Johnston Atoll. Various levels of admixture were observed in all locations.

Population structure of kole

After the initial trimming, filtering and demultiplexing, we retained a total of 69,387 loci. After following the additional filtering steps as implement in PoPoolation2, we retained 4,292 loci that were called in all pools and were used for all downstream analyses.

Population pairwise F_{ST} values are summarized in Table S1 and presented as a heat map (Fig. 3). All pairwise comparisons showed differentiation between locations with F_{ST} values ranging from 0.019 between Ni‘ihau and Lana‘i, up to 0.041 between Kure and Kaua‘i. The average F_{ST} across all populations was 0.029. Within the MHI, Moloka‘i and O‘ahu were the more differentiated that the other islands and less differentiated with the islands in the NWHI. Within the NWHI, Gardiner had relative low differentiation between all islands (F_{ST}: 0.022 - 0.029).

DISCUSSION

Patterns of dispesal across the archipelago

Studies of connectivity along the Hawaiian Archipelago have added to our understandng on how dispersal patterns are shaped and the mechanisms that influence how biodiversity is exchanged in the marine realm. By undertaking a genomics approach we have uncovered patterns of dispersal that have not been previously observed in Hawaiian marine reef fishes. The

island-by-island structure observed in *kole* is inconsistent with patterns of structure found in other fishes and has only been observed in one other species, vermetid gastrodpods (Faucci et al., unpublished), whose genetic structure can be explained by their crawl-away larvae mode of reproduction. The high structure of *kole* contrasts with previous research which, using a single mtDNA maker, found a genetically homogenous population across most of the range and only showing structure at Maro Reef and Pearl and Hermes Atoll (Eble *et al.* 2009). By utilizing RADseq we were able to illuminate fine scale population structure that the use of a single marker was unable to reveal.

The results of this study also show patterns that corroborate previously identified patterns of connectivity. An allozyme analysis of *manini* showed structure between Hawai‘i Island and O‘ahu (Planes & Fauvelot 2002). Furthermore, the multi-species genetic breaks identified by Toonen *et al.* (2011) found that each island in the MHI is genetically distinct, a pattern consistent with our findings for *manini* (Table 3). Along the rest of the range, connectivity shows partial concordance with trends in other species. We recovered a highly connected population extending from French Frigate Shoals to Midway Atoll, which also includes Johnston Atoll. However, the furthest northwestern ecological break was found to occur between Midway Atoll and Pearl & Hermes (to the east) in most other species (Toonen *et al.* 2011), however, we found a break between Midway Atoll and Kure (to the west). Nonetheless, the long expanses of connectivity in the middle of the archipelago is recurring pattern in Hawaiian marine species.

The high connectivity of *manini* between Johnston Atoll and the archipelago is a pattern documented in other fish species (Craig *et al.* 2010; DiBattista *et al.* 2011; Fernandez-Silva *et al.* 2015). Johnston Atoll is the nearest land mass to the Hawaiian Archipelago, 885 km southwest of French Frigate Shoals, and is included in the Hawaiian biogeographic province based on high

faunal similarity. Many endemic Hawaiian fishes are found there (Randall, 2007; Briggs and Bowen, 2012), and it is likely a stepping stone for Indo-Pacific biodiversity to colonize into Hawaii (Bowen, 2016). Johnson Atoll has been implicated as the source of propagules in the middle of the archipelago (Rivera *et al.* 2004; Gaither *et al.* 2011b; Andrews *et al.* 2014). Dispersal from Johnston Atoll into the archipelago is further supported by biophysical models. Kobayashi (2006) identified two potential corridors into the archipelago from Johnston Atoll for species with PLDs greater than 40 days - one being French Frigate Shoals (Kobayashi 2006).

Two anamolous patterns were observed: no significant population partitions between O‘ahu and the NWHI, and between Kure and the MHI (Table 3, Figure 2). It is not clear what biological or physical drivers are facilitating this pattern, which has not been observed in other species.

Factors influencing dispersal

The two surgeonfish under investigation share similar life histories, including similar pelagic larval duration (PLD). The PLD for *kole* is estimated at 50-60 days (based on sisters species *C. striatus* in Doherty *et al.* (1995) and the PLD for *manini* ranges from 54-70 (Randall 1961; Longenecker & Langston 2008). Despite exhibiting comparable PLDs, the patterns of genetic structure and connectivity across the Hawaiian Archipelago are quite dissimilar: *manini* shows genetic subdivisions between each of the islands that make up the MHI as well as Kure, while the islands found along the nearly 1300 km distance between French Frigate Shoals and Midway, along with Johnston Atoll, were not genetically distinguishable from one another. Conversely, *kole* shows genetic structure between every pairwise comparison along the archipelago. This corroborates previous studies that show that PLD alone is not a strong

predictor of genetic dispersal (Weersing & Toonen 2009; Selkoe & Toonen 2011), particularly as it relates to Hawaiian endemic fishes (Selkoe *et al.* 2014).

Nonetheless, the disparity between these patterns of connectivity for the two surgeonfishes is grounded in some aspect of the dispersal capability of each species. Previous researchers have hypothesized that endemic Hawaiian fishes are descendents of poor dispersers (Hourigan & Reese 1987; Eble *et al.* 2009). Based on this scenario, after a rare colonization event into Hawai‘i they were unable to maintain connectivity with the widespread Pacific population. Although interspecific patterns varies, studies that have investigated population structure of endemic Hawaiian fishes, including groupers, damselfish, and surgeonfishes, have shown that endemic species have higher levels of population structure across the archipelago, relative to widespread species (Rivera *et al.* 2004; Ramon *et al.* 2008; Eble *et al.* 2009; Tenggardjaja *et al.* 2016). An exception to this pattern are the three endemic butterflyfishes which were all found to be genetically homogenous (Craig *et al.* 2010). Widespread species show very little to no structure across the archipelago (Craig *et al.* 2007; Eble *et al.* 2009; DiBattista *et al.* 2011; Eble *et al.* 2011a; Andrews *et al.* 2014).

The population structure of *manini* has traits associated with endemic species as well as wide-ranging species in that structure is observed across most of the MHI and Kure, but is genetically homogenous across the remainder of the NWHI and Johnston Atoll. Recall that due to diagnostic physical characters, Hawaiian *manini* are classified by some authors as a subspecies (Randall 1961; Randall 2007). However, research based on allozymes which characterized genetic structure of *manini* across the entire Indo-Pacific range found that the Hawaiian population was genetically distinct from the rest of the range, albiet with F_{ST} values that indicate an isolated population rather than a species designation (Planes & Fauvelot 2002). A thorough

analysis using more sophisticated genetic techniques would need to be conducted to properly characterize Hawaiian *manini* as truly an endemic species. The conflicting patterns of dispersal in *manini* are not unknown. High population structure between small distances was documented in other archipelagos and even in lagoons (Planes *et al.* 1996; Planes *et al.* 1998a). Additionally. recent evidence from a parentage analysis conducting on O‘ahu found that the majority of larval settle less than 30 m from their spawning grounds, even in the face of strong currents (Coleman, unpublished). However, *manini* has also shown the ability to maintain connectivity across vast distances (Planes & Fauvelot 2002; Mirams *et al.* 2011). The dispersal ability of *manini* is clearly not static and is likely influenced by a variety of abioic and biotic factors.

Habitat preference and larval behavior are known to play a key role in dispersal and settlement queues (Jones 2015). The ecosystem of the MHI differs greatly from the relatively pristine ecosystem of the NWHI which was designated as the Papahānaumokuākea Marine National Monument in 2006, thereby limiting anthropogentic impacts. The MHI are made of up of high islands with steady freshwater run off that transport nutrients into surrounding water. Whereas, the NWHI are the oldest land in the archipelago and consist of low islands and atolls. The human impact in the MHI has also lead to degraded reefs, overfishing, and pollution, among other pressures, and have distorted many of the natural processes in this region. Also, many species that are present in the NWHI are rare or not found in the MHI such as *Acropora* corals (Grigg *et al.* 1981), a common species found throughout the Indo-Pacific, as well as many endemic fishes (Kosaki *et al.* 2016). However, differences in ecosystems does not seem to explain why O‘ahu *manini* shows high connectivity with the NWHI or why Kure shows connectivity with the MHI. Although, identifying the underlying factors promoting or inhibiting

dispersal across the archipelago remain elusive, this research provides further support for the theme that Hawaiian endemic species exhibit limiting disepersal.

CONCLUSIONS

Here, we have highlighted the utility of using two different genomic methods for assessing population structure to describe dispersal and connectivity in a contemporary framework. Although not directly compariable, they have both shown the ability to uncover fine scale connectivity patterns. A comparative analysis in *kole* against the targeted locus approach, which showed highly connected populations, and a genomic approach, which revealed higher levels of isolation across the archipelago than have been previously described, underscores the need to include genomics to identify contemporary patterns of dispersal.

As the field of population genetics continues to evolve, a suite of tools are becoming more readily available to evaluate patterns of connectivity (Germer *et al.* 2000; Andrews & Luikart 2014; Puritz *et al.* 2014b). The exponential increase in data will continue to revolutionize our ability to identify many factors that influence species ability to diversify including ecological important traits (Hohenlohe 2014), historic role of hybridization in shaping biodiversity (Meier *et al.* 2017), genetic basis for species interactions and adaptation (Allendorf *et al.* 2010; Hohenlohe *et al.* 2010), among others.

The utility of conducting a targeted marker analysis is not obsolete, and can effectively be used in concert to describe evolutionary and contemporary patterns of connectivity, along with the associated mechanisms facilitating these patterns (Gaither *et al.* 2015). As we move forward in assessing connectivity across the Hawaiian Archipelago, it may be worthwhile to revist many of these past studies that used a targeted loci, and integrate a genomics perspective

to uncover contemporary patterns of dispersal and identify the mechanisms that shaped the evolution of Hawaii's unique biodiversity.

Figure 3.1. Colletion locations and sample sizes of *Acanthurus triostegus* (left) and *Ctenocheatus strigosus (right)*. Solid line designates the Northwestern Hawaiian Islands which, in 2006, was designated the Papāhanaumokuākea Marine National Monument. Filled darker areas represent current coastlines while light areas represent the maximum historical above-water island area. Sample sizes for each species are in parentheses (*A. triostegus*, *C. strigosus*). *Photo credit: Keoki Stender*

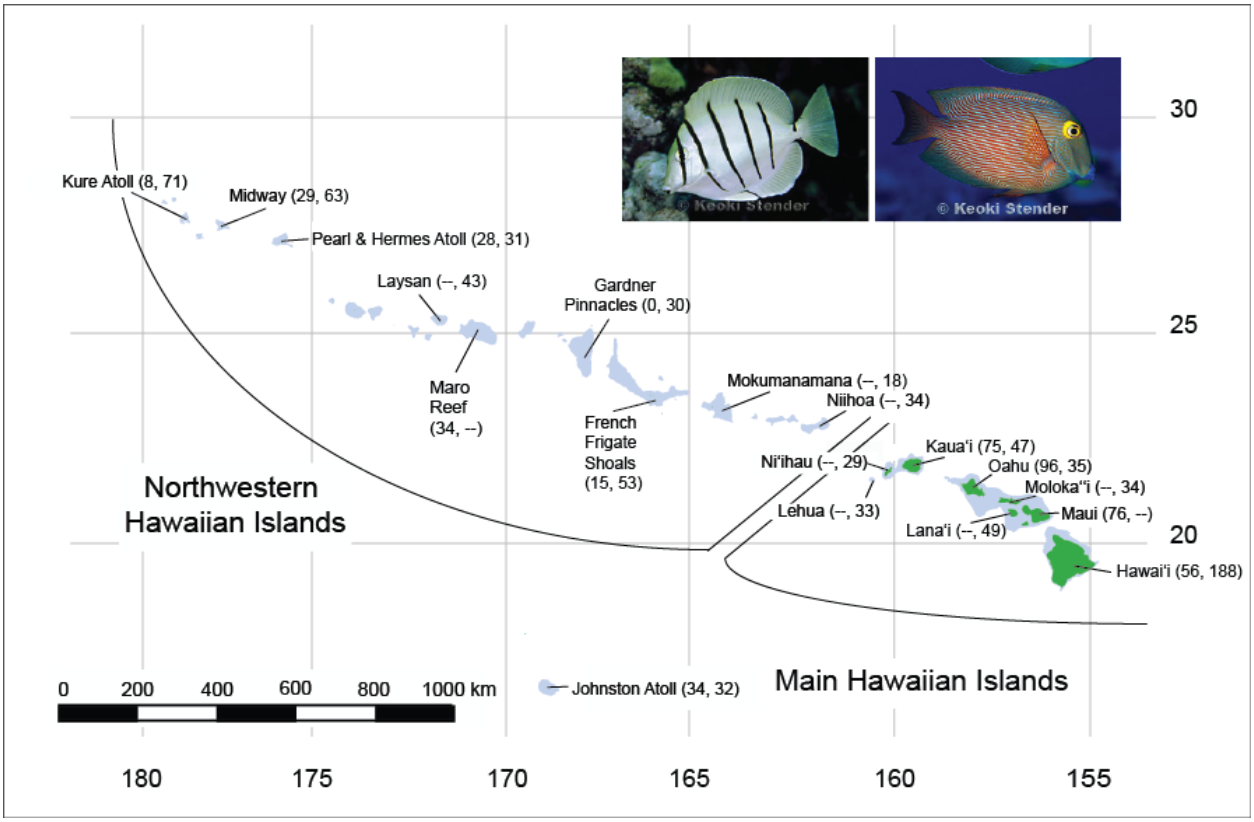

Figure 3.2. STRUCTURE bar plot (k=2) for Hawaiian populations of *Acanthurus triostegus* using 761 SNPs. Abbreviations: Hawaii, HAW; Maui, MAU; O‘ahu, OAH; Kaua‘i, KAU; French Frigate Shoals, FFS; Maro Reef, MAR; Pearl and Hermes Atoll, PH; Midway Atoll, MID; Johnston Atoll, JOH; Kure Atoll, KUR.

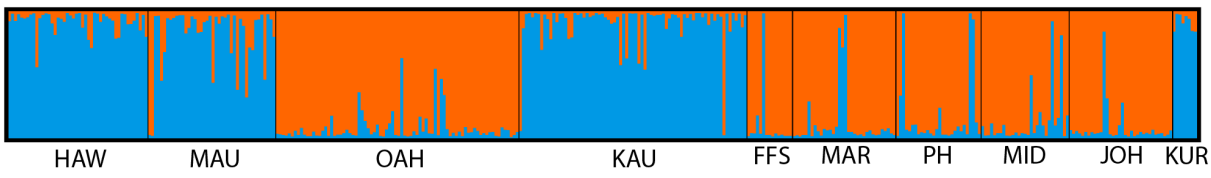

Figure 3.3. Heat map of F_{ST} values for *Ctenochaetus striogsus* based on Pool-seq libraries (No.of SNPs=4,292). Abbreviations: Hawai‘i Island, HAW; Moloka‘i, MOK; Lana‘i, LAN; O‘ahu, OAH; Kauai, KAU; Ni‘ihau, NII; Lehua Rock, LEH; Nihoa, NIH; Mokukmanana, MOK; Frigate Shoals, FFS; Gardiner, GAR; Laysan, LAY; Pearl and Hermes Atoll, PH; Midway Atoll, MID; Kure Atoll, KUR; Johnston Atoll, JOH.

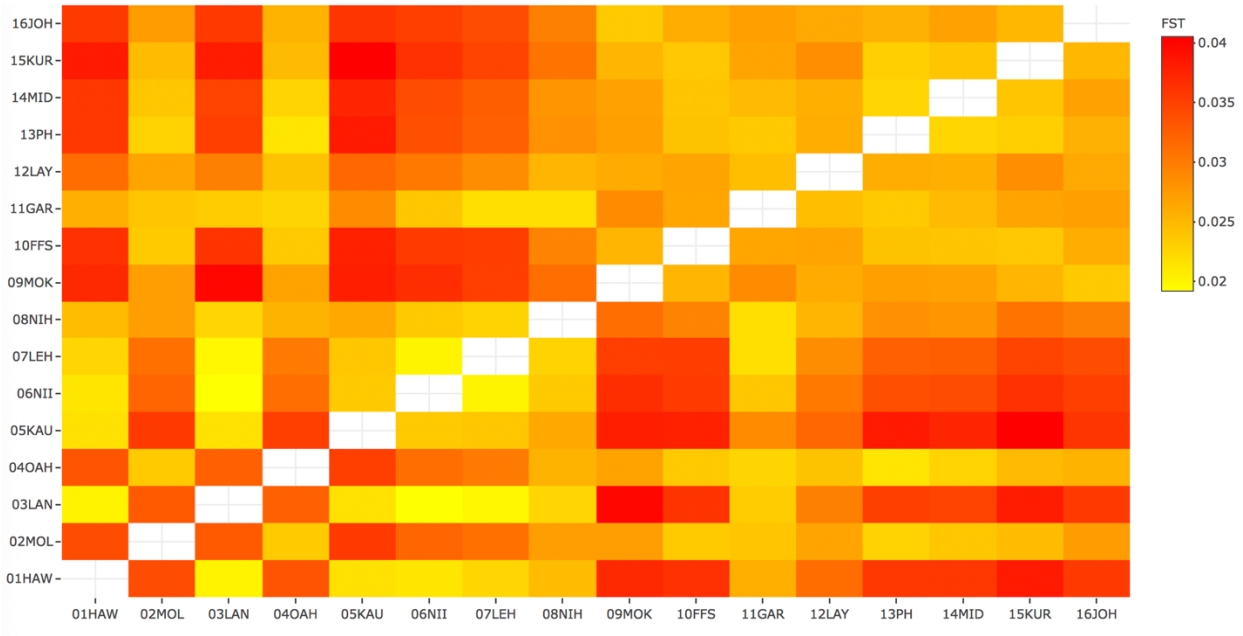

Table 3.1. Molecular diversity indices for populations of *Acanthurus triostegus* based on 761 SNPs. Number of individuals sequenced (n), average number of alleles per locus (N_a), effective number of alleles (N_{eff}), Observed heterozygosity (H_O), heterozygosity between populations (H_S), total heterozygosity (H_T), and inbreeding coefficient (G_{IS}) are presented.

Sample Location	*n*	N_a	N_{eff}	H_O	H_S	H_T	G_{IS}
Hawaii	56	1.340	1.054	0.047	0.037	0.037	-0.251
Maui	76	1.281	1.046	0.040	0.031	0.031	-0.281
Oahu	96	1.453	1.055	0.048	0.039	0.039	-0.248
Kauai	75	1.461	1.061	0.050	0.043	0.043	-0.176
French Frigate Shoals	15	1.195	1.056	0.050	0.038	0.038	-0.318
Maro Reef	34	1.325	1.055	0.049	0.038	0.038	-0.284
Pearl and Hermes	28	1.287	1.054	0.048	0.037	0.037	-0.293
Midway	29	1.308	1.056	0.051	0.039	0.039	-0.299
Kure	8	1.168	1.067	0.055	0.045	0.045	-0.217
Johnston	34	1.303	1.055	0.047	0.038	0.038	-0.253
All locations	451	1.312	1.056	0.049	0.039	0.039	-0.262

Table 3.2. Summary of collections for Ctenochaetus strigosus across the Hawaiian Archipelago

Sample Location	n
Hawai‘i Island	188
Moloka‘i	34
Lana‘i	49
O‘ahu	35
Kaua‘i	47
Ni‘ihau	29
Lehua	33
Nihoa	34
Mokumanamana	18
French Frigate Shoals	53
Gardiner	30
Laysan	43
Pearl and Hermes	31
Midway	63
Kure	71
Johnston Atoll	32
Total	790

Table 3.3. Matrix of pairwise F_{ST} statistics for 10 populations of *Acanthurus triostegus* based on 761 SNPs. Bolded numbers indicate significance at $p<0.05$. Italicized numbers indicate significance after controlling for false discovery rates at $\alpha = 0.05$ (as per Narum, 2006). The corrected $\alpha = 0.009$. Owing to low sample size Ni‘ihau has been excluded from the analysis. Abbreviations: Hawai‘i Island, HAW; Maui, MAU; O‘ahu, OAH; Kauai, KAU; French Frigate Shoals, FFS; Maro Reef, MARO; Pearl and Hermes Atoll, PH; Midway Atoll, MID; Kure Atoll, KUR; Johnston Atoll, JOH.

	HAW	MAU	OAH	KAU	FFS	MARO	PH	MID	JOH
HAW	--								
MAU	***0.026***	--							
OAH	***0.071***	***0.043***	--						
KAU	**0.007**	***0.022***	***0.051***	--					
FFS	***0.054***	***0.043***	<0.001	***0.048***	--				
MARO	***0.054***	***0.038***	0.006	***0.053***	<0.001	--			
PH	***0.048***	***0.033***	0.004	***0.044***	<0.001	<0.001	--		
MID	***0.048***	***0.026***	0.002	***0.045***	<0.001	<0.001	<0.001	--	
KUR	0.005	***0.057***	***0.073***	<0.001	***0.064***	***0.065***	***0.070***	***0.065***	--
JOH	***0.065***	***0.050***	0.001	***0.060***	<0.001	<0.001	0.005	0.002	***0.076***

Table 3.4. Results of the analysis of molecular variance (AMOVA) based on 761 SNPs for *Acanthurus triostegus*. Bolded values denote significance at $P < 0.05$

Source of Variation	F-statistic	% variation	F-value	Std.Dev.	P-value
Within Individual	F_{IT}	1.214	-0.214	0.04	--
Among Individual	F_{IS}	-0.248	-0.256	0.041	1.000
Among Population	F_{ST}	**0.033**	**0.033**	**0.008**	**0.001**

Table S3.1 Matrix of pairwise F_{ST} statistics for 16 populations across the Hawaiian Archipelago of *Ctenochaetus strigosus* using pooled DNA libraries. Abbreviations: Hawai‘i Island, HAW; Moloka‘i, MOK; Lana‘i, LAN; O‘ahu, OAH; Kauai, KAU; Ni‘ihau, NII; Lehua Rock, LEH; Nihoa, NIH; Mokumanamana, MOK; Frigate Shoals, FFS; Gardiner, GAR; Laysan, LAY; Pearl and Hermes Atoll, PH; Midway Atoll, MID; Kure Atoll, KUR; Johnston Atoll, JOH.

	HAW	MOL	LAN	OAH	KAU	NII	LEH	NIH	MOK	FFS	GAR	LAY	PH	MID	KUR
HAW	--														
MOL	0.0341	--													
LAN	0.0202	0.0329	--												
OAH	0.0333	0.0235	0.0324	--											
KAU	0.0217	0.0356	0.0216	0.0351	--										
NII	0.0212	0.0320	0.0192	0.0312	0.0236	--									
LEH	0.0226	0.0311	0.0198	0.0302	0.0238	0.0200	--								
NIH	0.0247	0.0273	0.0225	0.0254	0.0264	0.0237	0.0228	--							
MOK	0.0371	0.0273	0.0398	0.0268	0.0380	0.0367	0.0353	0.0313	--						
FFS	0.0364	0.0234	0.0361	0.0236	0.0377	0.0355	0.0354	0.0294	0.0253	--					
GAR	0.0258	0.0239	0.0233	0.0227	0.0288	0.0238	0.0219	0.0220	0.0288	0.0265	--				
LAY	0.0313	0.0267	0.0297	0.0241	0.0317	0.0303	0.0285	0.0253	0.0261	0.0267	0.0245	--			
PH	0.0358	0.0228	0.0351	0.0213	0.0383	0.0337	0.0322	0.0283	0.0271	0.0242	0.0236	0.0259	--		
MID	0.0359	0.0238	0.0348	0.0227	0.0373	0.0340	0.0326	0.0279	0.0269	0.0240	0.0249	0.0258	0.0226	--	
KUR	0.0384	0.0248	0.0381	0.0249	0.0405	0.0364	0.0348	0.0308	0.0253	0.0237	0.0267	0.0285	0.0231	0.0239	--
JOH	0.0356	0.0274	0.0356	0.0255	0.0360	0.0351	0.0341	0.0298	0.0235	0.0260	0.0270	0.0263	0.0257	0.0268	0.0252

Table S3.2.

Table S3.2. Evanno output identifying the ideal number of clusters present across populations						
# K	Reps	Mean Ln P(K)	St.dev Ln P(K)	Ln'(K)	|Ln''(K)|	Delta K
1	5	-33918.76	1.3813	NA	NA	NA
2	5	-32588.04	5.8833	1330.72	1246.84	211.92935
3	5	-32504.16	10.26	83.88	1170.84	114.116742
4	5	-33591.12	1246.8203	-1086.96	1701.54	1.364704
5	5	-32976.54	1382.7738	614.58	2957.68	2.138947
6	5	-35319.64	3776.1411	-2343.1	2846.96	0.753934
7	5	-34815.78	4854.4566	503.86	10977.02	2.261225
8	5	-45288.94	11566.0673	-10473.16	18067.32	1.562097
9	5	-37694.78	9498.7453	7594.16	4895.78	0.515413
10	5	-34996.4	1430.4382	2698.38	NA	NA

CHAPTER 4 GENOMIC ASSESSMENT OF THE LARVAL ODYSSEY: SEL-RECRUITEMETN AND BIASED SETTLEMENT IN THE HAWAIIAN SURGEONFISH *ACANTHURUS TRIOSTEGUS SANDVICENSIS*

ABSTRACT

The gap between spawning and settlement location of marine fishes, wherein the larvae occupy an oceanic phase, is a great mystery in both their natural history and conservation. Recent genomic approaches offer some promise, most especially in linking parent to offspring with assays of nucleotide polymorphisms. Here, we apply this methodology to the endemic Hawaiian manini (*Acanthurus triostegus sandvicensis*), a surgeonfish with a long pelagic larval stage of ~54 – 77 days. We collected 606 adults and 607 juveniles from 23 locations around the island of Oʻahu, Hawaiʻi. Based on 399 SNPs, we assigned 68 of these juveniles back to a parent (11.2% assignment rate). The western and northern sides of the island, which are subject to westerly currents, had little or no detected recruitment. In contrast, the majority of juveniles (94%) sampled along the eastern shore originated on this side of the island, primarily within semi-enclosed Kāneʻohe Bay. Nearly half of the assigned juveniles were found in the southern part of Kāneʻohe Bay, with local settlement likely facilitated by extended water residence time in this region. Several instances of self-recruitment were observed along the eastern and southern shores. Cumulatively, these findings indicate that most dispersal is between adjacent regions on the eastern and southern shores. Regional management efforts for manini and possibly other reef fishes will be effective only with collaboration among adjacent coastal communities, consistent with the traditional *moku* system of native Hawaiian resource management.

INTRODUCTION

Long-range dispersal in reef fishes is limited to the pelagic larval stage, as most coral reef organisms maintain confined home ranges as juveniles and adults (Leis 1991; Leis & McCormick 2002; Hellberg 2009). The pelagic larval phase lasts from weeks to months in marine fishes thereby making it difficult to determine the spawning location of origin. Identifying this gap in knowledge and understanding the extent of connectivity between locations on a local scale is essential for proper stewardship of these coastal resources throughout their life cycles (Johnson *et al.* 2018).

The convict surgeonfish (*Acanthurus triostegus sandvicensis*), known locally as *manini*, is a common surgeonfish in the Hawaiian Islands (Randall 2010). *Manini* is heavily targeted by recreational fisheries (e.g. sport, leisure, or subsistence) and is described as an exploited species by Hawai‘i’s Division of Aquatic Resources (Longenecker *et al.* 2008). In Hawai‘i, it is estimated that up to 24% of the households participated in recreational fishing and that nearly 1.2 million kg year^{-1} of reef-associated catch, 84% of which comes from non-commercial fisheries, are extracted from the Main Hawaiian Islands (McCoy *et al.* 2018). Therefore, the pressure of recreational fisheries on coastal resources is substantial, and as a result target species in the Main Hawaiian Islands have 50% lower biomass than in the uninhabited Northwest Hawaiian Islands (Friedlander *et al.* 2018). In addition to extensive exploitation, current regulations may be ineffective because the minimum legal size is smaller than the average minimum size at sexual maturity, resulting in a majority of females entering the fishery before reproduction (Longenecker *et al.* 2008; Schemmel & Friedlander 2017).

Under-managed fisheries can lead to overfishing and can inhibit long-term sustainability for communities that rely on these resources for subsistence. In response to this potential threat, legislation was passed in 1994 that allowed state and local agencies to create community-based subsistence fishing areas (CBSFAs), which have so far been implemented on the islands of Kauaʻi and Molokaʻi. Community members on Oʻahu have recognized that maintaining food security will require new management approaches or a return to *moku,* traditional Hawaiian management strategies to protecting coastal resources. As a result, *manini* was identified by Native Hawaiian community leaders as a species of concern, prompting this effort to identify connectivity patterns between offspring and adults. To advance this goal, one of the first steps is to determine the source of fish populations.

Several methods can be used to identify connectivity and dispersal patterns among reef fishes (Jones 2015). A traditional population genetics approach is effective to characterize connectivity across distances of 100s to 1,000s of km; however, it is usually ineffective at smaller spatial scales, such as individual islands or archipelagos, where it is difficult to detect signals of isolation within the existing pool of genetic diversity (Saenz-Agudelo *et al.* 2009). Chemical tagging has proven to be an effective method (Jones *et al.* 1999), but it is labor intensive, expensive, and has only been applied a handful of times. Hydrodynamic and biophysical models have potential to identify general patterns of larval dispersal (Kobayashi 2006; Jones 2015). This computational approach simulates the movement and dispersal of virtual particles and incorporates physical characteristics of the surrounding environment as well as complex biological components to make predictions of larval dispersal. Ideally these models are concordant with empirical data (Galindo *et al.* 2010; Leray *et al.* 2010; White *et al.* 2010; Bowen 2016) particularly by matching genetic connectivity to oceanic circulation models. However,

despite some success, model predictions often fail to match what is observed in nature (Selkoe *et al.* 2016a).

Genetic parentage analyses have proven to be powerful tool for identifying fine scale connectivity. However, these analyses are limited, and may explain no more than 26% of the variation in true connectivity within a population (Christie *et al.* 2017). Nonetheless, parentage analyses have described connectivity in a variety of taxa including butterflyfishes (Abesamis *et al.* 2017), clownfishes (Jones *et al.* 1999; Saenz-Agudelo *et al.* 2012), gobies (D'aloia *et al.* 2013), groupers (Almany *et al.* 2013), snappers (Harrison *et al.* 2012), and surgeonfish (Christie *et al.* 2010). In the case of the yellow tang (*Zebrasoma flavescens*), Christie *et al.* (2010) provided direct evidence of connectivity within an existing network of marine protected areas (MPAs) around Hawai‘i Island. In the clownfish *Amphiprion percula*, Planes *et al.* (2009) found high levels of local recruitment to natal reefs in Kimbe Bay, Papua New Guinea, as well as recruitment to adjacent locations within a network of MPAs. These studies show the efficacy of parentage analyses as a tool for characterizing dispersal patterns across small spatial scales.

With advances in genomic technology, single nucleotide polymorphisms (SNPs) are becoming increasingly popular for parentage analyses. Comparable studies have consistently shown that SNPs more accurately assign parent-offspring pairs when compared to microsatellites (Hauser *et al.* 2011; Andrews *et al.* 2018; Flanagan & Jones 2018; Thrasher *et al.* 2018) with as few as 100 SNPs being sufficient to resolve parentage (Flanagan & Jones 2018). Here, we use SNP-based parentage analysis to describe dispersal and connectivity patterns of manini around the island of O‘ahu, with a focus on Kāne‘ohe Bay on the eastern side of the island. This is the largest semi-enclosed bay in the main Hawaiian Islands with an area of 45 km^2, and a popular fishing spot with well-described oceanographic properties and biotic communities (Bahr *et al.*

2015). Our results are intended to inform community-based management efforts by identifying propagule sources and sinks of the larvae of a locally important fish species, as well as highlight areas that may be particularly vulnerable to excessive fishing pressure.

MATERIAL AND METHODS

Study species

Manini is an herbivorous surgeonfish found throughout the Indo-Pacific that specializes on benthic algae. It often occurs in large schools along reef flats and the outer reef. Schemmeland Friedlander (2017) recently described aspects of the reproductive biology of *manini* across the Hawaiian Islands: On O‘ahu, group spawning occurs before dusk where aggregations of 25-800 individuals form a few days before the new and full moon. Spawning takes place at depths from 7-30 m and peaks during February-June. However, spawning occurs throughout the year and is likely to be highly variable across the Hawaiian Islands. The pelagic larval duration (PLD) is estimated to range from 54 days (Longenecker & Langston 2008) up to approximately 77 days (Randall 2005a), a longer interval than most surgeonfishes (Leis & McCormick 2002; Eble *et al.* 2009).

Sampling and DNA extraction

Between May 2015 and July 2017, a total of 1,213 tissue samples of *manini* (606 adults; 607 juveniles) were collected from 23 locations around the island of O‘ahu, Hawai‘i using pole spears with SCUBA or snorkeling (Table 4.1, Fig. 4.1). Individuals < 121 cm were classified as juveniles (i.e., reproductively immature) based on the average of size of maturity for males and females (Randall 1961; Longenecker *et al.* 2008; Schemmel & Friedlander 2017). Tissues were

transferred to 95% ethanol and stored at room temperature. Genomic DNA was extracted using Omega Bio-Tek E-Z 96® Tissue DNA Kit (Norcross, GA, USA) and following the manufacturers protocol. Genomic DNA was resuspended in nanopure water. High molecular weight DNA was confirmed by visualizing on a 1.5% agarose gel with GelRed® (Biotium, Inc. Fremont, CA, USA).

Library preparation and sequencing

Restriction-site associated DNA (RAD) library preparation and sequencing was conducted by the Texas A&M core lab, starting with 150 ng of high-molecular weight genomic DNA per sample and following the double-digest RAD (ddRAD) protocol (Peterson *et al.* 2012). Briefly, this process included digesting each sample with *Msp*I and *Eco*RI (New England Biolabs, Ipswich, MA, USA) followed by cleaning each sample with PEG solution using retained beads. Samples were then normalized to equimolar concentration followed by ligation of sequencing adapters. After digestion and ligation, a PCR was performed using dual-indexed primers. Fragments of between 325 bp and 400 bp were selected using BluePippin (Sage Science, Beverley, MA, USA). Following size selection, a Fragment Analyzer was run to visualize library size range followed by qPCR to determine molarity of libraries. The resulting libraries were sequenced on a Illumina HiSeq® 4000 (150 paired-end reads, performed by NYU Langone Health Genome Technology Center). Sequence data were demultiplexed based on the barcodes from the adpaters using *process_radtags* found in the software package *STACKS* v. 2.41 (Catchen *et al.* 2011; Catchen *et al.* 2013). Each individual library was sequenced across two or three independent runs to increase the number of sequence reads for each sample and to ensure congruence in assignments between parents and offspring.

Genotyping and de novo assembly of RADseq libraries

Raw reads obtained from Illumina runs were assessed for sequence quality, AT/GC content and overrepresented and duplicate sequences using *FastQC* v. 0.10.1 (Andrews 2010). As a reference genome is not available for *A. triostegus*, after initial quality assessment, a *de novo* reference genome was assembled using *Rainbow* v. 2.0.4 (Chong *et al.* 2012) as performed in the dDocent pipeline (Puritz *et al.* 2014a; Puritz *et al.* 2014b). After generating the reference genome and mapping reads, SNP detection was performed using *FreeBayes* v. 1.10.54 (Garrison & Marth 2012). A second filtering step was performed where variants were excluded if they were not genotyped in 50% of individuals, had a minimum quality score of 30, and minor allele count of 3. Individuals with 10% missing genotypes were also excluded. Additionally, SNPs that did not meet the Hardy-Weinberg Equilibrium assumption were excluded.

Genetic parentage analysis

We conducted a parentage analysis using CERVUS v. 3.0.7 (Marshall *et al.* 1998; Kalinowski *et al.* 2007). This program calculates the likelihood that each candidate is the parent, taking into account population allele frequencies and genotype errors. An allele frequeny analysis was conducted to determine the suitability of loci for downstream assessment. CERVUS requires a parentage analysis simulation to determine the feasibility of the analysis given the set of loci and to calculate the critical likelihood ratios (LOD) to provide confidence in parent-offspring assignments. For the simulation to determine the critical LOD scores, we used 100,000 offspring (as recommended by the authors of CERVUS), an estimated genotyping error rate of 0.01, a proportion of loci typed across all individuals of 0.6868, and using an estimated number

of candidate parents sampled. To get the estimated number of sampled parents we first needed the population size of *manini* around O‘ahu which was estimated to be 1,058,942 +/- 265,902 (Mean +/- SE) based on 228 surveys from 2010-2016 (pers. comm. Ivor Williams, NOAA). The number of typed loci was 200 which was determined after the final number of SNPs was resolved. The genotype of each offspring was then compared to each candidate parent and a random individual in the population to calculate a likelihood ratio. This ratio is presented as a LOD score, the natural logarithm of calculated likelihood ratio. A positive LOD score indicates that a candidate parent is more likely to be the true parent, whereas a negative LOD score indicates the candidate parent is unlikely to be the true parent. Parent-offspring assignments were accepted at a 95% confidence level. Dispersal distances were estimated using the distance calculator tool from sea-seek.com (https://www.sea-seek.com/tools/tools.php).

RESULTS

SNP analysis

After initial trimming, filtering and demultiplexing, we retained a total of 80,955 loci. Following the second filtering step which accounted for coverage, minimum allele frequency and presence among the individuals included in the dataset, we identified 399 loci that met all the criteria for downstream analyses. When individual libraries were analyzed, they showed patterns consistent with the analyses of the concatenated dataset. In other words, parent-offspring assignments were the same when analyzed as individual libraries and when using the concatenated dataset. However, the concatenated dataset included a few more individuals due to the strict filtering process; therefore, the results present here are only for the individual library analyses.

Parental assignment

Of the 607 juveniles screened for DNA parentage analysis, we assigned 68 juveniles back to a parent (Table 4.2, Fig. 4.1), and the geographic distribution of assignments was highly uneven. No assignments were detected on the north shore of Oahu. Along the western side only one location, Kahe (KAH in Fig. 4.1), had offspring recovered in which all the adults were collected in Maunalua Bay (MB in Fig. 4.1). Along the south shore, the only assigned juvenile was a case of self-recruitment at China Walls (CW in Fig. 4.1). The largest concentration of assigned individuals was found along the eastern side, particulary within Kāneʻohe Bay, accounting for 94% of the assignments in this study (Fig. 4.2). The juveniles recovered along the eastern shore mostly originated from locations on the east side of the island with a few individuals originated from the south shore and one instance of dispersal from Kahe on the west. There were several instances of self-recruitment in addition to the one at China Walls; one at Kailua (KAI in Fig. 4.1), and five within Kāneʻohe Bay.

Distance between parent and detected offspring ranged from 0.25 km between Reef 14 and 16 in South Kāneʻohe Bay, to 78 km between Kahe and Kailua. The average distance between spawning and settlement was 27 km. The highest proportions of juveniles were found to have dispersed 10-15 km (27%) and between 25-30 km (23%). These trends can be attributed to the exportation of juveniles from Lāʻie and Kailua into Kāneʻohe Bay.

DISCUSSION

This study assigned 68 (11.2%) sampled juveniles back to a parent. This is a remarkably high recover rate considering this species has potential to remain in the planktonic phase for

nearly two months. For comparison, Christie (2010) sampled approximately 1,100 adults and juveniles of another surgeonfish, the yellow tang (*Zebrasoma flavescens*), along the windward coast of Hawai‘i Island and recovered four (0.36%) parent-offspring pairs. However, the low assignment may be atttributed to sampling a smaller proportion of the adult population. The population of yellow tang around Hawai‘i Island is estimated to be 4.2 million individuals, four times the estimated population size of manini around O‘ahu. In studies where the population sizes are much lower, a higher assignment rate is expected (Christie *et al.* 2017). Parentage analyses conducted on clownfishes (*Amphiprion polymnus*, *A. percula*) in Papua New Guinea assigned ~20% and 64% of sampled juveniles back to their parents (Saenz-Agudelo *et al.* 2009; Berumen *et al.* 2012). However, studies of butterflyfish (*Chaetodon vagabundus*) and groupers (*Plectropomus areolatus*) in the same region had a recovery rate of 8% and 10%, respectively (Berumen *et al.* 2012; Almany *et al.* 2013). Along the Great Barrier Reef of Australia, Harrison *et al.* (2012) had a recovery rate of 12% for a grouper (*Plectropomus maculatus*) and 16% for a snapper (*Lutjanus carponotatus*).

Bernardi *et al.* (2012) demonstrated that planktonic larval fish from a single spawning event may remain in close proximity, perhaps using shared sensory and behavioral mechanisms (Dixson *et al.* 2008). Despite the potential for these sophisticated behaviors, our initial screen revealed that no two individuals shared the same parent, providing no evidence of siblings at any of our 23 sample sites.

As exhibited above, the ability to assign juveniles back to their parent is highly variable and is influenced by many factors including life history strategies, physical processes of the surrounding environment, and the proportions of the adult population sampled. In the case of the clownfish and Great Barrier Reef studies, the adult population was heavily sampled, increasing

the success of assigning a juvenile to a parent. Additionally, clownfish are demersal spawners with a relatively short planktonic larval phase of ~11 days (Almany *et al.* 2007).

The distance between spawning and settlement of juvenile fishes also varies. On Hawai‘i Island, Christie *et al.* (2010) detected dispersal distances as high as 184 km which is attributed to a combination of passive transport and active behavioral mechanisms. No evidence of self-recruitment was observed in that study. In contrast, clownfish exhibit high levels of self-recruitment as well as shorter disperal distance (35 km) (Planes *et al.* 2009). In the current study, we assigned juveniles to parents that were separated by as little as ~0.25 km, to locations as far as 79 km apart (Kahe to Kailua; Table. 4.2). However, the majority of offspring were recovered within 30 km of their spawning location. The general patterns of limited dispersal we observed in *manini* are similar to many dispersal kernels obtained for marine fishes, which show a high proportion of recruitment close to spawning site that tapers down as distance increases (Jones 2015). In one of the most thorough dispersal kernal studies, D'Aloia *et al.* (2015) detected a mean dispersal of only 1.7 km in Belizean gobies (*Elacatinus lori*), with no dispersal event detected <16.4 km.

The dispersal patterns of *manini* around O‘ahu are quite complex and cannot be explained by any single factor. *Manini* are known to spawn in pairs as well as large groups (Robertson 1983). Subtidal habitat and marine physical processes vary around the island, both of which may influence dispersal patterns. Additionally, larval behavior will influence settlement, and larval *manini* can delay metamorphosis as needed to recruit to appropriate habitat (Randall 1961; McCormick 1999). Below we discuss the general patterns of disperal along each coast of O‘ahu, and discuss the potential mechanisms influencing dispersal.

North, west, and south O‘ahu

The assignment of parent-offspring pairs on the north and western side of O‘ahu is strikingly low when compared to other areas of the island. There is only one instance of dispersal from Kahe, located on the southwestern side of O‘ahu, to Kailua, and another instance from Maunalua Bay to Kahe. No additional parent-offspring assignments were detected along the entire western and northern coasts. There are several possible reasons for this trend. First, sampling effort along the western and northern side of the island are lower compared to other regions (Table 4.1). Collection efforts along the northern coast were constrained by higher wave energy with winter swells often > 7 m in height (Fletcher *et al.* 2008) , making many potential sites inaccessible. When collecting was possible, few juveniles were located. The number of juveniles collected from both western and northern coasts account for only 12% of the total juvenile collection. As an effort was made to collect both adults and juveniles at each collection location, the ratio of adults to juveniles may reflect the biological reality of the presence and absence of each size class. However, to verify this, a more systematic survey of *manini* adults and juveniles would be required.

The low recovey on the northern and western side may also be attributed to the currents surrounding O‘ahu that drive dispersal in a westerly direction. The North Hawaiian Ridge Current (NHRC) flows in a west-northwesterly direction adjacent to the northern coast of O‘ahu (Firing 1996). On the southern coast of O‘ahu, the Hawaii Lee Current (HLC) flows northwest following the Hawaiian Ridge from Maui to Kaua‘i (Lumpkin 1998). Therefore, propagules of *manini* orginating on the west or north coasts may be carried west towards Kauai. Notably, planktonic species that are typically found nearshore (within 1 km) on the eastern side of Oahu are more common offshore along the western coast of O‘ahu (Hassett & Boehlert 1999).

The dispersal observed from Maunalua Bay to Kahe is consistent with the flow of the HLC. However the dispersal from Kahe, and all locations along the south shore, to the eastern side of Oʻahu is against the HLC. The maximum flow of the HLC reaches 20 cm/s, although there is some interannual fluctation in the strength of the current (Lumpkin 1998). Nonetheless, the HLC would be predicted to be major barrier to dispersal in an easterly direction which makes us speculate that other physical or biological mechanisms are facilitating dispersal towards the eastern side of Oʻahu.

East Oʻahu and Kāneʻohe Bay

On the eastern side of Oʻahu, the tide floods to the southeast and ebbs to the northwest. Smaller scale circulation features are also established by headlands. Hence some of the fine scale coastal processes support the settlement of juveniles in both a north and south direction, as well as importation of larva into Kāneʻohe Bay. The successful assignments in this study are overwhelmingly concentrated on the eastern side of Oʻahu, primarily into Kāneʻohe Bay, accounting for 94% of the successfully assigned juveniles.

Kāneʻohe Bay, located on the northeast coast of Oʻahu, is a semi-enclosed estuarine system characterized by shallow patch reefs and an average depth of 10 m (Jokiel 1991). It is bounded by a barrier reef on the seaward (northeastern) side with two major channels out to the ocean. The bay has an extensive history of anthropogenic modifications including dredging, filling and increased sedimentation from runoff, all which which have severely altered the natural configuration, bathymetry, and even the currents that exist in the bay (Bahr *et al.* 2015).

In a partner study, Jerolmon (2016) modeled *manini* larval settlement along the eastern coast and within Kāneʻohe Bay. The overall patterns of simulated connectivity were dominated

by settlement and retention within Kāneʻohe Bay. Propagules that originated within the bay or entered the bay had a high chance of being retained. This pattern is consistent with the highest assignment of juveniles occurring within Kāneʻohe Bay.

The circulation patterns are highly variable between northern and southern reaches of Kāneʻohe Bay (Bathen 1968; Lowe *et al.* 2009), with water residence times ranging from <1 day on the outer reef to >1 month at the semi-enclosed southern part of the bay (Bathen 1968; Ostrander *et al.* 2008; Lowe *et al.* 2009). The northern half of the bay has a much more active circulation pattern with high levels of exchange between the bay and offshore waters. The southern part of the bay is characterized by reduced circulation due to flow restrictions which are absent in the northern part of the bay. Hence, the South Bay has been identified as a potential hotspot for retention and self-recruitment due to the high-water residence time (Lowe et al. 2009). Indeed, South Bay had the highest rates of parental assignments, accounting for > 50% of all the recovered offspring matches found in this study. Additionally, the highest rates of self-recruitment were in the South Bay. These patterns are consistent with the model predictions of Jerolmon (2016).

There are two instances of dispersal outside of the bay, both of which followed a northern trajectory towards Hauʻula and Lāʻie (HAU and LAIE in Fig. 4.1). These individuals originated at the northernmost collection site within Kāneʻohe Bay which is subjected to more oceanic conditions and where water residence time can be <1 day (Lowe *et al.* 2009). Unlike the southern part of the bay, physical processes in this northernmost part of the bay appear to reduce larval retention, a finding which is also consistent with Jerolmon (2016).

Fishery implications

Despite high assignment of juveniles within Kāneʻohe Bay, we observed that the size-class distribution of the manini population there appears to be skewed toward juvenile size-classes, which is reflected in the adult to juvenile sampling ratio (Table 4.1). This reflects trends from personal observations and conversations with local fishers, in which heavy fishing pressure on adult and subadult *manini* throughout the bay is likely to have reduced adult fish numbers and skewed size-class ratios within the population. Many Hawaiian coastal fishes, including *manini*, are harvested at rates that may not be sustainable (Smith 1993; Friedlander 2004; Friedlander *et al.* 2018). In the Main Hawaiian Islands, it is estimated that biomass of reef-associated catches is nearly 1.2 million kg per year. For *manini*, legal size limits are below the average size at maturity meaning females can be removed from the population before they can contribute offspring (Schemmel & Friedlander 2017).

One of the motivations for this study is to inform management efforts at the community level to ensure that subsistence fisheries for *manini* persist sustainably into the future. Inherent in this motivation is knowing which populations are currently at risk of overfishing. Although there were a few instances of self-recruitment inside Kāneʻohe Bay, the major source of recruitment originates outside the bay. This may be a result of heavy fishing pressure; however, an investigation of the abundance of adult *manini* inside the bay would need to be conducted to clarify these patterns. Nonetheless, in the context of fisheries and ensuring long-term sustainability, the population of *manini* inside Kāneʻohe Bay seems to be largely sourced by adjacent areas outside of the Bay. Accordingly, management efforts would need to ensure healthy population levels to the north and south of Kāneʻohe Bay, in order to accommodate heavy fishing pressure inside the bay. These findings provide a geographic scale at which both

communities and agencies may cooperate to target management efforts to promote sustainability in subsistence and recreational fisheries.

Patterns observed elsewhere for Manini

Manini are a ubiquitous feature of reefs from the East Pacific to the Western Indian Ocean, and several previous studies have provided genetic assessments of dispersal. Lessiosand Robertson (2006) reported very limited genetic connectivity on the scale of eastern versus central Pacific (mtDNA Φ_{ST} =0.355). Planes *et al.* (1996) reported a pattern of isolation by distance between proximal islands in French Polynesia. These authors concluded that most dispersal is between adjacent regions, and that long-distance dispersal is rare or sporadic. More directly pertinent to this study is the allozyme analysis of *manini* within the lagoon at New Caledonia, which revealed significant population structure (F_{ST}= 0.049) on a scale of a few hundred kms, unusual for a reef fish (Planes *et al.* 1998a). Planes *et al.* (1998b) reached a similar finding in Taiaro Lagoon in French Polynesia (F_{ST}= 0.055 between lagoon and ocean), concluding that *manini* could close their life cycle within the lagoon (an area of 6 km^2), which has no regular connection to the ocean. These results are in substantial agreement with our finding of limited larval dispersal on the scale of eastern and southern O‘ahu. Collectively, these studies reinforce the conclusion that a long pelagic larval duration does not invariably translate into extensive dispersal (Weersing & Toonen 2009; Selkoe & Toonen 2011).

CONCLUSIONS

Understanding the early life history, ecology and dynamics of *manini* is critical in projecting the success of adult populations and thus strategizing a method to ensure their

sustainability. We hope that by illuminating some of the pathways of dispersal and settlement around O‘ahu, we have provided one of the necessary components to properly inform conservation and management strategies. This study indicates that local recruitment is low along the northern and western shores of Oahu, which are subject to westerly currents, but is much higher in the protected waters of Kāne‘ohe Bay on the eastern side. Kāne‘ohe Bay acts as a sink for propagules from reefs as far as 60 km away. However, it is also a source of recruitment to regions along the eastern shore. Therefore, the community-based subsistence fishing areas (CBSFAs), a return to *moku* traditional Hawaiian management strategies, would be most effective on the eastern and southern coasts of Oahu. Indeed, given the scale of *manini* dispersal observed in previous studies, *moku* management seems a good fit to *manini* fisheries in general. The possible exceptions are the *manini* on the western and northern shores, where currents may disperse propagules beyond the coastal waters of Oahu.

Finally, we note that when demographic composition was skewed towards adults in our study sites, we found very low larval retention. When skewed towards juveniles, we found very high larval retention. This may be an artifact of fishing pressure, but also may provide a simple observational test that can indicate areas of productivity (in terms of high recruitment) without lethal sampling and expensive lab work.

Figure 4.1. Map of O‘ahu collection sites and dispersal pathways. Lines and arrows indicate the pathway of dispersal from parent to offspring. Dashed lines are a single dispersal event. Solid lines indicate dispersal paths shared by two or more larvae, and line thickness is proportional to the number of individuals that followed a given path. Red sqares indicated self recruitment events. Collection site codes are provided in Table 4.1.

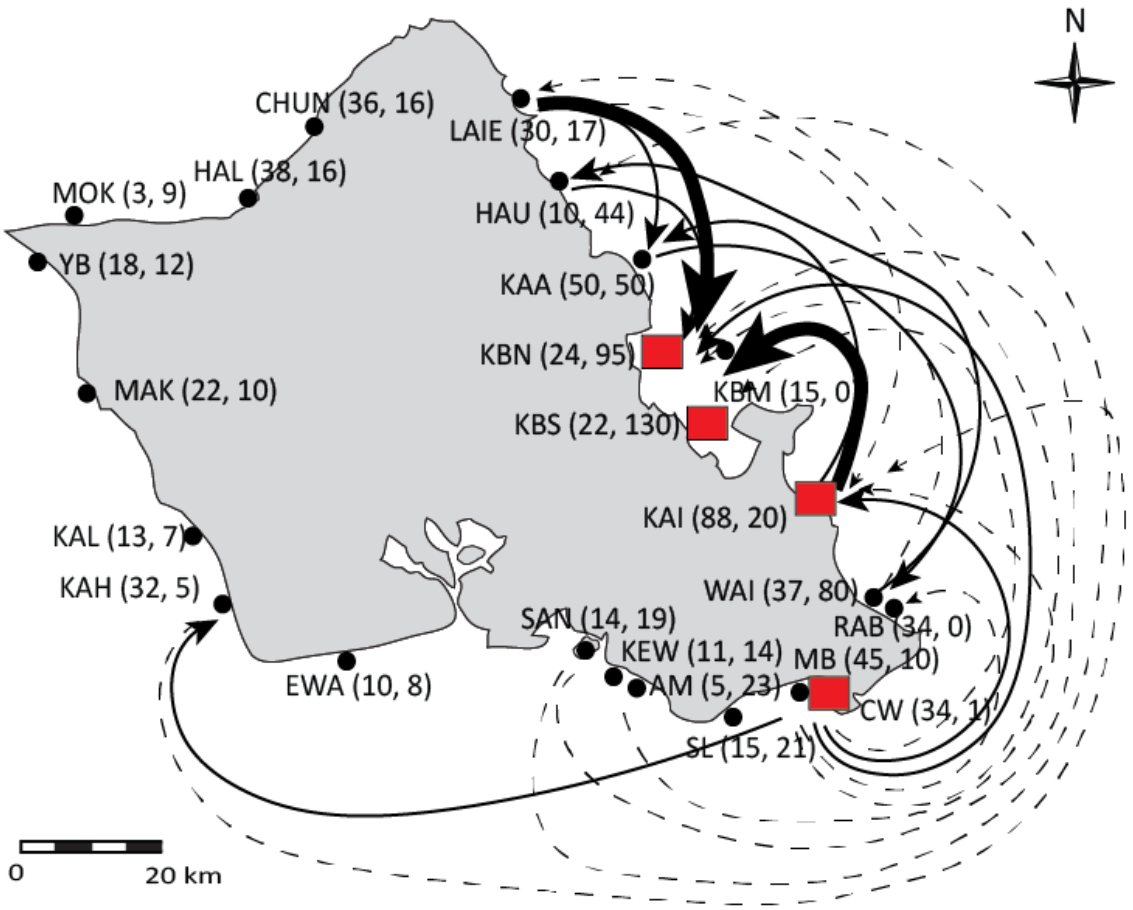

Figure 4.2. Map of Kāne‘ohe Bay colletion sites. Black dots indicate sampled reefs. Reefs IDs are based on nomenclature of Roy (1970). Only reefs where parents or offspring were recovered are listed, however dots denote the location where collections occurred. Lines and arrows indicate the pathway of dispersal from parent to offspring. Dashed lines indicated a single dispersal event.

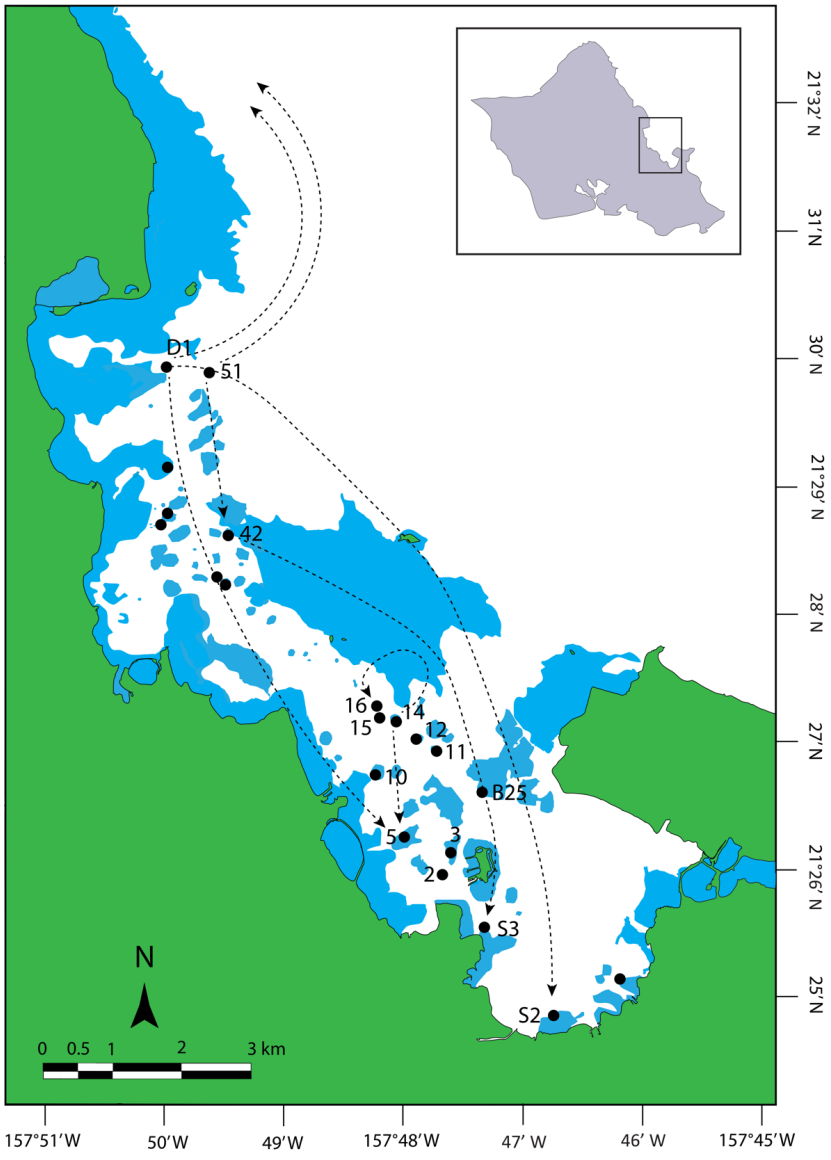

Table 4.1. Locations where *Acanthurus triostegus* was collected around Oʻahu, Hawaiʻi. Collection numbers are separated by adults and juveniles. The locations and number of assigned juveniles that were collected at each location is noted, as well as number of self-recruiting events.

Sampling Location	Site code	# of adults	# of juveniles	# assigned to location	# Self recruiting
North					
Mokulēʻia	MOK	3	9		
Haleʻiwa	HAL	38	16		
Chuns	CHUN	36	16		
East					
Lāʻie	LAIE	30	17	2	
Hauʻula	HAU	10	44	5	
Kaʻaʻawa	KAA	50	50	7	
Mouth of Kāneʻohe Bay	KBM	15	0		
Kāneʻohe Bay, North	KBN	24	95	1	1
Kāneʻohe Bay, South	KBS	22	130	35	4
Kailua	KAI	88	20	6	1
Waimānalo	WAI	37	80	2	
Rabbit Island	RAB	34	0		
South					
China Walls	CW	34	1		1
Maunalua Bay	MB	45	10		
Shangri La	SL	15	21		
Ala Moana	AM	5	23		
Kewalo	KEW	11	14		
Sand Island	SAN	14	19		
Ewa	EWA	10	8		
West					
Kahe	KAH	32	5	3	
Kalanianaole	KAL	13	7		
Mākaha	MAK	22	10		
Yokohama Bay	YB	18	12		
Total		606	607	61	7

Table 4.2. Pathways and distances of dispersal between collection sites. Abbreviations: KBS, Kāneʻohe Bay, South; KBN, Kāneʻohe Bay, North; * denotes self-recruitment, in addition to self-recruitment events in north and south Kåne'hoe Bay.

Location of parent-offspring pairs			
Parent	Offspring	# of occurrences	Dispersal Distance (km)
Lāʻie	Kaʻaʻawa	3	12.5
Lāʻie	KBN, Reef 42	1	23.0
Lāʻie	KBS Reef S2	1	30.0
Lāʻie	KBS, Reef B25	6	26.7
Lāʻie	KBS, Reef 11	2	25.6
Lāʻie	KBS, Reef 14	3	25.9
Lāʻie	KBS, Reef 15	1	25.9
Lāʻie	Kailua	1	35.2
Kaʻaʻawa	KBS, Reef 14	1	14.2
Kaʻaʻawa	Waimānalo	2	30.6
KBN, Reef D1	KBS Reef S2	1	10.2
KBN, Reef D1	KBS, Reef 5	1	6.8
KBN, Reef D1	Hauʻula	2	13.4
KBN, Reef 42	KBS, Reef S3	1	6.1
KBN, Reef 51	KBN, Reef 42	1	1.9
KBN, Reef 51	Lāʻie	1	23.0
KBS, Reef 14	KBS, Reef 5	1	1.4
KBS, Reef 16	KBS, Reef 14	1	0.2
Kailua	Kailua	1	*
Kailua	Kaʻaʻawa	3	23.2
Kailua	KBS	1	16.8
Kailua	KBS, Reef S3	1	17.9
Kailua	KBS, Reef 10	1	15.0
Kailua	KBS, Reef 14	4	14.0
Kailua	KBS, Reef 15	8	14.3
Waimānalo	KBS, Reef 15	1	23.0
Waimānalo	Hauʻula	2	35.6
Rabbit Island	Kaʻaʻawa	1	33.2
China Walls	Lāʻie	1	58.3
China Walls	KBS Reef 2	1	40.4
China Walls	KBS, Reef 5	1	38.9
China Walls	Kailua	2	25.6

China Walls	China Walls	1	*
Maunalua Bay	KBS, Reef 10	1	38.7
Maunalua Bay	Kailua	1	26.5
Maunalua Bay	Kahe	3	51.9
Ala Moana	Kailua	1	42.8
Sand Island	Hauʻula	1	70.9
Sand Island	KBS Reef 2	1	59.6
Kahe	Kailua	1	78.0

CHAPTER 5 CONCLUSIONS

The overall aim of my dissertation was to use molecular methods to gain a better understanding of connectivity in reef fishes. By assessing connectivity across wide-ranges to fine scales I aimed to contribute to our growing knowledge regarding the mechanisms that influence dispersal and connectivity in coral reef fishes and how these patterns promote biodiversity in both an evolutionary and a contemporary framework. In utilizing phylogeographic, traditional population genetic, and parentage analyses for my research, I was able to bring further attention to the insight that can be obtained by conducting theses analyses.

Insights from range-wide phylogeographic analyses

By assessing phylogeographic patterns across an entire species range, many studies have recovered highly diverged populations and cryptic diversity (Gaither *et al.* 2011a; Szabo *et al.* 2014; Fernandez-Silva *et al.* 2015; Dudoit *et al.* 2018). My research was set apart from other studies due to evaluating genetic patterns for a member of a monotypic genus. The regal angelfish, *Pygoplites diacanthus*, is the sole species in the genus *Pygoplites*. Phylogenetic analyses of the family Pomacanthidae (marine angelfishes), which is comprised of seven genera, estimated the age of *Pygoplites* to be 10 million years old (Alva-Campbell *et al.* 2010). The ability for a species to persist for millions of years without the pressure to diversify initiated the question as to the evolutionary or ecological traits that promoted species cohesion across time.

A range-wide genetic evaluation of *P. diacanthus* revealed cryptic diversity that was previously unknown. First, diverged populations were recovered that indicated isolation between the Red Sea, the Indian Ocean, and the Pacific Ocean. These genetic breaks coincided with

previously identified barriers to gene flow that have been observed across a variety of taxa and are recognized as major biogeographic barriers (Rocha *et al.* 2007; Toonen *et al.* 2016). Second, we endorsed sub-species status for the Indian Ocean and Red Sea populations based on shallow diagnostic distinctions based on genetics and morphology, raising the name to *P. d. flavescens,* while the Pacific Ocean population retained the name *P. d. diacanthus*. Furthermore, we provided evidence indicating hybridization between the different sub-species at the Christmas Island which was eventually recognized as a hybridization hotspot (Hobbs & Allen 2014).

Although my results corroborated previously identified patterns, new insights were also recovered. The phylogenetic progression showed the Red Sea population diverging from the Pacific Ocean and then two dispersal events from the Red Sea into the Indian Ocean leading to separate lineages. The two cryptic Indian Ocean lineages do not show evidence of admixture and it is unclear what mechanisms are facilitating the isolation between the lineages. A thorough study of the ecology and reproductive strategies within these lineages would need to be conduct to identify the mechanism promoting isolation despite occupying sympatric ranges. Further evidence of isolated was observed within the Pacific Ocean lineage where the French Polynesian population (Moorea) was found to be genetically differentiated from all other Pacific populations. An analysis of the circulation pattern in the region indicated that the westward flow of the Southern Equatorial Current and eddies created in the wake of Tahiti contributed to the formation of a strong counterclockwise flow around the island promoting the local retention of larvae while inhibiting emigration from other islands.

Range-wide phylogeographic studies contribute to our understanding into the mechanisms that lead to divergence within species. An insight that may otherwise not be apparent if sampling was limited to a smaller part of the species range. These patterns may

corroborate previously findings, as well as identify patterns that have not be documented elsewhere leading to further analyses of connectivity within these regions and identifying the processes promoting and inhibiting diversification.

Archipelago-wide connectivity and the efficacy of genomic sequencing

An archipelago-wide genetic assessment provided insight into how the use of genetic markers can lead to interpreting patterns of gene flow in a historical or contemporary framework. Assessing connectivity patterns of *Acanthurus triostegus* (*manini*) and *Ctenochaetus strigosus* (*kole*) provides the first archipelago wide population genetics analyses based on genomic sequencing. Previous studies used a targeted loci approach based on single *mt*DNA markers or a combination of *mt*DNA and selected *n*DNA or microsatellites. This research also provides the first report of island-by-island isolation for a reef fish (*kole*) across the archipelago, which counters connectivity patterns based on mtDNA (Eble *et al.* 2009). Using an mtDNA marker, *kole* was shown to have genetic homogeneity across most of the archipelago with an isolated population between Maro Reef and Pearl and Hermes Atoll in the northwest end of the archipelago. By conducting a genomics based genetic assessment, which utilizes hundreds to thousands of loci, we showed that population structure was much more fine scale. The results of our genomic-based connectivity study highlights the efficacy of genomic sequencing to characterize contemporary connectivity.

A comparison between connectivity patterns of *kole* and *manini* also provided insight into the dispersal potential for presumably low and high dispersing species. *Kole* is a Hawaiian endemic and it is hypothesized that Hawaiian endemics are descendants of low dispersers who colonized Hawai‘i but were unable to maintain sufficient connectivity with the rest of the Pacific

population. This hypothesis is supported by the high levels of population structure exhibited by Hawaiian endemic species when compared to species that are wide ranging who often show no structure across most of the archipelago. *Manini* has a Indo-Pacific distribution and the sole study investigating connectivity of *manini* within Hawai‘i showed structure between Hawai‘i Island and O‘ahu (Planes & Fauvelot 2002). However, connectivity studies in *manini* have shown that they have enigmatic dispersal potential. An assessement across a large portion of their range showed connectivity across thousands of kilometers (Planes & Fauvelot 2002); however, structure was observed between neighboring islands (Planes *et al.* 1998b), and even within a lagoon (Planes *et al.* 1998a). In the Hawaiian Archipelago manini was found to exhibit a combination of these traits where *manini* showed population structure between each of the Main Hawaiian Islands (MHI), but showed genetic homogeneity across most of the remaing 2/3 of the archipelago in the Northwestern Hawaiian Islands (NWHI) and Johnston Atoll.

By assessing connectivity across the entire archipelago I was able to provide insight into the different patterns that can be recovered depending on the scope of connectivity being assessed, as well as how endemic and wide-ranging species can differ in dispersal potential. The different patterns based on genomic sequencing compared to mtDNA suggest genomic sequencing is a much powerful tool to assess contemporary levels of connectivity and suggests it would be worthwhile to revist to past connectivity studies which used targeted loci and reassess them in a genomics framework. Comparing dispersal potential between endemic and wide-ranging species showed that endemic species are much more limited in their disperal ability. These findings provide insight into our understanding of how endemic species evolve in isolation.

Fine scale connectivity

Assessing connectivity of *manini* across O‘ahu is the first study to evaluate connectivity using a parentage analysis at the island scale for a reef fish, and provides valuable insight to our understanding of dispersal as well as assisting to inform management strategies. The signal for genetic differentation is not fine scale enough for a traditional population genetics approach; however a parentage analysis provides a means to characterize dispersal and connectivity at the island scale. Although, genetic connectiity of *manini* was found to be structured island-by-island in the MHI, the parentage analysis shows that even within an island connectivity is not homogenous between different regions of the island.

An island scale connectivity assessment resulted in identifying several patterns. First, dispersal in *manini* is not consistent in terms of direction and distance. We assigned juveniles to less than 250 meters from their presumed spawning location, up to as far away as 78 km on the opposite end of the island. The direction of dispersal coincided with known ciruclation patterns, however, several instances were discovered that counterd the direction of known physical processes suggesting unknown biological, ecological, and phyiscal processes have a greater influence on dispersal and settlement.

Second, despite a high diseprsal potential, *manini* larvae do not travel far from their natal reef. This is a trait common among many reef fishes (Jones 2015) and may explain why each island in the MHI are isolated from one another. The windward side of O‘ahu in particular is a region of local retention and the final destination for many *manini* larvae, particulalry within Kāne‘ohe Bay. Within Kāne‘ohe Bay local retention is attributed to phyiscal processes within the bay which inhibit dispersal out of the bay. However, while some of the patterns of dispersal into

the bay coincide with prediction made by biophyiscal models (Lai et al., unpublished), other patterns are unclear and require further investigation.

Finally, the identifying dispersal patterns are directly relevant to establishing a competent community-based management strategy. The motivations for section of my research was to inform management efforts of at the community level to ensure adequate Kāneʻohe Bay and east coast populations of *manini* persist for future generations. We identified that the population of *manini* inside Kāneʻohe Bay seems to be largely sourced by populations outsides of the bay. Based on these findings, I suggest that management efforts would need to coordinate with communities all around Oʻahu to ensure they retain healthy population levels which in turn would seed into Kāneʻohe Bay.

WORK CITED

Abesamis RA, Saenz-Agudelo P, Berumen ML*, et al.* (2017) Reef-fish larval dispersal patterns validate no-take marine reserve network connectivity that links human communities. *Coral Reefs* **36**, 791-801.

Ahti P, Coleman RR, DiBattista J*, et al.* (2016) Phylogeography of Indo-Pacific reef fishes: Sister wrasses, *Coris gaimard* and *C. cuvieri,* in the Red Sea, Indian Ocean, and Pacific Ocean. *Journal of Biogeography* ***Online early***.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Review Genetics* **October 2011**, 697-709.

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

Almany GR, Hamilton RJ, Bode M*, et al.* (2013) Dispersal of grouper larvae drives local resource sharing in a coral reef fishery. *Current biology* **23**, 626-630.

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

Amos B, Hoelzel AR (1991) Long term preservation of whale skin for DNA analysis. *Rept. Intl. Whaling Commission* **Special Issue 13**, 99-103.

Andrews KR, Adams JR, Cassirer EF*, et al.* (2018) A bioinformatic pipeline for identifying informative SNP panels for parentage assignment from RADseq data. *Molecular Ecology Resources* **18**, 1263-1281.

Andrews KR, Karczmarski L, Au WW*, et al.* (2010) Rolling stones and stable homes: social structure, habitat diversity and population genetics of the Hawaiian spinner dolphin (*Stenella longirostris*). *Molecular Ecology* **19**, 732-748.

Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. *Molecular Ecology* **23**, 1661-1667.

Andrews KR, Moriwake VN, Wilcox C*, et al.* (2014) Phylogeographic Analyses of Submesophotic Snappers *Etelis coruscans* and *Etelis "marshi"* (Family Lutjanidae)

Reveal Concordant Genetic Structure across the Hawaiian Archipelago. *PLOS One* **9**, e91665.

Andrews KR, Williams A, Fernandez-Silva I*, et al.* (2016) Phylogeny of deepwater snappers (Genus *Etelis*) reveals a cryptic species pair in the Indo-Pacific and Pleistocene invasion of the Atlantic. *Molecular Phylogenetics & Evolution* ***(Accepted)***.

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Bahr KD, Jokiel PL, Toonen RJ (2015) The unnatural history of Kāneʻohe Bay: coral reef resilience in the face of centuries of anthropogenic impacts. *PeerJ* **3**, e950.

Bailey G (2009) The Red Sea, coastal landscapes, and hominin dispersals. In: *The Evolution of Human Populations in Arabia* (eds. Petraglia MD, Rose J), pp. 15-37. Springer, Dordrecht, Netherlands.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16**, 37-48.

Bathen KH (1968) A descriptive study of the physical oceanography of Kaneohe Bay, Oahu, Hawaii. Report no. 14. Hawaii Institute of Mainre Biology, Honolulu, HI.

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

Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. *Annals of statistics* **29**, 1165-1188.

Bernardi G, Beldade R, Holbrook SJ, Schmitt RJ (2012) Full-sibs in cohorts of newly settled coral reef fishes. *PLOS One* **7**, e44953.

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

Berumen ML, Almany GR, Planes S*, et al.* (2012) Persistence of self-recruitment and patterns of larval connectivity in a marine protected area network. *Ecology and Evolution* **2**, 444-452.

Bird CE, B.S. Holland, B.W. Bowen, R.J. Toonen. (2011) Diversification of endemic sympatric limpets (Cellana spp.) in the Hawaiian Archipelago. *Molecular Ecology* **20**, 2128-2141.

Blum SD (1989) Biogeography of the Chaetodontidae: an analysis of allopatry among closely related species. *Environmental Biology of Fishes* **25**, 9-31.

Bonjean F, Lagerloef GSE (2002) Diagnostic model and analysis of the surface currents in the tropical Pacific Ocean. *Journal of Physical Oceanography* **32**, 2938-2954.

Bowen B, Gaither MR, DiBattista JD*, et al.* (2016) Comparative phylogeography of the ocean planet. *Proceedings of the National Academy of Sciences* ***(Accepted)***.

Bowen BW (2016) The three domains of conservation genetics: Case histories from Hawaiian waters. *Journal of Heredity* **107**, 309-317.

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

Briggs JC, Bowen BW (2012) A realignment of marine biogeographic provinces with particular reference to fish distributions. *Journal of Biogeography* **39**, 12-30.

Briggs JC, Bowen BW (2013) Marine shelf habitat: Biogeography and evolution. *Journal of Biogeography* **40**, 1023-1035.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**, 3124-3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, genomes, genetics* **1**, 171-182.

Chong Z, Ruan J, Wu C-I (2012) Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics* **28**, 2732-2737.

Chow S, Hazama K (1998) Universal PCR primers for S7 ribosomal protein gene introns in fish. *Molecular Ecology* **7**, 1247-1263.

Christie M (2010) Parentage in natural populations: novel methods to detect parent-offspring pairs in large data sets. *Molecular Ecology Resources* **10**, 115-128.

Christie MR, Meirmans PG, Gaggiotti OE, Toonen RJ, White C (2017) Disentangling the relative merits and disadvantages of parentage analysis and assignment tests for inferring population connectivity. *ICES Journal of Marine Science* **74**, 1749-1762.

Christie MR, Tissot BN, Albins MA*, et al.* (2010) Larval connectivity in an effective network of marine protected areas. *PLOS One* **5**, e15715.

Clark PU, Dyke AS, Shakun JD*, et al.* (2009) The last glacial maximum. *Science* **325**, 710-714.

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

Coleman RR, Gaither MR, Kimokeo B*, et al.* (2014) Large-scale introduction of the Indo-Pacific damselfish *Abudefduf viagiensis* into Hawai'i promotes genetic swamping of the endemic congener *A. abdominalis*. *Molecular Ecology* **23**, 5552–5565.

Cowman PF, Bellwood DR (2013) The historical biogeography of coral reef fishes: global patterns of origination and dispersal. *Journal of Biogeography* **40**, 209-224.

Cox CB, Moore PD (2000) *Biogeography: an ecological and evolutionary approach, 6th edn.* Blackwell Science, Oxford.

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

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

Cutler DJ, Jensen JD (2010) To Pool, or Not to Pool? *Genetics* **186**, 41-43.

D'aloia C, Bogdanowicz S, Majoris J, Harrison R, Buston P (2013) Self-recruitment in a Caribbean reef fish: a method for approximating dispersal kernels accounting for seascape. *Molecular Ecology* **22**, 2563-2572.

D'Aloia CC, Bogdanowicz SM, Francis RK*, et al.* (2015) Patterns, causes, and consequences of marine larval dispersal. *Proceedings of the National Academy of Sciences* **112**, 13940-13945.

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

Dennenmoser S, Vamosi SM, Nolte AW, Rogers SM (2017) Adaptive genomic divergence under high gene flow between freshwater and brackish-water ecotypes of prickly sculpin (Cottus asper) revealed by Pool-Seq. *Molecular Ecology* **26**, 25-42.

DiBattista JD, Berumen ML, Gaither MR*, et al.* (2013) After continents divide: comparative phylogeography of reef fishes from the Red Sea and Indian Ocean. *Journal of Biogeography* **40**, 1170-1181.

DiBattista JD, Choat JH, Gaither MR*, et al.* (2015) On the origin of endemic species in the Red Sea. *Journal of Biogeography* **43**, 13-30.

DiBattista JD, Roberts M, Bouwmeester J*, et al.* (2016) A review of contemporary patterns of endemism in the Red Sea. *Journal of Biogeography* **43**, 423-439.

DiBattista JD, Waldrop E, Bowen BW*, et al.* (2012) Twisted sister species of pygmy angelfishes: discordance between taxonomy, coloration, and phylogenetics. *Coral Reefs* **31**, 839-851.

DiBattista JD, Wilcox C, Craig MT, Rocha LA, Bowen BW (2011) Phylogeography of the Pacific Blueline Surgeonfish, *Acanthurus nigroris*, Reveals High Genetic Connectivity and a Cryptic Endemic Species in the Hawaiian Archipelago. *Journal of Marine Biology* **2011**, 17.

Dixson DL, Jones GP, Munday PL*, et al.* (2008) Coral reef fish smell leaves to find island homes. *Proceedings of the Royal Society B: Biological Sciences* **275**, 2831-2839.

Doherty PJ, Planes S, Mather P (1995) Gene Flow and Larval Duration in Seven Species of Fish from the Great Barrier Reef. *Ecology* **76**, 2373-2391.

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

Drew JA, Allen GR, Erdmann MV (2010) Congruence between mitochondrial genes and color morphs in a coral reef fish: population variability in the Indo-Pacific damselfish Chrysiptera rex (Snyder, 1909). *Coral Reefs* **29**, 439-444.

Dudoit Aa, Iacchei M, Coleman RR*, et al.* (2018) The little shrimp that could: phylogeography of the circumtropical *Stenopus hispidus* (Crustacea: Decapoda), reveals divergent Atlantic and Pacific lineages. *PeerJ* **6**, e4409.

Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. . *Conservation Genetics Resources* **4**, 359-361.

Eble J, Toonen R, Bowen B (2009) Endemism and dispersal: comparative phylogeography of three surgeonfishes across the Hawaiian Archipelago. *Marine Biology* **156**, 689-698.

Eble JA, Bowen BW, Bernardi G (2015) Phylogeography of coral reef fishes. In: *Ecology of fishes on coral reefs* (ed. Mora C), pp. 64-75. University of Hawaii Press, Honolulu, HI.

Eble JA, Rocha LA, Craig MT, Bowen BW (2011a) Not all larvae stay close to home: Insights into marine population connectivity with a focus on the brown surgeonfish (*Acanthurus nigrofuscus*). *Journal of Marine Biology* **2011**, 12.

Eble JA, Toonen Robert J., Sorenson L*, et al.* (2011b) Escaping paradise: larval export from Hawaii in an Indo-Pacific reef fish, the yellow tang Zebrasoma flavescens. *Marine Ecology Progress Series* **428**, 245-258.

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

Fernandez Silva I, Randall JE, DiBattista J*, et al.* (2015) Phylogeography of the Indo-Pacific goatfish *Mulloidichthys flavolineatus*: Isolation in peripheral biogeographic provinces and a cryptic evolutionary lineage in the Red Sea. *Journal of biogeography. In press*.

Fernandez-Silva I, Randall JE, Coleman RR*, et al.* (2015) Yellow tails in the Red Sea: phylogeography of the Indo-Pacific goatfish *Mulloidichthys flavolineatus* reveals isolation in peripheral provinces and cryptic evolutionary lineages. *Journal of Biogeography* **42**, 2402-2413.

Firing E (1996) Currents observed north of Oahu during the first five years of HOT. *Deep Sea Research Part II: Topical Studies in Oceanography* **43**, 281-303.

Fischer MC, Rellstab C, Leuzinger M*, et al.* (2017) Estimating genomic diversity and population differentiation – an empirical comparison of microsatellite and SNP variation in Arabidopsis halleri. *BMC Genomics* **18**, 69.

Flanagan SP, Jones AG (2018) The future of parentage analysis: from microsatellites to SNPs and beyond. *Molecular Ecology*.

Fletcher CH, Bochicchio C, Conger CL*, et al.* (2008) Geology of Hawaii reefs. In: *Coral Reefs of the USA* (eds. Riegl B, Dodge RE), pp. 435-487. Springer.

Frankham R, Briscoe DA, Ballou JD (2002) *Introduction to Conservation Genetics* Cambridge University Press, Cambridge.

Friedlander AM (2004) Status of Hawaii's coastal fisheries in the new millennium.

Friedlander AM, Donovan MK, Stamoulis KA*, et al.* (2018) Human-induced gradients of reef fish declines in the Hawaiian Archipelago viewed through the lens of traditional management boundaries. *Aquatic Conservation: Marine and Freshwater Ecosystems* **28**, 146-157.

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

Futschik A, Schlötterer C (2010) The Next Generation of Molecular Markers From Massively Parallel Sequencing of Pooled DNA Samples. *Genetics* **186**, 207-218.

Gaither MR, Bernal MA, Coleman RR*, et al.* (2015) Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Molecular Ecology* **24**, 1543-1557.

Gaither MR, Bowen BW, Bordenave T-R*, et al.* (2011a) Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the indo-pacific barrier with contemporary overlap in the coral triangle. *BMC Evolutionary Biology* **11**, 189-204.

Gaither MR, Bowen BW, Toonen RJ*, et al.* (2010a) Genetic consequences of introducing allopatric lineages of Bluestriped Snapper (*Lutjanus kasmira*) to Hawaii. *Molecular Ecology* **19**, 1107-1121.

Gaither MR, Jones SA, Kelley C*, et al.* (2011b) High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian Archipelago. *PLOS One* **6**, e28913.

Gaither MR, Rocha LA (2013) Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. *Journal of Biogeography* **40**, 1638-1648.

Gaither MR, Schultz JK, Bellwood DR*, et al.* (2014) Evolution of pygmy angelfishes: Recent divergences, introgression, and the usefulness of color in taxonomy. *Molecular Phylogenetics & Evolution* **74**, 38-47.

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

Galindo HM, Pfeifer-Herbert AS, A MM*, et al.* (2010) Seascape genetics along a steep cline: using genetic patterns to test predictions of marine larval dispersal. *Molecular Ecology* **19**, 3692-3707.

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907*.

Germer S, Holland MJ, Higuchi R (2000) High-throughput SNP allele-frequency determination in pooled DNA samples by kinetic PCR. *Genome research* **10**, 258-266.

Grigg RW, Wells JW, Wallace C (1981) *Acropora* in Hawaii. Part 1. History of the Scientific Record, Systematics, and Ecology - Part I. *Pacific Science* **35**, 1-13.

Guindon S, Dufayard JF, Lefort V*, et al.* (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* **59**, 307-321.

Guindon S, Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* **5**, 696-704.

Guo B, Li Z, Merilä J (2016) Population genomic evidence for adaptive differentiation in the Baltic Sea herring. *Molecular Ecology* **25**, 2833-2852.

Harrison HB, Williamson DH, Evans RD*, et al.* (2012) Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Current biology* **22**, 1023-1028.

Hassett R, Boehlert G (1999) Spatial and temporal distributions of copepods to leeward and windward of Oahu, Hawaiian Archipelago. *Marine Biology* **134**, 571-584.

Hauser L, Baird M, Hilborn R, Seeb LW, Seeb JE (2011) An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. *Molecular Ecology Resources* **11**, 150-161.

Hellberg M (2009) Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution, and Systematics* **40**, 291-210.

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

Hinton S (1962) Longevity of Fishes in Captivity as of September 1956. *Zoologica* **47**, 105-116.

Hobbs J-PA, Allen GR (2014) Hybridisation among coral reef fishes at Christmas Island and the Cocos (Keeling) Islands. *Raffles Bulletin of Zoology, Supplement* **30**.

Hodge JR, Bellwood DR (2016) The geography of speciation in coral reef fishes: the relative importance of biogeographical barriers in separating sister species. *Journal of Biogeography. (Online early)*.

Hohenlohe PA (2014) Ecological genomics in full colour. *Molecular Ecology* **23**, 5129-5131.

Hohenlohe PA, Bassham S, Etter PD*, et al.* (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genetics* **6**, e1000862.

Hourigan TF, Reese ES (1987) Mid-ocean isolation and the evolution of Hawaiian reef fishes. *Trends in Ecology & Evolution* **2**, 187-191.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science* **294**, 2310-2314.

Iacchei M, O'Malley JM, Toonen RJ (2014) After the gold rush: population structure of spiny lobsters in Hawaii following a fishery closure and the implications for contemporary spatial management. *Bulletin of Marine Science* **90**, 331-357.

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

Jerolmon C (2016) *A Physical-Biological Model of Oahu's Windward Coastal Waters: Implications for Larval Transport* Masters, University of Hawaii at Manoa.

Johns GC, Avise JC (1998) A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome *b* gene. *Molecular Biology and Evolution* **15**, 1481-1490.

Johnson DW, Christie MR, Pusack TJ, Stallings CD, Hixon MA (2018) Integrating larval connectivity with local demography reveals regional dynamics of a marine metapopulation. *Ecology* **99**, 1419-1429.

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

Jones GP (2015) Mission impossible: unlocking the secrets of coral reef fish dispersal. In: *Ecology of fishes on coral reefs* (ed. Mora C), pp. 16-28. Cambridge University Press, Cambridge, UK.

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**, 1099-1106.

Keller I, Wagner C, Greuter L*, et al.* (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology* **22**, 2848-2863.

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

Kobayashi DR (2006) Colonization of the Hawaiian Archipelago via Johnston Atoll: a characterization of oceanographic transport corridors for pelagic larvae using computer simulation. *Coral Reefs* **25**, 407-417.

Kofler R, Orozco-terWengel P, De Maio N*, et al.* (2011a) PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLOS One* **6**, e15925.

Kofler R, Pandey RV, Schlötterer C (2011b) PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**, 3435-3436.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* **15**, 1179-1191.

Kosaki RK, Pyle RL, Leonard JC*, et al.* (2016) 100% endemism in mesophotic reef fish assemblages at Kure Atoll, Hawaiian Islands. *Marine Biodiversity* **47**, 1-2 doi: 10.1007/s12526-12016-10510-12525.

Kozak GM, Brennan RS, Berdan EL, Fuller RC, Whitehead A (2014) Functional and populaiton genomic divergence within and between tow species of killifish adapted to different osmotic niches. *Evolution* **68**, 63-80.

Kulbicki M, Parravicini V, Bellwood DR*, et al.* (2013) Global biogeography of reef fishes: a hierarchical quantitative delineation of regions. *PLOS One* **8:e81847**.

Leichter JJ, Aildredge AL, Bernardi G*, et al.* (2013) Biological and physical interactions on a tropical island coral reef transport and retention processes on Moorea, French Polynesia. *Oceanography* **26**, 52-63.

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

Leis JM, McCormick MI (2002) The Biology, Behavior; and Ecology of the Pelagic, Larval Stage of Coral Ree Fishes. In: *Coral reef fishes: dynamics and diversity in a complex ecosystem* (ed. Sale PF), pp. 171-199. Academic Press, San Diego.

Lemer S, Planes S (2014) Effects of habitat fragmentation on the genetic structure and connectivity of the black-lipped pearl oyster, *Pinctada margaritifera,* populations in French Polynesia. *Marine Biology* **161**, 2035-2049.

Leray M, Beldade R, Holbrook SJ*, et al.* (2010) Allopatric divergence and speciation in coral reef fish: The three-spot dascyllus, *Dascyllus trimaculatus*, species complex. *Evolution* **64**, 1218-1230.

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

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

Lo-Yat A, Meekan MG, Carleton JH, Galzin R (2006) Large-scale dispersal of the larvae of nearshore and pelagic fishes in the tropical oceanic waters of French Polynesia. *Marine Ecology Progress Series* **325**, 195-203.

Longenecker K, Langston R (2008) Life history compendium of exploited Hawaiian fishes. Fisheries Local Action Strategy Division of Aquatic Resources Final Report, Honolulu, HI.

Longenecker K, Langston R, Eble J (2008) Growth, mortality and reproduction of manini, *Acanthurus triostegus sandvicensis*. Fisheries Local Action Strategy Division of Aquatic Resources Final Report.

Lovejoy N (1999) *Systematics, Biogeography and Evolution of Needlefishes (Teleostei: Belonidae)* Cornell University, Ithaca, NY.

Lowe RJ, Falter JL, Monismith SG, Atkinson MJ (2009) A numerical study of circulation in a coastal reef-lagoon system. *Journal of Geophysical Research: Oceans* **114**.

Ludt WB, Rocha LA (2014) Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, 25-38.

Lumpkin CF (1998) *Eddies and currents of the Hawaiian Islands* PhD, University of Hawaii at Manoa.

Marshall TC, Slate J, Kruuk LE, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Mol Ecol* **7**, 639-655.

McCoy KS, Williams ID, Friedlander AM*, et al.* (2018) Estimating nearshore coral reef-associated fisheries production from the main Hawaiian Islands. *PLOS One* **13**, e0195840.

McMillan WO, Weigt LA, Palumbi SR (1999) Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). *Evolution*, 247-260.

Meeker ND, Hutchinson SA, Ho L, Trede NS (2007) Method for isolation of PCR-ready genomic DNA from zebrafish tissues. *Biotechniques.* **43**, 610-614.

Meier JI, Marques DA, Mwaiko S*, et al.* (2017) Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications* **8**, 14363.

Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes* **4**, 792-794.

Mirams A, Treml E, Shields J, Liggins L, Riginos C (2011) Vicariance and dispersal across an intermittent barrier: population genetic structure of marine animals across the Torres Strait land bridge. *Coral Reefs* **30**, 937-949.

Narum SR (2006) Beyond Bonferroni: less conservative analyses for conservation genetics. *Conservation Genetics* . **7**, 783-787.

Ostrander CE, McManus MA, DeCarlo EH, MacKenzie FT (2008) Temporal and spatial variability of freshwater plumes in a semi-enclosed estuarine-bay system. *Estuaries and coasts* **31**, 192-203.

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. *Ecological Applications* **13**, 146-158.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for *De Novo* SNP Discovery and Genotyping in Model and Non-Model Species. *PLOS One* **7**, e37135.

Pianka ER (1978) *Evolutionary ecology* Harper and Row, New York, USA.

Planes S (1993) Genetic differentiation in relation to restricted larval dispersal of the convict surgeonfish, *Acanthurus triostegus,* in French Polynesia. *Marine Ecology Progress Series* **98**, 237-246.

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

Planes S, Galzin R, Bonhomme F (1996) A genetic metapopulation model for reef fishes in oceanic islands: the case of the surgeonfish, *Acanthurus triostegus*. *Journal of Evolutionary Biology* **9**, 103-117.

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

Planes S, Parroni M, Chauvet C (1998a) Evidence of limited gene flow in three species of coral reef fishes in the lagoon of New Caledonia. *Marine Biology* **130**, 361-368.

Planes S, Romans P, Lecomte-Finiger R (1998b) Genetic evidence of closed life-cycles for some coral reef fishes within Taiaro Lagoon (Tuamotu Archipelago, French Polynesia). *Coral Reefs* **17**, 9-14.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Puglise KA, Hinderstein LM, Marr JCA, Dowgiallo MJ, Martinez FA (2009) *Mesophotic coral ecosystems research strategy: International workshop to prioritize research and management needs for mesophotic coral ecosystems, Jupiter, Florida, 12-15 July, 2009* US Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service.

Puritz JB, Hollenbeck CM, Gold JR (2014a) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ* **2**, e431.

Puritz JB, Matz MV, Toonen RJ*, et al.* (2014b) Demystifying the RAD fad. *Molecular Ecology* **23**, 5937-5942.

Ramon M, Nelson P, Martini E, Walsh W, Bernardi G (2008) Phylogeography, historical demography, and the role of post-settlement ecology in two Hawaiian damselfish species. *Marine Biology* **153**, 1207-1217.

Randall JE (1961) Contribution to the Biology of the Convict Surgeonfish of the Hawaiian Islands, *Acanthurus triostegus sandvicensis*. *Pacific Science* **15**, 215-272.

Randall JE (1994) Twenty-two new records of fishes from the Red Sea. *Fauna of Saudi Arabia* **14**, 259-275.

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

Randall JE (2005a) *Reef and shore fishes of the South Pacific* University of Hawai'i Press, Honolulu, HI.

Randall JE (2005b) *Reef and Shore Fishes of the South Pacific: New Caledonia to Tahiti and the Pitcairn Islands* University of Hawaii Press, Honolulu, HI.

Randall JE (2007) *Reef and shore fishes of the Hawaiian Islands* University of Hawai'i Sea Grant College Program, Honolulu, HI.

Randall JE (2010) *Shore Fishes of Hawaii* University of Hawaii Press, Honolulu.

Reece JS, Bowen BW, Joshi K, Goz V, Larson A (2010a) Phylogeography of two moray eels indicates high dispersal throughout the Indo-Pacific. *Journal of Heredity* **101**, 391-402.

Reece JS, Bowen BW, Smith DG, Larson AF (2010b) Molecular phylogenetics of moray eels (Murenidae) demonstrates multiple origins of a shell-crushing jaw (Gymnomuraena, Echidna) and multiple colonizations of the Atlantic Ocean. *Molecular Phylogenetics & Evolution* **57**, 829 – 835.

Rivera MAJ, Kelley CD, Roderick GK (2004) Subtle population genetic structure in the Hawaiian grouper, *Epinephelus quernus* (Serranidae) as revealed by mitochondrial DNA analyses. *Biological Journal of the Linnean Society* **81**, 449-468.

Robertson DR, Karg F, Leao de Moura R, Victor BC, Bernardi G (2006) Mechanisms of speciation and faunal enrichment in Atlantic parrotfishes. *Molecular Phylogenetics & Evolution* **40**, 795-807.

Rocha LA, Bass AL, Robertson DR, Bowen BW (2002) Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology* **11**, 243-251.

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

Rocha LA, Lindeman KC, Rocha CR, Lessios HA (2008) Historical biogeography and speciation in the reef fish genus *Haemulon* (Teleostei: Haemulidae). *Molecular Phylogenetics & Evolution* **48**, 918-928.

Ronquist F (2004) Bayesian inference of character evolution. *Trends in Ecology & Evolution* **19**, 475-481.

Roy KJ (1970) Change in Bathymetric Configuration, Kaneohe Bay, Oahu 1882-1969, Hawaii Institute of Geophysics, University of Hawaii, Report 70-15: 26 pp.

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

Saenz-Agudelo P, Jones GP, Thorrold SR, Planes S (2012) Patterns and persistence of larval retention and connectivity in a marine fish metapopulation. *Molecular Ecology* **21**, 4695-4705.

Schemmel EM, Friedlander AM (2017) Participatory fishery monitoring is successful for understanding the reproductive biology needed for local fisheries management. *Environmental Biology of Fishes* **100**, 171-185.

Schultz JK, Pyle RL, DeMartini E, Bowen BW (2006) Genetic connectivity among color morphs and Pacific archipelagos for the flame angelfish, *Centropyge loriculus*. *Marine Biology* **151**, 167-175.

Selkoe K, Toonen RJ (2011) Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. *Marine Ecology Progress Series* **436**, 291-305.

Selkoe KA, Aloia CC, Crandall ED*, et al.* (2016a) A decade of seascape genetics: contributions to basic and applied marine connectivity. *Marine Ecology Progress Series* **554**, 1-19.

Selkoe KA, Gaggiotti OE, Andrews KR*, et al.* (2014) Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology* **23**, 3064-3079.

Selkoe KA, Gaggiotti OE, Treml EA*, et al.* (2016b) The DNA of coral reef biodiversity: predicting and protecting genetic diversity of reef assemblages. *Proceedings of the Royal Society of London B: Biological Sciences* **283**, 20160354.

Siddall M, Rohling EJ, Almogi-Labin A*, et al.* (2003) Sea-level fluctuations during the last glacial cycle. *Nature* **423**, 853.

Skillings DJ, Bird CE, Toonen RJ (2011) Gateways to Hawai ‘i: genetic population structure of the tropical sea cucumber *Holothuria atra*. *Journal of Marine Biology* **2011**.

Smith MK (1993) An ecological perspective on inshore fisheries in the main Hawaiian Islands. *Marine Fisheries Review* **55**, 34-49.

Song CB, Near TJ, Page LM (1998) Phylogenetic relations among percid fishes as inferred from mitochondrial cytochrome *b* DNA sequence data. *Molecular Phylogenetics & Evolution* **10**, 343-353.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* **73**, 1162-1169.

Swearer SE, Shima JS, Hellberg ME*, et al.* (2002) Evidence of self-recruitment in demersal marine populations. *Bulletin of Marine Science* **70**, 251-271.

Szabo Z, Snelgrove B, Craig MT, Rocha LA, Bowen BW (2014) Phylogeography of the Manybar Goatfish, *Parupeneus multifasciatus* reveals moderate structure between the Central and North Pacific and a cryptic endemic species in the Marquesas. *Bulletin of Marine Science* **90**, 493-512.

Taberlet P, Meyer A, Bouvet J (1992) Unusual mitochondrial DNA polymorphism in two local populations of blue tit *Parus caeruleus*. *Molecular Ecology* **1**, 27-36.

Tenggardjaja KA, Bowen BW, Bernardi G (2016) Reef Fish Dispersal in the Hawaiian Archipelago: Comparative Phylogeography of Three Endemic Damselfishes. *Journal of Marine Biology* **2016**, 17.

Thrasher DJ, Butcher BG, Campagna L, Webster MS, Lovette IJ (2018) Double-digest RAD sequencing outperforms microsatellite loci at assigning paternity and estimating relatedness: A proof of concept in a highly promiscuous bird. *Molecular Ecology Resources* **18**, 953-965.

Thresher RE, Brothers EB (1985) Reproductive ecology and biogeography of Indo-West Pacific Angelfishes (Pisces: Pomacanthidae. *Evolution* **39**, 878-887.

Timmers MA, Andrews KR, Bird CE*, et al.* (2011) Widespread dispersal of the crown-of-thorns sea star, *Acanthaster planci*, across the Hawaiian Archipelago and Johnston Atoll. *Journal of Marine Biology* **2011**.

Timmers MA, Bird CE, Skillings DJ, Smouse PE, Toonen RJ (2012) There's no place like home: Crown-of-Thorns outbreaks in the central Pacific are regionally derived and independent events. *PLOS One* **7**, e31159.

Toonen RJ, Andrews KR, Baums IB*, et al.* (2011) Defining boundaries for applying ecosystem-based management: A multispecies case study of marine connectivity across the Hawaiian Archipelago. *Journal of Marine Biology*, Article ID 460173.

Toonen RJ, Bowen BW, Iacchei M, Briggs JC (2016) Biogeography, Marine. In: *Encyclopedia of Evolutionary Biology. vol. 1* (ed. Kliman RM). Academic Press., Oxford.

Toonen RJ, Puritz JB, Forsman ZH*, et al.* (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* **1**, e203.

Vogler C, Benzie J, Lessios H, Barber P, Worheide G (2008) A threat to coral reefs multiplied? Four species of crown-of-thorns starfish. *Biol Lett* **4**, 696-699.

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

Waldrop E, Hobbs JP, Randall JE*, et al.* (2016) Phylogeography, population structure, and evolution of coral-eating butterflyfishes (Subgenus *Corallochaetodon*). *Journal of biogeography. In press*.

Weersing K, Toonen RJ (2009) Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series* **393**, 1-12.

Wetterstrand KA (2019) *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at:* www.genome.gov/sequencingcostsdata.

White C, Selkoe KA, Watson J*, et al.* (2010) Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences* **277**, 1685-1694.

Winterbottom R, Anderson RC (1997) A revised checklist of the epipelagic and shore fishes of the Chagos Archipelago, Central Indian Ocean. *Smith Institute of Ichthyology Ichthyological Bulletin* **66**, 1-28.

Wright S (1978) *Evolution and the Genetics of Populations: A Treatise in Four Volumes: Vol. 4: Variability Within and Among Natural Populations* University of Chicago Press, Chicago.

Wyrtki K, Kilonsky B (1984) Mean water and current structure during the Hawaii-to-Tahiti shuttle experiment. *Journal of Physical Oceanography* **14**, 242-254.