PHYLOGEOGRAPHY AND THE GENETICS OF INVASIVE SPECIES

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

DOCTOR OF PHILOSOPHY

IN

ZOOLOGY

December 2011

By

Michelle R. Gaither

Dissertation Committee:

Brian Bowen, Chairperson
Robert Toonen
Greta Aeby
David Carlon
Shannon Bennett

ACKNOWLEDGEMENTS

I would like to thank my committee members who provided essential guidance and encouragement throughout my graduate career. First I would like to thank Brian Bowen and Robert Toonen for taking a chance on me despite my long hiatus from academia and my complete lack of molecular experience. As the chair of my committee, Brian Bowen has offered endless support and mentorship that has truly made this dissertation possible. His gift for writing and his clear vision of the substance of a matter has greatly improved my skills as an academic. He has left an indelible mark on me as a scientist. I thank Robert Toonen for his guidance over the past six years, his endless encouragement, and countless discussions. Mostly I would like to thank Rob for the rarely employed, but sometimes essential, proverbial kick in the pants. I would like to thank Greta Aeby whose office, over the past several years, has been a place to vent emotions, discuss issues, and celebrate achievements and whose collaborations and encouragement introduced me to the world of parasitology. I thank David Carlon for providing alternative views, for his steady and reliable voice, and his always sensible advice. Finally, I would like to thank Shannon Bennett who, as a successful academic and mother, is an inspiration, proving that with the right tools and support women *can* do it all. Thank you to the facility, staff, and postdoctoral researchers at the Hawai‘i Institute of Marine Biology. I especially would like to thank: Steve Karl, Luiz Rocha, Matthew Craig, Joseph DiBattista, Zoltan Szabo, and Kimberly Andrews for helpful discussions and manuscript reviews. I thank Jennifer Schultz and Marc Crepeau, who spent many hours in the laboratory teaching me how to be a molecular biologist. This work was made possible by the financial support of several agencies including: National Science Foundation, University of Hawai‘i Sea Grant, Northwestern Hawaiian Islands Coral Reef Reserve, Dai Ho Chun Fund, University of Hawai‘i

Art and Sciences Council, University of Hawai‘i Graduate Student Organization, University of Hawai‘i Ecology, Evolution, and Conservation Biology Specialization, Hawai‘i Division of Aquatic Resources, and the Western Pacific Fisheries Management Council.

On a more personal note I would like to thank all past and present graduate students and interns at the Toonen-Bowen laboratory who provided stimulating conversation and debate and made my graduate experience thoroughly enjoyable. I would like to thank my Hawai‘i ‘ohana: Joanne Birch, Karen Brown, Kim Peyton, Kim Tice, Ilene Haraki, Gale Haraki, Nitsa McCarthy, and Team Jazz for their love and support. I want to thank my life partner, Shelley Jones, whose unwavering support, encouragement, and love puts the world into perspective. Finally, I wish to thank the women of my family: my Grandmother Ruth, whose strong character and love of hard work motivated me, my Aunt Robin, whose natural confidence and doggedness inspired me, and lastly my Mother whose unconditional love lifted me up and gave me the opportunity and determination to be me. This dissertation is dedicated to them.

ABSTRACT

Alien species can affect ecosystem structure through habitat alteration, competition with native species, hybridization, or direct predation upon natives. Predicting which species are probable invaders, understanding how they spread, and which ecosystems are most vulnerable is of immense scientific and practical interest. For my dissertation I applied phylogeographic and population level analyses to study introduced fishes in Hawai‘i to examine how these species spread and what factors limit their ranges. Here I capitalized on the 1950’s introduction of *Lutanjus kasmira*, *L. fulvus*, and *Cephalopholis argus* to Hawai‘i. All three species are firmly established in Hawai‘i. *L. kasmira* is by far the most successful of the three fishes having spread over 2,500 km and reaching the far northwest end of the archipelago in just 34 years. *C. argus*, on the other hand, has spread only to the middle of the archipelago at French Frigate Shoals while *L. fulvus* is restricted (so far) to the Main Hawaiian Islands. Conducting range-wide genetic surveys for each speices and using genetic structure across the natural range as a proxy for dispersal ability I found a remarkable correlation between genetic structure and invasion success. *L. kasmira*, the most widespread of the three species in the non-native range, demonstrates little genetic structure across nearly 20,000 km of its natural range, indicating that this species is able to cross large stretches of open ocean, find suitable habitat, settle, and reproduce. In contrast, *L. fulvus*, the least widespread of the three species in Hawaii, showed significant population structure at all geographic scales indicating that this species is successfully dispersing over only short geographic distances. *C. argus* shows an intermediate pattern. While the association between invasiveness and dispersal ability seems intuitive, this is the first time the relationship has been examined using empirical data. Understanding how invasive species

spread and what affects they have at the ecosystem level will allow more informed management of these altered systems as well as predict the consequences of future invasions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Introduction

Many human activities, such as plant and animal culture, trans-oceanic shipping, and the mass movement of people, promote the spread of species outside their natural geographic ranges. While many introduced species never become established, those that persist can have grave effects on human health, serious economic impacts, and pose a significant threat to biodiversity and ecosystem function (Mack et al. 2000, Kolar & Lodge 2001). Predicting which species are probable invaders, understanding how they spread, and which ecosystems are most vulnerable is of immense scientific and practical interest (Mack et al. 2000). Studies of introduced species are often impeded by the fact that, in most cases, the identity of the source population and the number of founding individuals are unknown. For my dissertation research I capitalized on the unusually well documented introduction of three marine fishes to Hawai‘i. I combined molecular techniques, ecological data, and government records to characterize the introductions and to shed light on the question of what makes a good invader. The circumstances of these introductions offer unprecedented opportunities to study species introductions pertinent to management of marine resources including the Papahānaumokuākea Marine National Monument (Fig. 1.1).

The Hawaiian fish introductions

Situated 4,000 km from the nearest continental landmass, Hawai‘i is the most isolated archipelago in the world. The Hawaiian Islands are geologically quite young with Kauai, the oldest of the Main Hawaiian Islands, being about 5.8 million years old (Clague & Dalrymple 1987). Their geographic isolation, coupled with the young geologic age of the islands has contributed to the depauperate nature of the Hawaiian fauna. There are 2,700 shorefish species recorded in the Indo-Malaysian region but only 622 are documented in Hawai‘i (Randall 1998, 2007). Almost completely lacking is the suite of shallow water snappers and groupers that are common throughout the Indo-Pacific and that support thriving artisanal fisheries. In an effort to

enhance local fisheries, the Hawai‘i Division of Fish and Game (HDFG) introduced twelve species of snapper and groupers to the Hawaiian Islands (Oda & Parrish 1982, Randall 1987). Of those, three became established: the Bluestripe Snapper *Lutjanus kasmira*, Blacktail Snapper *L. fulvus*, and Peacock Hind *Cephalopholis argus*.

These introductions are exceptional in that they were discrete and well documented events with known numbers of founders and source populations (Table 1.1, HDFG records). Government records indicate that these species became established and began to spread soon after release. *L. kasmira* is by the far the most successful of the three, reaching high densities on many reefs in Hawai‘i. This species has spread throughout the islands including to Midway Atoll in the far northwest of the archipelago (Oda & Parrish 1982, Randall 2007). *C. argus*, on the other hand, has spread only to the middle of the archipelago at French Frigate Shoals and *L. fulvus* is restricted to the Main Hawaiian Islands (Fig. 1.1).

Phylogeography: the tool of invasion biologists

Phylogeography is the study of the historical processes that are responsible for the contemporary distribution of genetic lineages within a species (Avise 2000). Understanding the geographic distribution of genetic diversity in the native range allows invasion biologists to identify source populations, to identify levels of genetic diversity in a species' natural range, and to pinpoint potential introduction pathways (Zardus & Hadfield 2005, Grapputo et al. 2005, Roman 2006, Rosenthal et al. 2008). Molecular tools can also be employed to test fundamental questions about how populations establish, disperse, and ultimately define range boundaries.

The impact of founder population size on genetic diversity

It is expected that the often dramatic decrease in effective population size that accompanies founder events leads to a decreased genetic diversity, decreased heterozygosity, and

increased expression of recessive deleterious alleles. If a colonizing population survives the founding event, lower genetic diversity is thought to limit the evolutionary potential of the newly established population, reducing its ability to respond to or adapt to changing environmental conditions and ultimately make it more susceptible to extinction (Frankham 2004). However, the accumulation of data indicates that genetic bottlenecks in founder populations are not as common as expected (Roman & Darling 2007, Bossdorf et al. 2005, Wares et al. 2005). Interpreting patterns of genetic diversity in introduced populations is confounded by the fact that in most cases the source population and the number of founding individuals are unknown. Under these conditions, researchers must reconstruct the history of introductions by combining molecular and geographic data to identify source populations (Wares et al. 2005). In some cases high genetic diversity in the introduced range can be attributed to admixture of genetically divergent populations (Kolbe et al. 2004, Genton et al. 2005, Carmeron et al. 2008, Rosenthal et al. 2008). In cases where only a single source population can be identified, high genetic diversity in the introduced range is often attributed to either a large number of colonizers, rapid population expansion following the founder event, or both (Hassan et al. 2003, Stepien et al. 2005). Discussions concerning the effect of founder events on genetic diversity would be greatly informed if more empirical data concerning the effects of founder population size on genetic diversity were available. Intentional and well documented introductions, where the source population and founder population size are confidently known, offer powerful test cases.

The intentional introduction of *L. kasmira*, *L. fulvus*, and *C. argus* to the Hawaiian Islands provides a rare opportunity to directly evaluate the effects of founder population size on genetic diversity in recently established populations. For my dissertation I capitalize on two unique aspects of these introductions: 1) the introductions occurred in well documented events

with known numbers of founders and source populations, 2) the source populations for *L. kasmira* (Nuka Hiva in the Marquesas Islands and Moorea in the Society Islands) are genetically distinct populations, allowing us to identify their descendents. I employed both mitochondrial and nuclear sequence data and conducted range wide phylogeographic surveys of *L. kasmira*, *L. fulvus* (Chapter 2, published in Journal of Biogeography), and *C. argus* (Chapter 3, published in BMC Evolutionary Biology) to assess natural patterns of genetic diversity in these three species. Finally, I characterized the introduction of *L. kasmira* to Hawai‘i (Chapter 4, published in Molecular Ecology) to assess the impact of founder population size on genetic diversity and to determine what molecular tools can add to our knowledge about the spread of *L. kasmira* throughout the Hawaiian archipelago. The overarching aim of my research is to explore the utility of genetic markers to expand our understanding of species introductions and to determine how phylogeography can inform the field of invasion biology.

REFERENCES

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

Bossdorf O, Auge H, Lafuma L, Rogers WE, Siemann E, Prati D (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, **144**, 1-11.

Cameron EK, Bayne EM, Coltman DW (2008) Genetic structure of invasive earthworms *Demdrpbaema octaedra* in the boreal forest of Alberta: insights into introduction mechanisms. *Molecular Ecology*, **17**, 1189-1197.

Clague DA, Dalrymple GB (1987) The Hawaiian-Emperor volcanic chain: Part 1. Geologic evolution. U.S. Geological Survey Professional Paper, **1350**, 5-54.

Frankham R (2005) Invasion biology: resolving the genetic paradox in invasive species. *Heredity*, **94**, 385-385.

Genton BJ, Shykoff A, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular Ecology*, **14**, 4275-4285.

Grapputo A, Boman S, Lindstrom L, Lyytinen A, Mappes J (2005) The voyage of an invasive species across continents: genetic diversity of North American and European Colorado potato beetle populations. *Molecular Ecology*, **14**, 4207-4219.

Hassan M, Harmelin-Vivien M, Bonhomme F (2003) Lessepsian invasion without bottleneck: example of two rabbitfish species (*Siganus rivulatus* and *Siganus luridus*). *Journal of Experimental Marine Biology and Ecology*, **291**, 219-232.

Kolar C, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution*, **16**, 199-204.

Kolbe JJ, Glor RE, Schettino LR, Lara AC, Larson A, Losos JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 480-483.

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M, Bazzaz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications*, **10**, 689-710.

Oda DK, Parrish JD (1982) Ecology of commercial snappers and groupers introduced to Hawaiian reefs. *Proceedings of the Fourth International Coral Reef Symposium*, Manila, 59-67.

Randall JE (1987) Introductions of marine fishes to the Hawaiian Islands. *Bulletin of Marine Science*, **47**, 356-400.

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

Randall JE (2007) *Reef and Shore Fishes of the Hawaiian Islands*. Sea Grant College Program, University of Hawai‘i, Honolulu.

Roman J (2006) Diluting the founder effect: cryptic invasions expand a marine invader’s range. *Proceedings of the Royal Society B*, **273**, 2453-2459.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution*, **22**, 454-464.

Rosenthal DM, Ramakrishnan A, Cruzan MB (2008) Evidence for multiple sources of invasion and intraspecific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. in North America. *Molecular Ecology*, **17**, 4657-4669.

Stepien CA, Randall JE, Rosenblatt RH (1994) Genetic and morphological divergence of a circumtropical complex of goatfishes: *Mulloidichthys vanicolensis, M. dentatus* and *M. martinicus*. *Pacific Science*, **48**, 44-56.

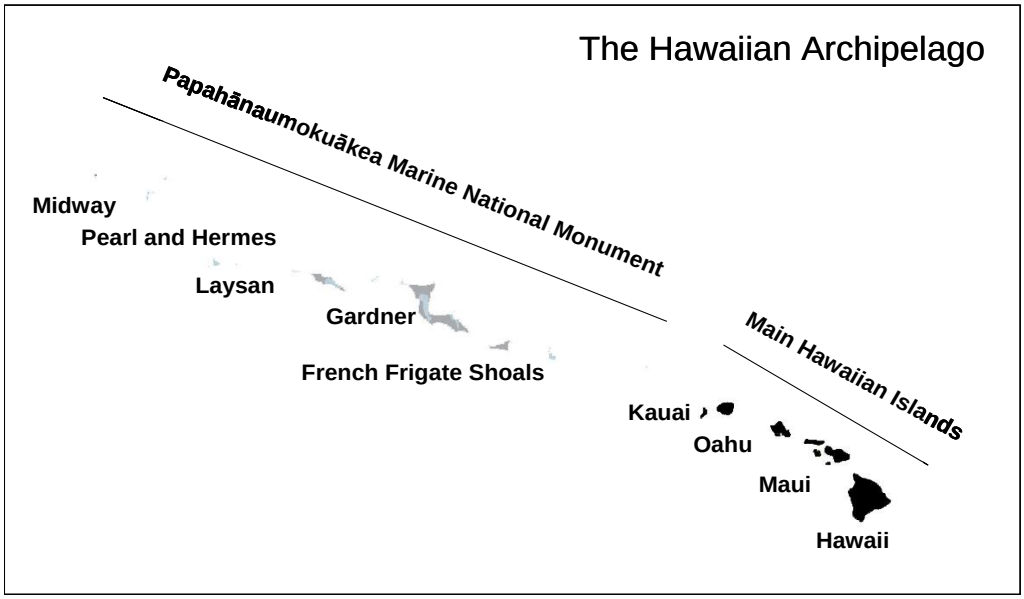
Wares JP, Hughes R, Grosberg K (2005) Mechanisms that drive evolutionary change. In: *Species Invasions: Insights Into Ecology, Evolution, and Biogeography* (ed. Sax DF, Stachowicz JJ, Gaines SD), pp 229-257, Sinauer Press, Sunderland.

Zardus JD, Hadfield MG (2005) Multiple origins and incursions of the Atlantic barnacle *Chthamalus proteus* in the Pacific. *Molecular Ecology*, **14**, 3719-3733.

Table 1.1. List of three species of reef fish released into Hawaiian waters between 1955 and 1961 by the Hawai‘i Division of Fish and Game. Source location, number released, and area of release are listed for each species.

Species	Source	Number Released	Area Released
Lutjanus kasmira	Nuku Hiva	2435	Oahu
	Moorea	728	
Cephalopholis argus	Moorea	1985	Oahu
		400	Hawai‘i Island
Lutjanus fulvus	Marquesas	23	Oahu
	Moorea	2021	
	Kanton	148	

Figure 1.1. Map of the Hawaiian Archipelago



CHAPTER TWO

Genetic evaluation of marine biogeographical barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*)

Published as: Gaither, M.R., R.J. Toonen, D.R. Robertson, S. Planes, and B.W. Bowen. 2010. Genetic evaluation of marine biogeographic barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus* spp.). Journal of Biogeography 37:133-147.

ABSTRACT

Aim In the Indo-Pacific, the mass of islands of the Indonesian archipelago constitute a major biogeographical barrier (the Indo-Pacific Barrier, IPB) separating the Pacific and Indian Oceans. Evidence for other, more localized barriers include high rates of endemism at the Marquesas and other isolated peripheral islands in the Pacific. Here we use mitochondrial-sequence comparisons to evaluate the efficacy of biogeographical barriers on populations of the snappers *Lutjanus kasmira* and *Lutjanus fulvus* across their natural ranges.

Location Pacific and Indian Oceans.

Methods Mitochondrial cytochrome *b* sequence data were obtained from 370 individuals of *L. kasmira* and 203 individuals of *L. fulvus* collected from across the species ranges. Allele frequency data for two nuclear introns were collected from *L. kasmira*. Phylogenetic and population level analyses were used to determine patterns of population structure in these species and to identify barriers to dispersal.

Results *Lutjanus kasmira* lacks genetic structure across the IPB and throughout 12,000 km of its central Indo-Pacific range. In contrast, *L. fulvus* demonstrates high levels of population structure at all geographical scales. In both species, highly significant population structure was due primarily to phylogenetic distinctiveness of their Marquesas Islands populations (*L. kasmira*, d = 0.50-0.53%; *L. fulvus*, d = 0.87-1.50%). Coalescence analyses of the *L. kasmira* data indicate that populations at opposite ends of its range (western Indian Ocean and the Marquesas) are the oldest. Coalescence analyses for *L. fulvus* are less robust but also indicate colonization from the Indian to the Pacific Ocean.

Main conclusions The IPB does not act as a biogeographical barrier to *L. kasmira*, and, in *L. fulvus*, its effects are no stronger than isolating mechanisms elsewhere. Both species demonstrate a strong genetic break at the Marquesas. Population divergence and high endemism in that

archipelago may be a product of geographical isolation enhanced by oceanographic currents that limit gene flow to and from those islands, and adaptation to unusual ecological conditions. *L. kasmira* shows evidence of Pleistocene population expansion throughout the Indo-central Pacific that originated in the western Indian Ocean rather than the Marquesas, further demonstrating a strong barrier at the latter location.

INTRODUCTION

Biogeographical provinces are defined by level of endemism and recognized habitat discontinuities (Briggs, 1974; Brown & Lomolino, 1998; Bellwood & Wainwright, 2002; Robertson *et al*., 2004; Floeter *et al*., 2008). In the open ocean, biogeographical barriers can shape species distributions by physically restricting dispersal (strong ocean currents) or by limiting dispersal success due to the distance between areas of suitable habitat. Examples of tropical marine biogeographical barriers include the expanses of open ocean that separate the central and eastern Pacific (Robertson *et al*., 2004; Lessios & Robertson, 2006) and the eastern and western Atlantic (Banford *et al*., 1999), the Amazon-Orinoco outflow that separates the Brazilian and Caribbean provinces (Rocha *et al*., 2002), and the large cold upwelling area off Southwest Africa that separates the tropical Atlantic and Indian Oceans (Briggs, 1974; Rocha *et al*., 2007).

The Indo-Pacific Barrier (IPB), is a widely recognized partition (based on faunal distributions) that separates the Pacific and Indian Ocean provinces (Briggs, 1974). While the location of the boundary between these provinces is debated (see Hobbs *et al*., 2009) this barrier is most often associated with the Sunda Shelf between Asia and Australia (Fleminger, 1986; Barber *et al*., 2006). During Pleistocene glacial cycles the repeated lowering of sea levels (as low as 120 m below present levels) imposed a nearly complete barrier between the two oceans.

Strong upwelling in the region probably enhanced the effectiveness of the IPB by reducing the availability of suitable habitat for tropical marine organisms (Galloway & Kemp, 1981; Fleminger, 1986; Voris, 2000; Naish *et al*., 2009). Historical and contemporary restrictions to dispersal between the Pacific and Indian Oceans are indicated by the confinement of many demersal species primarily to one ocean or the other (Woodland, 1983; McMillan & Palumbi, 1995; Randall, 1998; Briggs, 1999; but see Hobbs *et al.*, 2009). More recently, the effects of the IPB and other barriers have been assessed using genetic data. Taxa that disperse as adults show genetic continuity between the Pacific and Indian Oceans, indicating that they cross the IPB with regularity (e.g. whale shark, *Rhincodon typus*, Castro *et al*., 2007; wahoo, *Acanthocybium solandri*, Theisen *et al*., 2008). Studies of demersal organisms that lack vagile adults have found intraspecific genetic differentiation across the IPB in many fishes (Lacson & Clark, 1995; Planes & Fauvelot, 2002; Bay *et al.*, 2004; Lu *et al*., 2006; Menezes *et al*., 2006; Craig *et al*., 2007) and invertebrates (Lavery *et al*., 1995, 1996; Williams & Benzie, 1998; Benzie, 1999; Duda & Palumbi, 1999; Barber *et al.*, 2000 Lessios *et al*., 2001, 2003) with few exceptions (Lessios *et al*., 2001; Bowen *et al*., 2001; Horne *et al*., 2008).

In addition to the biogeographical barriers that separate tropical marine provinces, barriers within provinces can promote more localized isolation. Randall (1998, 2001, 2007) discussed five areas in Oceania that have high levels of endemism among shore fishes: Hawaii (25.0%), Easter Island (22.2%), the Marquesas (11.6%), Lord Howe and Norfolk Island (7.2%), and Rapa (5.5%). These peripheral island groups are all geographically isolated, being 500–1500 km from the nearest neighbouring archipelago (Randall, 1998). Genetic studies of widely distributed species have demonstrated genetic divergences among these five insular areas in milkfish (*Chanos chanos;* Winans, 1980), convict surgeonfish (*Acanthurus triostegus*; Planes &

Flauvelot, 2002), yellowfin goatfish (*Mulloidichthys vanicolensis*; Stepien *et al*., 1994), crown-of-thorns starfish (*Acanthaster planci*; Benzie & Stoddart, 1992), and four species of coral (Ayre & Hughes, 2004). However, only one genetic study has involved comparison of Marquesan populations (*Acanthurus triostegus*; Planes & Fauvelot, 2002).

The common bluestripe snapper, *Lutjanus kasmira* (Forsskål, 1775), and the blacktail snapper, *Lutjanus fulvus* (Schneider, 1801) (Lutjanidae), are close relatives (Miller & Cribb, 2007) that occupy the same geographical range, from the Marquesas Islands in the central Pacific to the east coast of Africa (Fig. 2.1). *L. kasmira* inhabits a wide depth range from shallow water to at least 265 m (Allen & Talbot, 1985; Randall, 1987; Oda & Parrish, 1987; Friedlander *et al*., 2002). Little has been published on the ecology of *L. fulvus*, but anecdotal reports from fishermen indicate that *L. fulvus* occupies a narrower depth range than *L. kasmira* (1–40 m). Both species are nocturnal predators. Data from two studies indicate that *L. kasmira* reaches sexual maturity at 1–2 years (Rangarajan, 1971; Morales-Nin & Ralston, 1990), and engages in mass spawning (Suzuki & Hioki, 1979). The pelagic larval duration (PLD) is unknown for both *L. kasmira* and *L. fulvus*, although other lutjanids have PLDs of 20–44 days (Zapata & Herron, 2002; Denit & Sponaugle, 2004).

Here we analysed DNA sequence data from *L. kasmira* and *L. fulvus* to assess biogeographical partitions in the Indo-Pacific. We resolved patterns of genetic isolation among reefs and archipelagos of the Indo-Pacific to address the following questions, 1) Are patterns of genetic structure concordant with known biogeographical boundaries? 2) Specifically, is there evidence of genetic structure across the Indo-Pacific Barrier? 3) What processes might account for geographical patterns of genetic structure in these fishes?

MATERIALS AND METHODS

A total of 370 individuals of *Lutjanus kasmira* from twelve locations and 203 individuals of *L. fulvus* from six locations were collected from across the species range in the Pacific and Indian Oceans (Fig. 2.1) by scuba divers using polespears. Tissue samples (fin clips or gill filaments) were preserved in either 95% ethanol (EtOH) or saturated NaCl solution (Seutin *et al.*, 1991) and stored at room temperature. DNA was isolated using DNeasy Tissue kits (Qiagen, Inc., Valencia, CA, USA) following the manufacturer's protocol for animal tissues, and stored at -20° C. In specimens of *L. kasmira* and *L. fulvus*, approximately 560 bp of mitochondrial cytochrome *b* (cyt*b*) were amplified using the primers H15020 (Meyer, 1994) and Cyt*b*-07L (Taberlet *et al*., 1992). In specimens of *L. kasmira*, we also used the primers Gh5F and Gh6R (Hassan *et al*., 2002), to amplify approximately 650 bp of intron 5 of the nuclear growth hormone (GH) gene and the primers ANTf1 and ANTr1 (Jarman *et al*., 2002), to amplify approximately 395 bp of the nuclear adenine nucleotide transporter translocase (ANT) intron. Polymerase chain reactions (PCR) for all three markers were carried out in a 20 µl volume containing 5–50 ng of template DNA, 0.25–0.5 µM of each primer, 1.5 mM $MgCl_2$, 2.5 mM of each dNTP, 0.5 units IMMOLASE™ DNA polymerase (Bioline Inc., Springfield, NJ, USA), 2.0 µl of 10x ImmoBuffer (Bioline Inc.) and deionized sterile water to volume. PCRs utilized the following cycling parameters: initial denaturation at 95°C and final extension at 72°C (ten minutes each), with an intervening 35 cycles of 30 seconds at 94°C, 30 seconds at the annealing temperature (48°C for cyt*b*; 60°C for GH; 58°C for ANT), and 1 minute at 72°C. Amplification products were purified using 0.75 units of Exonuclease I: 0.5 units of Shrimp Alkaline Phosphatase (ExoSAP, USB, Cleveland, OH, USA) per 7.5 µl PCR products at 37°C for 60 minutes, followed by deactivation at 80°C for 10 minutes. DNA sequencing was performed with

fluorescently-labelled dideoxy terminators on an ABI 3130XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) at the Hawaii Institute of Marine Biology EPSCoR Sequencing Facility. All samples were initially sequenced in the forward direction. Unique genotypes were confirmed by subsequent sequencing in the reverse direction.

Individuals of *L. kasmira* whose allelic states at the nuclear GH and ANT loci could not be resolved unambiguously from direct sequences, were cloned using a TA cloning method to identify the alleles. Prior to cloning, PCR products were purified using the Qiagen QIAquick PCR Purification kit following manufacturer's protocol and eluting in 30 µl of elution buffer per 20 µl PCR reaction. All products were cloned using the T-tailed vector pZErO-2 (Invitrogen Corp., Carlsbad, CA, USA) in *Escherichia coli* strain DB3.1. Ligation reactions included 1.6 µl 2x ligation buffer (Promega Corp., Madison, WI, USA), 5.0 ng of EcoRV digested and T-tailed pZErO-2 vector, 1.0 ng PCR product and 0.4 µl of T4 DNA ligase (Promega Corp., Madison, WI, USA) in 4 µl total volume. Heat shock transformation was conducted using 10 µl of α-Select Chemically Competent Cells (Bioline Inc., Springfield, NJ, USA) and 1 µl of ligated vector following manufacturer's protocol. LB agar plates with 50 µg/ml kanamycin were spread with 40 µl of transformation mixture. Following an overnight incubation at 37°C, cells from individual colonies were suspended in 20 µl of sterile water. Clones were amplified in 20 µl PCRs containing 0.1 µl of the cell suspension, 0.2 µM of each primer (M13F-GTAAAACGACGGCCAG and M13R-CAGGAAACAGCTATGAC), 2.0 mM $MgCl_2$, 0.15 mM dNTPs, 0.5 µg BSA, 0.5 units IMMOLASE™ DNA polymerase, and 2.0 µl of 10x PCR buffer (Bioline, Inc., Springfield, NJ, USA). PCR cycling conditions included an initial denaturing step at 94ºC for 3 min followed by 35 cycles of 94ºC for 30 sec, 60ºC for 30 sec and 72ºC for 30 sec,

followed by a final extension at 72ºC for 10 min. To verify the identity of alleles, clones were sequenced until at least two copies of each allele were observed.

Sequences for each locus were aligned, edited, and trimmed to a common length using the DNA sequence assembly and analysis software SEQUENCHER 4.6 (Gene Codes, Ann Arbor, MI, USA). In all cases, alignment was unambiguous with no indels or frameshift mutations at any locus. Unique haplotypes and alleles were identified with the merge taxa option in the phylogenetic analysis software MacClade 4.05 (Maddison & Maddison, 2002), and deposited in GenBank (ascension numbers: *L. kasmira* FJ754049-FJ754133 (cyt *b*), FJ754178-FJ754184 (GH intron), FJ754157- FJ754177 (ANT intron); *L. fulvus* FJ754134-FJ754156 (cyt *b*).

Data analyses

Mitochondrial DNA

Summary statistics for each species, including haplotype diversity (*h*) and nucleotide diversity (π), were estimated with algorithms in Nei (1987) as implemented in the statistical software package ARLEQUIN 3.11 (Excoffier *et al*., 2005). Fu's F_S (Fu, 1997) was calculated to test for evidence of population expansion using 10,000 permutations and excluding the two locations with small sample size (Lizard Island and Okinawa; Table 2.1). Significant negative values of F_S indicate an excess of low-frequency haplotypes, a signature characteristic of either selection or a recent demographic expansion (Fu, 1997). Haplotype frequencies and distance matrixes were used to estimate gene genealogies and statistical parsimony networks were constructed using the program TCS 2.21 (Clement *et al*., 2000). We used mismatch analyses to determine whether the number of pairwise differences among all DNA sequences reflected expanding or stable populations (Harpending, 1994; Schneider *et al*., 2000). Effective female population size (N_{ef}) was calculated from the cyt*b* data using the equation $\theta = 2N_{ef}\, v$, where v =

the mutation rate per generation for the entire sequence (v = number of bp × divergence rate within a lineage × generation time in years). Population age was calculated using the equation $\tau = 2vt$, where t is age in generations. We use a sequence divergence estimate of 1–2% per million years between lineages to estimate coalescence times (see Discussion). Generation times for most snappers are unknown, including *L. kasmira* and *L. fulvus*. Combining size at first maturity data for *L. kasmira* from the Andaman Sea (Rangarajan, 1971) with age-length relationships of *L. kasmira* from Hawaii (Morales-Nin & Ralston, 1990) we estimate generation time to be 3.7 years for this species. Because *L. kasmira* and *L. fulvus* are closely related congeners of similar size (Miller & Cribb, 2007) we provisionally apply this estimation of generation time to *L. fulvus* as well. Given the tentative nature of the estimate of generation time and mutation rate, the corresponding population age and N_{ef} values should be interpreted as first-order approximations (see Hudson & Turelli, 2003).

A likelihood approach, implemented in MODELTEST 3.7 (Posada & Crandall, 1998) was used to determine the mutational model which best fit the cyt*b* data. The GTR+I and the TIM+G models were found to be the best fit models by the Akaike information criterion for *L. kasmira* and *L. fulvus*, respectively. Because neither model is available in ARLEQUIN 3.11 (Excoffier *et al*., 2005), we used the most similar model available (Tamura & Nei, 1993). To test for hierarchical population genetic structure in each species, an analysis of molecular variance (AMOVA) was performed in ARLEQUIN using 20,000 permutations. An analogue of Wright's F_{ST}, which incorporates a model of sequence evolution (Φ_{ST}), was calculated for each data set and for pairwise comparisons among all locations with greater than five specimens. The average number of nucleotides (corrected; Tamura & Nei, 1993) that differ between populations was

calculated in ARLEQUIN. From these values we calculated average percent difference between populations, which we report here as sequence divergence (*d*).

Nuclear introns

Observed heterozygosity (H_O) and expected heterozygosity (H_E) were calculated for each locus and an exact test of Hardy-Weinberg equilibrium using 100,000 steps in a Markov chain was performed using ARLEQUIN. We tested for linkage disequilibrium between the two nuclear loci using the likelihood ratio test with 20,000 permutations as implemented in ARLEQUIN. Statistical parsimony networks for alleles at each locus were constructed using the program TCS. The multi-locus data set includes only those individuals that amplified at both loci. Using this data set we calculated mean H_E and F_{ST} for the entire data set and for pairwise comparisons between samples with greater than five specimens.

RESULTS

Lutjanus kasmira

Mitochondrial DNA

We resolved 475 bp of cytochrome *b* in 370 individuals of *L. kasmira* (Fig. 2.1). Due to low sample size, Lizard Island ($n = 5$) and Okinawa ($n = 2$) were excluded from all inter-population analyses. We observed 83 haplotypes with 56 of these haplotypes occurring in single individuals. The number of individuals (n), number of haplotypes (N_h), haplotype diversity (h), and nucleotide diversity (π) for each location are provided in Table 2.1. Overall nucleotide diversity in *L. kasmira* was low ($\pi = 0.004$) while the corresponding haplotype diversity was high ($h = 0.74$). Across all samples, $\pi = 0.001–0.005$ and $h = 0.50–0.88$. The most common haplotype (51.2% of specimens) was found at every location except the Marquesas (Fig. 2.2a).

In stark contrast, the Marquesas shared no haplotypes with any other location and formed an isolated cluster in the parsimony network (Fig. 2.2a).

Overall Φ_{ST} was 0.30 ($P < 0.001$). When we grouped samples by ocean basin (Pacific Ocean = Marquesas, Moorea, Kiritimati, Fiji, Guam, Okinawa and Lizard Island; Indian Ocean = Cocos-Keeling, Christmas Island, Diego Garcia, Seychelles and Sodwana Bay) we found no significant structure between the Pacific and Indian Oceans ($\Phi_{CT} = -0.003$, $P < 0.59$). Within Oceans we found significant structure in the Pacific Ocean ($\Phi_{ST} = 0.45$, $P < 0.001$) but not in the Indian Ocean ($\Phi_{ST} = 0.002$, $P = 0.33$). Pairwise comparisons indicate high levels of population structure between the Marquesas and every other sample location with significant pairwise Φ_{ST} = 0.53–0.63 ($P < 0.001$) (Table 2.2). When the phylogenetically distinct Marquesas sample (d = 0.50–0.53%) was removed from the analysis, the overall Φ_{ST} dropped to 0.012 ($P = 0.02$), we found low but significant structure between the Pacific and Indian Oceans ($\Phi_{CT} = 0.013$, P = 0.05), and Φ_{ST} within the Pacific Ocean was no longer significant ($\Phi_{ST} = 0.012$, $P = 0.12$). Within this data set, Moorea showed low but significant levels of structure when compared to all other sample locations (Φ_{ST} = 0.032–0.062; Table 2.2). The Indian Ocean locations of Diego Garcia and Sodwana Bay showed low but significant levels of population structure when compared to all Pacific Ocean locations (Φ_{ST} = 0.024–0.045, Φ_{ST} = 0.014–0.048, respectively; Table 2.2). The only other significant pairwise comparison was between Kiritimati and Cocos-Keeling ($\Phi_{ST} = 0.022$, $P = 0.015$).

Fu's F_S for the overall data set was -26.15 ($P < 0.001$) indicating an excess of low-frequency haplotypes (Table 2.1). The mismatch distribution (Fig. 2.3a) for the overall data set was unimodal (Harpending's raggedness index $r = 0.013$, $P = 0.99$) and resulted in values of τ = 3.23, $\theta_0 = 0.00$, and $\theta_1 = 3.05$. Based on a generation time of 3.7 years and a molecular clock

estimate of 1–2% divergence per 10^6 years between species (Lessios, 2008; Bowen *et al*., 2001; also see Discussion), we calculated a coalescence time of roughly 340,000 - 680,000 years. Initial female effective population estimate is $N_{ef0} = 0$ and current effective population estimate is N_{ef1} = 90,000–180,000. Coalescence times and demographic parameter estimates for each population are reported in Table 2.1.

Nuclear introns

We resolved 148 bp of the GH intron in 369 specimens and 168 bp of the ANT intron in 336 specimens of *L. kasmira* (Table 2.3). Seven polymorphic sites yielded 7 alleles at the GH locus and 16 polymorphic sites yielded 22 alleles at the ANT locus. Statistical parsimony networks showed that alleles at each locus are closely related (Fig. 2.2c, d). The most common allele at each locus was observed at all sample locations across the range. The second most common allele at both loci was observed primarily in the Marquesas Islands, but also at low frequencies elsewhere. The number of individuals (n), number of alleles (N_a), observed heterozygosity (H_O), expected heterozygosity (H_E), and the corresponding P-value for the exact test for Hardy-Weinberg equilibrium for each locus are listed in Table 2.3. Number of genotypes (N_g) and H_E are also listed for the multi-locus data set (Table 2.3). The samples from Diego Garcia were found to be out of Hardy-Weinberg equilibrium with an excess of homozygotes at both the GH and ANT loci (P = 0.03 and 0.03 respectively) while the samples from the Marquesas, Moorea, Guam were found to have an excess of homozygotes at only the ANT locus (P = 0.04, 0.03, and >0.01 respectively) (Table 2.3). Overall H_E was 0.25, 0.61 and 0.43 for the GH, ANT and multi-locus data sets respectively. Across all samples H_E = 0.00–0.59 for the GH intron, H_E = 0.38–0.83 for the ANT intron and mean H_E = 0.21–0.48 for the multi-locus nuclear data set. There was no indication of linkage disequilibrium between the two loci ($P > 0.05$).

Overall F_{ST} for the GH, ANT and the multi locus data set are $F_{ST} = 0.41$ ($P < 0.001$), $F_{ST} = 0.21$ ($P < 0.001$), and $F_{ST} = 0.28$ ($P < 0.001$) respectively. Using the multi-locus data set we grouped samples by ocean basin (Pacific Ocean = Marquesas, Moorea, Kiritimati, Fiji, Guam, Okinawa and Lizard Island; Indian Ocean = Cocos-Keeling, Christmas Island, Diego Garcia, Seychelles and Sodwana Bay). Similar to the cyt*b* data, we found no significant structure between the Pacific and Indian Oceans ($F_{CT} = -0.014$, $P = 0.77$), significant structure within the Pacific Ocean ($F_{ST} = 0.38$, $P < 0.001$), and no structure within the Indian Ocean ($F_{ST} = 0.004$, $P = 0.32$). Pairwise F_{ST} values for the multi-locus data set are reported in Table 2.2. Population comparisons indicate highly significant population structure between the Marquesas and every other sample location with pairwise $F_{ST} = 0.44–0.56$. When the Marquesas was excluded from the analysis, overall F_{ST} dropped to 0.004 ($P = 0.23$), we found low but significant structure between the Pacific and Indian Oceans ($F_{CT} = 0.004$, $P = 0.02$), and F_{ST} within the Pacific Ocean was no longer significant ($F_{ST} = -0.0001$, $P = 0.50$). The only other significant pairwise comparisons were between Moorea and Christmas Island ($F_{ST} = 0.048$, $P = 0.04$) and between Kiritimati and the Seychelles ($F_{ST} = 0.026$, $P = 0.034$).

Concordance between mtDNA vs. nuclear markers

We found strong agreement between our mitochondrial cytochrome *b* and multi-locus nuclear intron data sets for *L. kasmira*. The two marker types show the same pattern of the Marquesas population being highly divergent from all other populations with only low levels of structure elsewhere. The marker types differ only in the degree of genetic differentiation demonstrated with the Marquesas sharing no mitochondrial haplotypes with any other population (Fig. 2.2a) whereas the nuclear markers show strong shifts in allele frequency between populations with the most common allele at each locus being found in every population (Fig.

2.2c, d). These differences may be due to the fourfold lower effective population size of the mitochondrial genome (Avise, 2004), although other factors such as differing mutation rates probably contribute as well.

Lutjanus fulvus

Mitochondrial DNA

We resolved 480 bp of cytochrome *b* in 203 individuals of *L. fulvus* (Fig. 2.1), yielding 23 haplotypes, with 8 of these haplotypes observed in a single individual. Sample sizes and diversity estimates for each location are provided in Table 2.4. Similar to *L. kasmira*, overall nucleotide diversity in *L. fulvus* was low ($\pi = 0.006$) while the corresponding haplotype diversity was high ($h = 0.69$). Across all samples, $\pi = 0.0003–0.006$ and $h = 0.12–0.78$. Similar to *L. kasmira*, the most common haplotype (53.2% of specimens) was found at every location except the Marquesas (Fig. 2.2b).

Overall population structure was much higher in this species than in *L. kasmira* ($\Phi_{ST} = 0.64$, $P < 0.001$). All but three of the fifteen population-level pairwise comparisons were significant ($P < 0.05$; Table 2.5). The three non-significant comparisons were between the Philippines and Kiritimati, the Philippines and Kanton, and the Philippines and Cocos-Keeling. Similar to the findings for the *L. kasmira* data, the highest levels of structure were found at the Marquesas with significant pairwise values ranging from $\Phi_{ST} = 0.66$ to 0.91 ($P < 0.001$) (Table 2.5). When the Marquesas sample ($d = 0.87–1.50\%$) was removed from the analysis, the overall Φ_{ST} dropped to 0.11 ($P < 0.001$).

Fu's F_S for the overall data set was -5.85 ($P < 0.001$), indicating an excess of low-frequency haplotypes (Table 2.4). The mismatch distribution (Fig. 2.3b) for the overall data set was roughly unimodal (Harpending's raggedness index r = 0.067, $P = 0.53$) and resulted in

values of $\tau = 7.00$, $\theta_0 = 0.00$, and $\theta_1 = 3.02$. Based on a generation time calculated for *L. kasmira* (3.7 years) and a molecular clock estimate of 1–2% divergence per 10^6 years between lineages, we calculated a coalescence time of roughly 730,000–1,460,000 years. Initial female effective population estimate is $N_{ef0} = 0$ and current effective population estimate is N_{ef1} = 85,000–170,000. Coalescence times and demographic parameter estimates for each population are reported in Table 2.4.

DISCUSSION

Our survey of the snappers *Lutjanus kasmira* and *L. fulvus* revealed significant levels of genetic structure across their ranges. However, pairwise population comparisons indicated that most of this structure is due to the extreme isolation of the Marquesan population in each species. This location is divergent from all other samples (*L. kasmira*, d = 0.50–0.53%; *L. fulvus*, d = 0.87–1.50%). Once the phylogenetically distinct Marquesan population was removed from the analyses, population structure dropped dramatically in both taxa, revealing contrasting patterns of population structure between the two species. Our results indicate that *L. kasmira* is a highly dispersive fish that displays low but significant genetic structure between the Pacific and Indian Oceans, but no population structure across 12,000 km of its central range from Kiritimati in the central Pacific to Cocos-Keeling in the eastern Indian Ocean. In contrast, the congeneric *L. fulvus* demonstrated significant population structure at every geographical scale we examined across the same 12,000 km, including a break across the IPB that was of similar magnitude to other pairwise comparisons. Prior to dissecting these results, we address two caveats.

1. Several authors have calibrated molecular clocks for cytochrome *b* in marine fishes based on the closure of the Isthmus of Panama. Bowen *et al*. (2001) found a divergence rate of 1.8–2.2% Myr^{-1} for trumpetfishes (*Aulostomus* spp.). Lessios

(2008) summarized data for sixteen pairs of sister taxa distributed across the Isthmus and found sequence divergences of 1.7 to 21.6%. In the five pairs of sister taxa that Lessios (2008) included for which divergence is assumed to have been initiated at final closure of the Isthmus (3 Ma) divergence rates were 1.1–1.6% Myr^{-1} between lineages. Based on these studies, and to convey an appropriate caution, we choose the approximation of 1–2% Myr^{-1} between lineages (0.5–1.0 % within lineages). Regardless of the specific approximation we use our data always indicate a Pleistocene expansion. Thus for data interpretation we use the rank order of the coalescence times and do not attempt to make conclusions based on precise values.

2. Cases in which mitochondrial and nuclear data sets reveal fundamentally different patterns typically involve taxa with sex biased dispersal (turtles, Bowen *et al*., 2005; sharks, Pardini *et al*., 2001; dolphins, Möller *et al*., 2004). However, this pattern has never been documented in reef fishes. Here we found strong genealogical concordance among three independent loci in *Lutjanus kasmira*, and concordance (regarding the isolation of the Marquesas) across two co-distributed species (see Avise, 2004**.** In light of these findings we did not sequence the two nuclear introns in *L. fulvus*.

Effects of the Indo-Pacific Barrier

The shallow Sunda Shelf defines the corridor between the tropical Pacific and Indian Oceans. Lowering of sea level, as much as 120 m below present levels, occurred at least three times during the Pleistocene (Voris, 2000; Naish *et al*., 2009), greatly reducing shallow water habitat in tropical and sub tropical seas and probably causing dramatic declines in populations of shallow-depth fauna. In many species that are found in both oceans, a signature of historical

isolation can be detected in intraspecific genetic differentiation between oceans. A search of the literature uncovered studies of 18 species of marine fishes and invertebrates that were sampled across the IPB (Table 2.6). Of these, 83% (15 of 18) exhibited significant structure across the IPB. To date the only exceptions are the sea urchin *Diadema savignyi* (Lessios *et al*., 2001), the trumpetfish *Aulostomus chinensis* (Bowen *et al*., 2001), and the surgeonfish *Naso brevirostris* (Horne *et al*., 2008). Here we report that *L. kasmira* also shows no evidence of a significant genetic break across the IPB. Although we did find significant structure between the Pacific and Indian Ocean basins in *L. kasmira*, most of that divergence is due to the geographically remote island of Diego Garcia and Sodwana Bay at the end of the species range. The other three Indian Ocean locations, including the Seychelles in the western Indian Ocean, demonstrate no significant population differentiation when compared to the Pacific Ocean. *L. fulvus* has a significant population partition at the IPB, but the level of that differentiation was no higher than detected elsewhere (excluding the Marquesas). In both species, similar levels of divergence occur across the IPB as among other populations, indicating that this barrier does not have strong impact on either species.

Isolation of the Marquesas Islands

The Marquesas Islands are volcanic in origin and lie between 8°S, 140°W and 10°S, 138°W on the southeastern edge of Oceania, 475 km north of the Tuamotu Islands, 1050 km east of the Caroline Islands, and 4700 km west of the American mainland. The shore fish fauna has distinctly lower diversity than the rest of Polynesia (Randall, 2001). In contrast to the widely recognized IPB, which did not impede dispersal of either species of *Lutjanus*, the barrier at the Marquesas proved to be substantial. With no obvious geographic obstacles between the Marquesas and other South Pacific archipelagos, what could have produced the high level of

population differentiation in the Marquesas indicated by its 11.6% endemism rate, and the occurrence of strong genetic differentiation in three of five species of widely distributed reef fishes?

There have been three previous genetic studies of Marquesan shore-fishes, with one indicating strong differentiation of the Marquesan population (Planes & Fauvelot, 2002), and two others indicating no differentiation (Craig *et al*., 2007; Schultz *et al*., 2007). Our data show a strong phylogeographic break between the Marquesas and populations scattered throughout the rest of the range in both *Lutjanus* species. Those shared breaks stand in contrast to the very different patterns of intraspecific population structure across the central west Pacific in these species: very little structure in *L. kasmira* versus strong structure in *L. fulvus*. Yet both strongly and weakly dispersive snappers demonstrate an evolutionary genetic break at the Marquesas.

Randall (2001) attributed the endemism of the Marquesas to a combination of geographic isolation (a biogeographical barrier to dispersal) and unusually variable sea temperatures for an equatorial archipelago (ecological distinctiveness). However, the extensive Tuamotu archipelago is <500 km away from the Marquesas, a relatively small distance for a highly dispersive species such as *L. kasmira*, which shows strong genetic connectivity across much greater distances elsewhere throughout the Indo-central Pacific.

Directionality of the prevailing ocean current may amplify the effects of distance in the case of the Marquesas. The Southern Equatorial Current (SEC) flows from east to west between 4°N and 17°S (Wyrtki & Kilonsky, 1983; Bonjean & Lagerloef, 2002), with the Marquesas located on the edge of the strongest portion of that current (Fig. 2.4). The SEC originates 6000 km to the east of the Marquesas, where it draws water from an area with a temperate fish fauna. This combination of vast distance and inappropriate source fauna effectively isolates the

Marquesas from the east. The most likely direction of gene flow is from the Marquesas westwards. Our data indicate such directionality in gene flow in *L. fulvus*: the single population downstream from the Marquesas that we were able to sample (Kanton Island, 3500 km from the Marquesas) was the only location that shared haplotypes with the Marquesan population. The lack of Marquesan haplotypes of both *L. fulvus* and *L. kasmira* at two sites that are much closer but not downstream from the Marquesas, (Moorea to the south and Kiritimati to the north) is consistent with such a pattern of dispersal. Unfortunately we were unable to sample *L. kasmira* at Kanton. Further sampling of both species is needed at the reef systems nearest the Marquesas (the Tuamotus to the south and the Caroline Islands to the west) to test our hypothesis about the SEC, and establish the extent of any "leakage" from the Marquesas.

Unusual local environmental conditions may reinforce the biogeographical isolation of the Marquesan fish fauna (see Rocha & Bowen, 2008). Unlike the nearest Polynesian islands to the southwest, the Marquesas include an upwelling zone, have highly variable seawater temperatures, and little coral reef development. Those conditions may reduce the viability of colonists leaving or entering the Marquesas. Such a pattern of varying success has been documented in Atlantic reef fishes (Rocha *et al*., 2005). Thus, both a significant, although localized biogeographical barrier (distance enhanced by current directionality) and strongly divergent environmental conditions on each side of that barrier may have contributed to the genetic distinctiveness of Marquesan shore fish populations and the high endemism at those islands.

Glacial refugia and rapid expansions

Our cyt*b* data for both species exhibit low nucleotide diversity, high to moderate haplotype diversity (Tables 2.1, 2.4) and a star-shaped parsimony network (Fig. 2.2a, b), a

common mtDNA pattern in marine fishes (Grant & Bowen, 1998). These patterns are characteristics of a historical expansion in population size, a hypothesis supported by the unimodal mismatch distribution (Fig. 2.3a, b), the negative Fu's *F* values and the θ_0 and θ_1 values reported in Tables 2.1 and 2.4. This pattern may be a signature of severe population reductions attributed to the dramatic decreases in sea level during the Pleistocene period (Grant & Bowen, 1998). Rapid re-colonization of Oceania following sea level rise is evidenced by the low level of endemism in the region (Briggs, 1999) and a lack of intraspecific population structure across the central west Pacific in a diversity of marine organisms including snappers (*L. kasmira*; this study, excluding the Marquesas), trumpetfish (*Aulostomus chinensis*; Bowen *et al*., 2001), soldierfish (*Myripristis berndti*; Craig *et al*., 2007), angelfish (*Centropyge loriculus*; Schultz *et al*., 2007), parrotfish (*Chlorurus sordidus*; Bay *et al*., 2004), surgeonfish (*Naso* spp.; Klanten *et al*., 2007, Horne *et al*., 2008), sea urchins (Lessios *et al*., 2001), and sea stars (*Linkia laevigata*; Williams & Benzie, 1998). These data indicate that for many marine species the stretches of open water throughout the central and western Pacific Ocean are not significant barriers to gene flow. Dispersal may be facilitated by the relatively short distances between islands in that area (Keeney & Heist, 2006; Schultz *et al*., 2008). Departures from this pattern of high connectivity include snappers (*L. fulvus*; this study), damselfish (*Dascyllus trimaculatus*; Bernardi *et al*., 2001, *D. albisella*; Ramon *et al*., 2008), surgeonfish (*Acanthurus triostegus*; Planes & Fauvelot, 2002), and sea urchins (*Tripneustes gratilla*; Lessios *et al*., 2003, *Echinometra mathaei*; Palumbi *et al*., 1997) and are likely due to (unidentified) taxon specific factors.

Our *L. kasmira* data set show no evidence of genetic partitioning within the Indian Ocean, a result similar to that of populations of *Myripristis berndti* (Craig *et al*., 2007) and *Naso unicornis* (Horne *et al*., 2008) in which no significant structure in the mitochondrial genome was

found between Cocos-Keeling and the Seychelles. In contrast, populations of *N. brevirostris* from theses two locations demonstrated low but significant genetic structure (Horne *et al*., 2008). Because so few studies have sampled more than a few locations in the Indian Ocean, it is difficult to generalize about dispersal patterns in this ocean but initial data sets indicate high levels of gene flow, at least among reef fish populations.

Coalescence times for *L. kasmira* (Population age, Table 2.1) show a trend of decreasing age from west to east, with the oldest and most genetically diverse populations at the opposite ends of the species range (the Seychelles and Sodwana Bay in the western Indian Ocean and the Marquesas in the central Pacific). While our data are not conclusive, lines of evidence including coalescence times (Table 2.1), genetic diversity indices (Table 2.1 and multi-locus data from Table 2.3) and the mtDNA statistical parsimony network (Fig. 2.2a) all support a scenario of glacial refugia at the ends of the range with the remaining populations undergoing dramatic reductions or extirpations. As the glaciers retreated and sea level rose, the eastern Indian and Pacific Oceans seem to have been replenished from the western Indian Ocean. Mitochondrial sequence divergences indicate that the Marquesan population remained isolated from the rest of the species range throughout this period of geographic expansion. This pattern is of great interest because it indicates (1) that *L. kasmira* was able to expand eastward through the Indo-central Pacific (i.e. against the prevailing westward flows of major ocean currents); and (2) the mechanisms isolating the Marquesan population were strong enough to prevent it acting as a source for repopulation of depleted reefs elsewhere in the central Pacific.

Conclusions

Our genetic survey of the snappers *Lutjanus kasmira* and *L. fulvus* revealed a strongly contrasting pattern of population subdivision in these two species. While *L. kasmira*

demonstrated no population structure across the IPB, the Indian Ocean, and (most of the) Pacific Ocean, *L. fulvus* proved to be a highly structured species throughout the same range, with population structure across the IPB concordant with other pairwise comparisons. Despite differences in population structure, both fish demonstrate a remarkably strong phylogeographic break at the Marquesas Islands. Isolation of Marquesan populations may have arisen through a biogeographical barrier to inward dispersal (distance and contrarian ocean currents) and an unusual local environment that inhibits the survival of propagules from outside the archipelago. Coalescence analyses for *L. kasmira* prompt the hypothesis of glacial refugia in the western Indian Ocean and the Marquesas, with range expansion from eastern Africa through the Indo-central Pacific reinforcing the view that strong and enduring mechanisms have isolated the Marquesan population over evolutionary time.

REFERENCES

Allen, G.R. & Talbot, J.H. (1985) Review of the snappers of the genus *Lutjanus* (Pisces: Lutjanidae) from the Indo-Pacific, with the description of a new species. *Indo-Pacific Fishes*, **11**, 1-87.

Avise, J.C. (2004) *Molecular markers, natural history, and evolution*, 2nd edn. Sinauer Associates, Inc., Sunderland, MA.

Ayre, D.J. & Hughes, T.P. (2004) Climate change, genotypic diversity and gene flow in reef-building corals. *Ecology Letters*, **7**, 273-278.

Banford, H.M., Bermingham, E., Collette, B.B. & McCafferty, S.S. (1999) Phylogenetic systematics of the *Scomberomorus regalis* (Teleostei: Scombridae) species group: molecules, morphology and biogeography of spanish mackerels. *Copeia*, **3**, 596-613.

Barber, P.H., Erdmann, M.V. & Palumbi, S.R. (2006) Comparative phylogeography of three codistributed stomatopods: origins and timing of regional lineage diversification in the Coral Triangle. *Evolution*, **60**, 1825-1839.

Barber, P.H., Palumbi, S.R., Erdmann, M.V. & Moosa, M.K. (2000) A marine Wallace's line? *Nature*, **406**, 692-693.

Bay, L.K., Choat, J.H., van Herwerden, L. & Robertson, D.R. (2004) High genetic diversities and complex genetic structure in an Indo-Pacific tropical reef fish (*Chlorurus sordidus*): evidence of an unstable evolutionary past? *Marine Biology*, **144**, 757-767.

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

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

Benzie, J.A.H. & Stoddart, J.A. (1992) Genetic structure of crown-of-thorns starfish *Acanthaster planci* in Australia. *Marine Biology*, **112**, 631-639.

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

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

Bowen, B.W., Bass, A.L., Rocha, L.A., Grant, W.S. & Robertson, D.R. (2001) Phylogeography of the trumpetfishes (*Aulostomus*): ring species complex on a global scale. *Evolution*, **55**, 1029-1039.

Bowen, B.W., Bass, A.L., Soares, L. & Toonen, R.J. (2005) Conservation implications of complex population structure: lessons from the loggerhead turtle (*Caretta caretta*). *Molecular Ecology*, **14**, 2389-2402.

Briggs, J.C. (1974) *Marine zoogeography*. McGraw-Hill, New York.

Briggs, J.C. (1999) Extinction and replacement in the Indo-West Pacific Ocean. *Journal of Biogeography*, **26**, 777-783.

Brown, J.H. & Lomolino, M.V. (1998) *Biogeography*, 2nd edn. Sinauer Associates, Sunderland, MA.

Castro, A.L.F., Stewart, B.S., Wilson, S.G., Hueter, R.E., Meekan, M.G., Motta, P.J., Bowen, B.W. & Karl, S.A. (2007) Population genetic structure of Earth's largest fish, the whale shark (*Rhincodon typus*). *Molecular Ecology*, **16**, 5183-5192.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genalogies. *Molecular Ecology*, **9**, 1657-1659.

Craig, M.T., Eble, J.A., Bowen, B.W. & Robertson, D.R. (2007) High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae). *Marine Ecology Progress Series*, **334**, 245-254.

Denit, K. & Sponaugle, S. (2004) Growth variation, settlement, and spawing of gray snapper across a latitudinal gradient. *Transactions of the American Fisheries Society*, **133**, 1339-1355.

Duda, T.F., Jr. & Palumbi, S.R. (1999) Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations. *Marine Biology*, **134**, 705-710.

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

Fleminger, A. (1986) The Pleistocene equatorial barrier between the Indian and Pacific Oceans and a likely cause for Wallace's line. *UNESCO Technical Papers in Marine Science*, **49**, 84-97.

Floeter, S.R., Rocha, L.A., Robertson, D.R., Joyeux, J.C., Smith-Vaniz, W.F., Wirtz, P., Edwards, A.J., Barreiros, J.P., Ferreira, C.E.L., Gasparini, J.L., Brito, A., Falcón, J.M., Bowen, B.W. & Bernardi, G. (2008) Atlantic reef fish biogeography and evolution. *Journal of Biogeography*, **35**, 22-47.

Friedlander, A.M., Parrish, J.D. & DeFelice, R.C. (2002) Ecology of the introduced snapper *Lutjanus kasmira* (Forsskal) in the reef fish assemblage of a Hawaiian bay. *Journal of Fish Biology*, **60**, 28-48.

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

Galloway, R.W. & Kemp, E.M. (1981) Late Cenozoic environments in Australia. *Ecological Biogeography of Australia*, 51-80.

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

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

Hassan, M., Lemaire, C., Fauvelot, C. & Bonhomme, F. (2002) Seventeen new exon-primed intron-crossing polymerase chain reaction amplifiable introns in fish. *Molecular Ecology Notes*, **2**, 334-340.

Horne, J.B., van Herwerden, L., Choat, H.J. & Robertson, D.R. (2008) High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners. *Molecular Phylogenetics and Evolution*, **49**, 629-638.

Hobbs, J.-P.A., Frisch, A.J., Allen, G.R. & Van Herwerden, L. (2009) Marine hybrid hotspot at Indo-Pacific biogeographic border. *Biology Letters*, **5**, 258-261.

Hudson, R.R. & Turelli, M. (2003) Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **51**, 182-190.

Jarman, S.N., Ward, R.D. & Elliott, N.G. (2002) Oligonucleotide primers for PCR amplification of coelomate introns. *Marine Biotechnology*, **4**, 347-355.

Keeney, D.B. & Heist, E.J. (2006) Worldwide phylogeography of the blacktip shark (*Carcharhinus limbatus*) inferred from mitochondrial DNA reveals isolation of western Atlantic populations coupled with recent Pacific dispersal. *Molecular Ecology*, **15**, 3669-3679.

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

Lacson, J.M. & Clark, S. (1995) Genetic divergence of Maldivian and Micronesian demes of the damselfishes *Stegastes nigricans, Chrysiptera biocellata, C. glauca* and *C. leucopoma* (Pomacentridae). *Marine Biology*, **121**, 585-590.

Lavery, S., Moritz, C. & Fielder, D.R. (1995) Changing patterns of population structure and gene flow at different spatial scales in *Birgus latro* (the coconut crab). *Heredity*, **74**, 531-541.

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

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

Lessios, H.A. & Robertson, D.R. (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.

Lessios, H.A., Kessing, B.D. & Pearse, J.S. (2001) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution*, **55**, 955-975.

Lessios, H.A., Kane, J. & Robertson, D.R. (2003) Phylogeography of the pantropical sea urchin *Tripneustes*: Contrasting patterns of population structure between oceans. *Evolution*, **57**, 2026-2036.

Lu, C.-P., Chen, C.A., Hui, C.-F., Tzeng, T.-D. & Yeh, S.-Y. (2006) Population genetic structure of the swordfish, *Xiphias gladius* (Linnaeus, 1758), in the Indian Ocean and West Pacific inferred from the complete DNA sequence of the mitochondrial control region. *Zoological Studies*, **45**, 269-279.

Maddison, D.R. & Maddison, W.P. (2002) *MacClade 4: analysis of phylogeny and character evolution.* Sinauer Associates, Sunderland, MA.

McMillan, W.O. & Palumbi, S.R. (1995) Concordant evolutionary patterns among Indo-West Pacific butterflyfishes. *Proceedings of the Royal Society B: Biological Sciences,* **260**, 229-236.

Menezes, M.R., Ikeda, M. & Taniguchi, N. (2006) Genetic variation in skipjack tuna *Katsuwonus pelamis* (L.) using PCR-RFLP analysis of the mitochondrial DNA D-loop region. *Journal of Fish Biology,* **68**, 156-161.

Meyer, A. (1994) Shortcomings of the cytochrome *b* gene as a molecular marker. *Trends in Ecology and Evolution*, **9**, 278-280.

Miller, T.L. & Cribb, T.H. (2007) Phylogenetic relationships of some common Indo-Pacific snappers (Perciformes: Lutjanidae) based on mitochondrial DNA sequences, with comments on the taxonomic position of the Caesioninae. *Molecular Phylogenetics and Evolution,* **44**, 450-460.

Möller, L.M. & Beheregaray, L.B. (2004) Genetic evidence for sex-biased dispersal in resident bottlenose dolphins (*Tursiops aduncus*). *Molecular Ecology*, **13**, 1607-1612.

Morales-Nin, B. & Ralston, S. (1990) Age and growth of *Lutjanus kasmira* (Forskål) in Hawaiian waters. *Journal of Fish Biology*, **36**, 191-203.

Naish, T., Powell, R., Levy, R., Wilson, G., Scherer, R., Talarico, F., Krissek, L., Niessen, F., Pompilio, M., Wilson, T., Carter, L., DeConto, R., Huybers, P., McKay, R., Pollard, D., Ross, J., Winter, D., Barrett, P., Browne, G., Cody, R., Cowan, E., Crampton, J., Dunbar, G., Dunbar, N., Florindo, F., Gebhardt, C., Graham, I., Hannah, M., Hansaraj, D., Harwood, D., Helling, D., Henrys, S., Hinnov, L., Kuhn, G., Kyle, P., Läufer, A., Maffioli, P., Magens, D., Mandernack, K., McIntosh, W., Millan, C., Morin, R., Ohneisher, C., Paulsen, T., Persico, D., Raine, I., Reed, J., Riesselman, C., Sagnotti, L.,

Schmitt, D., Sjunneskog, C., Strong, P., Taviani, M., Vogel, S., Wilch, T. & Williams, T. (2009) Obliquity-paced Pliocene West Atlantic ice sheet oscillations. *Nature*, **458**, 322-328.

Nei, M. (1987) *Molecular evolutionary genetics*. Columbia University Press, New York.

Oda, D.K. & Parrish, J.D. (1987) Trophic ecology of the introduced ta'ape (*Lutjanus kasmira*) in Hawaii. *Bulletin of Marine Science*, **41**, 639-639.

Palumbi, S.R., Grabowsky, G., Duda, T., Geyer, L. & Tachino, N. (1997) Speciation and population genetic structure in tropical Pacific sea urchins. *Evolution*, **51**, 1506-1517.

Pardini, A.T., Jones, C.S., Noble, L.R., Kreiser, B., Hamish, M., Bruce, B.D., Stevens, J.D., Cliff, G., Scholl, M.C., Francis, M., Duffy, C.A.J. & Martin, A.P. (2001) Sex-biased dispersal of great white sharks. *Nature*, **412**, 139-140.

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

Posada, D. & Crandall, K.A. (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

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

Randall, J.E. (1987) Introductions of marine fishes to the Hawaiian Islands. *Bulletin of Marine Science*, **41**, 490-502.

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

Randall, J.E. (2001) Four new cardinalfishes (Perciformes: Apogonidae) from the Marquesas Islands. *Pacific Science*, **55**, 47-64.

Randall, J.E. (2007) *Reef and shore fishes of the Hawaiian Islands*. Sea Grant College Program, University of Hawaii, Honolulu.

Rangarajan, K. (1971) Maturity and spawning of the snapper, *Lutjanus kasmira* (Forskål) from the Andaman Sea. *Indian Journal of Fisheries*, **18**, 114-125.

Robertson, D.R., Grove, J.S. & McCosker, J.E. (2004) Tropical transpacific shorefishes. *Pacific Science*, **58**, 507-565.

Rocha, L.A. & Bowen, B.W. (2008) Speciation in coral-reef fishes. *Journal of Fish Biology*, **72**, 1101-1121.

Rocha, L.A., Bass, A.L., Robertson, D.R. & Bowen B.W. (2002) Adult habitat preferences, larval dispersal, and comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology*, **11**, 243-252.

Rocha, L.A., Robertson, D.R., Roman, J. & Bowen, B.W. (2005) Ecological speciation in tropical reef fishes. *Proceedings of the Royal Society Biological Sciences Series B*, **272**, 573-579.

Rocha, L.A., Craig, M.T. & Bowen, B.W. (2007) Phylogeography and the conservation of coral reef fishes. *Coral Reefs*, **26**, 501-512.

Schneider, S., Roessli, D. & Excoffier, L. (2000) *Arlequin: a software for population genetics data analysis ver 2.000*. *Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.*

Schultz, J.K., Pyle, R.L., DeMartini, E. & Bowen, B.W. (2007) Genetic connectivity among color morphs and Pacific archipelagos for the flame anglefish, *Centropyge loriculus*. *Marine Biology*, **151**, 167-175.

Schultz, J.K., Feldheim, K.A., Gruber, S.H., Ashley, M.V., McGovern, T.M. & Bowen, B.W. (2008) Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*). *Molecular Ecology*, **17**, 5336-5348.

Seutin, G., White, B.N. & Boag, P.T. (1991) Preservation of avian and blood tissue samples for DNA analyses. *Canadian Journal of Zoology*, **69**, 82-92.

Stepien, C.A., Randall, J.E. & Rosenblatt, R.H. (1994) Genetic and morphological divergence of a circumtropical complex of goatfishes: *Mulloidichthys vanicolensis*, *M. dentatus* and *M. martinicus*. *Pacific Science*, **48**, 44-56.

Suzuki, K. & Hioki, S. (1979) Spawning behavior, eggs, and larvae of the lutjanid fish, *Lutjanus kasmira*, in an aquarium. *Japanese Journal of Ichthyology*, **26**, 161-165.

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.

Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*, **10**, 512-526.

Theisen, T.C., Bowen, B.W., Lanier, W. & Baldwin, J.D. (2008) High connectivity on a global scale in the pelagic wahoo, *Acanthacybrium solandri* (tuna family Scombridae). *Molecular Ecology*, **17**, 4233-4247.

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

Williams, S.T. & Benzie, J.A.H. (1998) Evidence of a biogeographic break between populations of a high dispersal starfish: congruent regions within the Indo-West Pacific defined by color morphs, mtDNA, and allozyme data. *Evolution*, **52**, 87-99.

Winans, G.A. (1980) Geographic variation in the milkfish *Chanos chanos* I. Biochemical evidence. *Evolution*, **34**, 558-574.

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

Wyrtki, K. & Kilonsky, B. (1983) Mean water and current structure during the Hawaii-to-Tahiti shuttle experiment. *Journal of Physical Oceanography*, **14**, 242-254.

Zapata, F.A. & Herron, P.A. (2002) Pelagic larval duration and geographic distribution of tropical eastern Pacific snappers (Pisces: Lutjanidae). *Marine Ecology Progress Series*, **230**, 295-300.

Table 2.1. Molecular diversity indices and coalescence times for cytochrome *b* sequences from twelve populations of *Lutjanus kasmira*. Sample location, number of individuals sequenced (*n*), number of haplotypes (N_h), haplotype diversity (*h*), nucleotide diversity (π), Fu's *F* value and corresponding *P*-value in parenthesis, Harpending's raggedness indices (*r*) with corresponding *P*-value in parentheses, and mismatch distribution parameters τ, θ_0 and θ_1 as reported by ARLEQUIN 3.11 are listed.

Sample Location	*n*	N_h	*h*	π	Fu's F_s	*r*	τ	Population age (years)	θ_0	N_{ef0} (x 10^3)	θ_1	N_{ef1} (x 10^3)
Pacific Ocean												
Marquesas	47	19	0.84 ± 0.04	0.004 ± 0.003	-12.22 (<0.001)	0.046 (0.54)	1.75	180,000–370,000	0.00	0.0	824.40	23, 000–47,000
Moorea	49	11	0.54 ± 0.08	0.002 ± 0.001	-5.87 (<0.001)	0.075 (0.85)	1.07	110,000–230,000	0.33	9–19	1.23	40–70
Kiritimati	49	14	0.69 ± 0.07	0.003 ± 0.002	-7.88 (<0.001)	0.038 (0.90)	1.29	140,000–270,000	0.00	0.0	7.62	220–430
Fiji	31	9	0.50 ± 0.11	0.001 ± 0.001	-4.03 (<0.001)	0.086 (0.83)	0.77	80,000–160,000	0.25	7–14	1.52	40–90
Lizard Island	5	1	0.00 ± 0.00	0.000 ± 0.000	-	-	-	-	-	-	-	-
Guam	50	15	0.59 ± 0.08	0.002 ± 0.002	-8.13 (<0.001)	0.046 (0.91)	0.30	30,000–60,000	1.02	29–58	2.66	80–150
Okinawa	2	1	0.00 ± 0.00	0.000 ± 0.000	-	-	-	-	-	-	-	-
Indian Ocean												
Cocos-Keeling	35	12	0.54 ± 0.10	0.002 ± 0.002	-5.03 (<0.001)	0.092 (1.00)	∞	∞	0.00	0.0	∞	∞
Christmas Island	15	8	0.84 ± 0.09	0.003 ± 0.002	-0.84 (0.004)	0.051 (0.75)	0.91	96,000–190,000	0.90	26–51	∞	∞
Diego Garcia	34	15	0.88 ± 0.05	0.005 ± 0.003	-4.39 (<0.001)	0.069 (0.19)	2.05	210,000–430,000	0.71	19–40	15.45	440–880
Seychelles	19	10	0.74 ± 0.11	0.004 ± 0.003	-3.35 (<0.001)	0.065 (0.64)	3.06	320,000–640,000	0.00	0.0	3.45	100–200
Sodwana Bay	34	14	0.82 ± 0.06	0.004 ± 0.003	-3.81 (<0.001)	0.088 (0.22)	2.81	300,000–590,000	0.01	0.0	6.50	180–400
All samples	370	83	0.74 ± 0.03	0.004 ± 0.002	-26.15 (<0.001)	0.013 (0.99)	3.23	340,000–680,000	0.00	0.0	3.05	90–180
All samples except the Marquesas	323	64	0.67 ± 0.03	0.003 ± 0.002	-26.77 (<0.001)	0.034 (0.91)	2.27	240,000–480,000	0.00	0.0	2.14	60–120

Table 2.2. Pairwise *F* statistics for ten populations of *Lutjanus kasmira*. Pairwise Φ_{ST} values for cytochrome *b* data are below diagonal and pairwise F_{ST} values for the multi-locus nuclear data set are above diagonal. Values in bold are significant ($P < 0.05$).

Sample Location	Marquesas	Moorea	Kiritimati	Fiji	Guam	Cocos-Keeling	Christmas Island	Diego Garcia	Seychelles	Sodwana Bay
Marquesas	-	**0.559**	**0.549**	**0.492**	**0.533**	**0.496**	**0.498**	**0.527**	**0.441**	**0.530**
Moorea	**0.634**	-	-0.001	0.007	0.005	0.010	**0.048**	0.010	0.026	0.005
Kiritimati	**0.590**	**0.062**	-	-0.003	0.000	0.002	0.023	0.006	**0.026**	-0.001
Fiji	**0.618**	**0.049**	-0.006	-	-0.002	0.000	0.015	0.002	0.008	-0.002
Guam	**0.609**	**0.049**	-0.007	-0.011	-	0.003	0.014	-0.004	0.010	-0.009
Cocos-Keeling	**0.610**	**0.032**	**0.022**	0.002	0.004	-	0.010	-0.008	0.010	-0.004
Christmas Island	**0.564**	**0.051**	-0.007	-0.008	-0.005	0.017	-	0.016	0.021	0.018
Diego Garcia	**0.534**	**0.045**	**0.036**	**0.022**	**0.024**	0.001	0.006	-	0.013	-0.008
Seychelles	**0.543**	**0.045**	0.018	0.004	0.011	0.001	-0.005	-0.015	-	0.007
Sodwana Bay	**0.538**	**0.043**	**0.032**	**0.018**	**0.014**	0.002	0.008	-0.003	-0.005	-

Table 2.3. Molecular diversity indices and coalescence times for eleven populations of *Lutjanus kasmira* for the growth hormone (GH) and adenine nucleotide transporter translocase (ANT) nuclear introns. Sample location, number of individuals sequenced (n), number of alleles (N_a), observed heterozygosity (H_O), and expected heterozygosity (H_E) are listed for each intron. P-values are the result of exact tests for Hardy-Weinberg equilibrium using a Markov chain with 100,000 steps in ARLEQUIN 3.11. Number of genotypes (N_g) and mean H_E are listed for the multi-locus data set which includes only those individuals that amplified at both loci.

Sample	GH Intron					ANT Intron					Multi-locus		
Location	n	N_a	H_O	H_E	P-value	n	N_a	H_O	H_E	P-value	n	N_g	H_E
Pacific Ocean													
Marquesas	49	3	0.66	0.59	0.58	49	6	0.29	0.38	0.04	48	12	0.48
Moorea	50	3	0.10	0.10	1.00	47	5	0.32	0.38	0.03	47	8	0.24
Kiritimati	50	3	0.06	0.06	1.00	48	10	0.44	0.46	0.07	48	12	0.26
Fiji	31	3	0.13	0.12	1.00	29	8	0.64	0.55	0.90	28	9	0.34
Lizard Island	5	1	0.00	0.00	1.00	4	2	0.50	0.43	1.00	4	2	0.21
Guam	49	2	0.04	0.04	1.00	35	9	0.40	0.52	0.00	35	10	0.27
Okinawa	2	1	0.00	0.00	1.00	2	3	1.00	0.83	1.00	2	2	0.42
Indian Ocean													
Cocos-Keeling	30	5	0.13	0.19	0.06	25	10	0.40	0.51	0.10	25	13	0.33
Christmas Island	14	1	0.00	0.00	1.00	15	8	0.40	0.58	0.08	14	8	0.31
Diego Garcia	33	3	0.06	0.12	0.03	32	10	0.48	0.55	0.03	25	8	0.27
Seychelles	20	2	0.15	0.22	0.25	19	8	0.68	0.59	0.64	19	10	0.41
Sodwana Bay	36	3	0.06	0.06	1.00	34	11	0.47	0.51	0.51	34	12	0.29
All samples	369	7	0.16	0.25	<0.001	336	22	0.43	0.61	<0.001	329	34	0.43

Table 2.4. Molecular diversity indices and coalescence times for cytochrome *b* sequences from six populations of *Lutjanus fulvus*. Sample location, number of individuals (*n*), number of haplotypes (N_h), haplotype diversity (*h*), nucleotide diversity (π), Fu's *F* value and corresponding *P*-value in parenthesis, Harpending's raggedness indices (*r*) with corresponding *P*-value in parenthesis, and mismatch distribution parameters τ, θ_0 and θ_1 as reported by ARLEQUIN 3.11 are listed.

Sample Location	*n*	N_h	*h*	π	Fu's F_s	*r*	τ	Population age (years)	θ_0	N_{ef0} (x 10^3)	θ_1	N_{ef1} (x 10^3)
Pacific Ocean												
Marquesas	48	7	0.72 ± 0.04	0.002 ± 0.002	-2.06 (0.003)	0.142 (0.02)	1.15	120,000–240,000	0.00	0.0	∞	∞
Moorea	48	4	0.12 ± 0.06	0.0003 ± 0.001	-4.22 (<0.001)	0.588 (0.72)	3.00	300,000–630,000	0.00	0.0	0.15	4–8
Kiritimati	15	2	0.34 ± 0.13	0.001 ± 0.001	0.60 (0.94)	0.216 (0.39)	0.46	50,000–100,000	0.00	0.0	∞	∞
Kanton	46	10	0.66 ± 0.07	0.006 ± 0.003	-1.31 (0.002)	0.070 (0.68)	7.11	740,000–1,480,000	0.00	0.0	2.16	60–120
Philippines	37	9	0.66 ± 0.08	0.003 ± 0.002	-2.95 (<0.001)	0.042 (0.88)	2.10	220,000–440,000	0.00	0.0	2.34	70–130
Indian Ocean												
Cocos-Keeling	9	4	0.78 ± 0.11	0.003 ± 0.002	2.57 (0.139)	0.015 (0.44)	12.08	1,200,000–2,500,000	0.00	0.0	4.46	125–250
All samples	203	23	0.69 ± 0.03	0.006 ± 0.003	-5.85 (<0.001)	0.067 (0.53)	7.00	730,000–1,460,000	0.00	0.0	3.02	85–170
All samples except the Marquesas	155	18	0.51 ± 0.05	0.002 ± 0.002	-8.59 (<0.001)	0.010 (1.00)	∞	∞	0.00	0.0	533.95	15,000–30,000

Table 2.5. Pairwise *F* statistics for six populations of *Lutjanus fulvus.* Pairwise Φ_{ST} values for cytochrome *b* data are below diagonal and corresponding *P*-values are above diagonal. Values in bold are significant ($P < 0.05$).

Sample location	Marquesas	Moorea	Kiritimati	Kanton	Philippines	Cocos-Keeling
Marquesas	-	<0.001	<0.001	<0.001	<0.001	<0.001
Moorea	**0.912**	-	0.011	<0.001	0.002	<0.001
Kiritimati	**0.873**	**0.157**	-	0.073	0.205	0.010
Kanton	**0.660**	**0.153**	**0.086**	-	0.004	0.050
Philippines	**0.825**	**0.071**	0.022	**0.104**	-	0.533
Cocos-Keeling	**0.856**	**0.300**	**0.153**	0.114	-0.014	-

Table 2.6. Genetic surveys of population structure in reef organisms across the Indo-Pacific Barrier. Species scientific name, marker type, genetic structure between the Pacific and Indian Oceans (as reported), and references are listed.

Species	Marker	Genetic Structure	Reference
oldierfish:			
Myripristies berndti	mtDNA	Φ_{ST} = 0.58 (P < 0.001)	Craig *et al*., 2007
Trumpetfish:			
Aulostomus chinensis	mtDNA	Φ_{ST} = -0.027 (P = 0.805) not reported in reference	Bowen *et al.*, 2001
Surgeonfish:			
Acanthurus triostegus	Allozymes	F_{ST} = 0.21-0.25 (P < 0.05)	Planes & Fauvelot, 2002
Naso vlamingii	mtDNA	Φ_{ST} = 0.077 (P < 0.05)	Klanten *et al.*, 2007
Naso unicornis	mtDNA	Φ_{ST} = 0.018 (P = 0.02)	Horne *et al*., 2008
Naso brevirostris	mtDNA	Φ_{ST} = 0.030 (P = 0.08)	
Damselfish:			
Stegastes nigricans			
Chrysiptera biocellata	allozymes	Fixed allele differences	Lacson & Clark, 1995
Chrysiptera glauca			
Chrysiptera leucopoma			
Parrotfish:			
Chlorurus sordidus	mtDNA	Fixed haplotype differences	Bay *et al*., 2004
Seastar:			
Linkia laevigata	mtDNA	Φ_{ST} = 0.332 (P < 0.001)	Williams & Benzie, 1998
	allozymes	F_{ST} = 0.083 (P < 0.001)	
Acanthaster planci	allozymes	F_{ST} = 0.273 (P < 0.001)	Benzie, 1999
Sea urchins:			
Tripneustes gratilla	mtDNA	Φ_{ST} = -0.01–0.40 (P < 0.05 for 15 out of 32 pairwise comparisons)	Lessios *et al*., 2003
Diadema savignyi	isozymes	F_{ST} = -0.01–0.19 (P < 0.05 for 2 out of 21 pairwise comparisons)	Lessios *et al.*, 2001
Diadema setosum		F_{ST} = 0.36–0.73 (P < 0.05 for 16 out of 16 pairwise	
Coconut crab:			
Birgus latro	allozymes	F_{ST} = 0.082 (P < 0.001)	Lavery *et al.*, 1995, 1996
Tiger prawn:			
Penaeus monodon	nuclear intron	F_{ST} = 0.51 (P < 0.001)	Duda & Palumbi, 1999

Figure 2.1. Collection locations and sample sizes for *Lutjanus kasmira* (numbers) and *L. fulvus* (numbers in bold italics). Due to low sample size, locations marked with an asterisk (*) were excluded from population level analyses.

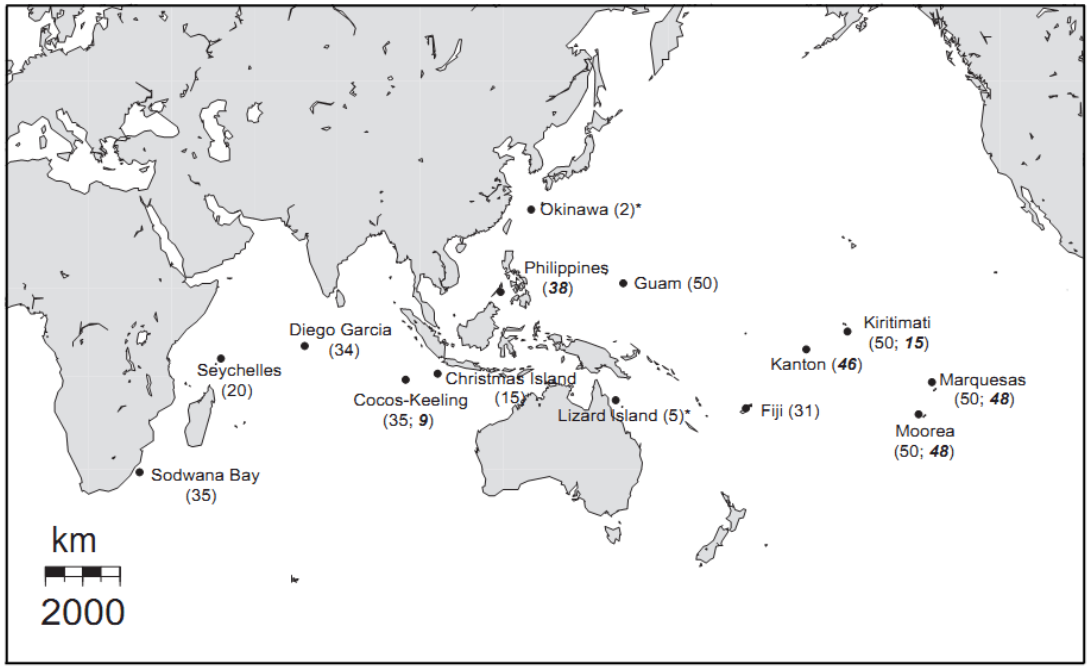


Figure 2.2. Statistical parsimony networks for *Lutjanus kasmira* (a, c, d) and *L. fulvus* (b) constructed using TCS 2.21 (Clement *et al*., 2000) for (a) 370 cytochrome *b* sequences (b) 203 cytochrome *b* sequences, (c) alleles at GH intron from 369 individuals, and (d) alleles at ANT intron for 336 individuals. Each circle represents one mitochondrial haplotype or nuclear allele with the area of each circle proportional to number of that particular haplotype or allele in the data set; dashes represent hypothetical haplotypes or alleles; colours represent collection location (see key).

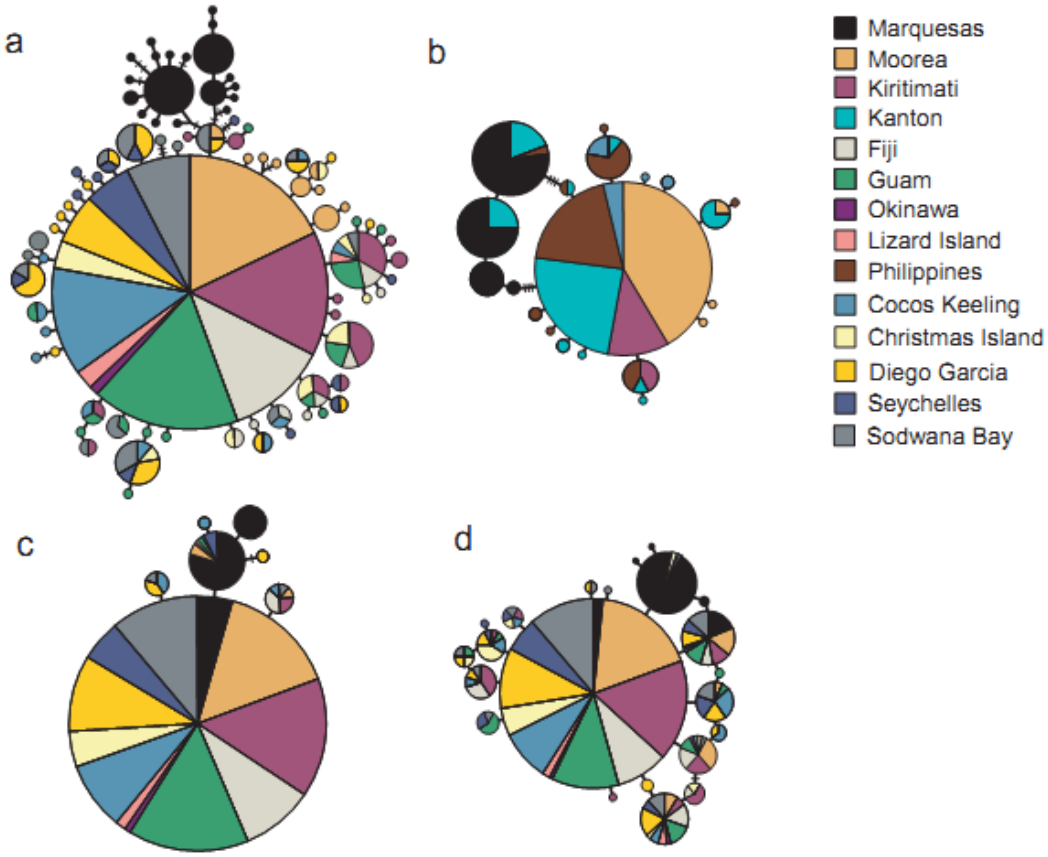


Figure 2.3. Mismatch distribution based on (a) 370 cytochrome *b* sequences from twelve populations of *Lutjanus kasmira*, (b) 203 cytochrome *b* sequences from six populations of *L. fulvus*. Black bars are the observed and white bars are the simulated pairwise differences as reported by ARLEQUIN 3.11. Harpending's raggedness indices and corresponding *P*-values for the complete data sets are shown.

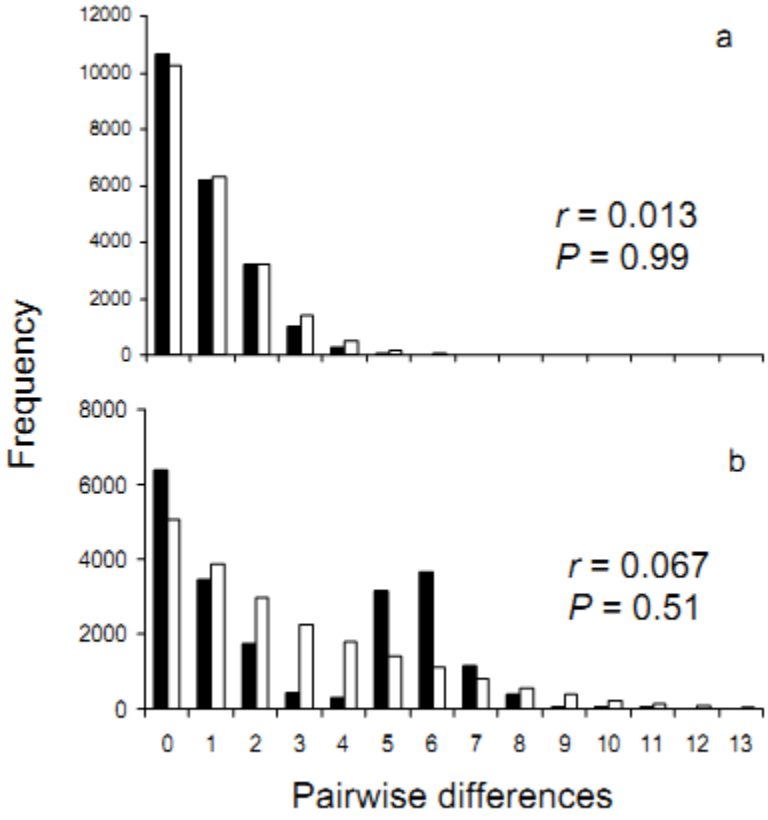
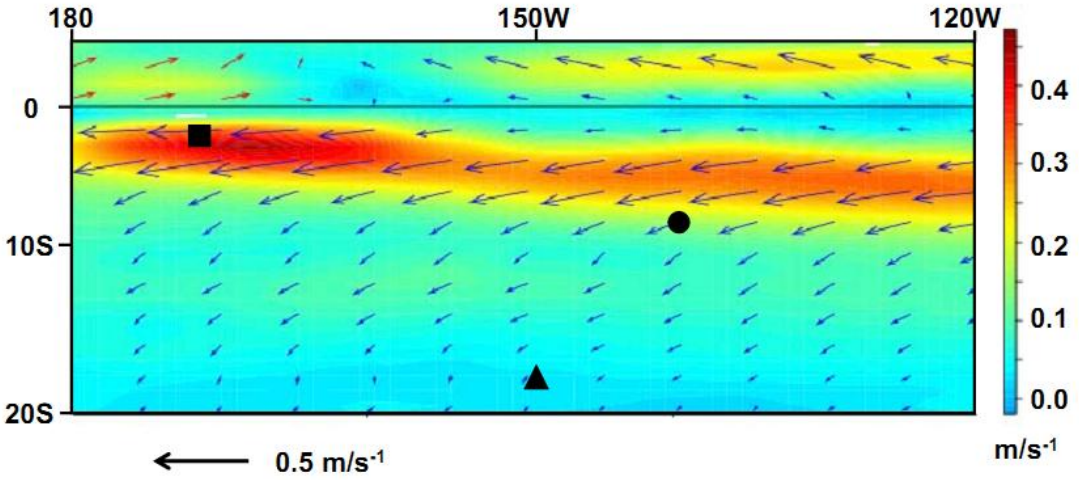


Figure 2.4. Mean ocean surface current vectors from May-September 1997-2007 when spawning in snappers is most likely to occur. Orientation of arrow indicates direction of current and length of arrow indicates current speed (0.514 m s^{-1} = 1 knot). Coloured scale bar represents mean current speed (m s^{-1}). Data obtained from http://www.oscar.noaa.gov; accessed 10 June 2009 (Bonjean & Lagerloef, 2002). Circle, Marquesas Island of Nuku Hiva; triangle, Moorea; square, Kanton Island.



CHAPTER THREE

Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle

Published as: Gaither, M.R., B.W. Bowen, T.-R. Bordenave, L.A. Rocha, S.J. Newman, J.A. Gomez, L. van Herwerden, M.T. Craig. 2011. Phylogeography of the reef fish *Cephalopholus argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle. BMC Evolutionary Biology 11:189 doi:10.1186/1471-2148-11-189

Abstract

Background: The Coral Triangle (CT), bounded by the Philippines, the Malay Peninsula, and New Guinea is the epicenter of marine biodiversity. Hypotheses that explain the source of this rich biodiversity include 1) the center of origin, 2) the center of accumulation, and 3) the region of overlap. Here we contribute to the debate with a phylogeographic survey of a widely distributed reef fish, the Peacock Grouper (*Cephalopholis argus*; Epinephelidae) at 21 locations (N = 550) using DNA sequence data from mtDNA cytochrome *b* and two nuclear introns (gonadotropin-releasing hormone and S7 ribosomal protein).

Results: Population structure was significant ($\Phi_{ST} = 0.297$, $P < 0.001$; $F_{ST} = 0.078$, $P < 0.001$; $F_{ST} = 0.099$, $P < 0.001$, for the three loci respectively) among five regions: French Polynesia, the central-west Pacific (Line Islands to northeastern Australia), Indo-Pacific boundary (Bali and Rowley Shoals), eastern Indian Ocean (Cocos/Keeling and Christmas Island), and western Indian Ocean (Diego Garcia, Oman, and Seychelles). A strong signal of isolation by distance was detected in both mtDNA (r = 0.749, $P = 0.001$) and the combined nuclear loci (r = 0.715, $P < 0.001$). We detected evidence of population expansion with migration toward the CT. Two clusters of haplotypes were detected in the mtDNA data ($d = 0.008$), corresponding to the Pacific and Indian Oceans, with a low level of introgression observed outside a mixing zone at the Pacific-Indian boundary.

Conclusions: We conclude that the Indo-Pacific Barrier, operating during low sea level associated with glaciation, defines the primary phylogeographic pattern in this species. These data support a scenario of isolation on the scale of 10^5 year glacial cycles, followed by population expansion toward the CT, and overlap of divergent lineages at the Pacific-Indian boundary. This pattern of isolation, divergence, and subsequent overlap likely contributes to species richness at the adjacent CT and is consistent with the region of overlap hypothesis.

Background

Current efforts to identify and preserve biodiversity are dependent upon our ability to locate hotspots and to understand how that diversity is generated. Conservation efforts must preserve not just standing biodiversity but also the mechanisms that produce it [1]. The Coral Triangle (CT), bounded by the Philippines, the Malay Peninsula, and New Guinea is the epicenter of marine biodiversity. Species diversity declines with distance from this region, both latitudinally and longitudinally, a pattern that applies to a broad array of taxa [2-8]. The generality of this pattern has led many to conclude that a common mechanism may be responsible for generating diversity in the CT. A number of hypotheses have been proposed to explain the source of the incredible number of species found in this region and these can be grouped into three categories: 1) center of origin, 2) center of accumulation, and 3) region of overlap.

The center of origin hypothesis was proposed by Ekman [9], who suggested that the CT is the primary source of biodiversity in the Indo-Pacific due to an unusually high rate of speciation in the region. He suggested that the decline in species richness with distance from the CT is an artifact of prevailing currents that impede outward dispersal [9]. The most common mechanism invoked to explain the proposed elevated speciation rate is the fracturing of populations as a result of the geological complexity of the region and eustatic sea level changes [10]. Others have suggested that increased rates of sympatric or parapatric speciation driven by different selection pressures in a heterogeneous environment could be contributing to the species richness of the CT [11,12]. Evidence for this argument includes the finding of fine scale population subdivisions within the CT [13-18].

In contrast, the center of accumulation hypothesis [19] proposes speciation in isolated peripheral locations with subsequent dispersal of novel taxa into the CT. The long history of the Pacific archipelagos, some of which date to the Cretaceous, and ocean current and wind patterns that favor dispersal toward the CT have been offered as a mechanism [19,20]. Finally, the region of overlap hypothesis [21] maintains that the high species diversity in the CT is due to the overlap of faunas from two biogeographic provinces: Indo-Polynesian and Western Indian Ocean [22]. The region roughly dividing these two provinces is west of the shallow Sunda and Sahul shelves of the East Indies. During the Pleistocene, sea level was as much as 130 m below present levels and produced a near continuous land bridge between Asia and Australia [23], greatly restricting dispersal between ocean basins in the region known as the Indo-Pacific Barrier (IPB). Isolation of conspecific populations across the IPB may have led to allopatric speciation and contributed to the distinction of the Pacific and Indian Ocean faunas. According to the region of overlap hypothesis, relaxation of the IPB following each Pleistocene glacial maximum has resulted in dispersal pathways between the Pacific and Indian Oceans with the CT representing the area of overlap between the two distinct biotas. The differences between the center of accumulation and region of overlap hypotheses are subtle. In both cases speciation occurs outside the CT with subsequent dispersal toward the CT. However, the region of overlap hypothesis is based on the premise that the isolating mechanism is the IPB with the faunas of the Pacific and Indian Oceans diverging during periods of restricted dispersal. In contrast, the center of accumulation hypothesis does not specify a mechanism of divergence nor is it associated with any biogeographic barrier. This hypothesis invokes speciation in peripheral locations, followed by dispersal to the CT on prevailing oceanic currents.

Contemporary species distributions are the most common line of evidence offered to examine these hypotheses yet no consensus has evolved. Mora et al. [7] examined the ranges of nearly 2,000 Indo-Pacific fishes and found that the midpoint of their ranges centered on the CT, a result they interpret as evidence for the center of origin hypothesis. Connolly et al. [24], using a midpoint domain model, found evidence for the accumulation of taxa in the CT due to species dispersing on oceanic gyres. Halas and Winterbottom [25] employed a novel phylogenetic approach to address the issue but found no conclusive evidence for any of the hypotheses. Evidence for a combined influence of all these processes in generating the high biodiversity in the CT has led many to conclude that the processes are not mutually exclusive and act simultaneously [3, 26-28].

Patterns of genetic variation in widely distributed species, while not often employed to address the source of biodiversity hotspots, provide a historical perspective that cannot be resolved with contemporary species distributions. Each hypothesis results in specific predictions about geographic positioning of new species and lineages within species [29]. The center of origin hypothesis predicts that the oldest populations (within new species) will be in the CT, possibly with decreasing haplotype diversity emanating from the center similar to the observed decline in species richness (sensu [30]). In contrast, the center of accumulation hypothesis predicts that the oldest populations (within new species) will be found peripheral to the CT accompanied with unidirectional dispersal toward the CT. Similar to the center of origin, the region of overlap hypothesis predicts that the most diverse (but not oldest) populations will be centered in the CT, however in this case the high diversity is the result of the overlap of divergent lineages from peripheral regions. While there have been a handful of intraspecific genetic studies that address the origin of diversity in the CT, the results are conflicting. Evidence

for the center of accumulation hypothesis has been found in the Lemon Damselfish (*Pomacentrus moluccensis*) [29] and the Yellow Tang (*Zebrasoma flavescens*) [31]. On the other hand, sea urchins [32] and wrasses [33] invoke a combination of the center of origin and the center of accumulation hypotheses. Of course all of these conclusions, including our own, are premised on the assumption that intraspecific genetic divergences translate into macroevolutionary (interspecific) partitions [34].

Here we contribute a range-wide phylogeographic study of a widely distributed grouper to test competing hypotheses concerning the origins of biodiversity in the CT. The Peacock Grouper, *Cephalopholis argus* (Bloch and Schneider 1801), is a demersal (bottom dwelling) reef fish of the family Epinephelidae. This species is found in reef habitat (2-40 m depth) from the Pitcairn group in the Pacific to east Africa and the Red Sea [35, Fig. 3.1]. Many members of the genus *Cephalopholis* display complex social behaviors such as territoriality, sequential hermaphroditism, and a haremic social system [36]. Long-range dispersal in this species, as in most coral reef organisms, is limited to the pelagic larval stage [37]. The pelagic larval duration for *C. argus* has not been determined but a 40-day average is proposed for Epinephelids [38]. We analyzed DNA sequence data to assess phylogeographic patterns across the range of this species to test three alternative hypotheses concerning the origin of the biodiversity in the CT. Explicitly we address the following questions: 1) does genetic diversity in the CT indicate an ancestral population with dispersal away from the CT as would be expected under the center of origin hypothesis, 2) is the ancestral diversity peripheral to the CT and accompanied with evidence of migration toward the CT as would be expected under the center of accumulation hypothesis, or 3) is the genetic diversity in the CT the result of mixing of divergent lineages across the IPB as would be expected under the region of overlap hypothesis?

Methods

A total of 550 *Cephalopholis argus* were collected from 21 locations across the species range in the Pacific and Indian Oceans including two locations at opposite ends of the CT (Philippines and Bali; Table 3.1). Most samples were collected by SCUBA divers using polespears or by fishers using lines. In some cases, samples were obtained from fish markets but only when we were confident they had been caught locally (within 100 km). Tissues samples (fin clips or gill filaments) were preserved in salt-saturated DMSO [39] and stored at room temperature. DNA was isolated using the modified HotSHOT method [40,41]. Approximately 870 bp of mitochondrial cytochrome *b* (Cyt*b*) were amplified using the primers CB6F (5'-CTCCCTGCACCTTCAAACAT-3') and CB6R (5'-GGAAGG TTAAAG CCC GTTGT-3') which we designed for this species. Additionally, approximately 375 bp of the third intron in the gonadotropin-releasing hormone (GnRH) gene were amplified using the primers GnRH3F and GnRH3R [42] and approximately 730 bp of the first intron of the S7 ribosomal protein (S7) gene were amplified using the primers S7RPEX1F and S7RPEX2R [43].

Polymerase chain reactions (PCRs) for all three markers were carried out in a 10 µl volume containing 2-15 ng of template DNA, 0.2-0.3 µM of each primer, 5 µl of the premixed PCR solution BioMix Red^{TM} (Bioline Inc., Springfield, NJ, USA), and deionized water to volume. PCR reactions utilized the following cycling parameters: initial denaturation at 95°C and final extension at 72°C (10 min each), with an intervening 35 cycles of 30 s at 94°C, 30 s at the annealing temperature (54°C for Cyt*b*; 58°C for GnRH and S7), and 45 s at 72°C. Amplification products were purified using 0.75 units of Exonuclease I/0.5 units of Shrimp Alkaline Phosphatase (ExoSAP; USB, Cleveland, OH, USA) per 7.5 µl PCR products at 37°C for 60 min, followed by deactivation at 80°C for 10 min. DNA sequencing was performed with

fluorescently-labeled dideoxy terminators on an ABI 3730XL Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) at the University of Hawaii's Advanced Studies of Genomics, Proteomics and Bioinformatics sequencing facility.

Sequences for each locus were aligned, edited, and trimmed to a common length using the DNA sequence assembly and analysis software GENEIOUS PRO 5.0 (Biomatters, LTD, Auckland, NZ). In all cases, alignment was unambiguous with no indels or frameshift mutations. Allelic states of nuclear sequences with more than one heterozygous site (GnRH = 43.1% and S7 = 48.4% of individuals) were estimated using the Bayesian program PHASE 2.1 [44,45] as implemented in the software DnaSP 5.0 [46]. We conducted six runs in PHASE for each data set. Each run had a unique random-number seed. Five runs were conducted for 1000 iterations with 1000 burn-in iterations. To ensure proper allele assignment, a sixth run of 10000 iterations was conducted. All runs returned consistent allele identities. GnRH and S7 genotyptes resulted in no more than 4 and 6 ambiguous sites per individual, respectively. PHASE was able to differentiate all alleles with > 95% probability at both loci except at single nucleotide positions in 4 individuals at GnRH and 10 individuals at S7 or 0.8% and 2.0% of samples, respectively. Unique haplotypes and alleles were identified with the merge taxa option in MacClade 4.05 [47] and deposited in GenBank (ascension numbers: *XXXX* for Cyt*b*, XXX for GnRH intron, XXXX for S7 intron.

Data analyses

Mitochondrial DNA

Summary statistics for *C. argus*, including haplotype diversity (*h*) and nucleotide diversity (π), were estimated with algorithms from Nei [48] as implemented in the statistical software package ARLEQUIN 3.5 [49]. To test whether haplotype and nucleotide diversities

differed between ocean basins (Pacific Ocean = Marquesas, Moorea, Kiritimati, Palmyra, Samoa/Tokelau, Baker/Howland, Kwajalein, Pohnpei, Saipan, Palau, Lizard Island, and Philippines; Indian Ocean = Sumatra, Bali, Scott Reef, Rowley Shoals, Christmas Island, Cocos/Keeling, Diego Garcia, Oman, and Seychelles) we calculated unpaired *t*-tests using the online calculator GraphPad (http://www.graphpad.com/quickcalcs/ttest1.cfm). The AIC implemented in jMODELTEST 0.1.1 indicated the TPM1uf+G as the best-fit model of DNA sequence evolution with a gamma value of 0.065. Median-joining networks were constructed using the program NETWORK 4.5 with default settings [50]. An intra-specific phylogeny was produced using maximum likelihood (ML) methods and default settings in the program RAxML 7.2.7 [51]. Trees were rooted using Cyt*b* sequences of two congenerics (*C. urodeta* and *C. taeniops*) obtained from GenBank (ascension numbers AY786426 and EF455990, respectively). Bootstrap support values were calculated using default settings with 1000 replicates. The ML tree topology was confirmed by neighbor-joining (NJ) and Bayesian Markov Chain Monte Carlo (MCMC) analysis using MEGA 4.0 [52] and MRBAYES 3.1.1 [53], respectively. The NJ tree was generated using the Tamura-Nei model of evolution [54] and a gamma parameter of 0.065. Bootstrap support values were calculated using 1000 replicates. The Bayesian analysis was run using the default four heated, one million step chains with an initial burn-in of 100,000 steps. We calculated the corrected average number of pairwise differences between mitochondrial lineages (*d*) in ARLEQUIN.

To determine whether the number of pairwise differences among all DNA sequences reflected expanding or stable populations [55], we calculated the frequency distribution of the number of mutational differences between haplotypes (mismatch analyses), as implemented in ARLEQUIN. To determine confidence intervals around this value we calculated Harpending's

raggedness index, r [55] which tests the null hypothesis of an expanding population. This statistic quantifies the smoothness of the observed pairwise difference distribution and a non-significant result indicates an expanding population. Fu's F_S [56], which is highly sensitive to population expansions was calculated using 10,000 permutations. Significant negative values of F_S indicate an excess of low-frequency haplotypes, a signature characteristic of either selection or a recent demographic expansion [56].

To test for hierarchical population genetic structure in *C. argus*, an analysis of molecular variance (AMOVA) was performed in ARLEQUIN using 20,000 permutations. Because the TPM1uf+G model of sequence evolution is not implemented in ARLEQUIN, we used the most similar model available [54] with a gamma value of 0.065. An analogue of Wright's F_{ST} (Φ_{ST}), which incorporates the model of sequence evolution, was calculated for the entire data set and for pairwise comparisons among all locations. We maintained $\alpha = 0.05$ among all pairwise tests by controlling for the false discovery rate as recommended by Benjamini and Yekutieli [57] and reviewed by Narum [58]. A Mantel test was performed to determine whether significant isolation-by-distance exists among populations by testing for correlation between pairwise Φ_{ST} values and geographic distance using the Isolation-by-Distance Web Service 3.16 [59]. Mantel tests were performed with 10,000 iterations on the data set that included negative Φ_{ST} values and again with negative Φ_{ST} values converted to zeros.

To estimate the time to coalescence we used the Bayesian MCMC approach implemented in BEAST 1.5.4 [60]. We conducted our analysis with a relaxed lognormal clock and under a model of uncorrelated substitution rates among branches. We used default priors under the HKY + G model of mutation (jMODELTEST) [61,62] and ran simulations for 10 million generations with

sampling every 1000 generations. Five independent runs were computed to ensure convergence and log files were combined using the program TRACER 1.5 [63].

Nuclear introns

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 using ARLEQUIN. To test whether H_E differed between ocean basins we calculated unpaired t-tests as described above. Linkage disequilibrium between the two nuclear loci was assessed using the likelihood ratio test with 20,000 permutations in ARLEQUIN. We tested for population expansions by calculating Fu's F_S [56], using 10,000 permutations in ARLEQUIN. Genotypes for each individual at the GnRH and S7 introns were compiled and used to calculate F_{ST} for the multi-locus data set and for pairwise comparisons between locations in ARLEQUIN. The false discovery rate among multiple comparisons was controlled as described above. Median-joining networks for alleles at each locus were constructed using the program NETWORK. We tested for correlation between pairwise F_{ST} values and geographic distance (isolation-by-distance) among all populations using the Isolation-by-Distance Web Service [60] as described above.

Migration

Migration rates between groups (*Nm*: where *N* is effective population size and *m* is migration rate) were calculated with the software MIGRATE 3.1.6 [64,65]. To minimize the parameters run, we pooled locations that showed no pairwise structure (i.e. those locations with a pairwise Φ_{ST} that did not significantly differ from zero) into demes defined by region (see Results). This program uses a Bayesian MCMC search strategy of a single, replicated, two million step chain. The default settings for priors were used with an unrestricted migration

model. Estimates of the number of immigrants per generation (*Nm*) were calculated by multiplying final estimates of Θ and M [66].

Results

Mitochondrial DNA

We resolved a 729 bp segment of cytochrome *b* in 550 individuals yielding 57 haplotypes with 34 of these haplotypes observed in single individuals (Table 3.1). Due to geographic proximity and a lack of genetic differentiation (as measured by pairwise Φ_{ST}) we grouped the specimens from the central Pacific locations of Samoa and Tokelau, and Baker and Howland Island. The number of individuals (N), number of haplotypes (N_h), haplotype diversity (*h*), and nucleotide diversity (π) for each location are provided in Table 3.1. Overall nucleotide diversity in *C. argus* was low (π = 0.005) while the corresponding haplotype diversity was high (*h* = 0.80). Across all samples, π = 0.001 - 0.009 and *h* = 0.38 - 0.96 with higher genetic diversity detected in the Pacific compared to the Indian Ocean (unpaired *t*-test, π: t = 2.22, df = 19, *P* = 0.039; *h*: t = 2.88, df = 19, *P* = 0.010). Using the program BEAST and implementing a molecular clock of 2% per million years [67-69] we estimated a coalescence time of approximately 930,000 years with bounds of the 95% highest posterior density intervals (HPD) of 0.5 and 1.5 million yrs, dates that correspond to the middle of the Pleistocene. The median-joining network revealed two clusters of haplotypes that are distinguished by three substitutions (*d* = 0.008, Fig. 3.2a). The two lineages, which we refer to as the Pacific and Indian Ocean lineages, were confirmed by the phylogenetic analyses (Fig. 3.3). Coalescence times for the two lineages were 580,000 (95% HPD = 0.25 - 1.0 million yrs) and 520,000 (95% HPD 0.22 - 0.88 million yrs) yrs, respectively. No haplotypes from the Pacific Ocean lineage were detected at the western Indian Ocean locations of Diego Garcia, Oman, and Seychelles while 10 of 291 samples from the Pacific

Ocean fell into the Indian Ocean lineage (Fig. 3.1). A region of extensive overlap was found at the Indian Ocean locations of Bali, Scott Reef, Rowley Shoals, Christmas Island, and Cocos/Keeling Islands (Fig. 3.1).

Overall Φ_{ST} was 0.297 ($P < 0.001$). When we grouped samples by ocean basin (as described in Methods) we found significant structure between the Pacific and Indian Oceans ($\Phi_{CT} = 0.242$, $P < 0.001$). Within oceans we found low but significant structure in the Pacific Ocean ($\Phi_{ST} = 0.036$, $P < 0.001$) and higher structure in the Indian Ocean ($\Phi_{ST} = 0.249$, $P < 0.001$). Pairwise comparisons indicate low levels of structure at the eastern edge of the range distinguishing Marquesas and Moorea (Table 3.2) but no structure across the entire central Pacific from Kiritimati to Lizard Island (Table 3.2). While there was no structure among locations in the western Indian Ocean (Diego Garcia, Oman, and Seychelles) and Sumatra, high levels of structure were observed between these locations and the eastern Indian Ocean (Christmas Island, Cocos/Keeling, Bali, Scott Reef, and Rowley Shoals).

Population expansion parameters for the overall dataset gave conflicting results. As expected with the presence of two divergent mitochondrial lineages, the mismatch distribution for the overall data set was bimodal (Fig. 3.4) and resulted in a significant raggedness index (r = 0.24, $P < 0.001$), a result that indicates a stable population. In contrast, Fu's F_S resulted in $F_S = -12.8$ ($P < 0.001$) signifying an excess of low-frequency haplotypes and an expanding population. Grouping locations that demonstrated no significant population structure (see Table 3.2) resulted in five groups: French Polynesia (FP) = Marquesas and Moorea; central-west Pacific (CW) = Kiritimati, Palmyra, Samoa/Tokelau, Baker/Howland, Kwajalein, Pohnpei, Saipan, Palau, Lizard Island, Philippines, and Scott Reef; Indo-Pacific boundary (IB) = Bali and Rowley Shoals; eastern Indian Ocean (EI) = Cocos/Keeling and Christmas Islands; western Indian Ocean (WI) =

Sumatra, Diego Garcia, Oman, and Seychelles. Despite their close geographic proximity and lack of genetic structure with many populations in the CW, Bali and Rowley Shoals were grouped separately because they demonstrate significant genetic structure when compared to Samoa/Tokelau and Baker/Howland. When analyzed separately, mismatch analyses resulted in non-significant raggedness indexes for each group (data not presented). Fu's F_S indicated expanding populations for FP (F_S = -4.8, P = 0.014), CW (F_S = -20.9, P < 0.001) and WI (F_S = -8.9, P < 0.001) but no evidence for population expansion was found for either IB (F_S = -0.43, P = 0.48) or EI (F_S = -0.48, P = 0.63).

Nuclear introns

We resolved 245 bp of the GnRH intron in 488 specimens and 393 bp of the S7 intron in 490 specimens (Table 3.1). Seven polymorphic sites yielded 11 alleles at the GnRH locus and 15 polymorphic sites yielded 20 alleles at the S7 locus. Median-joining networks for the GnRH and S7 introns revealed two prominent alleles at each locus that were found throughout the species' range (Fig. 3.2b, c). The number of individuals (n), number of alleles (N_a), observed heterozygosity (H_O), expected heterozygosity (H_E), and the corresponding P-value for the exact tests for HWE are listed in Table 3.1. The samples from the Marquesas and Samoa/Tokelau were found to be inconsistent with HWE expectations with an excess of homozygotes at the GnRH locus (P = 0.002 and 0.044 respectively) while the sample from Diego Garcia was found to have an excess of heterozygotes at the S7 locus (P = 0.047) (Table 3.1). Across all samples H_E = 0.26–0.80 for the GnRH intron and H_E = 0.28–0.79 for the S7 intron with higher values of H_E detected in the Pacific compared to the Indian Ocean (unpaired t-test, GnRH: t = 3.17, df = 19, P = 0.005; S7: t = 3.99, df = 19, P < 0.001). There was no indication of linkage disequilibrium between the two loci (P > 0.05).

Overall F_{ST} for GnRH, S7, and the multi-locus data set were $F_{ST} = 0.078$ ($P < 0.001$), $F_{ST} = 0.099$ ($P < 0.001$), and $F_{ST} = 0.127$ ($P < 0.001$), respectively. Analyses of these data reveal patterns of population structure that are concordant with the mitochondrial data set. Grouping samples by ocean basin (as above) revealed significant structure between the Pacific and Indian Oceans (GnRH, $F_{CT} = 0.056$, $P = 0.002$; S7, $F_{CT} = 0.103$, $P < 0.001$, multi-locus $F_{ST} = 0.154$, $P < 0.001$) and significant structure within ocean basins (Pacific Ocean: GnRH, $F_{ST} = 0.020$, $P = 0.025$; S7, $F_{ST} = 0.041$, $P < 0.001$; multi-locus $F_{ST} = 0.013$, $P = 0.039$; Indian Ocean: GnRH, $F_{ST} = 0.074$, $P < 0.001$; S7, $F_{ST} = 0.049$, $P < 0.001$; multi-locus $F_{ST} = 0.072$, $P < 0.001$). Pairwise F_{ST} for the multi-locus data set are reported in Table 3.2. Overall the nuclear data set measured lower levels of population structure compared to mtDNA. Using the multi-locus data set we found little population subdivision across the central Pacific and no structure in the western Indian Ocean. This data set did not detect the low levels of population structure at the Marquesas and Moorea revealed in the mtDNA data set nor were the Indian Ocean populations as divergent using these markers (Table 3.2).

As might be expected from loci with low numbers of closely related alleles, the mismatch distributions for the overall nuclear data set and for the five geographic groups (FP, CW, IB, EI, and WI) were unimodal and resulted in non-significant raggedness indices (overall dataset: GnRH, r = 0.32, $P = 0.082$; S7, r = 0.12, $P = 0.235$; data for geographic groups not shown). Fu's F_S calculations offered no evidence for expanding populations for either the overall dataset (GnRH, $F_S = -0.34$, $P = 0.521$; S7, $F_S = -3.93$, $P = 0.162$) or for the geographic groups (data not shown).

Migration

Migration analyses for the nuclear dataset proved to be uninformative. Posterior probabilities did not narrow on a single mode for several comparisons and confidence intervals were unreasonably large. We present only the mtDNA data here. Migration rates indicate that while the populations of *C. argus* at the ends of the range (FP and WI) contribute to genetic diversity across the central portion of the range (*Nm* per generation = 2.2 - 87.3 and 1.5 - 6.6, respectively), they rarely receive migrants (*Nm* per generation = 0.0 - 0.4 and 0.0, respectively; Fig. 3.5). There is evidence of considerable migration between the other groups (*Nm* per generation = 1.8 - 39.0; Fig. 3.5).

Isolation by distance (IBD)

Mantel tests showed a strong correlation between genetic distance (Φ_{ST} or F_{ST}) and geographic distance in the mtDNA (r = 0.749, $P = 0.001$) and the multi-locus nuclear (r = 0.715, $P < 0.001$) data sets. Replacing negative values of Φ_{ST} and F_{ST} with zeros did not affect the pattern or statistical significance. To test if genetic structure between ocean basins was driving IBD we conducted Mantel tests within oceans and found weaker but still significant correlations between genetic distance and geographic distance with the Cyt*b* data set (Pacific Ocean: r = 0.301, $P = 0.033$; Indian Ocean: r = 0.778, $P = 0.004$) but not the multi-locus nuclear data set (Pacific Ocean: r = -0.056, $P = 0.629$; Indian Ocean: r = 0.315, $P = 0.085$).

Discussion

The origin of the remarkable species richness of the Coral Triangle (CT) has fostered numerous and seemingly conflicting hypotheses. The center of origin hypothesis postulates that increased rates of speciation in the CT have resulted in high species diversity in the region [9]. In contrast, the center of accumulation hypothesis contends that taxa that have evolved peripherally accumulate in the CT due to prevailing currents [19]. Finally, the region of overlap hypothesis

states that the observed pattern is the result of admixture of the distinct biotas of the Pacific and Indian Oceans [21]. Despite the effort that has been dedicated to determining the mechanism driving the pattern of species diversity in the Indo-Pacific no consensus has emerged [7,24,25]. Our genetic survey of *C. argus* across 18,000 km of the Indo-Pacific lends some insight into this debate.

Cephalopholis argus demonstrates significant levels of genetic structure that indicate a historical partition between the Pacific and Indian Oceans (Table 3.2). Two mitochondrial lineages are distinguished by fixed differences (d = 0.008) indicating isolation for approximately one million years (95% HPD intervals are 0.5 - 1.5 million yrs), a time interval that corresponds to Pleistocene sea level fluctuations linked to Milankovitch climate cycles on the scale of 10^5 years [70]. Our analyses indicate expanding populations with migration toward the center of the range. The high genetic diversity of this species within and adjacent to the CT is a result of mixing Pacific and Indian Ocean lineages (Figs. 3.1, 3.5). These data support isolation of Pacific and Indian Ocean populations during prolonged and repeated sea level fluctuations of the Pleistocene followed by population expansion and colonization of the CT from populations in both the Pacific and Indian Oceans: a pattern that is consistent with predictions of the region of overlap hypothesis.

While incomplete lineage sorting is a serious problem for species level reconstructions our pattern of divergence across the IPB is corroborated by three independent markers. Additionally, the finding of isolation by distance across the species range is strong evidence that the patterns we present here are not driven by stochastic events.

Indo-Pacific Barrier-the mechanism of isolation

The Sunda shelf, surrounding the Malay Peninsula and western islands of Indonesia, and the Sahul shelf off northern Australia and New Guinea, separate the Pacific and Indian Oceans and together are known as the Indo-Pacific Barrier (IPB) [71]. Over the last 700,000 yr there have been at least three to six glacial cycles that lowered sea level as much as 130 m below present levels (Fig. 3.6, [23,72-74]). Species on the continental shelves were repeatedly subjected to widespread extirpations and presumably interruption of gene flow between Pacific and Indian Ocean populations. However, at glacial maxima the isolation of the two ocean basins was not complete. Associated with the change in sea level were concomitant changes in oceanographic current patterns, altered discharge of local rivers, with corresponding changes in temperature and salinity [75,76]. The narrow seaways that remained were likely under the influence of cooler upwelling, further limiting the availability of suitable habitat for tropical marine organisms [10,23,71,73,77]. In sum, the isolating mechanism between ocean basins may have been due to both ecological and geological factors.

The evidence for the impact of the IPB on shallow tropical marine organisms is extensive and compelling. Historical and contemporary restrictions to dispersal between the Pacific and Indian Oceans are indicated by the confinement of many demersal species primarily to one ocean or the other [3,21,78,79]. More recently, genetic data have been used to assess the IPB. Studies of demersal organisms that lack vagile adults have found intraspecific genetic differentiation across the IPB in many fishes [80-88] and invertebrates [89-96] with few exceptions [67,69,95,97,98]. Genetic analyses reveal signatures of isolation that correspond to Pleistocene sea level fluctuations across a diversity of taxa [82,85,93,97,99] including *C. argus*. This species demonstrates strong population structure between Pacific and Indian Ocean locations in both the mitochondrial and nuclear data sets. The mismatch distribution for *C. argus* is distinctly bimodal

(Fig. 3.4) which is characteristic of species under the influence of a strong biogeographic barrier [100,101]. The mid-Pleistocene age of the two mitochondrial lineages of *C. argus* coupled with assortment by ocean basins is compelling evidence that the divergence is a result of isolation on either side of the IPB.

Eastern Indian Ocean and the Coral Triangle: A region of overlap

Since the end of the last glacial maximum about 18,000 yrs ago, the land bridge that impeded dispersal between the Pacific and Indian oceans submerged and the rising sea level not only opened dispersal pathways but was also accompanied by an approximately 10 fold increase in suitable shallow reef habitat [4]. Woodland [21] was the first to propose that range expansions of species formed in isolation during Pleistocene glacial cycles contributed to the incredible species richness of the CT. His work on species distributions of rabbitfishes (Family Siganidae) and later the work of Donaldson [102] on hawkfishes (Family Cirrhitidae) offer supporting evidence. Range expansions are also indicated by the presence of a hybrid zone in the eastern Indian Ocean [103]. Cocos/Keeling and Christmas Islands lie 500 and 1,400 km, respectively, from the southern coast of the Indonesian Island of Java, and are a known region of overlap for Pacific and Indian Ocean fish faunas. Here, sister species that are otherwise restricted to different oceans inhabit the same reefs and in many cases hybridize [103-105]. Notably, we found nearly equal proportions of Pacific and Indian Ocean *C. argus* haplotypes in this hybrid zone (Fig. 3.1). These findings demonstrate that, at least in terms of intraspecific genetic diversity, the introgression is not restricted to Cocos/Keeling and Christmas Islands but instead extends well into Indonesia, the western Pacific, and to a lesser extent, the central Pacific.

If we provisionally assume that genetic divergences are the result of isolation across the IPB, we can estimate the degree of introgression since the last ice age. In some taxa, effective

migration between ocean basins is absent as evidenced by a lack of shared haplotypes between oceans (*Chlorurus sordidus* [82]; *Penaeus monodon* [93]). Other taxa reveal signatures of historical isolation but lack contemporary spatial structure (*Naso brevirostris* [97]). Pacific and Indian Ocean populations of *C. argus* share haplotypes but mixing is incomplete as evidenced by significant population structure between oceans, a pattern observed in several other species (*Myripristis berndti* [84]; *Naso vlamingii* [99]; *Nerita albieilla* [106]). *C. argus* is unique in that it demonstrates unidirectional dispersal out of the western Indian Ocean (WI) and French Polynesia (FP) toward the center of the range (Fig. 3.5) while populations in the CT and western Australia, the area near the Indo-Pacific boundary, demonstrate high levels of bidirectional dispersal, high genetic diversity, and extensive lineage overlap (Figs. 3.1, 3.5).

There is compelling evidence for the influence of the IPB on coral reef organisms from species level distributions to intraspecific lineage sorting. The degree of range expansion or lineage mixing after the last glacial maximum varies among taxa and may reflect species level differences such as dispersal ability, reproductive strategy, competitive ability, or habitat requirements.

Phylogeographic inferences: emerging patterns in Indo-Pacific reef fishes

Our data set allows for several phylogeographic inferences. Molecular diversity indices and the topology of the medium joining networks indicate that Indian Ocean populations harbor more genetic diversity. The position of the Indian Ocean lineage in the phylogenetic tree indicates that this lineage may be older but coalescence dates do not support this. Taken together these data may indicate that during low sea level stands, populations in the western Indian Ocean were less severely impacted than those in the Pacific. *C. argus* demonstrates no population structure across the nearly 8,000 km central Pacific range, from Kiritimati to Palau. However,

pairwise Φ_{ST} and F_{ST} values and migration rates indicate that populations at the eastern end of the range at Moorea and the Marquesas are isolated. This pattern of extensive population connectivity across the central Pacific with isolation at the ends of the Pacific range is emerging in reef fishes (reviewed in [86]; see also [107]).

Biogeographic inferences: the Western Indian Ocean Province

The isolation of the western Indian Ocean (WIO) is supported by both species distributions [22] and intraspecific genetic data [84,87,97] evidence that the microevolutionary divergences documented with DNA sequence data can lead to macroevolutionary partitions between species. Genetic analyses partition Indian Ocean populations of *C. argus* along an east-west gradient and indicate unidirectional dispersal out of the WIO. Despite being geographically a part of the Indian Ocean, the eastern Indian Ocean locations at Cocos/Keeling and Christmas Islands, and Western Australia are closely affiliated with the Pacific ichthyofauna with only 5% of reef fishes at Cocos/Keeling of exclusively Indian Ocean origin [108]. Instead, these islands are considered to be a part of the Indo-Polynesian Province that stretches from the eastern Indian Ocean to French Polynesia [3,22]. Diego Garcia in the Chagos Archipelago lies in the middle of the Indian Ocean 1,900 km east of the Seychelles and 2,400 km west of Cocos/Keeling. Fish surveys in the Chagos Islands delineated the archipelago into two distinct assemblages, with the northern portion sharing affinities with the eastern Indian Ocean and the southern portion (including Diego Garcia) more closely aligned with faunal assemblages further west [109]. The distinction of the ichthyofauna assemblage of the southern Chagos Archipelago coupled with a lack of intraspecific genetic structure in two species of reef fishes from Diego Garcia and sites further west (*Lutjanus kasmira* [98] and *C. argus* this study) indicate that Diego Garcia is a part of the Western Indian Ocean Province as described by Briggs [22]. Faunal surveys indicate that

the fishes of India and Sri Lanka have a strong affiliation with the Malay Peninsula [22]. Taken together, these data indicate that the western boundary of the Indo-Polynesian Province lies east of Oman and includes India, Sri Lanka, and the northern Chagos Archipelago. While we use species distributions and genetic data to define biogeographic provinces, the mechanisms that separate the eastern and western Indian Ocean are unknown and require more detailed genetic and oceanographic work.

Conclusions

Our genetic survey of the grouper *Cephalopholis argus* indicates that this species was strongly impacted by Pleistocene sea level fluctuations which resulted in the partitioning of this species into Pacific and Indian Ocean mitochondrial lineages that are distinguished by fixed differences (d = 0.008). Following the end of the last glacial maximum, connectivity between the Pacific and Indian Oceans resumed and *C. argus* populations expanded. Representatives of each mitochondrial lineage are now found in both oceans with the center of diversity occurring in the Coral Triangle: a pattern that we offer as support for the region of overlap hypothesis. In a recent review 15 out of 18 species demonstrated significant structure across the IPB, such that subsequent contact would constitute support of the region of overlap hypothesis [98]. However, the studies cited above, on a diverse array of marine taxa, offer equally compelling evidence for the other two competing hypotheses: the center of origin and center of accumulation. None of these hypotheses are mutually exclusive, and acting in concert, they are likely to explain the patterns of biodiversity in the Indo-Pacific.

References

1. Bowen BW, Roman J: **Gaia's handmaidens: the Orlog model for conservation biology.** Con Biol 2005, **19:**1037-1043.

2. Veron JEM: *Corals in space and time: The biogeography and evolution of the Scleratinia*. Sydney: University of New South Wales Press; 1995.

3. Randall JE: **Zoogeography of shorefishes of the Indo-Pacific region**. Zool Studies 1998, **37:**227-268.

4. Bellwood DR, Wainwright PC: **The history and biogeography of fishes on coral reefs**. In *Coral reef fishes: dynamics and diversity in a complex ecosystem*. Edited by Sale PF. San Diego: Academic Press; 2002:5-32.

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

6. Briggs JC: **Marine centers of origin as evolutionary engines.** J Biogeogr 2003, **30:**1-18.

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

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

9. Ekman S: *Zoogeography of the sea*. London: Sidgwick and Jackson; 1986.

10. McManus JW: **Marine speciation, tectonics, and sea-level changes in southeast Asia.** In *Proceedings of the Fifth International Coral Reef Congress*: 27 May-1 June 1985; Tahiti. Edited by Gabrie C, Vivien MH. Moorea: Antenne Museum-Ephe; 1985:133-138.

11. Briggs JC: **The marine East Indies: diversity and speciation.** J Biogeogr 2005, **32:**1517-1522.

12. Rocha LA, Bowen BW: **Speciation in coral reef fishes.** J Fish Biol 2008, **72:**1101-1121.

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

14. DeBoer TS, SubiMD, Subia MD, Ambariyanto, Erdmann, MV, Kovitvonga K, Barber: **Phylogeography and limited genetic connectivity in the endangered giant boring clam across the Coral Triangle.** Conserv Biol 2008, **22**:1255-1266.

15. Knittweis L, Krämer WE, Timm J, Kochzius M: **Genetic structure of *Heliofungia actiniformis* (Scleractinia: Fungiidae) populations in the Indo-Malay Archipelago: implication for the live coral trade management efforts.** Conserv Biol 2009, **10**:241-249.

16. Kochzius M, Seidel C, Hauschild J, Kirchhoff S, Mester P, Meyer-Wachsmuth I, Nuryanto A, Timm J: **Genetic population structure of the blue starfish *Linckia laevigata* and its gastropod ectoparasite *Thyca crystallina*.** Mar Eco Prog Ser 2009, **396**:211-219.

17. Nuryanto A, Kochzius M: **Highly restricted gene flow and deep evolutionary lineages in the giant clam *Tridacna maxima*.** Coral Reefs 2009, **28**:607-619.

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

19. Ladd HS: **Origin of the Pacific Island molluscan fauna.** Am J Sci 1960, **258A:**137-150.

20. Jokiel P, Martinelli FJ: **The vortex model of coral reef biogeography.** J of Biogeogr 1992, **19:**449-458.

21. Woodland DJ: **Zoogeography of the Siganidae (Pisces): an interpretation of distribution and richness patterns.** Bull Mar Sci 1983, **33:**713-717.

22. Briggs JC: *Marine zoogeography*. New York: McGraw-Hill; 1974.

23. Voris HK: **Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations**. J Biogeogr 2000, **27:**1153-1167.

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

25. Halas D, Winterbottom R: **A phylogenetic test of multiple proposals for the origins of the East Indies coral reef biota.** J Biogeogr 2009, **36:**1847-1860.

26. Wallace CC: **The Indo-Pacific center of coral diversity re-examined at species level.** In *Proceedings of the 8th international Coral Reef Symposium*: 24-29 June 1997; Panama. Edited by Lessios HA, Macintyre IG. Balboa: Smithsonian Tropical Research Institute; 1997: 365-370.

27. Allen GR, Adrim M: **Coral reef fishes of Indonesia.** Zool Studies 2003, **42:**1-72.

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

29. Drew J, Barber PH: **Sequential cladogenesis of the reef fish *Pomacentrus moluccensis* (Pomacentridae) supports the peripheral origin of marine biodiversity in the Indo-Australian archipelago.** Mol Phylogenet Evol 2009, **53:**335-339.

30. Bowen BW, Clark AM, Abreu-Grobois FA, Chavez A, Reichart H, Ferl RJ: **Global phylogeography of the ridley sea turtles (*Lepidochelys* spp.) inferred from mitochondrial DNA sequences.** Genetica 1998, **101:**179-189.

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

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

33. Bernardi G, Bucciarelli G, Costagliola D, Robertson DR, Heiser JB: **Evolution of coral reef fish *Thalassoma* spp. (Labridae): 1. Molecular phylogeny and biogeography.** Mar Biol 2004, **144:**369-375.

34. Avise JC: *Molecular markers, natural history, and evolution*. Sunderland: Sinauer Associates, Inc.; 2004.

35. Randall JE: *Reef and Shore Fishes of the South Pacific*. Honolulu: Sea Grant College Program University of Hawaii; 2005.

36. Shpigel M, Fishelson L: **Territoriality and associated behvaiour in three species of the genus *Cephalopholis* (Pisces: Serranidae) in the Gulf of Aquba, Red Sea.** J Fish Biol 1999, **38:**887-896.

37. Leis JM: **The pelagic phase of coral reef fishes: larval biology of coral reef fishes.** In *The Ecology of Fishes on Coral Reefs*. Edited by Sale PF. San Diego: Academic Press; 1991:183-230.

38. Linderman KC, Lee TM, Wilson WD, Claro R, Ault JS: **Transport of larvae originating in southwest Cuba and the Dry Tortugas: evidence for partial retention in grunts and snappers.** Proc Gulf Carib Fish Inst 2000, **52:**253-278.

39. Seutin G, White BN, Boag PT: **Preservation of avian and blood tissue samples for DNA analyses**. Can J Zool 1991, **69:**82-92.

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

41. Truett GE, Mynatt RL, Truett AA, Walker JA, Warman ML: **Preparation of PCR-quality mouse genomic DNA with hot sodium hydroxide and Tris (HotSHOT).** BioTechniques 2000, **29:**52-54.

42. Hassan M, Lemaire C, Fauvelot C, Bonhomme F: **Seventeen new exon-primed intron-crossing polymerase chain reaction amplifiable introns in fish.** Mol Ecol Notes 2002, **2:**334-340.

43. Chow S, Hazama K: **Universal PCR primers for S7 ribosomal protein gene introns in fish.** Mol Ecol 1998, **7:**1247-1263.

44. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** Am J Hum Genet 2001, **68:**978-989.

45. Stephens M, Donnelly P: **A comparison of Bayesian methods for haplotype reconstruction from population genotype data.** Am J Hum Genet 2003, **73:**1162-1169.

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

47. Maddison DR, Maddison WP: *MacClade 4: analysis of phylogeny and character evolution.* Sunderland: Sinauer Associates; 2002.

48. Nei M: *Molecular evolutionary genetics*. New York: Columbia University Press; 1987.

49. Excoffier L, Lischer HEL: **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.** Mol Ecol Res 2010, **10:**564-567.

50. Bandelt HJ, Forster P, Röhl A: **Median-joining networks for inferring intraspecific phylogenies.** Mol Biol Evol 1999, **16:**37-48.

51. Stamatakis A: **RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models.** Bioinformatics 2006, **22:**2688-2690.

52. Tamura K, Dudley J, Nei M, Kamar S: **MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0.** Mol Biol Evol 2007, **24:**1596-1599.

53. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny.** Bioinformatics 2001, **17:**754-755.

54. Tamura K, Nei M: **Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.** Mol Biol Evol 1993, **10:**512-526.

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

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

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

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

59. Jensen JL, Bohonak AJ, Kelley ST: **Isolation by distance, web service.** BMC Genetics 2005, **6:**13.

60. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.** BMC Evol Biol 2007, **7:**214-221.

61. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** Syst Biol 2003, **14:**685-695.

62. Posada D: **jModelTest: Phylogenetic model averaging.** Mol Biol Evol 2008, **25:**1253-1256.

63. **Molecular Evolution, Phylogenetics, and Epidemiology Website [**http://tree.bio.ed.ac.uk/software/tracer/].

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

65. Beerli P, Felsenstein J: **Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach.** Proc Nat Acad Sci 2001, **98:**4563-4568.

66. Beerli P: **How to use Migrate or why are Markov chain Monte Carlo programs difficult to use?** In *Population Genetics for Animal Conservation*. Edited by Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C. Cambridge: Cambridge University Press; 2009: 42-79.

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

68. Lessios HA: **The great American schism: divergence of marine organisms after the rise of the Central American Isthmus.** Ann Rev Evo Evol Sys 2008, **39:**63-91.

69. Reece JS, Bowen BW, Joshi K, Goz V, Larson AF: **Phylogeography of two moray eels indicates high dispersal throughout the Indo-Pacific.** J Hered 2010, **101:**391-402.

70. Hays JD, Imbrie J, Shackleton NJ: **Variations in the Earth's Orbit: Pacemaker of the Ice Ages**. Science 1976, **194:**1121-1132.

71. Fleminger A: **The Pleistocene equatorial barrier between the Indian and Pacific Oceans and a likely cause for Wallace's line.** UNESCO Technical Papers in Marine Science 1986, **49:**84-97.

72. Chappell J: **Relative and average sea level changes, and endo-, epi-, and exogenic processes on the earth.** In *Sea level, ice, and climatic change*. Int Ass Hydrol Sci Publ 1981, **131:**411-430.

73. Potts DC: **Evolutionary disequilibrium among Indo-Pacific corals.** Bull Mar Sci 1983, **33:**619-632.

74. Naish T, et al: **Obliquity-paced Pliocene West Atlantic ice sheet oscillations.** Nature 2009, **458:**322-328.

75. Tjia HD: **Java Sea.** In *The Encyclopedia of Oceanography*. Edited by Fairbridge RW. New York: Reinhold; 1966:424-429.

76. Van Andel TH, Heath GR, Moore TC, McGeary DFR. **Late Quaternary history, climate, and oceanography of the Timor Sea, northwestern Australia.** Am J Sci 1967, **265:**737-758.

77. Galloway RW, Kemp EM: **Late Cenozoic environments in Australia**. In *Ecological Biogeography of Australia*. Edited by Keast A. Junk: The Hague; 1981:51-80.

78. McMillan WO, Palumbi SR: **Concordant evolutionary patterns among Indo-West Pacific butterflyfishes.** Proc Roy Soc B 1995, **260:**229-236.

79. Briggs JC: **Extinction and replacement in the Indo-West Pacific Ocean.** J Biogeogr 1999, **26:**777-783.

80. Lacson JM, Clark S: **Genetic divergence of Maldivian and Micronesian demes of the damselfishes *Stegastes nigricans*, *Chrysiptera biocellata*, *C. glauca* and *C. leucopoma* (Pomacentridae).** Mar Biol 1995, **121:**585-590.

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

82. Bay LK, Choat JH, van Herwerden L, Robertson DR: **High genetic diversities and complex genetic structure in an Indo-Pacific tropical reef fish (*Chlorurus sordidus*): evidence of an unstable evolutionary past?** Mar Biol 2004, **144:**757-767

83. Menezes MR, Ikeda M, Taniguchi N: **Genetic variation in skipjack tuna *Katsuwonus pelamis* (L.) using PCR-RFLP analysis of the mitochondrial DNA D-loop region.** J Fish Biol 2006, **68:**156-161.

84. Craig MT, Eble JA, Robertson DR, Bowen BW: **High genetic connectivity across the Indian and Pacific Oceans in the reef fish *Myripristis berndti* (Holocentridae).** Mar Ecol Prog Ser 2007, **334:**245–254.

85. Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G: **Allopatric divergence and speciation in coral reef fish: the three-spot *Dascyllus, Dascyllus Trimaculatus*, species complex.** Evolution 2010, **64-5:**1218-1230.

86. Eble JA, Rocha LA, Craig MT, Bowen BW: **Not all larvae stay close to home: Long-distance dispersal in Indo-Pacific reef fishes, with a focus on the Brown Surgeonfish (*Acanthurus nigrofuscus*).** J Mar Bio, in press.

87. Winters KL, van Herwerden L, Choat HJ, Robertson DR: **Phylogeography of the Indo-Pacific parrotfish *Scarus psittacus*: isolation generates distinctive peripheral populations in two oceans.** Mar Biol 2010, **157:**1679-1691.

88. Fitzpatrick JM, Carlon DB, Lippe C, Robertson DR: **The west Pacific hotspot as a source or sink for new species? Population genetic insights from the Indo-Pacific parrotfish *Scarus rubroviolaceus*.** Mol Ecol 2011, **20**:219-234.

89. Lavery S, Moritz C, Fielder DR: **Changing patterns of population structure and gene flow at different spatial scales in *Birgus latro* (the coconut crab).** Heredity 1995, **74:**531-541.

90. Lavery S, Mortiz C, Fielder DR: **Indo-Pacific population structure and evolutionary history of the coconut crab *Birgus latro*.** Mol Ecol 1996, **5:**557-570.

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

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

93. Duda TF, Palumbi SR: **Population structure of the black tiger prawn, *Penaeus monodon*, among western Indian Ocean and western Pacific populations.** Mar Biol 1999, **134:**705-710.

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

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

96. Lessios HA, Kane J, Robertson DR: **Phylogeography of the pantropical sea urchin *Tripneustes*: Contrasting patterns of population structure between oceans.** Evolution 2003, **57:**2026-2036.

97. Horne JB, van Herwerden L, Choat HJ, Robertson DR: **High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish congeners.** Mol Phylogenet Evol 2008, **49:**629-638.

98. Gaither MR, Toonen RJ, Robertson DR, Planes S, Bowen BW: **Genetic evaluation of marine biogeographic barriers: perspectives from two widespread Indo-Pacific snappers (*Lutjanus* spp.).** J Biogeogr 2010, **37:**133-147.

99. Klanten OS, Choat JH, Van Herwerden L: **Extreme genetic diversity and temporal rather than spatial partitioning in a widely distributed coral reef fish.** Mar Biol 2007, **150:**659-670

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

101. Mora MS, Lessa EP, Cutrera AP, Kitttlein MJ, Vasallo AI: **Phylogeographical structure in the subterranean tuco-tuco *Ctenomys talarum* (Rodentia: Ctenomyidea): constrasting the demographic consequences of regional and habitat specific histories.** Mol Ecol 2007, **16:**3453-3465.

102. Donaldson TJ: **Distribution and species richness patterns of Indo-West Pacific Cirrhitidae: support for Woodland's hypothesis.** In *Proceedings of the Second International Conference on Indo-Pacific fishes*: 1986; Tokyo. Edited by Uyeno T, Arai R, Taniuchi T, Matsuura K. Tokyo: Ichthyological Society of Japan; 1986:623-628.

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

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

105. Craig MT: **The goldrim surgeonfish (*Acanthurus nigricans*; Acanthuridae) from Diego Garcia, Chagos Archipelago: first record for the central Indian Ocean.** Zootaxa 2008, **1850:**65-68.

106. Crandall ED, Frey MA, Grosberg RK, Barber PH: **Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods.** Mol Ecol 2008, **17**:611-626.

107. Schultz JK, Feldheim KA, Gruber SH, Ashley MV, McGovern TM, Bowen BW: **Global phylogeography and seascape genetics of the lemon sharks (genus *Negaprion*).** Mol Ecol 2008, **17:**5336-5348.

108. Allen GR, Smith-Vaniz WF: **Fishes of the Cocos (Keeling) Islands.** Atoll Res Bull 1994, **412:**1-21.

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

Table 3.1: Molecular diversity indices for 21 populations of *Cephalopholis argus*

	Cytb				*GnRH*					*S7*				
Sample Location	*N*	N_h	*h*	π	*N*	N_a	H_O	H_E	*P*-value	*N*	N_a	H_O	H_E	*P*-value
Marquesas (MQ)	50	8	0.65 ± 0.06	0.002 ± 0.001	34	3	0.21	0.44	0.002	48	9	0.52	0.60	0.596
Moorea (MR)	36	5	0.38 ± 0.10	0.001 ± 0.001	33	2	0.24	0.26	0.549	34	7	0.59	0.60	0.298
Kiritimati (KI)	32	8	0.74 ± 0.05	0.003 ± 0.002	33	3	0.55	0.49	0.335	28	7	0.36	0.40	0.388
Palmyra (PY)	29	6	0.65 ± 0.07	0.002 ± 0.002	30	3	0.30	0.33	0.217	27	6	0.33	0.33	0.337
Samoa/Tokelau (ST)	27	6	0.64 ± 0.07	0.001 ± 0.001	21	4	0.29	0.33	0.044	24	6	0.25	0.34	0.076
Baker/Howland (BH)	27	6	0.68 ± 0.06	0.002 ± 0.001	27	4	0.52	0.55	0.767	27	7	0.37	0.39	0.365
Kwajalein (KW)	22	10	0.86 ± 0.05	0.005 ± 0.003	23	5	0.48	0.52	0.846	23	7	0.26	0.48	<0.001
Pohnpei (PN)	15	6	0.74 ± 0.09	0.004 ± 0.002	15	3	0.20	0.38	0.055	15	8	0.67	0.67	0.709
Saipan (SP)	19	6	0.77 ± 0.07	0.003 ± 0.002	19	3	0.37	0.56	0.570	18	6	0.56	0.50	0.271
Palau (PA)	22	8	0.77 ± 0.07	0.004 ± 0.002	23	5	0.57	0.47	0.897	23	6	0.52	0.60	0.128
Lizard Island (LI)	12	5	0.67 ± 0.14	0.001 ± 0.001	10	3	0.40	0.35	1.000	7	3	0.29	0.28	1.000
Philippines (PI)	6	4	0.87 ± 0.13	0.009 ± 0.006	5	3	0.60	0.71	1.000	5	4	0.60	0.64	0.644
Bali (BA)	23	7	0.81 ± 0.05	0.005 ± 0.003	19	4	0.47	0.60	0.054	16	7	0.69	0.76	0.322
Scott Reef (SR)	42	8	0.73 ± 0.05	0.004 ± 0.002	41	7	0.56	0.58	0.262	39	8	0.72	0.64	0.897
Rowley Shoals (RS)	40	9	0.81 ± 0.04	0.005 ± 0.003	33	7	0.67	0.65	0.073	30	10	0.80	0.77	0.780
Christmas Island (CM)	49	10	0.83 ± 0.02	0.006 ± 0.004	47	7	0.55	0.57	0.262	49	11	0.71	0.70	0.717
Cocos/Keeling (CK)	40	9	0.87 ± 0.02	0.006 ± 0.004	30	10	0.83	0.78	0.161	29	7	0.72	0.71	0.599
Sumatra (SU)	4	3	0.83 ± 0.22	0.007 ± 0.005	6	5	0.83	0.82	0.703	5	4	0.60	0.71	1.000
Diego Garcia (DG)	33	10	0.87 ± 0.03	0.003 ± 0.002	33	8	0.82	0.80	0.295	33	9	0.76	0.66	0.047
Oman (OM)	9	5	0.81 ± 0.12	0.006 ± 0.003	4	4	0.75	0.64	1.000	7	6	0.71	0.79	0.869
Seychelles (SE)	13	10	0.96 ± 0.04	0.003 ± 0.002	13	5	0.62	0.80	0.054	13	5	0.77	0.70	0.739
All samples	550	55	0.80 ± 0.01	0.005 ± 0.003	499	11		0.58		500	20		0.67	

Sample location, number of individuals (N), number of haplotypes (N_h), haplotype diversity (*h*), and nucleotide diversity (π) are listed for the cytochrome *b* dataset. Number of alleles (N_a), observed heterozygosity (H_O), and expected heterozygosity (H_E) are listed for the gonadotropin releasing hormone (GnRH), and S7 ribosomal protein gene (S7) nuclear introns. *P*-values are the result of exact tests for Hardy-Weinberg equilibrium using a Markov chain with 100,000 steps in ARLEQUIN 3.5 [42].

Table 3.2: Pairwise *F* statistics for 21 populations of *Cephalophus argus*.

Location	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1.Marquesas	-	0.010	0.020	0.031	**0.077**	0.048	0.015	0.038	0.024	-0.001	-0.077	-0.018	**0.093**	0.021	**0.059**	**0.089**	**0.197**	**0.298**	**0.335**	**0.193**	**0.255**
2. Moorea	0.042	-	-0.021	0.004	0.035	0.023	0.003	0.026	0.004	0.007	-0.117	0.117	**0.095**	0.022	**0.064**	**0.093**	**0.200**	**0.399**	**0.340**	**0.188**	**0.261**
3. Kiritimati	0.039	0.059	-	0.002	0.017	-0.020	-0.029	0.036	-0.007	0.004	-0.044	0.041	**0.222**	0.026	**0.111**	**0.122**	**0.269**	**0.380**	**0.424**	**0.312**	**0.374**
4. Palmyra	0.036	0.021	0.006	-	-0.002	-0.027	-0.019	**0.088**	0.017	0.024	-0.077	0.161	**0.275**	**0.067**	**0.164**	**0.164**	**0.322**	**0.496**	**0.471**	**0.381**	**0.431**
5. Samoa/Tokelau	0.022	0.044	-0.002	0.002	-	0.004	0.012	**0.108**	0.039	0.057	-0.115	0.162	**0.277**	**0.098**	**0.190**	**0.183**	**0.329**	**0.483**	**0.476**	**0.372**	**0.426**
6. Baker/Howland	0.063	**0.104**	-0.023	0.032	-0.006	-	-0.004	**0.089**	0.027	0.041	-0.125	0.025	**0.241**	**0.077**	**0.164**	**0.169**	**0.309**	**0.361**	**0.457**	**0.341**	**0.399**
7. Kwajalein	**0.114**	**0.135**	0.015	0.030	0.059	0.036	-	0.032	-0.001	-0.001	-0.133	0.022	**0.151**	0.026	**0.097**	**0.104**	**0.229**	**0.328**	**0.381**	**0.242**	**0.304**
8. Pohnpei	0.019	0.015	-0.025	-0.014	-0.034	-0.019	0.029	-	0.050	0.002	-0.063	0.073	0.009	0.004	0.021	0.007	**0.079**	**0.280**	**0.225**	0.067	**0.135**
9. Saipan	**0.109**	**0.107**	-0.011	0.040	0.055	0.015	0.009	0.054	-	0.015	-0.089	0.048	**0.169**	0.017	**0.077**	**0.118**	**0.233**	**0.355**	**0.372**	**0.244**	**0.300**
10. Palau	**0.097**	**0.103**	0.013	0.002	0.056	0.050	-0.020	0.050	0.006	-	0.033	0.035	0.060	0.004	0.041	0.045	**0.145**	**0.286**	**0.292**	0.138	**0.202**
11. Lizard Is	0.008	-0.011	0.009	-0.015	-0.009	0.036	0.052	-0.020	0.040	-0.111	-	0.128	0.174	-0.045	0.040	0.033	**0.187**	**0.432**	**0.348**	0.222	**0.290**
12. Philippines	0.140	0.205	-0.039	-0.003	0.087	0.020	-0.095	0.072	-0.088	-0.093	0.070	-	0.043	0.003	-0.010	0.047	0.046	0.160	**0.147**	0.107	0.082
13. Bali	**0.169**	**0.179**	0.072	0.063	**0.126**	**0.112**	-0.015	0.096	0.039	-0.014	0.092	-0.077	-	**0.055**	0.006	-0.024	0.001	**0.151**	**0.093**	-0.028	0.022
14. Scott Reef	**0.103**	**0.085**	0.007	0.030	0.057	0.035	-0.001	0.050	-0.027	-0.012	0.042	-0.095	0.017	-	0.012	**0.046**	**0.124**	**0.224**	**0.257**	**0.115**	**0.177**
15. Rowley Shoals	**0.165**	**0.156**	0.062	0.067	**0.119**	**0.099**	0.003	0.100	0.022	-0.010	0.094	-0.084	-0.016	0.017	-	0.024	**0.059**	**0.137**	**0.155**	0.035	**0.084**
16. Christmas Is	**0.271**	**0.263**	**0.169**	**0.167**	**0.224**	**0.206**	0.067	**0.196**	**0.116**	0.064	**0.190**	-0.004	0.014	**0.095**	0.014	-	0.020	**0.168**	**0.134**	0.010	0.057
17. Cocos/Keeling	**0.391**	**0.387**	**0.287**	**0.276**	**0.345**	**0.327**	**0.150**	**0.303**	**0.226**	**0.158**	**0.294**	0.105	0.081	**0.201**	**0.092**	0.009	-	0.030	**0.041**	-0.027	-0.007
18. Sumatra	**0.830**	**0.906**	**0.731**	**0.751**	**0.849**	**0.805**	**0.535**	**0.835**	**0.676**	**0.595**	**0.836**	0.570	0.461	**0.615**	**0.459**	0.277	0.147	-	-0.019	-0.045	-0.008
19. Diego Garcia	**0.765**	**0.798**	**0.694**	**0.703**	**0.759**	**0.737**	**0.565**	**0.738**	**0.659**	**0.604**	**0.734**	**0.644**	**0.511**	**0.611**	**0.498**	**0.351**	**0.228**	0.039	-	0.014	-0.009
20. Oman	**0.790**	**0.861**	**0.695**	**0.718**	**0.801**	**0.760**	**0.521**	**0.620**	**0.648**	**0.580**	**0.775**	**0.379**	**0.472**	**0.592**	**0.453**	**0.299**	0.174	-0.034	-0.013	-	-0.028
21. Seychelles	**0.778**	**0.831**	**0.688**	**0.706**	**0.775**	**0.742**	**0.521**	**0.741**	**0.640**	**0.582**	**0.738**	**0.595**	**0.475**	**0.596**	**0.476**	**0.337**	**0.216**	0.109	-0.005	0.055	-

Pairwise Φ_{ST} values for cytochrome *b* data are below diagonal and pairwise F_{ST} values for multi-locus nuclear data set are above diagonal. Significant comparisons are in bold. We maintained $\alpha = 0.05$ among all pairwise tests by controlling for the false discovery [51,52]. The corrected $\alpha = 0.008$.

Figure 3.1. Map of study area. Pie charts represent the ratio of individuals at each location with either the Pacific or Indian Ocean lineage as defined in Figure 3.2. (Photo credit: Luiz Rocha.)

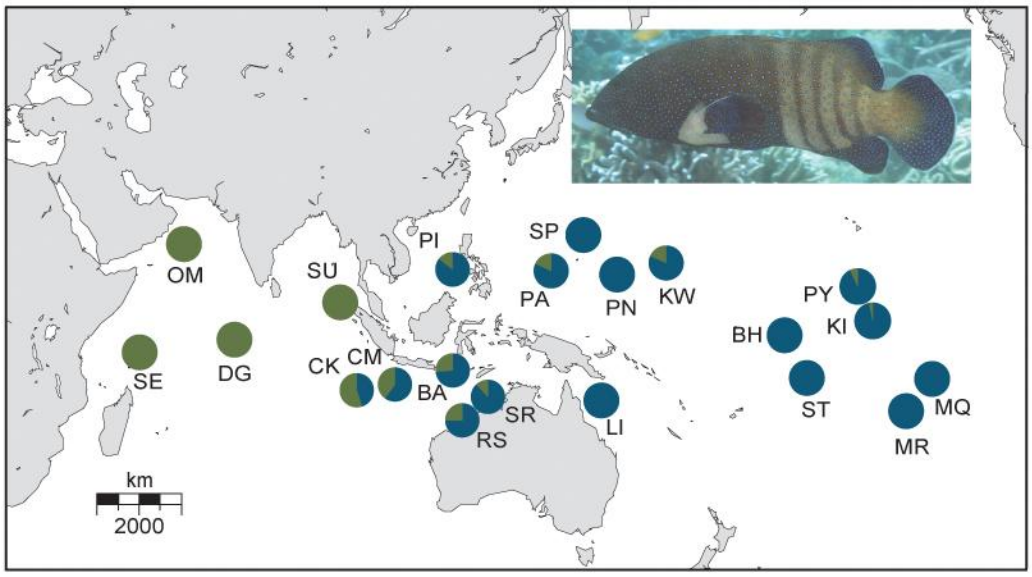


Figure 3.2. **Median-joining networks for *Cephalopholis argus*.** Networks were constructed using the program NETWORK 4.5 [43] for (a) 550 cytochrome *b* sequences (b) alleles at GnRH intron from 488 individuals, and (c) alleles at S7 intron for 490 individuals. Each circle represents one mitochondrial haplotype or nuclear allele with the area of each circle proportional to the number of that particular haplotype or allele in the data set; dashes represent hypothetical haplotypes or alleles; colors represent collection location (see key).

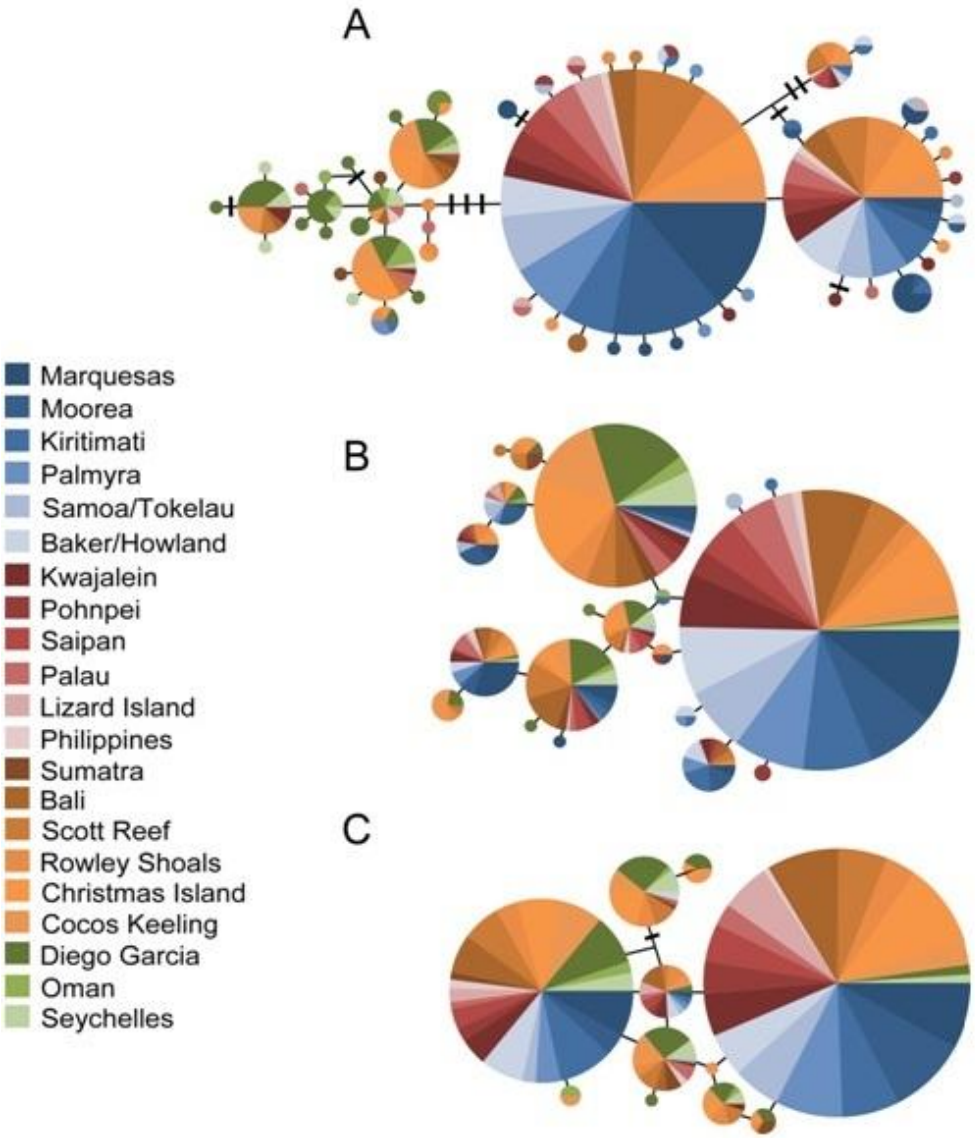


Figure 3.3. **Phylogenetic tree of C*ephalopholis argus* cytochrome *b* haplotypes.** The best maximum likelihood tree generated using program default settings in RAxML [44] and rooted using two congenerics (*C. urodeta* and *C. taeniops*). Bootstrap support values were calculated using default settings with 1000 replicates. For comparison neighbor-joining bootstrap values (1000 bootstrap replicates) and Bayesian posterior probabilities are presented. Colored bars delineate the Pacific and Indian Ocean lineages separated by three fixed differences (see figure 3.1).

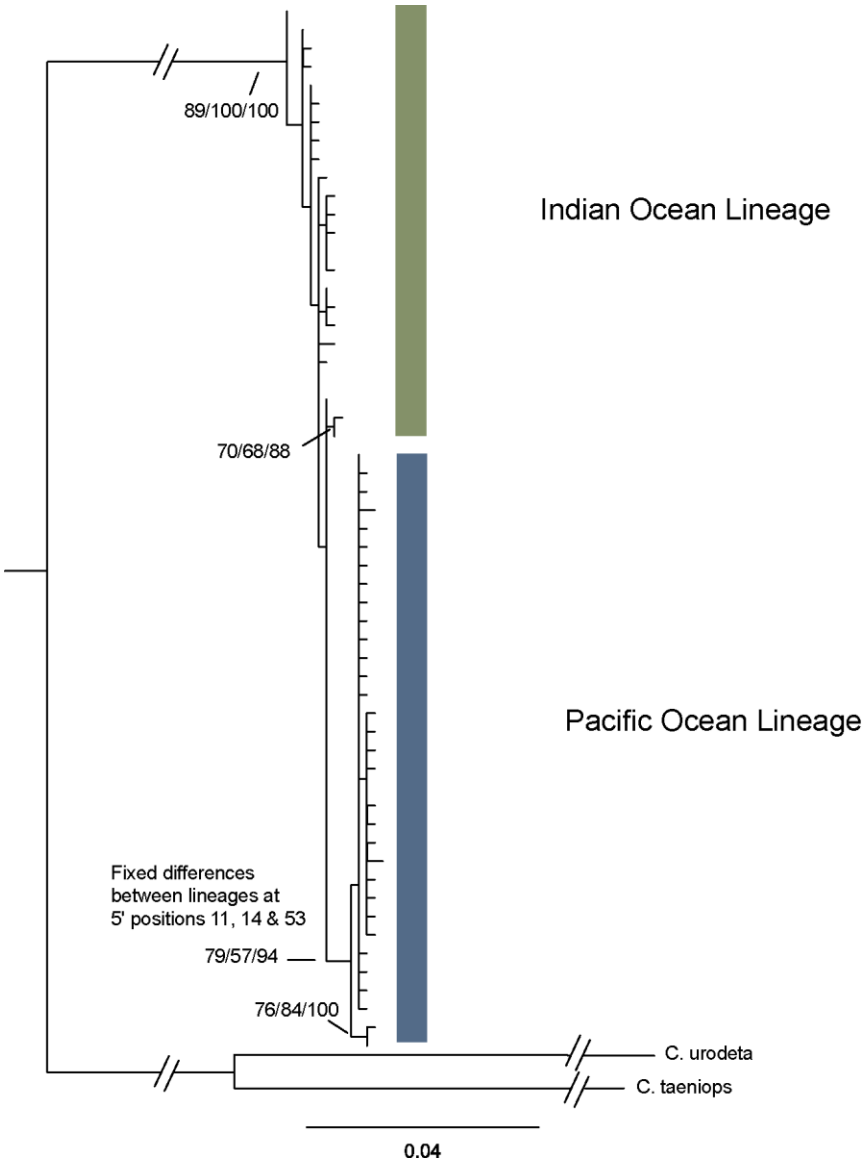


Figure 3.4. Mismatch distribution for *Cephalopholis argus*. Mismatch distribution based on 550 cytochrome *b* sequences from twenty-one populations. The dark colored line represents the observed and light colored line is the simulated pairwise differences as reported by DnaSP 5.0 [39]. The Harpending's raggedness index as calculated in ARLEQUIN 3.5 [42] and corresponding *P*-value are shown.

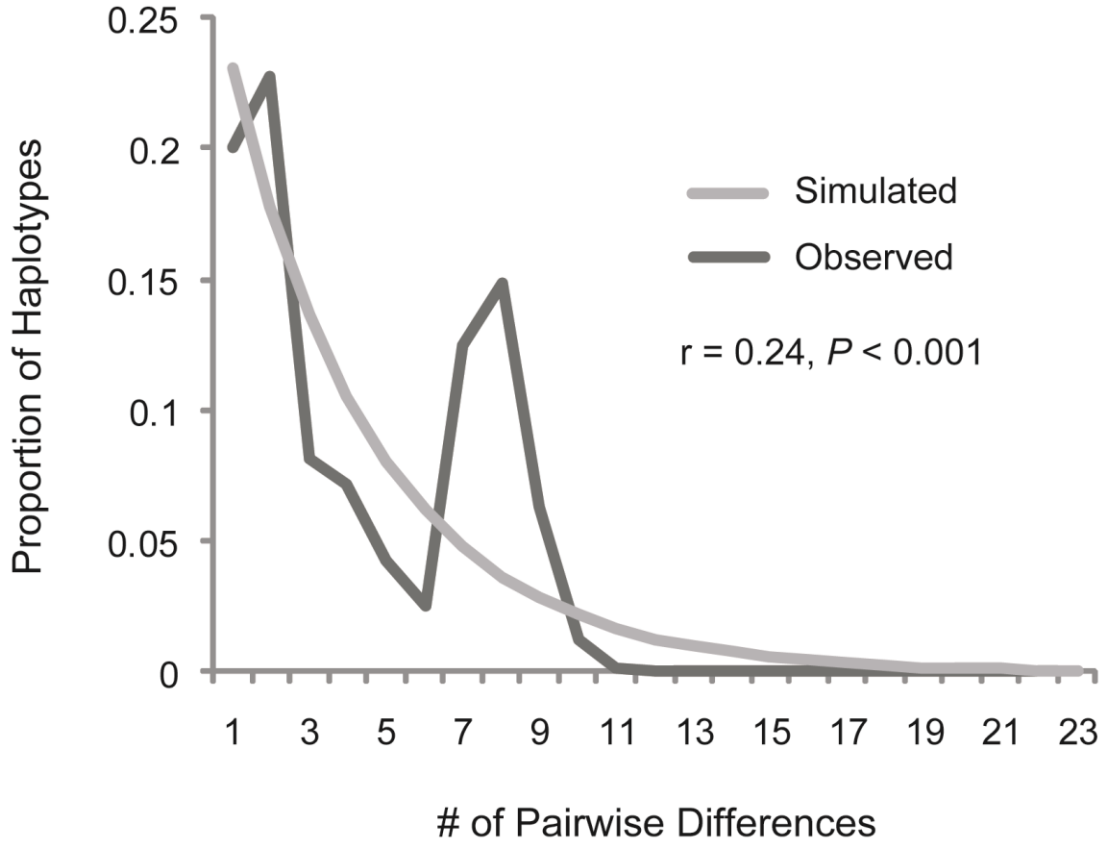


Figure 3.5. Migration rates for *Cephalopholis argus*. Migration rates (*Nm*: where *N* is effective female population size and *m* is migration rate) based on cytochrome *b* sequences calculated using MIGRATE 3.1.6 [54,55]. Locations with non-significant pairwise Φ_{ST} values were grouped (see Table 3.2). The direction of migration is indicated. Numbers of migrants per generation between geographic regions are reported with 95% confidence intervals in parentheses.

Groups	# migrants
East to West Migration	
FP → IB	16.7 (0.9 – 185.4)
FP → CW	87.3 (28.5 – 174.0)
FP → EI	2.2 (0.0 – 15.4)
CW → IB	7.1 (0.0 – 159.6)
CW → EI	2.1 (0.0 – 17.2)
IB → EI	1.8 (0.0 – 17.6)
FP → WI	0.0 (0.0 – 2.5)
CW → WI	0.0 (0.0 – 2.7)
IB → WI	0.0 (0.0 – 3.0)
EI → WI	0.0 (0.0 – 3.9)
West to East Migration	
WI → IB	2.7 (0.0 – 6.5)
WI → EI	1.5 (0.0 – 9.7)
WI → CW	6.6 (0.1 – 38.5)
EI→CW	11.7 (0.4 – 72.4)
EI→IB	39.0 (5.3 – 224.0)
IB → CW	8.8 (0.0 – 69.2)
CW → FP	0.4 (0.0 – 4.4)
IB → FP	0.2 (0.0 – 3.1)
EI → FP	0.2 (0.0 – 2.7)
WI → FP	0.0 (0.0 – 1.3)

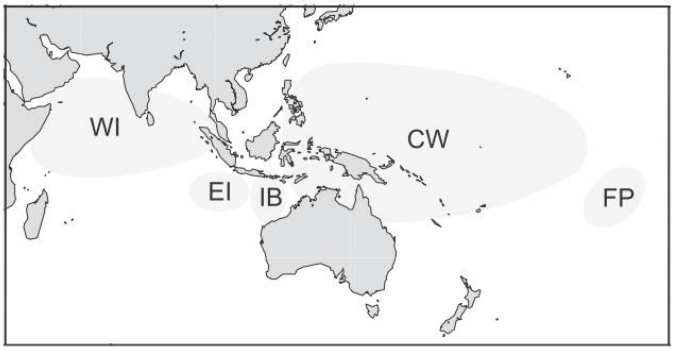
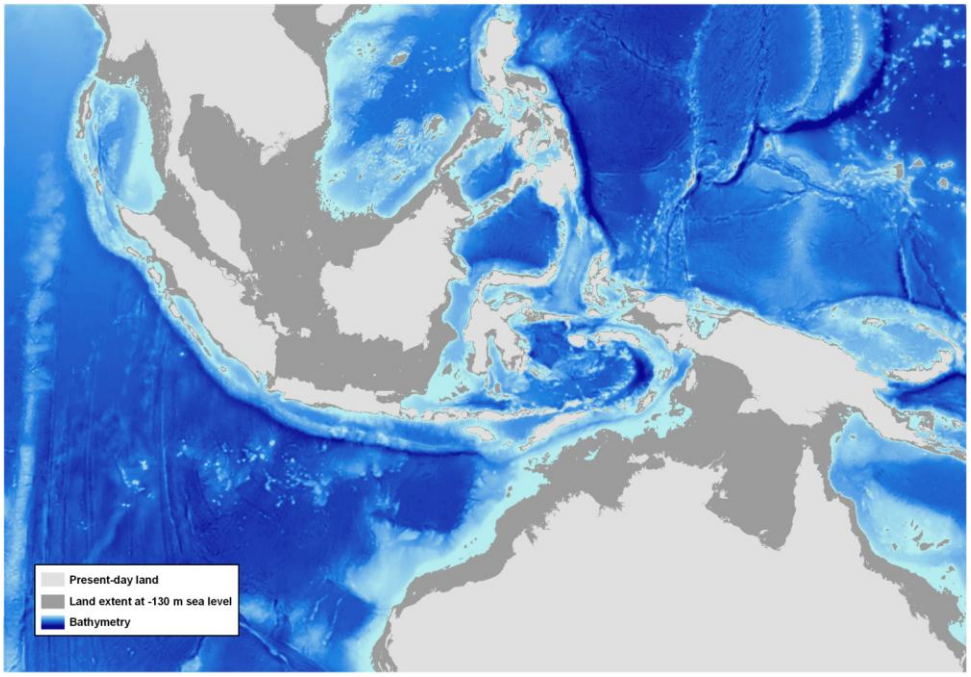


Figure 3.6. Map of Indo-Malaysia region during glacial maxima. Map shows the effect of lowered sea level on habitat in the region during Pleisotocene glacial maxima (Figure credit: Eric Franklin).



CHAPTER FOUR

Genetic consequences of introducing allopatric lineages of Bluestriped Snapper (*Lutjanus kasmira*) to Hawai‘i

Published as: Gaither, M.R., B.W. Bowen, R.J. Toonen, S. Planes, V. Messmer, J. Earle, D.R. Robertson. 2010. Genetic consequences of introducing allopatric lineages of Bluestripe Snapper (*Lutjanus kasmira*) to Hawai‘i. Molecular Ecology 19: 1107 - 1121.

Abstract

A half century ago the State of Hawaii began a remarkable, if unintentional, experiment on the population genetics of introduced species, by releasing 2431 Bluestriped Snappers (*Lutjanus kasmira*) from the Marquesas Islands in 1958 and 728 conspecifics from the Society Islands in 1961. By 1992 *L. kasmira* had spread across the entire archipelago, including locations 2000 km from the release site. Genetic surveys of the source populations reveal diagnostic differences in the mtDNA control region (d = 3.8%; ϕ_{ST} = 0.734, $P < 0.001$) and significant allele frequency differences at nuclear DNA loci (F_{ST} = 0.49; $P < 0.001$). These findings, which indicate that source populations have been isolated for approximately half a million years, set the stage for a survey of the Hawaiian Archipelago (N = 385) to determine the success of these introductions in terms of genetic diversity and breeding behavior. Both Marquesas and Society mtDNA lineages were detected at each survey site across the Hawaiian Archipelago, at about the same proportion or slightly less than the original 3.4:1 introduction ratio. Nuclear allele frequencies and parentage tests demonstrate that the two source populations are freely interbreeding. The introduction of 2431 Marquesan founders produced only a slight reduction in mtDNA diversity (17%), while the 728 Society founders produced a greater reduction in haplotype diversity (41%). We find no evidence of genetic bottlenecks between islands of the Hawaiian Archipelago, as expected under a stepping-stone model of colonization, from the initial introduction site. This species rapidly colonized across 2000 km without loss of genetic diversity, illustrating the consequences of introducing highly-dispersive marine species.

Introduction

The Bluestriped Snapper, *Lutjanus kasmira* (Forsskål, 1775), is a widely distributed coral reef fish with a natural range from South Africa to the Marquesas and Line Islands in the central

Pacific (Fig. 4.1). This natural range does not include the Hawaiian Islands, which has only a subset of the Indo-Pacific flora and fauna, lacking many taxa such as most shallow water snappers and groupers (Randall 2007). In an effort to fill a perceived empty ecological niche, and to enhance local fisheries, the Hawaii Division of Fish and Game (HDFG) introduced *L. kasmira*, among other reef fishes, to the Hawaiian Islands (Oda & Parrish 1982, Randall 1987). The introduction of *L. kasmira* was conducted in two discrete events. In preparation for introduction, juvenile fish (approximately 100-120 grams) were caught using hand lines and transferred to floating pens. When sufficient fish had been captured, the fish were transferred to the bait wells of the transport vessel and brought to Hawaii for transplant. In 1958, 2431 fish from Nuku Hiva in the Marquesas Islands, and, in 1961, 728 fish from Moorea in the Society Islands were released on Oahu (Fig. 4.1; 3.4: 1 ratio; HDFG records). *L. kasmira* quickly spread through the archipelago at a rate of about 60 km per year (Oda & Parrish, 1982; Randall 2007). In 1992, just 34 years after the initial introduction, *L. kasmira* was recorded at the far reaches of the archipelago at Midway Atoll (Randall *et al.* 1993) over 2000 km from the release site. This successful introduction provides a number of research opportunities relating to understanding founder/colonization processes.

The Marquesas and Society populations of *L. kasmira* are phylogenetically distinct with diagnostic differences in the mitochondrial genome (average sequence divergence for cytochrome *b* = 0.53%; Gaither *et al.* 2010). The introduction of *L. kasmira* to the Hawaiian Islands from two genetically divergent populations, resulted in the sympatric distribution of lineages that have been separated for about half a million years (Gaither *et al.* 2010). The genetic divergence between these two populations is a result of the phylogenetic distinction of the Marquesas population relative to other Indo-Pacific populations (Gaither *et al*. 2010). The

Marquesas have the third highest level of endemism among shorefishes (11.6%) in Oceania (Randall 2001) and high levels of genetic differentiation in three of five species of non-endemic reef fishes examined to date (Planes & Fauvelot 2002; Craig *et al*. 2007; Schultz *et al*. 2007; Gaither *et al.* 2010). The distinction of the Marquesas shorefish fauna has been attributed to a combination of geographic isolation (enhanced by the westerly Southern Equatorial Current) and adaptation to unusually variable sea temperatures (Randall 2001; Gaither *et al.* 2010). Because geographic isolation and ecological divergence may both promote speciation in fishes (Rogers & Bernatchez 2006; Rocha & Bowen 2008), we ask whether *L. kasmira* populations separated for half a million years can freely interbreed in sympatry. The genetic distinctiveness of the two source populations provides an opportunity to identify the descendants of the Hawaiian introductions, to assess their relative success in the archipelago, and to determine if the two genetic lineages are mixing.

Colonization events, including human-mediated introductions, often involve a severe reduction in population size and isolation from the larger parental population. The dramatic decrease in effective population size that accompanies such founder events is expected to lead to decreased genetic diversity (Nei *et al.* 1975). However, the accumulation of data indicates that genetic bottlenecks in introduced populations are not an invariable outcome (Roman & Darling 2007; Bossdorf *et al.* 2005; Wares *et al.* 2005). Interpreting patterns of genetic diversity in introduced populations is confounded by the fact that in most cases the source population and the number of founding individuals are unknown. Under these conditions, researchers must reconstruct the history of introductions by combining molecular and geographic data to identify source populations (Wares *et al.* 2005). In some cases high genetic diversity in the introduced range can be attributed to admixture of genetically divergent populations (Kolbe *et al.* 2004;

Genton *et al.* 2005; Carmeron *et al*. 2008; Rosenthal *et al.* 2008). In cases where only a single source population can be identified, high genetic diversity in the introduced range is often attributed to either a large number of colonizers, rapid population expansion following the founder event, or both (Hassan *et al.* 2003; Stepien *et al.* 2005). Discussions concerning the effect of founder events on genetic diversity would be greatly informed if more empirical data concerning the effects of founder population size on genetic diversity were available. Intentional and well documented introductions, where the source population and founder population size are confidently known, offer powerful test cases.

The intentional introduction of *L. kasmira* to the Hawaiian Islands provides a rare opportunity to directly evaluate the effects of founder population size on genetic diversity in recently established populations. Here we capitalize on two unique aspects of the introduction of *L. kasmira* to Hawaii: (i) the introduction occurred in two well documented events with known numbers of founders and source populations; (ii) the source populations (Nuka Hiva in the Marquesas Islands and Moorea in the Society Islands) are genetically distinct, allowing us to identify their descendents. We employ both mitochondrial and nuclear sequence data to ask the following questions 1) Did fish from both source populations become established in the Hawaiian Islands? If so, 2) how are their descendents distributed in the archipelago? 3) Were fish from both source populations equally successful at reproducing and colonizing the islands? 4) Are these genetically divergent populations interbreeding in the Hawaiian Islands? 5) Is there evidence of genetic bottlenecks at the introduction site or as the fish spread throughout the archipelago? The circumstances of this study offer unprecedented opportunities to study species introductions and invasions, pertinent to management of marine resources including the Papahānaumokuākea Marine National Monument (PMNM) that traverses 2000 km of the

northwestern (NW) Hawaiian Islands. At least 350 alien marine species occupy the inhabited Main Hawaiian Islands (Eldredge & Smith 2001) and few studies have addressed the threat these aliens pose to the uninhabited (and nearly pristine) ecosystems of the NW Hawaiian Islands. Hence an ongoing concern is the level of connectivity between the Main Hawaiian Islands and the NW Hawaiian Islands (see Eble *et al*. 2009). Here we document an extreme scenario of rapid colonization into the NW Hawaiian Islands, in numbers that are sufficient to retain the genetic diversity of parent populations.

Materials and methods

Study species

Lutjanus kasmira has broad habitat preferences, occupying hard substrata from shallow waters to at least 265 m (Randall 1987) and has a generalized predatory diet that includes fish, crustaceans, and cephalopods (Randall & Brock 1960; Oda & Parrish 1982). This species reaches sexual maturity at 1-2 years (Rangarajan 1971; Morales-Nin & Ralston 1990) and engages in mass spawning (Suzuki & Hioki 1979). Long-distance movement between isolated patches of adult habitat (reefs) occurs during a highly-dispersive pelagic larval phase that, in other species of *Lutjanus*, lasts 20-44 days (Zapata & Herron 2002; Denit & Sponaugle 2004).

Collections

A total of 385 specimens of *Lutjanus kasmira* were collected from 10 locations across the Hawaiian Archipelago by scuba divers using polespears (Table 4.1, Fig. 4.2). Specimens from the uninhabited NW Hawaiian Islands were obtained during research expeditions on the NOAA R/V Hi'ialakai, as part of an initiative by the Papahānaumokuākea Marine National Monument (http://hawaiireef.noaa.gov/) to monitor and characterize this vast protected area. Tissue samples (fin clips or gill filaments) were preserved in either 95% ethanol (EtOH) or saturated NaCl

solution (Amos & Hoelzel 1991), and stored at room temperature. Fifty *L. kasmira* samples from each of the Marquesas and Society source populations, previously analyzed in Gaither *et al.* (2009), were also used in this study.

DNA extraction, PCR amplifications and sequencing

All DNA extraction, PCR cycling, cloning, and sequencing protocols used here are identical to those in Gaither *et al.* (2009). The growth hormone (GH) and adenine nucleotide transporter translocase (ANT) intron sequences obtained from each of the Marquesas and Society populations in Gaither *et al.* (2009) were used in this study [GenBank accession numbers FJ754178-FJ754184 (GH intron), FJ754157-FJ754177 (ANT intron)]. All 385 specimens of *L. kasmira* collected from the Hawaiian Islands were sequenced at these two loci. Additionally, approximately 215 bp of the third intron in the gonadotropin-releasing hormone 3 (GnRH3-3) were amplified using the primers GnRH3F (5'-GCCCAAACCCAAGAGAGACTTAGACC-3') and GnRH3R (5'-TTCGGTCAAAATGACTGGAATCATC-3') (Hassan *et al.* 2002) and approximately 575 bp of the mitochondrial control region were amplified using the primers Lutjf1 (5'-GCACTCTGAAATGTCAAGTGAAAGG-3') and CRA (5'-TTCCACCTCTAACTCCCAAAGCTAG-3') (Lee *et al.* 1995) in all 484 samples (Hawaii = 385, Marquesas = 50, and Society = 49). PCR protocols and cycling conditions for both the GnRH3-3 intron and the mtDNA control region were carried out as described in Gaither *et al.* (2009) using an annealing temperature of 60°C. Due to the presence of multiple indels at the GnRH3-3 locus, that would require extensive cloning to phase alleles, analysis of this locus was restricted to the presence or absence of a ten bp indel near the reverse priming site. The presence of the indel was confirmed by cloning ten individuals and comparing alleles to direct sequences. The allelic state of the remaining individuals was inferred by direct sequencing.

Sequences for each locus were aligned and edited using SEQUENCHER 4.8 (Gene Codes, Ann Arbor, MI, USA) and trimmed to a common length. The mtDNA control region contained multiple indels which varied from 1-3 bp in length. Alignment of the mtDNA sequences was confirmed using default parameters in CLUSTAL W 1.81 (Thompson *et al.* 1994). Unique mtDNA haplotypes and nuclear alleles were identified with the merge taxa option in MACCLADE 4.05 (Maddison & Maddison 2002). All control region haplotypes and nuclear alleles unique to Hawaii were deposited in GenBank [accession numbers: GU123931-GU124148 (control region), GU192444-GU192447 ANT intron)]

Data analysis

Mitochondrial control region

Summary statistics including mtDNA haplotype diversity (h), and nucleotide diversity (π) were estimated using algorithms in Nei (1987) as implemented in ARLEQUIN 3.11 (Excoffier *et al.* 2005). A statistical parsimony network was constructed using the program TCS 1.21 (Clement *et al.* 2000). The resulting network was simplified using standard tie-breaking rules. In keeping with the cytochrome *b* data in Gaither *et al.* (2009) the control region sequences in the Marquesas and Society samples fell into two distinct lineages. Average percent difference between populations was calculated by dividing the average number of nucleotides (corrected; Tamura & Nei 1993) that differ between the two source populations (as calculated in ARLEQUIN) by the total number of base pairs. The average percent difference between populations is reported here as sequence divergence (d).

The number of individuals from the Hawaiian Islands that grouped with either the Marquesas or the Society mtDNA lineage was calculated, and deviations from the initial introduction ratio of 3.4 Marquesas/1.0 Society were tested using Fisher's exact test (Sokal &

Rohlf 1995). To test for the loss of genetic diversity in the introduced range, while controlling for unequal sample sizes (Leberg 2002), we estimated haplotype richness using rarefaction analysis. For this method we determined the haplotype frequency distribution for the largest sample in the comparison. From this larger sample we randomly subsampled haplotypes (size of subsample = N of smaller sample) with replacement 10,000 times to estimate the number of haplotypes that would occur in the smaller sample. We compared the distribution of the subsamples with the number of haplotypes found in the smaller sample. *P*-values were calculated based on the number of times in 10,000 subsamples that as many or more haplotypes were found in the larger sample as found in the smallest. Rarefaction curves plotting the number of individuals sampled against the expected number of mitochondrial haplotypes were constructed using ANALYTIC RAREFACTATION 1.4 (UGA Stratigraphy Lab website; http://www.uga.edu/~strata/software/).

The Akaike Information Criterion in MODELTEST 3.7 (Posada & Crandall 1998) was used to determine the mutational model that best fit the control region data. The best fit model is TVM+I+G with equal rates for all sites and a Ti/Tv ratio of 10.13. Because this model is not implemented in ARLEQUIN (Excoffier *et al.* 2005), the most similar model available (Tamura & Nei 1993) was employed using a gamma value of 0.77, a transversion weighting of 10.13 and a transition and deletion weight of 1.0. To test for population genetic structure in Hawaii *L. kasmira*, an analysis of molecular variance (AMOVA) was performed in ARLEQUIN using 20,000 permutations. An analogue of Wright's F_{ST} (ϕ_{ST}), which incorporates a model of sequence evolution, was calculated for the entire data set and for pairwise comparisons among all locations. We maintained $\alpha = 0.05$ among all pairwise tests by controlling for the false discovery rate as recommended by Benjamini & Yekutieli (2001) and reviewed by Narum (2006).

Nuclear Introns

Observed (H_O) and expected (H_E) heterozygosities were calculated for each locus and an exact test of Hardy-Weinberg equilibrium (HWE) using 100,000 steps in a Markov chain was performed using ARLEQUIN. Additionally, average H_E was calculated for the multi-locus data set. Linkage disequilibrium between the three nuclear loci was assessed using the likelihood ratio test with 20,000 permutations in ARLEQUIN. F_{ST} was calculated for the entire data set and for pairwise comparisons between locations. The false discovery rate among multiple comparisons was controlled as described above.

Tests for loss of genetic diversity in the introduced range were conducted with BOTTLENECK 1.2.02 (Piry *et al.* 1999) using the infinite alleles mutation (IAM) model (Kimura & Crow 1964). The loss of rare alleles was evaluated using the mode-shift test (Piry *et al.* 1999). The Wilcoxon signed rank test, which assumes that populations in mutation-drift equilibrium have an equal probability of heterozygote excess or deficit, was used to detect genetic bottlenecks (Cornuet & Luikart 1996).

Interbreeding

To determine if *L. kasmira* Hawaiian descendents from the Marquesas and Society Islands are interbreeding we employed the genealogical-frequency and individual-assignment methods of Nason *et al.* (2002). This method assigns individuals to one of six genealogical categories using multi-locus diploid data. Individuals are classified as either pure parental (P_1 and P_2), crosses between pure parentals (F_1), crosses between F_1 individuals (F_2) or backcrosses (BP_1 and BP_2). The program uses maximum-likelihood estimates to assign each individual to one of the six genealogical classes while providing estimates of statistical power for correct classification. For comparison, the Bayesian statistical model developed by Anderson &

Thompson (2000), which computes the posterior probability that an individual belongs to each of the hybrid classes (P_1, P_2, F_1, F_2, BP_1 and BP_2) was employed using the program default settings. A third method of testing for interbreeding utilized the chi-square (χ^2) goodness of fit (Sokal & Rohlf 1995) to test whether the nuclear alleles at each locus were randomly distributed among the two mitochondrial lineages in the introduced range.

Results

Distribution of descendents of the two source populations

Mitochondrial control region

We resolved a 521 bp segment of the mtDNA control region in 484 individuals yielding 218 haplotypes, with a few common haplotypes, 123 haplotypes observed in single individuals, and 41 haplotypes observed in two individuals. The number of specimens (N), the number of haplotypes (N_h), the number of haplotypes observed in single individuals (N_s), h, and π per location are listed in Table 4.1. There were no shared haplotypes between the two source populations (Marquesas and Society Islands). A statistical parsimony network demonstrated that samples from the two source populations fell into distinct lineages separated by 22 steps (Fig. 4.3; d = 3.8% between source populations). Haplotypes observed in one or two specimens (singletons and doublets) were found at every sample site and exclusion of these haplotypes from the parsimony network did not change the overall structure (Fig. 4.3). Haplotypes observed in the introduced range (Oahu, Kona, Hilo, Maui Nui, Kauai, Necker, French Frigate Shoals (FFS), Maro, Midway, and Kure) grouped with either the Marquesas or Society lineages (Fig. 4.3).

The overall ratio of the number of *L. kasmira* samples from the Hawaiian Islands that fell into the Marquesas lineage to those that fell into the Society lineage was 2.3:1. This value was significantly different than the introduction ratio of 3.4:1 (Fisher's exact text, P = 0.027).

Among the ten sample locations scattered across the Hawaiian Archipelago only Kona on Hawaii Island, differed significantly from the 3.4:1 introduction ratio, with a ratio of 1.3:1 (Fisher's exact text , $P = 0.036$) (Fig. 4.2). Once these samples were removed from the analysis the overall ratio (2.6:1) was not significantly different than the introduction ratio (Fisher's exact text, $P = 0.131$).

Nuclear introns

We resolved 148 bp of the GH intron in 482 specimens and 168 bp of the ANT intron in 471 specimens (Table 4.2). Three polymorphic sites yielded 4 alleles at the GH locus and 13 polymorphic sites yielded 15 alleles at the ANT locus. The GnRH3-3 intron was scored for the presence or absence of a 10 bp indel in 480 specimens. Summary statistics are listed in Table 4.2.

When all locations, from the native and introduced ranges, were grouped together there was a significant deviation from HWE (Hardy-Wienberg Equilibrium) expectations at the GH and ANT loci ($P = 0.02$ for each) (Table 4.2). In each case an excess of homozygotes was detected. When samples were divided by archipelago (Marquesas, Society, and Hawaiian Islands) the Marquesas and Society populations deviated from HWE expectations, with an excess of homozygotes, at the ANT locus ($P = 0.044$ and $P = 0.032$ respectively). No evidence of linkage disequilibrium between pairs of nuclear loci was detected ($P > 0.05$) within populations from each of the three archipelagos.

The number of nuclear alleles at each locus was similar for the two source populations (Tables 4.2, 4.6) however; there were strong shifts in allele frequencies between the Marquesas and Society Islands (Table 4.6). Populations in the introduced range had allele frequencies intermediate between the two source populations (Table 4.6). Three putative private alleles are found in each source population (Table 4.6; Marquesas = GH3, A1, A11; Society = GH4, A4,

A5). The presence of these alleles at widely separated locations in the introduced range (Table 4.6) provides additional evidence that descendents of both source populations spread throughout the archipelago. As expected for introduced populations of mixed lineages, many of the Hawaiian samples had a greater number of nuclear alleles and higher heterozygosities (H_O and H_E at ANT and GnRH3-3 loci) than either of the source populations (Table 4.2).

Population structure

Pairwise comparisons indicate significant population structure between the two source populations (Marquesas and Society Islands) with mtDNA $\phi_{ST} = 0.734$ ($P < 0.001$) and nDNA $F_{ST} = 0.49$ ($P < 0.001$) (Table 4.3). The Marquesas and Society populations were also significantly different than each of the ten Hawaiian populations (Oahu, Kona, Hilo, Maui Nui, Kauai, Necker, FFS, Maro, Midway, and Kure) (Table 4.2). In the introduced range, there was no population structure detected in the mtDNA (overall $\phi_{ST} = 0.001$, $P = 0.38$) or nDNA ($F_{ST} = 0.001$, $P = 0.30$) data sets (Table 4.3).

Interbreeding of the two populations in the Hawaiian Islands

The likelihood model of Nason *et al.* (2002) indicated that approximately 31% of the individuals from the Hawaiian Islands were F_1 X F_1 crosses (F_2 genealogical class) while the remainder (~69%) were F_2 X P_1 backcrosses (BP_1 genealogical class). The program did not assign any individual from the introduced range to either pure parental class (P_1 or P_2) or to the P_1 X P_2 cross (F_1 geneological class). This indicates that all assayed specimens of *L. kasmira* in the Hawaiian Islands are of mixed Marquesas and Society descent. The Anderson & Thompson (2002) model indicated similar results to the Nason model (data not shown).

The chi-square test corroborated the findings of the Nason *et al*. (2002) model, demonstrating that nuclear allele frequencies at the GH and ANT loci were not significantly

different than expected if the alleles were randomly distributed between the Marquesas (χ^2 = 0.227, P = 0.99; χ^2 = 0.157, P = 0.99 respectively) and Society (χ^2 = 0.256, P = 0.96; χ^2 = 0.803, P = 0.79) mitochondrial lineages in Hawaii. The Society mitochondrial lineage in Hawaii deviated significantly from a random distribution of alleles at the GnRH3-3 locus (χ^2 = 4.718, P = 0.03) while the Marquesas mitochondrial lineage did not (χ^2 = 1.729, P = 0.19).

Genetic consequence of founder event

The Marquesas and Society populations had high mtDNA haplotype diversity (h = 0.997 and 0.970 respectively). All ten populations in the introduced range (Oahu, Kona, Hilo, Maui Nui, Kauai, Necker, FFS, Maro, Midway, and Kure) had similarly high h values (h = 0.985 – 1.000; "All Data" Table 4.1). Using the parsimony network in Fig. 4.3 we divided the Hawaiian samples into either the Marquesas or Society mitochondrial lineage (Table 4.1). The mtDNA haplotype diversity values in the Hawaiian Islands ranged from 0.989 to 1.00 for the Marquesas lineage and from 0.733 to 1.00 for the Society lineage.

Due to the lower sensitivity of heterozygosity to losses of genetic diversity (Nei *et al*. 1975) we restricted our statistical comparisons of diversity loss to haplotype richness which we compare at the archipelago level. We observed 47 haplotypes in the Marquesas (N = 50) and 31 haplotypes in the Society Islands (N = 49) (Table 4.4). By creating haplotype frequency distributions for the larger Hawaiian sample sets and by randomly subsampling (10,000 times with replacement) these populations we found evidence of a small but significant decrease (17%) in haplotypes from the Marquesas lineage in Hawaii which had a mean of 39.0 haplotypes per subsample (N = 50, P-value < 0.001). A greater decrease (41%) was detected in the Society lineage in Hawaii with a mean of 18.3 haplotypes (N = 49, P-value < 0.001). The difference in

loss of haplotypes (17% versus 41%) was marginally significant (Fisher's exact test, $P = 0.058$) (Table 4.4).

Rarefaction curves, that plotted the number of individuals sampled against the expected number of mitochondrial haplotypes, were constructed (Fig. 4.4). Samples from the introduced range were separated by mitochondrial lineage and compared with their respective source population (Fig. 4.4). Due to the large confidence intervals (95%) there was no significant difference between expected number of mtDNA haplotypes in the native and introduced ranges at low sample sizes. However, as sample size increased ($N > 40$) the curves no longer over-lapped and a loss of mtDNA haplotypes in the introduced range became evident in the Society, but not the Marquesas lineage (Fig. 4.4).

Using the nuclear allele frequency data (Table 4.6) there was no evidence of a genetic bottleneck in any of the introduced populations using the Wilcoxon signed rank test or the mode shift test implemented in BOTTLENECK. However, it should be noted that three loci do not provide high levels of power for these analyses and due to the presence of shared alleles between the source populations and admixture of lineages in the introduced range we did not attempt to statistically compare allelic richness values using the nuclear data set.

Discussion

Establishment and spread of L. kasmira *throughout the Hawaiian Islands*

The introduction of *L. kasmira* to the Hawaiian Islands from two populations with diagnostic differences in the mitochondrial genome (Fig. 4.3) and several private nuclear alleles (Table 4.6) enables us to trace the fate of their descendents in the introduced range. Individuals from both source populations became established in the Hawaiian Islands and both mtDNA lineages are found on every island and atoll of the archipelago (Table 4.6, Figs 4.2, 4.3). The

mtDNA data (Fig. 4.2) indicate that the overall ratio of the two lineages across the Hawaiian Islands is less than the 3.4:1 introduction ratio. Although one Hawaii Island population had a ratio significantly less than 3.4:1, the overall ratio among the remaining populations (2.6:1) was not significant, albeit with a tendency in the same direction.

Hawaii Division of Fish and Game (HDFG) records show that the initial 2431 Marquesas fish began to reproduce and spread very soon after the initial release. When the 728 Society fish were introduced, three years after the Marquesas fish, the former had already spread to Hawaii Island about 500 km to the southeast of Oahu. This, in combination with the fact that maturity occurs at 1-2 years of age, indicates that the ratio at the time of the introduction of Society fish likely was greater than 3.4:1. After the three year head-start and rapid spread of the Marquesas lineage, it might be predicted that its descendents now would be proportionately more numerous and widespread than those of the Society lineage. This is not the case, however. Instead the data indicate that numerically the Society lineage was able to "catch up" with and even surpass, the Marquesas lineage, most notably on Hawaii Island. Given the estimated time of divergence between these two source populations (approximately half a million years) and the differences in their native environments (Gaither *et al*. 2010) it is possible that population specific adaptations endowed Society-lineage fish with a higher fitness in the Hawaiian environment. However, such an advantage would be quickly lost by interbreeding. Another likely advantage the Society lineage had over the Marquesas lineage at the time of introduction was the presence of an established population when the former were released. For the Society lineage this could have alleviated many of the adverse consequences associated with small population size, such as difficulty in finding suitable mates.

The genetic data indicate that there is now a single population of *L. kasmira* in the Hawaiian Islands. We found no population structure in the mitochondrial data set across the archipelago, and only one of forty-five pairwise comparisons of the nuclear data set was significant after control for false discovery rate. The lack of genetic structure coupled with the maintenance of genetic diversity across the archipelago implies that there was little or no loss of genetic lineages, as would be expected under a stepping stone model of colonization, as the fish spread through the islands. Instead our data indicate that either *L. kasmira* colonized each island in large enough numbers to capture most of the standing genetic diversity, or gene flow between the islands is sufficient to homogenize the geographic distribution of the genetic diversity, or both.

The success of *L. kasmira* in Hawaii, as indicated by HDFG catch records and corroborated here by our genetic data, is especially notable because most other introductions of reef fishes to the Hawaiian Islands have failed. In the 1950's the Hawaiian Division of Fish and Game introduced 11 non-native snappers and groupers (Oda & Parrish 1982; Randall 1987). Six of the eleven species were introduced in numbers greater than 1500 individuals (HDFG records) but 50 years later only three are regularly recorded in Hawaiian waters. Besides *Lutjanus kasmira*, these include the Blacktail Snapper (*Lutjanus fulvus*) and the Peacock Grouper (*Cephalopholus argus*). Notably, neither *L. fulvus* nor *C. argus* has colonized north of French Frigate Shoals (FFS; Fig. 4.2). While *L fulvus* is not a common fish in the lower Hawaiian Islands, *C argus* is more common there than it is in its natural range (Meyer 2008).

In the field of invasion biology, an intense debate revolves around the factors that promote successful colonization of new habitat (Allendorf & Lundquist 2003; Kolar & Lodge 2001), particularly the genetic factors (Frankham 2005; Golani *et al.* 2007; Zayed *et al.* 2007).

Two of the primary factors that are thought to contribute to invasion success are large founder populations and multiple introduction events (Colautti *et al.* 2006; Lockwood *et al.* 2005). Certainly these conditions apply to the introduction of *L. kasmira* in Hawaii, and it is possible that introductions of two genetically distinct populations have yielded a more robust fish than either parental stock (hybrid vigor; Allendorf & Luikart 2007). Other traits that might apply specifically to the invasion success of *L. kasmira* include mass spawning (Suzuki & Hioki 1979), broad habitat preference (2 – 265 m depth; Randall 1987), and a generalist diet (Randall & Brock 1960; Oda & Parrish 1982). Range-wide mtDNA surveys also indicate that this species has a much more dispersive larval stage than either *Lutjanus fulvus* or *Cephalopholus argus* (Gaither *et al.* 2010; unpubl. data) which may explain why *L. kasmira* has swiftly colonized the entire archipelago, while the other two species have not.

Interbreeding and outbreeding

The Marquesas and Society source populations of *L. kasmira* demonstrate an average mitochondrial control region sequence divergence of $d = 3.8\%$. The control region appears to evolve at 3-10% per million years in shore fishes (Bowen *et al.* 2006; Lessios *et al.* 2008). Using this range as a first order approximation we estimate that these two populations have been separated for about half a million years (380,000 to 1,300,000 years), a value that overlaps the estimate from cytochrome *b* data from the same samples (265,000 to 530,000 years; Gaither *et al.* 2010). This time interval is sufficient to produce gamete incompatibility in allopatric populations of sea urchins (Lessios 1984; Palumbi & Metz 1991). However, three to four million years is insufficient to prevent gamete compatibility in geminate species of gobies (Rubinoff & Rubinoff 1971) or to prevent hybridization in trumpetfish species (Bowen *et al.* 2001). Populations which diverge sufficiently in allopatry might resume mating upon secondary

contact but resulting offspring could have lower fitness than purebred offspring. Reinforcement theory predicts that due to lower fitness of hybrids, natural selection will favour the evolution of prezygotic isolating mechanisms to maximize fitness, and therefore drive further diversification (Coyne & Orr 2004). We see no evidence for reinforcement in the hybridization tests we performed here. Notably, our study was conducted approximately thirteen generations after the initial introduction (see Materials and methods for references). If reproductive barriers existed or preferential mating occurred during initial contact of these two lineages, the genetic signature has been lost. Furthermore, the mass spawning behaviour of this species (Suzuki & Hioki 1979) may have reduced the potential for assortative mating by eliminating active mate choice.

Genetic consequences of founding events

Contrary to expectations, alien species often retain high levels of genetic diversity in their introduced range (Roman & Darling 2007; Bossdorf *et al.* 2005; Wares *et al.* 2005). In cases where individuals from genetically divergent source populations are introduced to the same region (admixture), there may actually be an increase in genetic diversity in the introduced range (Kolbe *et al.* 2004; Genton *et al.* 2005; Roman & Darling 2007). This is the case for *L. kasmira* in the Hawaiian Islands. At the archipelago level, *L. kasmira* in Hawaii exhibit similar or slightly higher heterozygosities and a similar or greater number of nuclear alleles than either source population. The only possible exception to this pattern (Kure) may simply be a sampling artifact due to small sample size (N = 9).

Introductions that involve a large number of individuals (high propagule pressure) are less likely to suffer the loss of rare alleles and heterozygosity associated with founder events (Lockwood *et al.* 2005). What is unclear is how many individuals are required to prevent such a loss of genetic diversity. The answer to this question is dependent on both the genetic diversity

of the taxa involved and patterns of survivorship following the founder event. Dlugosch & Parker (2007) reviewed 80 surveys of molecular variation in introduced species. These include 11 cases of intentional introduction where the number of individuals is confidently known and derived from a single source population (Table 4.5). In these eleven cases, which cover a variety of taxa, a loss of genetic diversity was detected in all but one case involving the introduction of less than 250 individuals. The introduction of 2385 Peacock Groupers (*Cephalopholis argus*) to the Hawaiian Islands (Planes & Lecaillon 1998) is the only example in this review that involved a founder population of greater than 250 individuals (Dlugosch & Parker 2007) and for this species the authors found no loss of genetic diversity in the introduced range. As with *C. argus*, we found that 2431 *L. kasmira* were sufficient to prevent a major loss of mtDNA haplotypes (17%) (Marquesas Lineage, Table 4.1). In contrast, with a founder population size of 728 individuals (Society Lineage, Table 4.1), we detected a larger decrease in haplotype richness (41%) indicating, that at least for *L. kasmira* introduced to Hawaii, this founder size is at the level where we begin to detect substantial losses of genetic diversity. This conclusion should be tempered with the recognition that even a small number of reproductive adults can retain genetic diversity over the short term (Spencer *et al.* 2000) and that even very large populations may not retain genetic diversity if there is a high variance in reproductive sucess (Hedgecock 1994), as is generally the case for marine fishes (Grant & Bowen 1998; 2006).

Conclusion

The introduction of *L. kasmira* to the Hawaiian Islands is a remarkable case study for two reasons. First, the introduction of this species occurred in two discrete and well documented events with known chronology, numbers, and source populations. Second, this species was introduced from two genetically distinct populations at the Marquesas and Society Islands,

allowing us to trace the fate of their descendents in the introduced range. Using mitochondrial and nuclear sequence data we determined that individuals from both source populations became established in the archipelago, interbreed, and their descendents have colonized each island and atoll surveyed (Figs 4.2, 4.3). We found that 2431 *L. kasmira* were sufficient to prevent a substantial loss of mtDNA diversity while 728 individuals resulted in a 41% decrease in haplotype richness.

Previous reports document that *L. kasmira* colonized from the inhabited Main Hawaiian Islands to the farthest northwestern (NW) Hawaiian Islands, a distance of over 2000 km in just 34 years (Oda & Parrish 1982; Randall 1987). More recently, a range-wide genetic survey of *L. kasmira* demonstrated exception dispersal ability in this species (Gaither *et al*. 2010). Here we conclude that the rapid colonization across the NW Hawaiian is accompanied by maintenance of high levels of genetic diversity, indicating large numbers of colonists at every island along the way. The NW Hawaiian Island is now one of the largest marine protected areas in the world (Papahānaumokuākea Marine National Monument), and subject to large-scale efforts to prevent and eradicate alien introductions. In these circumstances, managers need to know whether the 350+ marine exotics in the inhabited islands of SE Hawaii pose a threat to the nearly pristine habitats of the NW Hawaiian Islands. Our data indicate that highly dispersive species such as *L. kasmira* may prove to be the most effective invaders, and add a new layer to the findings of Oda & Parrish (1982) and Randall (1987); not only can exotic species jump to the NW Hawaiian Islands, they can do so in great numbers and with robust genetic diversity.

References

Agerberg A, Jansson H (1995) Allozymic comparisons between three subspecies of the freshwater crayfish *Pacifastacus leniusculus* (Dana), and between populations introduced to Sweden. *Hereditas*, **122**, 33-39.

Allendorf FW, Luikart G (2007) *Conservation and the Genetics of Populations*. Blackwell Publishing, New York.

Allendorf FW, Lundquist LL (2003) Introduction: population biology, evolution, and control of invasive species. *Conservation Biology*, **17**, 24–30.

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217-1229.

Baker AJ, Moeed A (1987) Rapid genetic differentiation and founder effect in colonizing populations of common mynas (*Acridotheres tristis*). *Evolution*, **41**, 525-538.

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

Bonnet A, Thevenon S, Maudet F, Maillard JC (2002) Efficiency of semi-automated fluorescent multiplex PCRs with 11 microsatellite markers for genetic studies of deer populations. *Animal Genetics*, **33**, 343-350.

Bossdorf O, Auge H, Lafuma L, Rogers WE, Siemann E, Prati, D (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia*, **144**, 1-11.

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

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

Cameron EK, Bayne EM, Coltman DW (2008) Genetic structure of invasive earthworms *Demdrpbaema octaedra* in the boreal forest of Alberta: insights into introduction mechanisms. *Molecular Ecology*, **17**, 1189-1197.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657-1659.

Colautti RI, Grigorovich IA, MacIsaac HJ (2006) Propagule pressure: A null model for biological invasions. *Biological Invasions*, **8**, 1023–1037.

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

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics*, **144**, 2001-2014.

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.

Denit K, Sponaugle S (2004) Growth variation, settlement, and spawning of gray snapper across a latitudinal gradient. *Transactions of the American Fisheries Society*, **133**, 1339-1355.

Dlugosch KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology*, **17**, 431-449.

Eble JA, Toonen RJ, Bowen BW (2009) Endemism and dispersal: comparative phylogeography of three surgeonfish species across the Hawaiian Archipelago. *Marine Biology*, **156**, 689–698.

Eldredge LG, Smith CM (2001) *A guidebook of introduced marine species in Hawaii*. Bishop Museum Technical Report 21. Bishop Museum, Honolulu, HI.

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.

Frankham R (2005) Invasion biology: resolving the genetic paradox in invasive species. *Heredity*, **94**, 385-385.

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

Genton BJ, Shykoff A, Giraud T (2005) High genetic diversity in French invasive populations of common ragweed, *Ambrosia artemisiifolia*, as a result of multiple sources of introduction. *Molecular Ecology*, **14**, 4275-4285.

Golani D, Azzurro E, Corsini-Foka M, Falautano M, Andaloro F, Bernardi G (2007) Genetic bottlenecks and successful biological invasions: the case of a recent Lessepsian migrant. *Biology Letters*, **3**, 541–545.

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

Grant WS, Bowen BW (2006) Living in a tilted world: climate change and geography limit speciation in Old World anchovies (*Engraulis*; Engraulidae). *Biological Journal of the Linnean Society*, **88**, 673-689.

Hassan M, Harmelin-Vivien M, Bonhomme F (2003) Lessepsian invasion without bottleneck: example of two rabbitfish species (*Siganus rivulatus* and *Siganus luridus*). *Journal of Experimental Marine Biology and Ecology*, **291**, 219-232.

Hassan M, Lemaire C, Fauvelot C, Bonhomme F (2002) Seventeen new exon-primed intron-crossing polymerase chain reaction amplifiable introns in fish. *Molecular Ecology Notes*, **2**, 334-340.

Hedgecock D (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: *Genetics and Evolution of Aquatic Organisms* (ed. Beaumont AR), pp 122-134. Chapman and Hall, London.

Kimura M, Crow JF (1964) The number of alleles that can be maintained in a finite population. *Genetics*, **49**, 725-738.

Kolar CS, Lodge DM (2001) Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution*, **16**, 199–204.

Kolbe JJ, Glor RE, Schettino LR, Lara AC, Larson A, Losos JB (2004) Genetic variation increases during biological invasion by a Cuban lizard. *Nature*, **431**, 480-483.

Koskinen MT, Haugen TO, Primmer CR (2002) Contemporary Fisherian life-history evolution in small salmonid populations. *Nature*, **419**, 826-830.

Koskinen MT, Nilsson J, Veselov AJ, Potutkin AG, Ranta E, Primmer CR (2002) Microsatellite data resolve phylogeographic patterns in European grayling, *Thymallus thymallus*, Salmonidae. *Heredity*, **88**, 391-401.

Leberg PL (2002) Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology*, **11**, 2445-2449.

Le Page SL, Livermore RA, Cooper DW, Taylor AC (2000) Genetic analysis of a documented population bottleneck: Introduced Bennett's wallabies (*Macropus rufogriseus rufogriseus*) in New Zealand. *Molecular Ecology*, **9**, 753-763.

Lee W-J, Conroy J, Howell WH, Kocher TD (1995) Structure and evolution of teleost mitochondrial control regions. *Journal of Molecular Evolution*, **41**, 54-66.

Lessios HA (1984) Possible prezygotic reproductive isolation in sea urchins separated by the Isthmus of Panama. *Evolution*, **38**, 1144-1148.

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

Lockwood JL, Cassey P, Blackburn T (2005) The role of propagule pressure in explaining species invasions. *Trends in Ecology & Evolution*, **20**, 223–228.

Maddison DR, Maddison WP (2002) *MacClade 4: analysis of phylogeny and character evolution*. Sinauer Associates, Sunderland.

Meyer A (2008) *An ecological comparison of* Cephalopholis argus *between native and introduced populations*. PhD Thesis, University of Hawaii at Manoa.

Morales-Nin B, Ralston S (1990) Age and growth of *Lutjanus kasmira* (Forskal) in Hawaiian waters. *Journal of Fish Biology*, **36**, 191-203.

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

Nason JD, Heard SB, Williams FR (2002) Host-associated genetic differentiation in the goldenrod elliptical-gall moth, *Gnorimoschema gallaesolidaginis* (Lepidoptera: Gelechiidae). *Evolution*, **56**, 1475-1488.

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

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

Oda DK, Parrish JD (1982) Ecology of commercial snappers and groupers introduced to Hawaiian reefs. *Proceedings of the Fourth International Coral Reef Symposium*, Manila, 59-67.

Palumbi SR, Metz EC (1991) Strong reproductive isolation between closely related tropical sea urchins Genus *Echinometra*. *Molecular Biology & Evolution*, **8**, 227-239.

Piry S, Luikart G, Cornuet JM (1999) Bottleneck: a computer program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*, **90**, 502-503.

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, Lecaillon G (1998) Consequences of the founder effect in the genetic structure of introduced island coral reef fish populations. *Biological Journal of the Linnean Society*, **63**, 537-552.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817-818.

Randall JE (1987) Introductions of marine fishes to the Hawaiian Islands. Bulletin of Marine Science, **47**, 356-400.

Randall JE (2001) Four new cardinalfishes (Perciformes: Apogonidae) from the Marquesas Islands. *Pacific Science*, **55**, 47-64.

Randall JE (2007) *Reef and Shore Fishes of the Hawaiian Islands*. Sea Grant College Program, University of Hawaii, Honolulu.

Randall JE, Brock VE (1960) Observation on the ecology of epinepheline and lutjanid fishes of the Society Islands, with emphasis on food habits. *Transactions of the American Fisheries Society*, **89**, 9-16.

Randall JE, Earle JL, Pyle RL, Parrish JD, Hayes T (1993) Annotated checklist of the fishes of Midway Atoll, Northwestern Hawaiian Islands. *Pacific Science*, **47**, 356–400.

Rangarajan K (1971) Maturity and spawning of the snapper, *Lutjanus kasmira* (Forskal) from the Andaman Sea. *Indian Journal of Fisheries*, **18**, 114-125.

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

Roed KH, Soldal AV, Thorisson S (1985) Transferring variability and founder effect in Iceland reindeer, *Rangifer tarandus*. *Hereditas*, **103**, 161-164.

Rogers SM, Bernatchez L (2006) The genetic basis of intrinsic and extrinsic postzygotic isolation jointly promoting speciation in the lake whitefish species complex (*Coregonus clupeafornis*). *Journal of Evolutionary Biology*, **19**, 1979-1994.

Roman J, Darling JA (2007) Paradox lost: genetic diversity and the success of aquatic invasions. *Trends in Ecology & Evolution*, **22**, 454-464.

Rosenthal DM, Ramakrishnan AP, Cruzan MB (2008) Evidence for multiple sources of invasion and intraspecific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. in North America. *Molecular Ecology*, **17**, 4657-4669.

Rubinoff RW, Rubinoff I (1971) Geographic and reproductive isolation in Atlantic and Pacific populations of Panamanian *Bathygobius*. *Evolution*, **25**, 88-97.

Schmitt T, Cizek O, Konvicka M (2005) Genetics of a butterfly relocation: large, small and introduced populations of the mountain endemic *Erebia epiphron silesiana*. *Biological Conservation*, **123**, 11-18.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin: a software for population genetics data analysis ver 2.000*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.

Schultz JK, Pyle RL, DeMartini E, Bowen BW (2007) Genetic connectivity among color morphs and Pacific archipelagos for the flame anglefish, *Centropyge loriculus*. *Marine Biology*, **151**, 167-175.

Schumacher BD, Parrish JD (2005) Spatial relationships between an introduced snapper and native goatfishes on Hawaiian reefs. *Biological Invasions*, **7**, 925-933.

Sokal RR, Rohlf JF (1995) *Biometry: The Principles and Practice of Statistics in Biological Research*. W.H. Freeman and Company, New York.

Spencer CC, Neigel JE, Leberg PL (2000) Experimental evaluation of the usefulness of microsatellite DNA for detecting demographic bottlenecks. *Molecular Ecology*, **9**, 1517-1528.

St Louis VL, Barlow JC (1988) Genetic differentiation among ancestral and introduced populations of the Eurasian tree sparrow (*Passer montanus*). *Evolution*, **42**, 266-276.

Suzuki K, Hioki S (1979) Spawning behavior, eggs, and larvae of the lutjanid fish, *Lutjanus kasmira*, in an aquarium. *Japanese Journal of Ichthyology*, **26**, 161-165.

Tamura K, Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondiral DNA in humans and chimpanzees. *Molecular Biology & Evolution*, **10**, 512-526.

Taylor CE, Gorman GC (1975) Population genetics of a colonizing lizard: natural selection of allozyme morphs in *Anolis grahami*. *Heredity*, **35**, 241-247.

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

Wares JP, Hughes, R., Grosberg, K. (2005) Mechanisms that drive evolutionary change. In: *Species Invasions: Insights Into Ecology, Evolution, and Biogeography* (ed. Sax DF, Stachowicz JJ, Gaines SD), pp 229-257, Sinauer Press, Sunderland.

Webley LS, Zenger KR, English AW, Cooper DW (2004) Low levels of genetic variation within introduced Javan rusa deer (*Cervus timorensis russa*) in Australia. *European Journal of Wildlife Research*, **50**, 137-140.

Zapata FA, Herron PA (2002) Pelagic larval duration and geographic distribution of tropical eastern Pacific snappers (Pisces: Lutjanidae). *Marine Ecology Progress Series*, **230**, 295-300.

Zayed A, Constantin ŞA, Packer L (2007) Successful biological invasion despite a severe genetic load. *PLoS ONE* 2, e868. doi:10.1371/journal.pone.0000868.

Zeisset I, Beebee TJC (2003) Population genetics of a successful invader: the marsh frog *Rana ridibunda* in Britain. *Molecular Ecology*, **12**, 639-646.

Table 4.1. Molecular diversity indices for the mitochondrial control region sequences for the two source populations of *Lutjanus kasmira* and ten populations across the introduced range. Number of specimens (N), number of haplotypes (N_h), number of singletons (N_s), haplotype diversity (h), and nucleotide diversity (π) as reported by ARLEQUIN 3.11 are listed.

	All Data					Marquesas Lineage					Society Lineage				
	N	N_h	N_s	h	π	N	N_h	N_s	h	π	N	N_h	N_s	h	π
Source Populations															
Marquesas	50	47	45	0.997 (0.005)	0.019 (0.010)	50	47	45	0.997 (0.005)	0.019 (0.010)	-	-	-	-	-
Society	49	31	23	0.970 (0.012)	0.017 (0.009)	-	-	-	-	-	49	31	23	0.970 (0.012)	0.017 (0.009)
Introduced Range															
Oahu	50	40	30	0.992 (0.006)	0.033 (0.017)	40	33	26	0.991 (0.008)	0.023 (0.012)	10	7	4	0.933 (0.062)	0.018 (0.010)
Kona	50	41	35	0.989 (0.007)	0.038 (0.019)	28	26	25	0.992 (0.013)	0.019 (0.010)	22	15	10	0.957 (0.028)	0.018 (0.009)
Hilo	51	38	30	0.985 (0.008)	0.037 (0.018)	33	28	24	0.989 (0.011)	0.020 (0.010)	18	10	6	0.915 (0.041)	0.019 (0.010)
Maui Nui	39	32	28	0.985 (0.011)	0.034 (0.017)	29	24	21	0.980 (0.017)	0.018 (0.010)	10	8	7	0.933 (0.077)	0.019 (0.011)
Kauai	36	30	25	0.989 (0.010)	0.035 (0.018)	25	23	21	0.993 (0.013)	0.020 (0.011)	11	7	4	0.909 (0.066)	0.016 (0.009)
Necker	49	38	31	0.986 (0.008)	0.035 (0.018)	34	29	25	0.989 (0.010)	0.018 (0.010)	15	9	6	0.905 (0.054)	0.016 (0.009)
Maro	21	18	16	0.981 (0.023)	0.036 (0.018)	15	15	15	1.000 (0.024)	0.019 (0.011)	6	3	1	0.733 (0.155)	0.016 (0.010)
FFS	40	38	36	0.997 (0.006)	0.033 (0.017)	31	30	29	0.998 (0.009)	0.021 (0.011)	9	8	7	0.972 (0.064)	0.013 (0.008)
Midway	40	33	28	0.989 (0.009)	0.046 (0.023)	28	26	24	0.995 (0.011)	0.029 (0.015)	12	7	4	0.894 (0.063)	0.034 (0.018)
Kure	9	9	9	1.000 (0.052)	0.034 (0.019)	7	7	7	1.000 (0.076)	0.023 (0.013)	2	2	2	1.000 (0.500)	0.008 (0.009)
All Hawaii specimens	385	172	92	0.990 (0.001)	0.037 (0.018)	270	142	80	0.993 (0.001)	0.021 (0.011)	115	30	12	0.930 (0.011)	0.019 (0.010)
All specimens	484	218	123	0.991 (0.001)	0.037 (0.018)	320	170	99	0.993 (0.001)	0.021 (0.010)	164	48	24	0.946 (0.007)	0.018 (0.009)

Table 4.2. Number of specimens (N), number of alleles (N_a), heterozygosity observed (H_O), heterozygosity expected (H_E), and the corresponding P-value for an exact test of Hardy–Weinberg equilibrium (HWE) are listed for each nuclear intron. N_a and average H_E are listed for the multi-locus data set. Abbreviation: FFS = French Frigate Shoals.

	GH Intron					ANT Intron					GnRH3-3 Intron					Multi-locus	
	N	N_a	H_O	H_E	P	N	N_a	H_O	H_E	P	N	N_a	H_O	H_E	P	N_a	H_E
Source Populations																	
Marquesas	49	3	0.66	0.59	0.57	49	6	0.29	0.37	0.05	50	2	0.16	0.18	0.39	17	0.38
Society	50	3	0.10	0.10	1.00	47	5	0.31	0.38	0.03	50	2	0.37	0.49	0.14	16	0.32
Introduced Range																	
Oahu	49	4	0.59	0.58	0.73	49	8	0.59	0.57	0.95	50	2	0.56	0.48	0.37	23	0.54
Kona	50	4	0.74	0.60	0.14	50	9	0.64	0.64	0.39	50	2	0.54	0.47	0.37	23	0.57
Hilo	52	3	0.50	0.61	0.09	47	4	0.62	0.56	0.94	51	2	0.33	0.43	0.11	24	0.54
Maui Nui	39	3	0.56	0.61	0.48	39	6	0.62	0.59	0.22	39	2	0.54	0.50	0.75	19	0.57
Kauai	35	3	0.60	0.58	0.57	36	5	0.57	0.51	1.00	36	2	0.39	0.43	0.69	19	0.50
Necker	50	3	0.60	0.59	0.75	48	7	0.58	0.59	0.31	45	2	0.49	0.49	1.00	28	0.56
Maro	20	3	0.60	0.61	0.90	20	5	0.35	0.36	0.35	21	2	0.52	0.49	1.00	14	0.49
FFS	40	3	0.50	0.56	0.43	38	6	0.61	0.60	0.98	39	2	0.51	0.50	1.00	22	0.55
Midway	39	3	0.62	0.58	0.57	40	7	0.75	0.69	0.72	40	2	0.50	0.49	1.00	22	0.58
Kure	9	3	0.22	0.54	0.03	9	3	0.89	0.66	0.12	9	2	0.33	0.50	0.49	10	0.57
All Hawaii specimens	383	4	0.58	0.59	0.68	375	12	0.61	0.59	0.85	380	2	0.48	0.48	0.83	59	0.53
All specimens	482	4	0.54	0.58	0.02	471	15	0.55	0.61	0.02	480	2	0.44	0.46	0.23	65	0.53

Table 4.3. Pairwise *F*-statistics for the two source populations of *Lutjanus kasmira* and ten populations across the introduced range. Pairwise ϕ_{ST} values for control region data are below diagonal and pairwise F_{ST} values for the multi-locus nuclear data set are above diagonal. We maintained an alpha value of 0.05 among all pairwise tests by controlling for the false discovery rate as recommended by Benjamini & Yekutieli (2001) and reviewed by Narum (2006). Values in bold are significant at the corrected $\alpha = 0.010$. Abbreviation: FFS = French Frigate Shoals.

	Source Populations		Introduced Range									
Sample Location	Marquesas	Society	Oahu	Kona	Hilo	Maui Nui	Kauai	Necker	FFS	Maro	Midway	Kure
Marquesas	-	**0.490**	**0.089**	**0.089**	**0.067**	**0.102**	**0.039**	**0.104**	**0.144**	**0.060**	**0.114**	**0.163**
Society	**0.734**	-	**0.295**	**0.248**	**0.293**	**0.303**	**0.356**	**0.266**	**0.233**	**0.434**	**0.250**	**0.184**
Oahu	**0.081**	**0.525**	-	-0.005	-0.003	-0.003	0.001	-0.004	0.002	0.009	-0.002	0.007
Kona	**0.224**	**0.339**	0.042	-	-0.009	-0.007	0.002	-0.009	-0.004	0.020	-0.007	-0.014
Hilo	**0.187**	**0.395**	0.021	0.003	-	-0.002	-0.005	-0.002	0.007	0.016	-0.004	-0.000
Maui Nui	**0.107**	**0.503**	-0.006	0.022	0.008	-	0.008	0.001	0.001	0.007	-0.006	0.010
Kauai	**0.149**	**0.466**	0.005	0.011	-0.011	0.000	-	0.005	0.025	-0.001	0.011	0.025
Necker	**0.130**	**0.455**	-0.003	0.005	0.002	-0.007	-0.006	-	0.000	0.018	-0.006	-0.014
FFS	**0.083**	**0.527**	-0.011	0.033	0.019	-0.004	0.003	-0.011	-	**0.047**	-0.008	-0.020
Maro	**0.117**	**0.532**	-0.006	0.013	0.009	-0.004	0.003	-0.019	-0.025	-	0.028	0.066
Midway	**0.125**	**0.401**	0.004	0.005	-0.005	-0.007	-0.005	-0.007	0.003	-0.006	-	-0.014
Kure	**0.119**	**0.575**	-0.037	-0.004	-0.032	-0.036	-0.054	-0.037	-0.042	-0.041	-0.041	-

Table 4.4. Results of rarefaction analyses. Number of specimens (N), number of haplotypes (N_h), and mean number of haplotypes (*H*) (± standard deviation) estimated from 10,000 random subsamples (N = number of individuals sampled in the source population) of the Hawaiian lineages are listed. The % lost is the reduction in haplotypes when comparing the source population to the corresponding Hawaiian lineage.

Population	N	N_h	*H*	% lost	*P*-value
Marquesas					
Source	50	47			
Hawaiian lineage	270	142	39.0 ± 2.50	17.0%	< 0.001
Society					
Source	49	31			
Hawaiian lineage	115	30	18.3 ± 2.06	41.0%	< 0.001

Table 4.5. Table is modified from Dlugosch & Parker (2007). Studies of molecular variation in eleven intentionally introduced species. Only cases in which all individuals were derived from a single source population and the number of individuals released is confidently known are listed. Locations indicate the regions that served as the source (S) and introduced (I) areas. Number of individuals introduced (N_I) and marker type are listed. Values for allelic richness (A) and expected heterozygosity (H_E) are averages per locus and population.

Organism	Location (S / I)	N_I	Marker	A (S/I)	H_E (S/I)	Reference
Birds						
Common Myna *Acridotheres tristis*	India / Australia	~ 250	allozymes (21)	1.43 / 1.30	0.06 / 0.06	Baker & Moeed 1987
Eurasian Tree Sparrow *Passer montanus*	Germany / United States	20	allozymes (39)	1.50 / 1.33	0.101 / 0.078	St. Louis & Barlow 1988
Reptiles						
Jamaican Anole *Anolis grahami*	Jamaica / Bermuda	71	allozymes (24)	1.75 / 1.50	0.078 / 0.064	Taylor & Gorman 1975
Mammals						
Red-necked Wallaby *Macropus rufogriseus*	Australia / New Zealand	6-10	microsatellites (5)	8.4 / 4.6	0.767 / 0.586	Le Page *et al.* 2000
Caribou Rangifer tarandus	Norway / Iceland	35	allozymes (1)	8.0 / 3.0	0.729 / 0.332	RØed *et al.* 1985
Javan Rusa Deer Cervus timorensis russa	New Caledonia / Australia	7	microsatellites (10,24)	7.60 / 2.29	0.595 / 0.467	Bonnet *et al.* 2002 Webley *et al.* 2004
Insects						
Mountain Butterfly *Erebia epiphron silesiana*	E / W Sudetans (Czechia)	50*	allozymes (17)	1.59 / 1.47	0.100 / 0.116	Schmitt *et al.* 2005
Amphibians						
Marsh Frog *Rana ridibunda*	Hungary / Britain	12	microsatellites (5)	3.2 / 2.2	0.522 / 0.484	Zeisset & Beebee 2003
Crustaceans						
Signal Crayfish *Pacifastacus leniusculus*	Canada (Pitt Lake) / Sweden	~ 200	allozymes (4)	1.50 / 1.25	0.177 / 0.079	Agerberg & Jansson 1995
Fish						
Peacock Grouper *Cephalopholis argus*	Society / Hawaii	2385**	allozymes (9)	4.00 / 3.78	0.046 / 0/045	Planes & Lecaillon 1998
European Grayling *Thymallus thymallus*	NW Europe / Norway	"a small number"	microsatellites (17)	3.75 / 1.90	0.435 / 0.170	Koskinen *et al.* 2002a, b

* 50 inseminated females were translocated. This species is known to engage in multiple inseminations and to store sperm which would render that effective population size higher than expected from the census size.

** The number of individuals released is reported in this reference as 571 (released in 1956). This number has been corrected here to include the 1,814 fish that were released in 1961 from the same source population (HDFG records).

Table 4.6. Allele frequencies at three nuclear introns in the two source populations and across the introduced range.

	Source Populations		Introduced Range									
Allele	Marq.	Soc.	Oahu	Kona	Hilo	Maui Nui	Kauai	Necker	Maro	FFS	Mid	Kure
GH												
GH1	28	95	42	46	42	27	28	50	13	39	34	11
GH2	56	4	48	43	48	40	36	40	21	37	37	6
GH3	16	0	7	10	14	11	6	10	6	4	7	1
GH4	0	1	1	1	0	0	0	0	0	0	0	0
ANT												
A1	77	0	59	53	56	44	47	56	32	38	38	8
A2	7	73	25	28	27	23	13	23	4	30	24	7
A3	10	9	3	8	8	6	8	10	2	6	10	3
A4	0	8	1	3	3	3	2	3	1	1	2	0
A5	0	3	9	3	0	1	2	2	0	1	4	0
A6	0	0	1	2	0	0	0	1	0	1	0	0
A7	2	0	0	0	0	0	0	0	0	0	0	0
A8	0	0	0	1	0	0	0	0	0	0	0	0
A9	0	0	0	0	0	0	0	0	0	0	1	0
A10	0	1	0	0	0	0	0	0	0	0	0	0
A11	1	0	0	0	0	0	0	1	0	0	0	0
A12	1	0	0	0	0	0	0	0	0	0	0	0
A13	0	0	1	1	0	0	0	0	0	1	1	0
A14	0	0	1	0	0	1	0	0	1	0	0	0
A15	0	0	0	1	0	0	0	0	0	0	0	0
GnRH												
G1	10	41	40	37	31	35	22	36	17	31	34	7
G2	90	59	60	63	71	43	50	54	25	47	46	11

Figure 4.1. Map of the Indo-Pacific. *Lutjanus kasmira* were introduced to the Hawaiian Island of Oahu from two source locations: Nuku Hiva in the Marquesas Islands and Moorea in the Society Islands. The number of fish introduced from each location is shown. Blue shading represents *L. kasmira's* natural range.

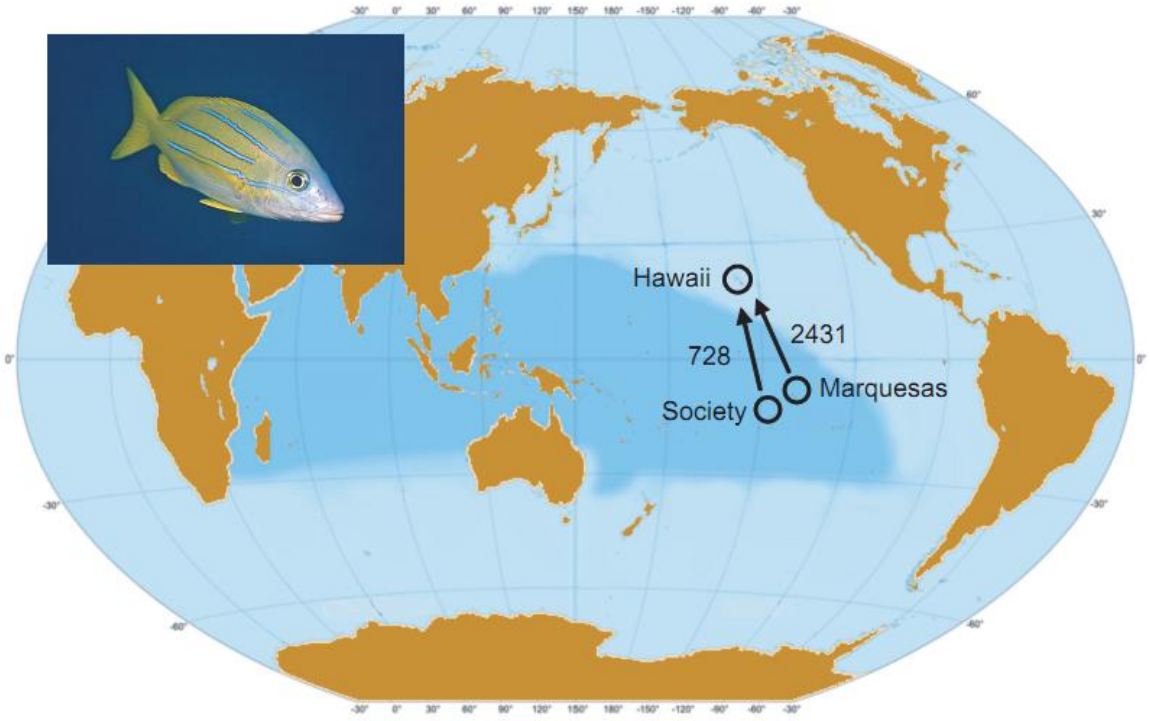


Figure 4.2. Map of the Hawaiian archipelago. Pie chart in left corner depicts the 3.4:1 introduction ratio of Marquesas fish (black) to Society fish (white). Pie charts for each sample location in Hawaii show the ratio of *Lutjanus kasmira* in either the Marquesas or Society lineage (see Fig. 4.3). Hilo and Kona are locations on opposite sides of Hawaii Island. The figure demonstrates that fish from both source populations are found at each sample location and are in roughly the same ratio as the original introduction ratio of 3.4:1. Abbreviation: FFS = French Frigate Shoals.

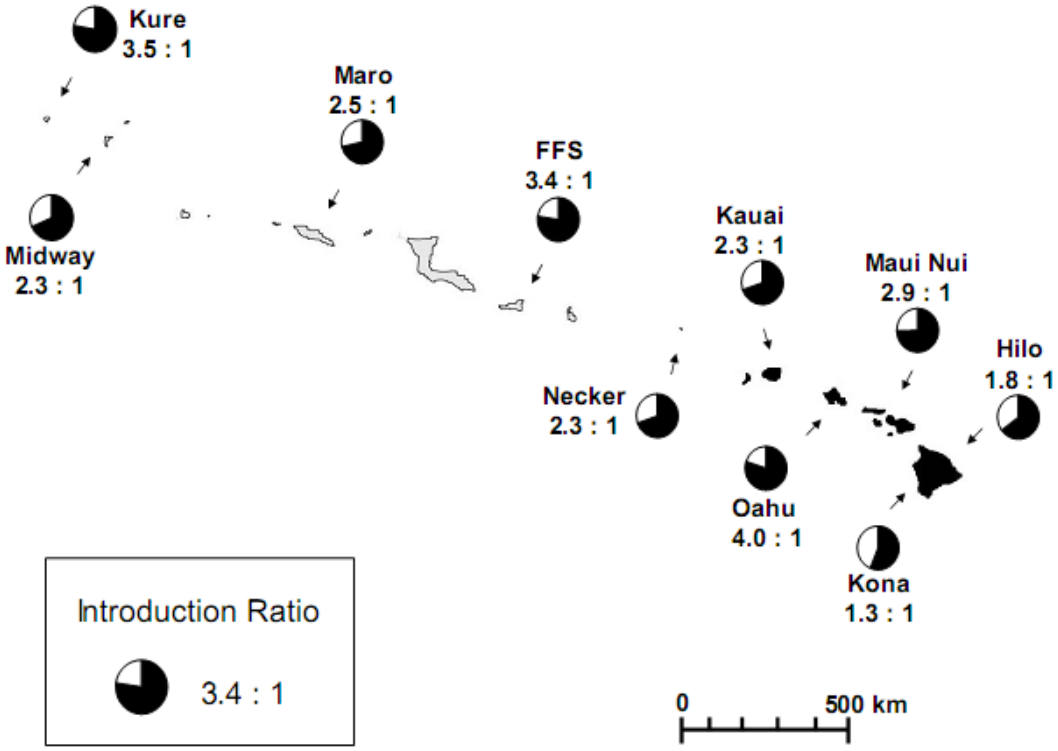


Figure 4.3. Statistical parsimony network for 484 control region sequences of *Lutjanus kasmira* constructed using TCS 2.21 (Clement *et al.* 2000). Each circle represents one mitochondrial haplotype with the area of each circle is proportional to number of that particular haplotype in the data set; dashes represent hypothetical haplotypes; colors represent collection location (see key). There were no shared haplotypes between the two source populations (Marquesas and Society) which formed two distinct lineages that are separated by 22 steps (average percent sequence divergence = 3.8% between source populations).

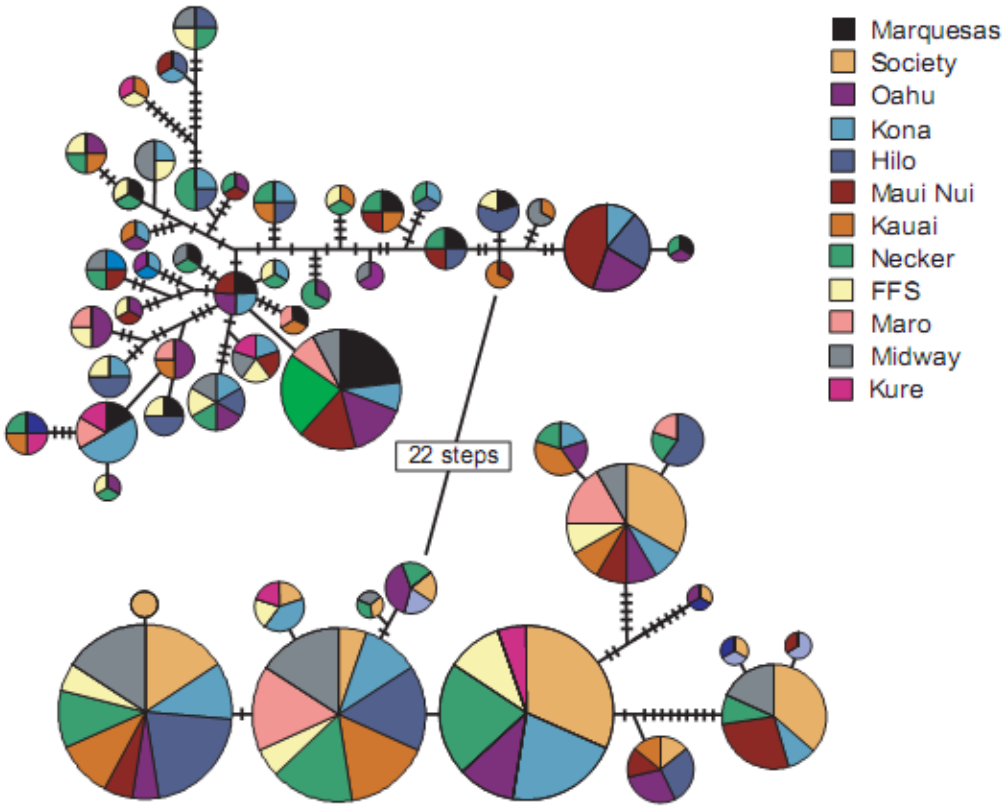
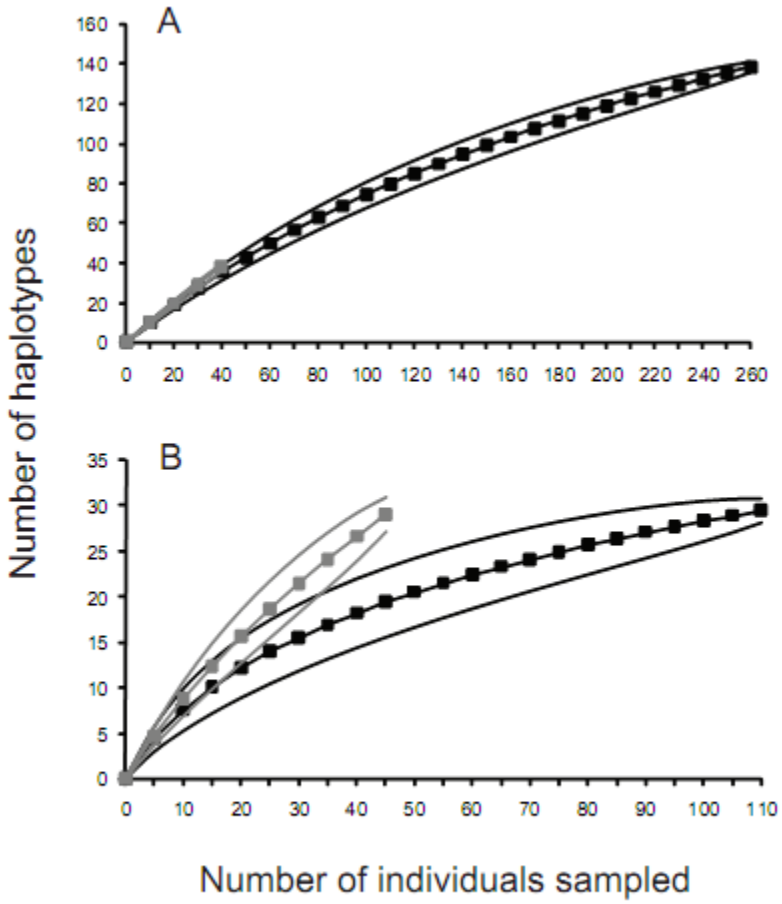


Figure 4.4. Rarefaction curves plotting the number of individuals sampled against the expected number of mitochondrial haplotypes were calculated using the Analytic Rarefactation1.4 software available at the UGA Stratigraphy Lab website (http://www.uga.edu/~strata/software/). Samples belonging to the Marquesas (A) and Society (B) lineages are plotted separately. Grey lines represent data for source populations, black lines represent data from the introduced range, and solid lines are 95% confidence intervals. The Society lineage in the introduced range is significantly different than the source population indicating a loss of rare haplotypes.



CHAPTER FIVE

Discussion

Phylogeographic surveys

The phylogeographic surveys of the snappers *Lutjanus kasmira* and *L. fulvus*, and the grouper *Cephalopholis argus* revealed contrasting patters of genetic differentiation (Fig. 5.1). While *L. kasmira* demonstrated no population structure across the Indo-Pacific Barrier (IPB), the Indian Ocean, and (most of) the Pacific Ocean, *L. fulvus* proved to be a highly structured species throughout the same range, with population structure across the IPB concordant with other pairwise comparisons (Chapter 2). In sharp contrast, the grouper *Cephalopholis argus* was strongly impacted by the IPB and Pleistocene sea level fluctuations that resulted in the partitioning of this species into Pacific and Indian Ocean mitochondrial lineages that are distinguished by fixed differences ($d = 0.008$; Chapter 3). Following the end of the last glacial maximum, connectivity between the Pacific and Indian Oceans resumed and *C. argus* populations expanded. Representatives of each mitochondrial lineage are now found in both oceans with the center of diversity occurring in the Coral Triangle.

Despite differences in population structure, both snappers demonstrate a remarkably strong phylogeographic break at the Marquesas Islands, a pattern not observed in *C. argus*. Isolation of Marquesan populations may have arisen through a biogeographical barrier to immigration (distance and contrarian ocean currents) and an unusual local environment that inhibits the survival of propagules from outside the archipelago. Coalescence analyses for *L. kasmira* prompt the hypothesis of glacial refugia in the western Indian Ocean and the Marquesas, with range expansion from eastern Africa through the Indo-central Pacific reinforcing the view that strong and enduring mechanisms have isolated the Marquesan population over evolutionary time (Chapter 2).

Lutjanus kasmira demonstrates little genetic structure across nearly 20,000 km of its natural range (Fig. 5.1). This degree of genetic similarity across such a large region indicates that

the larvae of this species are able to cross large stretches of open ocean, find suitable habitat, settle, and reproduce. In contrast, *C. argus* demonstrated little population structure within the Pacific or Indian Oceans but high levels of structure between oceans, indicating that there is restricted dispersal of this species between ocean basins. *L. fulvus* showed significant population structure in twelve of the fifteen possible pairwise comparisons between populations indicating that this species is successfully dispersing over only short geographic distances (Fig. 5.1).

The introduction of Lutjanus kasmira *to Hawaii*

Following introduction to the Island of O‘ahu, *L. kasmira* spread throughout the archipelago to the farthest Northwestern Hawaiian Island (NWHI) at Kure Atoll, a distance of over 2400 km in just 34 years (Fig. 5.2, Oda & Parrish 1982, Randall 1987). A range-wide genetic survey of *L. kasmira* demonstrated exceptional dispersal ability in this species (Chapter 2). My findings indicate that the rapid colonization across the Hawaiian Archipelago was accompanied by maintenance of high levels of genetic diversity, indicating large numbers of colonists at every island along the way (Chapter 4). The NWHI are now one of the largest marine protected areas in the world, the Papahānaumokuākea Marine National Monument (PMNM), and subject to large-scale efforts to prevent and eradicate alien introductions. Under these circumstances, managers need to know whether the 350+ marine exotics in the inhabited islands of Hawai‘i pose a threat to the nearly pristine habitats of the PMNM. My data indicate that highly dispersive species such as *L. kasmira* may prove to be the most effective invaders, and add a new layer to the findings of Oda and Parrish (1982) and Randall (1987); not only can exotic species jump to the NWHI, they can do so in great numbers and with robust genetic diversity (Chapter 4).

Predicting probable invaders

Predicting which species will be successful invaders is of immense scientific and practical interest. For my dissertation, I tested the hypothesis that species with the greatest dispersal ability in their native range will prove to be the most pervasive invaders. Comparing the differential success of the three alien species in Hawai‘i, as indicated by the extent to which they have spread in archipelago (Fig. 5.2), with the phylogeographic surveys, I found a strong correlation between genetic differentiation across the natural geographic range and invasion success (Figs. 5.1, 5.2). *L. kasmira*, the most successful invader in Hawai‘i, demonstrates little genetic structure across nearly 20,000 km of its natural range. In contrast, *L. fulvus*, the least widespread of the three species in Hawai‘i, demonstrated high levels of population structure over even small geographic distances. *C. argus* demonstrated an intermediate pattern.

I believe the observed pattern is likely due to differences in life history traits. All three species occupy roughly the same geographic range (Fig. 5.1) and have pelagic larval phases, during which long-distance movement is accomplished (Linderman et al. 2000; Denit & Sponaugle 2004), of similar duration. The interesting differences between these species are in juvenile habitat preferences and territoriality behavior of the adults. The juveniles of *L. kasmira* have broad habitat preferences, settling on virtually any hard substrata, whereas evidence indicates that *L. fulvus* juveniles settle in mangroves and move to coral reefs as subadults (Nakamura et al. 2008). Interestingly, mangroves are not native to Hawai‘i but several introduced species flourish in the Main Hawaiian Islands. Similar to *L. fulvus*, mangroves have not colonized further northwest than Kauai (Fig. 5.1). *C. argus* has the most complex life history of the three species. While the snappers are mass spawners, *C. argus* is a sequential hermaphrodite beginning life as a female. Males hold territories and guard harems of females.

Complex social behavior is likely to impart constraints on this species that we are working to better understand.

Significance

Though my research is focused on Hawaiian invasive species, the threats are global in scope. Issues of invasive species and their affects on naive ecosystems are an ever increasing threat in our modern world of trans-oceanic shipping and mass movement of people and cargo. Despite the lessons of the past, there are still discussions today about the utility of intentional introductions. The spread of alien species is of great concern to resource managers and conservationists concerned about ecosystem integrity. It is imperative that the spread of alien species be documented and the methods by which they disperse to new areas be determined. These Hawaiian introductions present a unique opportunity to combine historical and ecological data with advanced molecular techniques to predict which species are probably invaders and the possible consequences of new introductions. Understanding how invasive species spread and what affects they have at the ecosystem level will allow more informed management of these altered systems as well as predict the consequences of future invasions.

REFERENCES

Denit K, Sponaugle S (2004) Growth variation, settlement, and spawning of gray snappers across a latitudinal gradient. *Transactions of the American Fisheries Society*, **133**, 1339-1355.

Linderman KC, Lee TN, Wilson WD, Claro R, Ault JS (2000) Transport of larvae originating in southwest Cuba and the Dry Tortugas: evidence for partial retention in grunts and snappers. *Proceedings of the Gulf of and Caribbean Fisheries Institute*, **52**, 253-278.

Nakamura Y, Horinouchi M, Shibuno T, Tanaka Y, Miyajima T, Koike, et al. (2008) Evidence of ontogenetic migration from mangroves to coral reefs by black-tail snapper *Lutjanus fulvus*: stable isotope approach. *Marine Ecology Progress Series*, **355,** 257-266.

Oda DK, Parrish JD (1982) Ecology of commercial snappers and groupers introduced to Hawaiian reefs. *Proceedings of the Fourth International Coral Reef Symposium*, Manila, 59-67.

Randall JE (1987) Introductions of marine fishes to the Hawaiian Islands. *Bulletin of Marine Science*, **47**, 356-400.

Figure 5.1. Results of range wide genetic surveys for each of the three species introduced to Hawaii. The three species have similar geographic ranges (indicated with dark blue shading on maps). Circles on map encompass the largest regions in which no population differentiation was detected. Level of genetic structure is described and the extent of spread within the introduced range is given for each species.

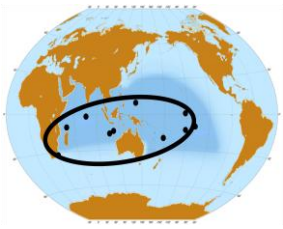
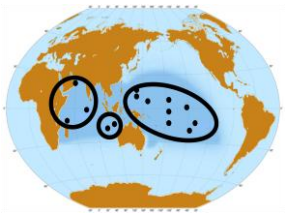
	Range wide genetic structure	Level of genetic structure	Distance of spread in Hawaii
Lutjanus kasmira		Low; only very low levels of genetic structure across 20,000 km	2,500 km; the entire archipelago
Cephalopholis argus		Moderate; no genetic structure in the Pacific or Western Indian Oceans but high levels of structure between Oceans.	1,200 km; to French Frigate Shoals
Lutjanus fulvus		High; high levels of genetic structure at all geographic scales	600 km; restricted to Main Hawaiian Islands

Figure 5.2. Map of the Hawaiian Archipelago. Shaded areas represent extent of range expansion of each introduced species.