Conservation Biology in Benthic Reef Fishes: Population Genomics of *Opistognathus aurifrons*

by

Diana Marcela Beltrán Rodríguez

A dissertation submitted in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY
in
MARINE SCIENCES
(Biological Oceanography)

Approved by:

_____	_____
Richard S. Appeldoorn, PhD President, Graduate Committee	Date
_____	_____
Nikolaos V. Schizas, PhD Member, Graduate Committee	Date
_____	_____
Manuel Valdés Pizzini, PhD Member, Graduate Committee	Date
_____	_____
Jorge R. Garcia Sais, PhD Member, Graduate Committee	Date
_____	_____
Ernesto Otero, PhD Chairperson of the Department	Date
_____	_____
Ivette Cruzado, PhD Representative of Graduate Studies	Date

Abstract

The world oceans are deteriorating at a fast pace with obvious effects on coral reefs, in which at least 50% of coral cover has disappeared. Conservation actions, such Marine Protected Areas are being implemented to alleviate marine ecosystems from stressors and allow populations to restore to healthy levels. Successful networks of MPAs can operate if the space among MPAs is smaller than the dispersal capacity of the species under protection. It is of most importance to estimate the scale of such connectivity among marine populations.

Similarly, populations in terrestrial species are often segregated across geography, usually forming independent evolutionary lineages. The addition of such information into conservation in the form of Evolutionary Significant Units (ESU) has helped maintain terrestrial species from extinction and restore natural populations. A similar approach is lacking on coral reef species despite the presence of clearly independently evolving populations. Here we present an in depth analysis of the molecular ecology of the common reef fish *Opistognathus aurifrons* to aid conservation on coral reefs by providing the scale of connectivity measured as the effective dispersal and the delineation of ESU segregated across the Caribbean.

We first designed twelve microsatellite markers with enough power to detect variations in allele frequencies along short geographical distances such across the Mona Passage. Using these markers across 260 individuals, we estimated that the maximum effective dispersal is 10 km. Our detailed data suggest that the distance among MPAs in the network around Puerto Rico is larger than the effective movement of the fish. We

found that MPAs exchange migrants likely via intermediate available unprotected habitat. At scales > 100 km connectivity among MPAs decrease, particularly across the Mona Passage, making Mona Island a genetic mosaic of the genetic variation from the eastern and western Caribbean. Driven by the result of small dispersal across fine scales, we tested whether the species is composed of smaller evolutionary significant units spread across the Caribbean. To increase our statistical power from that of microsatellites, we also genotyped individuals across eleven populations for over 18,000 single nucleotide polymorphisms. Our genomic data suggest a large degree of genetic segregation across the Caribbean with all major islands having unique genetic combinations, to the extreme of a monophyletic group in Florida. Our genome-wide analysis indicates that Belize, Florida, Curaçao and Puerto Rico form isolated groups, suggesting there are at least four Evolutionary Significant Units. Estimates of gene flow between pairwise comparisons also suggest that exchange among populations is limited (< 0.1%). Our findings stress the importance to generate a denser MPA network that ensures connectivity within a radius of at least ten kilometers. It also suggests isolated populations do not exchange gametes and have their own independent demographic dynamics, which should also be incorporated into management plans. Reef fish richness is driven by small cryptic benthic species such the yellowhead jawfish, adjusting MPAs to account for the short dispersal in these fishes and incorporating Evolutionary Significant Units into marine conservation, may enhance the maintenance of these vast reef biodiversity.

Resumen

Los océanos se están deteriorando a un ritmo acelerado, con efectos evidentes en los arrecifes de coral, al menos el 50% de la cobertura de coral ha desaparecido. Acciones de conservación, tales como las Áreas Marinas Protegidas (AMP) están siendo implementadas para aliviar a los ecosistemas marinos de los factores de estrés y permitir que las poblaciones puedan restaurarse a niveles saludables. Redes exitosas de AMPs pueden funcionar sí el espacio entre las AMP 's es menor que la capacidad de dispersión de las especies bajo protección. Es de vital importancia estimar la magnitud de dicha conectividad entre las poblaciones marinas.

Del mismo modo, las poblaciones de especies terrestres usualmente son separadas a lo largo la geografía y a menudo están formando linajes evolutivos independientes. La incorporación de dicha información en la conservación, en forma de Unidades significativas Evolutivas (ESU) han ayudado a salvaguardar las especies terrestres de la extinción y restaurar las poblaciones naturales. Un enfoque similar carecen las especies de los arrecifes de coral, a pesar de la presencia de poblaciones que claramente son evolucionariamente independientes. Aquí presentamos un análisis profundo de la ecología molecular del pez arrecifal *Opistognathus aurifrons* con el fin, de aportar a la conservación de los arrecifes de coral, proporcionando una escala de la medida de conectividad como la dispersión efectiva y la delineación de ESUs segregados en todo el Caribe.

Primero diseñamos doce marcadores de microsatélites con el poder suficiente para detectar variaciones en las frecuencias alélicas a lo largo de cortas distancias geográficas, tales como a través del Canal de la Mona. Usando estos marcadores a

través de 260 individuos de *O. aurifrons*, se estimó que la dispersión efectiva máxima es de 10 km. Nuestros datos detallados sugieren que la distancia entre las AMPs en la red alrededor de Puerto Rico es mayor que el movimiento efectivo de estos peces. Encontramos que las AMPs intercambian migrantes probablemente a través de hábitat intermedio disponible no protegido. A escalas > 100 kilometros la conectividad entre las AMP's disminuye, particularmente a través del Canal de la Mona, Haciendo de la Isla de Mona un mosaico de variación genética entre el este y el oeste del Caribe. Promovidos por el resultado de la pequeña dispersión de la yellowhead Jawfish a través de escalas geográficas finas, hemos probado si la especie se compone de pequeñas Unidades Significativas Evolutivas repartidas por todo el Caribe.
Para aumentar el poder estadístico que nos brindaron los microsatélites, también genotipamos individuos a través de once poblaciones en el Caribe, con más de 18.000 polimorfismos de nucleótidos simples (SNPs, por sus siglas en inglés). Nuestros datos genómicos sugieren un alto grado de segregación genética en todo el Caribe, mostrando en las principales islas que poseen unas combinaciones genéticas únicas, hasta el extremo de un grupo monofilético en la Florida. Nuestros datos genómicos indican que Belice, Florida, Curazao y Puerto Rico forman grupos aislados, lo que sugiere es, que hay al menos cuatro Unidades Significativas Evolutivas en la zona. Concordantemente, nuestras estimaciones de flujo génico entre las comparaciones por pares de las poblaciones. Nuestros resultados destacan la importancia de crear una red densa de AMPs que garantiza la conectividad dentro de un radio de al menos 10 kilómetros. También sugiere que las poblaciones aisladas no intercambian gametos y tienen sus propias dinámicas demográficas independientes, lo cual se deben incorporar

en los planes de manejo de las zonas protegidas. La riqueza de los peces del arrecife, es impulsada por pequeñas especies bentónicas crípticas como el yellohead jawfish, debe ser necesario el ajuste de las AMP’s, para tener en cuenta las especies peces de corta dispersión y la incorporación de Unidades Significativas Evolutivas en la conservación marina. Estos cambios pueden mejorar, asegurar y preservar la vasta biodiversidad

Copyrigth

Signed: Diana Beltran R

Date: December 10 /2015

Acknowledgements

My dissertation was possible by the generous support of various individuals and the stimulating and joyful atmosphere at the department of Marine Sciences of the University of Puerto Rico. I would like to thank each of these individuals for trusting on my project and the kind support at every step to successfully complete my Ph. D.

I like to first thank my advisor Richard Appeldoorn, whose unconditional support during the years, his patience and his wonderful guidance and resources increased the quality of my research and made me a more independent scientist. Manuel Valdés Pizzini was the first to open the door to studying conservation biology during my years at the CIEL. His knowledge on MPAs intellectually influenced me to initiate a dissertation related to marine reserves. I would also like to thank him for his substantial financial support through the school of Art and Sciences of the University of Puerto Rico, Mayagüez.

Approaching conservation questions with genetics had the imprint of Nikolaos Schizas, whose friendship and initial training in data analysis allowed me to use state of the art technologies to increase the power of my inferences about effective dispersal and visualize their applicability in marine management. I was also fortunate to meet Juan González Lagoa, a wonderful biologist, whose unconditional support was essential in times of difficulty and who has been my family in this beautiful island. I was fortunate to have his knowledge transferred to me during hundreds of informal talks about larval dispersal and marine biology.

.

To Carlos Prada infinite thanks for being the best adviser, my life partner, to guide me through this process and for his steady support over the years. I thank him for his patience to teach me how to best apply genetics to conservation and his vision to develop genomic tools at the right time, increasing the quality of my work. He was essential to the success of my dissertation and without whom this dissertation simply would have not been possible.

I would also like to thank the many colleagues who assisted me during long field collections, many of which were away from Puerto Rico. I thank Anibal Santiago, Duane Sanabria, Milton Carlo, Hector Ruiz and Martha Prada. I also thank Adeljean Ho who provided preserved material from Florida and The Bahamas. To all the workers of Marine Sciences Department all my gratitude, they made my life easier by helping me with all the university processes. I thank Negrito, Lalo, Zulma, Linnetthe, Jossie, Lilliveth. I thank Godoberto Lopez (RIP) my unforgettable diving body, whose happiness surrounded me. I learn from him how to be the best in your job even when it is a simple one. I keep his smile deep in my heart.

I would also like to thank all my CIMA fellows, those that already left and those who have just started. I thank you for making Magueyes Island, the classes and the laboratories such a pleasant place to learn about marine biology

Last but not least I thank my family. My mom Rosa María, my dad Edgard, my sisters: Claudia Catalina and Paola and brother Edgard Iván, who despite the distance have always been present supporting me. Thanks for teaching me the significance of

reaching my goals without compromising my values and be a better person every day. Thanks to their great efforts, I have achieved this goal.

Dedication

To

Rosa María y Edgard

Table of Contents

Table of Contents (Cont.)

Figure List

Table List

APPENDIX LIST

Chapter 1.

Introduction

The oceans have degraded in the last decades as a result of human activities (Mora 2008). This decline is critical in coastal areas such coral reefs, where coral cover has decreased more than 50 % worldwide (De'ath *et al.* 2012; Garner *et al.* 2003; Jackson *et al.* 2014). Conservation efforts are urgently needed to reduce such loss, recover depleted populations and restore natural habitats. Marine protected areas (MPAs) have been proposed to preserve reef fish biodiversity. When properly implemented and managed, MPAs can ameliorate, restore and achieve the long-term conservation of marine populations, promoting the recruitment success of overexploited species by protecting and increasing their population spawning potential (Roberts & Polunin 1991).

Determining the scale of connectivity among marine populations is critical to the design of successful networks of MPAs (Roberts 1997). Larval dispersal determines the degree of connectivity among marine populations, providing information on the ideal reserve size to achieve self-recruitment and the minimum spacing among reserves to maintain connectivity and diversity (Sale *et al.* 2005). Equally important to generate successful conservation strategies is to define the basic conservation units, such as species or independently evolving lineages. Such knowledge is required to properly estimate population sizes and to properly quantify species ranges and therefore measure

whether species are at any significant risk of extinction. In the following work, we present an analysis of the genetic ecology of the common benthic yellowhead jawfish (*Ophistognathus aurifrons*) to define the conservation units across a network of Caribbean MPAs, infer the population size of each unit and estimate the effective dispersal.

Dispersal, Marine Connectivity and Genetic Differentiation

Reproduction in marine species usually results in larvae that disperse in the plankton until development is completed and suitable habitat is found. If during this journey a significant number of larvae are recruited to another population, the receiving population is considered open, with immigration substantially contributing to its demographic dynamics (Sale *et al.* 2005). If instead populations are locally maintained by recruiting their own larvae, the populations are considered close. It is thus hypothesized that species that have longer planktonic larval durations a have higher probability of connecting populations than species with no or very short times as planktonic larvae. Understanding whether populations behave as open or closed and the geographic scale at which open populations are connected is central to adequately designing Marine Protected Areas (Jones *et al.* 2009). Routine mark-recapture studies are however challenging for marine populations as they often produce millions of planktonic larvae that quickly dilute in the sea and most of which die (Thorrold *et al.* 2006). Despite these complications some direct and indirect methods are available to measure the scale of connectivity in marine species.

Direct otolith chemical marking (*e.g*., oxytetracycline) and isotopic exposure to gravid females are powerful ways to track the origin of larvae and thus measure the scale of connectivity (Almany *et al.* 2007; Jones *et al.* 2007; Thorrold *et al.* 2006). Direct methods however often require collecting a high number of adults to inject isotopes, which is virtually impossible and may induce severe damage to the females. Also, marking a high number of individuals from an area to infer sources of recruitment is challenging, and searching for and processing recruits is expensive and labor intensive.

Alternatively, indirect measures provide a powerful way to infer connectivity among populations (Ayre & Hughes 2000; Baums *et al.* 2005; Underwood *et al.* 2007). Connected populations often share similar allele frequencies given their continues swap of genetic material, but if genetic exchange is restricted, or if migrants are unable to establish and reproduce in the new population, differences become evident and differentiation among populations develop at the genetic level (Hellberg 2009). Similarly, genetic data, such as that contained in hyper variable markers (*e.g.,* microsatellites) can even be used to carry out parentage analysis to estimate self-recruitment and evaluate the finest (<10 Km^2) scale of fish movement (Jones *et al.* 2009).

Not only biological but also physical models provide information on the scale of connectivity in marine systems (Cowen *et al.* 2006; Galindo *et al.* 2010; Paris *et al.* 2005). In fact, earlier physical models of passive particles suggested the idea of open marine populations over thousands of kilometers (Heck & McCoy 1978), and early genetic analyses have echoed the notion of connectivity over long distances (Grosberg

& Cunningham 2001; Lessios *et al.* 1998; Lessios & Robertson 2006; McFadden *et al.* 1997), an observation that coincides with the lack of obvious barriers in the sea and the long-range dispersal of marine species via planktonic larvae (Roberts 1997).

In the Caribbean for example, some progress has been made to measure connectivity. Shulman and Bermingham (Shulman & Bermingham 1995) initially proposed large (>500 km) dispersal and gene flow across the Caribbean for reef fishes with pelagic and non-pelagic eggs. In the French grunt (*Haemulon flavolineatum*) and the bluehead wrasse (*Thalassoma bifasciatum*) populations seem also connected over scales larger than 500 km (Purcell *et al.* 2006). Similarly, marine invertebrates such the elkhorn coral, *Acropora palmata,* show long distance dispersal. In this case only two populations seem to exist across the Caribbean with an admixed population in Puerto Rico and a possible barrier to gene flow around the Mona Passage.

More recently however, studies have challenged the idea of connectivity among marine populations over large scales (> 100 km). Over the past decade chemical tagging studies, detailed modeling of currents and larval survival have all reinforced the view that successful dispersers may travel far less than their apparent potential (Cowen & Sponaugle 2009; Hellberg 2007, 2009; Prada & Hellberg 2013; Swearer *et al.* 2002; Taylor & Hellberg 2003; Taylor & Hellberg 2006; Underwood *et al.* 2007).

Differentiation at smaller spatial scales is reported for many species, including several fish. Taylor and Hellberg (Taylor & Hellberg 2006), working with mDNA and nuclear

DNA, found evidence of strong differentiation of a cleaner goby, *Elacantinus evelynae*, around the Caribbean, suggesting thousands of generations of isolation despite modest geographical separation and larvae with a three week pelagic duration. Similarly, *Stegastes partitus* in Mesoamerican reefs shows a large degree of segregation (Salas *et al.* 2010). Even broadcast spawning corals such *A. cervicornis* differentiate at the scales of only 10's to 100's of km; thus populations around Caribbean islands are highly dependent on self-recruitment (Vollmer & Palumbi 2007). Patterns of segregation at small scales are echoed by other sessile taxa such the octocoral *Pseudopterogorgia elisabethae* (Gutiérrez-Rodríguez & Lasker 2004). At the extreme, fish populations separated by as little as 23 km have shown genetic isolation (Taylor & Hellberg 2003), and pervasive yet hidden barriers such as the Mona Passage (Baums *et al.* 2005; Taylor & Hellberg 2003; Taylor & Hellberg 2006) and Exuma Sound (Taylor & Hellberg 2006) have been uncovered. We are also just beginning to quantify the role of ecological barriers to marine connectivity. Recent studies show it is strong enough to even generate species separated by just one kilometer (Prada & Hellberg 2014; Serrano *et al.* 2014) . Complex coastal oceanographic processes, behavioral larval interactions and selection pressures in the habitat of recipient populations can result in this pattern of connectivity at small spatial scales (Cowen & Sponaugle 2009; Swearer *et al.* 2002; Warner 1997). In the following three chapters we present detailed genomic data on the scale of connectivity at large (> 500 km) and small (<100 km) scales in the common yellowhead jawfish.

Overview of Chapters

We first developed hypervariable molecular markers informative at ecological scales with sufficient power to unveil patterns of differentiation along short (< 100 km) geographical distances such across the Mona Passage (Chapter 2). We developed twelve microsatellite markers from an initial test of over fifty primer pairs using high throughput 454 pyrosequencing. In addition, our genomic data generated over tens of thousands of potential microsatellite markers that would contribute to enhance the available molecular tools to study this species. Resources that we made publically available for the coral reef fish community. Our markers amplified well across populations and were highly polymorphic within and among populations. Haplotypic diversity ranged from 0.6 to 0.9 with no significant deviations from Hardy-Weinberg proportions. We used these markers to understand genetic variation in the next two chapters.

To understand the scale of connectivity in benthic reef fish, in Chapter 3 we sampled populations across MPAs of the common and egg-brooding yellowhead jawfish (*O. aurifrons*) and evaluated genetic variation along a geographical continuous area. By screening over 200 individuals at eighteen molecular markers, we were able to determine that the effective dispersal is at most 10 km. Our pattern of genetic segregation suggests that genetic exchange occurs possibly through intermediate unprotected reefs via a stepping stone model. Given the low dispersal potential of the species, we found strong genetic breaks around the Mona Passage, suggesting Mona is an admixed population with genetic material from both Western and Eastern

Caribbean populations. Isolation across the Mona Passage, however, seems recent, as past reconstructions suggest migration occurred in both directions. Our comprehensive data along this fine scale suggest that the distance among MPAs in the network around Puerto Rico is larger than the effective movement of the fish.

Motivated by this small dispersal across fine scales, we studied the possibility of multiple Evolutionary Significant Units for conservation across the Caribbean (Chapter 4). We initially sampled over 260 individuals for twelve microsatellites and two mitochondrial loci. To increase our statistical power to measure differentiation among populations, we also genotyped a sub-sample of 95 individuals across 11 populations for over 18,000 single nucleotide polymorphisms. Our genomic data suggest a large degree of genetic segregation across the Caribbean, with all major islands having unique genetic combinations, to the extreme of a monophyletic group in Florida. Concordantly our estimates of gene flow from pairwise comparisons suggest that exchange among populations at large scales is marginal (> 0.1%) and self-recruitment large. Our genomic data indicate that Belize, Florida, Curaçao and Puerto Rico form isolated groups, suggesting there are at least four Evolutionary Significant Units.

In summary, our study shows that for this cryptic reef fish, effective dispersal is small and population maintenance comes from self-recruitment. Such little dispersal has over evolutionary scales isolated populations that do not exchange gametes. These isolated populations also have their own independent demographic dynamics warranting inclusion in management plans. Cryptic fish species like the yellowhead jawfish

comprise most of the reef fish biodiversity; adjusting MPA network designs to account for the short dispersal in these fishes and incorporating Evolutionary Significant Units into marine conservation may enhance the maintenance of reef biodiversity.

Chapter 2

Isolaton and Characterization of Twelve Microsatellite Loci to Study Connectivity in the Yellowhead Jawfish *Opistoganthus aurifrons*

Conservation Genet Resour
DOI 10.1007/s12686-014-0419-x

MICROSATELLITE LETTERS

Isolation and characterization of 12 microsatellite loci to study connectivity in the yellow jawfish *Opistognathus aurifrons*

Diana Beltran · Nikolaos V. Schizas · Carlos Prada · Richard Appeldoorn

Received: 24 December 2014 / Accepted: 29 December 2014

Abstract Marine Protected Areas (MPAs) are conservation strategies to preserve the degradation of marine ecosystems by allowing species to naturally recover. Central to MPA design is the assumption of connectivity in marine populations over hundreds of kilometers, but only in a few handful of species the scale of connectivity has been estimated. To facilitate the study of connectivity of reef fishes, we newly developed 12 microsatellite loci for the yellow jawfish *Opistognathus aurifrons*. We tested all microsatellite loci in eight Caribbean populations with various degrees of divergence. We found between 9 and 26 alleles per locus with polymorphism that ranged from 0.652 to 0.976. All loci were in Hardy–Weinberg equilibrium, except loci *1588* and *7983*. The described markers provide the most sensitive tools yet available to study connectivity at the finest spatial scale and evaluate if current networks of Caribbean MPAs maximize the potential for the recovery of reef fish populations.

Keywords Pyrosequencing · Marine Protected Areas · Coral reef · Conservation · Caribbean

D. Beltran (✉) · N. V. Schizas · R. Appeldoorn
Department of Marine Sciences, University of Puerto Rico, Mayagüez, PR 00667, USA
e-mail: diamarbeltran@gmail.com

N. V. Schizas
e-mail: nschizas@gmail.com

R. Appeldoorn
e-mail: richard.appeldoorn@upr.edu

C. Prada
Department of Biology, The Pennsylvania State University, University Park, PA 16802-5301, USA
e-mail: cprada1@psu.edu

Tropical marine ecosystems have degraded in the last decades as a result of human activities. This trend is marked in coastal habitats such as coral reefs, where populations of reef-dwelling organisms have collapsed and species commonly seen in the 1970s are rarely observed today. To mitigate such degradation, Marine Protected Areas (MPAs) have been implemented, and when properly designed can achieve the long-term conservation of marine populations. To generate networks of successful MPAs, populations must be biologically connected to the extent that the realized dispersal of managed species largely offsets the geographical distance between any two adjacent MPAs. As such, larval dispersal determines the degree of connectivity among marine populations, providing information on the ideal reserve size to achieve self-recruitment and the minimum spacing among reserves to maintain connectivity (Mora et al. 2006).

Traditionally, use of MPAs has been based on the premise that marine populations are open, with populations connected over hundreds of kilometers given the long larval planktonic duration (>30 days) of many marine species. However, recent studies suggest that for some reef species, connectivity is more common at the scale of tens of kilometers and populations segregated by habitats may even surprisingly have connectivity reduced to within hundreds of meters (Prada and Hellberg 2014). In light of these new studies, understanding the geographical scale at which marine populations are connected is critical to adequately design successful MPAs. Population genetic studies provide a powerful indirect approach to measure connectivity among populations (Jones et al. 1999). To adequately quantify the scale of connectivity across networks of MPAs in the Caribbean, we developed and tested 12 microsatellite loci for the common reef dweller, the yellow jawfish, *Opistognathus aurifrons*.

Published online: 09 January 2015

Table 1 Characterization of 12 microsatellite loci in 42 *Opistognathus aurifrons* individuals

Name	Code	Repeat motif	Primer sequence (5′–3′)	size (bp)	Ta (°C)	(N_A)	(H_O)	(H_E)	*P*HW
HLUJHHW01BHFL5	785	AGC	F: CAGAGGCTTGCCTTGAAGTG R: CGTTCGCTGCAGGTCATAC	186	59 and 53	20	0.905	0.934	0.313
HLUJHHW01CE13G	1588	AGC	F: GAAAGAGAAGTTGCCGCCTC R: CTCCCATTCCTGACTCACCC	264	59 and 53	14	0.929	0.776	0.004*
HLUJHHW01BM47D	1438	AAAG	F: CTTAGTTGGGATTGCGTGGC R: GGCCTCAGGAATTTCATCGC	224	59 and 53	14	0.929	0.884	0.254
HLUJHHW01CDWXY	8462	CCG	F: TGTTGTTGTTTCTCCGGCAC R: CTATTCTTTGGACACCGCGG	248	59 and 53	12	0.738	0.863	0.02
HLUJHHW01CP5G8	325	AAT	F: TTGGGTTGCAACTCTGTGTG R: TGCCATCTGTGTCCATTGTG	203	59 and 53	23	0.952	0.932	0.441
HLUJHHW01DLIPU	170	AGAT	F: TGACATCCACCACTGACAGG R: TATCGGCTGGTCCTTTCTGC	260	59 and 53	26	0.976	0.949	0.385
HLUJHHW01EM6DJ	516	AAGAT	F: TCTCCTCAGCCACCAAGAAG R: GGGTATCAGCACTGTTGTCC	150	59 and 53	14	0.967	0.887	0.005
HLUJHHW01CL23R	150	ACTG	F: GCGGCACACCTCTATTAAGC R: TCACGCAAACAGATGATAGCC	296	59 and 53	14	0.976	0.897	0.052
HLUJHHW01DN5ZM	246	AAG	F: GCAGCACGATCGAGAAACTG R: CAGATGGCCTCGTCAAACAC	289	56 and 53	12	0.976	0.889	0.048
HLUJHHW01EZ6IS	8537	CCG	F: CTGAAACTTCCCAACCAGCC R: CCTCGATGCTGCTTGATGTG	287	59 and 53	24	0.857	0.927	0.069
HLUJHHW01D9S09	1358	ACT	F: ACTCGACCCATGTTTCATCATC R: ACATCCACAGTTGTCACTTGC	247	59 and 53	25	0.976	0.945	0.335
HLUJHHW01C84TE	7983	CCG	F: CGGTATAGTGTGGGAGGGTC R: AAACTGGGATTGATGCGTGG	295	59 and 53	9	0.652	0.891	0.001*

Ta annealing temperature (°C), N_A number of alleles, H_O observed heterozygosity, H_E expected heterozygosity

* Deviation from Hardy–Weinberg equilibrium ($P > 0.0045$) after Bonferroni's correction, *P*HW: Hardy–Weinberg probability test

To isolate the microsatellite loci, we sampled 42 individuals (seven each from MPAs in Curaçao, Puerto Rico, Mona Island and Dominican Republic). To expedite microsatellite isolation, we pyrosequenced one diploid individual from Puerto Rico at the Duke Genome Facility. Briefly, after DNA extraction, we sonicated gDNA and prepared libraries following standard 454 chemistry protocols and quality control steps. We sequenced the library at a depth of 0.2× given a genome size of 1 GB for *O. aurifrons*. To find microsat-repeats with regions to develop primers, we biased our microsatellite search towards pairs of primers with melting temperatures around 59 °C, 2 % paired primer divergence and a primer size between 18 and 30 bp. After our initial filter, we then gave preference to microsatellites with >6 repeats and unique primers.

We extracted over 10,071 reads with microsatellite repeats. We randomly selected 48 sequences to synthesize primers and test them using at least three fish per population. Out of the 48, 16 worked consistently across populations. To facilitate genotyping, we added an M13-tail to the forward primer of each primer pair and followed a two-round PCR approach (Schuelke 2000). In the first PCR the targeted region that contains the microsatellite is amplified and in the second step the fluorescently labeled M13-tail is incorporated. After adding the M13-tail, 12 out of the 16 loci amplified consistently. Fragment scoring was performed in Geneious 7.1.7 (Kearse et al. 2012).

Allelic diversity ranged from 0.652 in locus *7983* to 0.976 in loci *170, 150, 246,* and *1358*. We found all loci under Hardy–Weinberg equilibrium, except loci *1588* and *7983* (Table 1). All microsatellites amplified well across populations without null alleles. The consistent amplifications along with the high levels of variation across these 12 microsatellites provide a powerful tool to evaluate the scale of connectivity among jawfish populations across MPAs and test whether such populations are connected and sustainable in the long-term.

Acknowledgments Funds came from the Puerto Rico Sea Grant, the Caribbean Coral Reef Institute, and the College of Arts and

Sciences-UPRM. We thank M. Hellberg at LSU for facilitating the genotyping of all samples. We thank Duane Sanabria, Anibal Santiago, Milton Carlo, and "Godito" for field assistance.

References

Jones GP, Milicich MJ, Emslie MJ, Lunow C (1999) Self-recruitment in a coral reef fish population. Nature 402:802–804

Kearse M et al (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649

Mora C et al (2006) Coral reefs and the global network of Marine Protected Areas. Science 312:1750–1751

Prada C, Hellberg ME (2014) Strong natural selection on juveniles maintains a narrow adult hybrid zone in a broadcast spawner. Am Nat 184:702–713

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–234

Chapter 3

Effective Dispersal of Reef Fish is Smaller than Current Spacing Among Marine Protected Areas

To be submitted to PLoS One

Abstract

The world's oceans are deteriorating at a fast pace, in particular coral reef ecosystems. Conservation measures, such as Marine Protected Areas, are being implemented to relieve some areas from local stressors and allow populations to restore to natural levels. Successful networks of MPAs can operate if the space among MPAs is smaller than the dispersal capacity of the species under protection. Here we studied connectivity patterns across populations in a series of MPAs across Puerto Rico and the Dominican Republic in the common yellowhead Jawfish, *Opistognathus aurifrons*. Using 18 molecular markers, we estimated that the maximum effective dispersal is 10 km. We found that MPAs exchange migrants likely via intermediate available unprotected habitats through stepping stone dispersal. At scales > 100 km such connectivity is decreased, particularly across the Mona Passage, making Mona Island a genetic mosaic of the genetic variation from the eastern and western Caribbean. Surprisingly, the Mona Passage is a leaky barrier and has historically permitted exchange of migrants in both directions. The MPA network studied is unable to maintain adequate levels of connectivity of these small benthic fishes if habitat in between them is

extirpated. Given that overall reef fish diversity is driven by species with life histories similar to that of the yellowhead jawfish, managers face a challenge to develop strategies that allow connectivity and avoid isolation of populations and their possible extinction.

Introduction

Tropical marine ecosystems have degraded in the last decades as a result of human activities (Mora 2008). This decline is marked in coastal areas, where coral reef organisms have decreased and species commonly seen in the 1970s are rarely observed today, especially commercially exploited species and apex predators (Steneck *et al.* 2009). One conservation initiative to restore marine populations is the designation of networks of Marine Protected Areas (MPAs), which in their most restrictive form are no-take. No–take MPAs are fishing-free spaces that, when properly implemented and managed, can restore populations if no other stressors are present. These MPAs act by increasing the population spawning potential of overexploited species (Roberts & Polunin 1991). Additional regulations may also relieve marine populations from local stressors by controlling land-based sources of pollution. Networks of MPAs are conceptually based on the idea that most marine populations are ecologically connected by the dispersion of planktonic larvae (Roberts 1997). The magnitude and spatial scale of larval dispersal determine the degree of connectivity among marine populations, providing information on the ideal reserve size and the minimum spacing among reserves to achieve both self-recruitment and maintain connectivity and diversity (Sale *et al.* 2005). Determining the scale of this effective movement of larvae is critical to designing successful MPA networks.

In theory, MPAs should work efficiently, as most marine systems lack obvious barriers to dispersal so that connectivity is achieved over large spatial scales (Roberts 1997). Reproduction in marine systems usually results in planktonic larvae that disperse

from days to months in the ocean, likely connecting populations over hundreds of kilometers. Physical models of passive particles have reinforced the idea of marine connectivity (Heck & McCoy 1978), and genetic inferences have provided indirect evidence of gene flow over thousands of kilometers (Grosberg & Cunningham 2001; Lessios *et al.* 1998; Lessios & Robertson 2006; McFadden *et al.* 1997). However, recent studies (Hellberg 2007; Hellberg *et al.* 2002; Swearer *et al.* 2002; Taylor & Hellberg 2003, 2006; Underwood *et al.* 2007; Warner 1997) have challenged previous assumptions of high connectivity among marine populations over large spatial scales (Heck & McCoy 1978; Roberts 1997).

In fact, when genetic data is incorporated into physical oceanographic models, populations seem locally maintained and connectivity among populations restricted (Baums & Paris 2006; Cowen *et al.* 2006; Galindo *et al.* 2010; Paris *et al.* 2005). Understanding whether populations are connected and the spatial scale at which this occurs is central to designing effective MPAs. Estimating larval dispersal is challenging, as common mark-recapture studies do not work efficiently in populations that produce millions of planktonic larvae that quickly dilute in the sea and almost all of which die (Thorrold *et al.* 2006). Fortunately, genetic surveys provide a powerful indirect estimate of population connectivity (Hellberg 2007). If populations are connected, genotypes should be randomly distributed across populations, but if genetic exchange is restricted, or if migrants are unable to establish and reproduce in the new population, differentiation among populations develops at the genetic level over time (Slatkin 1987). Similarly, multi-locus genetic data can capture subtle genetic variation across

geography, revealing the effective movement of migrants and their per generation dispersal (Rousset 1997).

Connectivity estimates inferred from genetic data (i.e., F_{ST}) are often the product of evolutionary processes across multiple time scales in the history of a species, including those at ecological scales, which are the most useful for designing MPA networks. Fortunately, we can measure connectivity at different time scales by applying different techniques. Gene genealogies within a coalescent framework (*i.e.*, Isolation with Migration, IMa2; n-Migrate) integrate migration rates over time scales and provide long-term estimates of migration. Conversely, assignment tests such as the ones implemented in BayesAss (Wilson & Rannala 2003) or the Bayesian admixture models of STRUCTURE (Pritchard *et al.* 2000) estimate contemporary connectivity patterns.

Here, we used multilocus sequences and microsatellites to test whether existing MPAs are genetically connected, allowing efficient movement of migrants among them, and large enough to sustain marine populations. We studied genetic variation across a longitudinal gradient of seven MPAs that spans the Mona Passage, a well-known biogeographic barrier between Puerto Rico and Dominican Republic. We first inferred genetic diversity and estimated effective population size in the common reef fish, *Opistognathus aurifrons* (Jordan & Thompson, 1905). We then quantified variation across geography and using *in-situ* fish densities estimated the per generation effective dispersal. Finally, we tested for variation between contemporary and historical connectivity across these MPAs.

Our data suggest a large effective population size, on the order of 7,170,477 breeding individuals, with an average effective movement of < 10 km per generation. MPAs are connected at scales smaller than 100 km, likely as a result of migrants moving across generations through intermediate unprotected reef areas. Contemporary patterns of connectivity suggest a decrease in connectivity across the Mona Passage, separating populations east of Mona Island (including Mona) from those west of it. Despite such a decrease, we observed considerable cross-population ancestry in the most recent generations. We also detected long-term historical genetic connectivity across the Mona Passage for both microsatellite and DNA sequences. Our data indicate that current MPAs, even within Puerto Rico, are too far apart to achieve long-term conservation of the huge diversity of small benthic fish species as represented by the yellowhead jawfish.

Materials and Methods

Studied Species

To estimate effective dispersal among MPAs, we chose the small (< 10 cm) benthic yellowhead jawfish (Jordan and Thompson, 1905), a common Caribbean species that occurs from Florida and the Bahamas to Barbados and northern South America (Colin 1971). The yellowhead jawfish is a sedentary benthic fish that lives in large decorated burrows (30-50 cm) on calcareous sand habitats at depths between 3 – 50 m. It feeds on zooplankton by hovering up to ~1 m above its burrow. Burrows also serve as mating nests, in which males invite females by lateral displays. Spawning occurs in the burrow and the male orally incubates the eggs (Colin 1971, 1973). Upon hatching, planktonic larvae are released, presumably in an advanced stage (post-flexion), that stay in the water column for up to three weeks (Ho *et al.* 2012; Young 1982). The species has moderate (~ F_{ST} = 0.3) average levels of genetic structure at the Caribbean scale (Ho *et al.* 2012). The ecology of the species has been studied (Colin 1971), with estimates of fish density in available in Puerto Rico (LeGore *et al.* 2006) and Panama (Hess 1993). Given our goal of capturing potential subtle genetic changes in a fish with pelagic larvae across a densely sampled geographical area, the benthic jawfish is ideal. While the yellowhead jawfish is representative of other fishes with the same biphasic life history (*i.e.*, planktonic larva, sedentary benthic adult), being an egg-brooder producing larvae with advanced swimming capability suggests its degree of larval dispersal may be reduced, which may result in lower dispersal and enhance our ability to detect potential genetic structure at fine spatial scales. Cryptic egg-brooding fishes represent a large portion of the biodiversity on coral reefs. Lastly, the species is harvested by aquarists

(Bejarano *et al.* 2015; LeGore *et al.* 2006), so the results will have direct application in testing if current MPAs would protect it and similar species.

Study Site, Sample Collection and Density of Fish and Habitats

To evaluate genetic connectivity at fine spatial scales, we sampled populations of the yellowhead jawfish in a network of MPAs along the Dominican Republic and Puerto Rico. Our sampling locations coincide with areas of major reef development and thus high species diversity. We focused on the southern (Caribbean) side of both Dominican Republic and Puerto Rico, where most of the reef development occurs. On average, the geographic distance between any two sampled MPAs is 90 km, with the largest distance between the two extremes being 486 km. A total of 260 individuals were collected from seven Marine Protected Areas along the Mona Passage (Fig. 3.1). Two in the Dominican Republic: Parque Submarino La Caleta (28), and Parque Nacional Natural del Este (29), and five in Puerto Rico: Mona Island Natural Reserve (57), Desecheo Island Marine Reserve (46), Tres Palmas Marine Reserve (16), La Parguera Natural Reserve (55), and Canal Luis Peña Marine Reserve (29) at Culebra Island. All samples were collected by scuba diving using clove oil as anesthetic at depths between 6 to 35 m. We collected morphometric measures for each sample and stored muscle tissue in 100% ethanol at -80 °C for genetic analysis. All specimens were vouchered at the Marine Sciences Fish Collection of the University of Puerto Rico and are available upon request.

Molecular Markers and Genotyping

We genotyped individuals at 18 molecular markers - one mitochondrial: the 5'end of the mitochondrial control region (mtCR); three nuclear encoding genes: nuclear recombination activating gene 1 (Rag1); RH2a-Opsin (RH2a); intron V from nuclear alpha-tropomyosinc (Atrop); two new anonymous nuclear markers: 1777E4; 4174E20; and twelve newly designed microsatellites (Beltran *et al.* 2014). All individuals were genotyped for all twelve microsatellites, 145 were also sequenced for Dloop and nuclear genes. We used previously developed primers for RAG and *Atrop* and developed new primers for the RH2a-Opsin and the other two new anonymous sequenced markers (Table S3.1).

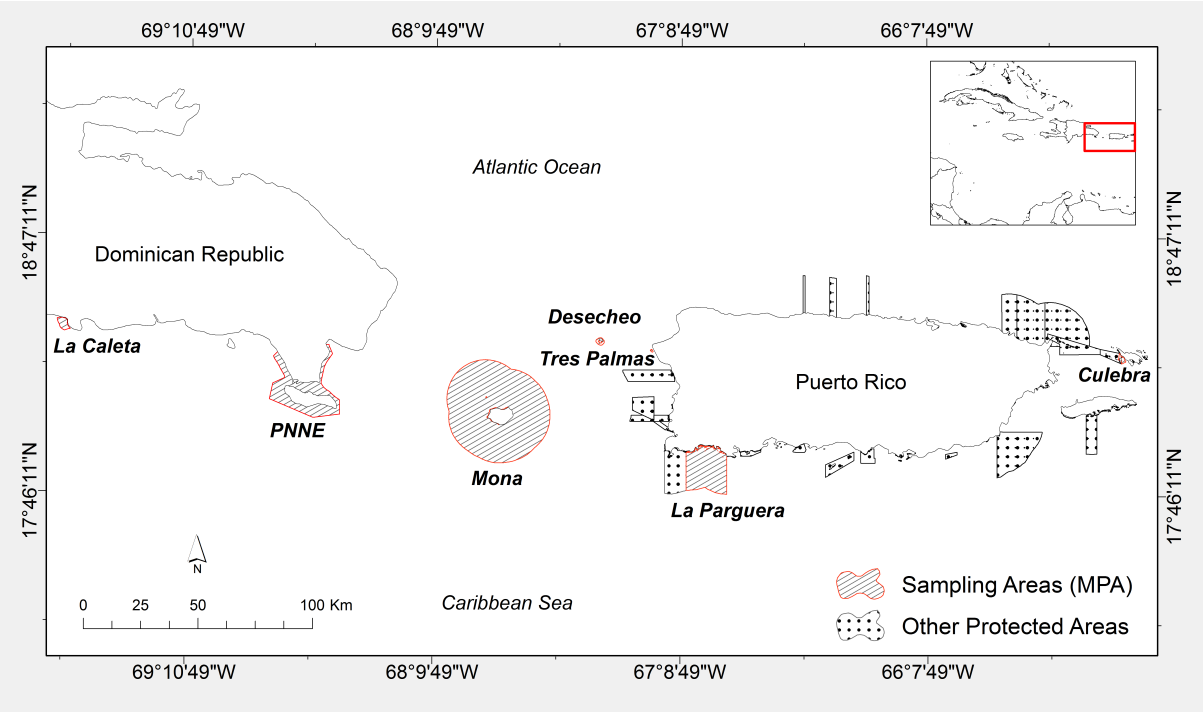

Figure 3.1 Geographic localities at Mona Passage and Culebra Island at Puerto Rico from which *Opistognathus aurifrons* were sampled. Note that in Desecheo Island Marine Reserve, Tres Palmas Marine Reserve and Canal Luis Peña Natural Reserve the size of the Reserve is too small to be perceived in the map.

To develop these new markers, we used 454 pyrosequencing of genomic DNA. Briefly, we extracted DNA from an individual from La Parguera PR. The extracted DNA was fragmented by sonication, and adapters were ligated and pyrosequenced at the Duke University Genomic Center. We generated a total of 1,081,023 sequence reads with an average read length of 558 bp. Given the large average fragment size, we BLAST all sequences against the Swiss-Prot database and identified 5,345 homologous protein coding sequences. From this pool we designed primers for 50 molecular markers; 12 consistently amplified across populations and were single copy nuclear markers. We used these twelve markers for our population genetic analysis.

To generate genotypic data, we extracted genomic DNA from muscle tissue using the QIAGEN DNeasy Kit following manufacturer's protocol. PCR amplifications were performed in a Bio-Rad thermal cycler following standard procedures with varying cycling conditions. A detail description of primers sequence, PCR amplification conditions, including annealing temperature for each marker is provided in Table S3.1.

To generate sequence data, we produced cycle-sequencing reactions in both directions to add fluorescent labels and analyzed them on an ABI 3130xl using the amplification primers. Sequences for each gene were assembled, edited, and aligned using Geneious R8 8.1.4 (Kearse *et al.* 2012). We resolved haplotypes using PHASE, version 2.1 (Stephens & Scheet 2005; Stephens *et al.* 2001). We used a recommended 80% probability as our cutoff (Harrigan *et al.* 2008). Once phased alleles were uncovered, we converted them to numeric haplotypes using FABOX 1.41 (Villesen 2007), and used

them as input in GENALEX 6.5 (Peakall & Smouse 2006, 2012), GENODIVE 2.0 (Meirmans 2014; Meirmans & Van Tienderen 2004) and STRUCTURE 2.3.2 (Pritchard *et al.* 2000).

To genotype individuals at microsatellite loci, we used our described two-step protocol (Beltran *et al.* 2014). First, we amplified the region containing the microsatellite sequence, and then we added the fluorescently labeled M13-tail. We genotyped individuals by multiplexing four microsatellites. To mix microsatellites, we included loci with different fragment sizes and used three fluorescent colors (HEX, NED, and FAM). We visualized all PCR amplicons using an automated sequencer (ABI 3130xl). We used an internal size standard (ROX-400) to estimate microsatellite sizes. We analyzed chromatograms with size standards and scored fragments in Geneious R8 8.1.4 (Kearse *et al.* 2012). For spurious scoring peaks, we repeated the PCRs and ran them individually on the sequencer as single samples. We also tried failed individuals at least five times before deeming them as failed. After our genotyping procedure, we generated 2,900 genotypes out of 3,120 possible.

To evaluate deviations from Hardy–Weinberg equilibrium in microsatellite data, we used MICROCHECKER 2.2.3 (Van Oosterhout *et al.* 2004). We found all loci under Hardy-Weinberg equilibrium, except loci 1588 and 7983. We quantified genetic differentiation and gene flow including and excluding these two loci, and results were qualitatively and quantitatively similar. Analyses for all loci are presented.

Genetic Diversity, Isolation by Distance and Effective Dispersal

We quantified genetic diversity across markers in GENODIVE 2.0 (Meirmans & Van Tienderen 2004) and estimate heterozygosity and number of alleles per population across all 12 microsatellites. To infer variation in sequences, we used DNAsp 5.10.1 (Librado & Rozas 2009) and estimate nucleotide (π) and segregating (θ) diversity, number of haplotypes and segregating sites. We estimated genetic variation across geography, by first calculating pairwise F_{ST}'s among locations as implemented in GENODIVE 2.0 (Meirmans & Van Tienderen 2004). We then estimated pairwise linear distance between locations in Google Earth 6.2 using the shortest nautical distance among populations. We test for normality of our observations via Shapiro-Wilk test (W = 0.96, p = 0.6). We then plotted genetic and geographical distances, estimated isolation by distance (IBD) and assessed significance of IBD assuming a linear regression in R (R Development Core Team 2013). We also used Pearson's non-parametric correlation to estimate slopes, while r^2 increased; the slope was on the same 3 e-5 range (Fig. S 3.1.

To estimate effective dispersal, we used the Rosset's (1997) approach:

$$\sigma = \sqrt{\frac{1}{4D_e m}}$$

where D_e is the effective density and m is the slope derived from our analysis of isolation by distance ($F_{ST}/(1-F_{ST})$). To estimate D_e we used both a direct density

estimate of fish per kilometer and one derived from genetic data by estimating effective population size.

Field density of the yellowhead jawfish was compiled from studies in Panama, Rincon and La Parguera (Hess 1993; LeGore *et al.* 2006). From those studies, we estimated density of fish to vary from 5,554 – 74,400 fish per km^2 in suitable habitats. We used these density estimates to estimate overall population size by adjusting values to the amount of habitat per reef area. We estimated differences in habitat areas by measuring habitat patch sizes from GIS maps. These estimates are likely biased upwardly and represent the upper bound of the highest fish density.

To measure effective dispersal from genetic data, we used effective population size (*Ne*) as inferred from IMa2 ($\Theta = 4\ Ne\ \mu$; where Θ is the amount of genetic variation in the population; *Ne* is the total effective size of the population and μ is the mutation rate). Once we had an estimate of the effective population size, we obtained fish density per km^2 by dividing by the total coral reef area across the Caribbean (26,000 km^2 (Burke & Maidens 2004)). Our density from genetic data is biased towards the lowest possible density, as long-term effective sizes are lower compared to census size (*e.g.*, for instance in humans *Ne* is only 10,000 and census is > 7 billion; (Hill 1981). In addition, we divided by the reef area across the Caribbean, but the yellowhead jawfish is a habitat specialist and occurs only on colonized pavement with sand channels close to coral reef areas, which in Puerto Rico make < 30 % of the reef area (LeGore *et al.* 2006). By having the direct census size as the upper bound of the highest density and

the genetic inference as the lowest possible density, we provide the widest range of effective dispersal estimates.

Genetic Differentiation and Contemporary Connectivity

To test for genetic differentiation across the sampled range, we used linkage disequilibrium among markers via Bayesian clustering as implemented in STRUCTURE 2.3.2 (Pritchard *et al.* 2000). We ran STRUCTURE in three different datasets: all data, only microsatellites or only nuclear. For each partition, we ran STRUCTURE without information of the origin of each individual, thereby reducing potential biases. We assumed an admixture model with a burn-in of 900,000 steps followed by 20 million iterations and 3 replicates per run. We ran STRUCTURE using a range of K values (number of inferred populations) from 1 to 7 (maximum number of populations) and then used STRUCTURE HARVESTER (Earl & VonHoldt 2012), which implements the Evanno method (Evanno *et al.* 2005) to infer the optimal number of populations (K's) in our dataset. We processed replicates with best K in CLUMMP (Jakobsson & Rosenberg 2007) using the default parameters. To generate figures we used Distruct 1.1 (Rosenberg 2004).

To further understand the distribution of genetic variation across geography, we calculated principal components in GENODIVE 2.0 (Meirmans & Van Tienderen 2004) and principal coordinates in Genalex 6.5 (Peakall & Smouse 2006, 2012) (Peakall & Smouse 2006, 2012). These multivariate approaches allowed us to plot the major axis of variation among both populations and individuals (Peakall & Smouse 2006, 2012).

To test for contemporary gene flow and inbreeding patterns among populations, we used the Bayesian MCMC statistical approach implemented in BayesAss 3 (Wilson & Rannala 2003). BayesAss is most powerful when migration rates are low (< 30% of the population), as it is the case across the Mona Passage, and migration occurred in the last generations (Wilson & Rannala 2003). Our STRUCTURE analysis suggests gene differentiation between Puerto Rico and the Dominican Republic at the Mona Passage. To analyze our data both in BayesAss and in IMa2, we pooled our data into three populations: Mona Island, West of Mona and East of Mona. For each analysis we ran BayesAss for a burn-in of 5,000,000 steps and inferred parameters using the next 100,000,000 iterations sampled at 20,000 intervals. To ensure appropriate mixing of the MCMC runs we adjust parameters to be within 20% -60% acceptance rates using information from preliminary runs. We ran the program at least three times for each dataset each starting from a different random seed. Mixing and convergence of MCMC runs were inspected using TRACER 1.6 (Rambaut *et al.* 2014). Replicates show similar results; here we present the one with the lowest deviance probability as suggested by (Meirmans 2014).

Historical Connectivity

To quantify historical connectivity across the Mona Passage, we used the isolation with migration model implemented in IMa2 (Hey 2010). To analyze our data in IMa2, we used the same population partition as in BayessAss. Priors in IMa2 were obtained in preliminary short runs varying all parameters. After we had confidence in our priors and mixing properties of the chains, we initiated three longer runs per comparison. We

ensured that our runs had a burn-in of at least 2,000,000 followed by 5,000,000 steps. Trend lines were visually inspected for proper mixing and convergence. To compare among different runs, the migration rate was scaled by the effective population size using the formula $M=Nm=\Theta m/2$. We plotted the density functions against the migration rate to visualize comparisons.

Results

Genetic Diversity

Sequenced markers were variable within populations, with the Dloop being the most variable (θ = 0.061 per site for all samples) and the RH2a-Opsin being the least variable (θ = 0.0038). We did not detect any departures from neutrality as inferred from Tajima's D and Fu's Fs (Table S3.2). Across the 260 individuals sampled at 12 microsatellite loci, the numbers of alleles varied from 13 to 16 with a mean of 14 (Table S3.3). Loci did not depart from Hardy-Weinberg expectations ($p > 0.05$; except loci 1588 and 7983) and were also segregating independently as we did not detect any linkage in pairwise comparisons ($p > 0.05$). Heterozygocity across microsatellite loci ranged from 8.05 to 0.889 with a mean of 0.874 ± 0.029.

Genetic variation segregates gradually across geography

Inbreeding coefficients among populations suggest genetic variation is largely partitioned by geography (Fig. 3.2). The isolation by distance (IBD) analysis suggests geography explains 57% of the genetic variation among populations across this MPA network ($r^2 = 0.57$; $p < 0.05$; Fig. 3.3). We recovered similar estimates when we used the microsatellite data alone ($r^2 = 0.42$; $p < 0.05$) or only the sequence data ($r^2 = 0.43$; $p < 0.05$). Our pattern of geographic isolation generates an IBD slope of 8×10^{-5} (3.6×10^{-5} when using Pearson's correlation) when using only the microsatellite data and a slightly lower value when using all 18 markers (1×10^{-5}). As a result of the geographic isolation, pairwise F_{ST}s were highest when comparing populations from West (La Caleta and PNNE, both locations in southeast Dominican Republic) and East of the Mona

Passage ($p < 0.002$). The Mona population also showed genetic differentiation when compared to all other populations ($p < 0.003$), except when compared with Tres Palmas (Table 3.1). Similar to the IBD pattern, the principal component analysis shows a gradient of populations that resembles the geographic distance among them (Fig. 3.3).

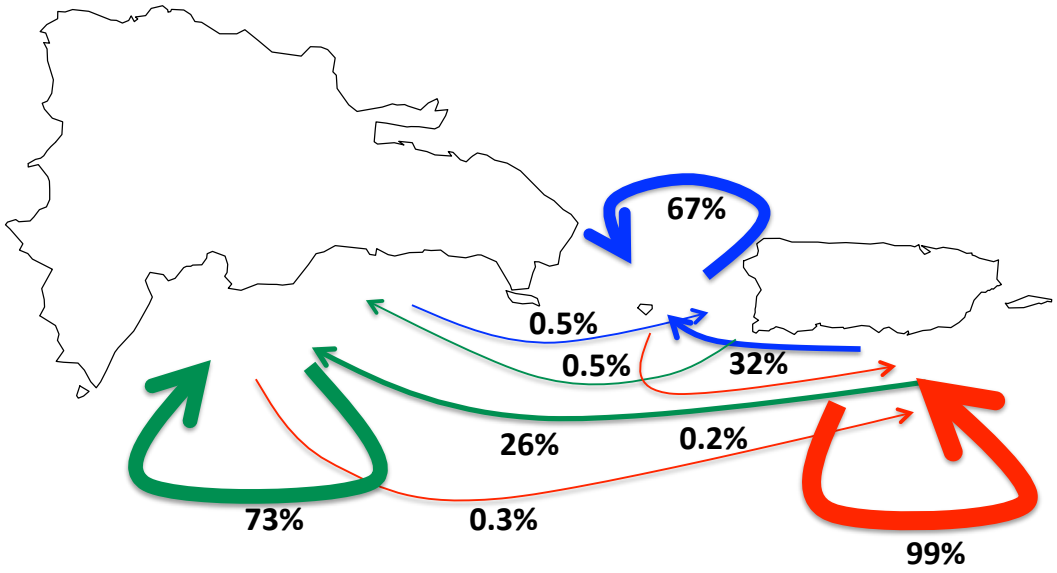

Figure 3.2. Cluster analysis results for *Opistognathus aurifrons* from seven sampling sites along Mona Passage as obtained from Structure. Values are the probability of membership of each area: west of Mona (green), Mona Island (blue), east of Mona (red).

Table 3.1 Spatial genetic structure across the Mona Passage as derived from 12 microsatellites. Bonferroni correction (0.05/21= 0.023). Significantly different pairwise comparisons are shown in red ($p < 0.0023$).

	Tres Palmas	La Parguera	PNNE	Mona	La Caleta	Desecheo	Culebra
Tres Palmas	--	0.208	0.273	0.361	0.001	0.329	0.482
La Parguera	0.208	--	0.001	0.001	0.001	0.002	0.392
PNNE	0.273	0.001	--	0.001	0.001	0.001	0.001
Mona Island	0.361	0.001	0.001	--	0.001	0.252	0.001
La Caleta	0.001	0.001	0.001	0.001	--	0.001	0.001
Desecheo	0.329	0.002	0.001	0.252	0.001	--	0.039
Culebra	0.482	0.392	0.001	0.001	0.001	0.039	--

PCA for individuals also shows the geographical partition of genetic variation seen in the F_{ST}'s and the subtle but noticeable genetic break west of Mona Island (Fig. 3.4). It also depicts the amount of overlap in genetic variation at Mona Island from individuals sampled in the Dominican Republic and those sampled in Puerto Rico. The first two principal components explain 72% and 18% of the genetic variation. Component one (x-axis on Fig. 3.4) separates populations from west to east, while component two (y-axis on Fig. 3.4) segregates populations within Puerto Rico.

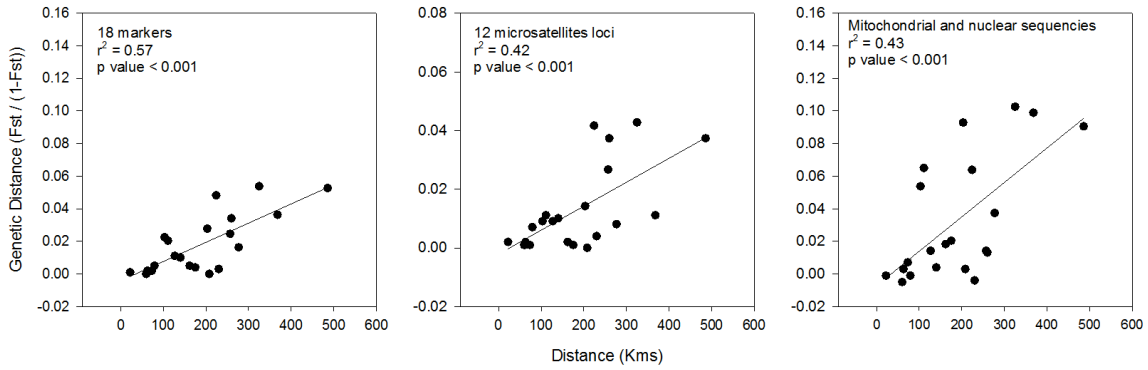

Figure 3.3. Isolation by Distance in *Opistognathus aurifrons* from seven sampling sites across the Mona Passage for different markers combination. A. For 18 markers indicating > 57%; B. 12 microsatellites indicating > 42%; C. Six markers (five nuclear genes + one mitochondrial genes) >43% of the variation of pairwise genetic distance among locations is explained by geographic distance, all with significant $p < 0.001$.

Effective Dispersal is less than ten kilometers

Density of fish estimates from field surveys ranged from 5,554 – 74,400 fish per km^2. As expected these estimates are biased towards high densities as they were done in habitats and locations in which the species is common. In contrast, our estimates from genetic data are low. Effective population size from genetic data varied from 789,787 to

20,660,273 with a mean of 7,170,477 breeding individuals across the whole Caribbean (26,000 km^2), with a density of 275 fish per km^2. Applying these density estimates to Rousset's equation, the effective dispersal for the yellowhead jawfish ranges from 0.2 to 3.39 km. If we use a smaller IBD slope obtained from samples only from Puerto Rico and the smallest density, we estimate an effective dispersal of 9.61 km.

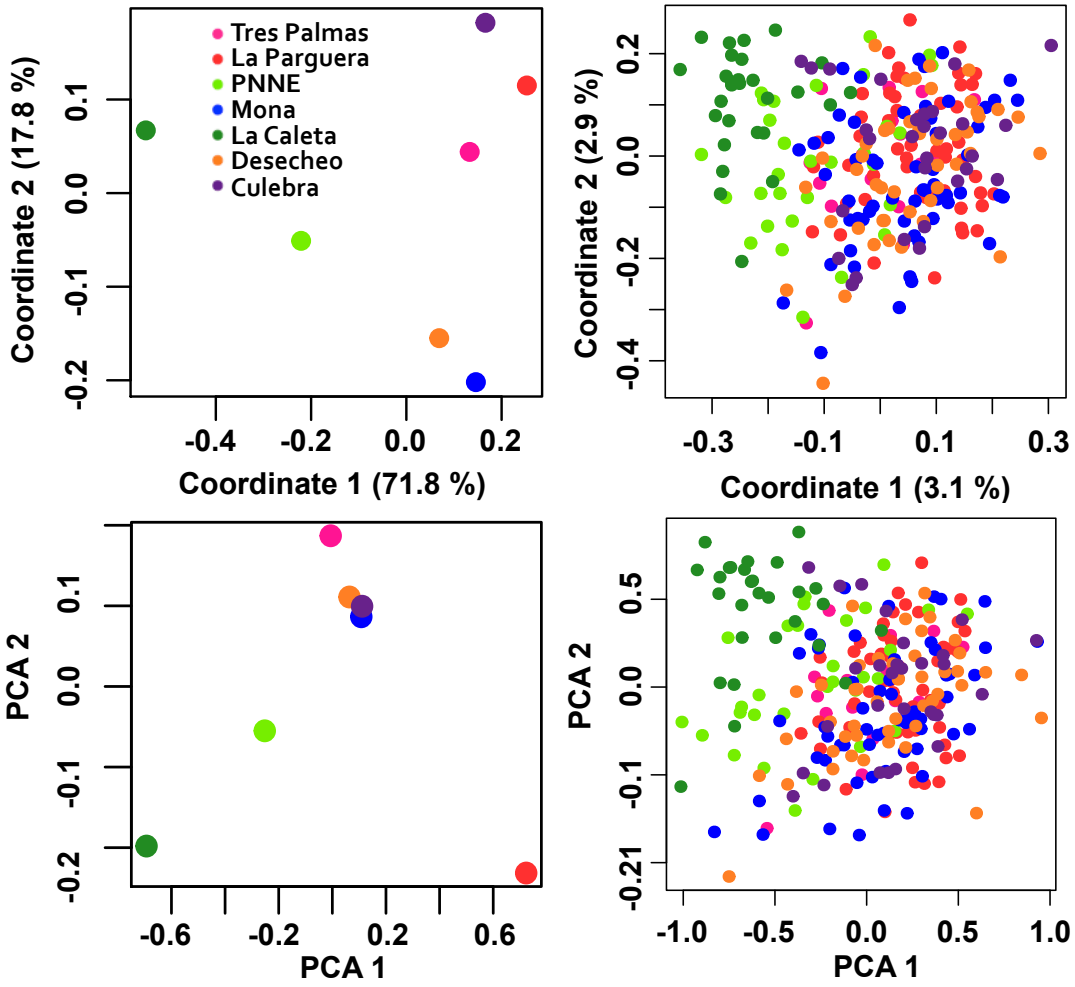

Figure 3.4. Multidimensional summary of genetic variation. Top panels show multidimensional scaling (MDS) in microsatellites genotypes for populations (left) and individuals (right). Bottom shows principal component analysis (PCA) of the same genotypes (left: populations; right: individuals).

Restricted contemporary gene flow across the Mona Passage

STRUCTURE analysis suggests the most likely number of populations is K=2, indicating the presence of two genetic pools across the sampled area, represented as green and red in Figure 3.5. One isolated population (the green) is more common West of the Mona Channel in the Dominican Republic (La Caleta and PNNE) and replaced gradually by the red genotype as one moves eastward to Puerto Rico. Estimates of the proportion of the number of individuals displaying the green genotype is significantly different when comparing populations from the Dominican Republic and those from Puerto Rico (Fisher Exact Test, $p < 0.001$). Genetic differentiation occurs between Puerto Rico and the Dominican Republic, with Mona Island being the westernmost point connecting the two isolated populations.

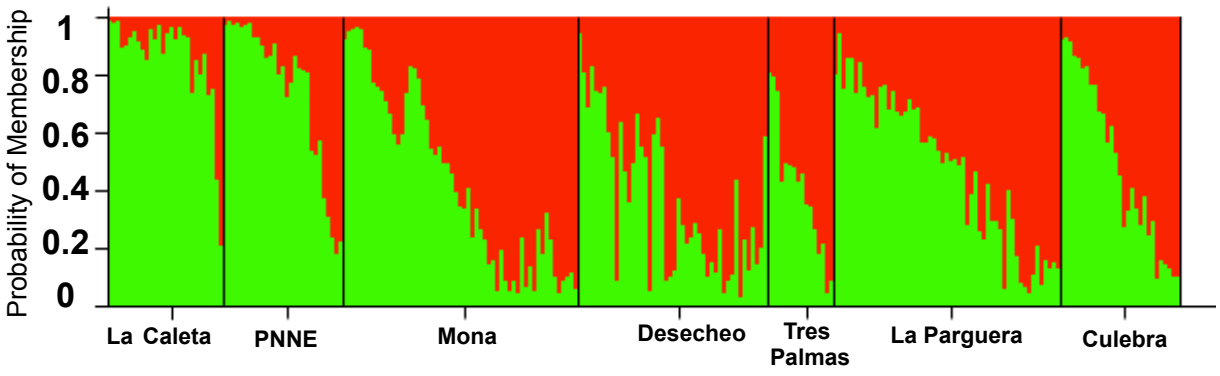

Figure 3.5. Graphical summary of clustering Analysis for *Opistognathus aurifrons* genotypes from seven Marine Protected Areas along Mona Passage using Structure. Probability of Membership of each sampling site (n=7) from individuals per site scored with 18 markers. Each vertical line represents an individual and the estimated proportion of the individual's genome from each inferred cluster.

Limited contemporary migration contrasts with rampant historical gene flow across the Mona Passage

We found extensive historical migration across the Mona Passage. Migration rates were higher when comparing both Puerto Rico and Dominican Republic against Mona Island. Mona receives on average ~ 162 gene copies per generation from Puerto Rico and sends 157; while it receives 43 from the Dominican Republic and sends 2. As predicted the smallest exchange occurs between Puerto Rico and the Dominican Republic with 22 migrants per generation going to Puerto Rico from the Dominican Republic and only 0.5 returning back

Discussion

The benefits of preserving marine reserves are evident when spacing among protected areas is small enough to allow larvae of neighboring populations to achieve genetic connectivity (Sale *et al.* 2005). It is thus critical to estimate the magnitude and direction of such exchange, evaluate if they represent contemporary bridges for the movement of migrants and quantify their historical potential. To understand connectivity among populations, we studied genetic variation and current and historical gene flow across a Caribbean network of MPAs in a common benthic reef fish with pelagic larvae. Analyzing variation at 18 genetic markers, we showed that: (i) The yellowhead jawfish has a huge amount of genetic variation with an effective population size on the order of millions of individuals. (ii) Effective migration per generation is less than ten kilometers generating a pattern of genetic segregation by geography, a pattern consistent for mitochondrial, nuclear sequences and microsatellites. (iii) Contemporary gene flow is restricted across the Mona Passage, segregating sites into two populations: those west of Mona, and all others. (iv) Restricted contemporary gene flow across the Mona Passage contrasts with historical gene flow in both eastern and western directions.

In the broadest sense our genetic analyses suggest populations within the studied MPA network connect through a stepping stone model via intermediate unprotected reef habitats. Effective dispersal within Puerto Rico of 10 km only allows neighboring populations to exchange migrants, and such limited connectivity results in a positive correlation between geographical and genetic distances as predicted from theory (Slatkin 1993). Across the Mona Passage, the effect dispersal distance was

substantially lower: 4 km. Below we present a perspective on the current spacing and connectivity among these already designated MPAs, the critical role of the Mona Island Reserve to preserve connectivity across the Passage and the contrasting patterns of historical and contemporary connectivity across this network of MPAs

Current MPAs are too far apart to allow connectivity in the absence of intermediate habitats

Our data suggest a pattern of genetic isolation with geography, with some populations separated only by < 100 km showing significant genetic differentiation. Our effective dispersal estimates also suggest the fish would have a neighborhood movement of less than 10 km per generation, providing an upper bound for the largest spacing among successful MPAs. Our dispersal estimates fall within previous estimates of dispersal such as for *Stegastes partitus* (9 km), *Hypoplectrus nigricans* (10 km) and *Hypoplectrus puella* (2-14 km) (Puebla *et al.* 2009; Puebla *et al.* 2012) and the anemone fish *Amphiprion clarkii* (4- 20 km) (Pinsky *et al.* 2010) but lower than in *Thalassoma bifasciatum* (27 km), *Haemulon flavolinueatum* (46 km) *Chaetodon capistratus* (52 km) (Puebla *et al.* 2012). The slope in our IBD estimates is also concordant with those inferred by Puebla (2012) in Belize and smaller (*i.e.* less genetic segregation) than those inferred by Pinski (2010) in the Philippines.

Dispersal estimates contrast with the mean distance among enforced MPAs (sampled in this study) within the network, which is 90 km and 37 km for all Management Areas, most designated yet unenforced. This is 9 X (or 4X) the ideal space among MPAs to ensure the direct connectivity desired to achieve a high level of protection. Our study

suggests that MPA networks with mean average space among MPAs of > 20 km are unsustainable in the long-term and may only be currently working thanks to the existence of intermediate unprotected habitats. We hypothesize that the current MPA network around Puerto Rico for small benthic fishes such the yellowhead jawfish may on evolutionarily scales expose isolated populations to extinction. Currently, connectivity among MPAs is possible through in unprotected intermediate habitats. Continued decline in reef health, however, may result in the disappearance of those habitats, resulting in fragmented isolated populations prone to demographic stochasticity and extinctions. MPA spacing of < 10 km, as suggested by earlier studies (Jones *et al.* 2009), would increase connectivity among adjacent MPAs and their long-term success.

Gene flow breaks down across the Mona Passage: Mona Island as a link between the east and the west of the Caribbean

We found a subtle genetic break at the Mona Passage, dividing populations east and west of it, with Mona Island being part of the eastern cluster with unequivocal western genetic traces in some individuals. Historically, however, gene flow has occurred between the two populations, and Mona Island is a critical link that has allowed long-term migration in both east and west directions. Our results fit the east-west pattern of differentiation, probably generated by the effect of deep ocean water in creating chaotic conditions and limiting larval migration across the Passage. For instance, in summer currents flow southward along the Dominican Republic and northward around Puerto Rico, while in winter surface water (up to 50 m) flows northward and deeper water (below 50 m) flow southward (Baums & Paris 2006). Such seasonal behavior of the currents is believed to be an important barrier that divides marine populations (Hellberg

2009), limiting larval dispersal across this narrow stretch during certain times of the year (Baums & Paris 2006). The genetic differentiation in the jawfish coincides with genetic studies of *Acropora* corals (Baums *et al.* 2005; Baums & Paris 2006; Vollmer & Palumbi 2007) octocorals (Gutiérrez-Rodríguez & Lasker 2004) goby fishes (Taylor & Hellberg 2003, 2006) and also with morphometric and color pattern variations in populations of fishes (Colin 1975; Dennis *et al.* 2005). The location of the genetic break, which divides eastern and western populations of Caribbean marine species varies, as Mege *et al.* (Mege *et al.* 2014) discovered *A. palmata* specimens with western genetic affinities in Puerto Rico. Our detailed sampling, however, revealed the biogeographic boundary occurs between Mona Island and the Dominican Republic.

Historical and contemporary gene flow

Our historical estimates of genetic variation in the yellowhead jawfish suggest that the current genetic break at the Mona Passage is a temporary barrier and connectivity could be restored as has happened during the historical demography of the species. Variation between contemporary and historical gene flow has also been found in the whitesnout anemonefish *Amphiprion mccullochi* in southeastern Australia (van der Meer *et al.* 2012). One caveat from such study is the use of the program n-migrate on microsatellite data for contemporary gene flow. n-migrate, similar to IMa2, uses the harmonic mean to integrate variation in gene flow over evolutionary times scales and provides an estimate of the long-term migration rate (Beerli & Palczewski 2010). In contrast, we used long-term migration for both sequence data and microsatellites under the IMa2 model as a measure of historical gene flow and STRUCTURE and BayesAss to estimate recent migration.

The Mona Passage is a temporary barrier where the major westward current does not impede the flux of migrants going eastward. This rampant gene flow across evolutionary scales is not surprising because the species has a pelagic larval duration of more than two weeks (15-21 days) (Young 1982), which could allow the fish in a single generation to travel 400 and 500 km (assuming passive dispersal and a current flow of 1 km/h) (Taylor & Hellberg 2003). While contemporary gene flow is of most use for conservation purposes (Jones *et al.* 2009), our results stress the role of historical migration in shaping contemporary patterns. In the context of MPA theory, this suggests that present-day barriers to gene flow may obfuscate the historical potential of dispersion of the species. Critically for MPA design, Mona Island represents a mosaic of the genetic variation from the east and the west and has historically been a stepping-stone to connect those populations. Preserving Mona is a large step forward towards keeping a condensed amount of the Caribbean genetic diversity and the likelihood of connectivity between the east and the west, if oceanographic conditions allow and conditions in the existing MPAs do not degrade further.

Conclusion

Our results suggest MPAs along the Mona Passage are genetically connected with their nearest neighboring populations, but the populations that are in the most remote parts of the Passage only exchange migrants through intermediate populations in unprotected habitats. Our findings stress the importance to create intermediate MPAs to generate a network that ensures connectivity within a radius of at least 10 km. Our contrasting pattern of contemporary genetic differentiation and historical gene flow highlights the importance of preserving populations, such the Mona population, at the edges of the current genetic breaks. Mona Island is a reservoir of genetic variation from both the eastern and western population and a critical link to allow historical pulses of gene flow. Species richness is driven by small benthic taxa such the yellowhead jawfish, to maintain high biodiversity on coral reefs, MPAs need to be closer.

Chapter 4

Multiple Evolutionary Significant Units:

Targets for marine conservation

Abstract

Populations segregate genetically over geographical areas and form evolutionary lineages, which have been identified in terrestrial vertebrates as critical when implementing conservation strategies. Such rationale is unusual in marine species because marine organisms disperse hundreds of kilometers as larvae, connecting populations separated even by deep-ocean barriers such as across the Pacific. In the last decades however there has been an increase in studies showing genetic differentiation in populations separated by as little as ten's of kilometers. These isolated populations warrant inclusion in management plans as Evolutionary Significant Units (ESUs). Using genotypes of over 260 individuals at mtDNA, twelve microsatellite loci and at > 10,000 single nucleotide polymorphisms, we studied geographical patterns of genetic variation in the common yellowhead jawfish. We found segregation across the Caribbean, with isolated groups in Belize, Florida, Curaçao and Puerto Rico representing at least four ESUs to consider across the Caribbean. We also found that genetic exchange is < 0.1% among groups. Our genomic data indicate that these ESUs should be incorporated to enhance marine management plans. The benthic yellowhead

jawfish shares its life history with the most specious families of reef fishes. Incorporating evolutionary knowledge into marine conservation in these rich groups may enhance the maintenance of reef fish biodiversity.

Introduction

Terrestrial vertebrates commonly segregate over the species's range forming independent lineages that are often designated as Evolutionary Significant Units (ESUs) for conservation. While each ESU has its own significance and management requirements, all united to preserve the species, i.e., species composed by multiple units can only be fully preserved if all evolutionary divergence is included into the management plan (Moritz 1994; Shaffer *et al.* 2015). The most salient feature of conservation based on units is the adverse effects of re-introducing individuals to rescue independently evolving sister lineages (Moritz 1994; Waples 1991). Conservation of ESUs in marine species have lagged behind because marine populations are thought to maintain genetic connectivity over large distances, thanks to the ability of larvae to disperse in the plankton for weeks to months and the lack of obvious barriers to dispersal (Roberts 1997). Indeed, studies have provided indirect evidence of genetically connected populations over thousands of kilometers (Grosberg & Cunningham 2001; Lessios *et al.* 1998; McFadden *et al.* 1997). Recent molecular studies, however, have shown marine populations are genetically isolated at the extreme of forming monophyletic groups (Eytan & Hellberg 2010).

In the Caribbean for example, gobies and blennies show strong differentiation around the Caribbean, suggesting thousands of generations of isolation despite modest (< 200 km) geographical separation and larvae with a three-week pelagic duration (D'Aloiaa *et al.* 2015; Eytan & Hellberg 2010; Taylor & Hellberg 2003). Such genetic differentiation is also reflected within Mesoamerican reefs in damselfishes (Salas *et al.* 2010). Genetic

isolation is not restricted to fish but also marine invertebrates such as *Acropora cervicornis* and *Pseudopterogorgia elisabethae*, which show differentiation at scales of 10-100 km with isolated populations around each major Caribbean island (Gutiérrez-Rodríguez & Lasker 2004; Vollmer & Palumbi 2007). At the extreme, populations separated by only 23 km are genetically differentiated (Taylor & Hellberg 2003) and pervasive yet hidden barriers such as the Mona Passage (Baums *et al.* 2005; Taylor & Hellberg 2003, 2006) and Exuma Sound (Taylor & Hellberg 2006) promote such genetic divergence. Extreme isolation in some marine species provides grounds to incorporate the Evolutionary Significant Unit concept to marine conservation, so that each unit receives distinct conservation priority. This is not only for protection focused on the individual species, which is, rarely done in species other than mammals, but also to define the minimum number of protected areas needed within a region (i.e., the wider Caribbean) to achieve full preservation of the phenotypic and genetic variation within species. Incorporating ESU information is now critical given the extinction of species before we even describe them, particularly for cryptic diversity (Victor 2015). It is thus vital to clearly delineate marine ESUs as done extensively in vertebrate terrestrial species (Moritz 1994). We explored this possibility in a commonly aquarium-harvested marine species: the yellowhead jawfish.

Traditionally, the identification of independently evolving lineages in marine species has been done using mitochondrial markers (Allendorf *et al.* 2010; Waples & Gaggiotti 2006). However, given the large effective populations sizes in marine species, their long and often convoluted history of sea level changes that seggregated and re-connected

populations, a single mtDNA marker, while informative fails to fully capture their evolutionary history and therefore to delineate these management units. In fact, mtDNA may be misleading if the gene genealogy does not correspond with the history of population split, which could occur due to the maternal inheritance of the mtDNA and the differential dispersal of sexes, natural selection in opposite direction to the genealogical tree, or more commonly by the random mutational and genealogical processes (Slatkin & Hudson 1991).

To overcome this evolutionary noise, researchers have used hyper-variable markers such as microsatellites, which by virtue of their huge variability and simultaneous use, can reveal even subtle genetic differentiation within species (Allendorf *et al.* 2010). Microsatellites, however, are often expensive to develop and there is uncertainty in their molecular model of evolution, which makes it difficult to reconstruct the evolutionary history of independent lineages. Studies using microsatellites are also difficult to replicate, so that data produced independently cannot be analyzed together, as scoring of fragment sizes varies among laboratories (Hobana *et al.* 2013). Fortunately, the drop cost in nucleic acid sequencing has boosted the use of genomic approaches to harness genetic variation in wild populations to inform conservation initiatives (Allendorf *et al.* 2010).

Genotyping of single nucleotide polymorphisms (SNPs) in wild populations has been accelerated by the use of Restriction Site Associated DNA sequencing (Miller *et al.* 2007) and its derivatives, GBS, (Elshire *et al.* 2011) and also to the feasibility of

obtaining SNP data through RNAseq (Allendorf *et al.* 2010). Despite the lower genetic diversity in each SNP compared to microsatellites, their well understood behavior of molecular evolution and the ease of sampling thousands of SNPs in a single assay, makes them powerful enough to reliably study evolutionary processes within species. Their statistical power comes from the fact that accuracy in demographic estimation, such as defining ESUs, increases more than linearly by increasing the number of markers (Beerli & Palczewski 2010; Brumfield *et al.* 2003; Felsenstein 2006; Kuhner *et al.* 2000). The slow mutation rate in SNP data often showing bi-allelic states also lessens the impact of misidentifying mutational models and alleviates noise from within population polymorphism to among population divergence (Meirmans & Hedrick 2010). Uniquely, SNP based approaches allow to evaluate and add adaptive genetic variation into conservation, such as for example defining ESUs (Bradbury *et al.* 2012; Bradbury *et al.* 2007; Narum *et al.* 2013).

We used mtDNA, microsatellite and SNP data to quantify genetic variation across the Caribbean in a common Caribbean reef fish widely used in the aquarium trade (LeGore *et al.* 2006). We first contrast patterns of variation at these three different markers, particularly at the Mona Passage, a known biographical barrier. We then provide fine scale estimates of genetic structure in populations that are within < 100 km apart and the genetic admixture that occur among these geographically close populations. Finally, we delineate the different ESUs across the Caribbean that could be incorporated into conservation actions. Our study demonstrates the immerse statistical power of genomic data to quantify fine scale geographical differentiation, estimate magnitude and direction

of migrations and its unique potential to define ESUs, thus enhancing coral reef fish conservation.

Materials and Methods

Study species

Opistognathus aurifrons is a common benthic reef fish heavily used for aquarium collections (LeGore *et al.* 2006). The yellowhead jawfish is territorial and build burrows in which it attracts females to mate. Male fish mouth-incubate fertilized eggs and release larvae in an advanced state of development, presumably post-flexion. Larvae stay for two to three weeks in the water column (Young 1982). Effective dispersal in this fish is only at most 10 km (Chapter 2) and as a result, it generates allele mtDNA haplotypic variation across the Caribbean (Ho *et al.* 2012). Despite such differentiation, monophyletic groups are yet to be found and with the randomness associated with sampling at a single locus, it is uncertain whether allele frequencies at the mtDNA are shared across the nuclear genome and therefore point to clear Evolutionary Significant Units. Here we complemented work by Ho *et al.* (2012) by sampling over 10,000 random SNPs across the nuclear genome.

Sampling, Molecular Markers and Genotyping

We collected *Opistognathus aurifrons* at 13 localities across the Caribbean: South Florida both in Key Largo and Miami; Bahamas at Bimini; Dominican Republic at Parque Submarino La Caleta and Parque Nacional Natural del Este; Puerto Rico at Mona, Culebra and Desecheo Islands, Tres Palmas and La Parguera; the USVI at St Thomas and also at both Aruba and Curacao (Fig. 4.1). To capture fish we dove to depths between 6 to 35 m and stunned fish with clove oil. Length of each fish was

recorded, each sample was vouchered at the Fish Collection of the University of Puerto Rico and muscle tissue stored in 100% ethanol at -80 °C.

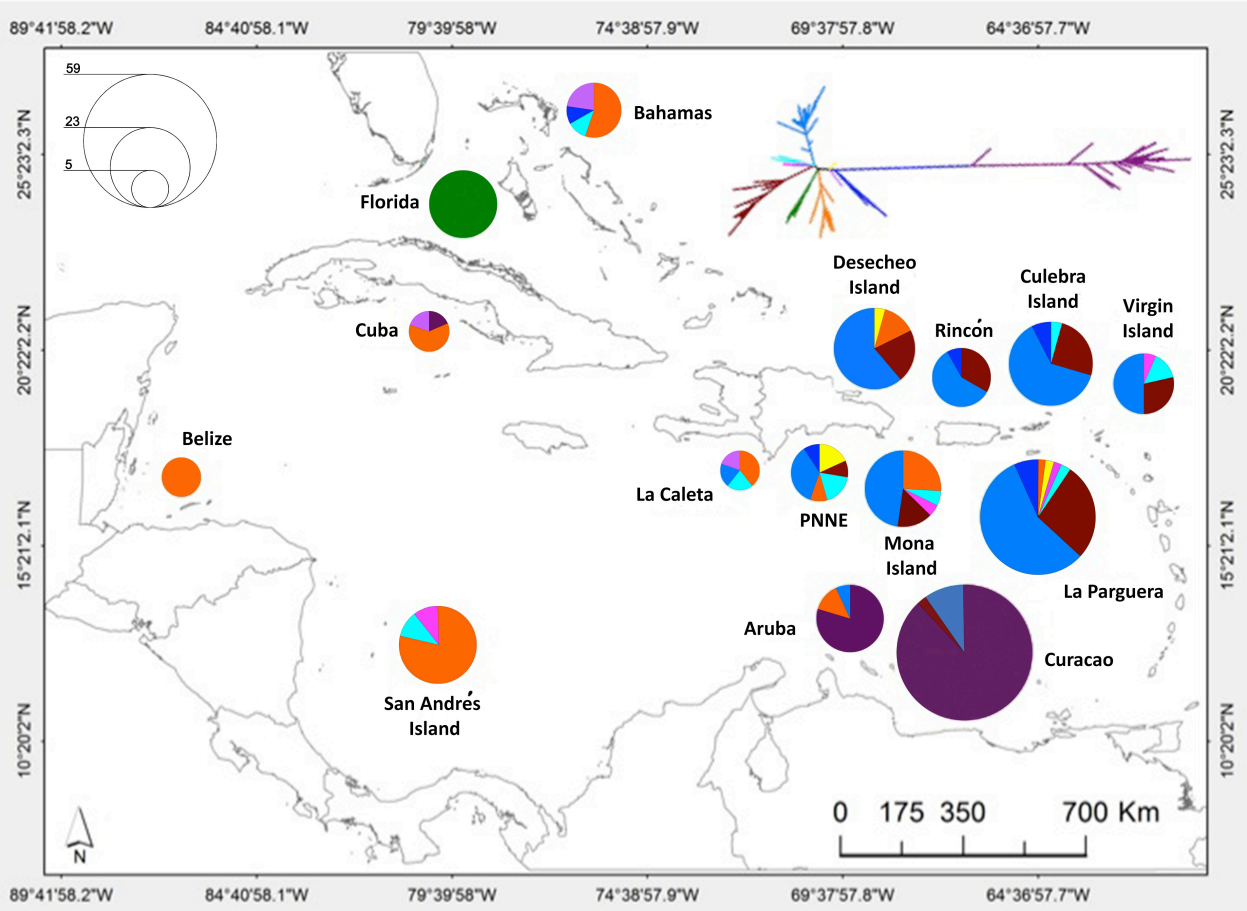

Figure 4.1. Sampling locations across the Caribbean. Different colors represent different haplotypes from concatenated mitochondrial (Dloop+COI) reconstruction (top right). Frequency distribution of mitochondrial haplotypes are shown in pie charts. Circles sizes represent sample size per location.

To genotype individuals, we first extracted genomic DNA from muscle tissue following the QIAGEN DNeasy Kit protocol. We used a Bio-Rad 4000 thermal cycler for PCR amplifications with varying cycling conditions depending on the marker. A detail description of primers sequence, PCR amplification conditions, including annealing temperature for each marker is provided in Table S4.1.

Using previously developed primers, we genotyped individuals for two mitochondrial genes; the 5' end of the mitochondrial control region (mtCR - Dloop) and the protein coding gene cytochrome oxidase I (COI) (Baldwin *et al.* 2009; Ivanova *et al.* 2007). We genotyped 269 individuals for Dloop and a subset (165) for COI. We produced cycle-sequencing reactions in both directions to add fluorescent labels and analyzed them on an ABI 3130xl using the amplification primers. Sequences for each gene was assembled, edited, and aligned using Geneious R8 8.1.4 (Kearse *et al.* 2012).

To genotype individuals at microsatellite markers, we used our earlier developed approach and markers (Beltran *et al.* 2014). We genotyped individuals by multiplexing four microsatellites simultaneously. To separate multiplexed alleles, we combined loci with different sizes and used three fluorescent colors (HEX, NED, and FAM). We visualize all PCR amplicons using an automated sequencer (ABI 3130xl), using an internal size standard (ROX-400) to estimate allele sizes. We analyzed chromatograms with size standards and scored fragments also in Geneious R8 8.1.4 (Kearse *et al.* 2012). For dubious scoring, we repeated PCRs and run them individually on the sequencer as single samples. We also tried failed individuals at least five times before deeming them as failed. After our genotyping procedure, we generated 3101 genotypes out of 3120 possible. In total, we genotyped 260 individuals at 12 loci, including those sampled for Dloop (except 9 individuals) and COI. We evaluated deviations from Hardy–Weinberg in microsatellites using MICROCHECKER 2.2.3 (Van Oosterhout *et al.* 2004). We found all loci under equilibrium, except loci 1588 and 7983. We carried out analysis of genetic differentiation and gene flow including and excluding these two loci

and results were qualitatively and quantitatively similar. Here we present results from the full dataset.

In addition, we obtained SNP genomic data across populations from a subset of 95 individuals using Genotyping by Sequencing (GBS) (Elshire *et al.* 2011). We included representatives from eleven populations (all, except Aruba and USVI). GBS was performed at the Cornell Institute of Genomic Diversity. GBS is a simple and well-standardized approach to collect massive amounts of loci. In brief, DNA samples are barcoded and adapter pairs ligated in dry. Restriction enzymes are then used to fractionate the genomic DNA. An optimization procedure determined the *PstI* (CTGCAG) was the best for the yellowhead jawfish. Once fragments are generated, adapters are ligated to the end of the fragments. Samples are then pooled and size selected to exclude unreacted adapters. Libraries were then amplified using an 18-cycle PCR with long primers complementary to the barcoded and common adapters, purified again using QIAquick, and via PicoGreen (Molecular Probes, Carlsbad, CA, USA) (Elshire *et al.* 2011). Samples were run on a 100- bp single-end Illumina HiSeq 2000 lane.

Raw sequence reads were processed using the UNEAK pipeline, an addition to TASSEL 3.0 (Bradbury *et al.* 2007) that used VCFtools v 0.1.11 (Danecek *et al.* 2011). UNEAK retains all reads with a barcode, cut site, and no missing data in the first 64 bp after the barcode. Reads are then grouped into tags by 100% identity, tags are aligned pairwise, and any tag pairs differing by one bp are called as potential SNPs. We

identified 1,528,449 total tags after merging with 639,274 total tag networks identified. To remove sequencing errors, individual samples and SNPs with a high amount of missing data, any alleles represented by fewer than 10% of the mean reads coming for the lane or a frequency of less than 3% were filtered out.

To further remove data artifacts, we filtered out SNPs that deviate from Hardy–Weinberg equilibrium ($p > 0.01$) in at least five populations. We also eliminated minor alleles with a frequency of < 0.05 across the entire dataset. We initially had over 58,756 SNPs for the 95 individuals; after removing individuals with < 10 % of the markers and SNPs with < 5% frequency we retained 18,616 SNPs for 90 individuals. To overcome any bias associated with having missing SNPs, we first generated a data file with no missing information. This file has 90 individuals genotyped at 525 SNPs. In the second dataset we have individuals that have at least 90 % of the SNPs and SNPs present in at least 70 % of the samples. This last dataset has 80 individuals and 10,059 SNPs. The output VCF files were manipulated in TASSEL (Bradbury *et al.* 2007) and PGDSPIDER v.2.0.5.0 (Lischer & Excoffier 2012) to generate the inputs in TreeMix (Pickrell & Pritchard 2012) GENALEX 6.5 (Peakall & Smouse 2006, 2012), GENODIVE 2.0 (Meirmans 2014; Meirmans & Hedrick 2010; Meirmans & Van Tienderen 2004) and STRUCTURE 2.3.2 (Pritchard *et al.* 2000).

Data Analysis

Alignment and Gene Genealogies

To discover geographically restricted monophyletic groups and establish relationships among alleles, we built gene genealogies using both Dloop and COI. To estimate the most likely model of evolution we used jModeltest (Posada & Crandall 2001). Given that both markers are mitochondrial and thus effectively linked, we reconstructed a concatenated gene genealogy via a maximum likelihood approach as implemented in RAxML (Stamatakis 2014). We did two analysis, we first inferred an unrooted tree with all sequences and then a topology including a close outgroup (*O. macrognathus*, NCBI number = JN193388).

Finer Genetic Differentiation

Moritz (1994) suggest ESUs may still exist even in the absence of monophyletic trees. To further delimit ESU in the yellowhead Jawfish, we estimated principal components in GENODIVE 2.0 (Meirmans & Van Tienderen 2004). To gain knowledge on how markers may provide alternative grouping among populations, we estimated PCAs for both microsatellite and SNP data. PCA allows efficient reduction of variation around the axes in our data, large in the case of SNPs, to plot only those reduced vectors (Peakall & Smouse 2006, 2012). We plot those axes in R (Team 2013). Given that one of our aims is to compare across markers, we performed PCA in individuals, so that we were able to quantify variation among individuals within populations when estimating PCA from microsatellites and SNPs.

As an alternative to delineate ESUs using variation in frequencies at alleles, we used association among loci to cluster populations via Bayesian probability as implemented in STRUCTURE 2.3.2 (Pritchard *et al.* 2000). We ran STRUCTURE in the microsatellite and the two SNP data sets. For each partition, we ran STRUCTURE without information of the origin of each individual, thereby reducing potential biases by a *priori* user categorization of samples into populations. We assumed an admixture model with a burn-in of 10% of the iteration steps. For microsatellite data we ran STRUCTURE for at least 10 million iterations and 3 replicates per run. For the SNPs, given the large amount of data, we restricted the runs to only 900,000 steps but increased the replicates to 4. We did not notice extreme variations in likelihood scores across replicates. We ran STRUCTURE assuming at least eight different K values (number of inferred populations) from one to eight and then used CLUMPAK (Kopelman *et al.*), which implements the Evanno method (Evanno *et al.* 2005) to infer the optimal number of populations (K's) in our dataset, to combined replicates and generate figures.

Isolation by Distance (IBD)

To further disentangle patterns of genetic differentiation across geography, we estimated pair-wise genetic differentiation using F_{ST}s as implemented in GENODIVE 2.0 (Meirmans & Van Tienderen 2004) We also calculated pairwise shortest Euclidian nautical distance for each pair of locations using Google Earth 6.2. The resulting pairwise matrix of both genetic and Euclidian distances was then plotted in R (R Development Core Team 2013). We fit a linear regression model to determine significance of the Isolation by Distance at the 0.05. To evaluate deviations from different markers, we estimated IBD in both microsatellite and SNP data. We restricted

the analysis for the populations that overlap in both datasets, which is restricted for the patterns along the Mona Passage. We used Pearson correlations to estimate deviations in the correlations between datasets.

To estimate migration across locations, we used BayesAss using all eleven populations. Given that BayesAss is designed for small datasets (< 300 loci) we ran it on bootstrap replicates from our SNP dataset. For each run we discarded the first 5 million steps and estimated parameters from 100 million steps sampled every 20,000 after this initial burn in. To provide a proper filter after the MCMC run, we adjusted parameters to be within 20% - 69% acceptance rates using information from preliminary runs. We ran three replicates from different seeds for each run. We assessed mixing and convergence with TRACER 1.6 (Rambaut *et al.* 2014). Replicates show similar results; here we present the one with the lowest deviance probability as suggested by (Meirmans 2014).

Results

Our phylogenetic reconstruction using the concatenated COI and Dloop on 269 individuals suggests that most of the genetic variation in the yellowhead jawfish is segregated across geography (Fig. 4.1). Notably, mtDNA variation in Florida is exclusive (monophyletic group – green clade), while in Belize lacks variation among individuals fixed for a single allele that is present at much lower frequencies in other areas in the Caribbean. This clade (orange in Fig. 4.1) has a west to east transition, being common in San Andres intermediate in Puerto Rico and less so in the USVI. The second major feature in our reconstruction is that > 90% of the alleles in Aruba and Curaçao are private (purple clade in Fig. 4.1). This clade is > 3% and 10% divergent, for COI and Dloop, respectively, to alleles elsewhere in the Caribbean.

Our PCA analysis also suggests strong differentiation across the Caribbean (Figs. 4.2). Results from microsatellites and both SNP datasets are similar. Our PCAs have higher resolution, and at least five tight groups can be discerned (Bahamas, Florida, Curaçao, La Caleta and Puerto Rico) and suggest they are independent Evolutionary Significant Units (Fig. 4.2). We found that samples from PNNE, while close to Puerto Rico are quite spread, with high inter-sample variability. Our dataset with > 10k SNPs has by far the highest resolution with reduced within group variation and increase among group distance.

To further describe genetically isolated groups, we used linkage disequilibrium across markers. Our STRUCTURE analysis suggests genetic differentiation at four populations

in the Caribbean. We found that as shown by our phylogenetic reconstruction and the PCA, Florida, Curaçao, Bahamas and Puerto Rico form isolated gene pools (Fig. 4.3). Samples from the Dominican Republic, however, are intermixed as individuals from La Caleta and have ancestries closer to Bahamian samples, while individuals sampled at the PNNE (only 127 km apart from La Caleta) are more similar to Puerto Rican samples with a minor admixture from the Bahamas. The five sampling sites in Puerto Rico exhibit a tight genetic group (sky blue in Fig. 4.3).

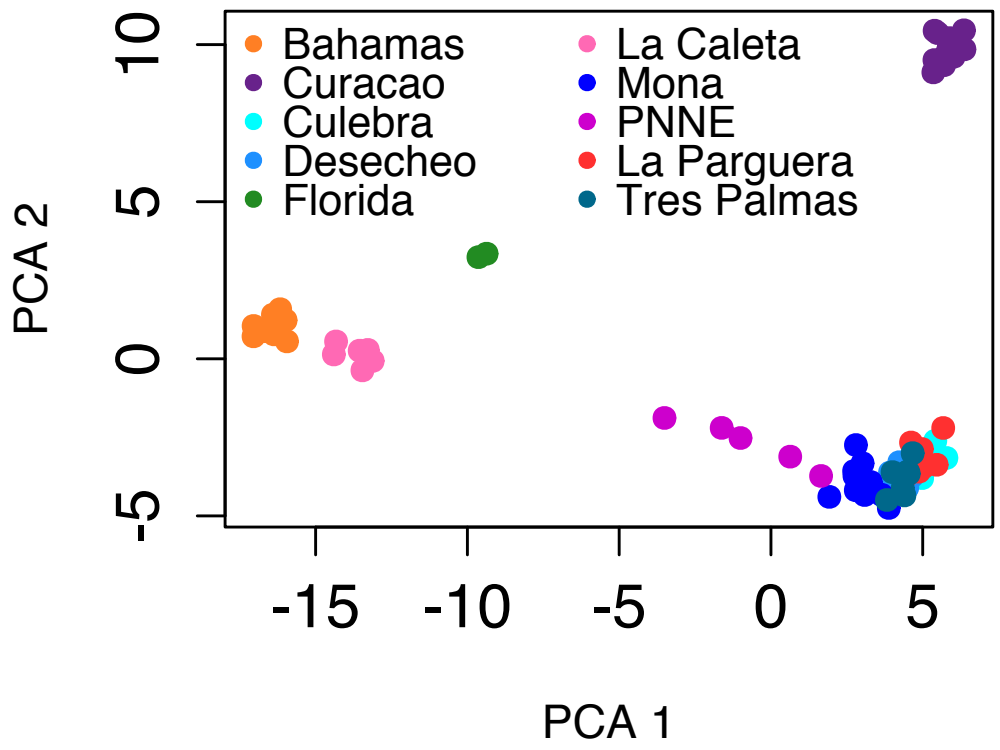

Figure 4.2. Principal Component Analysis (PCA) from all individuals at eleven sampling sites along Caribbean Region with 10,059 SNPs.

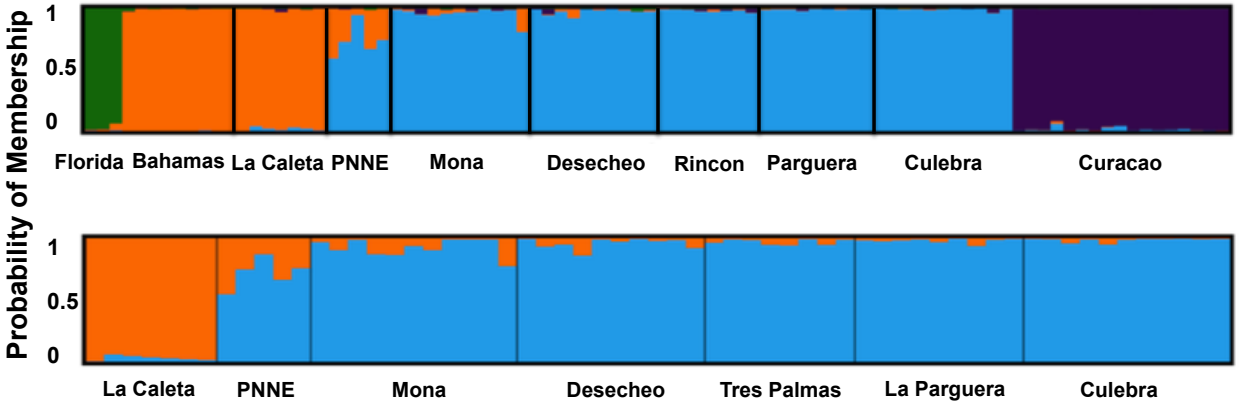

Figure 4.3. Graphical summary of clustering analysis for *Opistognathus aurifrons* genotypes from eleven areas along Caribbean using Structure. Probability of Membership of each sampling site (n=11) from individuals per site scored with 525 markers. Each vertical line represents an individual and the estimated proportion of the individual's genome from each inferred cluster.

Our Isolation-by-Distance (IBD) pattern of geographical differentiation based on pairwise F_{ST}'s also suggest a large role of geography in driving much of the genetic variation (~30%) coincident for both microsatellite and SNP markers ($p < 0.01$ in both cases) (Fig. 4.4). Patterns for the microsatellites at this smaller scale are similar, with geography explaining 15% of the genetic variation, yet the slopes from the two markers are different. For SNP data the slopes are $6\ e^{-5}$ and $8\ e^{-5}$, but only $2\ e^{-5}$ for microsatellites (Figure 4.4).

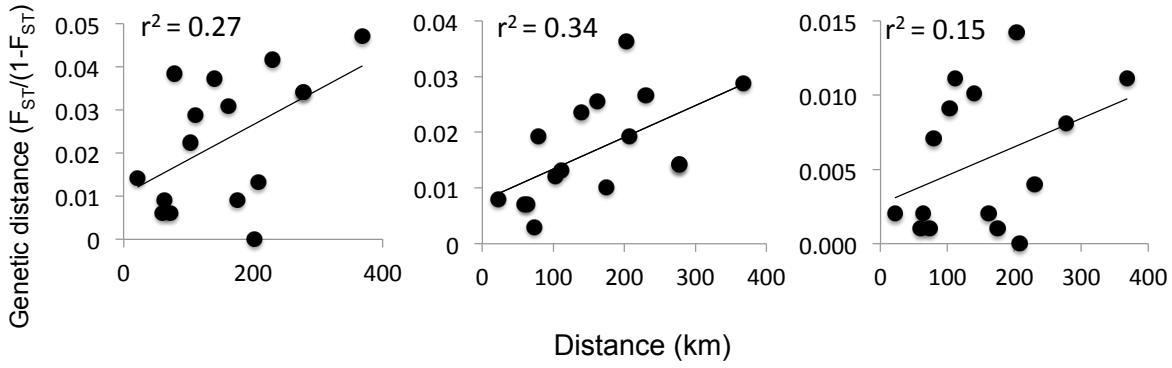

Figure 4.4. Isolation by Distance (IBD) in *O. aurifrons* from six sampling sites along Mona Passage (PNNE, Mona, Desecheo, Tres Palmas, La Paguera and Culebra). IBD from 10,059 SNPs (left); IDB from 525 SNPs (center) and IDB from 12 microsatellites (right).

At finer scales, our pairwise F_{ST} estimates from SNP data suggest population differentiation at all levels (except between Desecheo - Tres Palmas and Mona - Tres Palmas). The change in allele frequencies detected by the pairwise F_{ST} approach suggest high geographic structure even at scales < 100 km. Our comparison of the genetic differences via F_{ST}s between microsatellites and those generated by SNPs, suggest SNP data provides higher resolution. For instance, the power to detect significance of pairwise F_{ST}s from microsatellites disappeared in six of the population pair-wise comparisons (particularly comparisons between nearby populations; *e.g.,* Parguera - Tres Palmas), while it was completely captured by the SNP data (Table 4.1).

To study patterns of gene flow between pairs of populations across the Caribbean, we used the Bayesian MCMC statistical approach implemented in BayesAss 3 (Wilson & Rannala 2003). Using our bootstrapping procedure with the SNP data, we found little genetic exchange among populations in the Caribbean, even populations

geographically close such as those around Puerto Rico. This is in stark contrast to our earlier estimates from the microsatellite data (Chapter 3), where we found that at least 20% of the individuals from Mona and PNNE come from migrants from nearby locations.

Table 4.1. Diference in spatial genetic structure for 10059 Snps along Mona Passage polulations. Pairwise comparison along Mona Passage (p-values). Bonferroni correction = 0.05/21= 0.023. In red the values with signifficant gentic structure p-value<0.0023.

F_{ST} Mona Passage Data SNPs

	Culebra	Desecheo	La Caleta	Mona	PNNE	La Parguera	Tres Palmas
Culebra	0	-0.008	0.23	0.016	0.015	-0.008	-0.04
Desecheo	**-0.008**	0	0.208	-0.019	-0.008	0.014	-0.058
La Caleta	**0.23**	**0.208**	0	0.206	0.138	0.254	0.171
Mona	**0.016**	**-0.019**	**0.206**	0	0.012	0.024	-0.026
PNNE	**0.015**	**-0.008**	**0.138**	**0.012**	0	0.042	-0.034
La Parguera	**-0.008**	**0.014**	**0.254**	**0.024**	**0.042**	0	-0.027
Tres Palmas	**-0.04**	-0.058	**0.171**	-0.026	**-0.034**	**-0.027**	0

F'st Mona Passage Data _12 microsatellites

	Culebra	Desecheo	La Caleta	Mona	PNNE	La Parguera	Tres Palmas
Culebra	0	-0.043	0.242	0.022	0.033	-0.043	-0.066
Desecheo	-0.043	0	0.221	-0.023	0.033	0.006	-0.078
La Caleta	**0.242**	**0.221**	0	0.299	0.049	0.301	0.199
Mona	**0.022**	-0.023	**0.299**	0	0.051	0.064	-0.021
PNNE	**0.033**	**0.033**	**0.049**	**0.051**	0	0.087	-0.017
La Parguera	-0.043	**0.006**	**0.301**	**0.064**	**0.087**	0	-0.045
Tres Palmas	-0.066	-0.078	**0.199**	-0.021	-0.017	-0.045	0

Discussion

Our results suggest genetic segregation across geography in the yellowhead jawfish, with each sampled population showing private alleles and or drastic allele frequency changes at both the nuclear and mitochondrial genomes. The isolated populations have experienced minimal contemporary genetic exchange. The species is also characterized by abundant genetic variation, which prevents hypervariable markers such as microsatellites to achieve enough power to detect finer (< 50 km) differentiation. Phylogenetic, allele frequency variation and PCA analyses unambiguously suggest the presence of at least four Evolutionary Significant Units: one restricted to the southern Caribbean in Aruba and Curaçao, one in the northwest that forms a monophyletic group in Florida and two in the mid-Caribbean - one in The Bahamas that meets a second one mostly found in Puerto Rico. The high degree of segregation across the Caribbean suggests that incorporating this evolutionary information into conservation, routinely done in terrestrial vertebrates (Shaffer *et al.* 2015), may enhance the likelihood of preserving reef fish biodiversity either by invoking a species or location-focused management. Below we highlight the historical biogeographic evidence that supports the yellowhead jawfish ESU division, the importance of incorporating ESUs into coral reef conservation and the relative contribution of small benthic fish to reef fish biodiversity.

Biogeography and Evolutionary Significant U**nits**

Biogeograhically, the Greater Caribbean divides into three regions: A northern area that covers the Gulf of Mexico and Florida with the presence of temperate fauna; a Central

area that includes the West-Antilles, Central America and Bermuda and a third region in northern South America (Briggs & Bowen 2013; Robertson & Cramer 2014). Our genetic data recapitulates these biogeographical divisions with, for example, Florida monophyletic and completely isolated from nearby populations such as The Bahamas or Belize. The fixed mtDNA differences are concordat with our nuclear PCA structure derived from more than 10,000 SNPs (green color on Fig. 4.2), and our Treemix inference (Fig. 4.5). The Florida lineage is constrained to the Northern biogeographical region while Belize and The Bahamas are in the central region.

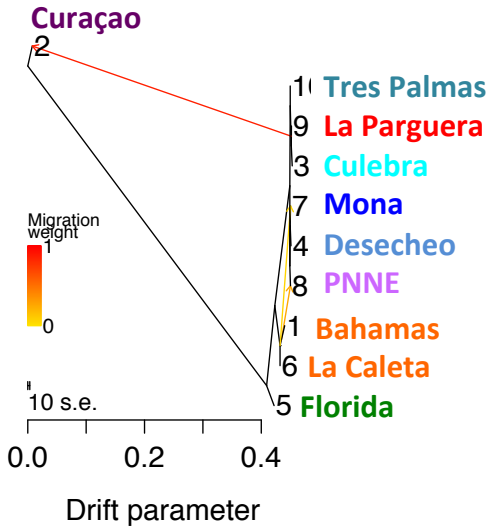

Figure 4.5. Relationships among populations based on variation on alellic frequencies at 10,059 SNPs with 500 replicates. Red and yellow connections depict possible migrations between pairs of populations.

The genetic segregation between Florida and The Bahamas occurs in just under 100 km. Such differentiation is not restricted to *O. aurifrons* but is shared by a larger number of fish species that generate unique ichthyological communities in each location (Böhlke & Robins 1959; Briggs 1974). The division of these biogeographical provinces is likely

maintained by strong poor-mixing currents across Florida (Briggs 1974). The Florida Current is a fast flowing and laminar current with little mixing (Briggs 1974). For example, Caribbean water, well defined by its salinity, flanks the Florida Current on its east side from the Bahamian archipelago. On the west side it is flanked by the also well-characterized Continental/Gulf water. The two clearly different water masses border the Florida current without mixing, suggesting planktonic larvae may not cross these three layers to go from Florida to The Bahamas or vice versa (Briggs 1974). An additional aspect that separates The Bahamian and Floridian faunas is that each has had incidental colonizations from different sources. In Florida populations and species have come from western Cuba through the Yucatán Channel, while Bahamian inhabitants came from stepping stone dispersal more directly from The Dominican Republic and eastern Cuba (Voss & Voss 1960). The Bahamas-Florida division is maintained by a strong laminar poor mixing current that generates a significant barrier to adult and larval fish dispersal and unique historical colonization events.

A second sharp genetic break in *O. aurifrons* occurs between Curaçao and Aruba and the rest of the Caribbean, which recapitulates the separation between the southern and central Caribbean regions (Robertson & Cramer 2014). In Curacao and Aruba > 90 % of the mtDNA alleles are private and no admixed individuals exist with Puerto Rican populations (Figs 4.2 and 4.3). The unique biotas in the Southern region are due to both historical colonization and little connectivity due to the large separation of deep ocean between the southern region and the major Antilles. Historically the southern end was colonized primarily by species coming from the eastern Pacific, most of which then

spread across the Caribbean. Not all species spread and some were retained, likely including divergent lineages within species. A result of this differential dispersion is the lower species richness in both the Central and northern regions compared to the south (Ludt & Rocha 2015). In *O. auriferous*, the major Curaçao COI clade is ~ 3% divergent from the more broadly distributed central lineage, separated for at least 1.5 million years assuming a mutation rate of 2% per million years (Bermingham *et al.* 1997). The Curaçao lineage has thus been isolated since geological times and may represent one of the lineages that came from the eastern Pacific and was retained in the southern Caribbean while its conspecifics spread through the Caribbean forming geographically isolated groups.

A third group revealed by our genetic data is a division in the central region, with a continental population around Belize that extends through The Bahamas and reaches the Dominican Republic and a second more insular one around Puerto Rico. Patterns of surface currents suggest that these two clades are likely separated by mesoscale gyres that form along the Mesoamerican area, decreasing larval mixing with the rest of the Caribbean and leading to genetic divergence among populations (Cowen *et al.* 2006). In fact, genetic data from both fish and invertebrates suggest this is a common pattern. Both *Hypoplectus* and *Elacanthinus* fishes form isolated populations around Belize (Puebla *et al.* 2008; Taylor & Hellberg 2003) and damselfishes segregate between Belize and Panama (Salas *et al.* 2010). In reef forming corals, Mesoamerican populations are segregated to insular central populations such Puerto Rico (Foster NL *et al.* 2012).

In addition to genetic variation, *O. aurifrons* also harbors rich phenotypic diversity across geography. Bohlke & Thomas 1961; Ho *et al.* 2012, Shows that morphological and melanistic patterns vary across the Caribbean with Bahamas populations heavily patterned and well-differentiated from those in Florida, which lack melanistic traits. Color patterns in St. Thomas- USVI are intermediate and though differentiated are more similar to fish from Aruba. The color from Aruba is intermediated to the ones in the USVI and The Bahamas. Both genetic and phenotypic diversity vary across locations, and each population is likely under unique demographic dynamics. Management strategies will benefit by incorporating these differences into their conservation plans by accounting for these geographically-based demographic differences.

Why should we incorporate ESU into marine conservation?

In the past two decades conservation genetics and genomics have greatly improved conservation of vertebrate species, though more common in mammals, it has been most useful for cryptic small taxa such frogs and lizards (Shaffer *et al.* 2015). Genetics has in general reinforced our previous understanding of biodiversity but has uniquely contributed to our knowledge of the overwhelming cryptic diversity. It has thus helped identify appropriate taxonomic and population units for protection and management, such as Evolutionary Significant Units (Allendorf *et al.* 2010). Genetics has also been useful to establish conservation priorities by informing on locations of hidden barriers to migration and corridors to maintain connected populations. Based on the multiple criteria proposed by (Crandalla *et al.* 2000; Moritz 1994; Waples 1991) the four

yellowhead jawfish lineages should be considered ESUs. The addition of ESU information could help conservation of small cryptic reef fishes in the Caribbean in at least three ways.

First it provides proper bounds on the actual populations. If we considered the yellowhead jawfish as a single interbreeding population we are likely overestimating the genetic and phenotypic diversity. The emphasis is different from a management perspective because we would change from managing a cosmopolitan species able to explore different geographical regions to a complex of at least four lineages, each geographically specialize and locally adapted that do not exchange genetic material with unique demographic dynamics. In the latter case re-introducing individuals from different ESUs would result in restoration failures because local fish would be unlikely to breed with introduced fish and phenotypic variation may also not match local requirements. Dividing the population into four different ESUs also emphasizes that each unit is more fragile given its smaller effective population size, which makes them more vulnerable to stochastic processes and extinction. Patterns found in cryptic benthic fish already suggest that geographically separated populations form Evolutionary Significant Units and that each geographical locale has a slightly varied version of the Caribbean cryptic fish diversity. Each ESU thus has individual evolutionary potential to, for example, respond to climate change, and thus should be individually preserved.

The second aspect is incorporating the historical perspective into management. Our results suggest that the populations in Curaçao seem ancestral to all others and have captured the deepest points of divergence. When interpreted in the light of the biogeographic patterns proposed by Ludt, 2014 it suggests that the southern region is the origin of this species and that it harbors this ancient diversity. Our data also suggest that Florida has lower genetic diversity, nested within the diversity found in the central region. If measures of phylogenetic diversity were used to prioritize conservation in *O. aurifrons*, the southern region would certainly rank first, with less priority given for Florida.

The third aspect is the identification of patterns of migrations and connectivity across the Caribbean. In our study, while there is strong genetic differentiation, the central region shows some degree of connectivity making it a particularly attractive area if the overall goal is to prioritize connectivity via stepping stone dispersal. If viewed in combination with other fish genetic studies, the central area likely connects or in geological times connected Puerto Rico with Belize and Bahamas through the Dominican Republic and Cuba (Eytan & Hellberg 2010; Puebla *et al.* 2008; Taylor & Hellberg 2006). Thus the central region is a geographical extent with the highest number of individuals in a somewhat cohesive unit. An in-depth study of the contemporary and historical migrations as well as the roles of genetic drift and natural selection would help to further refine these patterns.

Cryptic benthic fish diversity is vast and functionally critical on reefs

Small cryptic fish with demersal eggs contributes disproportionally to coral reef fish biodiversity. Gobiidae for example is the most speciose family of fishes, with over 91 species on Caribbean reefs (Floeter *et al.* 2008). We counted the number of species of Caribbean reef fishes and found that fishes with demersal benthic eggs, usually < 15 cm in length contribute up to 40 % of the total Caribbean reef fish biodiversity. The average species richness is 15.5 species per family, yet 11 of the 13 families with small cryptic fishes exceed this average (Fig. 4.6). Small cryptic taxa also often have tight interactions with bigger reef dwellers. For example, benthic cryptic fishes engage in symbiotic interactions with corals to clear up competing macroalgae, thus boosting coral survivorship and recruitment (Dixson & Hay 2012). Cryptic ichthyofauna also maintains the lower trophic levels of the food-web. Together they circulate over 25% of the total amount of carbon and energy in reef systems and provide unique trophic links between smaller invertebrates and detritus and larger carnivorous fish (Smith-Vaniz *et al.* 2006). Part of this circulation is related to their mutualistic role with bigger fish to clean them off parasites, enhancing the latter's survival (Cheney 2009). The fast generation time in these small fishes also promotes rapid energy movement in reef systems (Allen *et al.* 1992).

Cryptobenthic brooding fish with small bodies contribute to the highest (> 80 %) number of Caribbean endemics (Floeter *et al.* 2008; Rocha 2003). Our study suggests that cryptic fish also contribute to subtle variations in fauna within species in the Caribbean. Quantifying these differences and incorporating them into the management plans of marine reserves would ensure the maintenance of biodiversity on reefs.

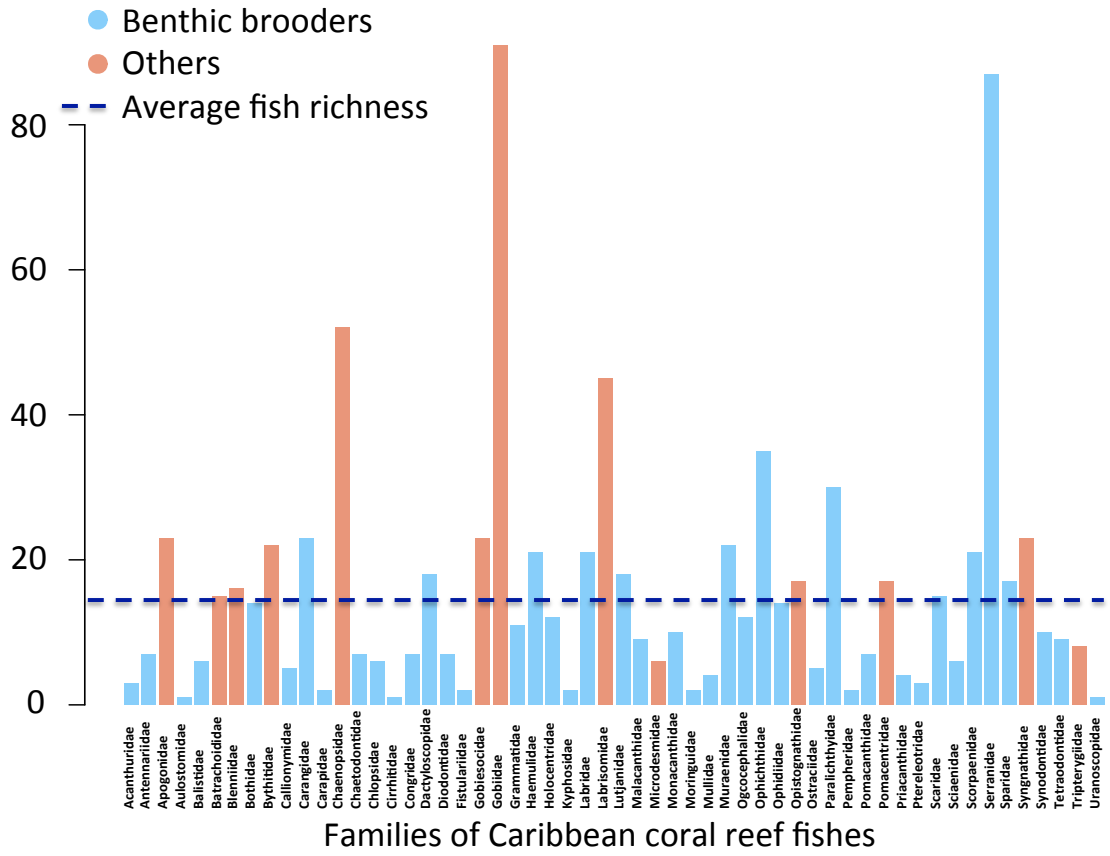

Figure 4.6. Number of species per Caribbean reef fish families.

Literature Cited

Allen LG, Bouvier LS, Jensen RE (1992) Abundance, diversity, and seasonality of cryptic fishes and their contribution to a temperate reef fish assemblage off Santa Catalina Island, California. *Bulletin of the Southern California Academy of Sciences* **9**, 55-69.

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. *Nature Reviews* **11**, 697-709.

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

Ayre DJ, Hughes TP (2000) Genotypic diversity and gene flow in brooding and spawning corals along the Great Barrier Reef, Australia. *Evolution* **54**, 1590-1605.

Baldwin CC, Weigt LA, Smith DG, Mounts JH (2009) Reconciling Genetic Lineages with Species in Western Atlantic Coryphopterus (Teleostei: Gobiidae). *Smithsonian Contributions to the Marine Sciences*, 112-138.

Baums IB, Miller MW, Hellberg ME (2005) Regionally isolated population of an imperiled Caribbean coral, *Acropora palmata*. *Molecular Ecology* **14**, 1377-1390.

Baums IB, Paris CB (2006) A bio-oceanographic filter to larval dispersal in a reef-building coral. *Limnology and Ocenography* **51**, 1969-1981.

Beerli P, Palczewski (2010) Unified Framework to Evaluate Panmixia and Migration Direction Among Multiple Sampling Locations. *Genetics* **185**, 313--326.

Bejarano S, Lohr K, Hamilton S, Manfrino C (2015) Relationships of invasive lionfish with topographic complexity, groupers, and native prey fishes in Little Cayman. *Marine Biology* **162**, 253–266.

Beltran D, Schizas NV, Prada C, Appeldoorn RS (2014) Isolation and characterization of 12 microsatellite loci to study connectivity in the yellow jawfish *Opistognathus aurifrons*. *Conservation Genetic Resources*.

Bermingham E, McCafferty SS, Martin AP (1997) Fish biogeography and molecular clocks: perspectives from the Panamanian Isthmus. In: *Molecular Systematics of Fishes* (ed. TD Kocher CS), pp. 113–128. Academic, San Diego.

Böhlke JE, Robins CR (1959) Studies on Fishes of the Family Ophidiidae: II. Three New Species from the Bahamas. *Proceedings of the Academy of Natural Sciences of Philadelphia* **111**, 37-52.

Bradbury IR, Hubert S, Higgins B*, et al.* (2012) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications*, 450-461.

Bradbury PJ, Zhang Z, Kroon DE*, et al.* (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633-2635.

Briggs JC (1974) *Marine Zoogeography* McGraw-Hill, New York.

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

Brumfield R, Nickerson D, Beerli P, Edwards SV (2003) The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* **18**, 249-256.

Burke L, Maidens J (2004) *Reefs at Risk in the Caribbean.* World Resources Institute (WRI), Washington, DC. Online at: http://www.wri.org/publication/

Cheney KL (2009) Blue and Yellow Signal Cleaning Behavior in Coral Reef Fishes. *Current Biology* **19**, 1283–1287.

Colin PL (1971) Interspecific relationship of the yellowhead jawfish, *Opistognathus aurifrons* (Pisces, Opistognathidae). *Copeia* **1971**, 469-473.

Colin PL (1973) Burrowing behaviour of the yellowhead jawfish, *Opistognathus aurifrons*. *Copeia*, 84-90.

Colin PL (1975) *Neon Gobies.* T.F.H. Publications, Inc, Neptune City, New Jersey.

Cowen RK, Paris CB, Srinivasan A (2006) Scaling of connectivity in marine populations. *Science* **311**, 522-527.

Cowen RK, Sponaugle S (2009) Larval Dispersal and Marine Population Connectivity. *Annual Review of Marine Science* **1**, 443–466.

Crandalla KA, Bininda-Emondsb ORP, Macec GM, Wayned RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* **15**, 290-295.

D'Aloiaa CC, Bogdanowiczb SM, Francisc RK*, et al.* (2015) Patterns, causes, and consequences of marine larval dispersal. *Proceedings of the Natural Academy of Sciences* **45**, 1-6.

Danecek P, Auton A, Abecasis G*, et al.* (2011) The Variant Call Format and VCF tools. *Bioinformatics* **27**, 2156-2158.

De'ath G, Fabricius K, Sweatman H, Puotinen M (2012) The 27-year decline of coral cover on the Great Barrier Reef nd its causes. *Proceedings of the Natural Academy of Sciences* **109**, 17995-17999.

Dennis GD, Smith-Vaniz WF, Colin PL, Hensley DA, McGehee MA (2005) Shore Fishes from Islands of the Mona Passage, Greater Antilles with Comments on Their Zoogeography. *Caribbean Journal of Science* **41**, 716-743.

Dixson DL, Hay ME (2012) Corals Chemically Cue Mutualistic Fishes to Remove Competing Seaweeds. *Science* **338**, 804-807.

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

Elshire RJ, Glaubitz JC, Sun Qi*, et al.* (2011) A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *Plos One* **6**, 1-10.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.

Eytan RI, Hellberg ME (2010) Nuclear and Mitochondrial Sequence Data reveal and Conceal Different Demographic Histories and Population Genetic Processes in Caribbean Reef Fishes. *Evolution* **64**, 3380-3397.

Felsenstein J (2006) Accuracy of Coalescent Likelihood Estimates: Do We Need More Sites, more Sequences, or More Loci? *Molecular Biology and Evolution* **23**, 691-700.

Floeter SR, Rocha LA, Robertson DR*, et al.* (2008) Atlantic reef fish biogeography and evolution. *Journal of Biogeography* **35**, 22–47.

Foster NL, Paris CB, Kool JT*, et al.* (2012) Connectivity of Caribbean coral populations: complementary insights from empirical and modelled gene flow. *Molecular Ecology* **21**, 1143–1157.

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

Garner TA, Cote IM, Gill JA, Grant A, Watkinson AR (2003) Long-term region-wide declines in Caribbean corals. *Science* **301**, 958–960.

Grosberg RK, Cunningham CW (2001) Genetic structure in the sea: from populations to communities. In: *Marine Community Ecology* (ed. M. D. Bertness SG, and M. E. Hay), pp. 61-84. Sinauer Associates, Sunderland, MA.

Gutiérrez-Rodríguez C, Lasker HR (2004) Microsatellite variation reveals high levels of genetic variability and population structure in the gorgonian coral Pseudopterogorgia elisabethae across the Bahamas. *Molecular Ecology* **13**, 2211-2221.

Heck KL, McCoy ED (1978) Long distance dispersal and the reef building corals on the Eastern Pacific. *Marine Biology* **48**, 349-356.

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

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

Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science* **70**, 273-290.

Hess HC (1993) Male mouthbrooding in jawfishes (Opistognathidae): constraints on polygyny. *Bulletin of Marine Science* **52**, 806-818.

Hey J (2010) Isolation with Migration Models for More Than Two Populations. *Molecular Biology and Evolution* **27**, 905-920.

Hill WG (1981) Genetics Research. *Estimation of effective population size from data on linkage disequilibrium* **38**, 209–216.

Ho ALFC, Pruett CL, Lin J (2012) Population genetic structure, coloration, and morphometrics of yellowhead jawfish *Opistognathus aurifrons* (Perciformes: Opistognathidae) in the Caribbean region. *Marine Ecology Progress Series* **444**, 275-287.

Hobana S, Arntzenb JW, Bertorellea G*, et al.* (2013) Conservation Genetic Resources for Effective Species Survival (ConGRESS): Bridging the divide between conservation research and practice. *Journal for Nature Conservation* **21**, 433-437.

Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) BARCODING Universal primer cocktails for fish DNA barcoding. *Molecular Ecology Notes* **7**, 544–548.

Jackson JBC, Donovan MK, Cramer KL, Lam VV (2014) Status and Trends of Caribbean Coral Reefs: 1970-2012 (ed. Global Coral Reef Monitoring Network I, Gland, Switzerland).

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

Jones GP, Almany GR, Russ GR*, et al.* (2009) Larval retention and connectivity among populations of corals reef fishes: advances and challenges. *Coral Reefs* **28**, 307-325.

Jones GP, Srinivasan M, Almany GR (2007) Population connectivity and conservation of marine biodiversity. *Oceanography* **20**, 42-53.

Kearse M, Moir R, Wilson A*, et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649.

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

Kuhner MK, Beerli P, Yamato J, Felsenstein J (2000) Usefulness of Single Nucleotide Polymorphism Data for Estimating Population Parameters. *Genetics* **156**.

LeGore RS, Hardin MP, García-Sais JR (2006) Puerto Rico Marine Ornamental Fishery Evaluation Phase II: Wild Population Assessments. 1-63.

Lessios HA, Kessing BD, Robertson DR (1998) Massive gene flow across the world's most potent marine biogeographic barrier. *Proceedings of the Royal Society B* **265**, 583-588.

Lessios HA, Robertson DR (2006) Crossing the impassable: genetic connections in 20 reef fishes across the Eastern Pacific Barrier. *Proceedings of the Royal Society, London series* **B**, 2201–2208.

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

Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* **28**, 298-299.

Ludt WB, Rocha LA (2015) Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography* **42**, 25-38.

McFadden CS, Grosberg RK, Cameron BB, Karlton DP, Secord D (1997) Genetic relationships within and between solitary and clonal forms of the sea anemone Anthopleura elegantissima revisited: Evidence for the existence of two species. *Marine Biology* **128**, 127-139.

Mege P, Schizas NV, Garcia J, Hrbek T (2014) Genetic seascape of the threatened Caribbean elkhorn coral, Acropora palmata, on the Puerto Rico Shelf. *Marine Ecology Progress Series* **36**, 195-209.

Meirmans PG (2014) Nonconvergence in Bayesian estimation of migration rates. *Molecular Ecology Resources* **14**, 726–733.

Meirmans PG, Hedrick PW (2010) Assessing population structure: FST and related measures. *Molecular Ecology* **11**, 5–18.

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

Miller MR, Atwood TS, Eames BF*, et al.* (2007) RAD marker microarrays enable rapid mapping of zebrafish mutations. *Genome Biology* **8**, 105.

Mora C (2008) A clear human footprint in the coral reefs of the Caribbean. *Proceedings of the Royal Society B* **275**, 767-773.

Moritz C (1994) Defining Evolutionary Significant Units for Conservation. *Tree* **9**, 373-375.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**, 2841-2847.

Paris CB, Cowen RK, Claro R, Lindeman KC (2005) Larval transport pathways from Cuban snapper (Lutjanidae) spawning aggregations based on biophysical modeling. *Marine Ecology Progress Series* **296**, 93-106.

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

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28**, 2537-2539.

Pickrell JK, Pritchard JK (2012) Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *Plos Genetics* **8**, 1-17.

Pinsky ML, Montes HR, Palumbi SR (2010) Using Isolation by Distance and effective Density to Estimate Dispersal Scales in Anemonefish. *Evolution* **64**, 2688–2700.

Posada D, Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. *Systematic Biology* **50**, 580–601.

Prada C, Hellberg ME (2014) Strong natural selection on juveniles maintains a narrow adult hybrid zone in a marine broadcast spawner. *American Naturalist* **184**, 702-713.

Pritchard JK, Stephens M., Donnelly P (2000) Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155**, 945-959.

Puebla O, Bermingham E, Guichard F (2008) Population genetic analyses of *Hypoplectrus* coral reef fishes provide evidence that local processes are

operating during the early stages of marine adaptive radiations. *Molecular Ecology* **17**, 1405–1415.

Puebla O, Bermingham E, Guichard F (2009) Estimating dispersal from genetic isolation by distance in a coral reef fish (*Hypoplectrus puella*). *Ecology* **90**, 3087–3098.

Puebla O, Bermingham E, Mcmillan WO (2012) On the spatial scale of dispersal in coral reef fishes. *Molecular Ecology* **21**, 5675–5688.

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

Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) Tracer v1.6. *Available from* http://beast.bio.ed.ac.uk/Tracer.

Roberts CM (1997) Connectivity and Management of Caribbean coral reefs. *Science* **278**, 1454-1457.

Roberts CM, Polunin NVC (1991) Are marine reserves effective in managment of the reef fisheries? *Reviews in Fish Biology and Fisheries* **1**, 65-91.

Robertson DR, Cramer KL (2014) Defining and Dividing the Greater Caribbean: Insights from the Biogeography of Shorefishes. *Plos One* **9**, e102918.

Rocha LA (2003) Patterns of distribution and processes of speciation in Brazilian reef fishes. *Journal of Biogeography* **30**, 1161-1171.

Rosenberg NA (2004) Distruct: a program for the graphical display of population structure *Molecular Ecology Notes* **4**, 137-138.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219–1228.

Salas E, Molina-Ureña H, WalterRyan P, D. HD (2010) Local and regional genetic connectivity in a Caribbean coral reef Fish. *Marine Biology* **157**, 437–445.

Sale PF, Cowen RK, Danilowicz BS*, et al.* (2005) Critical science gaps impide use of no-take fishery reserves. *Trends in Ecology & Evolution* **20**, 74-80.

Serrano X, Baums IB, O'Reilly K*, et al.* (2014) Geographic differences in vertical connectivity in the Caribbean coral Montastraea cavernosa despite high levels of horizontal connectivity at shallow depths. *Molecular Ecology* **23**, 4226–4240.

Shaffer HB, Gidiş M, McCartney-Melstad E*, et al.* (2015) Conservation genetics and genomics of amphibians and reptiles. Annual Reviews of Animal *Biosciences* **3**, 113-138.

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

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

Slatkin M (1993) Isolation by Distance in Equilibrium and Non-equilibrium populations. *Evolution* **47**, 264-279.

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

Smith-Vaniz WF, Jelks HL, Rocha LA (2006) Relevance of cryptic fishes in biodiversity assessments: a case study at Buck Island Reef National Monument, St. Croix. *Bulletin of Marine Science* **79**, 17-48.

Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, 1-2.

Steneck RS, Paris CB, Arnold SN*, et al.* (2009) Thinking and managing outside the box: coalescing connectivity networks to build region-wide resilience in coral reef ecosystems. *Coral Reefs* **28**, 367-378.

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

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

Taylor MS, Hellberg ME (2006) Comparative phylogeography in a genus of coral reef fishes: biogeographic and genetic concordance in the Caribbean. *Molecular Ecology* **15**, 695-707.

Team RDC (2013) R: A language and environment for statistical computing R Foundation for Statistical Computing. Vienna, Austria.

Thorrold SR, Jones GP, Planes S, Hare JA (2006) Transgenetional marking of embrionic otoliths in marine fishes using barium stable isotopes. *Canadian Journal of Fisheries and Aquatic Sciences* **63**, 1193-1197.

Underwood JN, Smith LD, Van Oppen MJH, Gilmour JP (2007) Multiple scales of genetic connectivity in a brooding coral on isoled reefs following catastrophic bleaching. *Molecular Ecology* **16**, 771-784.

van der Meer MH, Hobbs J-PA, Jones GP, van Herwerden L (2012) Genetic Connectivity among and Self-Replenishment within Island Populations of a Restricted Range Subtropical Reef Fish. *Plos One* **7**

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

Victor B (2015) How many coral reef fish species are there? Cryptic diversity and the new molecular taxonomy. In: *Ecology of Fishes on Coral Reefs* (ed. Mora C), pp. 76-87. Cambridge University Press, Cambridge.

Vollmer SV, Palumbi SR (2007) Restricted gene flow in the Caribbean staghorn coral Acropora cervicornis: Implications for the recovery of endangered reefs. *Journal of Heredity* **98**, 40-50.

Voss GL, Voss NA (1960) An Ecological Survey of the Marine Invertebrates of Bimini, Bahamas, with a Consideration of their Zoogeographical Relationships. *Bulletin of Marine Science* **10**, 96-116.

Waples RS (1991) Pacific Salmon, *Oncorhynchus spp*., and the Definition of "Species" Under the endangered Species Act. *Marine Fisheries Review* **53**, 11-22.

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

Warner RR (1997) Evolutionary ecology: how to reconcile pelagic dispersal with local adaptation. *Coral Reefs* **16S**, 115-128.

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

Young FA (1982) The yellowhead jawfish: breeding the marine mouthbrooder in captivity. *Freshwater and Marine Aquarium* **4**, 50−51.

Appendix A

Supplementary Information for Chapter 3

Table S3.1. Characterization of 18 genetic markers: Two mitochondrial genes; five nuclear genes, and twelve microsatellites.

	Name	Marker	Primer sequence 5' – 3'	Size	Ta (°C)	From
1.	D-loop	L15995 H16498	AATTCTCACCCCTAGCTCCCAAAG CCTGAAGTAGGAACCAGATG	223	52	Ivannova, 2007
2.	COI	FISH-BCL FISH-BCH	TCAACYAATCAYAAAGATATYGGCAC TAAACTTCAGGGTGACCAAAAAATCA	603	50	Baldwine, et al., 2009
3.	ATROP	A-TROPL A-TROPH	GAGTTGGA TCGCGCTCAGGAGCG CGGTCAGCCTCCTCAGCAATGTGCTT	332	60	Hickerson, and Cunningham, 2005
4.	RAG1	RAG1F1 RAG1R1	CTG AGC TGC AGT CAG TAC CAT AAG ATG T CTG AGT CCT TGT GAG CTT CCA TRA AYT T	354	58	Lí and Ortí, 2004
5.		1777E4F 1777E4R	AGGAGYTGGTGAACCAGAGCAAAGC AGATCRGCCTGAATSAGCCAGTT	153	60	Lí et al., 2010
6.		4174E20F 4174E20R	CTYTCGCTGGCTTTGTCTCAAATCA CTTTTACCATCKCCACTRAAATCCAC	365	58	Lí et al., 2010
7.	RH2a-Opsin	Gen 20F Gen 20F	TATGGTGAAACCGAACATGC CAAAAAGGATGGTGGACAGG	659	58	This study
8.	785	HLUJHHW01BHFL5	F: CAGAGGCTTGCCTTGAAGTG R: CGTTCGCTGCAGGTCATAC	186	59 y 53	Beltrán et al., 2015
9.	1588	HLUJHHW01CE13G	F: GAAAGAGAAGTTGCCGCCTC R: CTCCCATTCCTGACTCACCC	264	59 y 53	Beltrán et al., 2015
10.	1438	HLUJHHW01BM47D	F: CTTAGTTGGGATTGCGTGGC R: GGCCTCAGGAATTTCATCGC	224	59 y 53	Beltrán et al., 2015
11.	8462	HLUJHHW01CDWXY	F: TGTTGTTGTTTCTCCGGCAC R: CTATTCTTTGGACACCGCGG	248	59 y 53	Beltrán et al., 2015
12.	325	HLUJHHW01CP5G8	F: TTGGGTTGCAACTCTGTGTG R: TGCCATCTGTGTCCATTGTG	203	59 y 53	Beltrán et al., 2015
13.	170	HLUJHHW01DLIPU	F: TGACATCCACCACTGACAGG R: TATCGGCTGGTCCTTTCTGC	260	59 y 53	Beltrán et al., 2015
14.	516	HLUJHHW01EM6DJ	F: TCTCCTCAGCCACCAAGAAG R: GGGTATCAGCACTGTTGTCC	150	59 y 53	Beltrán et al., 2015
15.	150	HLUJHHW01CL23R	F: GCGGCACACCTCTATTAAGC R: TCACGCAAACAGATGATAGCC	296	59 y 53	Beltrán et al., 2015
16.	246	HLUJHHW01DN5M	F: GCAGCACGATCGAGAAACTG R: CAGATGGCCTCGTCAAACAC	289	56 y 53	Beltrán et al., 2015
17.	8537	HLUJHHW01EZ6IS	F: CTGAAACTTCCCAACCAGCC R: CCTCGATGCTGCTTGATGTG	287	59 y 53	Beltrán et al., 2015
18.	1358	HLUJHHW01D9S09	F: ACTCGACCCATGTTTCATCATC R: ACATCCACAGTTGTCACTTGC	247	59 y 53	Beltrán et al., 2015
19.	7983	HLUJHHW01C84TE	F: CGGTATAGTGTGGGAGGGTC R: AAACTGGGATTGATGCGTGG	295	59 y 53	Beltrán et al., 2015

Ta: annealing temperature (°C), Size in base pair

Table S3.2. Genetic Diversity for nuclear and mitochondrial genes; S=Number of polymorphic (segregating) sites; Pi= Nucleotide diversity, Theta (per site) from Eta. Tajima's D test for all populations were No significant.

	Locus length	S	Pi	Theta	Tajima's D	Fu's Fs statistic
D-loop						
All populations	223	61	0.03853	0.061	-1.1582	-112.91
Culebra	223	34	0.03696	0.04522	-0.6839	-10.112
Parguera	223	45	0.038	0.05538	-1.1084	-21.341
Tres Palmas	223	26	0.03809	0.04405	-0.6080	-3.479
Desecheo	223	29	0.03581	0.04101	-0.4857	-8.085
Mona	223	36	0.04387	0.05274	-0.6774	-4.888
PNNE	223	27	0.04156	0.0472	-0.5540	-2.521
La caleta	223	18	0.03991	0.04183	-0.3399	-0.526
COI: Cytochrome c oxidase subunit I						
All populations	603	30	0.00577	0.01087	-1.5408	-21.864
Culebra	603	18	0.00623	0.00841	-0.9757	-5.332
Mona	603	14	0.00549	0.00664	-0.64419	-3.687
PNNE	603	12	0.00519	0.00679	-1.0279	-2.857
La caleta	603	6	0.00464	0.00478	-0.1009	-2.371
ATROP: intron V from nuclear alpha-tropomyosinc						
All populations	232	8	0.00373	0.004	-0.14781	-2.561
Culebra	232	7	0.00399	0.00459	-0.33727	-1.729
Parguera	232	5	0.00378	0.00354	0.16962	-0.613
Tres Palmas	232	5	0.00391	0.00403	-0.09056	-1.197
Desecheo	232	5	0.00371	0.0035	0.14972	-0.61
Mona	232	6	0.00383	0.0043	-0.29818	-2.508
PNNE	232	2	0.00263	0.00165	1.34546	-0.218
La caleta	232	2	0.00281	0.00213	1.03299	0.345
RH2a-Opsin						
All populations	252	5	0.00093	0.00378	-1.58354	-9.207
Culebra	252	4	0.00098	0.00373	-1.75924	-4.334
Mona	252	3	0.00155	0.00283	-1.00286	-2.976
PNNE	252	0	0	0	0	0
La caleta	252	0	0	0	0	0
4174E20						
All populations	365	19	0.00712	0.00986	-0.78926	-1.975
Culebra	365	7	0.00397	0.00447	-0.30655	-1.849
Mona	365	5	0.00332	0.00326	0.04436	-0.779
PNNE	365	10	0.00861	0.00752	0.49681	-0.466
La caleta	365	17	0.02179	0.01651	1.48883	0.338
1777E4						
All populations	153	18	0.00532	0.02363	-2.20584	-16.235
Culebra	153	0	0	0	0	0
Mona	153	3	0.00236	0.00461	-1.06242	-1.736
PNNE	153	14	0.01699	0.02579	-1.24878	-4.616
La caleta	153	6	0.01284	0.01512	-0.70456	-1.191
RAG1 nuclear recombination activating gene 1						
All populations	354	11	0.00107	0.00585	-2.09418	-12.009
Culebra	354	5	0.00137	0.00351	-1.63789	-3.96

Parguera	354	3	0.00213	0.003	-1.03446	-1.466
Mona	354	4	0.00074	0.00269	-1.75617	0
PNNE	354	1	0.00026	0.00077	-1.1624	-0.957
La caleta	354	0	0	0	0	0

Table S3.3. Characterization of 18 genetic markers: Two mitochondrial genes; five nuclear genes and twelve microsatellites.

	Name	Marker	Primer sequence 5' – 3'	Size	Ta (°C)	From
1.	D-loop	L15995 H16498	AATTCTCACCCCTAGCTCCCAAAG CCTGAAGTAGGAACCAGATG	223	52	Ivannova, 2007
2.	COI	FISH-BCL FISH-BCH	TCAACYAATCAYAAAGATATYGGCAC TAAACTTCAGGGTGACCAAAAAATCA	603	50	Baldwine, et al., 2009
3.	ATROP	A-TROPL A-TROPH	GAGTTGGA TCGCGCTCAGGAGCG CGGTCAGCCTCCTCAGCAATGTGCTT	332	60	Hickerson, and Cunningham, 2005
4.	RAG1	RAG1F1 RAG1R1	CTGAGCTGCAGTCAG TACCATAAG ATGT CTGAGTCCTTGTGAGCTTCCATRAAYT T	354	58	Lí and Ortí, 2004
5.		1777E4F 1777E4R	AGGAGYTGGTGAACCAGAGCAAAGC AGATCRGCCTGAATSAGCCAGTT	153	60	Lí et al., 2010
6.		4174E20F 4174E20R	CTYTCGCTGGCTTTGTCTCAAATCA CTTTTACCATCKCCACTRAAATCCAC	365	58	Lí et al., 2010
7.	RH2a-Opsin	Gen 20F Gen 20F	TATGGTGAAACCGAACATGC CAAAAAGGATGGTGGACAGG	659	58	This study
8.	785	HLUJHHW01BHFL5	F: CAGAGGCTTGCCTTGAAGTG R: CGTTCGCTGCAGGTCATAC	186	59 y 53	Beltrán et al., 2015
9.	1588	HLUJHHW01CE13G	F: GAAAGAGAAGTTGCCGCCTC R: CTCCCATTCCTGACTCACCC	264	59 y 53	Beltrán et al., 2015
10.	1438	HLUJHHW01BM47D	F: CTTAGTTGGGATTGCGTGGC R: GGCCTCAGGAATTTCATCGC	224	59 y 53	Beltrán et al., 2015
11.	8462	HLUJHHW01CDWXY	F: TGTTGTTGTTTCTCCGGCAC R: CTATTCTTTGGACACCGCGG	248	59 y 53	Beltrán et al., 2015
12.	325	HLUJHHW01CP5G8	F: TTGGGTTGCAACTCTGTGTG R: TGCCATCTGTGTCCATTGTG	203	59 y 53	Beltrán et al., 2015
13.	170	HLUJHHW01DLIPU	F: TGACATCCACCACTGACAGG R: TATCGGCTGGTCCTTTCTGC	260	59 y 53	Beltrán et al., 2015
14.	516	HLUJHHW01EM6DJ	F: TCTCCTCAGCCACCAAGAAG R: GGGTATCAGCACTGTTGTCC	150	59 y 53	Beltrán et al., 2015
15.	150	HLUJHHW01CL23R	F: GCGGCACACCTCTATTAAGC R: TCACGCAAACAGATGATAGCC	296	59 y 53	Beltrán et al., 2015
16.	246	HLUJHHW01DN5M	F: GCAGCACGATCGAGAAACTG R: CAGATGGCCTCGTCAAACAC	289	56 y 53	Beltrán et al., 2015
17.	8537	HLUJHHW01EZ6IS	F: CTGAAACTTCCCAACCAGCC R: CCTCGATGCTGCTTGATGTG	287	59 y 53	Beltrán et al., 2015
18.	1358	HLUJHHW01D9S09	F: ACTCGACCCATGTTTCATCATC R: ACATCCACAGTTGTCACTTGC	247	59 y 53	Beltrán et al., 2015
19.	7983	HLUJHHW01C84TE	F: CGGTATAGTGTGGGAGGGTC R: AAACTGGGATTGATGCGTGG	295	59 y 53	Beltrán et al., 2015

Ta: annealing temperature (°C), Size in base pairs.

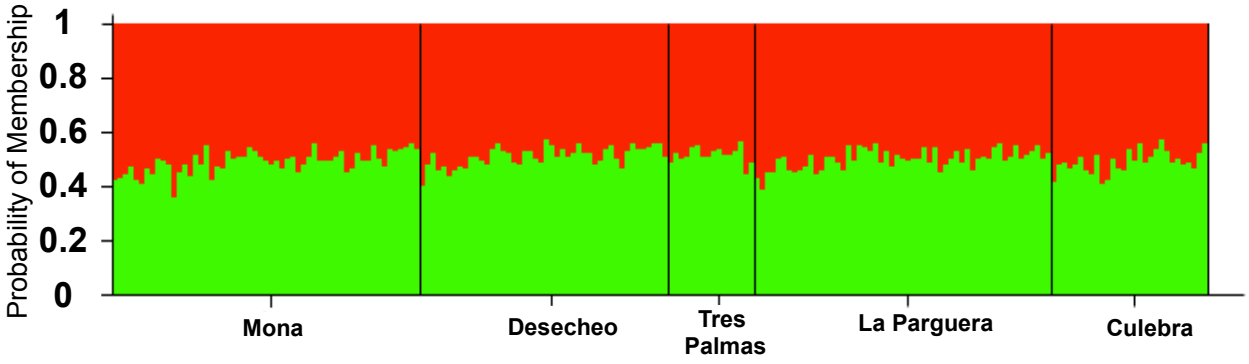

Figure S3.1. Graphical summary of clustering Analysis for *Opistognathus aurifrons* genotypes from five Puerto Rico Marine Protected Areas along Mona Passage using Structure. Probability of Membership of each sampling site (n=7) from individuals per site scored with 12 microsatellites. Each vertical line represents an individual and the estimated proportion of the individual's genome from each inferred cluster.

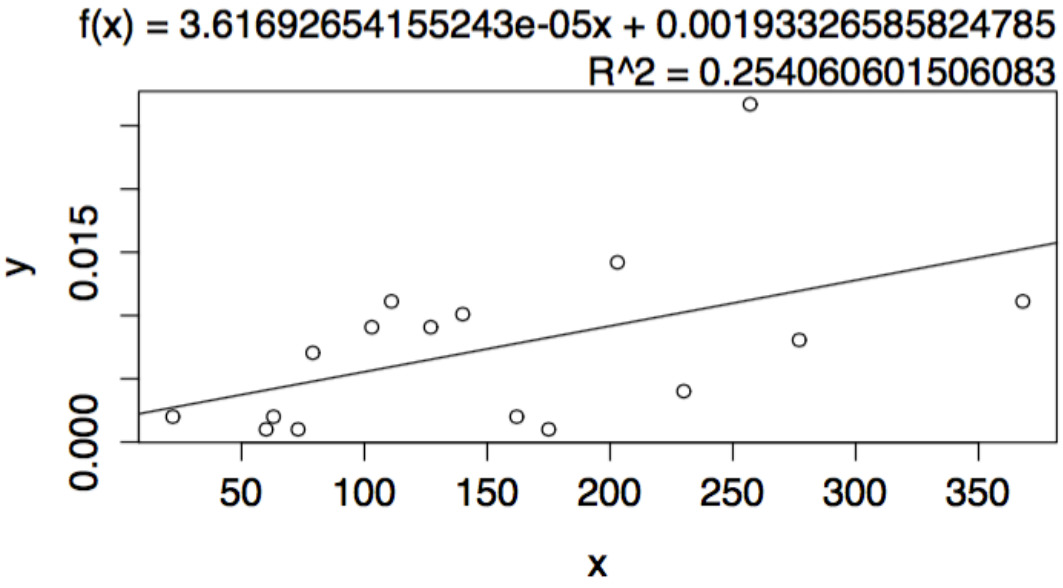

Figure S3.2. Person Correlation Analysis for *Opistognathus aurifrons* genotypes from seven Puerto Rico and Dominican Republic Marine Protected Areas along Mona Passage. Each sampling site (n=7) from individuals per site scored with 12 microsatellites.

Appendix B

Supplementary Information for Chapter 4

Table S4.1. Characterization of 18 genetic markers: Two mitochondrial genes; five nuclear genes, and twelve microsatellites.

Name	Marker	Primer sequence 5' – 3'	Size	Ta (°C)	From
D-loop	L15995 H16498	AATTCTCACCCCTAGCTCCCAAAG CCTGAAGTAGGAACCAGATG	223	52	Ivannova, 2007
COI	FISH-BCL FISH-BCH	TCAACYAATCAYAAAGATATYGGCAC TAAACTTCAGGGTGACCAAAAAATCA	603	50	Baldwine, et al., 2009
785	HLUJHHW01BHFL5	F: CAGAGGCTTGCCTTGAAGTG R: CGTTCGCTGCAGGTCATAC	186	59 y 53	Beltrán et al., 2015
1588	HLUJHHW01CE13G	F: GAAAGAGAAGTTGCCGCCTC R: CTCCCATTCCTGACTCACCC	264	59 y 53	Beltrán et al., 2015
1438	HLUJHHW01BM47D	F: CTTAGTTGGGATTGCGTGGC R: GGCCTCAGGAATTTCATCGC	224	59 y 53	Beltrán et al., 2015
8462	HLUJHHW01CDWXY	F: TGTTGTTGTTTCTCCGGCAC R: CTATTCTTTGGACACCGCGG	248	59 y 53	Beltrán et al., 2015
325	HLUJHHW01CP5G8	F: TTGGGTTGCAACTCTGTGTG R: TGCCATCTGTGTCCATTGTG	203	59 y 53	Beltrán et al., 2015
170	HLUJHHW01DLIPU	F: TGACATCCACCACTGACAGG R: TATCGGCTGGTCCTTTCTGC	260	59 y 53	Beltrán et al., 2015
516	HLUJHHW01EM6DJ	F: TCTCCTCAGCCACCAAGAAG R: GGGTATCAGCACTGTTGTCC	150	59 y 53	Beltrán et al., 2015
150	HLUJHHW01CL23R	F: GCGGCACACCTCTATTAAGC R: TCACGCAAACAGATGATAGCC	296	59 y 53	Beltrán et al., 2015
246	HLUJHHW01DN5M	F: GCAGCACGATCGAGAAACTG R: CAGATGGCCTCGTCAAACAC	289	56 y 53	Beltrán et al., 2015
8537	HLUJHHW01EZ6IS	F: CTGAAACTTCCCAACCAGCC R: CCTCGATGCTGCTTGATGTG	287	59 y 53	Beltrán et al., 2015
1358	HLUJHHW01D9S09	F: ACTCGACCCATGTTTCATCATC R: ACATCCACAGTTGTCACTTGC	247	59 y 53	Beltrán et al., 2015
7983	HLUJHHW01C84TE	F: CGGTATAGTGTGGGAGGGTC R: AAACTGGGATTGATGCGTGG	295	59 y 53	Beltrán et al., 2015

Table S4.2. Diference in spatial genetic structure for 10059 markers along Caribbean Region polulations.

Indices of genetic diversity per population _80 indivoduals_10059_Snps				
Population	Number of alleles	Effective number of alleles	Heterogozygosity within populations (Hs)	Total heterogozygosity Ht
Bahamas	1.407	1.159	0.109	0.109
Curacao	1.506	1.175	0.117	0.117
Culebra	1.438	1.171	0.118	0.118
Desecheo	1.386	1.169	0.118	0.118
Key Largo	1.133	1.097	0.074	0.074
La Caleta	1.409	1.166	0.116	0.116
Mona	1.539	1.179	0.124	0.124
PNNE	1.398	1.188	0.132	0.132
La Parguera	1.492	1.176	0.122	0.122
Rincón	1.449	1.175	0.121	0.121

Table S4.3. Diference in spatial genetic structure for 10059 SNPs along Caribbean Region Polulations. Pairwise comparison along Caribbean Region (p-values). Bonferroni correction = 0.05/45= 0.00111. In red the values with signifficant gentic structure p-value<0.00111.

F'st Caribbean Region_**Data _80_10059**										
	Bahamas	Curacao	Culebra	Desecheo	Key Largo	La Caleta	Mona	PNNE	La Parguera	Rincon
Bahamas	0	0.336	0.294	0.279	0.297	0.044	0.268	0.202	0.315	0.241
Curacao	**0.336**	0	0.107	0.108	0.327	0.286	0.124	0.125	0.115	0.089
Culebra	**0.294**	**0.107**	0	-0.008	0.277	0.23	0.016	0.015	-0.008	-0.04
Desecheo	**0.279**	**0.108**	**-0.008**	0	0.232	0.208	-0.019	-0.008	0.014	-0.058
Key Largo	0.297	0.327	0.277	0.232	0	0.273	0.298	0.273	0.352	0.154
La Caleta	**0.044**	**0.286**	**0.23**	**0.208**	0.273	0	0.206	0.138	0.254	0.171
Mona	**0.268**	**0.124**	**0.016**	-0.019	0.298	**0.206**	0	0.012	0.024	-0.026
PNNE	0.202	**0.125**	**0.015**	-0.008	0.273	0.138	**0.012**	0	0.042	-0.034
La Parguera	**0.315**	**0.115**	**-0.008**	**0.014**	0.352	**0.254**	**0.024**	0.042	0	-0.027
Rincon	**0.241**	**0.089**	**-0.04**	**-0.058**	0.154	**0.171**	-0.026	-0.034	-0.027	0