Fall 2017

Tracing the Genetic Footprints of the Redbelly Yellowtail Fusilier, Caesio Cuning, Across Multiple Spatial and Evolutionary Scales

Amanda Susanne Ackiss
Old Dominion University

TRACING THE GENETIC FOOTPRINTS OF THE REDBELLY YELLOWTAIL FUSILIER, *CAESIO CUNING*, ACROSS MULTIPLE SPATIAL AND EVOLUTIONARY SCALES

by

Amanda Susanne Ackiss
B.A. May 2002, University of Virginia

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

ECOLOGICAL SCIENCES

OLD DOMINION UNIVERSITY
December 2017

Approved by:

Kent E. Carpenter (Director)

Daniel Barshis (Member)

David Gauthier (Member)

Christopher Bird (Member)

ABSTRACT

TRACING THE GENETIC FOOTPRINTS OF THE REDBELLY YELLOWTAIL FUSILIER, C*AESIO CUNING,* ACROSS MULTIPLE SPATIAL AND EVOLUTIONARY SCALES

Amanda Susanne Ackiss
Old Dominion University, 2017
Director: Dr. Kent E. Carpenter

Overfishing is one of the most pervasive threats to coral reef ecosystems, and management of these multi-species resources is hampered by limited species-specific population level information. The reefs in the western tropical Pacific Ocean, including the Coral Triangle, are the most bio-diverse in the world. Home to more than 400 million people, this region contains some of the most threatened coral reef ecosystems. Presented here is the first comprehensive analysis of the genetic structure of *Caesio cuning,* planktivorous fish inhabiting reefs in the Coral Triangle and western Pacific Ocean. Data from both classical Sanger and next-generation sequencing were analyzed across multiple spatial scales to test hypotheses regarding the biogeography, ecology, and population connectivity of this important food fish.

Across the Coral Triangle, mitochondrial DNA sequencing was used to examine broad-scale genetic patterns from 33 locations. Results show the presence of two clades found on either side of the Sunda Shelf, a biogeographic pattern attributed to vicariant isolation during low sea level stands during the Pleistocene Epoch. No evidence for isolation-by-distance across the sampling region was found, however, within the clade associated with the Pacific Ocean, AMOVA and BARRIER analyses indicate significant genetic differences between central Indonesia and the Philippines relative to eastern Indonesia.

Restriction site-associated DNA (RAD) sequencing data from five sites along the

Kuroshio Current were used to examine fine-scale gene flow from the core to the northern limit of the species range. Results indicate that *C. cuning* in this region conform to the predictions of the central-peripheral population model described by Mayr (1963). Edge effects were found in peripheral populations including decreasing effective sample size, increased relatedness, and disjunct peripheral populations.

Finally, since management of reef fish resources occurs at the country-level, RAD sequencing data from seven sites were used to examine fine-scale patterns of population structure within the Philippines. Results suggest that *C. cuning* from Palawan are genetically distinct from all other sites in the north, central, and eastern Philippines. Excluding Palawan, no isolation-by-distance was found despite significant genetic structure between many sites, indicating that gene flow within the Philippines is largely impacted by regional oceanographic features.

This thesis is dedicated to my family, who has whole-heartedly supported me as I traveled the long and winding road to my PhD.

ACKNOWLEDGMENTS

There are many people who have been essential to the successful completion of this dissertation. Foremost, I would like to thank my advisor, Dr. Kent Carpenter, for taking a chance on a graduate student with a Bachelor of Arts in Media Studies, for introducing me to the spectacular marine ecosystems of the Coral Triangle, and for providing me with the incredible opportunity to live and work there. I would like to thank my committee members, Dr. David Gauthier, Dr. Daniel Barshis, and Dr. Christopher Bird, for generously making their labs available to me through months (years) of library prep and for their support through many unforeseen obstacles. I would also like to thank Dr. Annette Meñez, Dr. Menchie Ablan-Lagman, Dr. Ngurah Mahardika, and Dr. Hilconida Calumpang for welcoming me into their labs while I was in the Philippines and Indonesia.

I am so thankful for the many years of shared field work, workshops, videoke, oyster roasts, crab feasts, pig roasts, bonfires, and trivia nights with my academic family, the current and past members of the Carpenter lab: Milli and Jonell Sanciangco, Eric Crandall, Adam Hanson, Jeremy Raynal, Mia Comoros-Raynal, Brian Stockwell, Demian Willette, Ellen Biesack, Jack Buchanan, Christi Linardich, Mike Harvey, Gina Ralph, Andy Hines, and Emilie Stump. Our assorted shenanigans got me through many difficult challenges. Tagay!

I am grateful for the financial support of the Department of Biological Sciences through graduate teaching assistantships and the Dominion Scholar Award and of my advisor through graduate research assistantships. This dissertation would not have been possible without funding support to Kent Carpenter from the National Science Foundation grants OISE-0730256 and DEB-1257632.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF GRAPHS

INTRODUCTION

The Coral Triangle encompasses six countries in Southeast Asia and is both the epicenter of marine biodiversity and a hotspot for threats to coral reef ecosystems (Roberts et al. 2002, Nañola et al. 2011). This region includes the largest area of tropical continental shelf in the world at the convergence of the Indian and Pacific Oceans and is home to over 400 million people. Coastal development, watershed-based pollution, marine-based pollution and damage, and climate change-induced thermal stress all pose direct threats to coral reefs, but overfishing has been listed as the most pervasive local threat to reefs (Burke et al. 2012). Coral reefs are easily exploitable resources, and reef fish and invertebrates are important sources of food and livelihood in the coastal communities of the Coral Triangle (McManus et al. 1992, McManus 1997). As the human population in this region increases annually by an estimated 1-2% (US Census Bureau 2017), anthropogenic pressures on coastal reef resources continue to rise. Informed management of coral reef ecosystems is a priority for the conservation and sustainability of these resources in the coming decades.

Pelagic larvae are the primary means of demographic and genetic connectivity among most marine populations, and an understanding of larval movement within a species' range is essential to the successful management of targeted species. The most common strategy for improving the biomass and abundance of fisheries – and in particular coral reef organisms - is marine reserves (Roberts & Polunin 1991, Russ & Alcala 1996, Gell & Roberts 2003, Sale 2006). One of the most critical needs for the development of effective reserve networks is a more thorough understanding of patterns of larval dispersal (Sale et al. 2005). Although genetic connectivity is not equivalent to demographic connectivity, genetic analyses have specifically

been suggested as a means to delimit the spatial distribution of stocks and can be of use in guiding conservation planning in marine ecosystems (Deriso & Quinn 1998, Palumbi 2003), and this approach has been specifically proposed as a management mechanism in the Coral Triangle (Carpenter et al. 2011).

Genetic methods have long promised to reveal the stock structure of marine species but limitations have followed available technology. Early genetic analyses were widely successful in detecting cryptic speciation, uncovering sex-biased dispersal, and illuminating historical patterns of biogeographic structure (Knowlton et al. 1993, Avise 2000, Knowlton 2000, Pardini et al. 2001, reviewed by Bowen et al. 2014), but several characteristics of classical markers limit their ability to resolve fine-scale patterns of gene flow. Enzyme variants called allozymes were the first genetic markers to be employed to examine population structure (Lewontin & Hubby 1966). However, allozymes carry less information than the underlying DNA sequence due to codon degeneracy. Synonymous mutations remain undetected, but even non-synonymous mutations can be missed if the electrical charge of a new allozyme is equivalent to the ancestral state.

With the advance of genetic sequencing and the development of polymerase chain reaction (PCR), maternally inherited mtDNA was considered for a time to be an 'ideal' genetic marker for population and phylogeographic analyses (Avise et al. 1987, Moritz et al.1987). Several properties led to this conclusion: mtDNA is found in relatively high concentrations in vertebrate tissues making it easy to isolate, the genome is a simple structure assumed to be non-recombining, it has a quarter of the effective population size of nuclear DNA, and it has a rapid evolutionary rate – at the time estimated to exceed that of nuclear DNA by a factor of five to ten (Brown et al. 1979, Vawter and Brown 1986). However, in the following decades the proliferation of mtDNA-based studies across a broad number of taxa revealed several flaws in

these assumptions. Large variations in mitochondrial mutation rates relative to nuclear DNA were measured among different vertebrate taxa (Martin et al. 1992), and evidence of heteroplasmy and recombination were found in many animal species including marine fish and invertebrates (Magoulas & Zouros 1993, Hoarau et al. 2000, Ladoukakis & Zouros 2001). These exceptions - coupled with the inherent limitations acknowledged from the start of a genome that typically reflects only matrilineal history - mean the use and interpretation of mtDNA-based population analyses must be governed with care. Mitochondrial DNA still remains one of the most accessible targets of population analysis and is often validated (Bowen et al. 2014), but population genetic studies are now widely considered incomplete without the inclusion of nuclear DNA (Zhang & Hewitt 2003, Ballard & Whitlock 2004).

Microsatellites held promise to provide nuclear DNA genotypes capable of detecting fine-scale structure. These non-coding, bi-parentally inherited, and highly-polymorphic genetic markers were viewed as the best method for local population differentiation (Queller et al. 1993, Park and Moran 1994, O'Reilly and Wright 1995). In reef species such as groupers, snappers and clownfish where newly recruited juveniles and adults can be successfully collected, parentage analysis via microsatellites was used to evaluate larval export and self-retention in marine reserves (Jones et al. 2005, Harrison et al. 2012). Despite this utility, microsatellites may be no better than mtDNA at resolving marine population structure (Lukoschek et al. 2008, Karl et al. 2012) and in some cases can be outperformed by mtDNA (Hoarau et al. 2004). In addition to variable performance, microsatellites come with their own suite of challenges. Early methods of microsatellite development were relatively time consuming and expensive, and with fragment analysis came problems of homoplasy, null alleles, allelic drop out, and reproducibility between laboratories and equipment (Selkoe and Toonen 2006).

The emergence of novel sequencing technology and new methods of analysis are beginning to fill in the gaps left by classical approaches in population genetics. The early front-runners in next-generation sequencing (NGS) platforms such as Roche 454 pyrosequencing, Illumina, and SoLID and Ion semiconductor sequencing (both Life Technologies) enabled scientists to sample tens of thousands of independent genetic markers from the genome of any given species (i.e. Hohenlohe et al. 2011, Kraus et al. 2011). In particular, these technologies allow for genome-wide sampling of non-model organisms (Ekblom & Galindo 2011, Seeb et al. 2011). Nascent applications to population genetics research have shown these high-throughput sequencing methods capable of unprecedented resolution (Novembre et al. 2008).

Most studies applying high-throughput sequencing for population-level analysis target panels of single nucleotide polymorphisms or SNPs. Common SNP genotyping methods include oligonucleotide microarrays (Lockhart et al. 1996), TaqMan® assays (McGuigan and Ralston 2002) and restriction site-associated DNA (RAD) sequencing (Miller et al. 2007, Baird et al. 2008). Of these, RAD sequencing has emerged as one of the most popular approaches for sampling the genome of non-model organisms, and several alternative RAD methodologies were developed (Peterson et al. 2012, Wang et al. 2012, Toonen et al. 2013). As with previous genetic markers, RAD sequencing and its various forms come with their own suite of challenges and limitations such as allelic dropout, PCR bias and duplicates, and computationally intensive bioinformatics (see Davey et al. 2012, Gautier et al. 2012, Andrews & Luikart 2014, Purtiz et al. 2014, Andrews et al. 2014).

As with sequencing technologies, new approaches to genetic analysis of populations are also evolving. The most commonly used metrics for measuring genetic structure are *F*-statistics (Wright 1931, Bird et al. 2011). However, when *F*-statistics are employed to identify stock

structure, they often yield low and non-significant values (Waples 1998). The large effective population sizes of marine organisms often lead to high genetic diversity that washes out the denominator of equations for the fixation index (F_{ST}) and its analogues (Hedrick 2005, Meirmans 2006, Bird et al. 2011), making it very difficult to reject panmixia. Moreover, estimates of gene flow arising from F_{ST} are typically based on the island model of migration (Wright 1943), which has several unrealistic assumptions including discrete generations, uniform and constant effective population sizes (N_e) for each subpopulation, uniform and constant migration (m), and no mutation (Waples 1998). Coalescent approaches to assessing N_e independently of gene flow (Beerli and Felsenstein 2001) have recently been shown to be effective at elucidating genetic structure in datasets that failed to reject both panmixia and the island model when using F_{ST} approaches (Crandall et al. 2012) but have not yet become standard in population genetic analyses. Haplotyping individuals at unique loci across the genome can provide an added analytical benefit since the polymorphic sites in a single genomic fragment can be considered linked. Each marker or tag effectively contains not only the genotype (SNPs) but also the genetic history (linkage of SNPs) of each individual. Both standard statistical and coalescent genetic analyses can then be applied to these regions to determine levels of genetic diversity, population structure, and gene flow as well as the phylogenetic lineage of sampled individuals for the identification of source populations and directional gene flow.

The increased accessibility and sampling power of NGS coupled with evolving approaches to analyzing genetic diversity and gene flow have now opened the door for scientists to greatly improve the ability to examine stock structure of threatened coral reef fisheries using genetic methods. Up to this point, allozyme, mitochondrial, and microsatellite markers have been the most commonly used genetic methods to study population connectivity and stock structure,

so the genomes in the majority of coral reef and fisheries species remain inadequately sampled. Utilizing a diverse array of markers has been suggested as the most successful approach to genetic studies of population biology (Sunnucks et al. 2000), therefore pairing classical and high-throughput genotyping of both mitochondrial and nuclear genomes with a combination of analytical methods is likely the best approach for examining genetic patterns within a coral reef fishery in the Coral Triangle.

Study organism

The redbelly yellowtail fusilier, *Caesio cuning* (Bloch 1791), is an ecologically and economically important fish in the Coral Triangle at risk to overfishing. Schooling, reef-associated fusiliers (Perciformes: Caesionidae) are planktivores that can be found feeding at the reef face. Fusiliers account for a sizable portion of harvested reef species in the Coral Triangle and are caught both on the reef and over sand bottom via a variety of fishing gear including spears, hand-lines, fish traps, trawls, drive-in nets and gill nets (Carpenter 1988). Fishing pressure has resulted in significant caesionid population declines in the region. In the Philippines, Alcala and Russ (1990) completed a visual census on Sumilon Island to document changes in reef fish density after protective management was removed for a quarter of the island's reefs. They measured a 64% decrease in caesionid density after an eighteen-month period of fishing by approximately 100 local fishermen from an adjacent island using hand-paddled canoes. Given that artisanal fishing of caesionids has been shown to cause precipitous drops in local abundance, a better understanding of stock structure is particularly important for the management of *C. cuning*. In the Philippines alone, the annual estimated catch of caesionids

in commercial and municipal fisheries is 22,000 metric tons (BAS 2010), but given the difficulty of estimating catch data in artisanal fisheries in this region, these data are likely greatly underestimated (Alcala & Russ 2002).

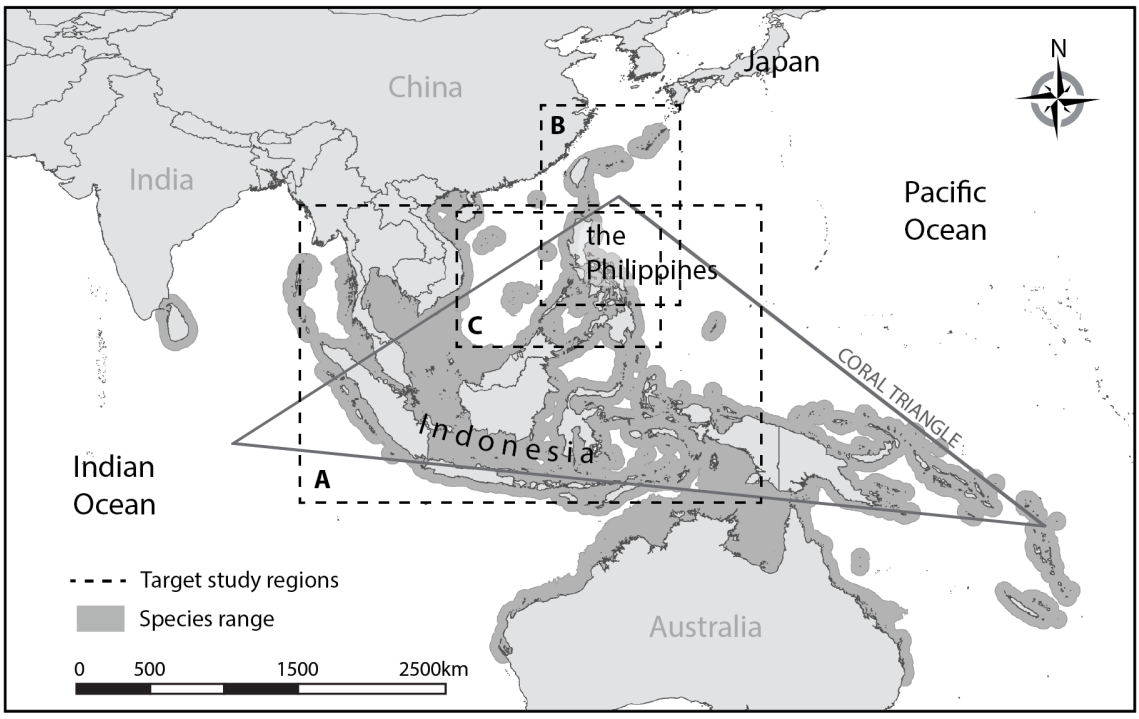


Fig. 1. The geography of study regions within the distribution of *Caesio cuning.* This research is an examination of multi-scale genetic patterns: A) across the core of the species range, B) in the northern periphery of the species range, and C) within the Philippines.

Like most of the targeted reef species in this region, little is known about the larval ecology of *Caesio cuning. Caesio cuning* is a broadcast-spawner with a range that extends from northern Australia to Japan and east from Sri Lanka to Vanuatu (Fig. 1). The closest relative with a known pelagic larval duration (PLD) is *Pterocaesio chrysozona* with an estimated PLD of 37-

47 d (Doherty et al. 1995). While there is no clear evidence to suggest strong larval behavior in *C. cuning* such as natal homing (Leis & Carson-Ewart 2003), Leis noted that *C. cuning* larvae are most often found over the mid- or inner-shelf in plankton tows, possibly accounting for the species' absence on nearby oceanic islands such as Guam that are inhabited by other fusiliers (Reader & Leis 1996). *Caesio cuning* and other fusiliers have been observed sleeping in crevices and holes in the reef structure, however, their level of fidelity to such shelter sites and individual reefs is unknown. The mobility of larval *C. cuning* coupled with their dependence on reef structure for shelter as adults and undefined movement suggests a varied spectrum of dispersal potential presenting a challenge to their effective management.

RESEARCH OBJECTIVES

The goal of this research is to pair classical population genetics approaches and high-throughput genotyping to illuminate aspects of the ecology, biogeography, and population connectivity of a non-model fishery species of the Coral Triangle. In order to address the unknown dispersal potential of *Caesio cuning* and to test the usefulness of applying similar methods of population analysis to the numerous other commercial and artisanal species of the Coral Triangle with unknown life history traits, this research utilizes both mtDNA analysis and restriction site-associated DNA (RAD) sequencing to examine genetic patterns across three spatial scales: across the Coral Triangle, from the core to the northern limit of the species range, and within a single country (Fig. 1).

Examining broad-scale biogeographic patterns

The population structure of a variety of Coral Triangle reef species has been examined, but whether genetic structure in the Indo-Pacific fusiliers existed was unknown prior to this research. Since there were no published studies on the genetics of any caesionid, the first part of this research focuses on applying the standard approach of sequenced mtDNA analysis across as much of the range of *Caesio cuning* as feasible, including all major regions in the Philippines and across the width of Indonesia. Of the many studies employing mtDNA markers to examine patterns of gene flow in demersal Indo-Pacific reef organisms such as snails, seastars, damselfishes and snappers (Crandall et al. 2008b; Vogler et al. 2008, Drew & Barber 2009, Gaither et al. 2010), none have examined genetic structure in a reef-associated zooplanktivore. With this in mind, the first part of this dissertation addresses the following hypothesis:

i. *Mid-water fusiliers contain detectable genetic structure across the Coral Triangle similar to that found in reef organisms known to be strongly site-associated.*

The analyses examine genetic connectivity and mitochondrial lineage divergence in *C. cuning* across the Coral Triangle by focusing on the following questions: (1) are multiple mitochondrial clades present in *C. cuning* across its range, (2) if so, is there evidence to indicate cryptic speciation or sex-biased dispersal, (3) are these mobile, mid-water planktivores impacted by the same barriers we commonly see in Indo-Pacific demersal species (i.e. Barber et al. 2006, Crandall et al. 2008a, Timm & Kochzius 2008) or do they exhibit the panmixia found in near-shore pelagics such as scad (Borsa 2003), and (4) are there geographic regions within the Coral

Triangle that can be identified as priority regions for further study to aid in proper management of this fishery?

Testing the central-peripheral population model

As a targeted species with a bipartite life history, the population structure of *C. cuning* is likely greatly influenced by the oceanographic patterns across its range. *C. cuning* larvae appear to be restricted to shallow water shelves. With only minor evidence of larval orientation within the water column, oceanography likely plays a major role in providing genetic connectivity across large expanses of water and disjunct reef systems. The northern offshoot of the North Equatorial Current, the Kuroshio Current, flows from the northern Coral Triangle to the very apex of the species' distribution in the Ryukyu Islands of Japan (Fig. 2). The first objective for this second area of research is to apply high-throughput genotyping via RAD tags in order to address a hypothesis on the role of this predominant western boundary current in sustaining genetic connectivity across large expanses of open water:

ii. *The Kuroshio Current maintains a genetically homogeneous stock of* C. cuning *along the eastern coast of the Philippines to Okinawa, the northern extent of the species' range.*

The central-peripheral population model (Mayr 1963) describes how populations behave at the edges of a species' range compared to core populations. The differences between the two are often a factor of gene flow. Range limits are caused by a combination of factors including dispersal, habitat specialization, and abiotic and biotic spatial variation in environments (MacArthur 1972, Brown 1984). Peripheral populations are generally predicted to be smaller,

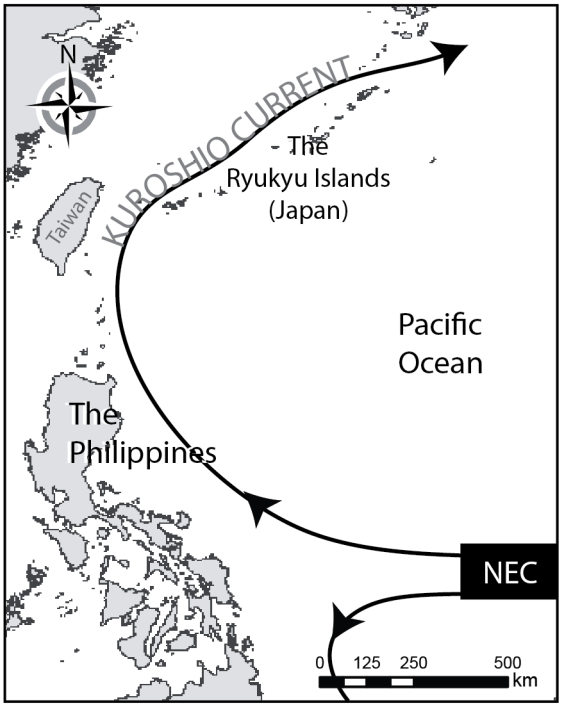


Fig. 2. The formation of the Kuroshio Current. At the central eastern edge of the Philippines, the northern branch of the bifurcation of the North Equatorial Current (NEC) becomes the Kuroshio Current. This major oceanographic feature flows past the northern limit of the *C. cuning* distribution.

and unlike their central counterparts, populations at the periphery typically only receive migrants from a single direction. This isolation puts them at increased risk of stochastic events like founder effects, bottlenecks, and genetic drift (Levin 1970, Hedrick et al. 1976). Decreased effective population sizes and genetic diversity and increased differentiation have been measured at species' peripheries (Yeh & Layton 1979, Vucetich & Waite 2003, Johannesson & André 2006), and the presence of these signatures make peripheral populations foci for conservation efforts (Lesica and Allendorf 1995). Gene flow plays an important part in determining the level of isolation in peripheral populations. The second objective is to determine if the Kuroshio

Current provides a conduit for gene flow that dampens these effects in peripheral populations of *C. cuning*:

iii. Due to gene flow via the Kuroshio Current, populations of C. cuning *at the periphery do not exhibit detectable genetic signatures of edge effects.*

A better understanding of the mechanisms maintaining populations at the periphery of the species' range provides us with valuable information on population connectivity in *C. cuning*. Classical and coalescent analyses of SNP data provides the genetic resolution and tools to address the following related questions: 1) what level and direction of gene flow (i.e. migration) – if any – do we see along the Kuroshio current, 2) do sites at the edge of the species' distribution appear to be genetically distinct, and therefore self-sustaining, or do they appear to rely on sites downstream for larval replenishment, and 3) can we detect a significant decrease in the effective population size (N_e) of a tropical marine species as we reach the limits of the species' range as is seen in terrestrial species (Vucetich & Waite 2003)?

Examining genetic connectivity within the Philippines

High-throughput NGS technology now affords us the sampling power and resolution to examine the effects of oceanography on the genetics and stock structure of a species across a spatial scale more conducive to developing functional management plans within the Coral Triangle. The final objective of this research is to apply RAD sequencing in order to address four main hypotheses regarding the role of oceanographic features as barriers to gene flow within a single country:

iv. *Within the Philippines, the Sulu Sea Throughflow divides stocks of* C. cuning *into genetically distinct east and west populations.*

The Sulu Sea Throughflow is a strong, southerly current that bifurcates the Philippines east and west of the Sulu Sea (Fig. 3). This current originates in the South China Sea and flows southward into central Indonesia and recent larval dispersal models predict this current to be a major conduit for larval dispersal from the Philippines into Indonesia (Kool et al. 2011). Within the Philippines, there is preliminary mitochondrial evidence that the Sulu Sea Throughflow is a barrier to gene flow east and west of the Sulu Sea in multiple organisms including hedgehog seahorses, three species of giant clam and at least two species of damselfish (Lourie et al. 2005, Deboer et al. 2014, Raynal et al. 2014, Hanson 2015).

v. *Periodic intrusion of the Kuroshio Current through the Luzon Strait provides connectivity between populations of* C. cuning *on the east and west coasts of Luzon.*

When the Northern Equatorial Current (NEC) hits the east coast of the Philippines, it bifurcates north and south into the Kuroshio and Mindanao Currents, respectively (Fig. 3). The Kuroshio Current is a powerful western boundary current that reaches mean maximum surface velocities of ~1.2 m s^{-1} (~104 km/day; Yang et al 2015). Kuroshio penetration into the South China Sea through the Luzon Strait occurs seasonally with the northeast-southwest monsoon (Metzger and Hurlbert 2001), and may provide a pathway for connectivity between sites on the east coast of Luzon and sites on the west in populations of *C. cuning* as has been observed in the rabbitfish, *Siganus fuscescens* (Ravago-Gotanco and Juinio-Meñez 2010).

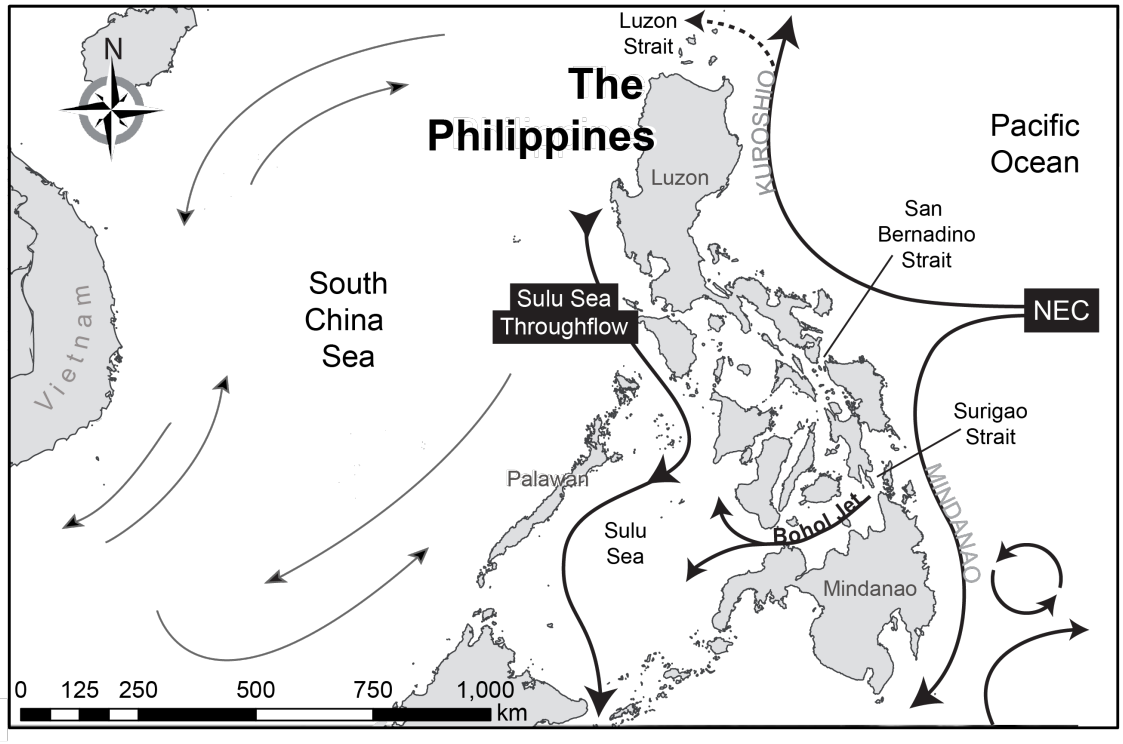


Fig. 3. Major oceanographic currents around the Philippine Islands. These include the Sulu Sea Throughflow, the Bohol Jet, and the North Equatorial Current (NEC).

vi. *High levels of gene flow are found in populations of* C. cuning *within the central Philippine islands, in part mediated by the Bohol Jet.*

The Surigao Strait is one of only two direct points of entry into the central Philippine islands from the Western Pacific (Fig. 3; Hurlbert et al. 2011), and is approximately 18 km wide with a sill depth of 63 m. The Surigao Strait is also the starting point for a predominant southwestward surface current that flows through the northern Bohol Sea with a mean surface velocity of 0.56 m s^{-1} (~48 km/day) known as the Bohol Jet (Hurlbert et al. 2011, Gordon et al. 2011). The relative speed of this westward jet current makes this region a potential area for high levels of connectivity. Given that the lower limit of migration necessary to overcome the effects

of genetic drift is one migrant on average per generation (Spietz 1974, Mills and Allendorf 1996), it is likely that the Bohol Jet is a conduit for gene flow in the southern central Philippines.

In addition to the Bohol Jet, the close proximity of islands (and therefore reefs) is likely to aid connectivity between populations in this region. A previous genetic study using SNPs generated from RAD sequencing found no significant differences across three sites sampled in the central Philippine islands in the parrotfish, *Scarus niger* (Stockwell et al. 2016).

vii. *North of the Surigao Strait, restricted oceanographic flow from the eastern seaboard into the central islands of the Philippines through the San Bernardino Strait has resulted in genetically distinct eastern and central populations.*

The San Bernardino Strait is the second of two major points of connection between inland Philippine seas and the Western Pacific Ocean and is approximately 17 km wide with an estimated sill depth of 23 m (Hurlbert et al. 2011, Gordon et al. 2011). Two islands located at the mouth of the strait, Capul and San Antonio, impede direct flow into inland bays, and coupled with low sill depth, it is likely that tidal currents predominantly impact this strait rather than strong surface flow. Reduced gene flow between eastern to central populations could maintain distinct populations in these regions.

Using the power of RAD sequencing to resolve fine-scale genetic structure, this chapter addresses these four hypotheses regarding the role of predominant oceanographic currents and restrictive land barriers as regional obstacles to gene flow with the following additional questions in mind: (1) is gene flow in a motile fusilier driven by influences mainly affecting larval

dispersal and (2) can we find evidence for significant genetic structure within the Philippines that may aid in the development of a regional management plan?

PRONOUNCED GENETIC STRUCTURES IN A HIGHLY MOBILE CORAL REEF FISH, *CAESIO CUNING*, IN THE CORAL TRIANGLE

Introduction

The concentration of marine biodiversity in the Coral Triangle poses both biogeographical questions and management challenges. Straddling the Indo-Malay-Philippine Archipelago and extending eastward to the Solomon Islands, the Coral Triangle is home to the highest diversity of marine organisms in the world (Briggs 1995, Carpenter & Springer 2005, Veron et al. 2009). Coral reef habitat in this region is extensive and complex, rivaling the Great Barrier Reef in area and spanning well over 25,000 islands. During the Pleistocene epoch (~2.5 mya to 12 kya), repeated glaciations caused radical changes to the regional geography as the Sunda and Sahul Shelves repeatedly rose above and fell below the surface of the water (Voris 2000). The exposure of these shelves significantly narrowed the gateway between the tropical Indian and Pacific Oceans, and sea level fluctuations during this epoch have been implicated in numerous studies as a driver of regional population differentiation and speciation across this region (e.g. Springer & Williams 1990, Mcmillan & Palumbi 1995, Barber et al. 2006, Crandall et al. 2008a,b, Vogler et al. 2008). More recently, oceanographic processes have been implicated in creating and maintaining genetic structure within this region. In particular, the Mindanao and Halmahera eddies, created at the convergence point of the Northern Equatorial Current and the New Guinea Coastal Current, have been hypothesized to limit larval dispersal, and isolate populations across the Maluku sea (Barber et al 2006, 2011, Kool et al. 2011).

Adapted from Ackiss et al. 2013 with permission from Inter-Research Science Center

Identifying regions of limited connectivity in species that span the Coral Triangle can lead to insights into the stock structure of fisheries for management, as well as mechanisms promoting lineage divergence in this region. Molecular techniques are particularly useful in highlighting regions where gene exchange does not occur (Hedgecock et al. 2007). Recent reviews indicate the presence of several genetic breaks shared by multiple species across this region, demonstrating that distinct geophysical processes can promote population structure and even lineage divergence within the Coral Triangle (Carpenter et al. 2011, Barber et al 2011). However, to date the vast majority of reef species showing pronounced genetic structure across the Coral Triangle have been demersal, such as clams, stomatopods, seastars, gastropods and clownfish (Barber et al. 2006, Crandall et al. 2008a,b, Deboer et al. 2008; Timm & Kochzius 2008, Nuryanto & Kochzius 2009). In contrast, relatively understudied near-shore pelagics give mixed results. The round scad mackerel, *Decapterus macrosoma*, shows very little genetic structure (Borsa 2003), while its congener *Decapterus russelli* shows up to three genetically structured populations (Rohfritsch & Borsa 2005).

Unfortunately the diversity that makes the Coral Triangle an area of evolutionary and biogeographic interest is vulnerable. The region is a hotspot for coral reef threats (Roberts et al. 2002, Nañola et al. 2011). The human population in this region increases annually by an estimated 1-2% (US Census Bureau 2017), and anthropogenic pressures on coral reef resources continue to rise. Coastal reefs are easily exploitable resources, and reef fish and invertebrates are important sources of food and livelihood in the coastal communities of Southeast Asia (McManus et al. 1992, McManus 1997). Informed management of coral reef ecosystems is a priority for the conservation and sustainability of coral reef resources in the coming decades.

The most accepted strategy for improving the biomass and abundance of reef organisms is establishment of marine reserves (Roberts & Polunin 1991, Russ & Alcala 1996, Sale 2006). Because dispersive larvae are the primary means of demographic and genetic connectivity among most populations, understanding patterns of larval dispersal has been identified as one of the most critical gaps in developing effective reserve networks (Sale et al. 2005). Although genetic connectivity is not equivalent to demographic connectivity, genetic methods can be of use in guiding conservation planning in marine ecosystems (Palumbi 2003). By identifying regions that are genetically and demographically independent, conservation planners can partition large marine ecosystems into smaller, more tractable management areas for which networks of marine reserves can be designed (Green & Mous 2008). This approach has been specifically proposed as a management mechanism in the Coral Triangle (Carpenter et al. 2011).

Schooling, reef-associated fusiliers (Perciformes: Caesionidae) are planktivores found feeding at the reef face and account for a sizable portion of harvested reef species in the Coral Triangle. They are caught via a variety of gear including hand-lines, fish traps, trawls, drive-in nets and gill nets (Carpenter 1988). In the Philippines alone, the annual catch of caesionids in commercial and municipal fisheries is approximately 22,000 metric tons (BAS 2010), but given the artisanal nature of most reef fisheries in this region, these catch data are likely greatly underestimated (Alcala & Russ 2002).

The redbelly yellowtail fusilier, *Caesio cuning* (Bloch 1791), is a caesionid commonly found in local markets across the Coral Triangle. It is a conspicuous mid-water member of Indo-Pacific reef ecosystems with a distribution that ranges from southern Japan to northern Australia and from Vanuatu to Sri Lanka. *C. cuning* are schooling, broadcast spawners, but beyond this, little is known about the larval ecology of the species. There are no known differences between

the life history strategies of males and females to suggest sex-biased dispersal. The closest relative with a known pelagic larval duration (PLD) is *Pterocaesio chrysozona* with an estimated PLD of 37-47 d (Doherty et al. 1995), and there is no evidence to suggest strong larval behavior such as homing (Leis & Carson-Ewart 2003) that may limit dispersal potential. As adults, *C. cuning* are highly mobile members of the coral reef ecosystem. While they can also be captured in trawls over soft bottom environments (Carpenter 1988) the extent of their movement remains unknown. *C. cuning* and other fusiliers have been observed sleeping in crevices and holes in the reef structure; however, their level of fidelity to such shelter sites and individual reefs is unclear. The mobility of larval *C. cuning* coupled with their dependence on reef structure for shelter and undefined movement as adults suggests a varied spectrum of dispersal potential.

The purpose of this study is to assess regional genetic connectivity and lineage divergence in *Caesio cuning* in order to address two major questions: (1) are mid-water, reef-associated planktivores impacted by the same barriers seen in demersal species or do they exhibit the panmixia found in near-shore pelagics and (2) if limitations to dispersal in *C. cuning* are present, can distinct geographic stocks be identified to aid in the management of fusiliers?

Methods

More than 630 *Caesio cuning* samples were collected from fish markets or by spear while SCUBA or skin diving from 33 localities in the Coral Triangle (Fig. 4). Only samples that were confirmed as being caught on nearby reefs were collected from local markets. Tissue samples were taken from the pectoral or caudal fin base and preserved in 95% ethanol.

DNA amplification and sequencing reactions were conducted at Boston University, the University of the Philippines Marine Science Institute, De La Salle University and Udayana

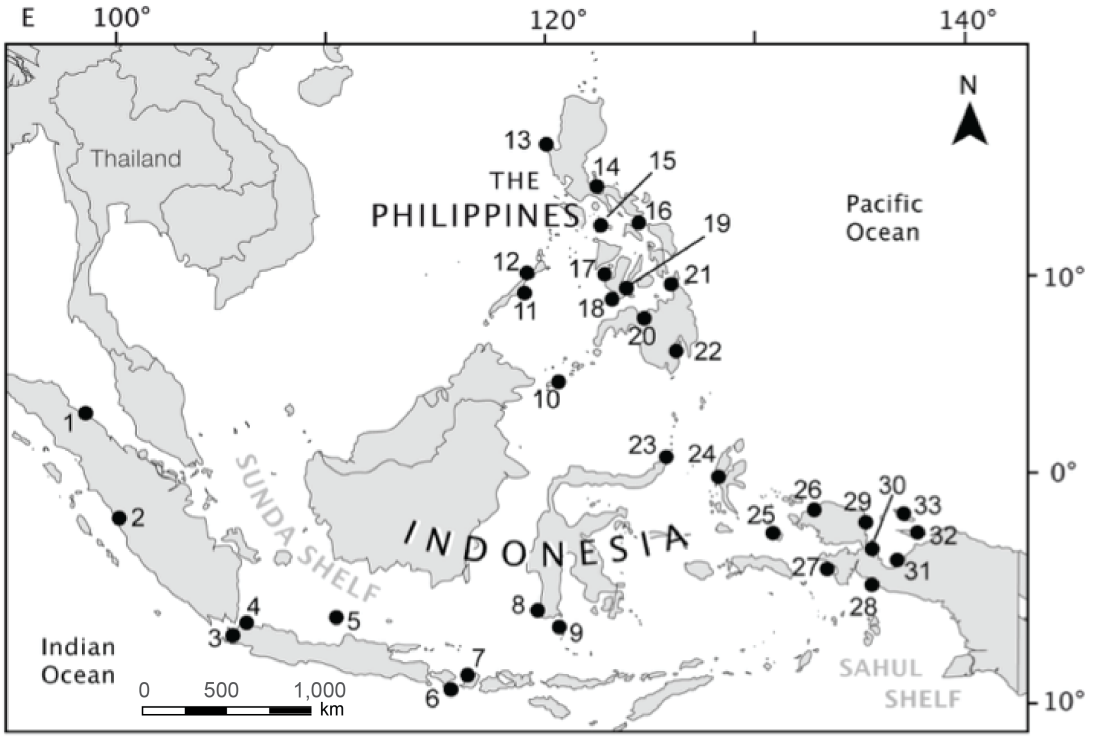


Fig. 4. Sampling localities across the core of the species range. Medan (1), Padang (2), Anyer (3), Seribu (4), Karimunjawa (5), Bali (6), Lombok (7), Makassar (8), Selayar (9), Tawi Tawi (10), Honda Bay (11), Ulugan Bay (12), Bolinao (13), Perez (14), Romblon (15), Sorsogon (16), Guimaras (17), Negros Occidental (18), Negros Oriental (19), Balingasag (20), Dinagat (21), Davao (22), Manado (23), Halmahera (24), Raja Ampat (25), Sorong (26), Fak Fak (27), Kaimana (28), Manokwari (29), Windesi – Teluk Cenderawasi (30), Karei – Teluk Cenderawasi (31), Yapen (32), Biak (33).

University. Whole genomic DNA was extracted using a 10% Chelex (Biorad) solution (Walsh et al. 1991). A 382-bp region of the mitochondrial d-Loop was amplified via polymerase chain reaction (PCR) using the forward and reverse primers CR-A and CR-E (Lee et al. 1995). PCR reactions were conducted in a 25 µL reaction consisting of 1 µL DNA extraction, 2.5 µL of 10x buffer, 2 µL MgCl2 (25 mM), 2.5 µL dNTPs (8 mM), 1.25 µL of each 10 µM primer, 1 µL of template, and 0.625 U of AmpliTaq (Applied Biosystems). Manual hot-start thermocycling

parameters were employed as follows: initial hold at 80°C, denaturation 94 °C (1min), main cycle 94°C (30 s), either 50 or 52°C (30 s) and 72°C (40 s) for 39 cycles, then a final extension of 72°C (7-10 min).

PCR products were electrophoresed on a 1% agarose gel and visualized with ethidium bromide or SYBR® Green staining. Successful PCR reactions were enzymatically prepared for sequencing by mixing 5 µl of PCR product with 0.5 U of Shrimp Alkaline Phosphatase and 5 U of Exonuclease and incubating for 30 minutes at 37°C followed by 15 minutes at 80°C. Forward and reverse sequencing reactions were performed with Big Dye terminator chemistry and run on an ABI 3730 automated DNA Sequencer (Applied Biosystems). Forward and reverse sequences were proofread in Sequencher™ 4.7 (Gene Codes Corporation, Ann Arbor, Michigan) and all resulting 382-bp fragments were aligned with ClustalX v2.0.12. The online tookit FaBox (Villesen 2007) was used to reduce the final alignment to unique haplotypes and create an input file for the population genetics data analysis program Arlequin 3.5.12 (Excoffier & Lischer 2010).

The species identity of sampled haplotypes was confirmed with a neighbor-joining tree run in PAUP* (Swofford 2003) that included the three most closely related species found across the sampling range as outgroups– *Caesio lunaris*, *Caesio teres* and *Caesio xanthonota*. The frequencies and phylogenetic relatedness of haplotypes were examined with a median-joining minimum spanning tree generated in NETWORK v4.6 (Bandelt et al. 1999). For each locality Arlequin and DnaSP v5 (Librado & Rozas 2009) were used to calculate standard genetic diversity indices, and the null hypothesis of neutrality was tested in the mitochondrial control region using Fu's F_S and Fu and Li's D^* tests, with significance determined by 1000 simulations of a neutral coalescent model. The latter two statistics were employed to evaluate the potential

effects of selection and demographic processes such as population expansion in the data (Fu 1997).

To investigate the presence of barriers to dispersal and gene flow, both *a priori* and *post hoc* analyses were performed. First, an analysis of molecular variance (AMOVA) and population pairwise Φ_{ST} were examined in Arlequin. The AMOVA framework was used to group sampling localities and test for hierarchical population structure within the dataset following *a priori* hypotheses based on previously measured phylogeographic breaks (Fig. 5; Table 2) as follows: absence of genetic structure, restricted gene flow east and west of the Makassar strait, a Sunda Shelf break at western Sumatra, the Philippines vs. Indonesia, east vs. west of the Maluku Sea, and a break at Cenderawasih Bay in Papua. All AMOVAs were run using sites with $n \geq 15$ and employed the Tamura and Nei's (1993) model of evolution, which was the model in Arlequin most equivalent to the best model for the dataset determined by jModelTest v1.0 (Posada 2008, Guindon & Gascuel 2003), the general time-reversible (Tavaré 1986). The significance of pairwise Φ_{ST} as well as among and within-population variance in the AMOVA framework was calculated using 30,000+ random permutations of the dataset. The p values for multiple pairwise comparisons were corrected using Benjamini and Hochberg's (1995) false discovery rate.

In addition, a *post hoc* spatial analysis of the pairwise Φ_{ST} matrix generated in Arlequin was employed using the program Barrier version 2.2 (Manni et al. 2004). Barrier characterizes the spatial relationship of sites from their GPS coordinates using Voronoi (1908) tessellation and Delaunay (1943) triangulation and applies Monmonier's (1973) maximum difference algorithm to a matrix of genetic distances (Φ_{ST} in this case) to identify genetic barriers across geographic space. The robustness of barriers were tested by resampling individuals within populations with

replacement using Excel and creating 100 bootstrapped replicates of the pairwise Φ_{ST} matrix in Arlequin.

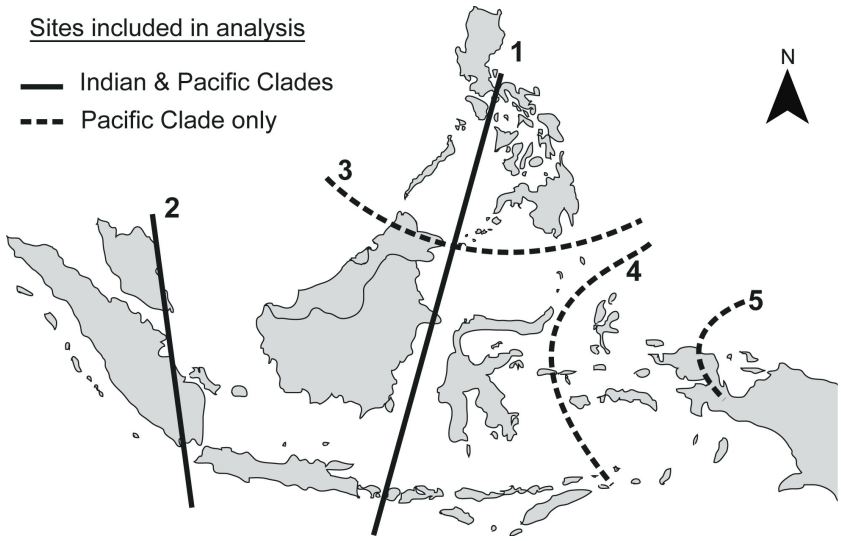


Fig. 5. AMOVA hypotheses. Lines indicate the approximate locations of regional genetic breaks found in the mtDNA of other well-sampled coral reef and near reef species across the Coral Triangle (see Table 2). Solid lines indicate partitions tested with a hierarchical analysis of molecular variance that included sites from both the Indian and Pacific clades; dashed lines indicate partitions tested within the Pacific clade only.

Since discrete genetic breaks can bias the results of analyses of Isolation by Distance (IBD) and the presence of isolation by distance can generate false positives in analyses of hierarchical structure (AMOVA) (Meirmans 2012), partial Mantel tests were employed that controlled for both optimal AMOVA clusters and geographic distance using the 'vegan' package for R (Oksanen et al. 2012; R Core Team 2012). Pairwise genetic distances (Φ_{ST}) among localities with n > 15 were imported from Arlequin, and negative pairwise Φ_{ST} values, a result of within-population variance exceeding among population variance, were set to zero. A geographic distance matrix was generated using a previously developed Python script that calculates shortest

distance over water from the GPS points of sample sites (Etherington 2011) in ArcGIS 9.3. A third distance matrix was created that reflected the hierarchical structure of the best AMOVA grouping by using a zero to code for localities within the same group and a one to code for localities in different groups. First, significant correlations were tested for between genetic and geographic distance using AMOVA group membership as a covariate. Then, the correlation between genetic distance and AMOVA grouping was tested using geographic distance as a covariate. Significance was tested with 10,000 random permutations, and the relationships among distances and clusters were plotted.

Results

A total of 625 fish were successfully sequenced at the mitochondrial control region, representing 20 study sites across Indonesia and 13 study sites in the Philippines. When aligned, 129 sites over the amplified 382-bp were polymorphic. There were 393 haplotypes, 308 of which were unique to a single individual. The highest frequency haplotype was shared by 18 individuals.

Phylogenetic Relatedness

The unweighted mean pairwise difference between haplotypes in the minimum spanning tree was 11.090 bp. All haplotypes from Medan and Padang, with the exception of a single individual from Padang, fell within a divergent clade separated from all other haplotypes by 8 mutational steps (Fig. 6a,b). A single individual sampled at Makassar, Sulawesi also fell within this divergent Indian Ocean clade. Regional clustering within the Pacific lineage shows some evidence that the distribution of haplotypes is non-random.

Table 1. Mitochondrial diversity indices for *Caesio cuning*. *n* = number of samples, hap = number of unique haplotypes, *h* = haplotype diversity, Π = nucleotide diversity, θ_s = theta estimated using the number of segregating sites, and Fu's F_S and Fu and Li's *D** = two neutrality statistics.

	Sampling Locality	***n***	**hap**	***h***	Π	θ_s	F_S	***D****
1	*Medan*	20	12	0.921 ± 0.039	0.017 ± 0.009	6.765	-1.641	-1.118
2	*Padang*	22	13	0.918 ± 0.036	0.017 ± 0.009	8.778	-2.168	-2.081*
3	*Anyer*	22	19	0.983 ± 0.021	0.026 ± 0.014	10.973	-7.154*	-0.322
4	*Seribu*	7	7	1.000 ± 0.076	0.024 ± 0.016	9.796	-1.725	-0.565
5	*Karimunjawa*	20	20	1.000 ± 0.016	0.036 ± 0.018	15.503	-10.469*	-0.072
6	*Bali*	26	22	0.982 ± 0.018	0.028 ± 0.015	10.482	-8.891*	-0.239
7	*Lombok*	16	15	0.992 ± 0.025	0.029 ± 0.016	11.452	-5.286*	-0.481
8	*Makassar*	18	18	1.000 ± 0.019	0.027 ± 0.016	13.665	-10.237*	-0.993
9	*Selayar*	20	15	0.942 ± 0.043	0.025 ± 0.014	10.429	-3.034	-0.794
10	*Tawi Tawi*	17	13	0.963 ± 0.033	0.028 ± 0.015	10.944	-1.984	-0.644
11	*Honda Bay*	26	23	0.991 ± 0.013	0.030 ± 0.016	11.793	-10.162*	-0.349
12	*Ulugan Bay*	21	19	0.991 ± 0.018	0.027 ± 0.015	10.562	-8.230*	-0.047
13	*Bolinao*	24	24	1.000 ± 0.012	0.028 ± 0.015	10.712	-16.723*	-0.527
14	*Perez*	25	24	0.997 ± 0.013	0.027 ± 0.014	11.388	-15.200*	-0.415
15	*Romblon*	17	17	1.000 ± 0.020	0.030 ± 0.016	10.649	-9.056*	-0.237
16	*Sorsogon*	19	18	0.994 ± 0.019	0.026 ± 0.014	10.872	-9.019*	-0.369
17	*Negros Occidental*	15	14	0.991 ± 0.028	0.027 ± 0.015	10.457	-5.352*	-0.767
18	*Guimaras*	26	25	0.997 ± 0.012	0.029 ± 0.015	12.579	-15.492*	-1.044
19	*Negros Oriental*	8	8	1.000 ± 0.063	0.031 ± 0.018	12.342	-1.933	-0.609
20	*Balingasag*	21	19	0.990 ± 0.017	0.029 ± 0.015	11.952	-8.981*	-0.446
21	*Dinagat*	46	44	0.998 ± 0.005	0.027 ± 0.014	13.197	-43.847*	-1.489
22	*Davao*	9	9	1.000 ± 0.052	0.026 ± 0.015	10.302	-2.911	-0.533
23	*Manado*	9	8	0.972 ± 0.052	0.026 ± 0.015	10.670	-1.157	-0.849
24	*Halmahera*	12	11	0.985 ± 0.040	0.029 ± 0.016	9.934	-2.627	0.494
25	*Raja Ampat*	13	10	0.949 ± 0.051	0.027 ± 0.015	10.312	-0.918	-0.633
26	*Fak Fak*	11	11	1.000 ± 0.039	0.023 ± 0.014	10.584	-4.636*	-0.797
27	*Sorong*	14	14	1.000 ± 0.027	0.025 ± 0.014	9.434	-6.906*	-0.409
28	*Kaimana*	16	16	1.000 ± 0.022	0.026 ± 0.014	9.644	-8.432*	0.037
29	*Manokwari*	8	8	1.000 ± 0.063	0.031 ± 0.018	12.727	-1.853	-0.436
30	*Windesi*	20	19	0.995 ± 0.018	0.026 ± 0.014	10.429	-9.444*	-0.591
31	*Karei*	13	13	1.000 ± 0.030	0.024 ± 0.013	10.634	-6.112*	-0.765
32	*Yapen*	21	19	0.991 ± 0.018	0.026 ± 0.014	9.728	-8.330*	-0.293
33	*Biak*	43	36	0.991 ±0.007	0.028 ± 0.014	13.174	-24.146*	-1.963*

* denotes significant values of Fu's F_s and Fu and Li's *D** (α=0.05).

Population Structure

Haplotype diversity was high, measuring at or near one for all localities (Table 1). The two sites from Sumatra - Medan and Padang - had lower nucleotide diversity (both 0.017 ± 0.009) compared to all other sites, which had nucleotide diversities ranging from 0.0242 to 0.0356. While high haplotype diversity and low nucleotide diversity could be an indication of recent population expansion, neither of these sites had significantly negative values for Fu's F_S (Table 1). Across all sampled localities, there were only two significant values for Fu and Li's D^* which is more sensitive to the effects of background selection (Fu 1997). However Fu's F_S, which is more sensitive to signatures of demographic expansion and genetic hitchhiking, was significantly negative at 11 of 13 sites in the Philippines and 14 of 20 sites in Indonesia, indicating that the departures from neutrality can be ascribed to one of these two processes (Fu 1997).

The results of a non-hierarchical AMOVA show significant genetic structuring in *Caesio cuning* across the Coral Triangle (Table 2; $\Phi_{ST} = 0.1421$, $p < 0.001$). Pairwise Φ_{ST} values calculated for each pair of sampling localities indicate that Medan and Padang are significantly different from all sites but each other and sites from eastern Indonesia appear more genetically similar to each other than most other sites (tables attached as supplemental).

A spatial analysis of the pairwise Φ_{ST} matrix in Barrier indicates that genetic variance among sites can be partitioned geographically. Bootstrapping analyses reached their highest confidence values when parameters were set to four barriers across the entire dataset (where $n \geq 15$). A barrier between the polygon space of Medan and Padang and all other sites is always the first to be placed by Barrier and carries unanimous bootstrap support (1.00) regardless of number of designated barriers (Fig. 7; barrier a). The second barrier is found in the region of Halmahera

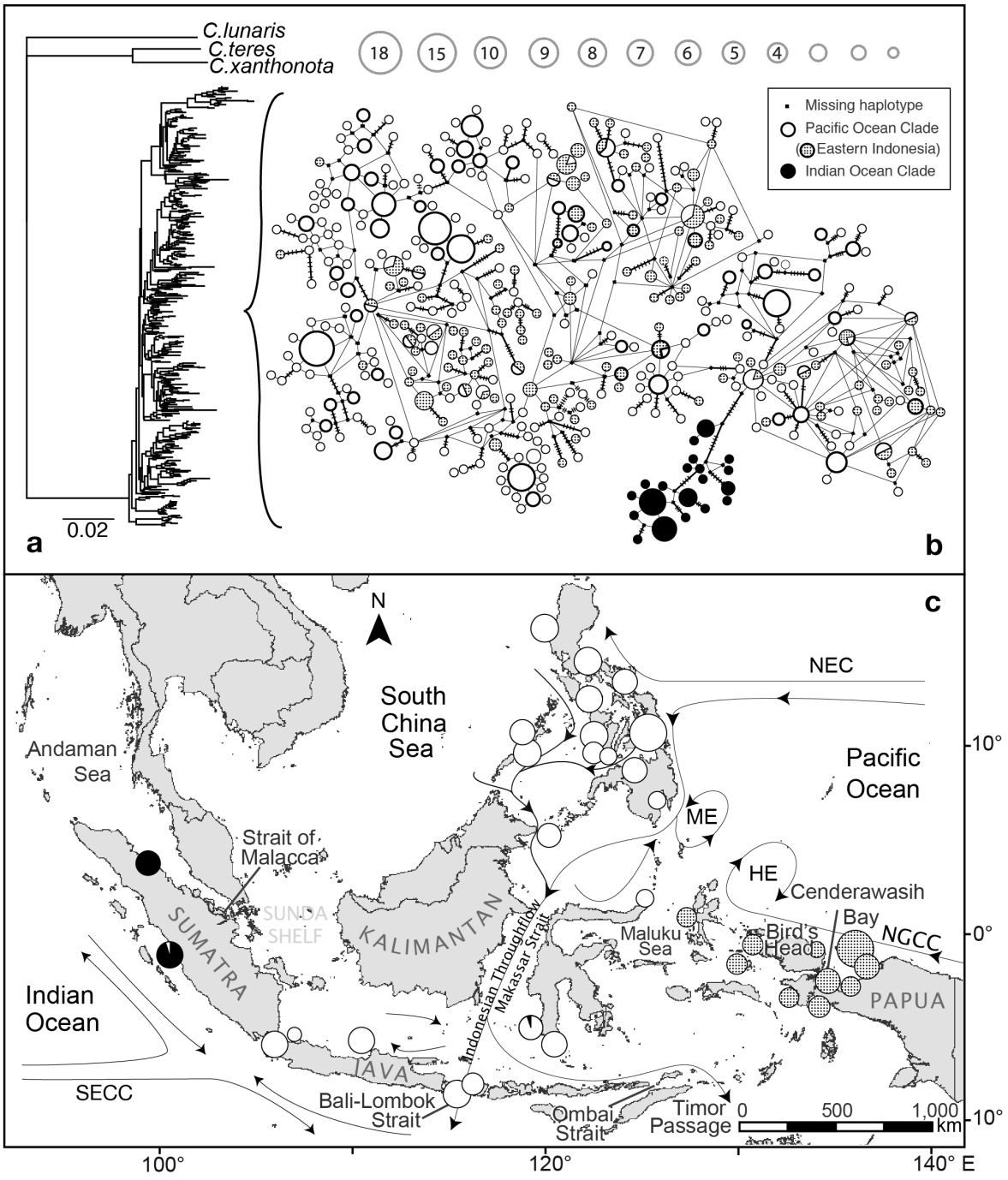


Fig. 6. Results of genetic analyses. a. Neighbor-joining analysis. b. Minimum spanning tree. Gray shading indicates the eastern Indonesian sites within the Pacific Clade. **c.** Geographic distribution of regional genetic structure. Area of circles is relative to total number of individuals sampled at each site. Major oceanographic features include the Northern Equatorial Current (NEC), New Guinea Coastal Current (NGCC), Halmahera Eddy (HE), Mindanao Eddy (ME) and Southern Equatorial Countercurrent (SECC).

Adapted from Ackiss et al. 2013 with permission from Inter-Research Science Center

Table 2. AMOVA summary. Unstandardized results of AMOVA tests with localities where n ≥ 15 using 30,000+ random permutations. Tested partitions are labeled 1-5 corresponding to illustrations in Fig. 5. The first three analyses include both lineages, while the lower three analyses examine genetic structure within the Pacific Clade. K values give the number of groupings tested. P-values ≤ 0.05 indicate significant statistics, and optimal partitions for each group of analyses are bolded. The last column lists pelagic and demersal species that exhibit phylogeographic breaks in mtDNA on which hypotheses for partitioning are based.

	Hypothesis	Sites	Statistic		p	% var	Partition Examples
Both Clades (Indian & Pacific)							
	k = 1	**23**	-	-	-	-	*Decapterus macrosoma* (Borsa 2003)
			-	-	-	-	
			Φ_{ST}	0.1421	0.001	14.21	
1	*k = 2; east vs. west of the Makassar Strait*	**23**	**Φ_{CT}**	0.0258	0.086	2.58	*Decapterus russelli* (Rofristch and Borsa 2009)
			Φ_{SC}	0.1312	0.001	12.78	
			Φ_{ST}	0.1537	0.001	84.64	
2	*k = 2; Western Sumatra vs. all other sites*	**23**	**Φ_{CT}**	**0.4796**	0.004	**47.96**	*Dascyllus trimaculatus* (Leray et al. 2010) *Acanthaster planci* (Vogler et al. 2008) *Tridacna crocea* (Deboer et al. 2008) *Nerita albicilla* (Crandall et al. 2008b)
			Φ_{SC}	0.0189	0.001	0.98	
			Φ_{ST}	0.4894	0.001	51.06	
Pacific Clade							
3	*k = 2; Philippines vs. Indonesia*	**21**	**Φ_{CT}**	0.0091	0.022	0.091	*Hippocampus kuda* (Lourie et al. 2005)
			Φ_{SC}	0.0140	0.001	1.39	
			Φ_{ST}	0.0229	0.001	97.71	
4	*k = 2; central CT vs. eastern Indonesia at Halmahera*	**21**	**Φ_{CT}**	**0.0450**	0.001	**4.50**	*Tridacna crocea* (Deboer et al. 2008) *Haptosquilla glyptocercus* (Barber et al. 2006)
			Φ_{SC}	0.0026	0.273	0.25	
			Φ_{ST}	0.0474	0.001	95.25	
5	*k = 2; central CT vs. eastern Indonesia at Cenderawasih Bay*	**21**	**Φ_{CT}**	0.0420	0.001	4.20	*Haptosquilla pulchella* (Barber et al. 2006) *Tridacna maxima* (Nuryanto and Kochzius 2009) *Protoreaster nodosus* (Crandall et al. 2008a)
			Φ_{SC}	0.0056	0.110	0.54	
			Φ_{ST}	0.0473	0.001	95.26	

Adapted from Ackiss et al. 2013 with permission from Inter-Research Science Center

and the Maluku Sea, which carries the next highest confidence values (0.78 to 0.80; Fig. 7; barrier b). The third barrier was complex and found in the Philippines with the most supported divisions between the southern Philippines and eastern Indonesia (0.49 to 0.60; Fig. 7; barrier c). The fourth barrier divided the Philippines from central Indonesia, but was supported by less than half of the bootstrap replicates (0.44; Fig. 7; barrier d). While the third and fourth barriers partition more variance in our dataset, neither carries strong enough bootstrap support to be viewed with any confidence.

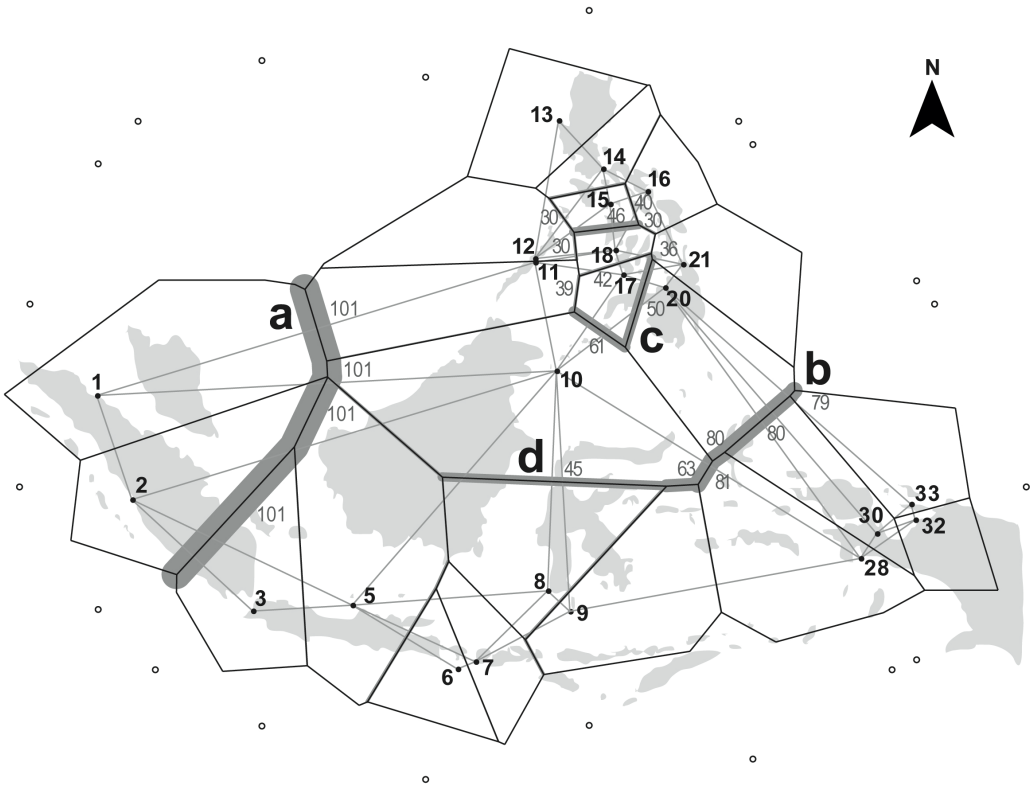


Fig. 7. Barrier analysis. Spatial analysis of sites (n ≥ 15) with four barriers designated (results labelled a-d) and corresponding confidence values labeled in gray (100 bootstrap replicates +1). Black polygons indicate Voronoi tessellation, gray lines indicate Delaunay triangulation. Thickness of barriers is relative to bootstrap support.

Adapted from Ackiss et al. 2013 with permission from Inter-Research Science Center

Hierarchical AMOVAs based on previously detected genetic clusters in other marine organisms showed concordance with the pairwise Φ_{ST} values and Barrier results. Grouping all sites east and west of the Makassar Strait (partition 1, Fig. 5) did not account for a significant portion of the genetic variance among groups measured at this locus whereas grouping the two western Sumatra sites separately from all others (partition 2, Fig. 5) accounted for 47.96% of the genetic variance ($\Phi_{CT} = 0.0258$, $p < 0.086$ vs. $\Phi_{CT} = 0.4796$, $p < 0.004$). Since the variance generated by spatially explicit, divergent clades can overwhelm signatures of structure within a dataset, Medan and Padang were removed from further AMOVA analyses. When the remaining sites from the Pacific Clade were split into a Philippines' group and an Indonesian group (partition 3, Fig. 5), the Φ_{CT} was significant but only explained 0.09% of the variance between groups ($\Phi_{CT} = 0.0091$, $p < 0.022$). Splitting sites east and west of the Maluku Sea (partition 4, Fig. 5) gave the optimal partition and accounted for 4.50% of the variance between groups ($\Phi_{CT} = 0.0450$, $p < 0.001$). When this partition was shifted to Cenderawasih Bay (partition 5, Fig. 5), it remained significant, accounting for slightly less variance between groups ($\Phi_{CT} = 0.0420$, $p < 0.001$). Of the five tested breaks across the Coral Triangle, *C. cuning* exhibits two commonly found in reef-associated, demersal species: a Sunda Shelf break at western Sumatra and a break near the Maluku Sea in eastern Indonesia.

Isolation by Distance

When all localities ($n \geq 15$) were included in the IBD analysis, points associated with the western Sumatran sites Medan and Padang clustered separately from other sites (Graph 1a). To avoid bias arising from their uniquely divergent lineage coupled with their location on the edge of our sampling range, these two localities were excluded from further IBD analyses. When a

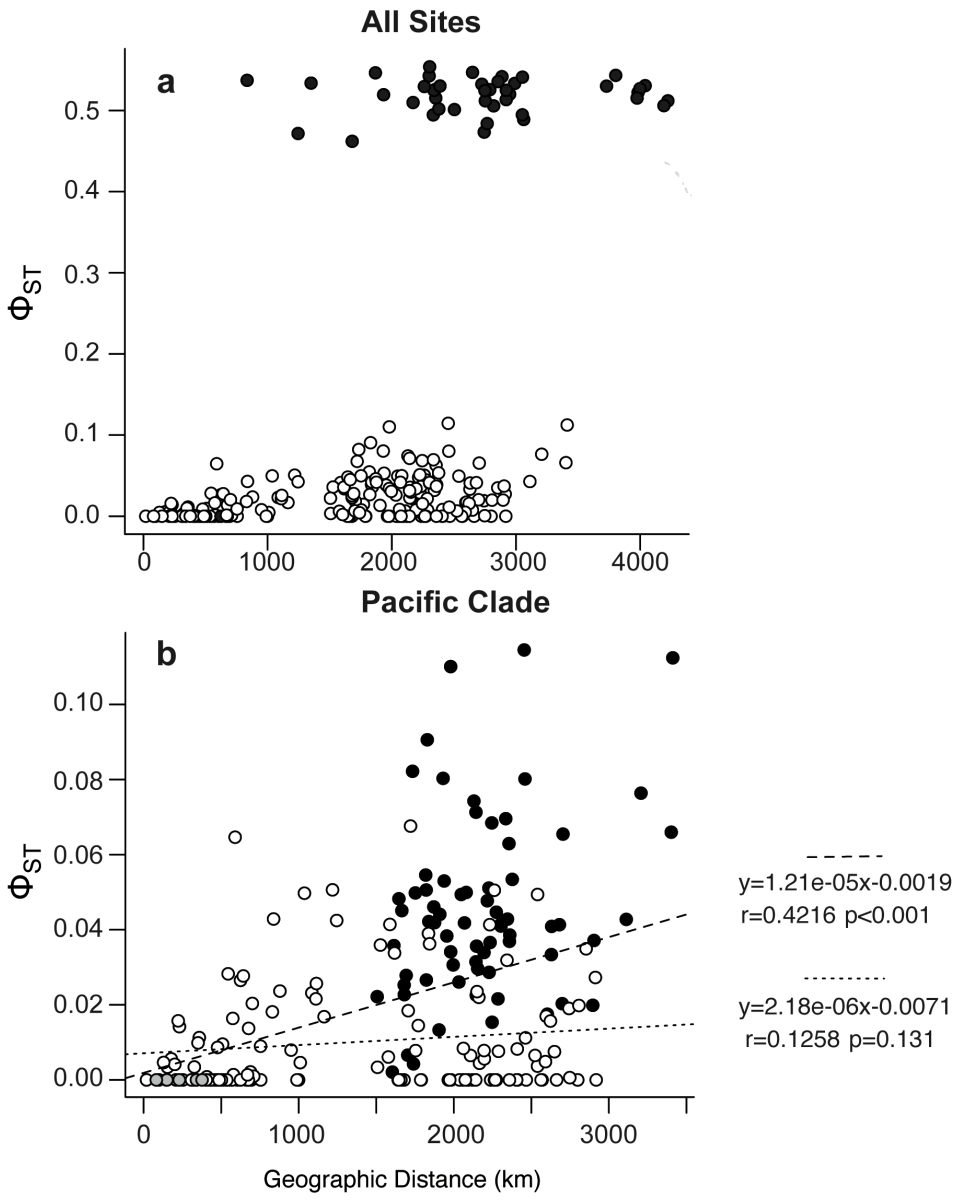
Mantel test was run of only the localities within the Pacific lineage, the results showed that there is a significant indication of IBD within this Pacific lineage (Graph 1b, dashed line). A Z of 8964.2023 and a correlation coefficient (r) of 0.4216 were measured with a corresponding p-value of less than 0.001.

Despite the correlation between genetic and geographic distance, the plot indicated that there were still sites nearly 3000 km apart within the Pacific lineage that exhibited no measurable genetic differences. Since the AMOVA analyses indicate the presence of hierarchical structure, partial Mantel tests were run to determine the nature of the significant correlation measured. A partial Mantel test examining the correlation of geographic distance to pairwise Φ_{ST} while accounting for the optimal AMOVA clusters (central Indonesia and the Philippines vs. sites in the Bird's Head region of Papua) resulted in a non-significant correlation coefficient (r) of 0.1642 ($p < 0.066$). A partial Mantel test examining the correlation of pairwise Φ_{ST} to the location of sites within one or the other of the two optimal sites while accounting for geographic distance resulted in an r of 0.5907 ($p < 0.001$), indicating the hierarchical clustering of sites explains a significant percentage of the variance in the dataset while isolation by distance does not. This is further supported by a Mantel test of only sites within the Philippines and central Indonesia cluster (no Mantel test was run on the eastern Indonesia cluster since all pairwise $\Phi_{ST} = 0$). A Z of 2093.5389 and a correlation coefficient (r) of 0.1258 was measured with a non-significant p-value of 0.131 (Graph 1b, dotted line).

Discussion

Patterns of genetic structure in a mid-water planktivore

Hierarchical genetic analyses revealed two significant regions of genetic structure across



Graph 1. Isolation By Distance. Comparison of pairwise Φ_{ST} to geographic distance for a. all sites with sample sizes greater than 15, showing clustering of Φ_{ST} Medan and Padang (black dots), and b. Pacific Clade only. Black dots are pairwise comparisons between sites belonging to different AMOVA clusters, white dots are comparisons between sites within the Philippines and central Indonesia cluster, and gray dots are comparisons between sites within the eastern Indonesia cluster (all $\Phi_{ST} = 0$). The dashed line is the regression for all sites in the Pacific Clade (significant due to presence of hierarchical structure), and the dotted line is the regression for only sites across the Philippines and central Indonesia (white dots only; non-significant).

Adapted from Ackiss et al. 2013 with permission from Inter-Research Science Center

the Coral Triangle in the coral reef fish, *Caesio cuning*. A sharp genetic break was observed across the Sunda Shelf barrier, echoing patterns reported from a diversity of reef taxa including crown-of-thorns seastars, damselfishes, snappers and snails (Vogler et al. 2008, Drew & Barber 2009, Gaither et al. 2010, Crandall et al. 2008b). Such population divergence across the Sunda shelf is frequently attributed to historical vicariance between Pacific and Indian Ocean populations during Pleistocene low sea level stands (e.g. Barber et al. 2000, Rohfritsch & Borsa 2005, Deboer et al. 2008). In addition, significant departures from neutrality, as measured by Fu's F_S, might indicate the lingering effects of a Pleistocene population expansion onto the Sunda and Sahul Shelves as sea levels rose during the Last Glacial Maximum. Similar departures have been seen in every species examined in this region so far (see Crandall et al. 2012). Shared phylogeographic patterns such as these result from broadly acting physical processes that shape genetic patterns in codistributed taxa (Avise 2000). However, the maintenance of these patterns in modern times, despite the lack of physical isolation, likely results from oceanographic currents or reproductive isolation between the two lineages.

During the northeast monsoon, the Southern Equatorial Counter Current (SECC) bifurcates off the coast of southern Sumatra (Schott & McCreary 2001). During the southwest monsoon, this reverses, and where Sumatra meets Java, a southeastern flow hits a northwesterly flowing current that is driven by the Indonesian Throughflow. Both monsoonal patterns have the potential to create a barrier to continuous gene flow at the site of bifurcation and conjunction (Fig. 6c), potentially reinforcing isolation during periods of lowered sea levels. Support for this hypothesis comes from a recent quantitative analysis using biophysical models coupled with matrix projection (Kool et al. 2011) that predicts the genetic isolation of populations in the Andaman Sea and western Sumatra.

While studies of many reef organisms indicate divergence between Pacific and Indian Ocean populations, only a few have sampled at a scale fine enough to illuminate the extent and location of overlap between these divergent lineages (see Barber et al. 2006, Crandall et al. 2008a,b, Deboer et al. 2008, Nuryanto and Kochzius 2009, Gaither et al. 2011). The overlap between divergent Indian and Pacific Ocean lineages in *Caesio cuning* is surprisingly small for such a potentially mobile fish. Haplotype distributions from our minimum spanning tree indicate very limited gene flow between the northern tip of Java and equatorial Sumatra – a distance of just over 800 km. No landmass or geographical feature separates the waters of Padang (Sumatra) from the two closest sample sites on Java, Anyer and Kepulauan Seribu, yet only a single individual unites the maternal lineages of Padang to these two sites (Fig. 6c). While regional oceanographic patterns could be limiting the genetic connectivity in *C. cuning* across this region, it is notable that across the same geographic range, the anenomefish *Amphiprion ocellaris* shows greater admixture of Indian and Pacific maternal lineages in the Java Sea (Timm & Kochzius 2008), and anenomefishes have a larval dispersal period of only 8 to 12 d (Fautin & Allen 1992) and larvae exhibit natal homing (Jones et al. 2005). Given the limited overlap of the two lineages, reproductive isolation between the clades cannot be ruled out as a possible explanation for the absence of gene flow in this region.

In addition to the phylogeographic break observed at the Sunda shelf, significant limits to genetic exchange were also seen in eastern Indonesia. At first pass, a significant correlation between genetic distance and over-water distance suggests that limits to gene flow in this region might be due a stepping-stone model of gene flow in which nearby localities exchange more migrants than they do with distant localities (Graph 1b). However, the partial Mantel tests clearly show that this appearance of isolation-by-distance is actually an artifact of hierarchical structure

between two genetic clusters, the junction of which is delimited by AMOVA and Barrier (Fig. 5 & 7; Table 2).

This genetic structuring across the Maluku Sea mirrors genetic structure and even pronounced phylogeographic breaks east and west of Halmahera found in two species of giant clam (Deboer et al. 2008, Nuryanto & Kochzius 2009) and 14 species of stomatopods (Barber et al. 2006, Barber et al. 2011), suggesting this region may be important for lineage divergence. While *Caesio cuning* populations on either side of Halmahera are not characterized by distinct clades as is seen in western Indonesia, the minimum spanning tree indicates some non-random, regional clustering of haplotypes. Frequency differences among related haplotypes within the Pacific Ocean clade may be caused by isolation facilitated by two eddies generated at the convergence point of the Northern Equatorial Current and the New Guinea Coastal Current, the Mindanao Eddy and the Halmahera Eddy (Fig. 6c). The Halmahera Eddy has previously been suggested as important for driving lineage divergence in the region of the Maluku Sea (Barber et al. 2006, 2011), however, both eddies direct a significant amount of flow back into the Pacific Ocean, so both may be contributing to genetic isolation observed in population genetic and computer modeling studies (Kool et al. 2011) conducted in this region.

The recovery of multiple regions of significant genetic structure in *Caesio cuning* is somewhat surprising because the high mobility potential of adults could result in genetic admixture, such as the signal of secondary contact seen in migratory *Decapterus macrosoma* (Borsa 2003). However, the concordance of these data with phylogeographic patterns of demersal reef species with larval dispersal as well as with predictions from biophysical models of larval dispersal (Kool et al. 2011) suggests that adult *C. cuning* are site-attached, and that the major avenue of genetic connectivity in *C. cuning* is via larval dispersal. If adults are truly site-

attached, *C. cuning* would be dependent on larval dispersal to maintain gene flow among populations across its range.

Implications for management

As the target of a significant artisanal fishery in the Coral Triangle, *Caesio cuning* is subject to anthropogenic population declines. A study of Sumilon Island in the Philippines documented changes in reef fish density after protective management was removed for a quarter of the island's reefs. Alcala and Russ (1990) measured a 64% decrease in caesionid density after an eighteen-month period of fishing by approximately 100 local fishermen from an adjacent island using hand-paddled canoes. Given that artisanal fishing of caesionids has been shown to cause precipitous drops in local abundance, a better understanding of stock structure is particularly important for the management of *C. cuning*.

The results of this study suggest that *Caesio cuning* populations in the Philippine and Indonesian portions of the Coral Triangle should be best viewed as at least three stocks. However, managing a reef fishery at this scale is complex because these stocks do not conform to national borders. No significant genetic divergence across sites in the Philippines and central Indonesia that are nearly 3000 km apart were seen (see pairwise Φ_{ST} table, Appendix A). This connectivity is likely facilitated by the Indonesian Throughflow, a strong current originating in the Western Pacific that flows between Kalimantan and Sulawesi and empties into the Indian Ocean via three major "chokepoints" – the Bali-Lombok Strait, the Ombai Strait and the Timor Passage (Fig. 6c). Dispersal simulations have predicted a net flow of larvae north to south via this pathway (Kool et al. 2011). The boundaries among stocks in western, central and eastern Indonesia all occur within Indonesian national borders, which potentially simplifies management

planning and authority. However, the absence of genetic structure between the Philippines and central Indonesia implies that the diversity and abundance of larvae produced from Philippine reefs could have an important impact on the sustainability and genetic diversity of reefs of central Indonesia. This interdependence between countries within the Coral Triangle emphasizes the importance of developing informed, multinational management plans such as the Coral Triangle Initiative (http://www.coraltriangleinitiative.org).

Further research should focus on fine scale sources and flow of larvae within areas of restricted gene flow in order to ensure continual replenishment of coral reef resources. In the case of *Caesio cuning*, future work should focus on areas with evidence of severely limited gene flow such as the junction of Sumatra and Java. Determining the nature of the limited overlap between the two mitochondrial clades will be key to proper management design in this region. Mitochondrial genetic studies do not have the power to detect reproductive isolation with certainty, so further study should incorporate bi-parentally inherited nuclear DNA. Multiple independent genetic markers such as SNPs from restriction site-associated sequencing could be applied to extended sampling in this area to detect whether it is cryptic speciation or barriers to genetic connectivity maintaining this break. It is particularly important to identify whether gene flow is restricted, since intense overfishing in such a region could result in temporary local extinctions. Until future research characterizes the exact nature and direction of genetic connectivity across this overlap zone, *C. cuning* can be considered to have a genetic pattern similar to other coral reef dependent vertebrates and invertebrates across the Sunda Shelf.

Acknowledgements

I thank the governments of the Philippines and Indonesia for supporting this research collaboration. I thank the University of the Philippines Marine Science Institute, De La Salle University, Udayana University, A. Juinio-Meñez and N. Mahardika for hosting laboratory work essential to this research. These collaborations were made possible by two grants from NSF: OISE-0730256 to K. E. Carpenter and P. H. Barber and NSF OCE-0349177 to P.H. Barber. Additional funding for collection and laboratory work from 18 of the 20 locations in Indonesia for this project was provided to S. Pardede from the Christensen Fellowship, Dr. T. Parkinson, and the Wildlife Conservation Society. The National Fisheries Research and Development Institute (NFRDI) and the U.S. Peace Corps in the Philippines, and the Indonesian Ministry for Research and Technology (RISTEK) and the Indonesian Institute of Sciences (LIPI) in Indonesia assisted with collection permits and/or field logistics. Sampling in Indonesia was covered under permits 1187/SU/KS/2006, 04239/SU.3/KS/2006, and 0119/FRPSM/VI/2009. This work was aided by G. Allen, G. Batin, J. Drew, M. Erdmann, A. Hanson, D. Pada, R. Rachmawati, J. Raynal, J. Sanciangco, C. Starger, E. Stump, and C. Yusuf.

TESTING THE CENTRAL-PERIPHERAL POPULATION MODEL ALONG THE KUROSHIO CURRENT

Introduction

The most predominant oceanographic feature of the Northwestern Pacific Ocean is the powerful Kuroshio Current. This current originates when the Northern Equatorial Current (NEC), driven by the North Pacific Gyre, hits and bifurcates north and south at approximately 13°N along the eastern coastline of the Philippines (Nitani 1972, Toole et al. 1990). The northern branch of this bifurcation becomes the Kuroshio Current, which flows along the western margin of the Pacific before turning east off the coast of Honshu, Japan to form the North Pacific Current. Like other western boundary currents in the northern hemisphere, the Kuroshio Current provides a steady influx of warm tropical water to sub-tropical areas resulting in the persistence of coral reef organisms at higher than normal latitudes (Veron 1992, Yamano et al. 2001).

Nested in the path of the Kuroshio Current, the Ryukyu Islands of southern Japan mark the northern extent of many ubiquitous Indo-Pacific coral reef species' ranges (Veron and Minchin 1992, Randall 1998). This trend is generally attributed to the steady decrease in sea surface temperature to beyond the thermodynamic threshold of tropical reef organisms at increasingly higher latitudes (Dana 1843, Vaughan 1919, Rosen 1975, Jokiel and Coles 1977, Crossland 1984). Hermatypic corals are obligatory habitat for most reef species and their prey, and at the latitude of the southernmost main island of Japan, shallow water reefs are classified as marginal due to low surface temperature, low aragonite saturation states, and low light (Kleypas et al. 1999). Without supportive habitat and beyond the extent of their fundamental niche, most tropical reef organisms do not succeed in establishing populations above the Ryukyus, though

range expansions are beginning to be documented with rising global sea surface temperatures (Yamano et al. 2011).

Peripheral populations such as those in the Ryukyu Islands can be prone to edge effects that significantly alter a population's genetic characteristics relative to central counterparts. This phenomenon is summarized in the central-peripheral population model (Mayr 1963). Under this model, populations at the center of a species range are contiguous, abundance is regulated by density-dependent factors, and gene flow is multidirectional. Combined, these forces result in larger populations with high genetic diversity. In contrast, peripheral populations are often fragmented and gene flow occurs from a single direction. Isolation, coupled with higher selective pressure in marginal habitats, results in reduced population sizes, low genetic diversity, and genetic divergence (Nei et al. 1975, Gould 2002). Empirical research indicates that the majority of species studied conform at some level to this model. In a review of 35 years of research on central-peripheral population effects, 70% of studies detected genetic differentiation between central and peripheral populations and 64% detected a decline in genetic diversity (Eckert et al 2008). Only 9% of the species represented in the review, however, were marine. Conservation scientists consider species with disjunct peripheral populations and low genetic diversity to have higher economic and evolutionary value than those with continuous peripheral populations (Lesica and Allendorf, 1995, Bunnell et al. 2004). Therefore, it is important to identify marine geographic regions and organisms prone to edge effects.

The Kuroshio Current provides a steady, northerly stream of warm surface water year round through the Ryukyu Islands, but the role this current may play in sustaining continuous central and peripheral populations in reef fishes remains unexamined. The pathway of the Kuroshio is an optimal geographic setting for the central-peripheral population model. The

Kuroshio forms along the tropical coast of the Philippines, traveling past nearly 1000 km of contiguous shoreline habitat before crossing a large expanse of deep water to a string of disjunct, sub-tropical island reef systems. However, it takes only a single effective migrant on average per generation to nullify disruptive genetic drift between two populations (Spieth 1974), and since the Kuroshio Current can reach mean maximum surface velocities of ~1.2 m s^{-1} (~104 km/day) (Yang et al 2015) there is great potential for genetic continuity via dispersive larvae.

As one of the many coral reef species whose northern distribution ends in the Ryukyu Islands, the redbelly, yellowtail fusilier *Caesio cuning* (Bloch 1791) is an ideal species to examine the role of the Kuroshio Current as a source of migrants to peripheral populations. Like the majority of marine organisms, *C. cuning* has a bipartite life history beginning as pelagic larvae and settling on coral reefs as juveniles. In contrast to other fusiliers, such as those in the genera *Pterocaesio* and *Dipterygonotus*, it has been noted that *C. cuning* larvae are primarily 'coastal', being nearly always found over the mid to inner continental shelf (Reader and Leis 1996). Observations of eleven *C. cuning* larvae in situ also showed some evidence that larvae can detect and orient themselves towards reefs (Leis and Ewart, 2003). Adults, while commonly found off the reef during the day, are non-migratory and dependent on reef structure for protection at night. These observations suggest that long distance dispersal in *C. cuning* is unlikely without the presence of a strong oceanographic conduit such as the Kuroshio Current. As an exploited species throughout its range, a better understanding of whether its peripheral populations are continuous or disjunct will be useful for the effective management of *C. cuning*. Here, restriction site-associated DNA (RAD) sequencing (Toonen et al. 2013) was used to examine fine-scale genetic signatures in *C. cuning* from central sites at the start of the Kuroshio Current in the Philippines to peripherally located sites at the northern edge of the species range in

the Ryukyu Islands. A panel of more than 2,500 neutral single nucleotide polymorphisms (SNPs) was used to address two major questions: (1) do we see evidence of the central-peripheral phenomenon from the center to the northern edge of the species range in *C. cuning*? and (2) what role does the Kuroshio Current play in facilitating gene flow to peripheral reef populations?

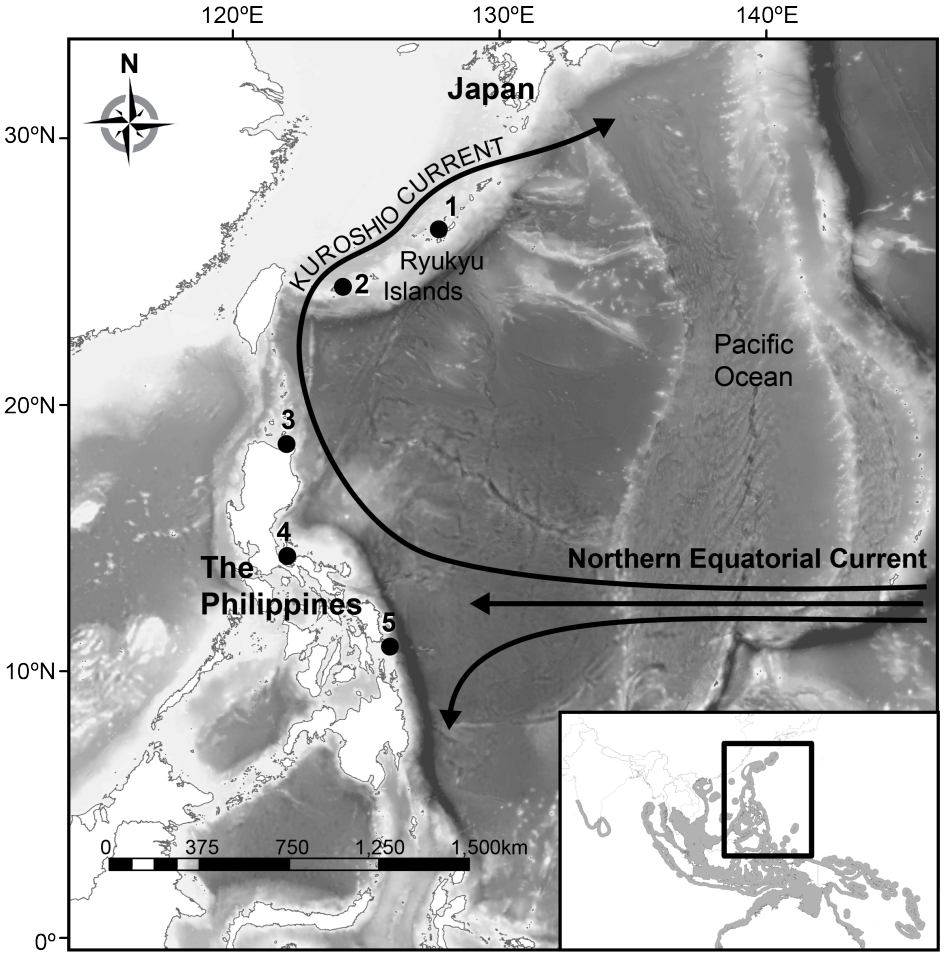


Fig. 8. Collection sites along the Kuroshio Current. Okinawa (1), Ishigaki (2), Santa Ana (3), Atimonan (4), Guiuan (5). INSET: The sampling area (box) within the *Caesio cuning* species range (grey shading).

Methods

Pectoral fin and muscle tissue were collected from 307 fish at regional markets and landings from three sites along the east coast of the Philippines and two sites in the Ryukyu Islands of Japan (Fig. 8). Fishes sampled from markets were confirmed by vendors to be harvested locally. To preserve genetic material, tissues were stored in either 95% ETOH or DNA/RNA Shield™ (Zymo Research). DNA was extracted using E-Z 96® Tissue DNA Kits (Omega Bio-tek, Inc.). Multiple, separate 100 uL elutions were used to filter low from high weight genomic material in order to target the most intact and high quality DNA. Elutions containing the least amount of degraded DNA were quantified with a fluorescence microplate reader, and aliquots containing 100 ng of DNA were cleaned using AMPureXP beads (Beckman Coulter, Inc.) for input into library preparation.

RAD library preparation and sequencing

Restriction site-associated DNA (RAD) libraries for individual samples were prepared following a modified ezRAD protocol (Toonen et al. 2013) with the assistance of the Genomics Core Laboratory at Texas A&M University – Corpus Christi. Genomic DNA was cut at 5'-GATC sites using the isoschizomers *MboI* and *Sau3AI* (New England Biolabs). Cleaned template DNA bound by AMPureXP beads was eluted in 21.5 uL of water and digested overnight with 2.5 uL of CutSmart ® Buffer and 2.5 U of each restriction enzyme in 25 uL reactions. Digested DNA was rebound to beads using a 2X 3M NaCl, 20% PEG solution (Fisher et al. 2011, Faircloth & Glenn 2014) and cleaned before being input into an Illumina TruSeq Nano DNA HT Library Prep Kit at the end repair step. Following one-third volume reactions of kit protocols, DNA was bead size-selected for 350-bp fragments and dual-indexed adaptors

ligated. Libraries were enriched with an 8-cycle PCR reaction using kit reagents and manufacturer's cycling parameters. Ligation was confirmed and libraries quantified using qPCR amplification with a KAPA Library Quant Kit (KAPA Biosystems). To increase coverage and narrow the range of fragments sequenced, the library products were run through a BluePippin (Sage Science) using 2% agarose dye-free cassettes with internal standards for additional size-selection. An initial range of 400-550 bp fragments produced too many unique RAD tags and low sequencing coverage, therefore size selection was reduced to 500-550 bp. Size-selection was confirmed via a Fragment Analyzer™ (Advanced Analytical Technologies) and reduced libraries for 205 individuals were paired-end sequenced (PE 100) on an Illumina HiSeq 2500 and 4000 at 60 libraries per lane.

Read processing and SNP filtering

Raw forward and reverse reads were truncated to 90 bp and quality filtered using the Stacks subprogram process_radtags (Catchen et al. 2013) before being input into the dDocent pipeline (Purtiz et al. 2014) for *de novo* assembly, read mapping, and variant calling. Default parameters were modified based on performance, included setting the CD-HIT sequence similarity parameter © to 0.92 and increasing the minimum base phred score (q) to 20 for an allele to be included in variant calling with Freebayes (v1.0.2-58-g054b257).

Filtering parameters were modified from recommendations with dDocent content (https://github.com/jpuritz/dDocent/tree/master/tutorials). Preliminary filtering to remove samples with high amounts of missing data and erroneous variant calls was done with VCFtools (Danecek et al. 2011). Conditions under which variants were removed were as follows: quality value lower than 30, genotyped in fewer than 95% of individuals, a minor allele frequency of

less than 0.05, a mean coverage depth of less than 5, and sites with more than two alleles. Additional filtering with the program vcflib (https://github.com/ekg/vcflib) removed loci with a heterozygote allele balance (AB) below 0.25 and above 0.75, loci with reads from both strands, loci with large variation in mapping quality among alleles, loci in which the alternate allele was only supported by unpaired reads, and loci with a quality score less than ¼ the depth or a mean depth of well above the majority distribution of depths across loci since high coverage can lead to inflated quality scores or false heterozygotes (Li 2014). Remaining variants were decomposed in order to remove indels.

Genetic structure between populations will generate significant departures from Hardy-Weinberg Equilibrium (HWE) in a global dataset (Wahlund 1928), so SNPs were filtered for population-specific deviation from HWE using a custom perl script (https://github.com/jpuritz/dDocent/blob/master/scripts/filter_hwe_by_pop.pl). Samples with more than 10% missing data after primary filtering were removed. Potential contamination and paralogous loci were identified by two methods. First, heterozygosity for all samples and loci was examined for potential outliers. Samples with high heterozygosity (possible contamination) and loci with heterozygosity >0.6 (possible paralogs) were removed. In addition, samples were haplotyped at all RAD tags remaining after primary filtering (https://github.com/chollenbeck/rad_haplotyper/). Samples with more than the expected number of haplotypes at more than 5% of loci (possible contamination) and RAD tags with more than two samples containing more observed haplotypes than expected (possible paralogs) were removed. Multiple SNPs on a single RAD tag are linked by proximity, so one random SNP per paired-end tag was selected for the final panel using a custom bash script. Dataset conversions

from VCF to GENEPOP and Geste/Bayescan file formats were executed with the Java conversion tool PGDSpider v 2.0.8.3 (Lischer and Excoffier 2012).

Identifying loci under selection

Most population analyses are built on a statistical framework that carries the assumption that loci are neutral. The filtered panel of SNPs was tested for outliers using both frequentist and Bayesian approaches with the programs LOSITAN (Beaumont and Nichols 1996, Antao et al. 2008) and BayeScan v2.1 (Foll and Gaggiotti 2008), respectively. LOSITAN employs an F_{ST}-outlier approach for selection detection, evaluating the relationship between F_{ST} and H_e. LOSITAN was run using the infinite alleles mutation model with parameter settings of "neutral" and forced mean F_{ST} (recommended), 500,000 simulations, confidence interval of 0.95, and a subsample size of 40. BayeScan uses a Bayesian approach to selection detection that implements the multinomial-Dirichlet model and was run with default parameters. Multiple comparisons can lead to type I errors, so a false discovery rate (FDR) correction at α=0.05 was set in both programs, and any loci identified as outliers by either program after correction were removed to produce a neutral panel of SNPs. Both forward and reverse tags associated with outlier loci were run through the megablast search tool in BLASTN v2.6.1 to examine similarity to NCBI archived sequences in the nr/nt and wgs databases (Zhang et al. 2000, Morgulis et al. 2008).

Population genetic analysis

Measures of diversity within populations and pairwise genetic differentiation between populations (F_{ST}) were generated in the program GenoDive v2.0b27 (Meirmans and Van Tienderan 2004). Significance of F_{ST} values was tested using 10,000 permutations, and p-values

were corrected using Benjamini & Hochberg's method of FDR correction (Benjamini and Hochberg 1995). Genetic variability within and among sites was examined with a principal components analysis (PCA) conducted in the R package 'adegenet' (Jombart 2008, Jombart and Ahmed 2011).

The power of different methods to estimate the number of discrete genetic populations within a dataset can vary based on a species' dispersal characteristics and the spatial distribution of sampling (Murphy et al. 2008, Schwartz and McKelvey 2008), therefore the use of multiple methods to detect barriers is recommended (Blair et al. 2012). Two different Bayesian clustering programs were used to detect genetic clusters within the sampled region. The program STRUCTURE v2.3.4 (Pritchard et al. 2000, Falush et al. 2003) was used to identify distinct genetic clusters and assign individuals to populations under the assumption of population admixture and correlated allele frequency. Runs consisted of a burn-in period of 50,000 MCMC iterations followed by 100,000 iterations for inferred K of 1 to 5 and replicated ten times for consistency. Results were collated and likelihood scores visualized in STRUCTURE HARVESTER (Earl and vonHoldt 2012), and the log probability and ΔK statistic (Evanno et al. 2005) were used to determine the most likely number of clusters. Replicate Q-matrices for the optimal K were processed with the Greedy algorithm in CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) in order to produce a single optimal alignment. Individuals were assigned to the cluster with the largest percentage of membership. For comparison, the spatially-explicit clustering program GENELAND (Guillot et al. 2005a,b, Guillot 2008, Guillot and Santos 2009) uses codominant genotypes and sampling coordinates in a Poisson-Voronoi tessellation model to approximate the pattern of population spread across space. Ten replicates of the spatial model in GENELAND v4.0.8 were run through R command line (R Core Team 2015) with the correlated

allele frequency model using the author recommended starting parameters: number of possible clusters (K) set to vary between 1 and 10 with 100,000 Markov Chain Monte Carlo (MCMC) iterations, thinning of 100, burn-in of 200, maximum rate of the Poisson process fixed to 174, and maximum number of nuclei in the Poisson-Voronoi tessellation process fixed to 348. Samples from each location shared the same spatial coordinates to allow individuals with the same coordinates to be assigned to different populations, UTM coordinate uncertainty was set to 1km. The modal K from these initial runs was fixed for 20 additional replicates, and the run with the highest posterior probability was used to assign individuals to clusters.

Distinguishing spatial clustering from clinal genetic differentiation can be particularly challenging since tests for one can be biased by the presence of the other (Schwartz and McKelvey 2008, Guillot et al. 2009, Meirmans 2012). One suggested method of distinguishing a pattern of isolation-by-distance (IBD) from spatial clustering is running partial Mantel tests. A pairwise geographic matrix was generated by calculating the shortest overwater distances between collection sites. Analyses were completed in ArcGIS v10.1 using the Cylindrical Equal Area projected coordinate system and a 500 m cell size (2,500 m^2). Shortest overwater distances were estimated using the cost distance tool, assigning an equal 'cost' to each cell of water and eliminating all paths that go over land. Correlation of geographic distances to pairwise F_{ST} was examined with a basic Mantel test. A third matrix of cluster membership by genetic clusters identified in previous analyses was used for partial Mantel tests. Mantel and partial Mantel tests were run in the R package 'vegan' (Oksanen et al. 2016) and the significance of Mantel statistics was measured with 120 permutations of the matrices.

To examine sites for the presence of edge effects, estimates of contemporary effective population (N_e) size were calculated using a bias-corrected linkage disequilibrium (LD) method

in the program NeEstimator v2.01 (Do et al. 2014). In addition to neutrality, the LD N_e method assumes independence of loci, so to test the effects of possible linkage on the N_e estimates, N_e was calculated with and without pairwise comparisons of $r^2 \leq 0.5$ using custom R scripts (Gruenthal et al. 2014, scripts available from Candy et al. 2015). Coefficients of relatedness (r) for individuals at each site were estimated using seven different estimators in the R package 'related' (Pew et al. 2015). The same pattern of relatedness was found with all estimators, but since likelihood estimators have been shown to be more robust than moment estimators when populations are structured and a large number of polymorphic loci are available (Wang 2011, 2014), only the results of the dyadic and triadic maximum likelihood-based estimators are reported here. One pair of individuals, San_007 and San_008, had an r value of 0.97 which is consistent with the same individual being sampled twice so San_008 was removed from all further analyses (APPENDIX B, Table A).

Levels and direction of migration between sites were examined using Bayesian coalescent analysis in MIGRATE-N v.4.2.14 (Beerli 2009). In addition to panmixia of all sites, three main hypotheses of connectivity in *Caesio cuning* were tested with continuous and stepwise models examining potential migration patterns among sites (Fig. 9). Uni-directional models simulate south to north connectivity influenced primarily by the Kuroshio Current (Fig. 9, Model A and B). Mixed models simulate bi-directional connectivity between the three sites in the Philippines connected with nearly continuous shoreline habitat and migration to the islands of Ishigaki and Okinawa occurring south to north via the Kuroshio Current (Fig. 9, Models C, D, and E). Bi-directional models serve as alternatives simulating both northward and southward migration for all sites (Fig. 9, Model F and G).

To reduce ascertainment bias introduced by an analysis of only polymorphic loci, a dataset containing the first 200 RAD tags genotyped in 100% of 30 individuals from each site was used to test models. Each tag contained combined F and R reads for a length of 182 bp. Starting theta (θ) and mutation-scaled migration (M) priors were generated using a percent value of the prior with maximum distributions of θ set to 0.1 and M set to 5,000. Metropolis-Hastings sampling was done every 100 steps for half a million generations (5,000 recorded steps) with a burn-in period of 1 million steps for 10 replicates. Each replicate used four parallel chains, run with a static heating scheme (1 million, 3, 1.5, 1) and a swapping interval of 10. Model performance was compared using Bayes factors (Beerli and Palczewski 2010)

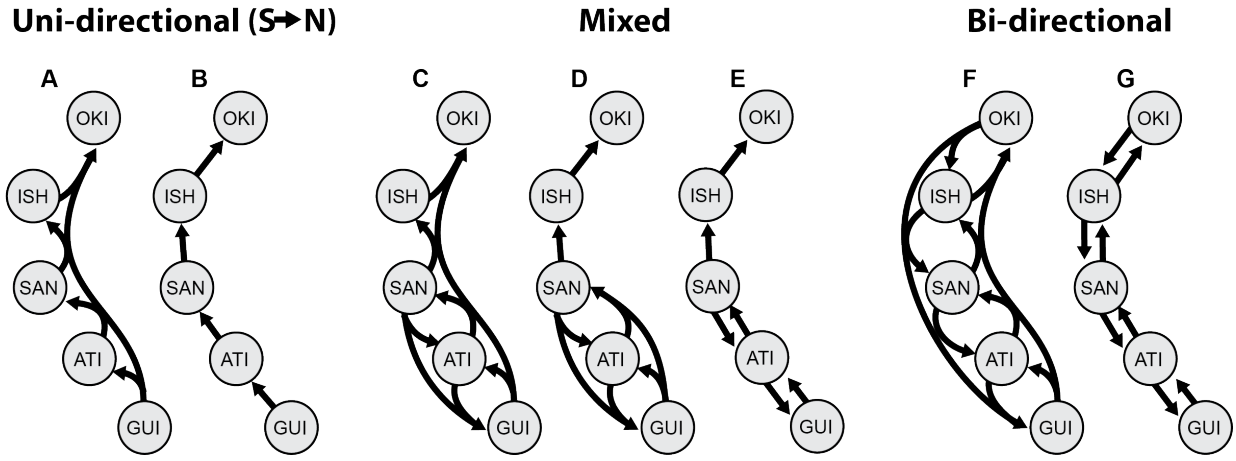


Fig. 9. Models of migration tested in MIGRATE-N. Uni-directional models: A) any site has migration to northern sites and B) nearest neighbor migration. Mixed models: C) full migration among the Philippines sites connected by roughly continuous shoreline with south to north migration to the Ryukyu Islands, D) full migration among Philippines sites with stepping stone migration south to north, and E) stepping stone migration between Philippines sites with south to north stepping stone migration to the Ryukyu Islands. Bi-directional (alternative) models: F) full migration among all sites and G) bi-directional nearest neighbor migration.

Results

A total of 205 sequenced libraries produced enough reads to be included in processing. During filtering, 21 samples were dropped due to low representation across loci, 9 samples were dropped for possible contamination, and 1 sample was a duplicate (Table 3). A total of 2,713 SNPs remained after filtering. LOSITAN and BayeScan identified 11 loci as candidates for positive selection after FDR correction. LOSITAN identified an additional 9 loci and BayeScan an additional 16 loci as outliers. All loci identified as outliers by either method were removed to produce a panel of 2,677 putatively neutral SNPs and a panel of 36 putatively selective SNPs successfully genotyped for 174 individuals across the five locations. Of the 36 outliers, approximately 14% (5 RAD tags) produced significant alignments to known genomic regions in other fishes, including *Dicentrarchus labrax*, *Oplegnathus fasciatus*, *Solea senegalensis*, *Stegastes partitus*, and *Larimichthys crocea*. One tag had a 94% identity match within the CYTH3 gene, and the rest produced 83-100% identity matches to genomic DNA regions as close as 2149bp from known or predicted coding regions. A summary of matches can be found in Table C in Appendix B. All analyses were performed on the neutral panel with the exception of the principal components analysis, which was run with both panels.

Summary statistics by site are reported in Table 3. Observed and expected heterozygosity (H_o / H_e) across all sites was 0.2594 and 0.2605, respectively. Site-specific H_o ranged from 0.2543 in Guiuan to 0.2617 in Okinawa, and the percentage of polymorphic loci observed in each site was greater than 99%. Values of G_{IS} ranged from -0.0119 in Okinawa to 0.0080 in Guiuan. The pairwise F_{ST} values ranged from 0.0004-0.0132. The largest value corresponded to the pairwise genetic differentiation between the two most distant sites, Guiuan and Okinawa. All

but the pairwise F_{ST} between Atimonan and Santa Ana were significant (Table 4). FDR correction did not reduce the p-value at which comparisons remained significant for α=0.05.

Table 3. Site statistics for *Caesio cuning* along the Kuroshio Current. Site code (abbv), number of tissues collected (n), individuals successfully genotyped and used in most analyses (N), percentage of observed alleles represented within site (A_o), observed heterozygosity (H_o), expected heterozygosity (H_e), and the inbreeding coefficient (G_{IS}, an analogue to F_{IS} based on G_{ST}: Nei 1973).

Location	abbv	GPS coordinates	n	N	A_o (%)	H_o	H_e	G_{IS}
Okinawa	OKI	26.47754 N, 127.80504 E	56	33	99.1	0.2617	0.2587	-0.0119
Ishigaki	ISH	24.37328 N, 124.08956 E	74	32	99.6	0.2594	0.2599	0.0019
Santa Ana	SAN	18.53085 N, 122.09347 E	51	39	99.9	0.2603	0.2602	-0.0005
Atimonan	ATI	14.00466 N, 121.92334 E	59	35	99.9	0.2615	0.2612	-0.0015
Guiuan	GUI	10.93783 N, 125.85169 E	67	35	99.7	0.2543	0.2563	0.0080

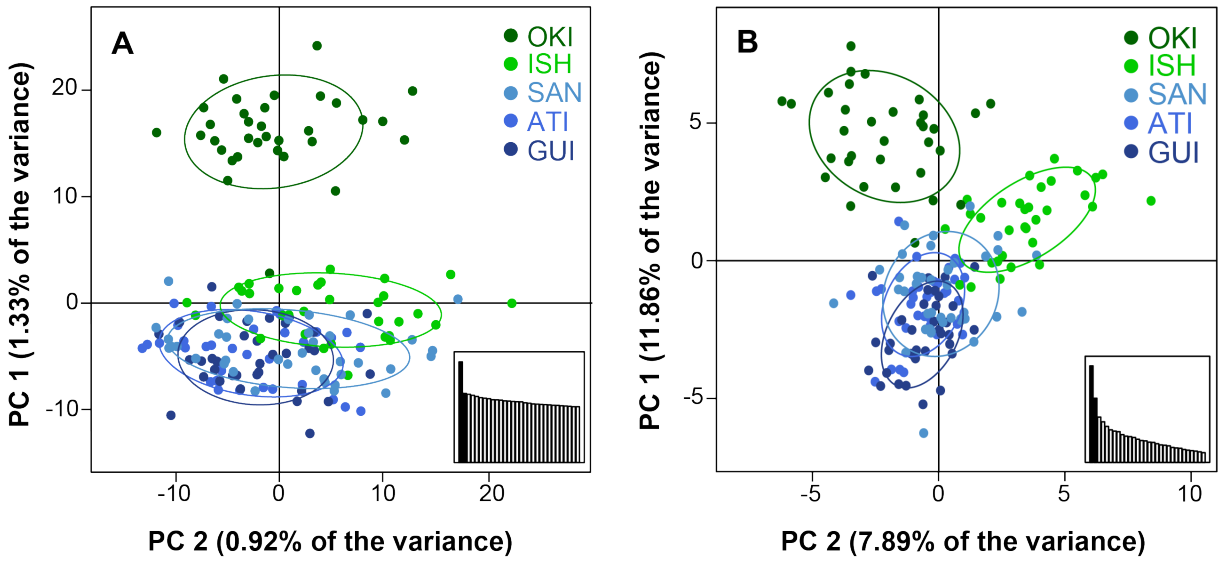
Table 4. Pairwise genetic and geographic distance values. Pairwise F_{ST} values are reported in the lower diagonal, and pairwise geographic distances (in kilometers) are reported in the upper diagonal.

Site	Okinawa	Ishigaki	Santa Ana	Atimonan	Guiuan
OKI	-	504	1080	1570	1718
ISH	0.0106**	-	694	1184	1495
SAN	0.0114**	0.0026**	-	541	997
ATI	0.0122**	0.0041**	0.0004	-	687
GUI	0.0132**	0.0046**	0.0008*	0.0009*	-

**significant, p=0.0001, *significant, p≤0.0355

A PCA using the neutral SNP panel separated samples into two groups along the axis of the first principal component (PC, Graph 2A). PC one accounted for 1.33% of the total variance

measured in the neutral panel of SNPs, which is a small percentage due to the large number of total axes but is more than twice the variation of the average axis. One group contained all but one of the samples from Okinawa and the other contained all other samples. A PCA performed on the 36 SNPs identified as outliers by LOSITAN and BayeScan resulted in three closely-spaced groups – one comprised of the samples from Okinawa, one comprised of the samples from Ishigaki, and one comprised of the Philippine sites (Graph 2B). PC one separates the two Ryukyu Island sites from the Philippines sites and accounts for more than 1/10 of the variance in these SNPs.



Graph 2. Principal components analysis of sites along the Kuroshio Current. The first principal component (PC 1) is graphed along the y-axis, and the second principal component (PC 2) is graphed along the x-axis. Insets contain the first 30 eigenvalues. A: 174 individuals at 2677 neutral SNP loci. B: 174 individuals at 36 SNP loci identified as being under positive selection.

STRUCTURE results over ten replicates were consistent and were similar to those from the principal components analysis (Fig. 10). Both the log-likelihood [L(K)] and ΔK indicated the optimal K=2. All but one individual from Okinawa was assigned to one cluster, and the remaining individual was assigned to a second cluster with all individuals from the other sites (Fig. 10). Initial GENELAND runs resulted in four estimates of K: 2, 3, 4, and 5. All runs with a modal K of 4 and 5 contained ghost populations (a cluster to which no individuals were assigned). Five out of the ten replicates had a modal K of 3, and six of ten replicates assigned individuals to 3 clusters regardless of modal K, so K=3 was used for the runs assigning

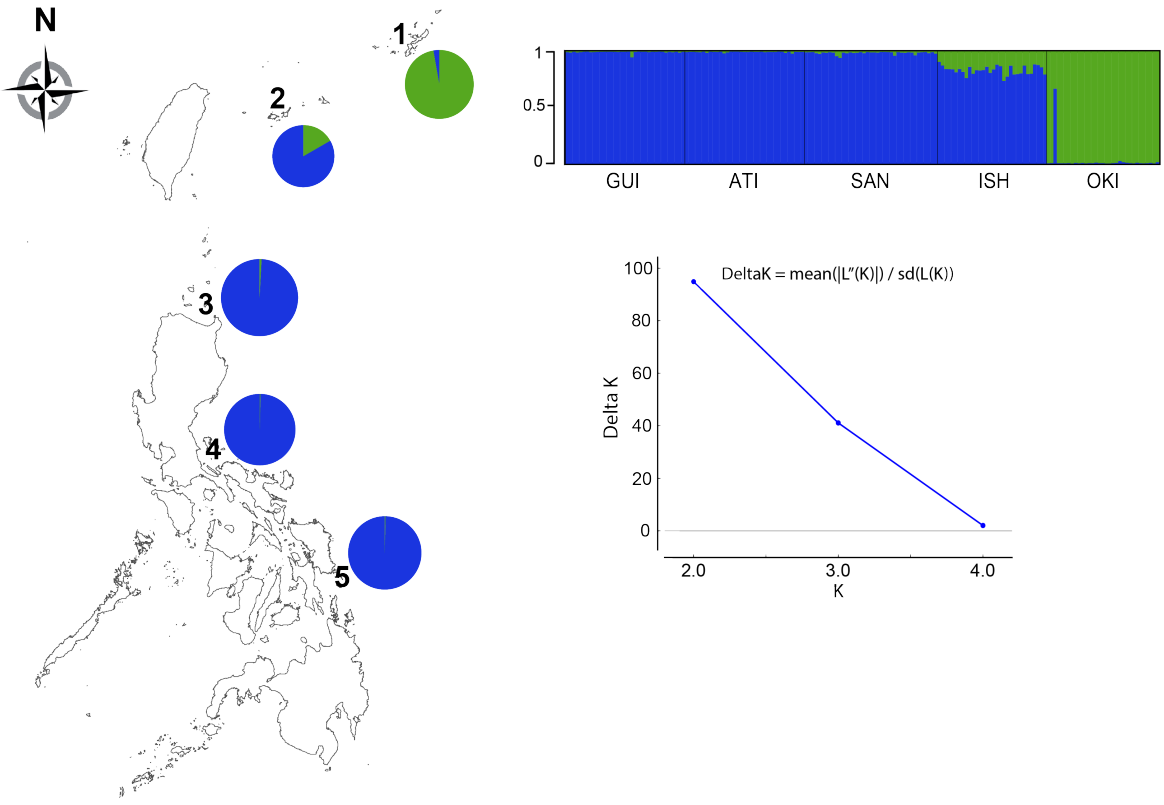


Fig. 10. Results from STRUCTURE. K=2 generated the largest delta K (graph) and log likelihood values. Plot indicates probability of membership to either lineage for each individual by location.

individuals to clusters. The run with the largest log-likelihood score assigned the individuals from the three Philippine sites to one cluster, the individuals from Ishigaki to a second cluster, and the individuals from Okinawa to a third cluster.

A basic Mantel test was significant for correlation between genetic and geographic distance (r=0.5340, p=0.0083). The concurrent PCA and STRUCTURE results were used to inform cluster assignment for partial Mantels. Pairwise distances between Okinawa and all other sites were assigned a 1 (different clusters) and all other population pairwise distances assigned a 0 (same cluster). The partial Mantel test for cluster-corrected IDB resulted in a lower Mantel statistic than the test for IBD-corrected clustering (r=0.7894 and r=0.9778, Table 5), but both were significant, indicating that clinal genetic differentiation and distinct genetic clustering are likely both contributing to structure observed between sites.

Table 5. Mantel and partial Mantel tests for isolation by distance (IBD). A matrix of cluster similarity was used for partial Mantel tests. Cluster assignment was informed by PCA and STRUCTURE results with Okinawa assigned to one cluster and all other sites to another. Significance of the Mantel statistic (r) was estimated with the maximum possible number of permutations (n=120).

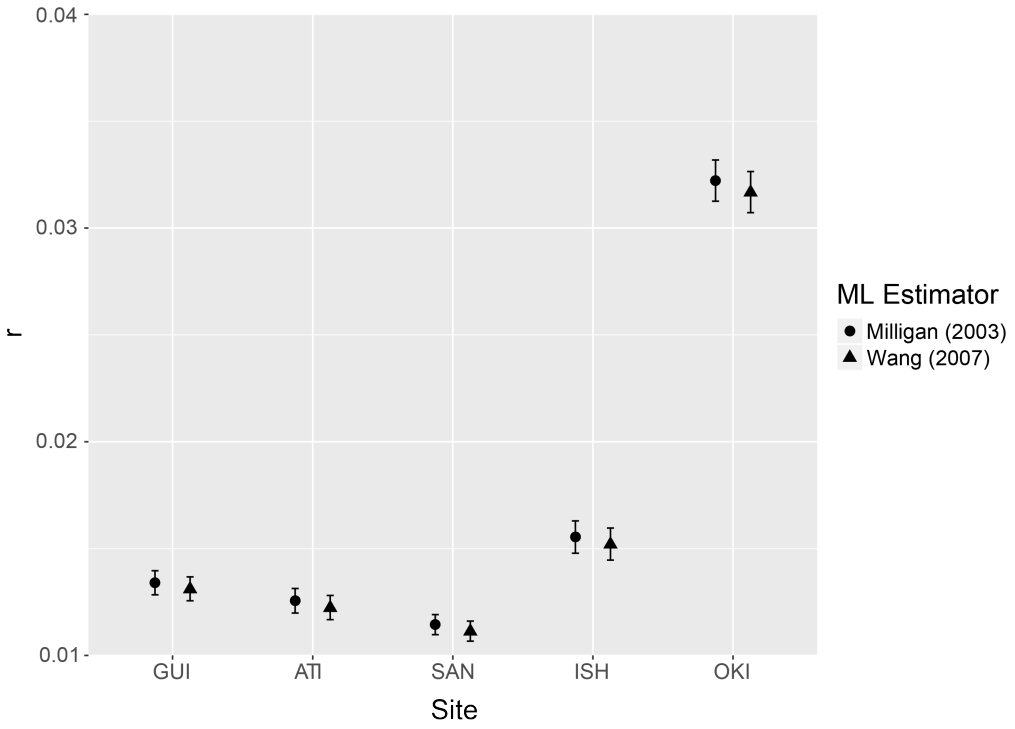
Test	x	y	z	r	p-value
Mantel					
IBD	F_{ST}	geographic distance	-	0.5340	0.0083
Partial Mantel					
IBD	F_{ST}	geographic distance	clusters	0.7894	0.0333
Clusters	F_{ST}	clusters	geographic distance	0.9778	0.0083

Estimates of effective population size (N_e) within sites ranged from 2,260-5,743 in Okinawa to 'infinite' in Atimonan and Guiuan (Table 6). In larger populations ($N > 500$), an 'infinite' estimate is often an indication that the harmonic mean sample size (S) is too small to generate an estimate of N_e using the LDN_e method (Waples and Do 2010, Do et al. 2014). S ranged from 30.4-38.5 for the five sites. When all pairwise locus comparisons were included in estimates ($r^2 < 1$), the smallest N_e was measured in Okinawa, with increasing estimates of N_e in Ishigaki and Santa Ana. Given that sample sizes similar among all sites (S=30.4-38.5), and these were adequate to generate finite estimates of N_e in Okinawa, Ishigaki and Santa Ana but not Atimonan and Guiuan, it follows that the true N_e in these two southern sites is larger than the true N_e of the northern sites. Exclusion of pairwise locus comparisons using a cutoff of $r^2 \leq 0.5$ did affect estimates of N_e, indicating the presence of some linkage between loci.

Table 6. N_e for *Caesio cuning* along the Kuroshio Current. S is the harmonic mean sample size for each population. Effective population size estimates (N_e) and 95% parametric confidence intervals (CI) were generated in R using the r^2 values from the linkage disequilibrium method in NeEstimator with a minor allele frequency cutoff of 0.05. N_e for each location was calculated with all pairwise comparisons ($r^2 < 1$) and using only comparisons with $r^2 \leq 0.5$.

Site	***S***	**N_e (CI)**	
		$r^2 < 1$	**$r^2 \leq 0.5$**
OKI	31.4	2,260 (1,591-3,887)	5,743 (2,783-Inf)
ISH	30.4	4,429 (2,424-25,212)	Inf (14,603-Inf)
SAN	38.5	5,963 (3,315-29,240)	14,298 (4,911-Inf)
ATI	33.2	Inf (6,516-Inf)	Inf (Inf-Inf)
GUI	34.5	Inf (33,851-Inf)	Inf (Inf-Inf)

Pairwise coefficients of relatedness (r) across all individuals ranged from 0.000-0.167. Most pairwise comparisons (99.9%) were deemed unrelated. The relatedness of one pair of individuals from Ishigaki and one pair of individuals from Okinawa was approximately that expected from first cousins (r= 0.167, both estimators, and r=0.158 and 0.159, Wang 2007 and Milligan 2003, respectively). An additional 11 Okinawa-Okinawa pairs, an Okinawa-Ishigaki pair, and an Atimonan-Guiuan pair had relatedness estimates with an upper 95% confidence limit at or above that expected from first cousins. The related pairs found within Ishigaki and Okinawa contributed to an increase in average relatedness measured among individuals within sites in the



Graph 3. Relatedness among individuals within population. Mean coefficients of relatedness (r) generated with two maximum likelihood (ML) estimators in the r package 'related.' Standard error of the mean is denoted by bars.

Ryukyu Islands, with the largest average relatedness measured within Okinawa (Graph 3). Values of mean relatedness among individuals within populations are reported in APPENDIX B (Table B).

Bezier log marginal likelihood scores indicate that the most supported model tested in MIGRATE-N was the full model (Model F) indicating variable levels of migration in any direction is possible. Estimates of mutation-scaled population size (Θ) ranged from 0.0451-0.0376. The largest estimated Θ was from individuals from Guiuan and the smallest was for individuals from Okinawa. Estimated levels of mutation-scaled migration (M) indicate very low levels of immigration between all sites, with the modal number of immigrants estimated to be one effective migrant every 52-58 generations in sites in the Philippines and Ishigaki, and one effective migrant approximately every 63 generations in Okinawa. Bezier log marginal likelihood scores, log Bayes factors, and model ranks are reported in Table 7. Values of Θ and modal number of migrants (N_m) with 95% confidence limits are reported in Table 8.

Table 7. Comparison of models of migration A-G. lmL = log marginal likelihood, LBF = log Bayes factor.

Type	Model	Bezier lmL	LBF	Rank
Uni-directional (S→N)	A	-80241.09	-2129.91	3
	B	-84381.66	-6270.48	8
Mixed	C	-79842.17	-1730.99	2
	D	-82386.99	-4275.81	5
	E	-83116.97	-5005.79	6
Bi-directional	**F**	**-78111.18**	**0.00**	**1**
	G	-83156.79	-5045.61	7
Panmixia	-	-82053.57	-3942.39	4

Table 8. Modal values of Θ and N_m from MIGRATE-N. Values are from the full model (Model F), which had the highest log marginal likelihood and log Bayes factor. Number of migrants (N_m) is estimated from mutation-scaled migration (M).

Site	Code	Θ (95% CI)
GUI	1	0.0451 (0.0408-0.0483)
ATI	2	0.0415 (0.0356-0.0447)
SAN	3	0.0406 (0.0371-0.0455)
ISH	4	0.0425 (0.0372-0.0458)
OKI	5	0.0376 (0.0343-0.0437)

$M_{i \rightarrow j}$	N_m (95% CI)
$M_{2 \rightarrow 1}$	0.0192 (0-1.0463)
$M_{3 \rightarrow 1}$	0.0192 (0-1.0463)
$M_{4 \rightarrow 1}$	0.0192 (0-1.0463)
$M_{5 \rightarrow 1}$	0.0192 (0-1.0463)
$M_{1 \rightarrow 2}$	0.0176 (0-0.8995)
$M_{3 \rightarrow 2}$	0.0176 (0-0.8995)
$M_{4 \rightarrow 2}$	0.0176 (0-0.8995)
$M_{5 \rightarrow 2}$	0.0176 (0-0.8995)
$M_{1 \rightarrow 3}$	0.0172 (0-0.8794)
$M_{2 \rightarrow 3}$	0.0172 (0-0.8794)
$M_{4 \rightarrow 3}$	0.0172 (0-0.8794)
$M_{5 \rightarrow 3}$	0.0172 (0-0.8794)
$M_{1 \rightarrow 4}$	0.0181 (0-0.9212)
$M_{2 \rightarrow 4}$	0.0181 (0-0.9212)
$M_{3 \rightarrow 4}$	0.0181 (0-0.9212)
$M_{5 \rightarrow 4}$	0.0181 (0-0.9212)
$M_{1 \rightarrow 5}$	0.0160 (0-0.8143)
$M_{2 \rightarrow 5}$	0.0160 (0-0.8143)
$M_{3 \rightarrow 5}$	0.0160 (0-0.8143)
$M_{4 \rightarrow 5}$	0.0160 (0-0.8143)

Discussion

Genetic patterns in central - peripheral populations

Despite the proximity of sampled sites to the Kuroshio Current, examination of genetic differentiation indicates disjunct peripheral populations in *Caesio cuning*. Across all sites, both

the PCA and STRUCTURE analysis show strong support for two genetic clusters, one in Okinawa and the other containing the individuals from Ishigaki and the Philippines. A distinct genetic lineage in the Ryukyu Islands is likely sustained in part by divergent selection in marginal environmental conditions at higher latitudes. The largest portion of variance in allele frequencies (PC 1) in loci under selection separated the individuals sampled from Ishigaki and Okinawa from those sampled from the east coast of the Philippines. Putative proximity to coding regions of many of these loci further validates that conditions in peripheral populations maintain a high level of selection, though megablast search results were not conclusive enough to make direct connections to metabolic processes related to thermoregulation.

The GENELAND results supporting three clusters are likely a result of both the scale of geographic distance between sampled sites and deviations from model assumptions. When coordinate uncertainty was changed to 10 km or even 50 km, GENELAND never assigned individuals within sites to different clusters despite both the PCA and STRUCTURE results indicating an individual sampled from Okinawa was most closely aligned to the genetic signature of Ishigaki. These results suggest a spatially-explicit method of clustering is likely to be most powerful when samples are more evenly distributed across the landscape. In addition, the presence of ghost populations in both the first (unknown K) and second (fixed K) MCMC outputs could be an indication of deviations from assumptions underlying the correlated frequency model. In a modification of the original underlying algorithm in GENELAND, the authors note that ghost populations in simulated data are extremely rare and suggest the presence of them can be a clue that data depart from model assumptions (Guillot et al. 2008). A departure such as isolation by distance, which partial mantel tests indicated is present across the five sampled sites, could result in overestimating the number of genetic clusters.

The presence of isolation by distance between sites supports evidence from observations of larval behavior that *C. cuning* do not regularly disperse long distances. This is the first time IBD has been documented in this species and is likely due to the resolution provided by a large SNP dataset. A previous study using mitochondrial DNA found no evidence for IBD though long-distance dispersal was hypothesized to be rare from the low levels of mixing observed between divergent mitochondrial clades (Ackiss et al. 2013).

Though sites in the Philippines belong to the same genetic lineage, pairwise genetic distances were significant between all but the two northern sites, Santa Ana and Atimonan, which may reflect a combination of geographic and oceanographic barriers to connectivity with Guiuan. Continual shoreline connects Santa Ana and Atimonan, but the ~17 km San Bernardino Strait divides the coastline between Atimonan and Guiuan. While there are likely equivalent lengths within the coastline that do not have appropriate reef habitat to support *C. cuning,* the presence of a consistent westward flow from the NEC through San Bernardino Strait (Han et al. 2008) could prevent larval passage across the strait. In addition, the latitudes at which the NEC bifurcation occurs can shift seasonally from 11° – 16.5° N (Qui and Lukas 1996), placing Guiuan within the southward flowing Mindanao Current. The bifurcation of the NEC has been hypothesized to restrict genetic connectivity of reef species along the east coast of the Philippines and genetic structure supporting this has been found in a giant clam species (*Tridacna crocea*) (Ravago-Gotanco et al. 2007).

While cluster analysis clearly differentiates Okinawa from the other sites, the effects of peripheral isolation were noticeable in estimates of relatedness and effective population size in all three of the most northern sites. Only a single pair of samples from the Philippines was related at approximately the first cousin level, and that pair was sampled from two separate sites.

No within-site related individuals were sampled until Ishigaki with a marked increase of within-site relatedness in Okinawa. In addition, finite estimates of N_e were only calculated for the three most northern sites with the smallest N_e calculated in Okinawa. Comparatively, the two most southern sites, Guiuan and Atimonan, only produced lower bound estimates for N_e.

The estimates of contemporary effective population size presented here should be viewed as spatially informative rather than as accurate estimates of true effective population size in these locations. Simulations using both microsatellites and SNPs showed precise estimation of N_e using the LD N_e method can be achieved with small populations ($N_e < 200$), but reliable estimates of N_e from large populations are difficult (Waples and Do 2010). The researchers point out, however, that even the ability to estimate a lower bound for N_e can provide useful information, and in the case of the five sites sampled along the Kuroshio Current, indicates a successively decreasing number of effective breeders contributing to local populations from the center to the periphery of the species range.

The extent of influence by the Kuroshio Current on population connectivity in this region appears highly variable between species. Only a handful of studies examine fine-scale genetic connectivity of tropical organisms from central to peripheral sites in this region. One analysis using of a species of seagrass (*Syringodium isoetifolium*) indicates that *S. isoetifolium* exhibits very similar central-peripheral population patterns to *C. cuning* (Kurokochi et al. 2015). A few support the Kuroshio Current as a homogenizing force across populations in the Philippines and the Ryukyu Islands. A microsatellite analysis of a seastar (*Acanthaster planci*) showed no differentiation between a site in the central Philippines and the Ryukyu Islands and no decrease in genetic diversity (Yasuda et al. 2009). Microsatellite analyses of three species of seagrass (*Enhalus acoroides, Cymodocea serrulata,* and *Cymodocea rotunda*) indicate that recruitment to

the Ryukyus occurs from populations in northeastern Philippines (Nakajima et al. 2014, Arriesgado et al. 2015, Arriesgado et al. 2016). However, reduced genetic diversity was only found in Ryukyu Island populations of *E. acoroides* and *C. serrulata*.

No clear relationship between the Kuroshio Current and patterns of genetic connectivity in *Caesio cuning* can be established from MIGRATE-N results. Analyses indicated there was potential for migration to and from any site, though effective migration rates were overall quite low. Estimated migration and mutation-scaled population sizes were higher in sites in the Philippines and Ishigaki relative to Okinawa, and this provides additional evidence of central-peripheral effects rather than any clear connection to the flow of the Kuroshio. Migration from sites in the Philippines to sites in the Ryukyus can only occur via larval transport in the Kuroshio Current, however migration against this current was equally likely. These results, however, should be viewed cautiously. MIGRATE-N assumes that each locus is independent (Beerli and Palczewski 2010), and the results of the N_e analyses indicate that there is some residual linkage among RAD tags. In addition, these estimates were generated from only a subset of the data. While phased haplotypes were generated using all SNPs surviving filtering on each tag, rare variants and other true polymorphisms whose presence would affect estimates of mutation rate were not included in the RAD tags used for analysis. Before any predictive conclusions can be made regarding the influence of the Kuroshio Current on peripheral populations of marine species in the Ryukyus, more fine-scale studies across diverse taxa are needed.

Implications for management

Even with continually increasing edge effects, there is no evidence that peripheral populations of *Caesio cuning* are currently experiencing detrimental genetic depression. No within- or between-site relatedness was measured at a level greater than first cousin, and G_{IS} in

all sites is close to 0 so there is no indication of inbreeding within any of the sampled populations. However, the divergence, lower N_e, and increased relatedness found in Okinawa do indicate that this population is potentially vulnerable to overfishing. Large declines in abundance of fusiliers (a loss of 64%) were documented in the central Philippines when a site that was previously closed to harvest was opened to artisanal fishing for a period of 18 months (Alcala and Russ 1990). If overharvest of *C. cuning* were to occur in the periphery, the increased genetic differentiation and edge effects puts these populations at risk of a bottleneck event and slower recovery times. Consistent monitoring of catch levels and abundance in Okinawa, Ishigaki, and the northern Philippines is recommended to prevent overharvest of this species in these sites.

Conclusion

This study represents the first comprehensive examination of genetic patterns in a coral reef fish from central to peripheral populations in the path of the Kuroshio Current. Results from a RAD sequencing analysis of samples from five locations indicate that *Caesio cuning* does exhibit disjunct populations towards the northern extent of its range in the Ryukyu Islands. Edge effects were found in the three most northern sites, including increased relatedness in the Ryukyus and decreasing effective population sizes from the tip of the Philippines to the Ryukyus. When considering management of this artisanal fishery species, it would be best to view populations at the apex of the species range as more vulnerable under high fishing pressure.

RAD SEQUENCING ILLUMINATES THE POPULATION DYNAMICS OF A CORAL REEF FISHERY ACROSS THE PHILIPPINE ARCHIPELAGO

Introduction

The Philippine archipelago sits at the apex of the Coral Triangle, an area straddling the tropical Indo-Pacific that contains both the highest levels of marine biodiversity and some of the most threatened reef resources in the world. Coastal development, watershed-based pollution, marine-based pollution and damage, and climate change-induced thermal stress all pose direct threats to coral reefs in this region, but overfishing has been listed as the most pervasive local threat to reefs (Burke et al. 2012). Artisanal fishing, in particular, is the primary cause of reef fish population declines (Newton et al. 2007). In the Philippines, small-scale municipal fisheries account for a greater percentage of the total fish production than commercial fishing and are estimated at more than 1.2 million metric tons per year (BFAR 2015). Declines in these small-scale fisheries have been documented over the past five decades (Muallil et al. 2014), and in some areas, precipitous declines have been seen. The estimated total biomass of demersal coastal fishes in Manila Bay from trawl survey data in 1993 was just 10% of the estimated total biomass from trawl survey data in 1948 (Stobutzki et al. 2006). The Philippines sits at the center of coral reef and shore fish species diversity in the Coral Triangle (Carpenter & Springer 2005, Allen 2008, Sanciangco et al. 2013), making management of these resources vital for the protection of both regional fisheries and biodiversity.

One of the major strategies for combatting both overfishing and biodiversity loss in the Philippines is the development of networks of marine protected areas (MPAs), the effectiveness of which requires well-informed design (McNeill 1994, Gaines et al. 2010). For a network of MPAs to protect and enhance populations, they must be spaced to mutually replenish protected

areas as well as provide recruitment to fished areas (Green et al. 2105). Most coral reef fishes have a bipartite life cycle, which includes a period when larvae are planktonic. Pelagic larval duration (PLD) can be anywhere from days to months (Victor and Wellington 2000, Shanks et al. 2003, Shanks 2009) during which time different species exhibit varying abilities to orient themselves in the water column (Leis and Carson-Ewart 2003, Gerlach et al. 2007). The optimal size, placement, and spacing of MPAs within networks, therefore, are heavily influenced by the range of dispersal potential of both adults and their pelagic larvae (Palumbi 2004, Green et al. 2015). A better understanding of the connectivity of populations of target species and the regional hydrodynamic processes that impact connectivity have been listed as crucial scientific gaps in knowledge for the effective design of marine reserves (Sale et al. 2005).

The redbelly yellowtail fusilier, *Caesio cuning* (Bloch 1791), is an ecologically and economically important reef fish in the Philippines. The annual estimated catch of caesionids in commercial and municipal fisheries is 22,000 metric tons (BAS 2010), though the difficulty of estimating catch data in artisanal fisheries means these data are likely greatly underestimated (Alcala & Russ 2002). Significant population declines have been documented indicating that overfishing can heavily impact this species. A visual census to measure changes in reef fish density at Sumilon Island after protective management was removed for a quarter of the island's reefs found a 64% decrease in caesionid density after an 18 month period of fishing by approximately 100 local fishermen from an adjacent island using hand-paddled canoes (Alcala and Russ 1990). Like the majority of reef fishes, *C. cuning* are broadcast spawners with pelagic larvae (Carpenter 1988) so there is great potential for regional hydrodynamic processes to influence the movement of this species. The PLD in *C. cuning* is unknown, however, making tools such as genetic analysis crucial to estimating the dispersal potential of the species.

The complex topography of the Philippine archipelago is characterized by several prominent oceanographic features with the potential to impact population connectivity in *Caesio cuning*, but the ability to detect such effects is dependent on using sufficiently powerful genetic markers. One example is the bifurcation of the Northern Equatorial Current (NEC) between the latitudes of 11° – 16.5° N on the east coast of the Philippines (Fig. 11, Nitani 1972, Toole et al. 1990, Qui and Lukas 1996). A previous population genetic analysis of *C. cuning* across 13 sites in the Philippines using mitochondrial DNA did not detect any significant genetic differentiation, including along the eastern seaboard (Ackiss et al. 2013). However in this dissertation, examination of population structure along the Kuroshio Current uncovered significant north-south differentiation in *C. cuning* in three sites across the region of the NEC bifurcation using thousands of single nucleotide polymorphism (SNP) loci from a restriction site-associated DNA (RAD) sequencing approach (Ackiss et al. in prep).

This study examines the relationship between four additional geographic and oceanographic features on *Caesio cuning* population structure across the Philippines (Fig. 11) : 1) The Sulu Sea Throughflow (SST); this strong, southerly current bifurcates the islands east and west of the Sulu Sea before exiting the Sibutu Passage. Preliminary evidence from mitochondrial DNA in several organisms including a seahorse, three species of giant clam, and a damselfish suggests that the SST is a barrier to gene flow across the Sulu Sea (Lourie et al. 2005, Deboer et al. 2014, Raynal et al. 2014). 2) Kuroshio Current intrusion through the Luzon Strait; the Kuroshio Current is a powerful western boundary current that reaches mean maximum surface velocities of ~1.2 m s^{-1} (~104 km/day; Yang et al 2015). Kuroshio penetration into the South China Sea through the Luzon Strait occurs seasonally with the northeast-southwest monsoon (Metzger and Hurlbert 2001), and may provide a pathway for connectivity between sites on the

east coast of the Philippines and sites on the west. 3) The Surigao Strait and Bohol Jet; the Surigao Strait is approximately 18 km wide with a sill depth of 63 m and is one of two deep water points of connection between the Western Pacific Ocean and inland Philippine Seas (Hurlbert et al. 2011). It is also the starting point for the Bohol Jet (Hurlbert et al. 2011, Gordon et al. 2011), a predominant southwestward surface current that flows through the northern Bohol Sea with a mean surface velocity of 0.56 m s^{-1} (~48 km/day). The relative speed of this westward

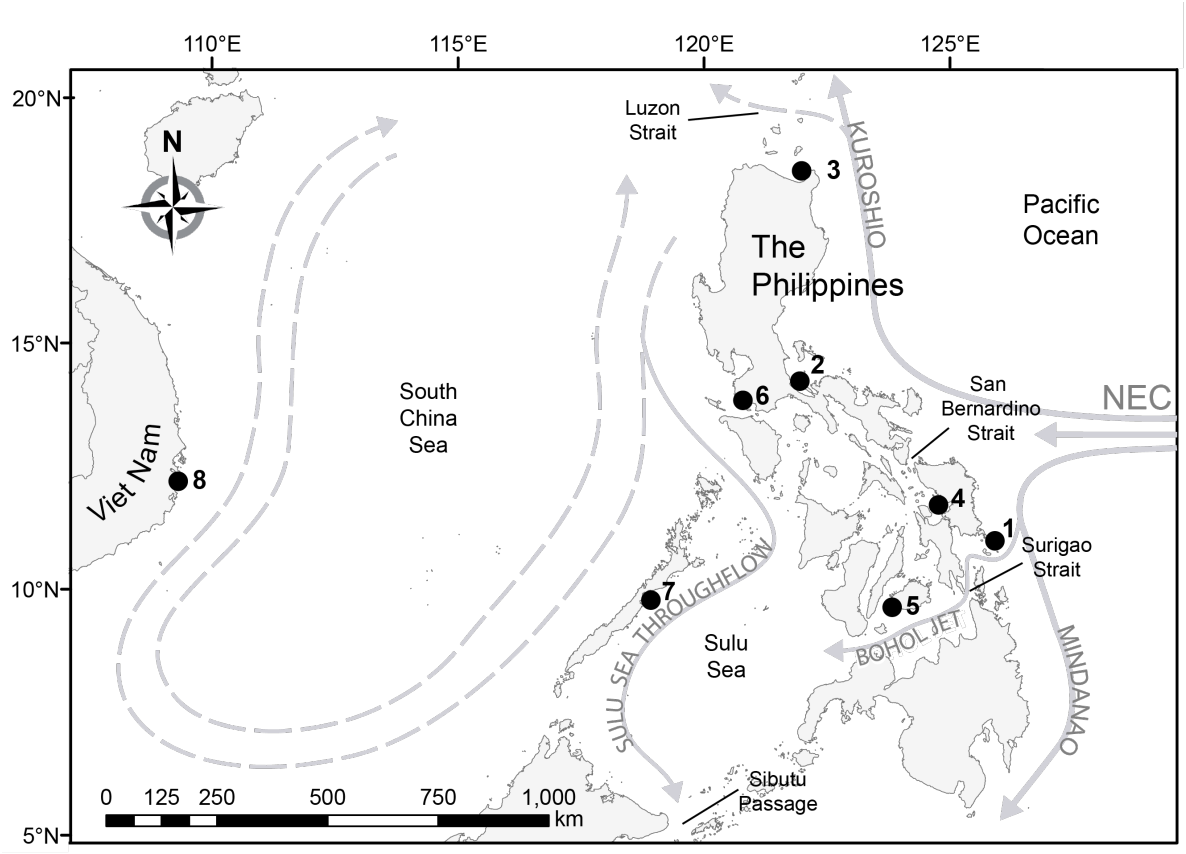


Fig. 11. Map of seven sampling sites in the Philippines with an outgroup from Viet Nam. 1: Guiuan, 2: Atimonan, 3: Santa Ana, 4: Catbalogan, 5: Bohol (Maribojoc), 6: Bauan, 7: Palawan (Honda Bay), 8: Nha Trang. Major currents are marked with solid lines, including the Northern Equatorial Current (NEC) and its offshoots the Kuroshio Current and the Mindanao Current. Major seasonal currents are marked with dashed lines.

current makes this region a potential area for high levels of connectivity. 4) The San Bernardino Strait; this strait is the second of two major points of entry into the central Philippine islands from the Western Pacific. The San Bernardino Strait is approximately 17 km wide with an estimated sill depth of 23 m (Hurlbert et al. 2011, Gordon et al. 2011). Low sill depth reduces current flow and potential connectivity between Pacific Ocean and central Philippine seas populations. The purpose of this study is to test the genetic connectivity between populations of *C. cuning* within the Philippines and across the South China Sea using RAD sequencing with the goal of gaining better insight into the processes that impact gene flow in this species.

Methods

Pectoral fin and muscle tissue were collected from 458 fish at regional markets and landings from seven sites in the Philippines. An additional 32 tissue samples from Nha Trang, Viet Nam were processed as an outgroup (Fig. 11). All fishes sampled from markets were confirmed by vendors to be harvested locally. To preserve genetic material, tissues were stored in either 95% ETOH or DNA/RNA Shield™ (Zymo Research). DNA was extracted using E-Z 96® Tissue DNA Kits (Omega Bio-tek, Inc.). Four to five separate 100 uL elutions were used to filter low from high weight genomic material in order to target the most intact and high quality DNA. Elutions containing the highest quality DNA were quantified with a fluorescence microplate reader, and aliquots of 100 ng of DNA were bound to AMPureXP beads (Beckman Coulter, Inc.) and washed with 80% ethanol before input into library preparation.

RAD library preparation and sequencing

Restriction site-associated DNA (RAD) libraries for individual samples were prepared following a modified ezRAD protocol (Toonen et al. 2013) with the assistance of the Genomics

Core Laboratory at Texas A&M University – Corpus Christi. Cleaned template DNA bound by AMPureXP beads was eluted in 21.5 uL of water and digested using the isoschizomers *MboI* and *Sau3AI* (New England Biolabs) overnight with 2.5 uL of CutSmart ® Buffer and 2.5 U of each restriction enzyme in 25 uL reactions. Digested DNA was rebound to beads using a 2X 3M NaCl, 20% PEG solution (Fisher et al. 2011, Faircloth & Glenn 2014) and washed with ethanol before being input into an Illumina TruSeq Nano DNA HT Library Prep Kit at the end repair step. Following one-third volume reactions of kit protocols, DNA was bead size-selected for 350-bp fragments and dual-indexed adaptors ligated. Libraries were enriched with an 8-cycle PCR reaction using kit reagents and manufacturer's cycling parameters. Ligation was confirmed and libraries quantified using qPCR amplification with a KAPA Library Quant Kit (KAPA Biosystems). To increase coverage and narrow the range of fragments sequenced, the library products were run through a BluePippin (Sage Science) using 2% agarose dye-free cassettes with internal standards for a size-selection of 500-550 bp. Size-selection was confirmed via a Fragment Analyzer™ (Advanced Analytical Technologies) and reduced libraries for 315 individuals paired-end sequenced (PE 100) on an Illumina HiSeq 2500 and 4000 at 60 libraries per lane.

Read processing and SNP filtering

Raw forward and reverse reads were truncated to 90 bp and quality filtered using the Stacks subprogram process_radtags (Catchen et al. 2013) before being input into the dDocent pipeline (Purtiz et al. 2014) for *de novo* assembly, read mapping, and variant calling. Default parameters were modified as required, included setting the CD-HIT sequence similarity

parameter (c) to 0.92 and increasing the minimum base phred score (q) to 20 for an allele to be included in variant calling with Freebayes (v1.0.2-58-g054b257).

Filtering parameters were modified from recommendations with dDocent content (https://github.com/jpuritz/dDocent/tree/master/tutorials). Preliminary filtering to remove samples with high amounts of missing data and erroneous variant calls was done with VCFtools (Danecek et al. 2011). Conditions under which variants were removed were as follows: quality value lower than 30, genotyped in fewer than 95% of individuals, a minor allele frequency of less than 0.05, a mean coverage depth of less than 5, and sites with more than two alleles. Additional filtering with the program vcflib (https://github.com/ekg/vcflib) removed loci with a heterozygote allele balance (AB) below 0.25 and above 0.75, loci with reads from both strands, loci with large variation in mapping quality among alleles, loci in which the alternate allele was only supported by unpaired reads, and loci with a quality score less than ¼ the depth or a mean depth of well above the majority distribution of depths across loci since high coverage can lead to inflated quality scores or false heterozygotes (Li 2014). Surviving variants were decomposed in order to remove indels.

Genetic structure between populations will generate significant departures from Hardy-Weinberg Equilibrium (HWE) in a global dataset (Wahlund 1928), so SNPs were filtered for population-specific deviation from HWE using a custom perl script (https://github.com/jpuritz/dDocent/blob/master/scripts/filter_hwe_by_pop.pl). Potential contamination and paralogous loci were identified by two methods. First, heterozygosity for all samples and loci was examined for outliers. Samples with high heterozygosity (possible contamination) and loci with heterozygosity >0.6 (possible paralogs) were removed. In addition, samples were haplotyped at all RAD tags surviving primary filtering

(https://github.com/chollenbeck/rad_haplotyper/). Samples with more than the expected number of haplotypes at more than 5% of loci (possible contamination) and RAD tags with more than two samples containing more observed haplotypes than expected (possible paralogs) were removed. Multiple SNPs on a single RAD tag are linked by proximity, so one random SNP per paired-end tag was selected for the final panel. Dataset conversions from VCF to GENEPOP and Geste/Bayescan file formats were executed with the Java conversion tool PGDSpider v 2.0.8.3 (Lischer and Excoffier 2012).

Identifying loci under selection

Loci under divergent or balancing selection can bias analyses of gene flow. The filtered panel of SNPs was tested for outliers via two different approaches with the programs LOSITAN (Beaumont and Nichols 1996, Antao et al. 2008) and BayeScan v2.1 (Foll and Gaggiotti 2008), respectively. LOSITAN employs an F_{ST}-outlier approach for selection detection, evaluating the relationship between F_{ST} and H_e. LOSITAN was run using the infinite alleles mutation model with parameter settings of "neutral" and forced mean F_{ST} (recommended), 500,000 simulations, confidence interval of 0.95, and a subsample size of 30. BayeScan uses a Bayesian approach to selection detection that implements the multinomial-Dirichlet model and was run with default parameters. Multiple comparisons can lead to type I errors, so a false discovery rate (FDR) correction at α=0.05 was set in both programs, and any loci identified as outliers by either program after correction were removed to produce a neutral panel of SNPs.

Population genetic analysis

Estimates of genetic diversity and inbreeding within populations as well as pairwise

genetic differentiation between populations (F_{ST}) were generated in the program GenoDive v2.0b27 (Meirmans and Van Tienderan 2004). Significance of F_{ST} values was tested using 10,000 permutations, and p-values were corrected using Benjamini & Hochberg's method of FDR correction (Benjamini and Hochberg 1995). Estimates of contemporary effective population (N_e) size were calculated using a bias-corrected linkage disequilibrium (LD) method in the program NeEstimator v2.01 (Do et al. 2014). In addition to neutrality, the LD N_e method assumes independence of loci, so to test the effects of possible linkage on the N_e estimates, N_e was calculated with and without pairwise comparisons of $r^2 \leq 0.5$ using custom R scripts (Gruenthal et al. 2014, scripts available from Candy et al. 2015). Coefficients of relatedness (r) for individuals at each site were estimated using the R package 'related' (Pew et al. 2015). Since likelihood estimators have been shown to be more robust than moment estimators when populations are structured and a large number of polymorphic loci are available (Wang 2011, 2014), the results of the dyadic and triadic maximum likelihood-based estimators are reported here.

Genetic variability within and among sites was examined with a principal components analysis (PCA) conducted in the R package 'adegenet' (Jombart 2008, Jombart and Ahmed 2011). The program STRUCTURE v2.3.4 (Pritchard et al. 2000, Falush et al. 2003) was used to identify distinct genetic clusters and assign individuals to populations under the assumption of population admixture and correlated allele frequency. Runs consisted of a burn-in period of 50,000 MCMC iterations followed by 100,000 iterations for inferred K of 1 to 5 and replicated ten times for consistency. Results were processed and likelihood scores visualized in STRUCTURE HARVESTER (Earl and vonHoldt 2012), and the log probability and ΔK statistic (Evanno et al. 2005) were used to determine the most likely number of clusters. Replicate Q-

matrices for the optimal K were processed with the Greedy algorithm in CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) in order to produce a single optimal alignment. Individuals were assigned to the cluster with the largest percentage of membership.

To examine the impact of the spatial spread of sites on genetic differentiation, tests for isolation-by-distance (IBD) were run with pairwise F_{ST} and geographic distance matrices. A pairwise geographic matrix was generated by calculating overwater distances between collection sites in ArcGIS v10.1 using the Cylindrical Equal Area projected coordinate system and a 500 m cell size (2,500 m^2). Shortest overwater distances were estimated using the cost distance tool, assigning an equal 'cost' to each cell of water and eliminating all paths that go over land. Correlation of geographic distances to pairwise F_{ST} was examined with a basic Mantel test in the R package 'vegan' (Oksanen et al. 2016). Partial Mantel tests are recommended as a means to distinguish IBD from heirarchical clustering which can bias tests for IBD (Meirmans 2012), so a third matrix of cluster membership by genetic clusters identified in previous analyses was used for partial Mantel tests. The significance of Mantel and partial Mantel statistics was measured with 719 and 999 permutations of the matrices, respectively.

Levels and direction of migration between sites were examined using Bayesian coalescent analysis in MIGRATE-N v.4.2.14 (Beerli 2009). Analyses focused on the four regions of interest: the Luzon Strait, the Sulu Sea, the San Bernardino Strait, the central Philippines and Bohol Jet. Subsets of three to four sites were chosen to test migration in each region to reduce run time and make sampling more tractable by limiting the number of potential model parameters. In addition to panmixia and the full model (all sites have the potential to exchange migrants with all other sites), the five most biologically appropriate models postulated from both

information on the predominant geographic and oceanographic features in each area and results of cluster analyses and pairwise genetic distances were tested.

To reduce ascertainment bias introduced by an analysis of only polymorphic loci, 200 RAD tags genotyped in 100% of individuals were used to test models for each subset of locations. Each tag contained combined F and R reads for a length of 182 bp. Starting theta (θ) and mutation-scaled migration (M) priors were generated from a percentage of the priors with maximum distributions of θ set to 0.1 and M set to 40,000. Metropolis-Hastings sampling was done every 1000 steps for 1 million generations (1,000 recorded steps) with a burn-in period of 1 million steps for ten replicates. Each replicate used four parallel chains run with a static heating scheme (1 million, 3, 1.5, 1) and a swapping interval of 10. Model performance was compared using Bayes factors (Beerli and Palczewski 2010).

Results

A total of 315 libraries were sequenced. During filtering 63 samples were dropped due to low read counts, a third of which belonged to Nha Trang, and 3 additional samples were dropped for possible contamination. After FDR corrections, 20 loci were identified by both LOSITAN and BayeScan as candidates for positive selection. LOSITAN identified an additional 12 loci and BayeScan an additional 2 loci under positive selection. All 34 loci identified as outliers were removed, and a final panel of 2,030 putatively neutral SNPs for the remaining 249 individuals was used for analysis.

Observed and expected heterozygosity ranged from 0.2493-0.2681 and 0.2499-0.2632, respectively (Table 9). Only 10 individuals from Nha Trang generated enough reads to be successfully genotyped, resulting in a lower percentage of observed alleles at this location. This

likely also contributes to the lower level of heterozygosity measured at this site. Values of G_{IS} for all sites were not largely deviant from 0, suggesting that inbreeding is not occurring in sampled locations. After FDR correction, a large portion of pairwise genetic distances remained significant (Table 10). Nha Trang and Puerto Princesa were significantly different from all other sites including each other with F_{ST} values ranging from 0.0055-0.0249. The two sites sampled on the northeast coast of the Philippines, Santa Ana and Atimonan, were significantly different from all other sites except each other. Sites within the central Philippines and Guiuan were not significantly different from each other with the exception of Bauan and Catbalogan.

Table 9. Statistics for sampled *Caesio cuning* populations. Site code (abbv), number of tissues collected (n), individuals successfully genotyped (N), percentage of observed alleles represented within site (A_o), observed heterozygosity (H_o), expected heterozygosity (H_e), and the inbreeding coefficient (G_{IS}) an analogue to F_{IS} based on G_{ST}: Nei 1973

Sampling Locality	**abbv**	**n**	**N**	**GPS Coordinates**	A_o **(%)**	H_o	H_e	G_{IS}
Santa Ana	SAN	51	39	18.53085 N, 122.09347 E	99.9	0.2681	0.2620	-0.0234
Atimonan	ATI	59	36	14.00466 N, 121.92334 E	99.8	0.2676	0.2632	-0.0165
Bauan	BAU	48	40	13.90696 N, 120.80441 E	99.9	0.2629	0.2611	-0.0068
Bohol	BOH	110	31	9.67222 N, 123.81573 E	99.6	0.2553	0.2568	0.0059
Catbalogan	CAT	52	28	11.75048 N, 124.84794 E	99.6	0.2615	0.2613	-0.0007
Guiuan	GUI	72	34	10.93783 N, 125.85169 E	99.6	0.2611	0.2585	-0.0098
Puerto Princesa	PP	66	31	9.81494 N, 118.82343 E	99.4	0.2587	0.2598	0.0030
Nha Trang (Viet Nam)	NT	32	10	12.22738 N, 109.21227 E	91.1	0.2493	0.2499	0.0107

Table 10. Pairwise comparisons. F_{ST} values (lower diagonal) and geographic distances in kilometers (upper diagonal). Significant F_{ST} values based on an FDR adjusted α= 0.0429 are in BOLD.

Site	**SAN**	**ATI**	**BAU**	**BOH**	**CAT**	**GUI**	**PP**	**NT**
SAN	-	541	788	1082	864	997	1210	1723
ATI	0.0007	-	871	762	544	687	1158	2132
BAU	**0.0019****	**0.0011***	-	682	555	714	555	1365
BOH	**0.0012***	**0.0014***	0.0002	-	306	304	619	1798
CAT	**0.0013***	**0.0017***	**0.0015***	0.0007	-	184	758	1925
GUI	**0.0013***	**0.0017***	0.0009	-0.0001	0.0008	-	910	1768
PP	**0.0102****	**0.0089****	**0.0055****	**0.0093****	**0.0082****	**0.0092****	-	1324
NT	**0.0249****	**0.0226****	**0.0192****	**0.0225****	**0.0221****	**0.0239****	**0.0092****	-

**p<0.0001, *p≤0.0345

Atimonan, Catbalogan, and Puerto Princesa generated estimates of effective population size (N_e) from 7,681-12,370 (Table 11). While all sites generated lower bound estimates of effective population size in the thousands and tens of thousands, all upper bound estimates were 'infinite.' With the LD N_e method, reliable estimates from large populations ($N > 500$) can be difficult to achieve (Waples and Do 2010), and 'infinite' estimates can be an indication that the sample size (S) is too small to generate an estimate of N_e for the number of loci used (Do et al. 2014). When only comparisons with $r^2 \leq 0.5$ were used, estimates were affected indicating some residual linkage between loci.

More than 99% of pairwise comparisons of relatedness involved unrelated individuals. Eight pairs of individuals generated estimates of relatedness at the first cousin level (r ~ 0.125) and one pair of individuals from Bohol were estimated to be related at the half-sibling level (r ~0.250) (Table A, Appendix C). Of the 8 putative first cousin pairs, 6 were among individuals

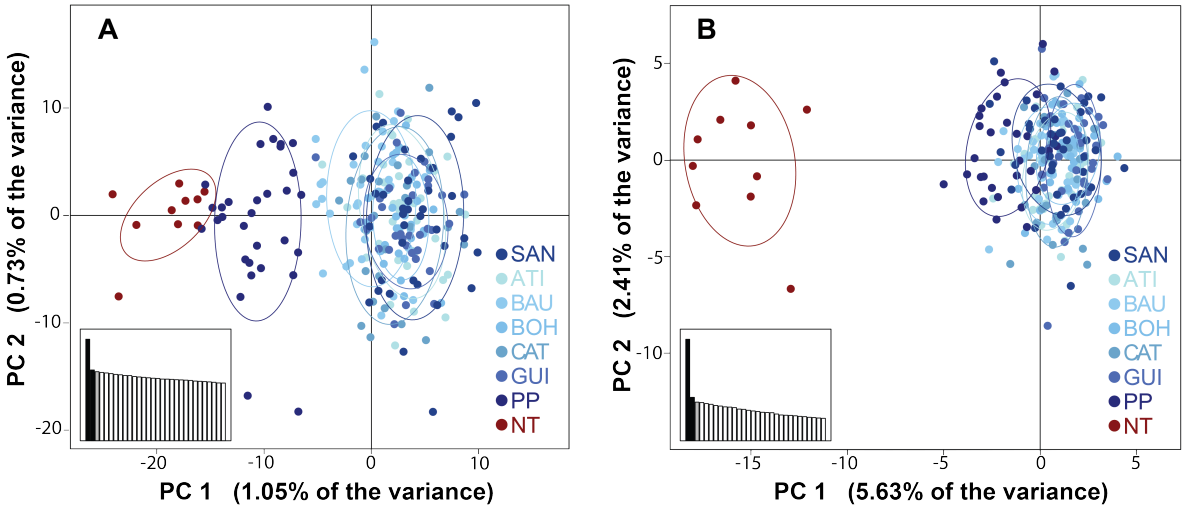
from Nha Trang. The remaining two included one pair of first cousins from Puerto Princesa and a Nha Trang-Puerto Princesa pair.

Table 11. N_e estimates for *Caesio cuning* in the Philippines. *S* is the harmonic mean sample size for each population. Effective population size estimates (N_e) and 95% parametric confidence intervals (CI) were generated in R using the r^2 values from the linkage disequilibrium method in NeEstimator with a minor allele frequency cutoff of 0.05. N_e for each location was calculated with all pairwise comparisons ($r^2 < 1$) and only comparisons with $r^2 \leq 0.5$.

Site	***S***	***N*$_e$ (CI)**	
		$r^2 < 1$	**$r^2 \leq 0.5$**
SAN	38.2	Inf (17,602-Inf)	Inf (Inf-Inf)
ATI	33.3	7,681 (2,895-Inf)	Inf (5,134-Inf)
BAU	38.5	Inf (9,736-Inf)	Inf (36,509-Inf)
BOH	28.1	Inf (4,333-Inf)	Inf (Inf-Inf)
CAT	26.6	12,370 (2,836-Inf)	Inf (Inf-Inf)
GUI	33.3	Inf (5,268-Inf)	Inf (Inf-Inf)
PP	27.9	8,441 (2,596-Inf)	Inf (Inf-Inf)

A PCA with 2030 neutral SNP loci indicates the presence of three distinct genetic clusters along the first principal component: one composed of individuals from Nha Trang, one of individuals from Puerto Princesa, and one containing the remaining individuals from the Philippines (Graph 4A). A PCA with 34 SNPs identified as under positive selection generates two clusters, one containing the 10 individuals for Nha Trang and the other containing the individuals from the Philippines (Graph 4B). Within the Philippines cluster, individuals from Puerto Princesa are slightly offset from the rest. The two main clusters separate along the first principal component (PC 1), accounting for nearly 6% of the total variance in loci under

selection indicating distinctive selective pressures in populations found along the Southeast Asian coastline compared to those in the Philippine Islands.



Graph 4. Principal components analysis of sites. The first principal component (PC 1) is graphed along the x-axis, and the second principal component (PC 2) is graphed along the y-axis. Insets contain the first 30 eigenvalues. A: 249 individuals at 2030 neutral SNP loci. B: 249 individuals at 34 SNP loci identified as being under positive selection.

The ability of STRUCTURE to correctly assign individuals proportionally decreases with smaller sample sizes (Waples and Gagiotti 2006) so analyses were run without the individuals from Nha Trang since only 10 individuals from that location were successfully genotyped. When all Philippines sites were included in the analysis, both log-likelihood values and ΔK supported an optimal K of 2 indicating the presence of two distinct lineages (Fig. 12). Individuals from Puerto Princesa were assigned to one lineage and all individuals from other sites were assigned to another. To test whether further clustering could be resolved within central, northern and

eastern sites, Puerto Princesa was removed and the STRUCTURE analysis was rerun. The log-likelihood value supported a K of 1 and ΔK supported a K of 2. To examine the proportions of two putative lineages across sites, cluster member coefficients were plotted for K=2 (Fig. 12). The ratios of Q values were similar in individuals from Santa Ana and Atimonan and in individuals from Guiuan, Bohol, and Bauan. Individuals from Catbalogan had Q value ratios that fell between these groups.

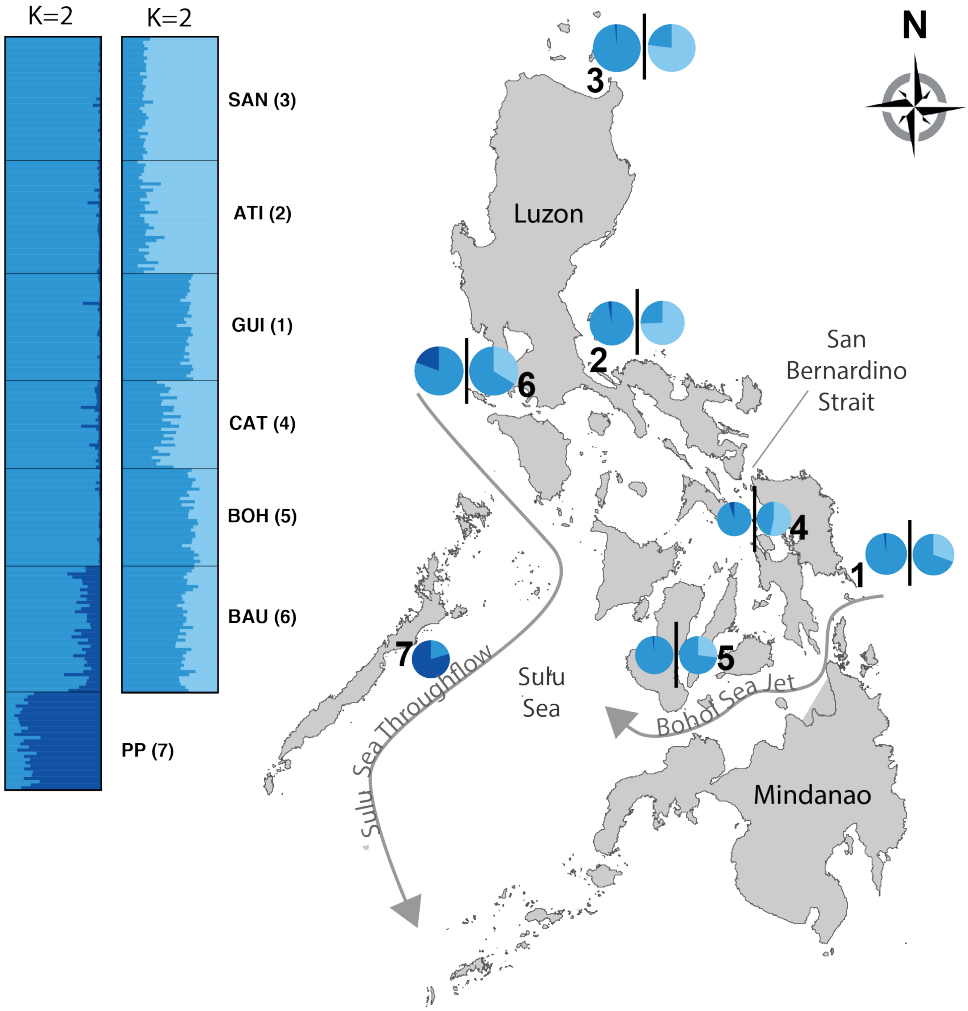


Fig. 12. STRUCTURE results for sites in the Philippines. Pie chart size is relative to number of individuals genotyped at each site. For sites with two pie charts: left pie chart corresponds to the STRUCTURE analysis with all sites (npop=7), right pie chart corresponds to the STRUCTURE analysis with Puerto Princesa excluded (npop=6).

An overall test for the presence of isolation by distance (IBD) among sites in the Philippines was non-significant (Table 12). Since the presence of two lineages was supported by both PCA and STRUCTURE results, partial Mantels were performed. A matrix of cluster similarity was used which assigned a 0 to pairwise values between sites in the central, north and eastern parts of the Philippines and a 1 to pairwise comparisons between Puerto Princesa and any other site. Tests for both IBD and hierarchical clustering were both significant (Table 12), however, the partial Mantel indicated only a slight correlation between genetic and geographic distance when correcting for clusters (r=0.5041) and a very large correlation between genetic distance and cluster when correcting for geographic distance (r=0.9666) indicating that the major driver of differentiation is hierarchical clustering. To confirm these results, a second Mantel test for IBD in sites excluding Puerto Princesa was run and was insignificant.

Table 12. Mantel and partial Mantel tests for isolation by distance (IBD) across seven sites in the Philippines. A matrix of cluster similarity was used for partial Mantel tests. Cluster assignment was informed by PCA and STRUCTURE results with Puerto Princesa assigned to one cluster and all other Philippine sites to another. Significance of the Mantel statistic (r) was estimated with 999 permutations.

Test	x	y	z	r	p-value
Mantel					
IBD	F_{ST}	geographic distance	-	0.051	0.4668
Partial Mantel					
IBD	F_{ST}	geographic distance	clusters	0.5041	0.0080
Clusters	F_{ST}	clusters	geographic distance	0.9666	0.0070

The results of pairwise genetic distances and PCA and STRUCTURE analyses were used to inform models of migration to be tested in MIGRATE-N in addition to panmixia and the full model (Fig. 13). Bauan, Bohol, and Puerto Princesa were used to test the impact of the SST on migration across the Sulu Sea. While not significantly different from one another in pairwise tests (F_{ST}=0.0002), Bauan and Bohol are significantly different from Puerto Princesa (F_{ST}=0.0055 and 0.0093, respectively), and PCA and STRUCTURE results indicate individuals from Puerto Princesa belong to a distinct lineage. Models tested include panmixia in the central Philippines sites of Bohol and Bauan with bi or unidirectional migration across the Sulu Sea (Model A, B & C), and migration south along the SST from Bauan to Puerto Princesa and either bidirectional migration between Bauan and Bohol or north to south migration from Bauan to Bohol (Model D & E). Bauan, Santa Ana, and Atimonan were used to test potential migration via Kuroshio intrusion through the Luzon Strait. Pairwise genetic distances indicate that Santa Ana and Atimonan are significantly different from Bauan but not from each other (F_{ST}=0.0007). Models tested include panmixia in Santa Ana and Atimonan with bidirectional migration with Bauan or north to south migration to Bauan (Model A & B), bidirectional migration between Santa Ana and Atimonan with east to west migration from both or only Santa Ana via Kuroshio incursion to Bauan through the Luzon Strait (Model C & D), and migration only south to north and east to west via the Kuroshio Current and Kuroshio incursion (Model E). Bauan, Catbalogan, Bohol and Guiuan were used to test patterns of migration into the central Philippines from the Bohol Jet as well as between centrally located sites. Pairwise tests indicate no significant differentiation between central sites with the exception of Bauan and Catbalogan (F_{ST}=0.0015). The F_{ST} value was extremely low between Guiuan and Bohol (-0.0001), potentially reflecting connectivity mediated by the Bohol Jet. Models tested include panmixia of Guiuan and Bohol

with bidirectional migration between all sites or north to south migration from Bauan to Bohol and bidirectional migration between other central sites or south to north migration to Catbalogan and bidirectional migration between other central sites (Model A, B & C), bidirectional migration between all sites with east to west migration via the Bohol Jet from Guiuan to Bohol (Model D), and bidirectional migration between Bauan and Bohol and Bauan and Catbalogan with south to north migration to Catbalogan from Bohol and Guiuan and east to west migration from Guiaun to Bohol mediated by the Bohol Jet (Model E). Finally, Atimonan, Catbalogan, and Guiuan were used to test for migration through the San Bernardino Strait. Results of pairwise differentiation indicate all sites are significantly different from each other except for Guiuan and Catbalogan (F_{ST} = 0.0008). STRUCTURE results indicate potential for some migration from Atimonan to Catbalogan through the San Bernardino Strait. Models tested include panmixia

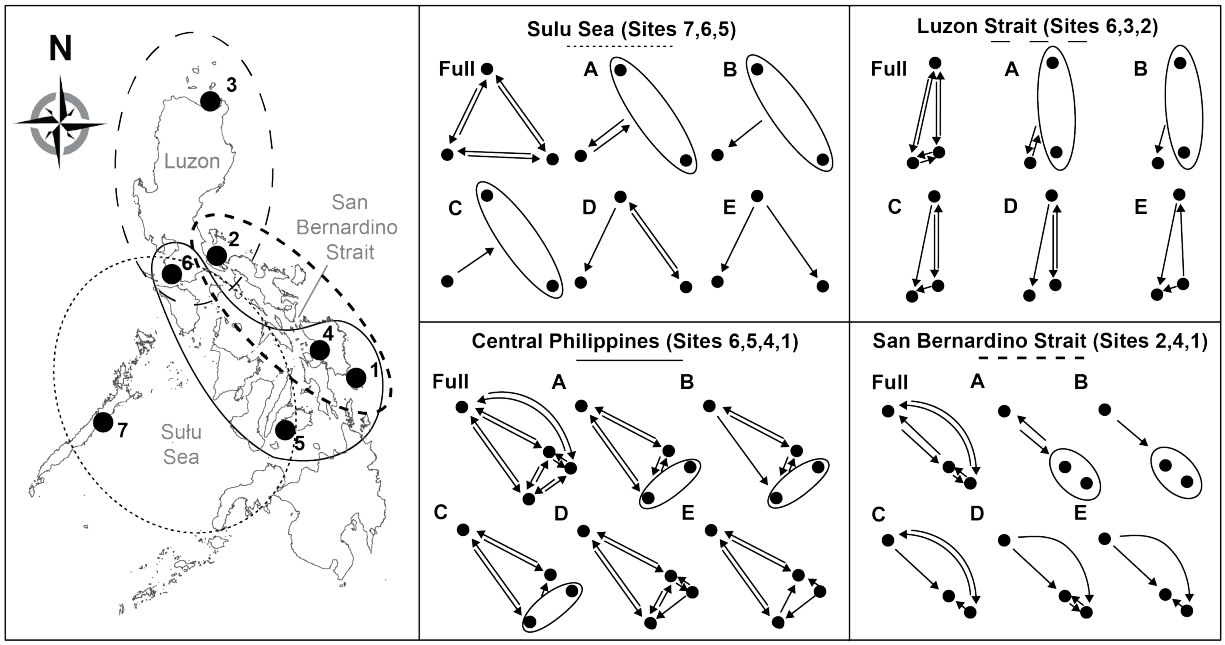


Fig. 13. Tested models of migration in four regions of interest in the Philippines.

between Catbalogan and Guiuan with bi or unidirectional migration with Atimonan (Model A & B), unidirectional migration from Atimonan and Guian to Catbalogan with bidirectional migration between Atimonan and Guiuan (Model C), and unidirectional migration through the San Bernardino Strait from Atimonan to Catbalogan, north to south migration via the Mindanao Current from Atimonan to Guiuan and either bi or unidirectional migration between Catbalogan and Guiuan (Model D & E).

MIGRATE-N results show that in all four subsets, the full model of migration is the most supported indicating there is potential for some level of migration between all sites in the Philippines (Table 13). In the San Bernardino Strait subset, the Bezier log marginal likelihood score for the second ranked model was almost nearly as high as the full model (LBF=-7.66). In this model, Atimonan provides migrants via the San Bernardino Strait and Mindanao Current to both Guiuan and Catbalogon, which share the potential for migration with each other (Model D, Fig. 13). Despite the potential for migration among all sites, estimated numbers of migrants (N_m) between sites are quite low. Across all subsets, the estimated rate of migration between sites in any direction is one migrant per 7 to 17 generations (see Migration Tables, Appendix C). Estimated thetas were impacted by the total number of individuals included in the analysis (n=84 for subsets of 3 populations, n=112 of subsets of 4 populations) with larger numbers of total individuals producing larger estimates of theta. Therefore, these rates of migration should be viewed with extreme caution since these model subsets invariably lack sampling in additional sites that contribute migrants with potential to impact estimates of theta.

Table 13. Comparison of models of migration across four regions in the Philippines. lmL = log marginal likelihood, LBF = log Bayes factor.

Region	Model	Bezier lmL	LBF	Rank
Luzon				
	A	-68089.82	-227.54	4
	B	-68206.17	-343.89	6
	C	-68096.97	-234.69	5
	D	-68350.74	-488.46	7
	E	-68019.94	-157.66	3
	Full	**-67862.28**	**0**	**1**
	Panmixia	-67979.58	-117.30	2
Sulu Sea				
	A	-68123.53	-227.07	4
	B	-68257.20	-360.74	5
	C	-68065.39	-168.93	3
	D	-68435.00	-538.54	7
	E	-68268.97	-390.51	6
	Full	**-67896.46**	**0**	**1**
	Panmixia	-68055.44	-158.98	2
San Bernardino Strait				
	A	-68325.59	-251.36	6
	B	-68267.72	-193.49	5
	C	-68340.45	-266.22	7
	D	-68081.89	-7.66	2
	E	-68254.23	-180.00	4
	Full	**-68074.23**	**0**	**1**
	Panmixia	-68201.03	-126.80	3
Central Philippines				
	A	-74385.48	-1111.79	3
	B	-74573.41	-1299.72	4
	C	-74647.75	-1374.06	5
	D	-73650.28	-376.69	2
	E	-74648.48	-1374.79	6
	Full	**-73273.69**	**0.00**	**1**
	Panmixia	-74901.52	-1627.83	7

Discussion

Overall results from a population analysis using 2,030 putatively neutral SNPs across the seven sites sampled in the Philippine archipelago illuminate distinct patterns of structure not found in previous analyses using a mitochondrial DNA marker (Ackiss et al. 2013). The most predominant pattern found was two divergent lineages on either side of the SST. Both PCA and STRUCTURE analyses support individuals in Puerto Princesa belonging to one lineage and individuals in all other sampled sites belonging to another. Evidence from geological studies shows that Palawan, the island where Puerto Princesa is located, is tectonically associated with mainland Southeast Asia (Briais et al. 1993, Hall 1998, 2002) and assemblages of terrestrial fauna support a connection between Palawan and the island of Borneo during Pleistocene low sea level stands (Heaney 1985). Given the close association of the ten individuals in the Viet Nam outgroup to the lineage present in Puerto Princesa, there is strong potential for the lineage present in Puerto Princesa to be of South China Sea origin. Relatedness analysis found a putative first cousin pair between Nha Trang and Puerto Princesa, lending further support to a South China Sea lineage. STRUCTURE Q-plots indicate the site in southwestern Luzon, Bauan, may also experience some gene flow from this lineage, but throughout the rest of the Philippines the separation of these two divergent clades is most likely maintained by the presence of the SST. Whether the separate grouping of individuals from Nha Trang is an indication of a third lineage or differentiation from Puerto Princesa due to the effects of isolation by distance can not be determined without further sampling in the region of the South China Sea.

Excluding Puerto Princesa, the presence of further population structure within the Philippine lineage was indicated by several significant pairwise F_{ST} values. Of the three sites sampled on the island of Luzon, the sites on the northeastern tip and the east coast (Santa Ana

and Atimonan) are not significantly different from one another, however both are significantly different from Bauan in the southwestern part of the island. This is supported by similar membership coefficients (Q) in Santa Ana and Atimonan compared to Bauan in the STRUCTURE analysis of the Philippine lineage. This pattern of genetic differentiation between individuals on the east and west of Luzon was also recently found in three species of seagrass (Nakajima et al. 2014, Kurokochi et al. 2015, Arriesgado et al. 2016), indicating that at least in these species Kuroshio intrusion through the Luzon Strait does not produce homogenizing levels of gene flow between populations on the east and west coasts of Luzon.

Populations of *Caesio cuning* in the central Philippine islands exhibited high levels of connectivity, a pattern found in another coral reef fish, the parrotfish *Scarus niger,* using RAD-generated SNP data (Stockwell et al. 2016). The two sites found along the path of the Bohol Jet, Guiuan and Bohol, were undistinguishable as subpopulations in pairwise comparisons (generating a negative F_{ST} value). The pairwise comparison between Guiuan and Bauan was also insignificant despite no less than 714 km between Guian and Bauan. These results suggest that the Bohol Jet and its interface with inland sea currents helps to maintain the high level of connectivity found in this region. The only significant relatedness found within the Philippine lineage was a putative half sibling pair within Bohol. Several full sibling pairs were found in the parrotfish *Chlorurus bleekeri* in the nearby island of Siquijor (Stockwell et al. in prep), which may be an indication of hydrodynamics that entrain larvae in this region.

Whether substantial amounts of dispersal and gene flow occur through the San Bernardino Strait cannot be determined from these data. In STRUCTURE plots, Catbalogan has a larger Q for the dominant lineage found in the eastern Luzon sites Atimonan and Santa Ana than other central sites. In addition, the model that came in a very closely ranked second in

MIGRATE-N had unidirectional migration from Atimonan to Catbalogan. These results suggest there is potential for gene flow through the San Bernardino Strait from these northeastern localities, however, both sites were significantly different from Catbalogan in pairwise comparisons.

MIGRATE-N failed to support models of migration associated with oceanographic features, even for sites on either side of the SST, and this may be due to both the time scale over which MIGRATE-N estimates migration and fundamental faults inherent in haplotyped RAD tags that have been generated post-filtering. The structured coalescent employed in MIGRATE-N looks for the most recent common ancestor and integrates over all possible genealogies and migration events (Beerli and Felsenstein 1999, 2001), thus MIGRATE-N generates migration rates over potentially thousands of generations rather than contemporary migration rates. In addition, these estimates were generated from RAD tags that contain only a subset of the true data. Phased, haplotypic RAD tags were generated using all SNPs surviving filtering on each tag instead of just one. However, rare variants and other true polymorphisms whose presence would affect estimates of mutation rate were removed during the filtering process in order to generate a SNP panel that is informative across all sites and does not include sequencing errors or invalid base calls. In light of these potentially confounding issues, the results from MIGRATE-N runs presented here should be viewed with extreme caution.

The inability to generate estimates of effective population size as well as low levels of relatedness overall within the Philippines are indications of relatively large populations. Simulations using both microsatellites and SNPs showed precise estimation of both N_e and upper bound estimates of N_e using the LD N_e method can be difficult from populations where $N>1000$ (Waples and Do 2010). Low levels of relatedness were supported by the G_{IS} values generated by

site, which suggest there are no measurable levels of inbreeding occurring at any of the sites sampled. It is worth noting that the largest amount of relatedness in individuals was found in the 10 outgroup individuals from Viet Nam. Of the nine putatively related pairs, seven involved individuals from Nha Trang, six of which were Nha Trang-Nha Trang pairs. This may be an indication of smaller population sizes, however, 10 individuals was insufficient for estimation of N_e. The large separation along principal component one from the PCA using divergent SNPs of individuals from Nha Trang and individuals from the Philippines indicates distinct selective pressures on coastal coral reefs in mainland Southeast Asia, the presence of which could potentially be impacting overall population sizes relative to counterparts in the Philippine archipelago.

The detection of regional limits to gene flow coupled with no evidence for isolation by distance among sites supports the hypothesis that hydrodynamic features such as the SST, Bohol Jet and NEC are drivers of connectivity or structure within populations of *Caesio cuning* in the Philippines. However, given the association of the divergent lineage in Puerto Princesa with the South China Sea, another potential source of the population divergence observed across the sampled region is basin isolation. The same Pleistocene sea level lows that provided coastline connection between the island of Palawan and Borneo also created relatively isolated sea basins across the Coral Triangle (Voris 2000) including the South China Sea and the Sulu Sea. Patterns of divergence that are spatially associated with sea basins in several coral reef species across this region have been hypothesized to have been caused by extended periods of basin isolation during this epoch (Barber et al. 2002, Ravago-Gotanco and Juinio-Meñez 2010, Carpenter et al. 2011, Raynal et al. 2014).

Testing origins of population structure in this study of the Philippine archipelago is limited by the locations sampled. Future analyses would be improved with the addition of sites not included in this analysis, particularly from the southern island of Mindanao and on the South China Sea side of the Philippines. In the current study, the Bohol Jet serves as a conduit for connectivity. However, this strong surface current could also act as a potential barrier to gene flow between sites north and south of the Bohol Sea. In addition, understanding the extent and spread of the South China Sea lineage in *C. cuning* along the western coast of the Philippines could help management planners accurately demark specific stocks or regions tied to genetic descent.

The observed patterns of population structure in *Caesio cuning* provide resource managers with valuable information for the delineation of genetic neighborhoods on which to focus management strategies. Particular consideration should be given to areas where concordant patterns have been found in other species, such as the east coast of Luzon north of the NEC bifurcation. The distinct genetic signatures found in *C. cuning* north of the NEC bifurcation provides additional support to an already established pattern in at least two other targeted reef species (Ravago-Gotanco et al. 2007, Magsino and Juinio-Meñez 2008). High levels of connectivity in the central islands north of the Bohol Jet have also been found in two parrotfish and a rabbitfish (Stockwell et al. 2016, in prep), all targeted municipal fisheries, suggesting this region should be treated as a network of reefs. In addition, the discovery of a divergent lineage in Puerto Princesa east of the SST suggests that the island of Palawan should also be treated as a distinct region for the development of management strategies. Future analyses incorporating un-sampled locations and more taxa will help to further refine the delineation of coral reef neighborhoods on which to focus the development of reserve networks in the Philippines.

Conclusions

Genetic analysis of seven sites in the Philippine archipelago revealed the presence of two divergent lineages, one associated with the South China Sea and one associated with the islands in the central, northern, and eastern parts of the Philippines. Additional structure within the Philippine lineage correlated to present day oceanographic features and suggests that hydrodynamic processes act as both drivers and maintainers of structure or connectivity between different regions in the Philippines. The large disparity between a previous genetic analysis of connectivity in the Philippines using mitochondrial DNA that found no population structure and the results presented in the current study using RAD sequencing data emphasizes the importance of using the best possible genetic tools available to inform conservation and management decisions.

DISCUSSION

This research represents the first comprehensive genetic analysis of the population structure in the fusilier fish, *Caesio cuning,* which is an important municipal fishery on reefs in the Coral Triangle and western Pacific Ocean. Genetic data from both classical Sanger and next-generation sequencing were examined across three spatial scales with multiple analytical approaches to address several hypotheses regarding the biogeography, ecology and population connectivity of *Caesio cuning*.

Establishing the genetic patterns of a highly mobile reef species

Despite its vagile lifestyle, genetic analyses of *Caesio cuning* across the Coral Triangle using mitochondrial DNA revealed significant regions of genetic structure. Two distinct clades were observed on either side of the Sunda Shelf, a pattern reported from several highly site-associated reef taxa (Vogler et al. 2008, Crandall et al. 2008b, Drew & Barber 2009, Gaither et al. 2010). Shared phylogeographic patterns such as these result from broadly acting physical processes that shape genetic patterns in codistributed taxa (Avise 2000, Carpenter et al. 2011). The presence of two mitochondrial clades in this region is generally attributed to historical vicariance between Pacific and Indian Ocean populations during Pleistocene low sea level stands with the exposure of the Sunda and Sahul shelves (e.g. Barber et al. 2000, Rohfritsch & Borsa 2005, Deboer et al. 2008).

It is uncommon for reef species to exhibit two mitochondrial clades that maintain relatively distinct spatial boundaries, which makes the lack of overlap found between divergent Indian and Pacific Ocean lineages in *Caesio cuning* surprising. The anenomefish *Amphiprion*

ocellaris, which exhibits natal homing behavior and has a short pelagic larval duration of 8-12 days, shows greater admixture of Indian and Pacific maternal lineages in the Java Sea than *C. cuning* (Fautin & Allen 1992, Jones et al. 2005, Timm & Kochzius 2008). Low levels of admixture in modern times despite the lack of vicariant isolation such as that seen in *C. cuning* clades, likely results from oceanographic currents, but reproductive isolation between the clades cannot be ruled out as a possible explanation for the absence of gene flow in this region.

In addition to the Sunda shelf break, significant limits to genetic exchange were also found in eastern Indonesia. Genetic structure east and west of Halmahera is found in several coral reef invertebrates (Barber et al. 2006, Deboer et al. 2008, Nuryanto & Kochzius 2009, Barber et al. 2011), suggesting this region may be important for lineage divergence in the Coral Triangle. While *Caesio cuning* populations in this region are not characterized by distinct clades as seen across the Sunda Shelf, partial Mantel tests indicate that this genetic signature is not isolation-by-distance but hierarchical structure between two genetic clusters. The presence of two strong eddies at this junction, the Halmahera Eddy and the Mindanao Eddy, may be the cause of significant haplotypic frequency differences detected in this region. Dispersal simulations of pelagic larvae from coral reef organisms in central and eastern Indonesia support the hypothesis that these eddies drive genetic isolation in this region (Kool et al. 2011).

The recovery of multiple regions of significant genetic structure in *Caesio cuning* is somewhat unexpected because of the high mobility potential of adults. However, the concordance of predictions from biophysical models of larval dispersal with these data as well as with phylogeographic patterns of site-attached reef species such as giant clams (Deboer et al. 2008, Nuryanto & Kochzius 2009) suggests that the major avenue of genetic connectivity in *C. cuning* is via larval dispersal.

Testing the central-peripheral population model with high-resolution genetic markers

RAD sequencing of *Caesio cuning* in five sites from central to peripheral populations along the Kuroshio Current uncovered both the presence of isolation by distance and two distinct genetic clusters, one in Okinawa and the other containing the individuals from Ishigaki and the Philippines. The presence of isolation by distance was somewhat surprising in light of the proximity of sampled sites to the Kuroshio Current, however, previous documentation of the rarity of *C. cuning* off the continental shelf may indicate limits to the ability of larvae to successfully disperse across the deep water between the Philippines and Ishigaki (Reader and Leis 1996). A distinct genetic lineage in Okinawa is likely sustained in part by divergent selection in marginal environmental conditions at higher latitudes. A principal components analysis of loci under selection separated the individuals sampled in the Ryukyus and those sampled from the east coast of the Philippines, and proximity to coding regions of many of these loci lends additional support to distinct patterns of selection in peripheral populations.

Sites in the Philippines belong to the same genetic lineage, but pairwise genetic distances indicated significant genetic differentiation across the eastern seaboard in the region of the bifurcation of the Northern Equatorial Current (NEC). The bifurcation of the NEC has previously been hypothesized to restrict genetic connectivity of reef species north and south along the east coast of the Philippines, and genetic structure supporting this has been found in other species including a rabbitfish and a giant clam (Magsino and Juinio-Meñez 2008, Ravago-Gotanco et al. 2007).

The effects of peripheral isolation were noticeable in estimates of relatedness and effective population size in the three of the most northern sites. No within-site related individuals

were sampled until Ishigaki with a marked increase of within-site relatedness in Okinawa. In addition, finite estimates of N_e were only calculated for the three most northern sites with the smallest N_e calculated in Okinawa. Comparatively, the two most southern sites, Guiuan and Atimonan, only produced lower bound estimates for N_e. However, even the ability to estimate a lower bound for N_e can provide useful information (Waples and Do 2010), and in the case of the five sites sampled along the Kuroshio Current, indicates a successively decreasing number of effective breeders contributing to local populations from the center to the periphery of the species range.

Even with continually increasing edge effects, no evidence was found that indicates peripheral populations of *Caesio cuning* are currently experiencing detrimental genetic depression. No within- or between-site relatedness was measured at a level greater than first cousin, and G_{IS} in all sites is close to 0 so there is no indication of inbreeding within any of the sampled populations. However, the divergence, lower N_e, and increased relatedness found in Okinawa do indicate that this population is potentially vulnerable to overfishing. If overharvest of *C. cuning* were to occur in the periphery, the increased genetic differentiation and edge effects puts these populations at risk of a bottleneck event and slower recovery times. Consistent monitoring of catch levels and abundance in Okinawa, Ishigaki, and the northern Philippines is recommended to prevent overharvest of this species in these sites.

Using RAD sequencing to examine stock structure and connectivity in the Philippines

Genetic analysis of seven sites sampled in the Philippine archipelago and an outgroup from Viet Nam found two divergent lineages on either side of the Sulu Sea Throughflow (SST). STRUCTURE analyses of Philippine sites assigned individuals from Puerto Princesa to one

lineage and individuals from all other sites to another. The close association of the ten individuals in the Viet Nam outgroup to those from Puerto Princesa in a principal components analysis (PCA) indicates the lineage in Puerto Princesa is likely of South China Sea origin, a finding supported by the presence of a putative first cousin pair between Nha Trang and Puerto Princesa. The maintenance of these two divergent clades in the Philippines is most likely due to the presence of the SST.

Within the Philippine lineage, significant pairwise F_{ST} values and STRUCTURE membership coefficients indicated that individuals from the two sites on the east coast of Luzon, Santa Ana and Atimonan, are significantly different from all other sites but each other. This pattern of genetic differentiation indicates that Kuroshio intrusion through the Luzon Strait does not produce homogenizing levels of gene flow between populations on the east and west coasts of Luzon. It also supports the findings from genetic analysis along the Kuroshio Current, which produced the same pattern of differentiation in *C. cuning* on the eastern seaboard with a panel of nearly 3,000 SNPs.

Contrasting patterns of connectivity were found through the two main straits that connect the eastern seaboard to the central Philippines. Despite its location on the eastern seaboard outside of the Surigao Strait, Guiuan did not have a single significant pairwise F_{ST} value with any site sampled in the central Philippine islands. This is most likely due to the influence of the Bojol Jet (Gordon et al. 2011) and its interface with inland sea currents. The pairwise F_{ST} between the two sites found directly on the path of the Bohol Jet, Guiuan and Bohol, was negative (an indication of more variation between individuals within sites than between sites). Comparatively, tests of gene flow from the east coast to the central Philippines through the relatively shallow San Bernardino Strait were inconclusive due to conflicting results. In

STRUCTURE plots, Catbalogan showed some similarity to the eastern Luzon sites Atimonan and Santa Ana, but both sites were significantly different from Catbalogan in pairwise comparisons. In addition, while MIGRATE-N failed to support models of migration associated with oceanographic features overall, the model that came in a very closely ranked second in this region had unidirectional migration from Atimonan to Catbalogan.

Overall levels of relatedness in the Philippines was low, and a large percentage of relatedness found could be attributed to the 10 individuals from Viet Nam. Of the nine putatively related pairs, seven involved individuals from Nha Trang, six of which were Nha Trang-Nha Trang pairs. This could be a sign of smaller population sizes, however, the number of individuals successfully genotyped was insufficient to attempt estimation of N_e. The largest proportion of variance in the PCA of divergent SNPs separated individuals from Nha Trang and the Philippines, which could be an indication of distinct selective pressures in coastal coral reef fishes in mainland Southeast Asia. If selective processes were sufficiently detrimental, they could be impacting overall population sizes relative to counterparts in the Philippine archipelago. Conversely, the inability to generate estimates of N_e in the Philippines sites infers relatively large population sizes (N>1000; Waples and Do 2010). Despite the level of relatedness in individuals from Viet Nam, the G_{IS} values suggested there are no measurable levels of inbreeding occurring at any of the sites sampled.

Overall, the presence of two distinct lineages provides specific regions of genetic descent that could be treated as separate stocks by resource managers. Within the Philippine lineage, the recovery of significant regions of differentiation with no evidence for isolation by distance among sites supports the theory that hydrodynamic features are the main drivers of connectivity or structure within populations of *C. cuning* in the Philippines. Therefore, the design of marine

protected area networks should give additional consideration to areas with evidence of isolating oceanographic features such as the eastern seaboard north of the bifurcation of the NEC.

CONCLUSIONS AND FUTURE DIRECTIONS

The use of multiple genetic markers to examine connectivity in populations of *Caesio cuning* across different spatial scales in the species range proved successful at illuminating aspects of the species' biogeographic history as well as spatial and oceanographic influences that impact gene flow and effective population size. Two distinct maternal lineages in *C. cuning* were found across the Coral Triangle on either side of the Sunda Shelf. Using RAD sequencing to examine patterns of gene flow within the Pacific mitochondrial lineage uncovered the presence of isolation by distance along the Kuroshio Current, the presence of measureable edge effects from centrally to peripherally located populations of *C. cuning*, and three further regions of divergence including in the Ryukyu Islands at the edge of the species range and east and west of the Sulu Sea Throughflow (SST).

These findings provide a multitude of additional hypotheses to examine for future studies of genetic patterns in populations of *C. cuning*. For example, the extent of overlap of the Indian and Pacific Ocean maternal lineages appeared limited relative to other coral reef fishes. An examination using nuclear-based genetic markers across the region of overlap in Indonesia and Malaysia could confirm whether this pattern exists due to limitations to pelagic larval dispersal or reproductive barriers. In addition, analysis of *C. cuning* across the Philippines was limited to the seven sites included in this study. The addition of sites from the southern island of Mindanao

and on the South China Sea side of the Philippines would clarify the extent and spread of the South China Sea lineage in *C. cuning* along the western coast of the Philippines could help management planners demark specific stocks or regions tied to genetic descent.

Here, the use of coalescent analyses only achieved marginal success at providing useful information about populations of *Caesio cuning*. MIGRATE-N did detect reduced levels of migration to areas impacted by edge effects, but in general failed to support models of migration associated with oceanographic features. The structured coalescent employed in MIGRATE-N looks for the most recent common ancestor and integrates over all possible genealogies and migration events (Beerli and Felsenstein 1999, 2001). With local population sizes estimated to be greater than 1,000 in *C. cuning*, MIGRATE-N is potentially estimating migration rates over thousands or even tens of thousands of generations rather than contemporary migration rates. Performance may also have been affected by the filtering of rare variants before the generation of phased RAD tags whose presence would affect estimates of mutation rate.

Future estimation of migration rates and models with restriction site associated DNA (RAD) data could benefit from several modified approaches. If a coalescent analysis such as MIGRATE-N is employed, separate filtering to generate the haplotyped RAD tags required for input that does not include the removal of rare variants would help reduce errors associated with estimations of mutation rate. In addition, a non-equillibrium approach that uses multilocus genotypes to assign individuals to source populations and extract information on migration within the last few generations (Rannala and Mountain 1997, Wilson and Rannala 2003) may prove more successful in a species like *C. cuning* with relatively large effective population sizes.

LITERATURE CITED

Ackiss AS, Pardede S, Crandall ED, Ablan-Lagman M, Carmen A, Barber PH, Carpenter KE (2013) Pronounced genetic structure in a highly mobile coral reef fish, *Caesio cuning*, in the Coral Triangle. Mar Ecol Prog Ser 480:185-197

Alcala A, Russ G (1990) A direct test of the effects of protective management on abundance and yield of tropical marine resources. ICES J Mar Sci 47:40

Alcala A, Russ G (2002) Status of Philippine coral reef fisheries. Asian Fish Sci 15:177-192

Allen GR (2008) Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. Aquat Conserv 18:541-556

Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. Mol Ecol 23:1661-1667

Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JE, Luikart G (2014) Trade-offs and utility of alternative RASseq methods: reply to Puritz et al. Mol Ecol 23:5943-5946

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. BMC Bioinformatics 9:323

Arriesgado DM, Kurokochi H, Nakajima Y, Matsuki Y, Uy WH, Fortes MD, Campos WL, Nadaoka K, Lian C (2015) Genetic diversity and structure of the tropical seagrass *Cymodocea serrulata* spanning its central diversity hotspot and range edge. Aquat Ecol 49:357-72

Arriesgado DM, Kurokochi H, Nakajima Y, Matsuki Y, Uy WH, Fortes MD, Campos WL,

Nadaoka K, Lian C (2016) Population genetic diversity and structure of a dominant tropical seagrass, *Cymodocea rotundata*, in the Western Pacific region. Mar Ecol 37: 786–800

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

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37-48

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

Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2002) Sharp genetic breaks among populations of Haptosquilla pulchella (Stomatopoda) indicate limits to larval transport patterns, causes, and consequences. Mol Ecol 11:659-674

Barber PH, Erdmann MV, Palumbi SR (2006) Comparative phylogeography of the three codistributed stomatopods: origins and timing of regional lineage diversification in the coral triangle. Evolution 60:1825-1839

Barber PH (2009) The challenge of understanding the Coral Triangle biodiversity hotspot. J Biogeogr 36:260-265

Barber PH, Cheng SH, Erdmann SE, Tengardjaja K, Ambariyanto (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. Crustacean Iss 19:129-156

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:e3376

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proc Roy Soc London B 263:1619-1626

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

Beerli, P (2009) How to use migrate or why are markov chain monte carlo programs difficult to use? In: Bertorelle O, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C (eds) Conservation Biology, Volume 17. Cambridge University Press, Cambridge UK

Beerli, P. and M. Palczewski (2010) Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics 185:313–326

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B Met 57:289-300

Bird CE, Karl SA, Smouse PE, Toonen RJ (2011) Detecting and measuring genetic differentiation. Phylogeogr Pop Genet Crustacea, 19(3), pp.l-55.

Blair C, Weigel DE, Balazik M, Keeley AT, Walker FM, Landguth E, Cushman SA, Murphy M, Waits L, Balkenhol N (2012) A simulation-based evaluation of methods for inferring linear barriers to gene flow. Mol Ecol Resour 12: 822-833

Bloch ME (1791) Naturgeschichte der auslandischen fische, Berlin

Borsa P (2003) Genetic structure of the round scad mackeral *Decapterus macrosoma* (Carangidae) in the Indo-Malay archipelago. Mar Biol 142:575-581

Bowen BW, Shanker K, Yasuda N, Celia M, Malay MCMD, von der Heyden S, Paulay G, Rocha LA, Selkoe KA, Barber PH, Williams ST (2014) Phylogeography unplugged: comparative surveys in the genomic era. Bull Mar Sci 90:13-46.

Briais A, Patriat P, Tapponnier P (1993) Updated interpretation of magnetic anomalies and seafloor spreading stages in the South China Sea: Implications for the Tertiary tectonics of Southeast Asia. J Geophys Res-Earth. 98:6299-328

Briggs JC (1995) Global biogeography. Elsevier, Amsterdam; New York

Brown WM, George Jr M, Wilson AC (1979) Rapid evolution of animal mitochondiral DNA. P Nat Acad Sci USA 76:1967-1971

Brown JH (1984) On the relationship between abundance and distribution of species. Am Nat 124: 255-279

Bureau of Agricultural Statistics (2010) Fisheries statistics of the Philippines 2007-2009. p 1-416

Burke L, Reytar K, Spalding M, Perry A (2012) Reefs at risk revisted in the Coral Triangle. World Resources Institute, Washington, D.C.

Candy JR, Campbell NR, Grinnell MH, Beacham TD, Larson WA, Narum SR (2015) Population differentiation determined from the putative neutral and divergent adaptive genetic markers in Eulachon (*Thaleichthys pacificus*, Osmeridae), an anadromous Pacific smelt. Mol Ecol Res 15:1421-1434

Carpenter KE (1988) FAO Species Catalogue. Vol. 8. Fusilier fishes of the world. An annotated and illustrated catalogue of caesionid species known to date. In: Fischer W, Scialabba N (eds) FAO Fish Synop, Volume 8, FAO, Rome

Carpenter KE, Springer VG (2005) The center of the center of marine shore fish biodiversity: the Philippine Islands. Environ Bio Fish 72:467-480

Carpenter KE, Barber PH, Crandall ED, Ablan-Lagman MCA and others (2011) Comparative

phylogeography of the coral triangle and implications for marine management. J Mar Biol 2011:396982

Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22: 3124-3140

Crandall ED, Jones ME, Muñoz MM, Akinrobe B, Erdmann MV, Barber PH (2008a) Comparative phylogeography of two seastars and their ectosymbionts within the coral triangle. Mol Ecol 17:526-5290

Crandall ED, Frey M, Grosberg R, Barber PH (2008b) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. Mol Ecol 17:611-626

Crandall ED, Treml EA, Barber PH (2012a) Coalescent and biophysical models of stepping-stone gene flow in neritid snails. Mol Ecol 21: 5579–5598

Crandall ED, Sbrocco EJ, DeBoer TS, Barber PH, Carpenter KE (2012b) Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the last glacial maximum. Mol Biol Evol 29: 707–719

Crossland CJ (1984) Seasonal variations in the rates of calcification and productivity in the coral *Acropora formosa* on a high-latitude reef. Mar Ecol Prog Ser 15: 135-140

Dana JD (1843) On the temperature limiting the distribution of corals. Am J Sci 45: 130-131

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker R, Lunter G, Marth G, Sherry ST, McVean G, Durbin R, and 1000 genomes Project Analysis Group (2011) The Variant Call Format and VCFtools. Bioinformatics 27: 2156-2158

Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML (2012) Special features of RAD Sequencing data: implications for genotyping. Mol Ecol 22:3151-3164

Deboer TS, Subia M, Erdmann MV, Kovitvongsa K, Barber PH (2008) Phylogeography and

limited genetic connectivity in the endangered boring giant clam across the Coral Triangle. Conserv Biol 22:1255-1266

Delaunay B (1934) Sur la sphère vide. Bulletin de l'Académie des Sciences de l'URSS, Classe des sciences mathématiques et naturelles 6: 793–800

Do C, Waples RS, Peel D, Macbeth GM, Tillet BJ, Ovenden JR (2014) NeEstimator V2: re-implementation of software for the estimation of contemporary effective populations size (N_e) from genetic data. Mol Ecol Resour 14: 209-214

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

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

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359-361

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1-15

Etherington TR (2011) Python based GIS tools for landscape genetics: visualising genetic relatedness and measuring landscape connectivity. Methods Ecol Evol 2:52-55

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics under Linux and Windows. Mol Ecol Resour 10:564-567

Faircloth BC, Glenn TC (2014) Protocol: preparation of an AMPure XP substitute (AKA Serapure). doi: 10.6079/J9MW2F26

Falush, D., Stephens, M., and Pritchard, J. K. (2003a). Inference of population structure: extensions to linked loci and correlated allele frequencies. Genetics 164:1567–1587

Fautin DG, GR Allen (1992) Field guide to anemone fishes and their host sea anemones. Western Australian Museum, Perth, Australia

Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L, Berlin AM, Blumenstiel B, Cibulskis K, Friedrich D, Johnson R, Juhn F, Reilly B, Shammas R, Stalker J, Sykes SM, Thompson J, Walsh J, Zimmer A, Zwirko Z, Gabriel S, Nicol R, Nusbaum C (2011) A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol 12:R1

Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180: 977-993

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147:915

Gaines SD, White C, Carr MH, Palumbi SR (2010) Designing marine reserve networks for both conservation and fisheries management. P Nat Acad Sci 107: 18286-18293

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

Gaither MR, Bowen BW, Bordenave T-R, Rocha LA and others (2011) Phylogeography of the

reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the indo-pacific barrier with contemporary overlap in the coral triangle. Evol Biol 11:189

Gautier M, Gharbi K, Cezard T, Foucaud J, Kerdelhué, Pudlo P, Cornuet J-M, Estoup A (2012) The effect of RAD allele droput on the estimation of genetic variation within and between populations. Mol Ecol 22:3165-3178

Gell FR, Roberts CM (2003) Benefits beyond boundaries: the fishery effects of marine reserves. Trends Ecol Evol 18: 448-455

Gerlach G, Atema J, Kingsford MJ, Black KP, Miller-Sims V (2007) Smelling home can prevent dispersal of reef fish larvae. P Nat Acad Sci 104: 858-863

Gordon AL, Sprintall J, Ffield A (2011) Regional oceanography of the Philippine Archipelago. Oceanography 24:14-27

Gould SJ (2002) The structure of evolutionary theory. Belknap Press of Harvard University Press, Cambridge, Massachusetts

Green A, Mous P (2008) Delineating the Coral Triangle, its ecoregions and functional seascapes.Version 5.0. TNC Coral Triangle Program Report 1/08

Green AL, Maypa AP, Almany GR, Rhodes KL, Weeks R, Abesamis RA, Gleason MG, Mumby PJ, White AT (2015) Larval dispersal and movement patterns of coral reef fishes, and implication for marine network design. Biol Rev 90:1215-1247

Gruenthal KM, Witting DA, Ford T, Neuman MJ, Williams JP, Pondella DJ, Bird A, Caruso N, Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. Mol Ecol 18:4734–4756

Gruenthal KM, Witting DA, Ford T, Neuman MJ, Williams JP, Pondella DJ, Bird A, Caruso N,

Hyde JR, Seeb LW, Larson WA (2014) Development and application of genomic tools to the restoration of green abalone in southern California. Conserv Genet 15:109–121

Guillot G, Estoup A, Mortier F, Cosson JF (2005a) A spatial statistical model for landscape genetics. Genetics 170:1261–1280

Guillot G, Mortier F, Estoup A (2005b) GENELAND: a computer package for landscape genetics. Mol Ecol Notes 5:712–715

Guillot G (2008) Inference of structure in subdivided populations at low levels of genetic differentiation. The correlated allele frequencies model revisited. Bioinformatics 24:2222–2228

Guillot G, Santos F (2009) A computer program to simulate multilocus genotype data with spatially autocorrelated allele frequencies. Mol Ecol Resour 9: 1112–1120

Guillot G, Leblois R, Coulon A, Frantz AC (2009) Statistical methods in spatial genetics. Mol Ecol 18: 4734–4756

Guindon S, Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Sys Biol 52: 696-704

Hall R (1998) The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. Biogeography and geological evolution of SE Asia. p 99-131

Hall R (2002) Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. J Asian Earth Sci 20:353-431

Han W, Moore AM, Di Lorenzo E, Gordon AL, Lin J (2008) Seasonal surface ocean circulation and dynamics in the Philippine Archipelago region during 2004-2008. Dynam Atmos Oceans 47:113-147

Heaney LR (1985) Zoogeographic evidence for middle and late Pleistocene land bridges to the

Philippine Islands. Mod Quat Re 9:127-44

Hedgecock D, Launey S, Pudovkin A, Naciri Y, Lapègue S, Bonhomme F (2007) Small effective number of parents (Nb) inferred for a naturally spawned cohort of juvenile European flat oysters *Ostrea edulis*. Mar Biol 150:1173-1182

Hedrick PW, Ginevan ME, Ewing EP (1976) Genetic polymorphism in heterogenous environments. Ann Rev Ecol Sys 7:1-32

Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633-1638

Hohenlohe PA, Amish SJ, Catchen JM, Allendorf FW, Luikart G (2011) Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Mol Ecol Res 11:117-122

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL (2000) Heteroplamsy and evidence for recombination in the mitochondrial control region of the flatfish *Platyichthys flesus*. Mol Biol Evol 19:2261-2264

Hoarau G, Piquet AM-T, van der Veer HW, Rjinsdorp AD, Stam WT, Olsen JL (2004) Population structure of plaice (*Pleuronectes platessa* L.) in northern Europe: a comparison of resolving power between microsatellites and mitochondrial DNA data. J Sea Res 51:183-190

Hurlburt HE, Metzger EJ, Sprintall J, Riedlinger SN, Arnone RA, Shinoda T, Xu X (2011) Circulation in the Philippine archipelago simulated by 1/12 and 1/25 global HYCOM and EAS NCOM. Oceanography 24:28-47

Hyde JR, Seeb LW, Larson WA (2014) Development and application of genomic tools to the restoration of green abalone in southern California. Conserv Genet 15:109–121

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

Johannesson K, André C (2006) Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. Mol Ecol 15:2013-2029

Jokiel PL, Coles SL (1977) Effects of temperature on the mortality and growth of Hawaiian reef corals. Mar Biol 43:201-208

Jombart, T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070-3071

Jones GP, Planes S, Thorrold SR (2005) Coral reef fish larvae settle close to home. Curr Biol 15:1314-1318

Karl SA, Toonen RJ, Grant WS, Bowen BW (2012) Common misconceptions in molecular ecology: echoes of the modern synthesis. Mol Ecol 21:4171-4189

Kinlan B, Gaines S (2003) Propagule dispersal in marine and terrestrial environments: A community perspective. Ecology 84: 2007–2020

Kleypas JA, McManus JW, Meñez LAB (1999) Environmental Limits to Coral Reef Development: Where Do We Draw the Line? Amer Zoo 39: 146-159

Knowlton N, Weight LA, Solorzano LA, Mills DK, Bermingham E (1993) Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the isthmus of Panama. Science 260:1629-1632

Knowlton N (2000) Molecular genetic analyses of species boundaries in the sea. Hydrobiologia 420:73-90

Kraus RHS, Kerstens HHD, Van Hooft P, Crooijmans RPMA, Van Der Poel JJ, Elmberg J, Vignal A, Huang Y, Li N, Prins HH, Groenen MMA (2011) Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyphynchos*). BMC Genomics 12: 150

Kool JT, Paris CB, Barber PH, Cowen RK (2011) Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. Glob Ecol Biogeogr 5:695-706

Kurokochi H, Matsuki Y, Nakajima Y, Fortes MD, Uy WH, Campos WL, Nadaoka K, Lian C (2016) A baseline for the genetic conservation of tropical seagrasses in the western North Pacific under the influence of the Kuroshio Current: the case of *Syringodium isoetifolium*. Conserv Genet 17:103-110

Ladoukakis ED, Zouros E. (2001) Direct evidence for homologous recombination in mussel (*Mytillus galloprovincialis*) mitochondrial DNA. Mol Biol Evol 18:1168-75

Lee W, Conroy J, Howell W, Kocher T (1995) Structure and evolution of teleost mitochondrial control regions. J Mol Evol 41:54-66

Leis JM, Carson-Ewart BM (2003) Orientation of pelagic larvae of coral-reef fishes in the ocean. Mar Ecol Prog Ser 252:239-253

Levin DA (1970) Developmental instability and evolution in peripheral isolates. Am Nat 104: 343-353

Lewotin RC, Hubby JL (1966) A molecular approach to the study of genetic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics 54:595-609

Li, H (2014) Towards better understanding of artifacts in variant calling from high-coverage

samples. Bioinformatics 30:2843-2851

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

Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittman M, Wang C, Kobayashi M, Norton H, Brown EL (1996) Expression monitoring by hybridization to density oligonucleotide arrays. Nat Biotechnol 14:1675-1680

Lourie SA, Green DM, Vincent ACJ (2005) Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: *Hippocampus*). Mol Ecol 14:1073-1094

Lukoschek V, Waycott M, Keogh JS (2008) Relative information content of polymorphic microsatellites and mitochondrial DNA for inferring dispersal and population genetic structure in the olive sea snake, *Aipysurus laevis*. Mol Ecol 17:3062-3077

MacArthur RH (1972) Geographical ecology. Harper & Row, New York

Magsino RM, Juinio-Meñez MA (2008) The influence of contrasting like history traits and oceanic processes on genetic structuring of rabbitfish populations *Siganus argenteus* and *Siganus fuscescens* along the eastern Philippine coasts. Mar Biol 154:519-532

Magoulas A, Zouros E (1993) Restriction-site heteroplasmy in anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. Mol Biol Evol 10:319-325

Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic)

variation: how barriers can be detected by "Monmonier's algorithm." Hum Biol 76:173-190.

Martin AP, Naylor GJP, Palumbi SR (1992) Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature 357: 153-155

Mayr E (1963) Animal species and evolution. Havard Univ. Press

McManus JW (1997) Tropical marine fisheries and the future of coral reefs: a brief review with emphasis on Southeast Asia. Coral Reefs 16:121-127

McManus JW, Nañola CL, Reyes RB, Kesner KN (1992) Resource ecology of the Bolinao coral reef system. ICLARM Stud Rev 22, ICLARM, Philippines

Mcmillan WO, Palumbi SR (1995) Concordant evolutionary patterns among Indo-West Pacific butterflyfishes. Proc R Soc Biol Sci 260:229-236

McNeill SE (1994) The selection and design of marine protected areas: Australia as a case study. Biodivers Conserv 3:586-605

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

Meirmans PG (2012) The trouble with isolation by distance. Mol Ecol 21: 2839-2846

Metzger EJ, Hurlburt HE (2001) The nondeterministic nature of Kuroshio penetration and eddy shedding in the South China Sea. J Phys Oceanogr 31:1712-32

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification using restriction site associated DNA (RAD) markers. Genome Res 17:240-248

Milligan BG (2003) Maximum-likelihood estimation of relatedness. Genetics 163:1153-1167

Monmonier MS (1973) Maximum-difference barriers: An alternative numerical regionalization

method. Geogr Anal 5:245-61

Morgulis A, Coulouris G, Raytselis Y, Madden TL, Agarwala R, Schäffer AA (2008) Database indexing for production MegaBLAST searches. Bioinformatics 24:1757-1764

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann Rev Ecol Syst 18:269-92

Muallil RN, Mamauag SS, Cababaro JT, Arceo HO, Aliño PM (2014) Catch trends in Philippine small-scale fisheries over the last five decades: The fishers′ perspectives. Mar Policy 47:110-7

Murphy MA, Evans JS, Cushman SA, Storfer A (2008) Representing genetic variation as continuous surfaces: an approach for identifying spatial dependency in landscape genetic studies. Ecography 31:685–697

Nakajima Y, Matsuki Y, Lian C, Fortes MD, Uy WH, Campos WL, Nakaoka M, Nadaoka K (2014) The Kuroshio Current influences genetic diversity and population genetic structure of a tropical seagrass, *Enhalus acoroides*. Mol Ecol 23:6029-44

Nañola CL, Aliño PM, Carpenter KE (2011) Exploitation-related reef fish species richness depletion in the epicenter of marine biodiversity. Environ Biol Fish 90:405-420

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

Newton K, Cote IM, Pilling GM, Jennings S, Dulvy NK. (2007) Current and future sustainability of island coral reef fisheries. Curr Biol. 17:655-658

Nitani H (1972) Beginning of the Kuroshio. Kuroshio: Physical Aspects of the Japan Current. p 129-163

Novembre J, Johnson T, Byrc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS,

Bergmann S, Nelson MR, Stephens M, Bustamente CD (2008) Genes mirror geography within Europe. Nature 456:98-101

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

Oksanen JF, Blanchet G, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H (2012) vegan: community ecology package. R package version 2.0-4. http://CRAN.R-project.org/package=vegan

O'Reilly PT, Wright JM (1995) The evolving technology of DNA fingerprinting and its application to fisheries aquaculture. J Fish Biol 47:29-55

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. Ecol Appl 13:146-158

Palumbi SR (2004) Marine reserves and ocean neighborhoods: the spatial scale of marine populations and their management. Annu Rev Environ Resour 29:31-68

Pardini AT, Jones CS, Noble LR, Kreiser B, Malcom H, Bruce BD, Stevens JD, Cliff G, Scholl MC, Francis M, Duffy CAJ, Martin AP (2001) Sex-biased dispersal of great white sharks. Nature 412:139-140

Park LA, Moran P (1994) Developments in molecular genetic techniques in fisheries. Rev Fish Biol Fish 4:272-299

Petersen BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive methods for de novo SNP discovery and genotyping in model and non-model species. PLOS ONE 7:e37135

Pew J, Muir PH, Wang J, Frasier TR (2015) related: an R package for analyzing pairwise relatedness from codominant molecular markers. Mol Ecol Resour 15:557–561

Philippine Fisheries Profile (2015) Quezon City: Department of Agriculture – Bureau of Fisheries and Aquatic Resources. Available from: http://www.bfar.da.gov.ph/publication.jsp?id=2345#post

Posada D (2008) jModeltest: Phylogenetic Model Averaging. Mol Biol Evol. 25: 125-1256

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

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431 https://doi.org/10/7717/peerj.431

Puritz JB, Matz MV, Toonen RJ, Weber JN, Bolnick DI, Bird CE (2014) Demystifying the RAD fad. Mol Ecol 23:5937-5942

Qiu B, Lukas R (1996) Seasonal and interannual variability of the North Equatorial Current, the Mindanao Current, and the Kuroshio along the Pacific western boundary. J Geophys Res C 101:12315-12330

R Core Team (2012) R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org

R Core Team (2015) R: A language and environment for statistical computing. R foundation for statistical computing. Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org

Rannala B, Mountain JL (1997) Detecting immigration by using multilocus genotypes. P Nat Acad Sci USA 94:9197-9201

Randall JE (1998) Zoogeography of Shore Fishes of the Indo-Pacific Region. Zool Stud 37: 227-268

Ravago-Gotanco RG, Magsino RM, Juinio-Meñez MA (2007) Influence of the North

Equatorial Current on the population genetic structure of *Tridanca crocea* (Mollusca: Tridacnidae) along the eastern Philippine seaboard. Mar Ecol Prog Ser 336:161-168

Ravago-Gotanco RG, Juinio-Meñez MA (2010) Phylogeography of the mottled spinefoot *Siganus fuscescens*: Pleistocene divergence and limited genetic connectivity in the Philippine archipelago. Mol Ecol 19: 4520-4534

Raynal JM, Crandall ED, Barber PH, Mahardika GN, Ablan-Lagman MCA, Carpenter KE (2014) Contemporary oceanographic features versus Pleistocene land barriers: multiple explanations for population genetic structure in the humbug damselfish (*Dascyllus aruanus*) in the Coral Triangle. Bull Mar Sci 90:513-532

Reader SE, Leis JM (1996) Larval development in the lutjanid subfamily Caesioninae (Pisces): the genera *Caesio*, *Dipterygonotus*, *Gymnocaesio*, and *Pterocaesio*. Bull Mar Sci 59:310-369

Roberts CM, Polunin NVC (1991) Are marine reserves effective in management of reef fisheries? Rev Fish Biol Fish 1:65-91

Roberts CM, McClean CJ, Veron JEN, Hawkins JP and others (2002) Marine biodiversity hotspots and conservation priorities for tropical reefs. Science 295:1280

Rofritsch A, Borsa P (2005) Genetic structure of Indian scad mackerel *Decapterus russelli*: Pleistocene vicariance and secondary contact in the Central Indo-West Pacific seas. Heredity 95:315-326

Rosen BR (1975) The distribution of reef corals. Rep Underwater Assoc 1:1-16

Rowe HC, Renaut S, Guggisberg A (2011) RAD in the realm of next-generation sequencing technologies. Mol Ecol 20:3499-3502

Russ G, Alcala A (1996) Do marine reserves export adult fish biomass? Evidence from Apo

Island, central Philippines. Mar Ecol Prog Ser 132:1-9

Sale PF, Cowen RK, Danilowicz BS, Jones GP and others (2005) Critical science gaps impede use of no-take fishery reserves. Trends Ecol Evol 20:74-80

Sale PF (2006) The science we need to develop for more effective management. In: Sale PF (ed) Coral reef fishes: dynamics and diversity in a complex ecosystem. Academic Press p 361-376

Sanciangco JC, Carpenter KE, Etnoyer PJ, Moretzsohn F (2013) Habitat availability and heterogeneity and the Indo-Pacific warm pool as predictors of marine species richness in the tropical Indo-Pacific. PLoS One 8:e56245

Schott FA, McCreary JP (2001) The monsoon circulation of the Indian Ocean. Prog Oceanogr 51:1-123

Schwartz MK, McKelvey KS (2008) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. Conserv Genet 10:441–452

Seeb JE, Carvalho G, Hauser L, Naish K, Roberts S, Seeb LW (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. Mol Ecol Res 11:1-8

Selkoe KA, Toonen RJ (2006) Microsatelliates for ecologists: a practical guide to using and evaluationg microsatellite markers. Ecol Lett 9:615-629

Shanks AL, Grantham BA, Carr MH (2003) Propagule dispersal distance and the size and spacing of marine reserves. Ecol Appl 12:S159-S169

Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. Biol Bull 216:373-385

Spieth PT (1974) Gene flow and genetic differentiation. Genetics 78: 961-965

Springer VG, Williams JT (1990) Widely distributed Pacific plate endemics and lowered sea-level. Bull Mar Sci 47:631-640

Stobutzki IC, Silvestre GT, Talib AA, Krongprom A, Supongpan M, Khemakorn P, Armada N, Garces LR (2006) Decline of demersal coastal fisheries resources in three developing Asian countries. Fish Res 78:130-42

Stockwell BL, Larson WA, Waples RK, Abesamis RA, Seeb LW, Carpenter KE (2016) The application of genomics to inform conservation of a functionally important reef fish (*Scarus niger*) in the Philippines. Conserv Genet 17:239-249

Sunnucks P (2000) Efficient genetic markers for population biology. Trends Ecol Evol 15:199-203

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, MA

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

Tavaré S (1986) Some probablistic and statistical problems on the analysis of DNA sequences. In: Muira RM (ed) Lectures in Mathematics in the Life Sciences, Vol 17. American Mathemathical Society, p 57-86

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

Toole JM, Millard RC, Wang Z, Pu S (1990) Observations of the Pacific North Equatorial Current bifurcation at the Philippine coast. J Phys Oceanogr. 20:307-318

Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR, Bird CE

(2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ 1:e203

US Census Bureau, International Programs: International Data Base. Revised: 27 September 2016, Version: Data:16.0804 Code:12.0321. Web. 15 March, 2017 https://www.census.gov/population/international/data/idb/informationGateway.php

Vaughan TW (1919) Corals and the formation of coral reefs. Smithsonian Institution Annual Report, 1971, pp 189-238

Vawter L, Brown WM (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science 234:194-196

Veron JEN (1992) Hermatypic corals of Japan. Australian Institute of Marine Science, Townsville, Queensland

Veron JEN, Minchin PR (1992) Correlations between sea surface temperature, circulation patterns and the distribution of hermatypic corals in Japan. Cont Shelf Res 12:835-857

Veron JEN, DeVantier LM, Turak E, Green AL, Kininmonth S, Stafford-Smith M, Peterson N (2009) Delineating the coral triangle. Galaxea, J Coral Reef Studies 11:91-100

Victor BC, Wellington GM (2000) Endemism and the pelagic larval duration of reef fishes in the eastern Pacific Ocean. Mar Ecol Prog Ser 205: 241-248

Villesen P (2007) FaBox: an online toolbox for fasta sequences. Mol Ecol Notes 7:965-968

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

Voris H (2000) Special Paper 2: Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J Biogeogr 27:1153-1167

Voronoi G (1908) Nouvelles applications des paramètres continus à la théorie des formes

quadratiques. J Reine Angewandte Mathematik 133:97–178

Vucetich JA, Waite TA (2003) Spatial patterns of demography and genetic processes across the species' range: null hypotheses for landscape conservation genetics. Conserv Genet 4:639-645

Wahlund S (1928) Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachet. Hereditas 11:65-106

Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10:506-513

Wang J (2007). Triadic IBD coefficients and applications to estimating pairwise relatedness. Genet Res 89:135-153

Wang J (2011) Unbiased relatedness estimation in structured populations. Genetics 187:887-901

Wang S, Meyer E, McKay JK, Matz MV (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat Methods 9:808-810

Wang J (2014) Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. J Evol Biol 27:518-530

Waples R (1998) Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. J Heredity 89: 438–450

Waples RS, Gagiotti 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. Mol Ecol 15:1419-1439

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary N_e using highly

variable genetic markers: a largely untapped resource for applied conservation and evolution. Evol Appl 3:244-262

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

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159

Wright S (1943) Isolation by distance. Genetics 28:11

Yamano H, Hori K, Yamauchi M, Yamagawa O, Ohmura A (2001) Highest-latitude coral reef at Iki Island, Japan. Coral Reefs 20:9-12

Yamano H, Sugihara K, Nomura K (2011) Rapid poleward range expansion of tropical reef corals in response to rising sea surface temperatures. Geophys Res Lett 38: L04601

Yang YJ, Jan S, Chang MH, Wang J, Mensah V, Kuo TH, Tsai CJ, Lee CY, Andres M, Centurioni LR, Tseng YH (2015) Mean structure and fluctuations of the Kuroshio East of Taiwan from in situ and remote observations. Oceanography 28:74-83

Yasuda N, Nagai S, Hamaguchi M, Okaji K, Gerard K, Nadaoka K. (2009) Gene flow of *Acanthaster planci* (L.) in relation to ocean currents revealed by microsatellite analysis. Mol Ecol 18:1574-1590

Yeh FC, Layton C (1979) The organization of genetic variability in central and marginal populations of lodgepine *Pinus contorta* spp. *latifolia*. Can J Genet Cytol 21:487-503

Zhang Z, Schwartz Z, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203-14

Zhang D-X, Hewitt GM (2003) Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects. Mol Ecol 12:563-584

APPENDIX A

CORRECTED AND UNCORRECTED PAIRWISE Φ_{ST} VALUES

Supplemental Table A1. Uncorrected pairwise differences for the first 15 sampling sites. The lower diagonal contains Φ_{ST} values; the upper diagonal contains the corresponding p-values. Significant Φ_{ST} values highlighted in *blue*.

Locality	Medan	Padang	Anyer	Seribu	Karimunjawa	Bali	Lombok	Selayar	Makassar	Tawi Tawi	Ulugan Bay	Bolinao	Honda Bay	Perez	Romblon
Medan	*	0.508	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Padang	-0.008	*	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Anyer	0.532	0.534	*	0.577	0.279	0.111	0.180	0.041	0.370	0.956	0.262	0.061	0.644	0.042	0.158
Seribu	0.586	0.588	-0.016	*	0.629	0.427	0.214	0.542	0.578	0.495	0.534	0.198	0.746	0.128	0.246
Karimunjawa	0.465	0.473	0.007	-0.015	*	0.077	0.380	0.112	0.749	0.325	0.616	0.283	0.810	0.316	0.605
Bali	0.536	0.539	0.022	0.000	0.024	*	0.442	0.073	0.579	0.054	0.314	0.615	0.538	0.173	0.625
Lombok	0.518	0.521	0.017	0.022	0.003	-0.001	*	0.088	0.919	0.080	0.810	0.418	0.387	0.687	0.511
Selayar	0.543	0.548	0.037	-0.013	0.022	0.028	0.032	*	0.317	0.025	0.502	0.165	0.195	0.024	0.250
Makassar	0.501	0.508	0.003	-0.012	-0.012	-0.007	-0.025	0.007	*	0.109	0.761	0.659	0.608	0.459	0.596
Tawi Tawi	0.527	0.530	-0.031	-0.010	0.006	0.041	0.037	0.054	0.028	*	0.180	0.028	0.550	0.021	0.142
Ulugan Bay	0.523	0.527	0.008	-0.007	-0.007	0.005	-0.017	-0.005	-0.012	0.018	*	0.652	0.774	0.632	0.812
Bolinao	0.529	0.536	0.027	0.021	0.007	-0.007	0.000	0.015	-0.009	0.042	-0.007	*	0.350	0.352	0.999
Honda Bay	0.495	0.501	-0.009	-0.027	-0.014	-0.005	0.003	0.014	-0.007	-0.007	-0.013	0.003	*	0.369	0.650
Perez	0.508	0.515	0.032	0.034	0.004	0.013	-0.011	0.040	0.000	0.048	-0.007	0.003	0.003	*	0.342
Romblon	0.520	0.528	0.018	0.019	-0.007	-0.008	-0.004	0.012	-0.007	0.023	-0.017	-0.035	-0.010	0.005	*
Guimaras	0.480	0.488	0.014	0.007	-0.013	0.046	0.008	0.039	0.010	0.011	-0.004	0.018	-0.004	0.000	0.011
Sorsogon	0.533	0.539	-0.005	-0.017	-0.008	0.004	0.008	0.022	-0.007	0.006	-0.011	-0.004	-0.023	-0.014	0.002
N. Occidental	0.516	0.526	0.044	0.058	0.003	0.017	-0.010	0.063	-0.002	0.060	-0.004	0.015	-0.004	-0.022	0.007
N. Oriental	0.543	0.548	-0.013	-0.017	0.002	-0.027	-0.017	-0.033	-0.014	-0.006	-0.040	-0.032	-0.023	0.005	-0.045
Dinagat	0.492	0.496	0.019	0.042	0.008	0.030	-0.007	0.037	-0.006	0.043	-0.002	0.007	0.011	-0.007	0.008
Balingasag	0.519	0.523	-0.023	-0.021	-0.007	0.005	-0.010	0.022	-0.010	-0.015	-0.016	-0.009	-0.016	0.001	-0.009
Davao	0.561	0.567	0.016	0.005	0.009	-0.003	-0.014	-0.011	-0.019	0.040	-0.033	-0.021	-0.002	-0.012	-0.021
Manado	0.517	0.531	0.128	0.152	0.036	0.129	0.077	0.123	0.082	0.107	0.085	0.090	0.080	0.053	0.077
Halmahera	0.538	0.543	0.115	0.146	0.042	0.105	0.053	0.074	0.049	0.124	0.053	0.067	0.089	0.067	0.062
Raja Ampat	0.593	0.598	0.195	0.236	0.114	0.173	0.128	0.165	0.117	0.216	0.137	0.130	0.165	0.142	0.125
Sorong	0.570	0.571	0.041	0.064	0.012	0.076	0.043	0.024	0.021	0.057	0.021	0.040	0.044	0.062	0.027
Manokwari	0.540	0.547	0.021	0.018	-0.011	0.044	0.002	0.038	-0.013	0.032	0.012	0.022	0.018	0.033	0.018
Windesi TC	0.531	0.541	0.074	0.094	0.020	0.067	0.035	0.039	0.012	0.089	0.029	0.033	0.049	0.040	0.021
Karei TC	0.563	0.569	0.085	0.089	0.045	0.100	0.066	0.052	0.039	0.099	0.052	0.081	0.071	0.091	0.071
Yapon	0.515	0.527	0.108	0.129	0.037	0.112	0.063	0.054	0.050	0.108	0.044	0.065	0.072	0.055	0.050
Biak	0.502	0.509	0.064	0.068	0.020	0.078	0.039	0.031	0.018	0.079	0.024	0.041	0.047	0.042	0.029
Fak Fak	0.580	0.588	0.126	0.143	0.048	0.131	0.103	0.069	0.057	0.153	0.075	0.081	0.094	0.085	0.076
Kaimana	0.528	0.533	0.043	0.077	0.018	0.074	0.027	0.026	0.008	0.056	0.017	0.042	0.040	0.037	0.034

Supplemental Table A2. Uncorrected pairwise differences for the last 18 sampling sites. The lower diagonal contains Φ_{ST} values; the upper diagonal contains the corresponding p-values. Significant Φ_{ST} values highlighted in *blue*.

Locality	Guimaras	Sorsogon	N. Occidental	N. Oriental	Dinagat	Balingasag	Davao	Manado	Halmahera	Raja Ampat	Sorong	Manokwari	Windesi TC	Karei TC	Yapen	Biak	Fak Fak	Kaimana
Medan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Padang	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Anyer	0.183	0.527	0.039	0.598	0.071	0.945	0.251	0.003	0.000	0.000	0.049	0.209	0.001	0.004	0.000	0.000	0.000	0.032
Seribu	0.338	0.619	0.080	0.600	0.066	0.707	0.409	0.018	0.007	0.000	0.060	0.301	0.010	0.014	0.006	0.028	0.004	0.027
Karimunjawa	0.789	0.639	0.380	0.409	0.210	0.617	0.304	0.099	0.041	0.000	0.226	0.637	0.116	0.020	0.035	0.076	0.024	0.137
Bali	0.006	0.338	0.168	0.806	0.013	0.313	0.456	0.001	0.001	0.000	0.001	0.080	0.001	0.000	0.000	0.000	0.000	0.001
Lombok	0.257	0.291	0.600	0.670	0.620	0.652	0.602	0.033	0.048	0.002	0.048	0.419	0.057	0.012	0.017	0.025	0.003	0.124
Selayar	0.030	0.128	0.019	0.857	0.013	0.100	0.568	0.006	0.022	0.000	0.149	0.119	0.052	0.041	0.029	0.043	0.028	0.117
Makassar	0.235	0.614	0.474	0.651	0.645	0.667	0.717	0.016	0.038	0.000	0.129	0.648	0.207	0.041	0.019	0.101	0.021	0.289
Tawi Tawi	0.215	0.328	0.026	0.476	0.011	0.687	0.141	0.023	0.003	0.000	0.033	0.161	0.002	0.006	0.001	0.000	0.001	0.025
Ulugan Bay	0.518	0.694	0.505	0.924	0.508	0.833	0.890	0.014	0.035	0.000	0.146	0.269	0.071	0.015	0.032	0.053	0.011	0.168
Bolinao	0.095	0.509	0.180	0.879	0.212	0.667	0.765	0.013	0.010	0.000	0.026	0.193	0.038	0.001	0.006	0.007	0.006	0.023
Honda Bay	0.517	0.936	0.492	0.751	0.136	0.834	0.477	0.021	0.001	0.000	0.022	0.233	0.005	0.002	0.002	0.002	0.001	0.029
Perez	0.415	0.794	0.915	0.374	0.716	0.408	0.625	0.047	0.008	0.000	0.003	0.094	0.016	0.001	0.008	0.005	0.002	0.030
Romblon	0.212	0.379	0.328	0.954	0.247	0.634	0.724	0.030	0.026	0.000	0.114	0.253	0.127	0.003	0.027	0.041	0.013	0.063
Guimaras	*	0.559	0.354	0.363	0.178	0.755	0.213	0.095	0.005	0.000	0.021	0.393	0.024	0.008	0.013	0.011	0.004	0.060
Sorsogon	-0.005	*	0.627	0.603	0.696	0.868	0.725	0.018	0.004	0.000	0.009	0.172	0.012	0.003	0.009	0.008	0.003	0.065
N. Occidental	0.004	-0.011	*	0.282	0.516	0.236	0.523	0.071	0.027	0.000	0.002	0.175	0.034	0.003	0.023	0.016	0.004	0.052
N. Oriental	0.004	-0.012	0.014	*	0.366	0.871	0.914	0.082	0.138	0.005	0.218	0.599	0.278	0.127	0.086	0.181	0.027	0.387
Dinagat	0.009	-0.008	-0.004	0.005	*	0.526	0.543	0.010	0.015	0.000	0.030	0.302	0.053	0.003	0.004	0.004	0.017	0.301
Balingasag	-0.012	-0.019	0.012	-0.033	-0.003	*	0.468	0.016	0.006	0.000	0.089	0.496	0.013	0.002	0.004	0.009	0.002	0.065
Davao	0.018	-0.020	-0.009	-0.055	-0.005	-0.003	*	0.029	0.056	0.003	0.024	0.103	0.066	0.016	0.086	0.078	0.049	0.132
Manado	0.039	0.079	0.054	0.076	0.071	0.086	0.106	*	0.235	0.006	0.006	0.157	0.131	0.033	0.475	0.056	0.027	0.121
Halmahera	0.075	0.086	0.068	0.045	0.048	0.078	0.079	0.025	*	0.136	0.207	0.381	0.555	0.207	0.473	0.237	0.249	0.554
Raja Ampat	0.159	0.178	0.150	0.135	0.126	0.157	0.150	0.162	0.045	*	0.000	0.010	0.008	0.000	0.007	0.001	0.013	0.002
Sorong	0.045	0.053	0.079	0.023	0.035	0.029	0.067	0.124	0.025	0.130	*	0.594	0.309	0.380	0.063	0.388	0.106	0.579
Manokwari	0.003	0.023	0.026	-0.013	0.009	-0.005	0.049	0.042	0.007	0.115	-0.013	*	0.723	0.900	0.142	0.545	0.085	0.866
Windesi TC	0.038	0.048	0.040	0.014	0.022	0.048	0.049	0.037	-0.010	0.093	0.009	-0.023	*	0.661	0.540	0.971	0.290	0.949
Karei TC	0.065	0.077	0.085	0.033	0.057	0.074	0.094	0.093	0.020	0.158	0.003	-0.039	-0.013	*	0.116	0.438	0.060	0.821
Yapon	0.051	0.070	0.054	0.048	0.050	0.081	0.051	-0.008	-0.006	0.111	0.042	0.033	-0.007	0.031	*	0.490	0.169	0.529
Biak	0.035	0.041	0.046	0.022	0.027	0.044	0.037	0.049	0.011	0.107	0.001	-0.008	-0.019	-0.001	-0.003	*	0.275	0.861
Fak Fak	0.090	0.095	0.093	0.081	0.056	0.100	0.083	0.108	0.022	0.108	0.034	0.053	0.009	0.059	0.025	0.010	*	0.306
Kaimana	0.030	0.032	0.041	0.004	0.005	0.030	0.037	0.043	-0.012	0.110	-0.008	-0.033	-0.028	-0.024	-0.007	-0.016	0.010	*

Supplemental Table B1. Pairwise differences corrected with Benjamini and Hochberg's method of control of the false discover rate for the first 16 sampling sites. Adjusted p≤0.01686. Significant Φ_{ST} values highlighted in *blue*.

Locality	Medan	Padang	Anyer	Seribu	Karimunjawa	Bali	Lombok	Selayar	Makassar	Tawi Tawi	Ulugan Bay	Bolinao	Honda Bay	Perez	Romblon	Guimaras
Medan	*	0.508	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Padang	-0.008	*	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Anyer	0.532	0.534	*	0.577	0.279	0.111	0.180	0.041	0.370	0.956	0.262	0.061	0.644	0.042	0.158	0.183
Seribu	0.586	0.588	-0.016	*	0.629	0.427	0.214	0.542	0.578	0.495	0.534	0.198	0.746	0.128	0.246	0.338
Karimunjawa	0.465	0.473	0.007	-0.015	*	0.077	0.380	0.112	0.749	0.325	0.616	0.283	0.810	0.316	0.605	0.789
Bali	0.536	0.539	0.022	0.000	0.024	*	0.442	0.073	0.579	0.054	0.314	0.615	0.538	0.173	0.625	0.006
Lombok	0.518	0.521	0.017	0.022	0.003	-0.001	*	0.088	0.919	0.080	0.810	0.418	0.387	0.687	0.511	0.257
Selayar	0.543	0.548	0.037	-0.013	0.022	0.028	0.032	*	0.317	0.025	0.502	0.165	0.195	0.024	0.250	0.030
Makassar	0.501	0.508	0.003	-0.012	-0.012	-0.007	-0.025	0.007	*	0.109	0.761	0.659	0.608	0.459	0.596	0.235
Tawi Tawi	0.527	0.530	-0.031	-0.010	0.006	0.041	0.037	0.054	0.028	*	0.180	0.028	0.550	0.021	0.142	0.215
Ulugan Bay	0.523	0.527	0.008	-0.007	-0.007	0.005	-0.017	-0.005	-0.012	0.018	*	0.652	0.774	0.632	0.812	0.518
Bolinao	0.529	0.536	0.027	0.021	0.007	-0.007	0.000	0.015	-0.009	0.042	-0.007	*	0.350	0.352	0.999	0.095
Honda Bay	0.495	0.501	-0.009	-0.027	-0.014	-0.005	0.003	0.014	-0.007	-0.007	-0.013	0.003	*	0.369	0.650	0.517
Perez	0.508	0.515	0.032	0.034	0.004	0.013	-0.011	0.040	0.000	0.048	-0.007	0.003	0.003	*	0.342	0.415
Romblon	0.520	0.528	0.018	0.019	-0.007	-0.008	-0.004	0.012	-0.007	0.023	-0.017	-0.035	-0.010	0.005	*	0.212
Guimaras	0.480	0.488	0.014	0.007	-0.013	0.046	0.008	0.039	0.010	0.011	-0.004	0.018	-0.004	0.000	0.011	*
Sorsogon	0.533	0.539	-0.005	-0.017	-0.008	0.004	0.008	0.022	-0.007	0.006	-0.011	-0.004	-0.023	-0.014	0.002	-0.005
N. Occidental	0.516	0.526	0.044	0.058	0.003	0.017	-0.010	0.063	-0.002	0.060	-0.004	0.015	-0.004	-0.022	0.007	0.004
N. Oriental	0.543	0.548	-0.013	-0.017	0.002	-0.027	-0.017	-0.033	-0.014	-0.006	-0.040	-0.032	-0.023	0.005	-0.045	0.004
Dinagat	0.492	0.496	0.019	0.042	0.008	0.030	-0.007	0.037	-0.006	0.043	-0.002	0.007	0.011	-0.007	0.008	0.009
Balingasag	0.519	0.523	-0.023	-0.021	-0.007	0.005	-0.010	0.022	-0.010	-0.015	-0.016	-0.009	-0.016	0.001	-0.009	-0.012
Davao	0.561	0.567	0.016	0.005	0.009	-0.003	-0.014	-0.011	-0.019	0.040	-0.033	-0.021	-0.002	-0.012	-0.021	0.018
Manado	0.517	0.531	0.128	0.152	0.036	0.129	0.077	0.123	0.082	0.107	0.085	0.090	0.080	0.053	0.077	0.039
Halmahera	0.538	0.543	0.115	0.146	0.042	0.105	0.053	0.074	0.049	0.124	0.053	0.067	0.089	0.067	0.062	0.075
Raja Ampat	0.593	0.598	0.195	0.236	0.114	0.173	0.128	0.165	0.117	0.216	0.137	0.130	0.165	0.142	0.125	0.159
Sorong	0.570	0.571	0.041	0.064	0.012	0.076	0.043	0.024	0.021	0.057	0.021	0.040	0.044	0.062	0.027	0.045
Manokwari	0.540	0.547	0.021	0.018	-0.011	0.044	0.002	0.038	-0.013	0.032	0.012	0.022	0.018	0.033	0.018	0.003
Windesi TC	0.531	0.541	0.074	0.094	0.020	0.067	0.035	0.039	0.012	0.089	0.029	0.033	0.049	0.040	0.021	0.038
Karei TC	0.563	0.569	0.085	0.089	0.045	0.100	0.066	0.052	0.039	0.099	0.052	0.081	0.071	0.091	0.071	0.065
Yapon	0.515	0.527	0.108	0.129	0.037	0.112	0.063	0.054	0.050	0.108	0.044	0.065	0.072	0.055	0.050	0.051
Biak	0.502	0.509	0.064	0.068	0.020	0.078	0.039	0.031	0.018	0.079	0.024	0.041	0.047	0.042	0.029	0.035
Fak Fak	0.580	0.588	0.126	0.143	0.048	0.131	0.103	0.069	0.057	0.153	0.075	0.081	0.094	0.085	0.076	0.090
Kaimana	0.528	0.533	0.043	0.077	0.018	0.074	0.027	0.026	0.008	0.056	0.017	0.042	0.040	0.037	0.034	0.030

Supplemental Table B2. Pairwise differences corrected with Benjamini and Hochberg's method of control of the false discover rate for the last 17 sampling sites. Adjusted p≤0.01686. Significant Φ_{ST} values highlighted in *blue*.

Locality	Sorsogon	N. Occidental	N. Oriental	Dinagat	Balingasag	Davao	Manado	Halmahera	Raja Ampat	Sorong	Manokwari	Windesi TC	Karei TC	Yapen	Biak	Fak Fak	Kaimana
Medan	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Padang	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Anyer	0.527	0.039	0.598	0.071	0.945	0.251	0.003	0.000	0.000	0.049	0.209	0.001	0.004	0.000	0.000	0.000	0.032
Seribu	0.619	0.080	0.600	0.066	0.707	0.409	0.018	0.007	0.000	0.060	0.301	0.010	0.014	0.006	0.028	0.004	0.027
Karimunjawa	0.639	0.380	0.409	0.210	0.617	0.304	0.099	0.041	0.000	0.226	0.637	0.116	0.020	0.035	0.076	0.024	0.137
Bali	0.338	0.168	0.806	0.013	0.313	0.456	0.001	0.001	0.000	0.001	0.080	0.001	0.000	0.000	0.000	0.000	0.001
Lombok	0.291	0.600	0.670	0.620	0.652	0.602	0.033	0.048	0.002	0.048	0.419	0.057	0.012	0.017	0.025	0.003	0.124
Selayar	0.128	0.019	0.857	0.013	0.100	0.568	0.006	0.022	0.000	0.149	0.119	0.052	0.041	0.029	0.043	0.028	0.117
Makassar	0.614	0.474	0.651	0.645	0.667	0.717	0.016	0.038	0.000	0.129	0.648	0.207	0.041	0.019	0.101	0.021	0.289
Tawi Tawi	0.328	0.026	0.476	0.011	0.687	0.141	0.023	0.003	0.000	0.033	0.161	0.002	0.006	0.001	0.000	0.001	0.025
Ulugan Bay	0.694	0.505	0.924	0.508	0.833	0.890	0.014	0.035	0.000	0.146	0.269	0.071	0.015	0.032	0.053	0.011	0.168
Bolinao	0.509	0.180	0.879	0.212	0.667	0.765	0.013	0.010	0.000	0.026	0.193	0.038	0.001	0.006	0.007	0.006	0.023
Honda Bay	0.936	0.492	0.751	0.136	0.834	0.477	0.021	0.001	0.000	0.022	0.233	0.005	0.002	0.002	0.002	0.001	0.029
Perez	0.794	0.915	0.374	0.716	0.408	0.625	0.047	0.008	0.000	0.003	0.094	0.016	0.001	0.008	0.005	0.002	0.030
Romblon	0.379	0.328	0.954	0.247	0.634	0.724	0.030	0.026	0.000	0.114	0.253	0.127	0.003	0.027	0.041	0.013	0.063
Guimaras	0.559	0.354	0.363	0.178	0.755	0.213	0.095	0.005	0.000	0.021	0.393	0.024	0.008	0.013	0.011	0.004	0.060
Sorsogon	*	0.627	0.603	0.696	0.868	0.725	0.018	0.004	0.000	0.009	0.172	0.012	0.003	0.009	0.008	0.003	0.065
N. Occidental	-0.011	*	0.282	0.516	0.236	0.523	0.071	0.027	0.000	0.002	0.175	0.034	0.003	0.023	0.016	0.004	0.052
N. Oriental	-0.012	0.014	*	0.366	0.871	0.914	0.082	0.138	0.005	0.218	0.599	0.278	0.127	0.086	0.181	0.027	0.387
Dinagat	-0.008	-0.004	0.005	*	0.526	0.543	0.010	0.015	0.000	0.030	0.302	0.053	0.003	0.004	0.004	0.017	0.301
Balingasag	-0.019	0.012	-0.033	-0.003	*	0.468	0.016	0.006	0.000	0.089	0.496	0.013	0.002	0.004	0.009	0.002	0.065
Davao	-0.020	-0.009	-0.055	-0.005	-0.003	*	0.029	0.056	0.003	0.024	0.103	0.066	0.016	0.086	0.078	0.049	0.132
Manado	0.079	0.054	0.076	0.071	0.086	0.106	*	0.235	0.006	0.006	0.157	0.131	0.033	0.475	0.056	0.027	0.121
Halmahera	0.086	0.068	0.045	0.048	0.078	0.079	0.025	*	0.136	0.207	0.381	0.555	0.207	0.473	0.237	0.249	0.554
Raja Ampat	0.178	0.150	0.135	0.126	0.157	0.150	0.162	0.045	*	0.000	0.010	0.008	0.000	0.007	0.001	0.013	0.002
Sorong	0.053	0.079	0.023	0.035	0.029	0.067	0.124	0.025	0.130	*	0.594	0.309	0.380	0.063	0.388	0.106	0.579
Manokwari	0.023	0.026	-0.013	0.009	-0.005	0.049	0.042	0.007	0.115	-0.013	*	0.723	0.900	0.142	0.545	0.085	0.866
Windesi TC	0.048	0.040	0.014	0.022	0.048	0.049	0.037	-0.010	0.093	0.009	-0.023	*	0.661	0.540	0.971	0.290	0.949
Karei TC	0.077	0.085	0.033	0.057	0.074	0.094	0.093	0.020	0.158	0.003	-0.039	-0.013	*	0.116	0.438	0.060	0.821
Yapon	0.070	0.054	0.048	0.050	0.081	0.051	-0.008	-0.006	0.111	0.042	0.033	-0.007	0.031	*	0.490	0.169	0.529
Biak	0.041	0.046	0.022	0.027	0.044	0.037	0.049	0.011	0.107	0.001	-0.008	-0.019	-0.001	-0.003	*	0.275	0.861
Fak Fak	0.095	0.093	0.081	0.056	0.100	0.083	0.108	0.022	0.108	0.034	0.053	0.009	0.059	0.025	0.010	*	0.306
Kaimana	0.032	0.041	0.004	0.005	0.030	0.037	0.043	-0.012	0.110	-0.008	-0.033	-0.028	-0.024	-0.007	-0.016	0.010	*

APPENDIX B

RELATEDNESS AND ALIGNMENT SUMMARY TABLES

Table A. Coefficients of relatedness (r) from the r package 'related' for San_007 and San_008. Estimates of relatedness are consistent with the individual San_007 being sampled twice.

Sample pair	H_O	Estimators	r	95% CI
San_007-	0.272	Milligan (2003)*	0.967	0.948-0.987
San_008	0.262	Wang (2007)**	0.967	0.948-0.991

*dyadic likelihood estimator, **triadic likelihood estimator

Table B. Mean coefficients of relatedness (r) from the r package 'related' for individuals within populations.

Location	r	
	Milligan (2003)	Wang(2007)
OKI	0.0322	0.0317
ISH	0.0155	0.0152
SAN	0.0114	0.0111
ATI	0.0126	0.0122
GUI	0.0134	0.0131

*dyadic likelihood estimator, **triadic likelihood estimator

Table C. BLASTN alignment summary for RAD tags associated with 5 of 36 outlier loci.

Query ID	Hit ID	Hit Accession #	% Identity	Alignment Length	Mismatches	Gap opens	Query start	Query end	Hit Start	Hit end	e-value	bit score
dDocent_Contig_11633_F	gi|374428415|emb|FQ310507.3|	FQ310507.3	93.827	81	3	2	1	81	2843967	2844045	2.97E-24	121
dDocent_Contig_20147_F	gi|397776255|gb|JQ780820.1|	JQ780820.1	95.918	49	1	1	12	59	2003	1955	1.81E-11	78.7
dDocent_Contig_20147_F	gi|374428414|emb|FQ310506.3|	FQ310506.3	95.918	49	1	1	12	59	1588073	1588121	1.81E-11	78.7
dDocent_Contig_20147_F	gi|374428414|emb|FQ310506.3|	FQ310506.3	93.878	49	2	1	12	59	11848600	11848552	8.44E-10	73.1
dDocent_Contig_20147_F	gi|374428414|emb|FQ310506.3|	FQ310506.3	83.117	77	7	6	12	86	11175136	11175064	1.41E-07	65.8
dDocent_Contig_20147_F	gi|374428414|emb|FQ310506.3|	FQ310506.3	83.117	77	7	6	12	86	11893116	11893044	1.41E-07	65.8
dDocent_Contig_20147_F	gi|374428414|emb|FQ310506.3|	FQ310506.3	97.222	36	1	0	23	58	12070153	12070188	1.83E-06	62.1
dDocent_Contig_20147_F	gi|374428414|emb|FQ310506.3|	FQ310506.3	97.222	36	1	0	24	59	12571976	12572011	1.83E-06	62.1
dDocent_Contig_20147_F	gi|712042627|gb|KJ546039.1|	KJ546039.1	93.878	49	2	1	12	59	1609	1657	8.44E-10	73.1
dDocent_Contig_20147_F	gi|1104859873|dbj|LC056058.1|	LC056058.1	95.556	45	1	1	16	59	28686	28642	3.04E-09	71.3
dDocent_Contig_20147_F	gi|1065196124|gb|KU236380.1|	KU236380.1	100	37	0	0	23	59	919	883	1.09E-08	69.4
dDocent_Contig_20147_F	gi|657596110|ref|XM_008304520.1|	XM_008304520.1	95.349	43	2	0	17	59	3836	3878	1.09E-08	69.4
dDocent_Contig_20147_F	gi|429508162|gb|JQ710660.1|	JQ710660.1	100	37	0	0	23	59	84	48	1.09E-08	69.4
dDocent_Contig_20147_F	gi|1108984874|ref|XM_019270578.1|	XM_019270578.1	97.222	36	1	0	23	58	1051	1086	1.83E-06	62.1
dDocent_Contig_20147_R	gi|374428414|emb|FQ310506.3|	FQ310506.3	88.889	63	7	0	18	80	10013362	10013300	1.81E-11	78.7
dDocent_Contig_34240_R	gi|1108997576|ref|XM_010731103.2|	XM_010731103.2	94.505	91	1	3	1	91	7863	7777	2.95E-29	137
dDocent_Contig_34240_R	gi|657541357|ref|XM_008278362.1|	XM_008278362.1	96.154	78	1	2	14	91	8443	8368	6.39E-26	126
dDocent_Contig_41922_R	gi|992195141|ref|XM_015604406.1|	XM_015604406.1	90.698	86	8	0	2	87	1507	1592	1.38E-22	115
dDocent_Contig_7771_R	gi|374428414|emb|FQ310506.3|	FQ310506.3	94.34	53	2	1	1	52	10385797	10385849	5.05E-12	80.5
dDocent_Contig_7771_R	gi|374428414|emb|FQ310506.3|	FQ310506.3	89.286	56	5	1	1	55	8161016	8160961	1.09E-08	69.4
dDocent_Contig_7771_R	gi|374428416|emb|FQ310508.3|	FQ310508.3	91.071	56	4	1	1	55	13849368	13849313	2.35E-10	75
dDocent_Contig_7771_R	gi|374428415|emb|FQ310507.3|	FQ310507.3	87.5	56	6	1	1	55	8852433	8852488	5.08E-07	63.9

APPENDIX C

Relatedness and Migration Tables

Table A. Coefficients of relatedness (r) with 95% confidence intervals for pairs of individuals from the r package 'related'. Only pairs with r≥0.125 are reported.

Pair	r	
	Milligan (2003)*	**Wang(2007)****
BOH088-BOH089	0.3280 (0.2257-0.4119)	0.3257 (0.2469-0.4245)
NT011-NT013	0.1596 (0.1017-0.2207)	0.1593 (0.1027-0.2214)
PP007-PP012	0.1476 (0.0940-0.2032)	0.1470 (0.0947-0.2042)
NT009-NT019	0.1335 (0.0838-0.1864)	0.1298 (0.0809-0.1847)
NT012-NT015	0.1289 (0.0860-0.1887)	0.1246 (0.0888-0.1955)
NT009-NT011	0.1285 (0.0761-0.1863)	0.1281 (0.0771-0.1883)
NT011-NT030	0.1268 (0.0730-0.1802)	0.1262 (0.0737-0.1812)
NT015-PP046	0.1249 (0.0559-0.1778)	0.1248 (0.0588-0.1814)
NT009-NT012	0.1246 (0.0672-0.1823)	0.1240 (0.0680-0.1837)

*dyadic likelihood estimator, **triadic likelihood estimator

Table B. Modal values of Θ and N_m from MIGRATE-N for the Luzon Strait data subset. Values are from the full model, which had the highest log marginal likelihood and log Bayes factor. Number of migrants (N_m) is estimated from mutation-scaled migration (M).

Site	Code	Θ (95% CI)
BAU	1	0.0190 (0.0165-0.0215)
SAN	2	0.0195 (0.0168-0.0221)
ATI	3	0.0192 (0.0166-0.0217)

$M_{i \rightarrow j}$	$4N_m$ (95% CI)
$M_{2 \rightarrow 1}$	0.0634 (0-3.2985)
$M_{3 \rightarrow 1}$	0.0634 (0-3.2985)
$M_{1 \rightarrow 2}$	0.0650 (0-3.3800)
$M_{3 \rightarrow 2}$	0.0650 (0-3.3800)
$M_{1 \rightarrow 3}$	0.0641 (0-3.3332)
$M_{2 \rightarrow 3}$	0.0641 (0-3.3332)

Table C. Modal values of Θ and N_m from MIGRATE-N for the Sulu Sea data subset. Values are from the full model, which had the highest log marginal likelihood and log Bayes factor. Number of migrants (N_m) is estimated from mutation-scaled migration (M).

Site	Code	Θ (95% CI)
BAU	1	0.0229 (0.0173-0.0202)
BOH	2	0.0221 (0.0167-0.0196)
PP	3	0.0212 (0.0163-0.0188)

$M_{i \rightarrow j}$	$4N_m$ (95% CI)
$M_{2 \rightarrow 1}$	0.0672 (0-3.4961)
$M_{3 \rightarrow 1}$	0.0672 (0-3.4961)
$M_{1 \rightarrow 2}$	0.0650 (0-3.3800)
$M_{3 \rightarrow 2}$	0.0650 (0-3.3800)
$M_{1 \rightarrow 3}$	0.0626 (0-3.2535)
$M_{2 \rightarrow 3}$	0.0626 (0-3.2535)

Table D. Modal values of Θ and N_m from MIGRATE-N for the Central Philippines data subset. Values are from the full model, which had the highest log marginal likelihood and log Bayes factor. Number of migrants (N_m) is estimated from mutation-scaled migration (M).

Site	Code	Θ (95% CI)
BAU	1	0.0430 (0.0396-0.0463)
BOH	2	0.0426 (0.0379-0.0464)
CAT	3	0.0426 (0.0389-0.0368)
GUI	4	0.0473 (0.0383-0.0473)

$M_{i \rightarrow j}$	$4N_m$ (95% CI)
$M_{2 \rightarrow 1}$	0.5737 (0-29.8341)
$M_{3 \rightarrow 1}$	0.5737 (0-29.8341)
$M_{4 \rightarrow 1}$	0.5737 (0-29.8341)
$M_{1 \rightarrow 2}$	0.5684 (0-29.5568)
$M_{3 \rightarrow 2}$	0.5684 (0-29.5568)
$M_{4 \rightarrow 2}$	0.5684 (0-29.5568)
$M_{1 \rightarrow 3}$	0.5676 (0-29.5152)
$M_{2 \rightarrow 3}$	0.5676 (0-29.5152)
$M_{4 \rightarrow 3}$	0.5676 (0-29.5152)
$M_{1 \rightarrow 4}$	0.5836 (0-30.3472)
$M_{2 \rightarrow 4}$	0.5836 (0-30.3472)
$M_{3 \rightarrow 4}$	0.5836 (0-30.3472)

Table E. Modal values of Θ and N_m from MIGRATE-N for the San Bernardino Strait data subset. Values are from the full model, which had the highest log marginal likelihood and log Bayes factor. Number of migrants (N_m) is estimated from mutation-scaled migration (M).

Site	**Code**	**Θ (95% CI)**
ATI	1	0.0200 (0.0173-0.0227)
CAT	2	0.0194 (0.0168-0.0220)
GUI	3	0.0192 (0.0168-0.0217)

$M_{i \rightarrow j}$	**$4N_m$ (95% CI)**
$M_{2 \rightarrow 1}$	0.0668 (0-3.4719)
$M_{3 \rightarrow 1}$	0.0668 (0-3.4719)
$M_{1 \rightarrow 2}$	0.0648 (0-3.3679)
$M_{3 \rightarrow 2}$	0.0648 (0-3.3679)
$M_{1 \rightarrow 3}$	0.0639 (0-3.3228)
$M_{2 \rightarrow 3}$	0.0639 (0-3.3228)

APPENDIX D

TABLE OF IACUC PROTOCOL APPROVAL NUMBERS

Permit Name	Permit Holder	Animals Specified	IACUC Approval Number	Year(s)
Coral Triangle Partnerships for International Research and Education	Kent E. Carpenter	Fishes	11-019	2011-2013
Collaborative Proposal: Documenting Diversity in the Apex of the Coral Triangle: Inventory of Philippine Marine Biodiversity	Kent E. Carpenter	Fishes	13-013	2014

APPENDIX E

PERMISSION FOR REPRINTING PUBLISHED MATERIAL

OLD DOMINION UNIVERSITY
IDEA FUSION

AMANDA ACKISS <aacki002@odu.edu>

Request permission for reuse of published article - MEPS

Ian Stewart <ian.stewart@int-res.com> Tue, Jan 31, 2017 at 3:22 AM
To: AMANDA ACKISS <aacki002@odu.edu>

Dear Amanda

Asserting that Inter-Research is the copyright holder, you are hereby licensed to reproduce an author-formatted and modified version of the article Ackiss et al. (2013) Mar Ecol Prog Ser 480:185–197 https://doi.org/10.3354/meps10199 in your PhD thesis to be submitted to Old Dominion University.

The original published source should be clearly acknowledged.

With best wishes

Ian Stewart
Permissions
Inter-Research
[Quoted text hidden]

VITA

Amanda Susanne Ackiss
Department of Biological Sciences
Old Dominion University, Norfolk, VA 23529

EDUCATION

Bachelor of Arts (Media Studies), University of Virginia, Charlottesville, VA 2002

EXPERIENCE

Graduate Research Assistant 2008-2017

- USAID PEER 3-S Mekong Basin Project
- NSF PIRE/USAID PEER Advanced Genomics Workshop, Manila
- Verde Island Passage Marine Biodiversity Survey, Philippines
- NSF Pan-Pacific Advanced Studies Institute
- Coral Triangle NSF PIRE Project (Philippines/Indonesia/Vietnam/Thailand)
- Euteleost Tree of Life Project

Graduate Teaching Assistant 2011-2016

- Ichthyology (BIOL 420/520)
- Introduction to Biology (BIOL 115/122N/124N)
- Introduction to Human Biology Lab (BIOL 109)
- Environmental Sciences Lab (BIOL 108)

PEER-REVIEWED PUBLICATIONS

Ackiss AS, Pardede S, Crandall ED, Ablan-Lagman M, Carmen A, Barber PH, Carpenter KE (2013) Pronounced genetic structure in a highly mobile coral reef fish, *Caesio cuning*, in the Coral Triangle. Mar Ecol Prog Ser 480:185-197

OTHER PUBLICATIONS

Ackiss AS, Santos TR (2011) Patterns of Gene Flow in the Coral Triangle of the Red-Toothed Triggerfish, *Odonus niger* (Rüppell 1836). Conference paper: Philippine Association of Marine Sciences