# Old Dominion University ODU Digital Commons

**Biological Sciences Theses & Dissertations** 

**Biological Sciences** 

Fall 2017

# Population Structure of Lethrinus Lentjan (Lethrinidae, Percoidei) Across the South China Sea and the Philippines Is Detected With Lane-Affected RADSeq Data

Ellen E. Biesack Old Dominion University, melonfelon@gmail.com

Follow this and additional works at: https://digitalcommons.odu.edu/biology\_etds

Part of the Ecology and Evolutionary Biology Commons, Genetics Commons, and the Natural Resources and Conservation Commons

#### **Recommended Citation**

Biesack, Ellen E.. "Population Structure of Lethrinus Lentjan (Lethrinidae, Percoidei) Across the South China Sea and the Philippines Is Detected With Lane-Affected RADSeq Data" (2017). Master of Science (MS), Thesis, Biological Sciences, Old Dominion University, DOI: 10.25777/ky0g-w988 https://digitalcommons.odu.edu/biology\_etds/21

This Thesis is brought to you for free and open access by the Biological Sciences at ODU Digital Commons. It has been accepted for inclusion in Biological Sciences Theses & Dissertations by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

# POPULATION STRUCTURE OF LETHRINUS LENTJAN

# (LETHRINIDAE, PERCOIDEI) ACROSS THE SOUTH

#### CHINA SEA AND THE PHILIPPINES IS DETECTED

# WITH LANE-AFFECTED RADSEQ DATA

by

Ellen E. Biesack B.S. May 2012, University of North Carolina Wilmington

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

#### MASTER OF SCIENCE

# BIOLOGY

OLD DOMINION UNIVERSITY December 2017

Approved by:

Kent E. Carpenter (Director)

Daniel Barshis (Member)

Christopher E. Bird (Member)

#### ABSTRACT

# POPULATION STRUCTURE OF *LETHRINUS LENTJAN* (LETHRINIDAE, PERCOIDEI) ACROSS THE SOUTH CHINA SEA AND THE PHILIPPINES IS DETECTED WITH LANE-AFFECTED RADSEQ DATA

Ellen E. Biesack Old Dominion University, 2017 Director: Dr. Kent E. Carpenter

Southeast Asia includes the Coral Triangle, a marine biodiversity hotspot that supports important fishery resources experiencing varied threats. Patterns of speciation and population structure in the Coral Triangle have been examined to test hypotheses relating to the historical geologic processes that may have influenced this biodiversity phenomenon. This study investigates the genetic population structure of the Pink-ear Emperor Snapper, Lethrinus lentjan (Lacepède, 1802), across the Philippines and the South China Sea. The species is fished throughout the Coral Triangle by subsistence and commercial fishers and their landings have been in decline for several years, which could be indicative of depleted stocks. Six locations were sampled representing the northeastern, southeastern, central, and western Philippines, and central and southern Vietnam. This study used restriction-site associated DNA sequencing (RADSeq) to sample single nucleotide polymorphism markers from throughout the genome. Here, RADseq data is successfully used to detect structuring within the Philippines where a previous study of L. lentjan mitochondrial control region sequences did not. Genetic structure analyses revealed significant divergence along the boundaries of repeatedly isolated ocean basins, as observed in several Indo-Pacific species. Pleistocene vicariance is a suspected driving factor in lineage diversification for this species, supporting the hypothesis that the Coral Triangle is a center of origin. Before management or conservation strategies can be implemented, a stock

assessment should be completed for *L. lentjan*, including research on its life history, ecosystem services, and metapopulation dynamics. This study also used STRUCTURE analysis to detect a strong Illumina sequencer lane effect, where the genotype of a fish is associated with the sequencer lane used to generate the data. To control and remove this lane effect, the data was parsed into two clusters, a new method for salvaging lane-affected data. It is recognized that RADSeq is prone to lane effects that can result in erroneous conclusions, and several strategies are outlined for identifying, mitigating, and avoiding them in future high-throughput RADSeq studies.

Copyright, 2017, by Ellen E. Biesack and Kent E. Carpenter, All Rights Reserved.

This thesis is dedicated to my late friend Stanley J. Biel, who never ceased to impress upon me the value of getting an education.

"Never tell me the odds."

-Han Solo

#### ACKNOWLEDGMENTS

I must first and foremost acknowledge the support of my advisor, Dr. Kent Carpenter. It has been a once-in-a-lifetime opportunity to not only work as a member of both the Global Marine Species Assessment team and the molecular systematics laboratory, but also to participate in several influential conservation and fisheries projects with colleagues around the world. The access to Dr. Carpenter's vast breadth of experience and knowledge has made my time at Old Dominion incredibly valuable. I would also like to thank my committee members, Drs. Daniel Barshis and Christopher Bird, for lending their expertise to train me in the lab and in computing.

I owe much of my success, and my sanity, to my lab mates, who have given sound advice for traversing the hills and valleys of academia, and who have also fulfilled the important role of being great friends. I'd like to recognize the contributions of my fellow molecular ecologists, the soon-to-be-Drs. Amanda Ackiss and Brian Stockwell, and all the current and former GMSA lab members whose commitment to preserving biodiversity is my daily inspiration. Thanks also to my very supportive family and friends, in North Carolina, Norfolk, and elsewhere, for their unwavering love and support.

Several people contributed to sample and data collection for this project and deserve thanks: A. Lizano, T. Nañola, J. Casaban, E. Gumiran, B. Dang, Q. Vu, O. Truong, T. Nguyen, and J. Williams. Additional thanks must go out to Drs. Jonathon Puritz and Garrett McKinney for their bioinformatics advice and help. The main funding for this project came from the National Science Foundation's Partnerships in International Research and Education grant (OISE-0730256). Additional funding came from USAID's Partnerships for Enhanced Engagement in Research grant "Conservation genetics for improved biodiversity and resource management in a changing Mekong Delta", NSF's Division of Environmental Biology grant (12576304), and the U.S. Smithsonian Institution National Museum of Natural History.

# TABLE OF CONTENTS

Page
------

LIST OF TABLES	X
LIST OF FIGURES	xi
INTRODUCTION	1
GENETIC STRUCTURE IN THE CORAL TRIANGLE	2
STUDY SPECIES AND LIFE HISTORY CHARACTERISTICS	5
MANAGEMENT AND PHYLOGEOGRAPHIC OBJECTIVES	7
LANE EFFECTS IN RADSEQ	10
MATERIALS AND METHODS	12
SAMPLING DESIGN	
SAMPLE COLLECTION	
GENETIC LIBRARY PREP AND SEQUENCING	
SNP DISCOVERY	15
DIAGNOSING A LANE EFFECT	16
STATISTICAL ANALYSES OF POPULATION STRUCTURE	16
ESTIMATING MIGRATION	19
RESULTS	
PARSING THE LANE EFFECT	22
POPULATION STRUCTURE	
MIGRATION	
DISCUSSION	
RADSEQ CHALLENGES	
RADSEQ AND LANE EFFECTS	32
POPULATION STRUCTURE IN LETHRINUS LENTJAN	
POPULATION STRUCTURE ACROSS THE PHILIPPINES	
POPULATION STRUCTURE ACROSS THE SOUTH CHINA SEA	
POPULATION STRUCTURE ALONG THE COAST OF VIETNAM	
CONTRASTS WITH PREVIOUS STUDY OF LETHRINUS LENTJAN	39
RECOMMENDATIONS FOR MANAGEMENT	40
CONCLUSIONS	42
REFERENCES	44

APPENDICES	
A. AMOVA GROUPINGS TESTED FOR POOL 1	
VITA	64

# LIST OF TABLES

Table	Page
1. Sampling information	13
2. Summary of SNP panels and AMOVA and Mantel results for Pools 1 and 2	23
3. Summary of genetic diversity indices for Pools 1 and 2	24
4. Results of relatedness analysis	24
5. Pairwise differences for neutral loci and loci under selection for Pools 1 and 2	25
<ul><li>6. Bezier approximation scores for models of migration tested in Migrate-n for Pools</li><li>1 and 2</li></ul>	29
7. Estimates of parameters returned by Migrate-n for selected models of migration for Pools 1 and 2	30

# LIST OF FIGURES

Figure	Page
1. Map of the Coral Triangle with major oceanographic features and sample sites	2
2. STRUCTURE plots for pool assignment and Pools 1 and 2	17
3. Models of migration tested in Migrate-n	21
4. Principle components and principle coordinates analyses of neutral SNP loci and loci under selection for Pools 1 and 2	28

#### **INTRODUCTION**

Despite apparent widespread larval dispersal and few persistent physical barriers apart from dry land, patterns of genetic diversification reveal themselves in many marine species, from sedentary invertebrates (Lessios et al. 2000; Uthicke & Benzie 2003) to highly mobile ichthyofauna (Colborn et al. 2001; Block et al. 2005; Selkoe et al. 2014). One extensively studied and yet poorly understood regional example of this phenomenon is the Coral Triangle (Barber et al. 2000; Barber 2009), the most biodiverse marine realm on the planet (Roberts et al. 2002; Hoeksema 2007). It has been hypothesized that this concentration of biodiversity is due to lineage diversification resulting from the isolation of marine basins during historical periods of low sea level (Benzie 1998; Crandall et al. 2008b; Barber et al. 2011), the accumulation of varied forms from peripheral Pacific islands (Jokiel & Martinelli 1992), the collision of four tectonic plates (Santini 2002), or the existence of unique ocean currents and geographic features resulting in complex habitats and abundant refugia (Stehli & Wells 1971; Sanciangco et al. 2013). Studying patterns of population genetic structure can shed light on which of these mechanisms is most influential in this remarkable region that includes the nations of Indonesia, Malaysia, Papua New Guinea, the Philippines, the Solomon Islands, and Timor l'Este (Fig. 1).



**Fig. 1** Map of the Coral Triangle with major oceanographic features and sample sites. Sample sites are indicated with solid circles and are numbered: 1) Atimonan, 2) Cebu, 3) Mati, 4) Puerto Princesa, 5) Nha Trang, and 6) Phú Quốc. Single-pointed arrows indicate year-round flow, while double-pointed arrows represent seasonally reversing flow. Ocean currents are based on findings of Nan *et al.* 2015, Hu *et al.* 2009, and Inoue & Welsh 1993. Pleistocene lowest sea level stands are filled in light grey and are based on Voris 2000.

#### Genetic Structure in the Coral Triangle

The Philippine archipelago, situated at the apex of the Coral Triangle, is a "center of the center" of marine biodiversity with the highest concentration of marine species per unit area in the Coral Triangle (Carpenter & Springer 2005; Allen 2008; Sanciangco *et al.* 2013), and it has

been posited that historic basin isolation is a leading cause of its species richness (Benzie 1999a; Ravago-Gotanco & Meñez 2010; Carpenter *et al.* 2011). Contemporary flow from the North Equatorial Current (NEC) drives modern-day surface water currents throughout the Philippines, which in turn facilitate larval transport and gene flow through the complex system of islands (Inoue & Welsh 1993; Han *et al.* 2009; Fig. 1). This flow was presumably restricted during periods of low sea level, when several of the seas connecting the more than 7,000 presently emergent Philippine islands were partially or fully enclosed by land bridges (Voris 2000 - Fig. 1). At the lowest sea levels (120 meters below present levels), a nearly continuous, north-tosouth land mass formed from the eastern islands of the archipelago. This blocked direct flow of surface water from the Pacific Ocean to the inner Philippine seas repeatedly throughout the Pleistocene Epoch, and was present most recently during a glacial maximum 17,000 years ago (Voris 2000).

Ancient oceanographic patterns potentially affected gene flow, and evidence of their influence is revealed in contemporary patterns of population and species structure at scales larger than the Philippine archipelago (Benzie 1998; Crandall *et al.* 2014). Basin isolation influenced ocean currents and was integral to the formation of species in the Coral Triangle in particular (Timm *et al.* 2008). Intraspecific patterns of structure aligning with previously isolated basins across the region are also seen in several taxa (Barber *et al.* 2002; Kochzius *et al.* 2009). The Pleistocene isolation of seas, including the Celebes and Sulu Seas, has been cited as a driver of lineage diversification in rabbitfishes (Ravago-Gotanco & Meñez 2010), a damselfish (Raynal *et al.* 2013), and in species of seahorses (Lourie *et al.* 2005), which display east-to-west differentiation within the Philippines despite contemporary surface currents driven by the NEC (Fig. 1). Populations in the South China Sea over the Pleistocene-emergent Sunda Shelf are

genetically distinct from populations to the north in several taxa, an effect that has been ascribed to rapid expansion onto the shelf when it was inundated (Nelson *et al.* 2000; Crandall *et al.* 2012).

Modern-day surface currents flowing over areas of deep water may reinforce ancient patterns of isolation by preventing larvae and weak swimmers from crossing channels between shelf areas and shorelines (Barber *et al.* 2006; Bird *et al.* 2007). For example, some fish populations have maintained differentiation between the northern and southeastern Philippines because of the opposing directions of two prevailing ocean currents, the Kuroshio and the Mindanao currents, which result from the bifurcation of the NEC as it encounters the eastern face of the archipelago (Ravago-Gotanco *et al.* 2007; Magsino & Juinio-Meñez 2008; Fig. 1). Another barrier to dispersal located between the eastern and western South China Sea was found by modeling the spread of pelagic larvae (Treml *et al.* 2015) and may be maintained by strong currents carrying surface water between the Pacific and Indian oceans that are influenced by monsoonal variations in wind direction (Hu *et al.* 2000; Fig. 1).

Many of these surface currents obscure historical boundaries to gene flow after the inundation of physical barriers following Pleistocene glacial maxima, and several taxa apparently exhibit panmixia across the Coral Triangle (Crandall *et al.* 2008a; Horne *et al.* 2008; Kochzius *et al.* 2009). Modern oceanographic currents may allow highly mobile organisms to disperse over long distances that were previously impassible under lowered sea levels. Several open channels that could promote such dispersal exist throughout the region. For example, the Indonesian Throughflow carries surface water from the Pacific Ocean south through the Makassar Strait between Borneo and Sulawesi and empties into the Indian Ocean thousands of kilometers away (Fig. 1). Today, surface currents connect the inner Philippine seas, and boundaries between once

fragmented ocean basins are now corridors for dispersal (Treml *et al.* 2015). The Mindoro-Panay Throughflow may act as a vector for dispersal of parrotfish (*Scarus niger*) larvae from one island to another 400 km to the south (Stockwell *et al.* 2015). The Sulu Sea Throughflow has also been shown to transport larvae from the South China Sea to the Coral Triangle in oceanographic models (Kool *et al.* 2011; Treml *et al.* 2015; Fig. 1) and potentially serves as a barrier to eastwest transport of larvae across the Sulu Sea (Raynal *et al.* 2013).

# Study Species and Life History Characteristics

The population structure of the important foodfish *Lethrinus lentjan* (Lacepède, 1802) has been investigated in the Coral Triangle using non-coding mitochondrial control region sequences to inform stock delineation of the species (Hines 2013). Commonly known as the Pink-ear Emperor Snapper, *L. lentjan* has a large range extending from Madagascar and the Red Sea east to the Ryukus and Tonga (Carpenter & Allen 1989). It is reef-associated and captured in many shallow-water and small-scale fisheries throughout Southeast Asia. The mitochondrial study found large-scale patterns in population structure in *L. lentjan*, including a genetic break along the Sunda Shelf separating the Indian and Pacific oceans that has been observed in several other species (Arnaud *et al.* 1999; Benzie 1999b; Crandall *et al.* 2008a; Gaither *et al.* 2010; Ackiss *et al.* 2013). However, there is apparent panmixia within the Philippines and across much of central Asia, as found by traditional Sanger-sequencing techniques (Hines 2013). Despite sampling six sites throughout the Philippines, no structure was detected within the archipelago, and the best-supported AMOVA models grouped all Philippines sampling sites with others from Indonesia and central Vietnam.

Life history characteristics of *L. lentjan* suggest it has some potential, however, for range restriction and population differentiation. Young settle and develop in shallow coastal waters,

most often over sand flats and in mangrove swamps, before migrating to coral reefs and surrounding habitats (Carpenter & Allen 1989). Adults are moderately mobile, but may adhere to smaller home ranges than their morphology is capable of. Most species limit their dispersal when they find suitable habitat, despite the physical ability to disperse further (Endler 1977; Grüss *et al.* 2011), and two congeners, *L. harak* and *L. obsoletus*, show high fidelity to reefs defended by agonistic behavior (Nanami & Yamada 2009; Taylor & Mills 2013). *Lethrinus lentjan* likely displays similar behaviors, as these species have all been shown to share similar evolutionary history and morphology (Lo Galbo *et al.* 2001).

A potentially stronger influence on gene flow in *L. lentjan* is pelagic larval duration (PLD), a measure of how long marine organisms develop in a free-swimming and drifting larval stage. Several researchers have suggested a negative correlation between PLD and population structure (Doherty et al. 1995; Kinlan & Gaines 2003; Cunningham et al. 2009; Young et al. 2015), though other studies have been inconclusive (Bay et al. 2006; Weersing 2007; Riginos et al. 2011; Selkoe et al. 2014). The expectation is that organisms free to disperse as larvae will promote sharing of genetic material between populations, resulting in panmixia. The maintained exchange of one individual on average per generation is enough to prevent allele fixation at a divergent locus in two populations (Slatkin 1987), so in the absence of strong barriers, larval dispersal alone can oppose population differentiation, even over long distances. This mode of dispersal is particularly important for sedentary organisms, or those with very limited home ranges, such as L. lentian, to expand their species range. The PLD of L. lentian in the Great Barrier Reef has been estimated at 30 days (Wilson 1998), and a congener Lethrinus nebulosus has a PLD of 37 days. Individuals of the closely-related family Lutjanidae from Australia and the Indo-Pacific employ PLDs of 25 to 40 days (Brothers et al. 1983; Brothers & Thresher 1985). L.

*lentjan*'s protracted spawning period, assumed from Great Barrier Reef populations (Currey *et al.* 2013), would allow larvae produced throughout the year to use seasonally reversing currents to disperse throughout the region.

#### Management and Phylogeographic Objectives

Life history characteristics like home range, PLD, and spawning behaviors can help researchers design species-specific hypotheses for stock connectivity, which has consequences for conservation and fisheries management, especially for fishes that are heavily exploited. Globally, capture of *L. lentjan* has been steadily decreasing from its peak in 2009 with total landings around 10,300 metric tonnes (FAO Fisheries & Aquaculture Department 2016), which may suggest declines in wild stocks. Emperor snappers make up an important portion of subsistence and commercial fish landings in the Philippines, and *L. lentjan* is commonly seen in markets depending on location and time of year. The status of *L. lentjan* fisheries has been assessed in the Red Sea, Arabian Gulf, and India, with all reports concluding that the fish is not overexploited, at least in those areas (Kedidi *et al.* 1984; Grandcourt *et al.* 2011; Vasantharajan *et al.* 2015). Despite declines in global landings, its persistence in markets makes it questionable that *L. lentjan* is in immediate danger from overfishing alone in the Coral Triangle.

The Philippines is, however considered to be one of the most at-risk hotspots of coral reef biodiversity in the world (Roberts *et al.* 2002). Coral cover there has already declined to one of the lowest site averages in the Indo-Pacific (Bruno & Selig 2007). Coastal and reef areas are increasingly affected by sedimentation, destructive fishing practices (such as dynamite and poison fishing), tourism, harvest for food and curios, and pollution (Cesar 2000). Roughly half of reef-building coral species in the Coral Triangle are at an elevated risk of extinction (Carpenter *et al.* 2008). Populations in the South China Sea, bordering the Philippines to the west,

experience fishing pressure from several nations, and the destruction of shallow reefs and small islands for the construction of military outposts has raised concerns over the loss of stepping stone habitats and sources of larvae for fished species (McManus *et al.* 2010). An abundance of low-lying land in Vietnam makes it particularly vulnerable to sea level rise and saltwater intrusion (Dasgupta *et al.* 2009), which could result in range expansion of marine species into freshwater habitats. These impacts, in combination with exploitation, may be causing localized declines in species such as *L. lentjan* that have not been well studied and whose global conservation status may currently be of "Least Concern" (Carpenter *et al.* 2016).

Research on populations of exploited fishes is becoming increasingly important as nations rely more and more on marine life for their food and economies. Fisheries and a growing marine tourism industry contribute billions of dollars each year to the economies of Southeast Asia, and their populations rely largely on fish for their main source of protein (FAO Fisheries & Aquaculture Department 2016). Philippine marine fisheries in particular have seen a leveling off in landings despite increased fishing effort, an indication of overexploitation and depleted stocks (Silvestre & Pauly 1997). The national economy and the livelihoods of millions of Filipinos are at risk from coral reef destruction and overfishing (White *et al.* 2000). As previously mentioned, *L. lentjan* juveniles rely on mangrove nursery habitat, an ecosystem that is at risk from deforestation (Polidoro *et al.* 2016). Although adults may not be obligatorily dependent on reefs for their livelihood, functional coral reefs are the foundation of shallow water marine diversity in the Coral Triangle, and they sustain prey species for emperors and other fish.

Conservationists can use findings from lineage diversification studies to define intra- and interspecific management units, including ecozones (Spalding *et al.* 2007) and evolutionarily significant units (Fraser & Bernatchez 2001), which are being used increasingly to inform

conservation efforts. In the attempt to mitigate damage from potential threats, many governments and international organizations have made great progress in the development of marine protected areas (MPAs) based on these concepts in the Coral Triangle (White *et al.* 2000). Currently about 2.5% of Vietnam's Exclusive Economic Zone and 0.9% of the Philippines' is protected in some capacity (Marine Conservation Institute 2017). MPAs vary in their levels of protection, from notake to regulated harvest, and the goal is often to protect sources of propagules and spawning areas in order to return populations to virgin (unfished) levels (Harrison *et al.* 2012). MPAs designed to accommodate both metapopulation dynamics of multiple species of concern and relevant and damaging human impacts can tackle many issues at once (Crowder *et al.* 2000).

Fished species can benefit from site-based conservation strategies like MPAs, but they also require monitoring of stock levels to document the effects of harvest on populations. A stock assessment has yet to be completed for *L. lentjan* anywhere in the Coral Triangle and stock delineation is a necessary first step in assessing the status of populations, defining management units, and applying protections to conserve biodiversity and fishes as sources of food (Begg *et al.* 1999). Several characteristics that can be used for identifying stocks (Ihssen 1991) have been measured in localized ranges of *L. lentjan*, including growth rates (Toor 1964, Wassef 1991; Grandcourt 2002), trace metals in tissues (Al-Yousuf *et al.* 2000), and population genetics (Hines 2013). Stocks tend to act as discrete breeding units and their structure is often detectable by genetic studies, which are easier than traditional tag and recapture methods (Schwartz *et al.* 2006). Molecular methods have become more available in recent years to researchers in fisheries and conservation (Carvalho & Hauser 1994). *L. lentjan* is part of the snapper-grouper fisheries complex, members of which have been found to be adversely and swiftly affected by the introduction of fishers to previously unfished areas (Russ & Alcala 1996). As fishing efforts

increase with population growth in developing regions like Southeast Asia, robust fisheries science and management is imperative.

This study is one of the first to pursue a phylogeographic question using advanced genomic sequencing techniques in the Indo-Pacific region. It attempts to detect structuring of *L. lentjan* populations within the Philippines and across the South China Sea using restriction site-associated DNA sequencing, or RADSeq, where traditional Sanger sequencing of a mitochondrial gene showed panmixia (Hines 2013). It is hypothesized that *L. lentjan* will display differentiation at the genome-wide level in a pattern corresponding with historic basin isolation resulting from persistent limits to gene flow during Pleistocene fluctuations in sea level. The findings of this study should contribute to the development of specific hypotheses for how speciation and diversification arise in the Coral Triangle, and to designing solutions for the continuing loss of its marine biodiversity. The methods described in this study can be adopted for stock structure analysis for almost any harvested or otherwise threatened marine species, including non-model organisms such as *L. lentjan* (Toonen *et al.* 2013).

#### Lane Effects in RADseq

In addition to examining patterns of genetic structure of a popular southeast Asian foodfish, this study aims to explore the effects of biases in high-throughput sequencing that originate from discrepancies between different sequencing lanes, known as "lane effects" (Morris *et al.* 2011; Mastretta-Yanes *et al.* 2015; Stetter *et al.* 2017). These biases may cause erroneous results if they are ignored and resulting patterns of genetic structure are ascribed to real biological divergence (Gilad & Mizrahi-Man 2015; Leek *et al.* 2010; Xing *et al.* 2007). Thus far, the impact of these effects have mainly been examined for microarray data (referred to as "batch effects"), but their prevalence in genotyping by sequencing methods, such as Illumina sequencing, is gaining attention in the literature. This study will demonstrate a method for diagnosing and working with lane-affected data and will provide advice for designing sequencing projects with potential lane effects in mind.

#### MATERIALS AND METHODS

#### Sampling Design

Six study sites were chosen to test for differentiation among *Lethrinus lentjan* populations across the Philippines and South China Sea (Fig. 1). Sites were selected from the Philippines to represent the northern and southern Pacific coast (Atimonan and Mati, respectively), the central seas (Cebu), and the western Philippines (Puerto Princesa). Two sites in Vietnam were also sampled – one north of the Mekong River outflow (Nha Trang) and one south (Phú Quốc). These two sites were chosen to test for population structure across the South China Sea (Treml *et al.* 2015) and to take into account possible differentiation across the Mekong River outflow, as previously observed in mitochondrial DNA of *L. lentjan* (Hines 2013).

#### Sample Collection

*Lethrinus lentjan* can be purchased from fish markets in the Coral Triangle throughout much of the year. A total of 355 tissue samples from individual *L. lentjan* were collected from the six sites between 2013 and 2016 (Table 1). Fish were either purchased whole from markets and landings or fin clips were collected from particularly generous vendors. Tissues designated for genetic study were preserved in 95% molecular grade ethanol. When available, a manufactured buffer, such as RNAlater or DNA/RNA Shield (Thermo Fisher Sciencific, Waltham, MA and Zymo Research, Irvine, CA, respectively), was used. During transport, samples were stored at room (<23°C) or refrigerated temperatures (4°C), and kept out of direct sunlight, until they could be permanently stored in a freezer (-20°C).

Site Name	Site #	Short- hand	Latitude	Longitude	Number Collected	Collection Dates	Number Sequenced	Number Pool 1	Number Pool 2
Philippines									
Atimonan	1	Ati	14° 00' 19" N	121° 55' 11" E	54	Apr. 2014	41	34	6
Cebu	2	Ceb	11° 03' 33" N	124° 00' 42" E	62	Jan., Feb. 2015	44	6	37
Mati	3	Mat	06° 57' 9" N	126° 12' 22" E	46	May 2014, spring 2015	40	9	9
Puerto Princesa	4	PtP	09° 56' 27" N	118° 47' 42" E	77	Apr. 2014	45	22	22
Vietnam									
Nha Trang	5	NhT	12° 15' 41" N	109° 12' 01" E	86	Sep. 2013, summer 2014, spring 2016	35	19	6
Phú Quốc	6	PhQ	10° 08' 10" N	103° 57' 12" E	30	Spring 2016	30	21	0

**Table 1** Sampling information. Sample sites are listed with country, site number as appearing in Figure 1, shorthand used for sites in tables and figures, global position, the number of samples collected, the collection dates, the number successfully sequenced, and the number analyzed for Pools 1 and 2

#### Genetic Library Prep and Sequencing

DNA extraction was carried out with the Omega Bio-Tek Blood & Tissue Extraction Kit (Norcross, GA). A serial series of elutions were performed, where DNA is freed from the silica matrix with 100  $\mu$ L of elution buffer, and each elution was captured in a separate well from the others. Each DNA elution for each fish was visualized via electrophoresis on a 1.5% agarose gel stained with SYBR Safe (Invitrogen, Thermo Fisher Scientific). The elution with maximum representation of high-weight material (seen as a bright band between 7,000 and 10,000 base

pairs on the gel) and minimal degradation (seen as lack of a smear occurring below 1,000 base pairs on the gel) was selected for each fish. The chosen elutions were quantified using the AccuBlue High Sensitivity dsDNA Quantitation Kit (Biotium, Hayward, CA), following the provided protocol, and a FLUOstar Optima Plate Reader (BMG Labtech, Cary, NC). One hundred nanograms of extracted DNA was either concentrated using the Savant DNA120 SpeedVac vacuum centrifuge (Thermo Fisher Scientific) or diluted with water to reach the desired volume for input to the ezRAD protocol developed by Toonen *et al.* (2013). Following a 2:1 beads-to-sample volume cleanup protocol using Agencourt AMPure XP Beads (Beckman Coulter, Indianapolis, IN), samples were digested with the isoschizomeric restriction enzymes *MboI* and *Sau3AI* (New England BioLabs, Ipswich, MA). The library preparation protocol used one-third reactions of the Illumina TruSeq Nano HT Library Prep Kit (Illumina, San Diego, CA), with PCR amplification of libraries as described by Toonen *et al.* (2013).

Ligation of Illumina adapters and dual-indexed barcodes was confirmed via qPCR before pooling of samples for sequencing and final size selection with a Pippin Prep machine (Sage Science, Beverly, MA) at the Texas A&M University – Corpus Christi Genomic Core Laboratory. Preliminary size selection was completed for the range of 400 to 550 base pairs (bp) for a pilot sequencing run. This resulted in low coverage of reads and was determined to be too broad for the sampling done by the two enzymes selected. For the full sampled set, a size range of 500 to 550 base pairs, successfully tested on a species with a similar genome size to *Lethrinus* species with known c-values, was used. Paired-end (PE) 100 sequencing was run at New York University (NYU; New York City, NY) on five lanes of an Illumina HiSeq 2500 machine.

RAD tags were trimmed to 90 bp with the process radtags command from the program Stacks v.1.20 (Catchen et al. 2013) before being run through the dDocent pipeline for RAD data for *de novo* reference assembly, mapping of reads, and single nucleotide polymorphism (SNP) calling (Puritz et al. 2014). The reference was assembled from reads with coverage depth of 4x or more and represented in ten or more sampled individuals. SNPs were called by FreeBayes (Garrison & Marth 2012) with a minimum quality score of 20. A custom R code was used to remove polymorphisms represented with a minimum depth of 3 in less than about 60% of individuals from each site, varying slightly based on sample sizes. Individuals were assessed for percentage of missing data and those with greater than 20% missing genotypes were eliminated from the dataset. Putative SNP filtering was continued using VCFtools (Danecek et al. 2011) to select polymorphisms with a minimum mean depth of 5x and minor allele frequency of 5%. SNPs were also filtered to eliminate indels and loci genotyped on both forward and reverse reads using VCF filter and to eliminate possible paralogs using the perl script rad haplotyper (Hollenbeck 2017). One random SNP was selected from each contig following filtering, resulting in a final panel of putatively true, unlinked SNPs. The variant call format (VCF) file was then converted to a Genepop file in PGDSpider (Lischer & Excoffier 2012) for further analyses. Heterozygosity of loci and samples was examined using scripts developed by G. McKinney (University of Washington) and those with exceptionally low or high heterozygosity were eliminated. The programs Lositan (Beaumont & Nichols 1996; 50,000 simulations, infinite alleles model, false discovery rate [FDR] correction of 0.05) and Bayescan v.2.1 (Foll & Gaggiotti 2008; FDR correction 0.05) were used to identify SNPs under directional and balancing selection to be removed from the dataset for analyses.

A correlation between sequencing lane and genotype was tested for using STRUCTURE (Hubisz et al. 2009; Pritchard et al. 2000), and a significant lane effect was detected in the putatively neutral SNPs among individuals passing quality filtering. STRUCTURE analysis was run with a burn-in of 20,000 steps and 10 iterations of 50,000 Markov chain Monte Carlo (MCMC) repetitions for each K value, 1 through 6. The most likely K value was determined to be 2 in Structure Harvester using the Evanno method (Evanno et al. 2005; Earl et al. 2012). Simple majority (>50%) probability of assignment to one of the two clusters was used to categorize fish to "pools", each a combination of sequencing lanes whose samples cluster together (Fig. 2). Individuals that were erroneously assigned using this method to the wrong pool of origin were dropped from further analysis of the pools. Samples from each pool were run separately through *de novo* reference assembly, mapping, SNP discovery and filtering, including filters of loci not conforming to Hardy-Weinberg Equilibrium in at least one site (Hollenbeck 2016) and those genotyped at less than 95%. Loci were evaluated for heterozygosity and selective pressure and appropriately filtered as previously described. Preliminary STRUCTURE analyses were carried out to remove possibly misidentified specimens, identified as having >10%admixture of a second lineage not corresponding to lane assignment or geography (Ladner & Palumbi 2012). Potentially misidentified samples were dropped from the pools and SNP filtering was repeated, resulting in two neutral datasets for statistical analyses.

# Statistical Analyses of Population Structure

All statistical analyses for genetic structure were carried out separately for the two datasets. Point and maximum likelihood estimators of relatedness were determined in the R package "related", with 200 bootstraps for 95% confidence intervals, to determine the need for

removal of related individuals in order to avoid biases in genetic structure analyses (Goldberg & Waits 2010). Genetic diversity indices, pairwise differentiation, and analyses of molecular variance (AMOVA) for hierarchical structuring among sites were calculated for neutral SNPs in the program GenoDive (Meirmans & Van Tienderen 2004). AMOVA groupings were constructed for varying *K* values according to two hypotheses: *a priori* geography (based on land masses, oceanography, and previous phylogeographic studies), and *post hoc* STRUCTURE



**Fig. 2** Results of Structure analysis for A) all five sequencing lanes, B) Pool 1, consisting of three lanes and identified as majority blue individuals from Structure analysis of all lanes, and C) Pool 2, consisting of two lanes and identified as majority green individuals from Structure analysis of all lanes. The bar plots display posterior probabilities for cluster assignment, calculated from neutral RAD loci, among sampled individuals, grouped by sample sites Atimonan (Ati), Cebu (Ceb), Mati (Mat), Puerto Princesa (PtP), Nha Trang (NhT), and Phú Quốc (PhQ).

analysis. *A priori* hypotheses tested are as follows: based on the South China Sea as a barrier (Treml *et al.* 2015; Philippines sites clustered together and Vietnam sites clustered together), summer ocean currents (Inoue & Welsh 2009; Hu *et al.* 2000; Puerto Princesa clustered with Vietnam sites and Atimonan and Cebu each separate), winter ocean currents (Inoue & Welsh 1993; Hu *et al.* 2000; Vietnam sites clustered together, Cebu and Puerto Princesa clustered together, and Atimonan separate), the Sulu Sea as a barrier (Raynal *et al.* 2013; Puerto Princesa separate, Vietnam sites clustered, and remaining Philippines sites clustered), east-to-west differentiation (Raynal *et al.* 2013; Atimonan and Mati clustered, Cebu and Puerto Princesa clustered, and Vietnam sites clustered), east-to-west differentiation with Pacific differentiation along the bifurcation of the NEC (Raynal *et al.* 2013; Ravago-Gotanco & Meñez 2010; Cebu and Puerto Princesa clustered, Vietnam sites clustered, and Atimonan and Mati each separate), and historical basin isolation (Voris 2000; Vietnam sites clustered and Atimonan and Mati each separate).

Mantel tests based on Pearson's product-moment correlation (Legendre & Legendre 2012) for neutral loci were implemented in the R package "ecodist" (Goslee 2007) to detect isolation by distance, using the shortest over-water distances between sites found in ArcGIS (Esri, Redlands, CA).

STRUCTURE analyses were conducted with a burn-in of 20,000 steps and 10 iterations of 50,000 Markov chain Monte Carlo (MCMC) repetitions for each *K* value, 1 through the number of actual sampled sites (six for Pool 1 and five for Pool 2). The most likely *K* value was determined in Structure Harvester using the Evanno method (Evanno *et al.* 2005; Earl *et al.* 2012), the iterations were consolidated in CLUMPP (Jakobsson & Rosenberg 2007), and STRUCTURE plots were created using distruct (Rosenberg 2004). Principle components analyses (PCAs) and principle coordinates analyses (PCoAs) were completed in the R package "adegenet" (Jombart & Ahmed 2011) for neutral and loci under positive selection to visualize population structuring in a two-dimensional space.

#### **Estimating Migration**

Directional migration models were compared and historical migration rates among sites for each dataset were estimated separately in the program Migrate-n v.4.2.14 (Beerli 2009) for a maximum of 200 neutral loci genotyped in 100% of individuals, including RAD tags containing multiple polymorphisms. Coalescent theory, used by Migrate-n, expects true data, including linked loci, in order to estimate mutation rates. Since drastically differing sample sizes can bias estimates in Migrate-n, only the sites with 10 or more samples were included in analyses (Beerli pers. comm.). With this restriction, analyses of Pool 1 included the sites Atimonan, Puerto Princesa, Nha Trang, and Phú Quốc and those of Pool 2 included Cebu and Puerto Princesa. The filtered variant call format (VCF) file produced by VCFtools prior to selection of one random SNP per contig was converted to an IMa sequence file with rad haplotyper and then converted manually to a Migrate input file. Directional migration models were constructed *a priori* from knowledge of contemporary and historical geography and oceanography. Figure 3 displays the models tested for Pool 1, which were based on the following: full migration among all sites (Full Model), the Luzon Strait as the main corridor for dispersal between Vietnam and the Philippines (Nan et al. 2015; Model 1), north-to-south and east-to-west migration (Model 2), south-to-north and west-to-east migration (Model 3), full migration among open ocean sites and directional migration through the Philippine seas (Kool et al. 2011; Model 4), historical basin isolation (Voris et al. 2000; Model 5), ocean currents (Inoue & Welsh 1993) during the peak spawning period of L. lentjan (assuming April to June from Grandcourt et al. 2011; Model 6), larval

dispersal modeled by Treml *et al.* (2015; Model 7), and a panmixia model in which all sampled sites are part of one randomly mixed population. All models except for Model 6 assume a protracted spawning season, as observed in Great Barrier Reef *L. lentjan* populations (Currey *et al.* 2013), which may allow for bidirectional larval dispersal via seasonally reversing currents (Fig. 1). All possible models were tested for Pool 2, including a full model, Puerto Princesa to Cebu (eastward), Cebu to Puerto Princesa (westward), and panmixia (models not pictured). Priors were set at a maximum of 0.1 for mutation-scaled effective population size ( $\theta$ ) and 50,000 for mutation-scaled effective immigration rate (*M*), sampling every 1,000 steps for an MCMC chain of 20 million steps and a burn-in of 20 million. A static heating scheme with four chains was included with a swapping interval of 10. Each run was repeated 10 times and results were averaged across the 10 repetitions. Competing models were compared using Bezier approximation scores estimated by the program. The number of immigrants per generation for each site was calculated by multiplying the modes of  $\theta$  and *M* estimates, and dividing by 4 (Beerli 2009).



**Fig. 3** Models of migration tested in Migrate-n for Pool 1. Arrows point in the direction of migration among sampled sites Atimonan (Ati), Puerto Princesa (PtP), Nha Trang (NhT), and Phú Quốc (PhQ). Panmictic groups are circled in grey. Geographic locations of sampling sites are indicated on the map in the first panel.

#### RESULTS

#### Parsing the Lane Effect

SNP discovery and filtering for samples from all five sequencing lanes resulted in a putative SNP panel of 221 loci across 214 successfully sequenced samples. Samples represented all six populations in the Philippines and Vietnam (40 individuals from Atimonan, 44 from Cebu, 30 from Mati, 45 from Puerto Princesa, Philippines; 34 from Nha Trang, and 21 from Phú Quốc, Vietnam; Table 1). Two loci were removed for low heterozygosity (>0.05) but no loci affected by selection were identified by either Lositan or Bayescan. STRUCTURE analysis was carried out on panel of 219 SNPs for pool assignment (Fig. 2). Samples were correctly assigned to their actual pool using simple majority in >98% of cases. Three individuals were misassigned and dropped from further analyses. Preliminary STRUCTURE analyses on each pool identified potentially misidentified specimens from Cebu (n=1), Mati (n=14), and Nha Trang (n=9), all of which were dropped from the remaining analyses. Pool 1 consisted of 111 individuals from all six sites, originated from three sequencing lanes, formed a *de novo* reference of 64,843 contigs, and resulted in a panel of 415 filtered SNPs (Table 2). Pool 2 consisted of 80 individuals from five sites, originated from two sequencing lanes, formed a *de novo* reference of 88,148 contigs, and resulted in a panel of 9,168 filtered SNPs (Table 2). One locus from Pool 1's SNP panel was identified in BayeScan as being under positive selection and was removed from further analyses, and 41 loci were removed from Pool 2's SNP panel, identified collectively by Lositan and BayeScan (FDR corrected; 25 under positive selection and 16 under balancing selection; Table 2). Statistical analyses were continued on 414 and 9,127 putatively neutral loci for Pool 1 and 2, respectively. Parsing out pools because of lane effects resulted in reduced numbers of specimens

per site, including sites with less than 10 specimens that are considered not reliable in most population assessment analyses (Table 1).

**Table 2** Summary of analyzed SNP panels for Pools 1 and 2, including all putative SNPs passing quality filtering steps, putatively neutral SNPs filtered for selective pressure, and SNPs assumed to be under balancing and/or positive selection. Also displayed are results of an ungrouped AMOVA and Mantel correlation test for Pool 1

Site	Putative SNPs	Filtered SNPs	SNPs under selection	$\begin{array}{c} \textbf{AMOVA} \\ F_{\text{ST}} \end{array}$	AMOVA <i>p</i> -value	Mantel coefficient	Mantel <i>p</i> -value
Pool 1	415	414	positive: 1	0.0086	0.0001	0.5733	0.1667
Pool 2	9,168	9,127	positive: 25 balancing: 16				

#### Population Structure

Indices of genetic diversity for each pool, totaled over all individuals in each analysis, and by sampled site, are presented in Table 3. Pool 1 shows heterozygous excess and a corresponding negative inbreeding coefficient, possibly resulting from its relatively small SNP panel. Several putative half siblings were estimated by relatedness analysis, but these pairs were regarded as spurious as at least 500 SNPs are needed to confidently call half siblings (Mo *et al.* 2016). One putative half sibling pair was found within Cebu in Pool 2 (Table 4). Relatedness estimators detected no true full siblings in either pool, so no individuals were removed from genetic structure analyses (Goldberg & Waits 2010).

		-	Pool 1					Pool 2		
Site	Number of alleles	Effective # of alleles	$H_0$	$H_{\rm S}$	$G_{\rm IS}$	Number of alleles	Effective # of alleles	$H_{\rm O}$	$H_{\rm S}$	$G_{\rm IS}$
Atimonan	2.0000	1.3084	0.2625	0.2222	-0.1813	1.7927	1.4009	0.2761	0.2710	-0.0187
Cebu	1.7126	1.2962	0.2518	0.2102	-0.1976	1.9992	1.4208	0.2704	0.2730	0.0096
Mati	1.8816	1.3215	0.2729	0.2299	-0.1870	1.8802	1.4014	0.2627	0.2670	0.0159
Puerto Princesa	1.9638	1.2897	0.2437	0.2086	-0.1685	1.9856	1.4108	0.2655	0.2679	0.0088
Nha Trang	1.9493	1.2892	0.2455	0.2065	-0.1888	1.7780	1.3975	0.2645	0.2691	0.0172
Phú Quốc	1.9614	1.3067	0.2596	0.2187	-0.1867					
Overall	2.0000	1.2865	0.2560	0.2160	-0.1849	2.0000	1.3879	0.2678	0.2697	0.0067

**Table 3** Summary of genetic diversity indices by sampled site and overall for all neutral loci included in analyses of Pools 1 and 2, including number of alleles, effective number of alleles, observed heterozygosity ( $H_0$ ), expected heterozygosity ( $H_s$ ), and inbreeding coefficient ( $G_{IS}$ )

**Table 4** Results of relatedness analysis of Pool 2, including coefficients of relatedness (r) and 95% confidence intervals for the dyadic maximum likelihood estimator (Milligan 2003) and lynchrd moment estimator (Lynch & Ritland 1999) for putatively related pairs. Also presented is the site assignment of individuals of the pair and the inferred relationship

Sites	lynchrd (95% CIs)	dyadml (95% CIs)	Inferred relationship
Cebu-Cebu	0.225 (0.203-0.248)	0.234 (0.208-0.255)	half sibling

Pairwise comparisons of sites using neutral SNPs are displayed in Table 5 for sites with a sample size of 19 individuals or greater to avoid spurious conclusions resulting from sample selection bias from very low sample sizes. Comparisons for Pool 1 revealed significant differences among all sites (Table 5), with the smallest  $F_{ST}$  between Atimonan and Puerto
Princesa ( $F_{ST}$ =0.0031, p=0.0089) and the largest between Atimonan and Phú Quốc ( $F_{ST}$ =0.0140, p<0.0001). The pairwise genetic distance between Cebu and Puerto Princesa for Pool 2 was also found to be significant ( $F_{ST}$ =0.0015, p<0.0001; Table 5). When only Pool 2's 41 SNPs under selection were considered, Cebu and Puerto Princesa were more differentiated than detected in the neutral SNP analysis ( $F_{ST}$ =0.1766, p<0.0001; Table 5). AMOVAs and a Mantel test for isolation by distance were only carried out on neutral loci for Pool 1, since Pool 2 only had two populations with sample sizes over 19. The ungrouped AMOVA showed significant structuring across all examined sites ( $\alpha$ =0.05) for Pool 1 ( $F_{ST}$ = 0.0086, p=0.0001; Table 2). All *a priori* AMOVA groupings based on geography were not significant and are presented in Appendix A. The Mantel test was not significant (r=0.5733, p=0.1667; Table 2), suggesting that isolation by distance is not a major influence on population divergence for the examined sites.

**Table 5** Pairwise comparisons of sampled sites (n > 19) analyzed in each pool and for Pool 2's SNPs affected by selection, with  $F_{ST}$  values above the diagonals and p-values below. P-values significant at the 0.05 level are bolded

Pool 1	Atimonan	Nha Trang	Phú Quốc	Puerto Princesa	Pool 2	Cebu	Puerto Princesa	Pool 2 SNPs under selection	Cebu	Puerto Princesa
Atimonan		0.0063	0.0140	0.0031	Cebu		0.0015	Cebu		0.1766
Nha Trang	0.0002		0.0095	0.0078	Puerto Princesa	0.0001		Puerto Princesa	0.0001	
Phú Quốc	0.0001	0.0001		0.0139						
Puerto Princesa	0.0089	0.0001	0.0001							

Structure Harvester identified an optimal K value of 2 from STRUCTURE analysis for both pools using the Evanno Method, although care should be taken in interpreting these results as the maximum likelihoods calculated for each pool were highest for a K value of 1. Both analyses displayed distinct differences between Cebu and the other sites (Fig. 2). Cluster probability distributions in Cebu samples resemble those of Phú Quốc in Pool 1, but the sample size of Cebu in this analysis is very small (n=6) and the distance between the sites suggests that they are likely not strongly connected through gene flow. Nha Trang samples in Pool 1 appear to fall in between Cebu and the remaining sites, but Nha Trang is the most different site from Cebu in the analysis of Pool 2. All post-hoc AMOVAs based on STRUCTURE results for Pool 1 were found to be nonsignificant, and are presented in Appendix A.

Principle components analyses (PCAs) of neutral SNPs show similar patterns to STRUCTURE analysis (Fig. 4). The PCA of Pool 1 shows Cebu and Phú Quốc clustered closely together, with all sites displayed in a star pattern around Nha Trang. The first two axes explain about equal amounts of the variance (2.47% and 2.41%). As only one locus was found to be experiencing selective pressure, this analysis was not completed for Pool 1's loci under selection. The PCoA of Pool 1 shows a similar star-shaped pattern of the sites surrounding Nha Trang, but Cebu and Phú Quốc are also separated (Fig. 4). One individual from the half sibling pair in Pool 2 was removed from the PCAs and PCoAs due to its strong influence on the visualization of the samples. The PCA of Pool 2 shows a more pronounced separation of Nha Trang from all other sites (grids on the plot have dimensions of 20), and Cebu is separated from the remaining sites in the opposite direction along the axis that contains the most variance (eigenvalue 1.63%; Fig. 4). The PCoA of neutral SNPs for Pool 2 resembles the PCA, but further separates Atimonan from the remaining sites along axis two, perhaps due to its small sample size (n=6; Fig. 4). A PCA of Pool 2's SNPs under selection showed variation delegated mainly to the first axis (eigenvalue 14.86%), with Cebu and Puerto Princesa separating from all other sites and each other (although the plot has relatively smaller grid dimensions of 2) and all remaining sites clustered closely together (Fig. 4). The PCoA of the SNPs under selection show a very similar pattern to the PCA, with less pronounced differentiation at grid dimensions of 0.1 (Fig. 4).

## Migration

Of all models tested in Migrate-n for Pool 1, based on analysis of 188 loci represented in 100% of individuals, the full model was found to be the most likely (Bezier approximation score of -60875.92; Table 6). The full model assumes migration between all possible pairs of the four sites tested (Atimonan, Puerto Princesa, Nha Trang, and Phú Quốc), but historical migration rates were estimated to be less than 1 migrant per generation along all migration pathways (Table 7). Atimonan appears to be a conduit of migration, exchanging the largest numbers of migrants with the other sites. The full model was also found to be the most likely for Pool 2 (Bezier approximation score of -59422.60; Table 6) based on analysis of 200 loci represented in 100% of individuals. Migrate-n again estimated very limited genetic exchange between the two sites tested, Cebu and Puerto Princesa, at less than 1 migrant per generation in either direction (Table 7).

Migrate-n is designed to estimate historical migration, but in the event of very recent changes to connectivity patterns, estimates of migration can be biased (Samarasin *et al.* 2017). In particular, the program underestimates historical migration and estimates are closer to contemporary rates when connectivity has been reduced recently. We would expect connectivity in the study region to have increased since the last glacial maximum several thousand generations ago (Ludt and Rocha 2015; Gaither *et al.* 2011), and so this bias is likely not an



**Fig. 4** Principle components analyses (PCAs) of A) neutral RAD loci for Pool 1, B) neutral RAD loci for Pool 2, and C) loci under selection for Pool 2; and principle coordinates analyses (PCOAs) of D) neutral RAD loci for Pool 1, E) neutral RAD loci for Pool 2, and F) loci under selection for Pool 2. PCA ellipses contain 95% of the variance for sample sites Atimonan (Ati), Cebu (Ceb), Mati (Mat), Puerto Princesa (PtP), Nha Trang (NhT), and Phú Quốc (PhQ). Eigenvalues are displayed in barplots, with retained values filled in black (for displayed 2D axes) and grey.

issue, though it should be recognized that heavy exploitation may lead to reduced population

connectivity (Allendorf et al. 2008).

	Model	Bezier approximation score	Delta	Rank	
Pool	1				
	Full	-60875.92	0.00	1	
	Model1	-62668.48	-1792.56	3	
	Model2	-63457.84	-2581.92	7	
	Model3	-62814.84	-1938.92	4	
	Model4	-62499.59	-1623.67	2	
	Model5	-63445.87	-2569.95	6	
	Model6	-63741.36	-2865.44	5	
	Model7	-63672.18	-2796.26	8	
	Panmixia	-64006.30	-3130.38	9	
Pool 2					
	Full	-59422.60	0	1	
	Eastward	-59896.63	-474.03	3	
	Westward	-59996.63	-574.03	4	
	Panmixia	-59760.67	-338.07	2	

**Table 6** Bezier approximation scores for Migrate-n models tested for Pools 1 and 2, used to select the most likely model. Also displayed are deltas and ranks based on the scores

**Table 7** Parameters estimated by the full models for Pools 1 and 2 in Migrate-n. Displayed are the sites and migration pathways for which parameters are estimated, mutation-scaled effective population size ( $\theta$ ), mutation-scaled migration (M), and migrants per generation (calculated by multiplying  $\theta$  for the receiving population by M and dividing by 4; Beerli 2009) for each pool. Sample sites Atimonan (Ati), Nha Trang (NhT), Phú Quốc (PhQ), and Puerto Princesa (PtP) were investigated for Pool 1 and sites Cebu (Ceb) and Puerto Princesa (PtP) were investigated for Pool 2

	Site/Pathway	θ	95% CIs	М	95% CIs	Migrants per generation	95% CIs
Pool 1							
	Ati	0.0431	0.0410-0.0453				
	NhT	0.0158	0.0135-0.0179				
	PhQ	0.0174	0.0147-0.0200				
	PtP	0.0183	0.0156-0.0208				
	Ati → NhT			50.0	0-133.3	0.1975	0.0-0.5975
	Ati → PhQ			50.1	0-1100.0	0.2179	0.0-5.5000
	Ati → PtP			50.0	0-133.3	0.2288	0.0-0.6932
	NhT → Ati			16.7	0-866.7	0.1799	0.0-9.8089
	NhT $\rightarrow$ PhQ			16.7	0-866.7	0.0726	0.0-4.3335
	NhT $\rightarrow$ PtP			16.7	0-866.7	0.0764	0.0-4.5068
	PhQ → Ati			16.7	0-866.7	0.1799	0.0-9.8089
	PhQ → NhT			16.7	0-866.7	0.0660	0.0-3.8850
	PhQ → PtP			16.7	0-866.7	0.0764	0.0-4.5068
	PtP → Ati			16.7	0-866.7	0.1799	0.0-9.8089
	PtP → NhT			16.7	0-866.7	0.0660	0.0-3.8850
	$PtP \rightarrow PhQ$			16.7	0-866.7	0.0726	0.0-4.3335
Pool 2							
	Ceb	0.0431	0.0374-0.0419				
	PtP	0.0107	0.0085-0.0129				
	Ceb → PtP			16.7	0-866.7	0.0447	0.0-2.7886
	PtP → Ceb			16.7	0-866.7	0.1799	0.0-9.0722

### DISCUSSION

### RADSeq Challenges

RADSeq is growing rapidly in popularity due to its ease of use. We chose the ezRAD method because it does not require the purchase of expensive lab equipment or excessive optimization or training, as library prep can be carried out using commercially available kits (Puritz *et al.* 2014). The costs of these techniques have also dropped considerably in the past few years and far outpace traditional sequencing methods in terms of cost per sequence, time, and output (Davey & Blaxter 2011). But for all its advantages, RADSeq is not without its faults. Arnold *et al.* (2013) found that RADSeq tends to underestimate diversity as a result of polymorphisms at restriction sites, causing biased sampling. The PCR step of the library prep also may cause errors resulting from over-amplification, uneven amplification among loci and samples, and GC content biases (Mastretta-Yanes *et al.* 2015; Benjamini & Speed 2012; Aird *et al.* 2011).

We also discovered that some knowledge of the genome size of the study organism is required to select the correct restriction enzyme that does not cut too often or too rarely, so as to produce an amount of reads that can be covered a sufficient number of times by a lane of sequencing. The pilot run of libraries produced for this study was returned with very low coverage data, prompting the contraction of the size selection range. Further study of *Lethrinus lentjan* would benefit from obtaining an accurate estimate of genome size in order to better predict the effect of restriction enzymes with recognition sites of different lengths when utilizing RAD sequencing, so that adequate coverage can be achieved. Every laboratory technique is of course also subject to human and equipment error that is unpredictable and unquantifiable. The ease and low cost of RADSeq certainly outweigh its disadvantages for its many applications. *RADSeq and Lane Effects* 

A recognized complication encountered with Illumina sequencers is the occurrence of biases in base calls among sequencing lanes or different sequencing runs. As of yet, it is unclear if any one cause is to blame, but quality degradation at certain positions on sequenced reads has been observed to result in a lane effect among Illumina sequencing runs (Mastretta-Yanes *et al.* 2015).

Lane effects can, however, be detected and eliminated with knowledge of the utilized methods and study design. Marioni *et al.* (2008) describe a method of detecting lane-affected genes in RNA-seq data by examining read counts of genes across multiple lanes after eliminating expected confounding factors such as the tissue of origin of the samples. Similarly, with RADSeq, genotype data can be separated based upon sample site, and SNPs can be evaluated with locus-by-locus AMOVAs to identify those significant for lane assignment. Raw reads may also be examined with fastQC (Andrews 2010) and multiQC (Ewels *et al.* 2016) to identify regions or particular positions that are subject to quality degradation that may cause lane effects (Mastretta-Yanes *et al.* 2015). Phasing errors on the sequencer's flow cell can cause quality degradation, especially toward the end of long DNA fragments, that can cause bases to be called incorrectly even after undergoing quality filtering (Fuller *et al.* 2009).

In the present study, the above strategies were unable to efficiently mitigate the lane effect without unacceptable levels of data loss. Instead, the dataset was split according to lane assignment in order to be analyzed separately, a strategy used to analyze sequence data containing cryptic coral species (Ladner & Palumbi 2012). STRUCTURE and discriminant analyses like PCAs are sensitive to differences in allele frequencies that can originate from multi-species or lane-affected datasets, and are useful for confirming the presence of such confounding factors. One advantage of this method is that viable results could be generated without the additional cost of more library preparation and sequencing. Another benefit was the discovery of relatively large numbers of SNPs, offering more confidence in lineage and comparative analyses, especially for Pool 2 (Table 2). A SNP panel of this size was impossible to obtain with the lane effected data due to failures at several quality filters, including paralogy and deviations from Hardy Weinberg Equilibrium resulting from infrapopulation structure within the sampled sites (Huo *et al.* 2016). A drawback to this method was small sample sizes (<10) from Cebu in Pool 1, Atimonan and Nha Trang in Pool 2, and Mati in both pools, which may cause inaccuracies in population structure inference due to sample selection bias (Shringarpure & Xing 2014). In order to avoid erroneous conclusions, these sample sites were left out of most analyses.

A more prudent approach for future projects would be to consider the possibility that lane effects may arise in sequencing data, and plan the study design accordingly. The placement of samples on lanes should be randomized if more than one lane is required (Churchill 2002), or sampling sites or species (depending on if the study is on the population or phylogenetic level) should be equally stratified across multiple lanes so that a lane effect can be distinguished from actual biological divergence (Leek *et al.* 2010). When samples are stratified across lanes, a lane effect can be observed in two or more clades within sampling sites or species that correlate with sequencing lanes rather than geographically (see Fig. 2A). Libraries can also be pooled at lower concentrations so as to fit all samples on a single lane, and this pool can be sequenced multiple times to attain the desired depth of coverage, thereby averaging the effects of the different lanes across all samples. However, care would also need to be taken to ensure that sequencing run bias is not introduced with sequential sequencing strategies. Detailed records should always be kept for all laboratory methods, but especially of those that may produce batch effects such as changes in lab personnel or reagents, or lane effects such as pooling methods, sequencing facility changes, and lane assignment. This will allow for any curious patterns uncovered by STRUCTURE, PCAs, or any other genetic structure analyses to be compared with potentially confounding factors.

## Population Structure in Lethrinus lentjan

An overall objective of this study was to test if advanced genomic methods would detect population structure where traditional Sanger sequencing of mitochondrial DNA did not (Hines 2013). High-throughput sequencing methods, such as RADSeq, are more costly than traditional methods and therefore the coverage of sites in this study was restricted to test structure at broad spatial scales. Despite this limited site selection, the significant population structure results from this study offer a number of insights relating to basin isolation and the influence of hydrographic features on the population structure of *Lethrinus lentjan*.

The population structure observed in *L. lentjan* may be at least in part explained by a pattern of Pleistocene basin isolation, although potentially modified by present day surface currents. The Philippines experienced multiple basin isolation events during the Pleistocene because of its complex coastlines and shelf areas, and concordant genetic signatures are apparent despite contemporary ocean currents among these basins. Despite the persistence of a physical connection between the South China Sea and the Pacific Ocean during Pleistocene glacial maxima, there is little evidence for continued connectivity across these basins. Two major physical features, the outflow of the Mekong River and the Pleistocene emergent Sunda Shelf, may also have influenced population structure along the coast of Vietnam.

### *Population Structure across the Philippines*

Genetic structure among *L. lentjan* populations sampled within the Philippines is generally consistent with the Pleistocene isolation of ocean basins, which repeatedly restricted gene flow across the archipelago. Sampling effort represents the Pacific coast of the Philippines (Atimonan and Mati), the Visayan Sea in the center (Cebu), and the Sulu Sea in the west (Puerto Princesa), each positioned in bodies of water once separated by Pleistocene land bridges (Voris 2000; Fig. 1). Pairwise significant differences, a significant ungrouped AMOVA, and nonsignificant grouped AMOVAs (Tables 2, 5; Appendix A) provide evidence for restricted gene flow among these Philippine sites that is best explained by once separated ocean basins. Mantel tests failed to detect a significant correlation between  $F_{ST}$  and geographic distances between sites (Table 2), suggesting that populations are differentiated hierarchically rather than on a cline. Extending sampling efforts to include more than one site within each basin, representing locations such as the eastern Sulu Sea and the northern and southern Visayas, and increasing sample sizes at these sites could help verify the influence of basin isolation, or of more complex barriers like surface currents, on population structuring.

East-to-west differentiation within the Philippines, which has in other studies of fish species been ascribed to basin isolation during periods of low sea level (Lourie *et al.* 2005; Ravago-Gotanco & Meñez 2010; Raynal *et al.* 2013), may be maintained by contemporary ocean surface currents. Much of the differentiation, as shown in STRUCTURE and PCA results (Figs. 2, 4), appears to originate in the central Philippines, represented by the Cebu sampling site. The tendency for Cebu to differ so strongly could be explained by the repeated isolation of the inner seas throughout the Pleistocene Epoch, and this is supported by Migrate-n analysis of Pool 2 that estimates very limited migration between Cebu and Puerto Princesa in the Sulu Sea (Table 7).

However, even following the inundation of land bridges separating the Visayan Sea from other basins, the persistence of strong ocean currents continue to enforce these barriers. For example, the persistent Sulu Sea Throughflow that flows between Cebu and Puerto Princesa is a strong dispersant of larvae from north to south (Kool *et al.* 2011), but may also act as an east-to-west barrier (Raynal *et al.* 2013). The Bohol Sea Jet is inferred to block the dispersal of propagules, promoting genetic structuring of parrotfish populations in the southern Philippines (Stockwell *et al.* 2015). These potential barriers may be an alternative explanation to Pleistocene basin isolation for the dissimilarities between the Sulu and inner Philippine seas, and the presence of a sibling pair in Cebu provides evidence for the entrainment of larvae within the inner seas.

High connectivity of the inner seas of the Philippines is predicted by contemporary larval dispersal models (Treml *et al.* 2015) and consistent with the lack of genetic structure observed in the Philippines previously for *L. lentjan* (Hines 2013). In the Sulu Sea, multiple open channels provide potential for gene flow from surrounding seas to result in the similarities seen in STRUCTURE analysis between the Atimonan and Puerto Princesa populations in Pool 1 (Fig. 2). From the west, surface currents feed the Sulu Sea Throughflow, which carries water south along eastern Palawan Island, transporting larvae from the South China Sea to reefs in the western Sulu Sea and further south to join the Indonesian Throughflow (Kool et al. 2011). The intrusion of surface waters from the Kuroshio Current into the South China Sea (Nan *et al.* 2015) may allow for connectivity between Atimonan and Puerto Princesa on opposite sides of the archipelago, resulting in similarities observed in some analyses (Fig. 2; Table 5). Minimal historical migration rates among sites within the archipelago as predicted by Migrate-n (Table 7) contrastingly reflect the structure of Pleistocene isolated ocean basins that repeatedly restricted genetic exchange throughout the Philippines. Again, a non-significant Mantel test for isolation-

by-distance (Table 2) suggests that any existing migration among the basins sampled has not yet homogenized less distanced populations, and more complete sampling of the entirety of these ocean basins, including intervening sites between east and west Luzon would help to verify that this is the case.

Along the Pacific coast of the Philippines, analyses of the two pools are unable to come to a clear consensus on genetic structure between Atimonan and Mati. The forces of the bifurcation of the NEC, forming the north-flowing Kuroshio Current and the south-flowing Mindanao Current, have been cited as preventing populations on the east coast of the Philippines from seeding populations to the south in rabbitfishes (Ravago-Gotanco & Meñez 2010; Fig. 1). However, models by Treml *et al.* (2015), as previously mentioned, found no mechanisms for restriction of larval dispersal in the Philippines, including at this potential barrier. The small sizes of Mati samples for both pool analyses prevent us from shedding further light on this particular region (Table 1).

### Population Structure across the South China Sea

Basin isolation is consistent with patterns of differentiation between the South China Sea and the inner Philippine seas as observed in analyses of Pool 1. Nha Trang, located in the western South China Sea, and Phú Quốc, located in the Gulf of Thailand, are both significantly differentiated from Philippines sites by pairwise comparisons (Table 5). STRUCTURE and PCA results (Figs. 2, 4) reveal that Nha Trang is more similar to Philippine sites than Phú Quốc, sampled from a region that was exposed multiple times throughout the Pleistocene and has since been recolonized. The lack of differentiation between Cebu and Phú Quốc could be an artefact of the small sample size of Cebu in Pool 1.

Estimates of migration, along with STRUCTURE results, appear to support the

hypothesis that the South China Sea is a barrier to gene flow (Treml *et al.* 2015). Ocean currents often promote gene flow and mixing, and the intrusion of the Kuroshio Current into the South China Sea (Nan *et al.* 2015) is a possible mechanism for gene flow between Atimonan and the Vietnam sites. The Luzon Strait remained open during the lowest sea level stands of the Pleistocene, and today a strong, seasonally reversing current flows through the South China Sea to connect the Kuroshio Current to the Indian Ocean (Hu *et al.* 2000; Fig. 1). However, the restriction of gene flow across the South China Sea, as supported by larval dispersal models (Treml *et al.* 2015), is supported by our results showing clear differentiation between thoroughly sampled sites in the Philippines and Vietnam (Fig. 2; Table 5). Historic gene flow between the inner Philippine seas and the South China Sea is unsupported by estimates of migration rates in Migrate-n (Table 7). Corroboration between ancient and contemporary larval dispersal along this corridor provides evidence for the formative influence of restricted passages during glacial maxima between the Sulu and South China seas on genetic structure of the species.

# Population Structure Along the Coast of Vietnam

Differentiation between Phú Quốc and Nha Trang may be an artefact of population expansion onto the Sunda Shelf when it was flooded following the last glacial maximum (Crandall *et al.* 2012), or persistence of major physical barriers to dispersal, or both. Several studies have inferred a barrier between the waters over the Sunda Shelf and the South China Sea (Rocha *et al.* 2007; Carpenter *et al.* 2011) and this pattern suggests that Pleistocene sea levels may have been influential in diversification in this region. Significant differences between Phú Quốc and Nha Trang detected by all genetic structure analyses of Pool 1 (Figs. 2, 4; Table 5) and restricted migration between the two sites as estimated in Migrate-n (Table 7) support such a barrier consistent with a genetic break identified in the mitochondrial DNA of *L. lentjan* (Hines 2013). This observed population structure could be a result of factors influencing genetic makeup during recolonization, such as a founder effect (Crandall *et al.* 2012), or barriers to larval dispersal across the Mekong River outflow (Dai & Trenberth 2002; Rocha *et al.* 2002, Hines 2013). A barrier to larval dispersal in the South China Sea modeled by Treml *et al.* (2015) may be explained by the location of their sample site in the Gulf of Thailand to the west of the Mekong River outflow and over the submerged Sunda Shelf.

### Contrasts with Previous Study of Lethrinus lentjan

The first examination of population structure of *L. lentjan* in this region also detected a genetic break ascribed to the outflow of the Mekong River, which separates the Gulf of Thailand from Nha Trang and the Philippines (Hines 2013). Our genetic structure analyses also consistently find that Nha Trang and Phú Quốc are significantly different (Figs. 2, 4; Table 5), and our study also finds limited historical connections (Table 7).

Hines's study, in contrast with ours, found no significant differences among sites within the Philippines or indeed throughout much of the Coral Triangle (Hines 2013). The only significant pairwise differences found among Philippines sites are between Tacdugon, located in the northern Visayas and geographically the most distant from the other sites, and two southern sites off the island of Negros. The small sample size of this site (n=8) precludes us from making any inferences about its true relationship to the other sites. The difference in results between these two studies is likely explained by methodology.

Prior study of *L. lentjan*'s population structure in the Coral Triangle using 88 polymorphic sites on a single mitochondrial gene (Hines 2013) differs markedly from results from our RADseq panel of hundreds to thousands of neutral SNPs, and these differences may be due to the regions of the genome sampled and greater abundance of markers. The examination of

hundreds of independent loci offers potentially greater insight into the evolution of the whole genome and could be a source of differences between the results of these two studies (Kumar *et al.* 2012). The addition of even more loci, sites, and samples are all further improvements that could be made to our dataset to elucidate the interactions of *L. lentjan* populations in the South China Sea and adjacent oceans.

## Recommendations for Management

High degrees of structuring observed across the Philippines and South China Sea indicate that despite similarities among sampled sites throughout the region, the *L. lentjan* stock is somewhat subdivided. To maintain a truly subdivided stock, populations would have to be discrete breeding units with little or no exchange of individuals (Lacy 1987). Historical divisions between sampled sites, possibly reinforced through periods of low sea level and restricted gene flow, may have largely driven the population structure observed. Life history characteristics have also been found to be strong influencers on the genetic structure of populations in the region (Magsino & Juinio-Meñez 2008). The combination of L. *lentjan*'s relatively protracted PLD and its tendency to maintain a small home range as an adult may have influenced the patterns of diversification and connectivity detected by this study, in regions of restrictive and dispersant ocean currents, respectively.

Patchy government-supported protection and continuing exploitation may reduce connectivity of foodfish populations in Southeast Asia. The Philippines has one of the world's best networks of marine protected areas (MPAs), consisting of over 1,800 local and national reserves (Cabral *et al.* 2014). The majority of marine reserves in the Philippines, however, are very small (<1 km<sup>2</sup>) and not closely spaced enough to support ecologically meaningful connectivity (Horigue *et al.* 2015; Weeks *et al.* 2009). Weeks *et al.* (2009) recommend the establishment of large no-take areas and Horigue *et al.* (2015) provide strategies for scaling up local MPAs in order to fill in gaps along ecologically important pathways. *L. lentjan* populations would benefit from the expansion of the Philippines MPA network, as currently only 19% of mangrove habitat (Long & Giri 2011) and ~3% of coral reefs (Weeks *et al.* 2009) are protected, both habitats relied upon by the species. *L. lentjan* landings have been in decline in recent years and significant  $F_{ST}$  values indicate a fragmented stock, both of which are consequences of overexploitation.

Results from this study, a previous study of *L. lentjan* (Hines 2013), and larval dispersal models (Treml et al. 2015), all support a distinction between populations in the Gulf of Thailand and off the eastern coast of Vietnam. This study also identified strong differentiation between the inner Philippine seas and surrounding ocean basins, perhaps influenced by periods of low sea level during the Pleistocene Epoch. Based on measures of population differentiation from this study, we can conclude that the sampled populations across the Philippines and the South China Sea are undergoing only limited mixing, and thus recommend the *L. lentjan* stock to be considered subdivided by fisheries managers (Bird *et al.* 2007). However, further sampling should be carried out to contribute to this dataset and to obtain more resolution within each of the sampled ocean basins.

### CONCLUSIONS

This study hypothesized that *Lethrinus lentjan* would show differentiation aligning with Pleistocene glacial maxima ocean basins, separated repeatedly by continuous land masses that eventually became the Philippine archipelago. Abundant significant differences among sites suggest that despite modern day surface currents promoting mixing throughout the region, Pleistocene vicariance is a possible driving factor in lineage diversification for this species. The results contrast with a previous study of L. lentjan that found no significant differences among distant sites across the Coral Triangle. Historically restricted migration could explain continued population differentiation in L. lentjan in the Philippines and the South China Sea across the boundaries of previously isolated basins. Results support the hypothesis that the Coral Triangle, and the Philippines in particular, act as a center of origin, resulting in elevated diversity. As evidenced by significant differences among sampled populations, fishing efforts throughout the region could be exploiting different stocks that may react independently to current harvest levels and have different management needs. A comprehensive stock assessment of L. lentjan populations in the Philippines is required before management can be prescribed. Results from this study may be used to predict patterns of structure in other similar taxa, including small, potentially overfished snappers. Further study would be warranted for species that may be at risk of overexploitation so that their management can be tailored to their stock structure.

A lane effect was successfully mitigated using a method previously used to parse cryptic species from sequencing datasets. Some genetic structure analyses are sensitive to confounding factors resulting from inconsistencies across sequencing lanes and can be used to identify lane effects and divide datasets along them. Several strategies in study design are effective for avoiding the development of such biases, including equal stratification of sampling sites or species across sequencing lanes. Lane effects have not yet been explored in the literature for the most recent sequencing technologies, but it is likely a prevalent issue in high-throughput sequencing datasets. It is recommended that researchers consider the possibility that lane biases may arise in even the most advanced and most often used sequencing methods, such as Illumina sequencing, and use the recommended best practices in library preparation and study design.

#### REFERENCES

- Ackiss AS, Pardede S, Crandall ED, Ablan-Lagman MCA, Ambariyanto, Romena N, Barber PH,
   Carpenter KE (2013) Pronounced genetic structure in a highly mobile coral reef fish,
   *Caesio cuning*, in the Coral Triangle. *Marine Ecology Progress Series*, 480, 185-197.
- Aird D, Ross MG, Chen W-S et al. (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biology* **12**, R18.
- Al-Yousuf MH, El-Shahawi MS, Al-Ghais SM (2000) Trace metals in liver, skin and muscle of *Lethrinus lentjan* fish species in relation to body length and sex. *Science of the Total Environment*, 256, 87-94.
- Allen G (2008) Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **18**, 541-556.
- Allendorf FW, England PR, Luikart G, Ritchie PA, Ryman N (2008) Genetic effects of harvest on wild animal populations. *Trends in Ecology & Evolution*, **23**, 327-337.
- Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
- Arnaud S, Bonhomme F, Borsa P (1999) Mitochondrial DNA analysis of the genetic relationships among populations of scad mackerel (*Decapterus macarellus, D. macrosoma, and D. russelli*) in Southeast Asia. *Marine Biology*, **135**, 699-707.

- Arnold B, Corbett-Detig RB, Hartl D, Boblies K (2013) RADSeq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, 22, 3179-3190.
- Barber PH (2009) The challenge of understanding the Coral Triangle biodiversity hotspot. *Journal of Biogeography*, **36**, 1845-1846.
- Barber PH, Palumbi SR, Erdmann MV, Moosa MK (2000) Biogeography: 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 largal transport: patterns, causes, and consequences. *Molecular Ecology*, **11**, 659-674.
- Barber PH, Erdmann MV, Palumbi SR (2006) Comparative phylogeography of three codistributed stomatopods: origins and timing of regional lineage diversification in the Coral Triangle. *Evolution*, **60**, 1825-1839.
- Barber PH, Cheng SH, Erdmann MV, Tenggardjaja K, Ambariyanto (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. *Crustacean*, **19**, 129-156.
- Bay LK, Crozier RH, Caley MJ (2006) The relationship between population genetic structure and pelagic larval duration in the coral reef fishes on the Great Barrier Reef. *Marine Biology*, 149, 1247-1256.
- Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B*, **263**, 1619-1626.

- Beerli P (2009) How to use migrate or why are markov chain monte carlo programs difficult to use? In: *Conservation Biology, vol. 17: Population Genetics for Animal Conservation* (eds Bertorelle G, Bruford MW, Hauffe HC, Rizzoli A, Vernesi C), pp. 42–79.
  Cambridge University Press, Cambridge, UK.
- Begg GA, Friedland KD, Pearce JB (1999) Stock identification and its role in stock assessment and fisheries management: an overview. *Fisheries Research*, **43**, 1-8.
- Benjamini Y, Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. *Nucleic Acids Research*, **40**, e72.
- Benzie JAH (1998) Genetic structure of marine organisms and SE Asian biogeography. In:
   *Biogeography and Geological Evolution of SE Asia* (eds Hall R, Holloway JD), pp. 197-209. Backbuys Publishers, Leiden, the Netherlands.
- Benzie JAH (1999a) Genetic structure of coral reef organisms: ghosts of dispersal past. *American Zoologist*, **39**, 131-145.
- Benzie JAH (1999b) Major genetic differences between crown-of-thorns starfish (*Acanthaster planci*) populations in the Indian and Pacific oceans. *Evolution*, **53**, 1782-1795.
- Bird CE, Holland BS, Bowen BW, Toonen RJ (2007) Contrasting phylogeography in three endemic Hawaiian limpets (*Cellana* spp.) with similar life histories. *Molecular Ecology*, 16, 3173-3186.
- Block BA, Teo SLH, Walli A *et al.* (2005) Electronic tagging and population structure of Atlantic bluefin tuna. *Nature*, **434**, 1121-1127.

- Brothers EB, Williams DMB, Sale PF (1983) Length of larval life in twelve families of fishes at "One Tree Lagoon", Great Barrier Reef, Australia. *Marine Biology*, **76**, 319-324.
- Brothers EB, Thresher RE (1985) Pelagic duration, dispersal, and the distribution of Indo-Pacific coral-reef fishes. *Symposia Series for Undersea Research, NOAA's Undersea Research Program*, **3**, 53-69.
- Bruno JF, Selig ER (2007) Regional decline of coral cover in the Indo-Pacific: timing, extent, and subregional comparisons. *PLoS One*, **2**, e711.
- Cabral RB, Aliño PM, Balingit AC *et al.* (2014) The Philippine marine protected area (MPA) database. *Philippine Science Letters*, **7**, 300-8.
- Carpenter KE, Allen G (1989) FAO species catalogue No. 125 Vol.9 Emperor fishes and largeeye breams of the world (family Lethrinidae). An annotated and illustrated catalogue of lethrinid species known to date. Food and Agricultural Organization of the United Nations, Rome, Italy. 118 pp.
- Carpenter KE, Springer VG (2005) The center of the center of marine shore fish biodiversity: the Philippine Islands. *Environmental Biology of Fishes*, **72**, 467-480.
- Carpenter KE, Abrar M, Aeby G *et al.* (2008) One-third of reef-building corals face elevated extinction risk from climate change and local impacts. *Science*, **321**, 560-563.
- Carpenter KE, Barber PH, Crandall ED *et al.* (2011) Comparative phylogeography of the Coral Triangle and implications for marine management. *Journal of Marine Biology*, **2011**, 396982.

- Carpenter KE, Lawrence A, Myers R (2016) *Lethrinus lentjan. The IUCN Red List of Threatened Species 2016*, e.T16720036A16722340.
- Carvalho GR, Hauser L (1994) Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries*, **4**, 326-350.
- Catchen J, Hohenlohe P, Bassham S, Amores A, Cresko W (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124-3140.
- Cesar HSJ (2000) Coral Reefs: Their Functions, Threats and Economic Value. In: *Collected Essays on the Economics of Coral Reefs* (ed Cesar HSJ), pp. 14-39. CORDIO, Kalmar University, Kalmar, Sweden.
- Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. *Nature Genetics*, **32**, 490.
- Colborn J, Crabtree RE, Shaklee JB, Pfieler E, Bowen BW (2001) The evolutionary enigma of bonefishes (*Albula* spp.): cryptic species and ancient separations in a globally distributed shorefish. *Evolution*, **55**, 807-820.
- Crandall ED, Frey MA, Grosberg RK, Barber PH (2008a) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology*, 17, 611-626.
- Crandall ED, Jones ME, Muñoz MM, Akinronbi B, Erdmann MV, Barber PH (2008b) Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. *Molecular Ecology*, **17**, 5276-5290.

- Crandall ED, Sbrocco EJ, DeBoer TS, Barber PH, Carpenter KE (2012) Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the Last Glacial Maximum. *Molecular Biology and Evolution*, **29**, 707-719.
- Crandall ED, Treml EA, Liggins L *et al.* (2014) Return of the ghosts of dispersal past: historical spread and contemporary gene flow in the blue sea star *Linckia laevigata*. *Bulletin of Marine Science*, **90**, 399-425.
- Crowder LB, Lyman SJ, Figueira WF, Priddy J (2000) Source-sink population dynamics and the problem of siting marine reserves. *Bulletin of Marine Science*, **66**, 799-820.
- Cunningham KM, Canino MF, Spies IB, Hauser L (2009) Genetic isolation by distance and localized fjord population structure in Pacific cod (*Gadus macrocephalus*): limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of Aquatic Sciences*, **66**, 153-166.
- Currey LM, Williams AJ, Mapstone BD et al. (2013) Comparative biology of tropical Lethrinus species (Lethrinidae): challenges for multi-species management. *Journal of Fish Biology*, 82, 764-788.
- Dai A, Trenberth KE (2002) Estimates of freshwater discharge from continents: Latitudinal and seasonal variations. *Journal of Hydrometeorology*, **3**, 660-687.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.
- Dasgupta S, Laplante B, Meisner C, Wheeler D, Yan J (2009) The impact of sea level rise on developing countries: a comparative analysis. *Climatic Change*, **93**, 379-88.

- Davey JW, Blaxter ML (2011) RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, **9**, 416-423.
- 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.
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4, 359-361.
- Endler JA (1977) *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, New Jersey, USA.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611-20.
- Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**, 3047-8.
- FAO Fisheries & Aquaculture Department (2016) Global Production Statistics (online query). *Fishery Statistical Collections*, Accessed 1 June 2016. URL: http://www.fao.org/fishery/statistics/global-production/en.
- 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.
- Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, **10**, 2741-2752.

- Fuller CW, Middendorf LR, Benner SA et al. (2009) The challenges of sequencing by synthesis. *Nature Biotechnology*. 27, 1013-23.
- 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*). *Journal of Biogeography*, **37**, 133-147.
- Gaither MR, Bowen BW, Bordenave TR et al. (2011) Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle. *BMC Evolutionary Biology*, **11**, 189.
- Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. *arXiv*, preprint arXiv:1207.3907.
- Gilad Y, Mizrahi-Man O (2015) A reanalysis of mouse ENCODE comparative gene expression data. *F1000Research*, **4**.
- Goldberg CS, Waits LP (2010) Quantification and reduction of bias from sampling larvae to infer population and landscape genetic structure. *Molecular Ecology Resources*, 10, 304–13.
- Goslee SC, Urban DJ (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software*, **22**, 1-19.
- Grandcourt EM (2002) Demographic characteristics of a selection of exploited reef fish from the Seychelles: preliminary study. *Marine and Freshwater Research*, **53**, 123-30.

- Grandcourt E, Al Abdessalaam TZ, Francis F, Al Shamsi A (2011) Demographic parameters and status assessments of *Lutjanus ehrenbergii*, *Lethrinus lentjan*, *Plectorhinchus sordidus* and *Rhabdosargus sarba* in the southern Arabian Gulf. *Journal of Applied Ichthyology*, 27, 1139-290.
- Grüss A, Kaplan DM, Guénette S, Roberts CM, Botsford LW (2011) Consequences of adult and juvenile movement for marine protected areas. *Biological Conservation*, **144**, 692-702.
- Han W, Moore AM, Levin J, Zhang B *et al.* (2009) Seasonal surface ocean circulation and dynamics in the Philippine Archipelago region during 2004-2008. *Dynamics of Atmospheres and Oceans*, 47, 114-137.
- Harrison HB, Williamson DH, Evans RD et al. (2012) Larval export from marine reserves and the recruitment benefit for fish and fisheries. *Current Biology*, **22**, 1023-1028.
- Hines AB (2013) Comparative phylogeography of the emperor snappers *Lethrinus lentjan* and *Lethrinus harak* (Lethrinidae: Percoidei) in the Coral Triangle (Unpublished masters thesis). Old Dominion University, Norfolk, Virginia, USA.
- Hoeksema BW (2007) Delineation of the Indo-Malayan centre of maximum marine biodiversity:
  the Coral Triangle. In: *Biogeography, time, and place: distributions, barriers, and islands*(ed Renema W), pp. 117-178. Springer, Leiden, the Netherlands.
- Hollenbeck C (2016) filter\_hwe\_by\_pop.pl. *GitHub Repository*, Accessed 6 December 2016. URL: https://github.com/jpuritz/dDocent/blob/master/scripts/filter\_hwe\_by\_pop.pl.

Hollenbeck C (2017) rad\_haplotyper: a program for building SNP haplotypes from RAD sequencing data. *GitHub Repository*, Accessed 9 September 2016. URL: https://github.com/chollenbeck/rad\_haplotyper.

- Horigue V, Pressey RL, Mills M, Brotánková J, Cabral R, Andréfouët S (2015) Benefits and challenges of scaling up expansion of marine protected area networks in the Verde Island Passage, Central Philippines. *PloS One*, **10**, e0135789.
- Horne JB, van Herwerden L, Choat JH, Robertson DR (2008) High population connectivity across the Indo-Pacific: congruent lack of phylogeographic structure in three reef fish cogeners. *Molecular Phylogenetics and Evolution*, **49**, 629-638.
- Hu J, Kawamura H, Hong H, Qi Y (2000) A review on the currents in the South China Sea: seasonal circulation, South China Sea warm current and Kuroshio intrusion. *Journal of Oceanography*, **56**, 607-624.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322-1232.
- Huo GN, Liu L, He HB, Attwood SW (2016) An investigation into the potential effects of infrapopulation structure and other sources of sampling error, on population genetic studies of the transmission of Schistosoma japonicum (Trematoda: Digenea). *Parasites & Vectors*, 9, 165.
- Ihssen PE, Booke HE, Casselman JM, McGlade JM, Payne NR, Utter FM (1981) Stock identification: materials and methods. *Canadian Journal of Fisheries and Aquatic Sciences*, 38, 1838-1855.

- Inoue M, Welsh SE (1993) Modeling seasonal variability in the wind-driven upper-layer circulation in the Indo-Pacific region. *Journal of Physical Oceanography*, **23**, 1411-1436.
- 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, 1801-6.
- Jokiel P, Martinelli FJ (1992) The vortex model of coral reef biogeography. *Journal of Biogeography*, **19**, 449–458.
- Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070-3071.
- Kedidi SM, Abushusha T, Allam K (1984) Biology and stock assessment of the redspot emperor,
   *Lethrinus lentjan*, from waters adjacent to Tuwwal, Saudi Arabia [Red Sea]. Food and
   Agricultural Organization of the United Nations, Rome, Italy. 25 pp.
- Kinlan BP, Gaines SD (2003) Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology*, **84**, 2007-2020.
- Kochzius M, Seidel C, Hauschild J et al. (2009) Genetic population structures of the blue starfish Linckia laevigata and its gastropod ectoparasite Thyca crystalline. Marine Ecology Progress Series, 396, 211-219.
- Kool JT, Paris CB, Barber PH, Cowen RK (2011) Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. *Global Ecology and Biogeography*, **20**, 695-706.

- Kumar S, Banks TW, Cloutier S (2012) SNP discovery through next-generation sequencing and its applications. *International Journal of Plant Genomics*, **2012**, 831460.
- Lacy RC (1987) Loss of genetic diversity from managed populations: interacting effects of drift, mutation, immigration, selection, and population subdivision. *Conservation Biology*, 1, 143-158.
- Ladner JT, Palumbi SR (2012) Extensive sympatry, cryptic diversity and introgression throughout the geographic distribution of two coral species complexes. *Molecular Ecology*, **21**, 2224-38.
- Leek JT, Scharpf RB, Bravo HC *et al.* (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nature reviews. *Genetics*, **11**, 733-9.
- Legendre P, Legendre L (2012) Numerical Ecology, 3rd English Edition. Elsevier, Amsterdam, the Netherlands. 1006 pp.
- Lessios HA, Kessing BD, Pearse JS (2000) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. *Evolution*, **55**, 955-975.
- Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298-299.
- Lo Galbo AM, Carpenter KE, Reed DL (2001) Evolution of trophic types in emperor fishes (*Lethrinus*, Lethrinidae, Percoidei) based on cytochrome b gene sequence variation. *Journal of Molecular Evolution*, **54**, 754-762.
- Long JB, Giri C (2011) Mapping the Philippines' mangrove forests using Landsat imagery. **11**, 2972-81.

- Lourie SA, Green DM, Vincent ACJ (2005) Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: Hippocampus). *Molecular Ecology*, **14**, 1073–1094.
- Ludt WB, Rocha LA (2015) Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, **42**, 25-38.
- Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. *Genetics*, **152**, 1753-66.
- Magsino RM, Juinio-Meñez MA (2008) The influence of contrasting life history traits and oceanic processes on genetic structuring of rabbitfish populations *Siganus argenteus* and *Siganus fuscescens* along the eastern Philippine coasts. *Marine Biology*, **154**, 519–532.
- Marine Conservation Institute (2017) *MPAtlas*, Seattle, WA, Accessed 4 June 2017. URL: www.mpatlas.org [Accessed 04/06/2017].
- Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, 18, 1509-17.
- Mastretta-Yanes A, Arrigo N, Alvarez N, Jorgensen TH, Pinero D, Emerson BC (2015)
  Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15, 28-41.

- McManus JW, Shao KT, Lin SY (2010) Toward establishing a Spratly Islands International Marine Peace Park: ecological importance and supportive collaborative activities with an emphasis on the role of Taiwan. *Ocean Development & International Law*, **41**, 270-280.
- Meirmans PG, Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, **4**, 792-794.

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

- Mo SK, Liu YC, Wang SQ, Bo XC, Li Z, Chen Y, Ni M (2016) Exploring the efficacy of paternity and kinship testing based on single nucleotide polymorphisms. *Forensic Science International: Genetics*, **22**, 161-168.
- Morris GP, Grabowski PP, Borevitz JO (2011) Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. *Molecular Ecology*, 20, 4938-52.
- Nan F, Xue H, Yu F (2015) Kuroshio intrusion into the South China Sea: a review. *Progress in Oceanography*, **137**, 314-333.
- Nanami A, Yamada H (2009) Site fidelity, size, and spatial arrangement of daytime home range of thumbprint emperor *Lethrinus harak* (Lethrinidae). *Fisheries Science*, **75**, 1109-1116.
- Nelson JS, Hoddell RJ, Chou LM, Chan WK, Phang VP (2000) Phylogeographic structure of false clownfish, *Amphiprion ocellaris*, explained by sea level changes on the Sunda shelf. *Marine Biology*, **137**, 727-736.
- Polidoro BA, Carpenter KE, Collins L *et al.* (2016) The loss of species: mangrove extinction risk and geographic areas of global concern. *PLoS One*, **5**, e10095.

- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.
- Puritz J, Hollenbeck CM, Gold JR (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. *PeerJ*, **2**, e314v1.
- Ravago-Gotanco RG, Magsino RM, Juinio-Meñez MA (2007) Influence of the North Equatorial
   Current on the population genetic structure of *Tridacna crocea* (Mollusca: Tridacnidae)
   along the eastern Philippine seaboard. *Marine Ecology Progress Series*, 336, 161-168.
- Raynal JM, Crandall ED, Barber PH, Mahardika GN, Lagman MC, Carpenter KE (2013) Basin isolation and oceanographic features influence lineage diversification in the humbug damselfish (*Dascyllus aruanus*) in the Coral Triangle. *Bulletin of Marine Science*, 90, 513-532.
- Riginos C, Douglas KE, Jin Y, Shanahan DF, Treml EA (2011) Effects of geography and life history traits on genetic differentiation in benthic marine fishes. *Ecography*, **34**, 566-575.
- Roberts CM, McClean CJ, Veron JEN *et al.* (2002) Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science*, **295**, 1280-1284.
- Rocha LA, Bass AL, Robertson DR, Bowen BW (2002) Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology*, 11, 243-251.
- Rocha LA, Craig MT, Bowen BW (2007) Phylogeography and the conservation of coral reef fishes. *Coral Reefs*, **26**, 501-512.

- Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Resources*, **4**, 137-8.
- Russ GR, Alcala AC (1996) Marine reserves: rates and patterns of recovery and decline of large predatory fish. *Ecological Applications*, **6**, 947-961.
- Samarasin P, Shuter BJ, Wright SI, Rodd FH (2017) The problem of estimating recent genetic connectivity in a changing world. *Conservation Biology*, **31**, 126-135.
- 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.
- Santini F, Winterbottom R (2002) Historical biogeography of Indo-western Pacific coral reef biota: is the Indonesian region a centre of origin? *Journal of Biogeography*, **29**, 189-205.
- Schwartz MK, Luikart G, Waples RS (2006) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology and Evolution*, **22**, 25-33.
- Selkoe KA, Gaggiotti OE, ToBo Laboratory, Bowen BW, Toonen RJ (2014) Emergent patterns of population genetic structure for a coral reef community. *Molecular Ecology*, 23, 3064-3079.
- Shringarpure S, Xing,EP (2014) Effects of Sample Selection Bias on the Accuracy of Population Structure and Ancestry Inference. *G3: Genes*|*Genomes*|*Genetics*, **4**, 901–11.
- Silvestre GT, Pauly D (1997) Management of tropical fisheries in Asia: an overview of key challenges and opportunities. *ICLARM Conference Proceedings*, **53**, 8-25.

- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787-792.
- Spalding MD, Fox HE, Allen GR *et al.* (2007) Marine ecoregions of the world: a bioregionalization of coastal and shelf areas. *BioScience*, **57**, 573-583.
- Stehli FG, Wells JW (1971) Diversity and age patterns in hermatypic corals. *Systematic Zoology*, 20, 115-126.
- Stetter MG, Müller T, Schmid KJ (2017) Genomic and phenotypic evidence for an incomplete domestication of South American grain amaranth (*Amaranthus caudatus*). *Molecular Ecology*, **26**, 871-86.
- Stockwell BL, Larson WA, Waples RK, Abesamis RA, Seeb LW, Carpenter KE (2015) The application of genomics to inform conservation on a functionally important reef fish (*Scarus niger*) in the Philippines. *Conservation* Genetics, **17**, 239-49.
- Taylor BM, Mills JS (2013) Movement and spawning migration patterns suggest small marine reserves can offer adequate protection for exploited emperorfishes. *Coral Reefs*, **32**, 1077-87.
- Timm J, Figiel M, Kochzius M (2008) Contrasting patterns in species boundaries and evolution of anemonefishes (Amphiprioninae, Pomacentridae) in the centre of marine biodiversity. *Molecular Phylogenetics and Evolution*, **49**, 268-276.
- Toonen RJ, Puritz JB, Forsman ZH *et al.* (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, **1**, e203.
- Toor HS (1964) Biology and fishery of the pig-face bream, *Lethrinus lentian* Lacépéde. III. Age and growth. *Indian Journal of Fisheries*, **11**, 597-620.
- Treml EA, Roberts J, Halpin PN, Possingham HP, Riginos C (2015) The emergent geography of biophysical dispersal barriers across the Indo-West Pacific. *Diversity and Distributions*, 21, 465-476.
- Uthicke S, Benzie JAH (2003) Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. *Molecular Ecology*, **12**, 2635-2648.
- Vasantharajan M, Jawahar P, Venkatasamy M (2015) Recruitment pattern, virtual population analysis (VPA) and exploitation rate of *Lethrinus lentjan* (Lacepede, 1802) exploited in Thoothukudi Coast, Tamil Nadu, India. *International Journal of Marine Science*, 5, 1-4.
- Voris HK (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. *Journal of Biogeography*, **27**, 1153-1167.
- Wassef EA (1991) Comparative growth studies on *Lethrinus lentjan*, Lacépéde 1802 and *Lethrinus mahsena*, Forsskål 1775 (Pisces, Lethrinidae) in the Red Sea. *Fisheries Research*, **11**, 75-92.

Weeks R, Russ GR, Alcala AC, White AT (2009) Effectiveness of marine protected areas in the Philippines for biodiversity conservation. *Conservation Biology*, **24**, 531-40.

Weersing KA (2007) Population genetics, larval dispersal, and connectivity in marine systems (Unpublished Masters Thesis). University of Hawai'i, Honolulu, Hawai'i, USA.

- White AT, Vogt HP, Arin T (2000) Philippine coral reefs under threat: the economic losses caused by reef destruction. *Marine Pollution Bulletin*, **40**, 598-605.
- White AT, Aliño PM, Cros A et al. (2014) Marine protected areas in the Coral Triangle: progress, issues, and options. *Coastal Management*, **42**, 87-106.
- Wilson GG (1998) A description of the early juvenile colour patterns of eleven *Lethrinus* species (Pisces: Lethrinidae) from the Great Barrier Reef, Australia. *Records of the Australian Museum*, **50**, 55–83.
- Xing Y, Ouyang Z, Kapur K, Scott MP, Wong WH (2007) Assessing the conservation of mammalian gene expression using high-density exon arrays. *Molecular Biology and Evolution*, 24, 1283-5.
- Young EF, Belchier M, Hauser L *et al.* (2015) Oceanography and life history predict contrasting genetic population structure in two Antarctic fish species. *Evolutionary Applications*, 8, 486-509.

## APPENDIX A.

# AMOVA GROUPINGS TESTED FOR POOL 1

Results of AMOVA models tested for Pool 1 sampling sites, including site groupings for sampling sites Atimonan (Ati), Cebu (Ceb), Mati (Mat), Puerto Princesa (PtP), Nha Trang (NhT), and Phú Quốc (PhQ); among group *F*-statistic ( $F_{CT}$ ); standard deviation; 2.5 and 97.5% confidence intervals; variance; and *p*-values.

$F_{\rm CT}$	Std. dev.	2.5% CI	97.5% CI	Variance	<i>p</i> -value
-0.00167	0.00134	-0.00425	0.00092	-0.00002	0.49995
-0.00167	0.00134	-0.00425	0.00092	-0.00002	0.49995
-0.00014	0.00198	-0.00414	0.00371	0.00000	0.66973
0.00676	0.00191	0.00300	0.01055	0.00007	0.16298
	<i>F</i> <sub>CT</sub> -0.00167 -0.000167 -0.00014 0.00676	F <sub>CT</sub> Std. dev.   -0.00167 0.00134   -0.00167 0.00134   -0.00014 0.00198   0.00676 0.00191	F <sub>CT</sub> Std. dev.     2.5% CI       -0.00167     0.00134     -0.00425       -0.00167     0.00134     -0.00425       -0.00014     0.00198     -0.00414       0.00676     0.00191     0.00300	F <sub>CT</sub> Std. dev.     2.5% CI     97.5% CI       -0.00167     0.00134     -0.00425     0.00092       -0.00167     0.00134     -0.00425     0.00092       -0.00014     0.00198     -0.00414     0.00371       0.00676     0.00191     0.00300     0.01055	F <sub>CT</sub> Std. dev.     2.5% CI     97.5% CI     Variance       -0.00167     0.00134     -0.00425     0.00092     -0.00002       -0.00167     0.00134     -0.00425     0.00092     -0.00002       -0.00014     0.00198     -0.00414     0.00371     0.00000       0.00676     0.00191     0.00300     0.01055     0.00007

## Ellen E. Biesack

Department of Biological Sciences, Old Dominion University, Norfolk, Virginia 23529

#### **Education**

Bachelor of Science in Marine Biology, University of North Carolina Wilmington, May 2012

## **Employment**

Graduate Teaching Assistant, Old Dominion University Department of Biological Sciences, Aug. 2015-present

Research Assistant, Old Dominion University Department of Biological Sciences, Jan. 2014-Aug. 2017

Scientific Writing Consultant, Old Dominion University Department of Biological Sciences, Sep.-Dec. 2013

OCEANS/Oceans-17 Night Activity Counselor, UNC Wilmington MarineQuest, Jun.-Aug. 2013

Research Technician, UNC Wilmington Center for Marine Science, Feb.-Apr. 2013

Assistant Curator, UNC Wilmington Natural History Collections, Sep. 2010-Jul. 2012

Hatchery Technician, UNC Wilmington Center for Marine Science Oyster Research Hatchery, Feb. 2011-May 2011

Laboratory and Field Technician, UNC Wilmington Center for Marine Science Fish Ecology Lab, May-Jul. 2010

#### **Relevant Presentations**

- **Biesack, E.E.**, T.T. Oanh, & B.T. Dang. 2017. RAD-sequencing of foodfish *Lethrinus lentjan* reveals population structure in the South China Sea. Poster presented by K.E. Carpenter at the Next-Generation Sequencing and Bioinformatics Workshop at Can Tho University, Can Tho, Vietnam.
- **Biesack, E.E.** 2016. Marine Suture Zones as Evidence of Postglacial Recolonization in an Indo-Pacific Aquarium Fish. Oral presentation given at the 146th Annual AFS Meeting, Kansas City, MO.
- **Biesack, E.E.** 2015. Exploring the Population Structure of *Lethrinus lentjan*, a Popular Food Fish in Southeast Asia. Oral presentation given at the 145th Annual AFS Meeting, Portland, OR.

## **Publications**

- Dang, B.T. Q.H.D. Vu, K.E. Carpenter, V.T. Doan, T.O. Truong, T.L. Tran, A.S. Ackiss, E.E. Biesack, & B.L. Stockwell. Population genomics of the peripheral freshwater fish, *Polynemus melanochir* (Perciformes, Polynemidae) in a changing Mekong Delta. *In preparation*.
- Carpenter, K.E. & E.E. Biesack. 2016. Bony Fishes Introduction. In K.E. Carpenter & N. De Angelis, eds. The living marine resources of the Eastern Central Atlantic. Vol. 3: Bony Fishes Part 1. FAO Species Identification Guide for Fishery Purposes. Rome, FAO.