Title
Genomic Signatures of Natural Selection and Geographic Isolation in Corallivorous Snails

Author
Simmonds, Sara Elizabeth

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Genomic Signatures of Natural Selection and Geographic Isolation in Corallivorous Snails

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biology

by

Sara Elizabeth Simmonds

2016

ABSTRACT OF THE DISSERTATION

Genomic Signatures of Natural Selection and Geographic Isolation in Corallivorous Snails

by

Sara Elizabeth Simmonds

Doctor of Philosophy in Biology

University of California, Los Angeles, 2016

Professor Paul H. Barber, Chair

Our understanding of evolution in marine ecosystems is framed by theories of speciation developed in terrestrial environments. In the ocean, however, speciation processes are likely to be different than on land. A general lack of absolute barriers, and the vast distances certain organisms can travel as larvae, mean that populations likely diverge in the presence of gene flow. The objective of this dissertation is to examine the relative contribution of different mechanisms of divergence in the sea in order to deepen our understanding of speciation. We examined the population genetics of ectoparasitic snails (*Coralliophila radula*, *C. violacea*) that specialize on *Porites* corals, and occupy a vast geographic and environmental range across the Indo-Pacific. In Chapter One, we used a comparative phylogeographic approach to explore whether populations of both taxa diverged across common geographic barriers, or due to adaptation to the host. We found striking evidence of genetic structure with geography for both snail species, *and* structure concordant with host within *C. violacea* populations. These findings suggest that in addition to

historical sea level fluctuations, symbioses also contribute to diversification of these snails in the Coral Triangle. In Chapter Two, we used genome-wide data (SNPs) to investigate whether the ecological divergence we observed in *C. violacea* occurred via directional selection on different hosts and identify loci under selection. We saw genetic evidence of snail migration between hosts, as well as hybridization. By testing for F_{ST} outliers, we found loci under divergent selection, including a gene involved in the control of xenobiotic detoxification pathway gene expression, perhaps allowing snails to neutralize coral-specific toxins. These findings provide strong support for ecological divergence with gene flow, driven by adaptation to host. In Chapter Three, we focused on one ecomorph of *C. violacea* that inhabits coral reefs across a range of environmental conditions. Using genome-wide data and a global ocean-climate database, we identified signatures of geographic isolation and local adaptation. We saw four genetically distinct groups, consistent with results from Chapter One, with most divergence in peripheral populations. Searching for genetic associations with ocean climate variables, we found that the strongest driver of local adaptation was sea surface temperature variation. Our results show that local adaption to different environments likely reinforces neutral divergence, especially in peripheral populations.

The dissertation of Sara Elizabeth Simmonds is approved.

Thomas W. Gillespie

Peggy M. Fong

Paul H. Barber, Committee Chair

University of California, Los Angeles

2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor, Paul Barber. I am so grateful for the opportunity to work with a scientist who not only values biodiversity but whom also seeks to understand it and to protect one of the most extraordinary places on Earth. I am immensely indebted to you for your commitment to increasing the representation of women in science and for the chance to work as a mentor in the Diversity Project. Science is better when everyone has the opportunity to participate. Thank you also to the members of my committee, Peggy Fong, Thomas Gillespie and Thomas Smith for your guidance and all the comments that greatly improved my research and dissertation. A massive thank you to Chris Meyer for the insights and discussions that lead me to study ecological divergence in corallivorous gastropods. To Eli Meyer and Misha Matz, thank you for your endless patience while helping me troubleshoot laboratory protocols and debug scripts.

Conducting fieldwork in remote locations is not without risks, and we had our share of mishaps and injuries, but thankfully friends and strangers were there to help. I am forever indebted to Matthew Craig, Michelle Weber, Ketut, the Diver Alert Network and the doctors and nurses at the Hyperbaric Unit in Rumah Sakit Sanglah in Bali for saving my life when I needed to take a ride in the decompression chamber.

To all the Barber Lab undergraduates, graduate students, and post-docs, past and present, your support and friendship have been more than I could have hoped for. Thank you to Elizabeth Sbrocco and Ali Hamilton for showing me the ways of the Barber Lab and how to appease the PCR gods. Rita Rachmawati you are like family to me. Terima kasih untuk semuanya Mbak. Allison Fritts-Penniman, Abril Iñiguez, and Sam Cheng thank you for so many adventures, so

many songs and so much utter ridiculousness. Terima kasih banyak Dita Cahyani, Aji Wahyu Anggoro, Dian Pertiwi, dan Andre Sembiring dari Indonesian Biodiversity Research Center. All of you welcomed me to Indonesia with warm hearts, big smiles and lots of laughs. Thanks for putting up with us invading your space every summer and for sharing the immense diversity of Indonesia. Thank you to the NSF-PIRE team who introduced me to the complexity and joy of fieldwork in Indonesia and the Philippines, especially Amanda Ackiss, Brian Stockwell, and Liz Borda.

Thank you to all those that collected specimens for me including Rita Rachmawati, Allison Fritts-Penniman, Brian Stockwell, Michelle Weber, Mark Erdmann, Hayley Nuetzel and Demian Willette. Morover, thanks to those that provided assistance with fieldwork and permits: Allison Fritts-Penniman, Dita Cahyani, Aji Wahyu, Z.A. Muchlisin, Edi Rudi, Arie Muhardy, Febrian Hadinata, Pak Usef, Matthew Iacchei, Amanda Ackiss, Gianni Arlotti, David Smith, Michael Sweet and Richard Pooley. Diving support was provided by Bali Diving Academy, Critters@Lembeh, Lorenso's Cottages, Cubadak Paradiso Village, Komodo National Park, Freddies Santai Sumurtiga, Lumba-Lumba Diving Centre, Cendrawasih Bay National Park, Papua State University, Papua Diving, Kri Eco Resorts and the Korallion Lab. A special thank you to Zac Forsman and Clarissa Reboton for assistance with identifying *Porites* species. We would also like to thank the Indonesian Biodiversity Research Center at the Udayana University, the Institute for Environmental and Marine Sciences at Silliman University, Nha Trang University, and the Korallion Lab in the Maldives for institutional support.

Finally, I am so lucky to have a family who has always been so supportive of me and allowed me to follow my dreams. Thank you to my parents and my grandfather for encouraging my intense curiosity of nature as a child. To my sister, thank you for being my best friend and

biggest cheerleader. Without the enduring support and love of my husband and my little boy I would be lost. Thank you for all you do!

This work was funded by UCLA and grants from NSF-PIRE (OISE-0730256 to K. Carpenter and P.H. Barber and OISE-1243541 to F. Rohwer and P.H. Barber) and USAID (497-A-00-10-00008-00). We used the facilities at Vincent J. Coates Genomics Sequencing Laboratory in UC Berkeley, supported by NIH S10 Instrumentation Grants (S10RR029668 and S10RR027303). The Lemelson Foundation Fellowship, Conchologists of America, Sigma Xi and the UCLA Department of Ecology and Evolutionary Biology all generously provided additional funding to me. We gratefully acknowledge support from the Indonesia government including the Indonesian Ministry of Research and Technology (RISTEK), Indonesian Institute of Sciences (LIPI), BKSDA and the National Marine Park offices of Bunaken and Wakatobi. Sampling was covered under research permits obtained in Indonesia (RISTEK 2011, 198/SIP/FRP/SMNl/2012, 187/SIP/FRP/SM/VI/2013), Timor-Leste (Direccao nacional de Pescase Aquicultura 0042/DNPA/IOP/VII/11), the Maldives (Ministry of Fisheries and Aquaculture Permit No. (OTHR)30-D/INDIV/2013/116) and Hawai'i (Dept. of Land and Natural Resources SAP 2013-11).

I would like to thank my co-authors on the papers that have resulted from my dissertation. Chapter One is under revision for the *Journal of Biogeography* with my co-authors Samantha Cheng, Rita Rachmawati, Vincent Chou, Paul Barber, Nida Calumpong, and Ngurah Mahardika. S. Cheng and R. Rachmawati helped with sample collection, permits for collection and export, as well as edited the manuscript. V. Chou conducted a vast amount of lab work and processed samples to be sequenced. N. Calumpong kindly hosted me in her lab at Silliman University for

sample collection and lab work as well as edited the manuscript. N. Mahardika also hosted me at the IBRC and facilitated permits for local expeditions. P. Barber helped to develop the scope of the project and revised the manuscript. Chapter Two is currently in preparation for submission and I would like to thank my co-authors: A.L. Fritts-Penniman, S. H. Cheng, E. Meyer, P.H. Barber. A.L. Fritts-Penniman and S. H. Cheng helped to collect samples and prepared some RADseq libaries for sequencing. E. Meyer wrote custom scripts for RADseq data processing and advised on lab protocols. P.H. Barber was the PI on the project and edited the manuscript. Chapter Three is being prepared for submission and I would like to thank my co-authors: A.L. Fritts-Penniman, S. H. Cheng, E. Meyer, P.H. Barber. A.L. Fritts-Penniman and S. H. Cheng helped collect samples and prepared some RADseq libaries for sequencing. P.H. Barber was the PI on the project, helped to develop the scope of the research and edited the manuscript.
Thank you to John Wiley & Sons, Inc. for us granting permission to modify and reprint a figure from a paper by Ludt & Rocha (2015) published in the *Journal of Biogeography*.

VITA

2000	B.Sc. Zoology, University of Tasmania
2002	B.Sc. 1st Class Honours in Zoology, University of Tasmania
2002-2003	Junior Research Fellow, School of Zoology, University of Tasmania
2003-2004	Research Assistant, Dept. of Marine Sciences, University of Connecticut
2005	Teaching Assistant, *Evolution and Biodiversity*, California State University, Fullerton
2005-2007	Graduate Research Assistant, Dept. of Biological Science, California State University, Fullerton
2008	M.S. Biology, California State University, Fullerton
2007-2009	Research Associate II, Marine Science Institute, University of California, Santa Barbara
2009-2010	GAANN Fellow, University of California, Los Angeles
2010-2011	SEAS-FLAS Fellow, University of California, Los Angeles
2010-2011	Teaching Assistant, *Intro to Marine Science*, University of California, Los Angeles
2012	Laboratory Coordinator, *Intro to Marine Science*, University of California, Los Angeles
2010-2013	Teaching Assistant, *Molecular Ecology, Marine Ecology* and *Surveying Biodiversity*, Indonesian Biodiversity Research Center
2010, 2013	Lemelson Fellow, University of California, Los Angeles
2009-2014	Graduate Research Assistant, Dept. Ecology and Evolutionary Biology, University of California, Los Angeles

PUBLICATIONS

Sara E. Simmonds (née Koch)

Theses

Koch, S.E. (2002) Larval transport of the introduced northern Pacific seastar (*Asterias amurensis*) in the Derwent Estuary, Tasmania. Honours Thesis, University of Tasmania.

Koch, S.E. (2008) Exploring the use of statoliths of *Kelletia kelletii* as natural tags to estimate population connectivity across a species' range. M.S. Thesis, California State University, Fullerton.

Reports

Johnson C., **Koch. S.** & Patil J. (2004) Spatial and temporal patterns in the abundance of larvae of the northern Pacific seastar (*Asterias amurensis)* in the Port of Hobart and Storm Bay, SE Tasmania. Final Report to the Department of Sustainability and Environment, Victoria. 34p.

Peer Reviewed

Hopper J.V., White C., Lorda J, **Simmonds S.E.**, Kuris A.M., & Hechinger R.F. (2014) Reduced parasitism of a marine whelk, *Kelletia kelletii*, in its expanded geographic range. *Journal of Biogeography*. 41:1674-1684.

Simmonds S.E., Kinlan B.P., White C.C., Paradis G.L., Warner R.R., & Zacherl D.C. (2014) Geospatial statistics strengthen the ability of natural geochemical tags to estimate range-wide population connectivity in marine species. *Marine Ecology Progress Series*. 508:33-51.

Simmonds, S.E., Chou, V.S., Cheng S.H., Rachmawati R., Calumpong H.C., Mahardika G.N., & Barber P.H. Evidence of ecological divergence in the Coral Triangle. (in revision *Journal of Biogeography*).

PRESENTATIONS

Kinlan B.P., & **Koch S.E**. A novel statistical method to pinpoint larval origins and estimate range-wide connectivity in open-coast marine populations. Western Society of Naturalists, Monterey, California. November 2009.

Koch S.E., Kinlan B.P., Warner R.R., Gaines S.D., & Zacherl D.C. Building empirical estimates of larval dispersal and population connectivity in open-coast marine species. Contributed talk. Western Society of Naturalists. Monterey, California. November 2009.

Kinlan B.P., **SE Koch**, Warner R.R., & Zacherl D.C. Empirical estimation of larval connectivity across a species range from multivariate natural tag-recapture data. Contributed talk. Ocean Sciences Meeting. Portland, Oregon. January 2010.

Simmonds S.E. Host-specificity and ecological speciation in Indo-Pacific corallivorous gastropods. Invited seminar. Hawai'i Institute of Marine Biology. Coconut Island, Hawai'i. September 2011.

Simmonds S.E. Host-specificity and ecological speciation in Indo-Pacific corallivorous gastropods. Contributed talk. NSF PIRE synthesis meeting. Manila, Philippines. July 2012.

Simmonds S.E. Host-specificity and ecological speciation in Indo-Pacific corallivorous gastropods. Contributed talk. Pacific Advanced Studies Institute Genomics Workshop. Dumaguete, Philippines. July 2012.

Simmonds S.E., Rachmawati R., Cheng S.C., & Barber P.H. Host-specificity and ecological speciation in Indo-Pacific corallivorous gastropods. Contributed talk. Western Society of Naturalists. Monterey, California. November 2012.

Simmonds S.E., Fritts-Penniman A., Cheng S., Mahardika N., & Barber P.H. Adaptive divergence and genomic signatures of host-driven selection in coral-associated gastropods. Contributed poster. Gordon Research Conference, Speciation. Ventura, California. March 2015.

CHAPTER 1:

EVIDENCE OF ECOLOGICAL DIVERGENCE FROM CORAL-EATING SNAILS (GENUS *CORALLIOPHILA*) IN THE CORAL TRIANGLE

1.1 ABSTRACT

The majority of marine phylogeographic studies focus solely on physical processes shaping population genetics. However, the genetics of symbiotic taxa may also be influenced by their strong relationships with other organisms. Here, we examined the population genetic structure of coral-eating snails *Coralliophila radula* and *C. violacea*. These sister species are obligate parasites on corals and sympatric throughout much of their geographic ranges, from the Red Sea to the Eastern Pacific. We tested for genetic structure of snails in relation to coral hosts (*Porites* spp.) and known biogeographic barriers. We also examined the evolutionary relationships of coral hosts. *Coralliophila violacea* showed striking genetic structure within populations concordant with coral host lineage. Also, the most widely distributed host-associated group of *C. violacea* showed significant geographic structure, with divergence among Hawaiian populations, the Coral Triangle, and the Indian Ocean. Both snail species exhibited phylogeographic structure across the Sunda Shelf, a typical pattern in marine taxa of the region. Our findings suggest that symbiotic relationships may contribute to lineage diversification in the Coral Triangle, in addition to well-known processes like sea level fluctuations and physical oceanography.

1.2 INTRODUCTION

Our understanding of evolution in marine ecosystems is framed by theories developed in terrestrial environments (Miglietta et al., 2011). Historically, researchers have invoked geographic-based models of speciation without gene flow (i.e. allopatry) to explain the majority of diversity in terrestrial systems (Barraclough & Vogler, 2000). Such models are not a natural fit for the marine realm, however (Palumbi, 1994; Puebla, 2009). The ocean's fluid nature, and the prominence of dispersive life history stages in marine organisms, make it a unique evolutionary environment. Most marine organisms have planktonic larvae that increase the potential for gene flow between geographically separated regions. Even species with relatively modest mean dispersal distances can have dispersal kernels with long tails (Kinlan & Gaines, 2003). Long tails provide sufficient genetic connectivity to limit population divergence (Slatkin, 1987), even across broad geographic scales.

While uncommon, geographic barriers to gene flow in the ocean do exist, albeit with varying degrees of permeability. Landmasses are the most obvious, isolating biota in different ocean basins (Briggs & Bowen, 2013), both currently (e.g. Isthmus of Panama, see Lessios, 2008 for review) and in the past (e.g. Sunda Shelf, see Ludt & Rocha, 2015 for review). However, vast expanses of open ocean can isolate remote archipelagos such as Hawai'i (e.g. Polato et al., 2010; Iacchei et al., 2016; Waldrop et al., 2016) or populations spanning the Eastern Pacific Barrier (e.g. Baums et al., 2012). Additionally, large freshwater discharges like the Amazon can form barriers to gene flow for shallow-water marine species (Rocha, 2003).

Dispersal barriers are critical to the evolution and distribution of marine biodiversity, including in the world's most diverse marine ecosystem, the Coral Triangle (Barber et al., 2011;

Carpenter et al., 2011; Gaither et al., 2011; Gaither & Rocha, 2013). Low sea levels during the Pliocene and Pleistocene (Williams & Benzie, 1998; Ludt & Rocha, 2015), and more recent oceanographic features such as the Halmahera Eddy (Kool et al., 2011), create potent dispersal barriers for various reef organisms (see Barber et al 2011 and Carpenter et al., 2011 for reviews). Still, allopatric divergence alone may be insufficient to explain the Coral Triangle's extraordinary species diversity (Briggs, 1999, 2006). Processes such as ecological divergence and assortative mating can promote divergence with gene flow, but remain relatively unexplored in marine systems (Krug, 2011; Miglietta et al., 2011).

Ecological divergence is the evolution of reproductive isolation among populations driven by opposing selection in ecological niches or environments (Schluter & Conte, 2009). While widely documented in terrestrial ecosystems, ecological barriers to gene flow in the ocean have only recently been reported (Krug, 2011; Bird et al., 2012; Bowen et al., 2013). In terrestrial and freshwater systems, ecological divergence often takes place in sympatry via assortative mating in different microhabitats, or on different hosts in species with strong symbioses (Hatfield & Schluter, 1999; Matsubayashi et al., 2010). Evidence suggests that strong symbiotic relationships may similarly drive ecological divergence in the marine environment (Munday et al., 2004; Sotka, 2005; Faucci et al., 2007). Given the abundance of symbiosis in the ocean, ecological divergence on different hosts may influence marine diversity to a greater extent than previously thought.

Marine snails in the genus *Coralliophila* are symbionts of anthozoans (Oliverio et al., 2009). The sister species *C. radula* (A. Adams, 1853) and *C. violacea* (Kiener, 1836) are ectoparasites, exhibiting obligate relationships with corals in the family Poritidae (Fujioka &

Yamazato, 1983). These snails are sessile and feed suctorially on photosynthetic products sent by the corals to regenerate injured sites (Oren et al., 1998). As adults, the snails live in groups and rarely move (Soong & Chen, 1991; Oren et al., 1998). Dispersal is achieved via planktonic larvae brooded by protandrous hermaphrodite females (Soong & Chen, 1991). Both snail species have extensive geographic ranges, and occur sympatrically in coral reefs throughout the tropical and subtropical Indo-Pacific, from the Red Sea to the Eastern Pacific. They share almost identical life-history traits and ecological niches (i.e. coral host species, unpub. data S. Simmonds).

The goal of this study is to enhance our understanding of the evolutionary processes generating marine biodiversity in the Coral Triangle. Specifically, we test the hypothesis that co-distributed populations of *C. radula* and *C. violacea* will exhibit concordant patterns of phylogeographic structure, patterns that result from physical processes shaping the phylogeography of other marine organisms in the Coral Triangle (above). However, because parasitic relationships with coral hosts create the possibility of ecological divergence, we first test for the genetic structure that could result from ecological segregation among sympatric populations of snails utilizing different coral hosts.

1.3 MATERIALS AND METHODS

1.3.1 Field sampling

During 2011–2013, we collected *Coralliophila radula* and *C. violacea* from Indo-Pacific localities ($N = 14$ and $N = 17$ respectively, Table 1.1, Figure 1.1). These localities span the Sunda

Shelf Barrier, an area where phylogeographic structure is commonly observed (Barber et al., 2011), and also include known areas of isolation (i.e. Hawai'i). At each locality, we collected snails from multiple coral host species (N = 1–4), and 1–6 colonies of each coral species (Table 1.2, Figure 1.2) for a total of 235 *C. radula* and 328 *C. violacea* from 114 coral colonies representing 13 putative species. A portion of each snail's foot tissue was preserved in 95% ethanol and stored at room temperature for DNA analysis.

Porites corals are notoriously difficult to identify *in situ* due to their morphological plasticity and small corallites (Forsman et al., 2015). Genetically similar colonies can have vastly different morphologies and *vice versa* (Forsman et al., 2009, 2015). Therefore, to define coral species both morphologically and genetically, we collected detailed information about each snail's host: tagged photos of coral colonies *in situ*; took macro photos with a transparent ruler to measure corallites; and sampled tissues for genetic analysis.

1.3.2 DNA extraction and sequencing

We sequenced 1–15 snails from each coral colony. We extracted DNA using 10% Chelex™ (BioRad) (Walsh et al., 2013) or DNeasy® Blood and Tissue Kit (QIAGEN) and then amplified a 668 bp length fragment of COI mtDNA using primers HCO-2198, LCO-1490 (Folmer et al., 1994). Following an initial denaturation at 94°C for 1.5 min, the thermocycling parameters were: 94°C for 30 secs, 50°C for 30 secs, and 72°C for 45 secs for 35 cycles with a final 10 min extension at 72°C.

We extracted coral DNA using DNeasy® Blood and Tissue Kit (QIAGEN) and sequenced a 656 bp fragment of the nuclear ribosomal ITS region with primers developed by Forsman et al., (2009). Following an initial denaturation at 94°C for 1.5 min, thermocycling parameters were:

94°C for 45 secs, 50°C for 45 secs, and 72°C for 1 min for 30 cycles with a final 6 min extension at 72°C.

For both snail and coral samples, all PCR product cleanup and DNA sequencing was done by the UC Berkeley DNA Sequencing Facility.

1.3.3 Data analysis

Determining coral host identities and evolutionary relationships

To identify each coral species, test for cryptic diversity, and determine phylogenetic relationships among hosts, we aligned our sequences ($N = 20$; Table 1.3) with GenBank reference sequences representing all available *Porites* species within our sampling range. We included additional basal lineages from the Caribbean and *Goniopora* and *Stylaraea* as outgroups ($N = 37$; Table 1.3). We aligned coral sequences in Geneious v8.1.8 (Kearse et al., 2012) using the MAFFT plugin. We tested the appropriate substitution model using AIC in jModelTest v2.1.7 (Darriba et al., 2012). Maximum likelihood analyses were performed with PhyML v3 (Guidon et al., 2010) (TPM1uf+G substitution model, 9999 bootstrap replicates) and FastTree 2 (Price et al., 2010) using default settings. Bayesian analyses were performed using MrBayes v3.2.2 (Ronquist et al., 2012) (HKY model, 1,000,000 chain length, burn-in length 100,000, subsampling frequency 200, 4 heated chains). We visualized trees and annotated support values using TreeGraph 2 (Stöver & Müller, 2010).

Ecological and geographic analyses of genetic structure of Coralliophila

For the snails, we aligned and edited complementary sequences, and confirmed translations in Geneious. Sequences were aligned using the MAFFT plugin in Geneious. We trimmed final sequence alignments to 576 bp for *C. radula* and 617 bp for *C. violacea* and then reduced all

sequences to unique haplotypes using FaBox v1.41 (Villesen, 2007). We calculated standard diversity statistics (haplotype and nucleotide diversity) in ARLEQUIN v3.5 (Excoffier & Lischer, 2011).

We tested for divergence associated with coral hosts in sympatry, since genetic structure in the two *Coralliophila* species could be partitioned by host or geography. First, we created minimum spanning trees (MSTs) using Gephi v0.8.2 (Bastian et al., 2009) based on pairwise differences calculated in ARLEQUIN. We only included haplotypes from populations where *C. radula* and *C. violacea* were found on different coral host lineages at the same locality. We then plotted coral host lineage onto MSTs and ran analysis of molecular variance (AMOVA) in ARLEQUIN, partitioning the genetic data by coral host lineage. To test for the relative contributions of geographic divergence vs. divergence in sympatry, we calculated pairwise Φ_{ST} values among localities within host-associated lineages, and then calculated pairwise Φ_{ST} between host-associated populations within individual sampling locations.

To test for phylogeographic partitions, we constructed MSTs and then plotted resulting haplogroups onto geography. We then ran AMOVAs with and without *a priori* partitions to test for genetic structure related to divergence associated with isolation across the Sunda Shelf with significance determined by 100,000 random replicates in ARLEQUIN.

1.4 RESULTS

Of the two snail species, *C. radula* was less abundant than *C. violacea* (N = 235 vs. 328); found at fewer locations (N = 14 vs. 17; Table 1.2) and on fewer coral host species (8 vs. 12; Table 1.2). The two sister species of *Coralliophila* exhibited ecological niche overlap in the corals they

inhabited, having seven putative host species in common (Table 1.2). Also, within those shared hosts, *C. radula* and *C. violacea* co-occurred on about half of all sampled coral colonies.

1.4.1 Sequences and genetic diversity

We sequenced 20 *Porites* coral colonies for ITS (761 bp) and deposited them into GenBank (Table 1.3). We obtained 235 COI sequences from *C. radula* (567 bp) and 328 from *C. violacea* (617 bp), yielding a total of 192 and 296 unique haplotypes, respectively. Both snail species had high haplotype diversity (*C. radula*: $h = 0.966–1.00$ and *C. violacea*: $h = 0.900–1.00$, Table 1.4) in all populations except one: *C. violacea* (Pulau Keluang; $h = 0.667$, Table 1.4). Nucleotide diversity was low in both species (*C. radula*: $\pi = 0.011–0.024$ and *C. violacea*: $\pi = 0.011–0.040$, Table 1.4), although in all localities *C. violacea* had a higher average number of polymorphic sites ($N = 67$) than *C. radula* ($N = 41$).

1.4.2 Phylogenetics of coral host species

The phylogenetic tree of 46 *Porites* coral ITS sequences (Table 1.3, Figure 1.3) revealed that many morphologically distinct species were genetically indistinguishable from each other. For example, Clade 1 (Figures 1.2 and 1.3) had six morphologically different species including massive, branching and plating forms. Clade 3 included massive *P. lutea* and small nodular *P. brighami*. Clade 4 consisted of branching and plating species of different sizes (Figures 1.2 and 1.3; this study: *P. cylindrica*, *P. negrosensis*; GenBank: *P. lichen* and *P. randalli*). However, some clades did consist of morphologically similar species. For instance, Clade 2 had three species with nodular growth forms of varying size (*P. annae, P. evermanni* and *P. rus*) (Figures 1.2 and 1.3). Results also showed incongruities among samples of the same species collected from different localities. For example, *P. cylindrica* collected in this study from Indonesia did

not group with sequences from GenBank obtained from Samoa and Fiji (Figure 1.3). This mismatch could be a case of misidentification or cryptic diversity. Given the well-documented taxonomic challenges with *Porites* (Forsman et al., 2009, 2015) and these uncertainties, we opted to group coral host species by well supported genetic lineages (Figure 1.3) instead of morphological taxonomy. We used these groupings for all tests of genetic structure concerning coral host.

1.4.3 Analysis of ecological divergence in *Coralliophila*

To investigate the genetics of snails for structure in relation to coral host, we first built MSTs of haplotypes from six localities (Pemuteran, Nusa Penida, Pulau Mengyatan, Lembeh, Bunaken and Dumaguete) where snails from both *Porites* lineages were sampled (Table 1.2). There was no evidence of genetic structure with host in *C. radula*, which we confirmed with AMOVA analyses (*Porites* lineage 1 vs. 2) ($\Phi_{CT} = -0.018$, $P = 1.000$, Table 1.5).

In contrast, the MST of *C. violacea* on sympatric hosts (*Porites* lineage 1 vs. 2) showed two clades (A and B), largely concordant with host lineages (Figure 1.4). Clade A of *C. violacea* (Figure 1.4) was found predominately on six species of *Porites* (*P. lobata, P. attenuata*, *P. compressa*, *P. rus*, *P. solida*, *P. lutea*) all belonging to one genetic lineage, *Porites* lineage 1 (Figure 1.3). Clade B of *C. violacea* (Figure 1.4) was found on four different *Porites* species (*P. cylindrica*, *P. negrosensis*, *P. nigrescens* and *P. tuberculosus*) from *Porites* lineage 2 (Figure 1.3). AMOVA also showed marked genetic differentiation between hosts in sympatry ($\Phi_{CT} = 0.561$, $P = 0.003$, Table 1.5), but no structure among populations within host ($\Phi_{SC} = 0.003$, $P = 0.328$, Table 1.5). Despite these distinctions, we found occasional mismatches between *C. violacea* mtDNA background and their host (Figure 1.4). Some *C. violacea* collected from

Porites lineage 2 species fell in MST Clade A (mean = 11.7%). However, we identified only one mismatch the other way, when coral hosts were sympatric (Figure 1.3).

1.4.4 Phylogeographic analyses of *Coralliophila*

Because there was no observed ecological divergence in *C. radula,* we tested for phylogeographic structure using all haplotypes. The MST revealed three deeply divergent haplogroups separated by 18 or more steps that were concordant with geography (Figure 1.5). The red group was restricted to sites in the Indian Ocean (Figure 1.5). The blue group was the most common, present at all sites in the Pacific Ocean (Figure 1.5). The yellow group was rarest and found only within the Coral Triangle (Figure 1.5).

Due to the strong genetic associations by coral host lineage in *C. violacea*, we tested for geographic structure separately within samples collected from each coral lineage. The MST of *C. violacea* collected from *Porites* lineage 1 distinguished six haplogroups (Figure 1.6). The blue group was most common, present at all sites, and dominating at Coral Triangle sites (Figure 1.6). The red group dominated sites in the Indian Ocean (>75%, Figure 1.6). The yellow group was found almost exclusively within the Coral Triangle, with the exception of Pulau Weh in western Indonesia, and was concordant with snails found on hosts of the mismatched genetic lineage. The purple group was restricted to Hawai'i (Figure 1.6). There were also two rare, but divergent, haplogroups (turquoise, pink) only seen at sites (Hon Mun, Ticao Dili, and Bunaken) within the Coral Triangle (Figure 1.6).

Non-hierarchical AMOVAs of all haplotypes showed significant genetic structure in *C. radula* ($\Phi_{ST} = 0.531$, $P = 0.000$; Table 1.5) and *C. violacea* ($\Phi_{ST} = 0.213$, $P > 0.001$; Table 1.5). The percent variation in *C. radula* was almost equal among (53%) and within (47%) populations.

However, in *C. violacea* more variation was present within (79%) than among (21%) populations (Table 1.5).

Hierarchical AMOVA analyses comparing *C. radula* populations from the Indian Ocean and the Coral Triangle + Hawai'i (i.e. spanning the Sunda Shelf), revealed a prominent genetic break (Φ_{CT} = 0.735, P = 0.011, ~5% sequence divergence), with the most variation (74%) between ocean basins (Table 1.5). However, only 0.5% of the variation was among populations within oceans (Φ_{SC} = 0.018, P = 0.039, Table 1.5). More isolated locations like Hawai'i were significantly different from a few populations in the Coral Triangle (Dumaguete, Pemuteran, Pulau Mengyatan), but only marginally so (pair-wise Φ_{ST} = 0.120–0.140, Table 1.6).

Where sample sizes were sufficient, a non-hierarchical AMOVA of *C. violacea* from *Porites* lineage 2 in the Coral Triangle showed no significant genetic structure among populations (2% var; Φ_{ST} = 0.020, P = 0.091). However, a few pairwise Φ_{ST} distances were significant: populations in Lembeh were different from Komodo (Φ_{ST} = 0.048), and South Bali (Φ_{ST} = 0.109).

Because Hawaiian populations of *C. violacea* from *Porites* lineage 1 were distinct in the MST (Figure 1.6), we defined three partitions: 1) the Indian Ocean, 2) the Coral Triangle, and 3) Hawai'i for AMOVAs. Genetic structure was strong (Φ_{CT} = 0.427, P = 0.002, Table 1.7) with 43% of the variation among regions (Table 1.7). Snails from Hawai'i were the most genetically distinct, resulting in the highest pair-wise Φ_{ST} values (Φ_{ST} = 0.475–0.689, Table 1.7). Populations from Hon Mun in Vietnam were also genetically distinct from all other populations except Dili in Timor Leste (Φ_{ST} = 0.081–0.447, Table 1.7).

1.5 DISCUSSION

Although phylogeographic studies in the Coral Triangle typically focus on allopatric divergence, results from the corallivorous snail *C. violacea,* showed striking evidence for ecological divergence. Two groups of *C. violacea* were strongly concordant with the lineage of the *Porites* coral from which we collected the snails. Even within individual populations, there was genetic divergence among snails collected from different coral hosts. Given the high prevalence of symbioses on the reefs of the Coral Triangle, the recovery of ecological divergence in *C. violacea* suggests that ecology could be a major driver of lineage diversification in this mega-diverse ecosystem.

Even though *C. radula* did not exhibit ecological divergence, both snail species showed evidence of phylogeographic structure across the Sunda Shelf, as predicted for ecologically similar, sympatrically distributed sister taxa. This classic phylogeographic pattern is observed in a wide diversity of Indo-Pacific marine taxa (see Barber et al., 2011; Carpenter et al., 2011 for reviews), and is typically attributed to eustatic sea level fluctuations, This pattern provides evidence of allopatric divergence in these two snails within the Coral Triangle. In addition to structure across the Sunda Shelf, *C. violacea* populations in Hawai'i were also highly divergent, indicating divergence at both the center and in more isolated areas of this species' range.

1.5.1 Ecological barriers to gene flow

Within the Coral Triangle, two sympatric haplogroups of *C. violacea* were concordant with coral host lineages. Ecological divergence among populations inhabiting sympatric host taxa is commonly reported for terrestrial species, particularly phytophagous insects such as fruit flies (Bush, 1969), pea aphids (Peccoud et al., 2009), butterflies (Fordyce, 2010) and stick insects

(Nosil et al., 2012). However, marine studies have not typically found evidence for genetic structure among populations on different, sympatrically distributed hosts (e.g. Sotka et al., 2003; Johnston et al., 2012; Li & O'Foighil, 2012), with the exception of sponge-dwelling snapping shrimp (Duffy, 1996). Interestingly, genetic structure was not observed among snails on individual species of corals, but among groups of related species, with *C. violacea* Clade A mainly associated with three clades of *Porites* that include 12 sampled species. In contrast, Clade B was only associated with four species of *Porites*. It is important to note that two coral host species could not be sequenced in this or previous studies, so we cannot determine the phylogenetic relationship of *P. tuberculosus* or *P. nigrescens* within *Porites.* However, based on the genetic signature of snails collected from these species, we predict that *P. nigrescens* would group with *Porites* lineage 1, and *P. tuberculous* would group with *Porites* lineage 2. Outside of shared evolutionary history, it is unclear what unites or distinguishes these groups of corals, as they both include a diversity of morphologies.

While sympatric populations of parasites from different hosts can be genetically distinct, they still frequently exchange genes (Drès & Mallet, 2002). Indeed, the small number of mismatched mtDNA haplotypes seen on the *C. violacea* host MST could be the result of either incomplete lineage sorting, or migration, suggesting some genetic exchange. However, even reduced gene flow resulting from segregation by host can, over time, lead to speciation (Matsubayashi et al., 2010). Phylogenetic studies of symbiotic marine taxa have discovered host-specific cryptic species in anemone-dwelling snapping shrimp (Hurt et al., 2013) and anthozoan-associated hydroids (Montano et al., 2015), barnacles (Tsang et al., 2009), snails (Gittenberger & Gittenberger, 2011), nudibranchs (Faucci et al., 2007), and fishes (Munday et al., 2004).

The significant genetic patterns we report in *C. violacea* could be the result of host-associated haplogroups having distinct host preferences and experiencing differential selection. Previous studies have hinted at host preferences in *C. violacea* (Fujioka & Yamazato, 1983), as well as differential selection on different host morphologies (Chen et al., 2004). However, those studies did not characterize the coral genetics, making their results difficult to interpret in the context of this work. Our results emphasize the importance of collecting both host and symbiont data for DNA testing, especially given the challenges of coral taxonomy.

It is unclear what is driving the strong association between the divergent *C. violacea* haplogroups and their unique assemblages of coral hosts. However, possibilities include 1) larval settlement cues, 2) differences in nutritional quality of corals, or 3) secondary metabolites and pigments in the corals. Other coral-eating gastropods such as *Phestilla* nudibranchs use chemical cues from their hosts to induce larval settlement and metamorphosis (Ritson-Williams et al., 2009). Growth rates, recovery times (see Henry & Hart, 2005 for review), and the quality and quantity of available nutrients (Yamashiro et al., 1999; Baums et al., 2003), vary among coral species. *Porites* have small corallites and weak nematocysts, providing minimal physical protection from predators (Connell, 2012). Other corals with weaker physical protection (i.e. gorgonians, cup corals, soft corals) have chemical defenses (Wang et al., 2008) suggesting that secondary metabolites for chemical defense may be present in *Porites*. Whether secondary metabolites, settlement cues, or nutrients, the interactions between parasites and hosts are chemically mediated, suggesting a fruitful avenue of future research for understanding the ecological and evolutionary dynamics of host-parasite associations.

The ~2–3% COI sequence divergence we observed between groups of host-associated *C.*

violacea indicates that ecological divergence began ~1.4–3.6 Ma (Plio-Pleistocene), assuming a heuristic molecular clock with a conservative divergence rate of 0.7–1.2%/myr for molluscan COI (Marko, 2002). Using time-dependency of substitution rates in other marine invertebrates from the region (2.3–6.7%/Ma; Crandall et al., 2011), divergence could have started much sooner 0.3–1.3 Ma, within the Pleistocene (2.5 Ma–11.7 ka).

1.5.2 Geographic barriers to gene flow

Co-distributed species with equivalent ecologies and life histories should be impacted by broadly acting physical processes in similar ways (Avise, 2000). Both *C. radula* and *C. violacea* exhibited substantial genetic divergence between the Indian and Pacific Ocean basins spanning the Sunda Shelf. During the Pleistocene, sea levels repeatedly dropped by 100–150 m, cyclically exposing the Sunda and Sahul shelves (Voris, 2000) that created a partial barrier between the two oceans, lasting for up to ~15,000–30,000 years. Genetic structure among marine organism populations spanning the Sunda Shelf is typically attributed to these sea level changes (Gaither & Rocha, 2013). Numerous marine molluscs show phylogeographic structure across this region (Crandall et al., 2008a; DeBoer et al., 2008; Kochzius & Nuryanto, 2008; Nuryanto & Kochzius, 2009).

The 5% COI sequence divergence observed in *C. radula* suggests that separation across the Sunda Shelf began at the latest at the beginning of the Pleistocene (~2.5 Ma), assuming a heuristic molecular clock with a conservative divergence rate of 1%/myr for molluscan COI (Marko, 2002; 0.7–1.2%/myr). However, time-dependency of substitution rates in other marine invertebrates from this region yield estimates of 2.3–6.7%/Ma (Crandall et al., 2011), indicating that divergence could have occurred less than 1 Ma. Either way, these values place divergence

within periods of Pleistocene sea level fluctuations.

C. violacea likely colonized the Hawaiian archipelago around 0.5–1.7 Ma based on 3.5–4% COI sequence divergence between Hawaiian and other Indo-Pacific populations. The genetic isolation of Hawaiian populations of *C. violacea* is seen in many other Indo-Pacific species (summarized in Gaither et al., 2011), and the levels of population structure were similar as well (COI, Φ_{ST} = 0.08–0.89; Skillings et al., 2011). Surprisingly, there was only weak genetic structure between populations of the sister species *C. radula* in Hawai'i and a few Coral Triangle populations. It remains unclear why species with nearly identical ecological niches and life history strategies, that inhabit the same hosts, and overlap in the majority of their geographic ranges, would have concordant patterns in one part of their range, but discordant patterns in the other. It is unlikely that they differ drastically in planktonic larval duration, so possible explanations include different population demographics or the timing of colonization or expansion into different parts of their ranges. For example, as evidenced by star polytomies in MSTs and relative abundances, *C. radula* may have experienced recent population expansions in the Coral Triangle; whereas, *C. violacea* may have expanded in the Indian Ocean. Similarly, subtle ecological differences might structure populations in ways we cannot untangle without collecting information (e.g. microhabitat) for each specimen. For instance, while both species co-occur on the same coral hosts, they may specialize on different microhabitats or nutrients within a host.

1.5.3 A tale of two species

As with the discordant phylogeographic structure between *C. radula* and *C. violacea,* it is puzzling that *C. violacea* has diverged on different coral lineages, while *C. radula* has not. There

are several possible explanations. Each species might be responding differently to the same selective pressures, because of the different evolutionary histories affecting the genetic background upon which selection acts (Prunier et al., 2012). Perhaps *C. radula* does not have the same level of standing genetic variation. Alternatively, *C. radula* may, in fact, be diverging on different hosts such that selection *is* occurring in the face of gene flow at particular loci, but selection is too weak and migration too low or high for divergence to be evident in neutral loci (Thibert-Plante & Hendry, 2010). Previous studies reported adaptations to different hosts by herbivorous marine invertebrates, and genetically mediated differences in fitness on hosts (e.g. Sotka et al., 2003), yet found no genetic structure in mitochondrial DNA. Also, differences in fitness and selection between host-associated populations could be maintained under ongoing gene flow; a process demonstrated in numerous other systems (e.g. Mullen & Hoekstra, 2008). Genome-wide sequencing is needed to look for loci under selection, and estimate levels of gene flow between populations. Finally, the two snail species may not, in fact, be sister to each other, and another undiscovered species could change our understanding of the system's evolutionary dynamics.

1.6 CONCLUSIONS

Allopatric speciation was such a dominant model of speciation, that early terrestrial studies reporting sympatric speciation mediated by ecological differences (*ecological speciation*) were met with considerable skepticism (Bird et al., 2011). Today, a growing body of literature indicates ecological speciation is more common than previously thought. While studies of ecological divergence in the ocean are still in their infancy, the pervasiveness of obligate host

relationships in the marine environment suggests that ecologically mediated divergence and speciation could be important in the evolution of marine biodiversity, particularly in hyper-diverse regions like the Coral Triangle.

1.7 TABLES AND FIGURES

Table 1.1 Sampling localities for *Coralliophila radula*, *C. violacea*. Coordinates are in decimal degrees. Multiple sites were sometimes sampled at each locality. Locality numbers correspond to those in Figure 1.1. Regions were used for AMOVA analyses.

Locality	Region	Country	Province/state	Latitude	Longitude
1. Vav'varu	Indian Ocean	Maldives	North Province	5.419	73.358
2. Pulau Weh	Indian Ocean	Indonesia	Aceh	5.887	95.348
3. Pulau Keluang	Indian Ocean	Indonesia	Aceh	5.129	95.294
4. Pulau Pagang	Indian Ocean	Indonesia	Sumatra	-1.157	100.352
5. Hon Mun	Coral Triangle	Vietnam	Nha Trang	12.170	109.308
6. Pemuteran	Coral Triangle	Indonesia	Bali	-8.140	114.654
7. Nusa Penida	Coral Triangle	Indonesia	Bali	-8.675	115.513
8. Pulau Mengyatan	Coral Triangle	Indonesia	East Nusa Tenggara	-8.557	119.685
9. Wangi-Wangi	Coral Triangle	Indonesia	South Sulawesi	-5.269	123.519
10. Dili	Coral Triangle	Timor-Leste	Timor Island	-8.477	125.911
11. Lembeh	Coral Triangle	Indonesia	North Sulawesi	1.479	125.251
12. Bunaken	Coral Triangle	Indonesia	North Sulawesi	1.612	124.783
13. Dumaguete	Coral Triangle	Philippines	Negros Oriental	9.332	123.312
14. Ticao	Coral Triangle	Philippines	Luzon	12.628	123.706
15. Raja Ampat	Coral Triangle	Indonesia	West Papua	-0.559	130.672
16. Manokwari	Coral Triangle	Indonesia	West Papua	-0.888	134.085
17. Ka'a'awa	Hawai'i	USA	O'ahu	21.584	-157.887

Table 1.2 Number of mitochondrial cytochrome oxidase I (COI) sequences from *Coralliophila radula* and *C. violacea* at each locality collected from available coral hosts. Photo vouchers of coral species are shown in Figure 1.2. Localities in bold face are where snails from both *Porites* lineages were sampled.

Locality	Collection host	*C. radula*	*C. violacea*
1. Vav'varu	*P. lobata*	21	13
	P. rus	5	–
	P. sp1	17	–
2. Pulau Weh	*P. annae*	–	1
	P. lobata	4	22
	P. rus	4	3
3. Pulau Keluang	*P. lobata*	–	3
4. Pulau Pagang	*P. lobata*	–	5
5. Hon Mun	*P. lobata*	–	33
6. Pemuteran	*P. attenuata*	15	–
	P. cylindrica	5	17
	P. lobata	10	15
	P. negrosensis	3	5
7. Nusa Penida	*P. cylindrica*	–	5
	P. lobata	7	9
8. Pulau Mengyatan	*P. cylindrica*	–	8
	P. lobata	12	8
	P. nigrescens	–	2
9. Wangi-Wangi	*P. attenuata*	16	–
	P. cylindrica	–	1
10. Dili	*P. lobata*	3	2
11. Lembeh	*P. cylindrica*	–	12
	P. lobata	4	9
	P. tuberculosus	–	14
12. Bunaken	*P. cylindrica*	–	18
	P. lobata	6	23
13. Dumaguete	*P. cylindrica*	18	22
	P. lobata	15	32
14. Ticao	*P. lobata*	26	8
15. Raja Ampat	*P. attenuata*	20	7
	P. lobata	6	–
	P. rus	6	–
16. Manokwari	*P. lobata*	1	–
	P. lutea	5	12
17. Ka'a'awa	*P. compressa*	–	12
	P. evermanni	5	4
	P. lutea	1	1
	P. solida	–	2

Table 1.3 Coral ITS sequence species IDs, sample codes, localities and GenBank accession numbers for **a)** samples from this study and **b)** reference samples from GenBank.

Sample Code	Species ID	Locality	GenBank
a)			
MA002.06	*Porites lobata*	1. Vav'varu, Maldives	
MA003.15	*Porites lobata*	1. Vav'varu, Maldives	
MA007.10	*Porites rus*	1. Vav'varu, Maldives	
MA010.10	*Porites rus*	1. Vav'varu, Maldives	
MA014.11	*Porites rus*	1. Vav'varu, Maldives	
MA015.10	*Porites rus*	1. Vav'varu, Maldives	
ID2053.02	*Porites annae*	2. Pulau Weh, Indonesia	
VT0002.17	*Porites lobata*	5. Hon Mun, Vietnam	
ID2004.11	*Porites cylindrica*	6. Pemuteran, Indonesia	
ID2005.09	*Porites cylindrica*	6. Pemuteran, Indonesia	
ID2007.03	*Porites attenuata*	6. Pemuteran, Indonesia	
ID2024.02	*Porites cylindrica*	6. Pemuteran, Indonesia	
ID2029.03	*Porites negrosensis*	6. Pemuteran, Indonesia	
ID2037.02	*Porites cylindrica*	9 .Wangi-Wangi, Indonesia	
ID2023.16	*Porites attenuata*	9 .Wangi-Wangi, Indonesia	
ID2015.10	*Porites lobata*	15. Raja Ampat, Indonesia	
ID2018.05	*Porites lobata*	15. Raja Ampat, Indonesia	
ID2058.04	*Porites attenuata*	15. Raja Ampat, Indonesia	
ID2059.17	*Porites rus*	15. Raja Ampat, Indonesia	
HI006.04	*Porites compressa*	17. Ka'a'awa, Hawai'i, USA	
b)			
	Stylaraea punctata	Japan	AB907017
	Goniopora sp.	Fiji	FJ416593
	Porites lichen	Japan	AB907024
	Porites lobata	Australia	AY320308
	Porites lobata	Cook Islands	AY320317
	Porites lobata	Tahiti	AY320325
	Porites lobata	Galapagos	AY320338

Sample Code	Species ID	Locality	GenBank
	Porites lobata	Fiji	AY320348
	Porites asteroides	Gulf of Mexico	AY458021
	Porites divaricata	Belize	AY458037
	Porites divaricata	Belize	AY458038
	Porites furcata	Panama	AY458044
	Porites furcata	Panama	AY458045
	Porites sverdrupi	Mexico	AY458050
	Porites rus	Tahiti	AY458057
	Porites colonensis	Panama	AY458063
	Porites solida	Samoa	FJ416503
	Porites solida	Samoa	FJ416504
	Porites lichen	Samoa	FJ416506
	Porites lichen	Samoa	FJ416507
	Porites annae	Samoa	FJ416514
	Porites cylindrica	Samoa	FJ416527
	Porites randalli	Samoa	FJ416532
	Porites lutea	Samoa	FJ416549
	Porites lutea	Samoa	FJ416550
	Porites lutea	Samoa	FJ416551
	Porites cylindrica	Samoa	FJ416555
	Porites cylindrica	Samoa	FJ416556
	Porites compressa	Hawai'i	FJ416557
	Porites duerdeni	Hawai'i	FJ416558
	Porites evermanni	Hawai'i	FJ416559
	Porites rus	Hawai'i	FJ416576
	Porites brighami	Hawai'i	FJ416577
	Porites monticulosa	Hawai'i	FJ416578
	Porites cylindrica	Fiji	FJ416594
	Porites fontanesii	Yemen	HE585990
	Porites panamensis	Panama	KC178871

Table 1.4 *Coralliophila radula* and *C. violacea.* Population level summary statistics and neutrality test statistics.

	C. radula				*C. violacea*			
Locality	**N**	***h***	**π (%)**	**θs**	**N**	***h***	**π (%)**	**θs**
1. Vav'varu	43	0.996	0.011	15.716	13	0.987	0.013	4.573
2. Pulau Weh	10	0.978	0.013	10.251	26	0.988	0.015	16.510
3. Pulau Keluang	-				3	0.667	0.021	12.667
4. Pulau Pagang	-				5	0.900	0.032	21.600
5. Hon Mun	-				33	0.987	0.011	16.509
6. Pemuteran	33	0.966	0.014	17.248	37	1.000	0.035	21.598
7. Nusa Penida	6	1.000	0.017	11.825	14	1.000	0.035	25.156
8. Pulau Mengyatan	12	0.970	0.012	9.272	18	1.000	0.037	21.805
9. Wangi-Wangi	17	0.993	0.017	16.860	2	1.000	0.040	24.000
10. Dili	3	1.000	0.013	7.333	35	0.998	0.029	23.797
11. Lembeh	4	1.000	0.012	7.636	41	0.999	0.036	26.645
12. Bunaken	6	1.000	0.024	15.766	54	0.999	0.034	7.346
13. Dumaguete	33	0.998	0.013	15.030	8	1.000	0.021	0.000
14. Ticao	25	0.993	0.016	14.036	7	0.952	0.029	21.633
15. Raja Ampat	32	0.990	0.017	17.630	12	1.000	0.022	19.537
16. Manokwari	6	1.000	0.023	14.891	19	1.000	0.020	7.550
17. Ka'a'awa	5	1.000	0.016	8.640	13	0.987	0.013	4.573

Table 1.5 *Coralliophila* AMOVA results from ARLEQUIN testing hypotheses about **a)** Non-hierarchical, **b)** Host: sympatric populations of *C. radula* (5. Pemuretan, 13. Dumaguete) and *C. violacea* (5. Pemuteran, 6. Nusa Penida, 7. Pulau Mengyatan, 12. Bunaken, 13. Dumaguete) with snails from each coral host lineage, and **c)** Geography: *C. radula* (Indian Ocean, Coral Triangle + Hawai'i); *C. violacea* from *Porites* lineage 1 (Indian Ocean, Coral Triangle, Hawai'i). Significant values are bolded.

	C. radula			***C. violacea***		
Source of variation	**Fixation indices**	**P-values**	**% var.**	**Fixation indices**	**P-values**	**% var.**
a) Non-hierarchical						
Among populations	**Φ_{ST} 0.531**	**0.000**	53.10	**Φ_{ST} 0.213**	**0.000**	78.72
Within populations			46.90			21.28
b) Host						
Between hosts	Φ_{CT} -0.018	1.000	-1.08	Φ_{CT} **0.561**	**0.003**	56.14
Among populations	Φ_{SC} 0.022	0.055	2.22	Φ_{SC} 0.003	0.328	0.13
Within populations	Φ_{ST} 0.004	0.165	99.58	**Φ_{ST} 0.563**	**0.000**	43.73
c) Geography				***C. violacea* *Porites* lineage 1**		
Between regions	**Φ_{CT} 0.735**	**0.011**	73.49	**Φ_{CT} 0.427**	**0.002**	42.65
Among population	**Φ_{SC} 0.018**	**0.039**	0.46	**Φ_{SC} 0.056**	**0.000**	3.19
Within population	**Φ_{ST} 0.740**	**0.000**	26.04	**Φ_{ST} 0.458**	**0.000**	54.16

Table 1.6 *Coralliophila radula*. Pairwise population Φ_{ST} comparisons.

	Indian Ocean		Coral Triangle + Hawai'i											
Locality	**1**	**2**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
1. Vav'varu	0													
2. Pulau Weh	0.017	0												
6. Pemuteran	**0.773**	**0.741**	0											
7. Nusa Penida	**0.778**	**0.721**	-0.021	0										
8. Pulau Mengyatan	**0.790**	**0.760**	-0.012	-0.010	0									
9. Wangi-Wangi	**0.762**	**0.709**	-0.011	-0.050	-0.012	0								
10. Dili	**0.792**	**0.748**	-0.061	-0.117	-0.116	-0.089	0							
11. Lembeh	**0.802**	**0.763**	0.090	0.033	**0.117**	0.039	0.106	0						
12. Bunaken	**0.749**	**0.666**	0.017	-0.032	0.015	0.000	-0.099	0.061	0					
13. Dumaguete	**0.788**	**0.762**	0.006	0.004	0.002	0.009	-0.032	**0.094**	0.036	0				
14. Ticao	**0.761**	**0.715**	**0.033**	-0.029	0.024	0.006	-0.068	0.042	0.017	**0.026**	0			
15. Raja Ampat	**0.757**	**0.711**	**0.018**	-0.017	0.022	0.007	-0.048	0.012	0.000	0.001	-0.001	0		
16. Manokwari	**0.746**	**0.662**	0.026	-0.031	0.014	0.012	-0.121	0.093	-0.092	**0.073**	0.031	0.041	0	
17. Ka'a'awa	**0.775**	**0.719**	**0.119**	0.023	**0.122**	0.069	-0.028	0.114	0.043	**0.140**	0.005	0.071	0.013	0

Table 1.7 *Coralliophila violacea* collected from *Porites* lineage 1 hosts. Pairwise population Φ_{ST} comparisons.

	Indian Ocean				Coral Triangle											Hawai'i
Locality	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**
1. Vav'varu	0															
2. Pulau Weh	-0.021	0														
3. Pulau Keluang	-0.040	-0.016	0													
4. Pulau Pagang	0.085	0.082	-0.103	0												
5. Hon Mun	**0.340**	**0.375**	**0.385**	**0.447**	0											
6. Pemuteran	**0.227**	**0.287**	**0.201**	**0.279**	**0.170**	0										
7. Nusa Penida	**0.166**	**0.229**	0.102	**0.168**	**0.178**	0.002	0									
8. Pulau Mengyatan	**0.328**	**0.383**	**0.292**	**0.322**	**0.336**	**0.032**	0.039	0								
10. Dili	**0.357**	**0.448**	0.158	0.305	0.261	0.234	0.150	0.328	0							
11. Lembeh	**0.230**	**0.299**	**0.170**	**0.244**	**0.114**	-0.021	-0.030	0.039	0.104	0						
12. Bunaken	0.150	0.206	0.102	0.178	**0.111**	-0.011	-0.016	**0.064**	0.086	-0.028	0					
13. Dumaguete	**0.207**	**0.259**	**0.193**	**0.244**	**0.183**	-0.014	-0.018	-0.010	0.242	-0.005	0.015	0				
14. Ticao	**0.194**	**0.258**	0.146	**0.218**	**0.183**	-0.047	-0.078	0.001	0.182	-0.040	-0.039	-0.049	0			
15. Raja Ampat	**0.176**	**0.249**	0.076	**0.171**	**0.106**	0.066	0.035	**0.137**	-0.108	0.006	0.015	**0.089**	0.032	0		
16. Manokwari	**0.183**	**0.246**	0.134	**0.224**	**0.081**	0.016	-0.007	0.126	0.102	-0.003	0.000	0.037	-0.026	0.008	0	
17. Ka'a'awa	**0.671**	**0.708**	**0.640**	**0.657**	**0.689**	**0.635**	**0.597**	**0.648**	**0.475**	**0.602**	**0.579**	**0.624**	**0.615**	**0.538**	**0.599**	0

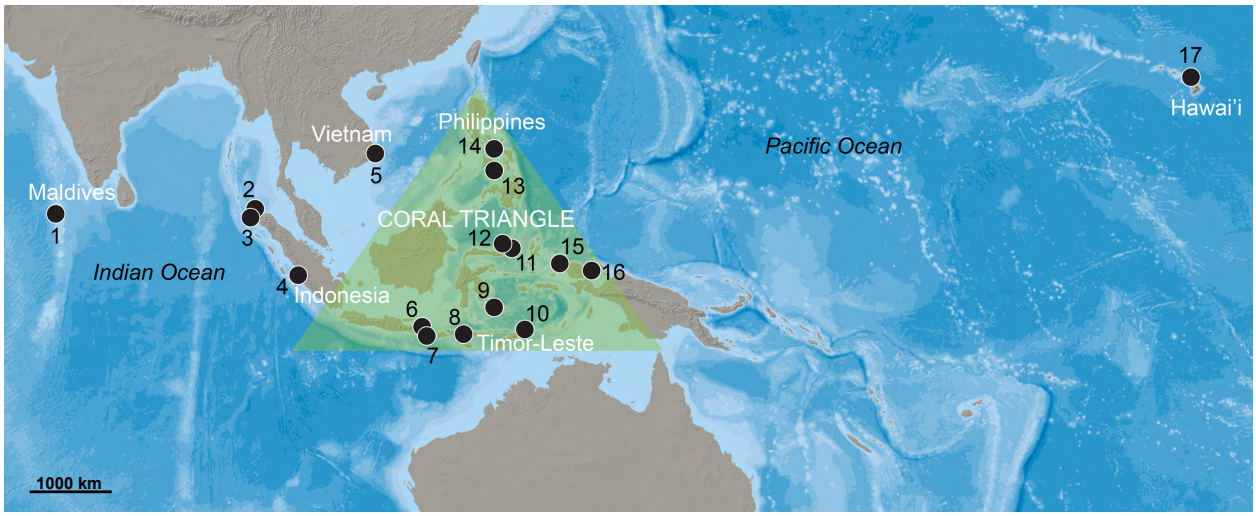


Figure 1.1 Population sampling localities across the Indo-West Pacific for ectoparasitic snails (*Coralliophila radula*, *C. violacea*) and on a suite of coral hosts (*Porites* spp.). Locality names and coordinates are shown in Table 1.1. Raster map made with Natural Earth.

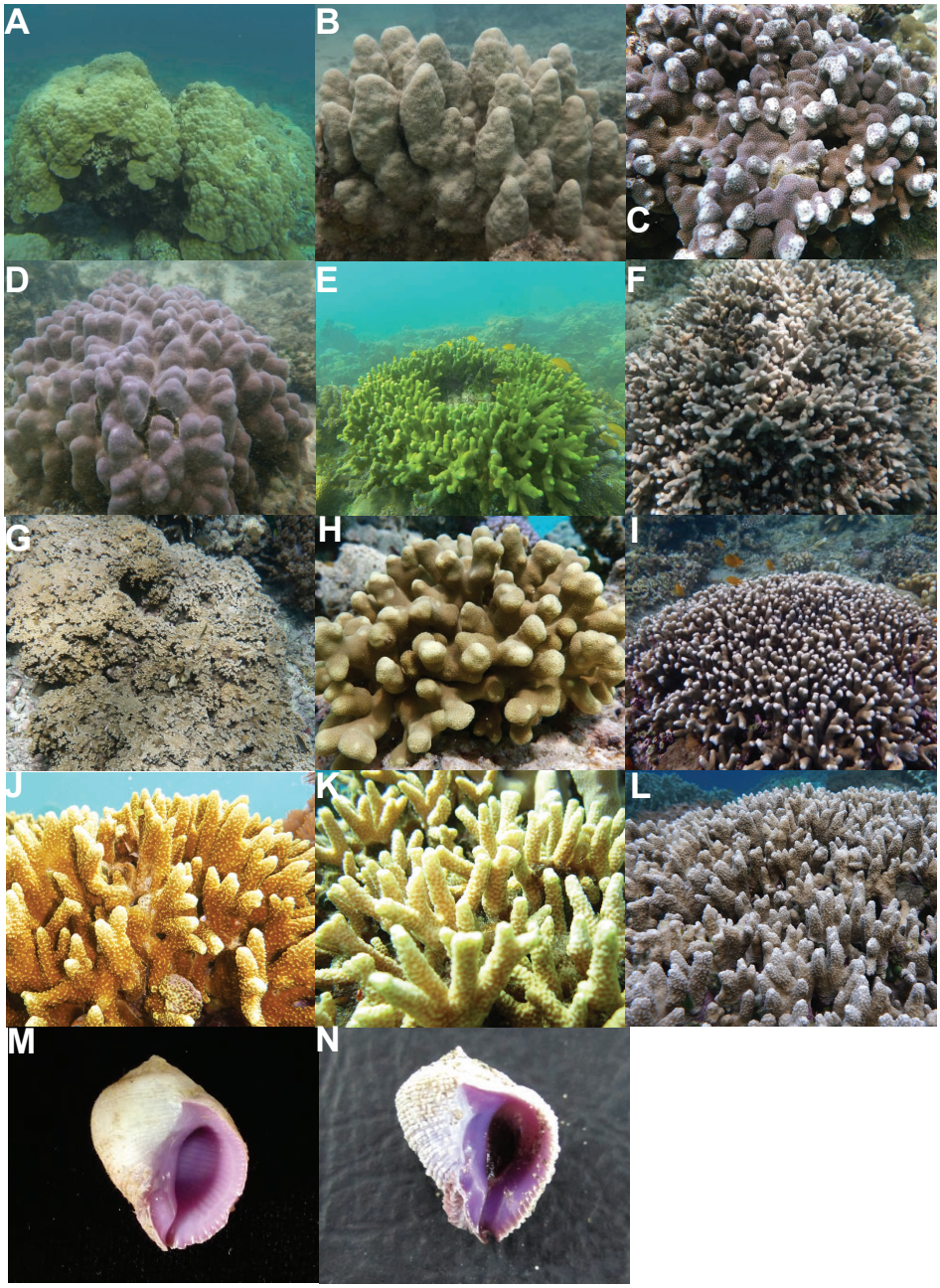


Figure 1.2 Photo vouchers of the coral host species **(a-l)** of **m)** *C. violacea* (Kiener, 1836) and **n)** *C. radula* (A. Adams, 1855). **a)** *Porites lobata* (Dana, 1846), Vav'varu, Maldives. **b)** *Porites solida* (Forskål, 1775), Hawai'i, USA. **c)** *P. annae* (Crossland, 1952), Aceh, Indonesia. **d)** *P. evermanni* (Vaughan, 1907), Hawai'i, USA. **e)** *P. attenuata* (Nemenzo, 1955), Bali, Indonesia. **f)** *P. compressa* (Dana, 1846), Hawai'i, USA. **g)** *P. rus* (Forskål, 1775), North Sulawesi, Indonesia. **h)** *Porites sp1*, Vav'varu, Maldives. **i)** *P. cylindrica* (Dana, 1846), Negros Oriental, Philippines. **j)** *P. nigrescens* (Dana, 1848), East Nusa Tenggara, Indonesia. **k)** *P. negrosensis* (Veron, 1990) Bali, Indonesia. **l)** *P. tuberculosis* (Veron, 2000) North Sulawesi, Indonesia. Photos and tissue samples were taken from each coral colony. Coral species were identified using Veron (2000), and Forsman et al. (2015) for *P. evermanni*. *Porites* sp1 did not match any known species descriptions.

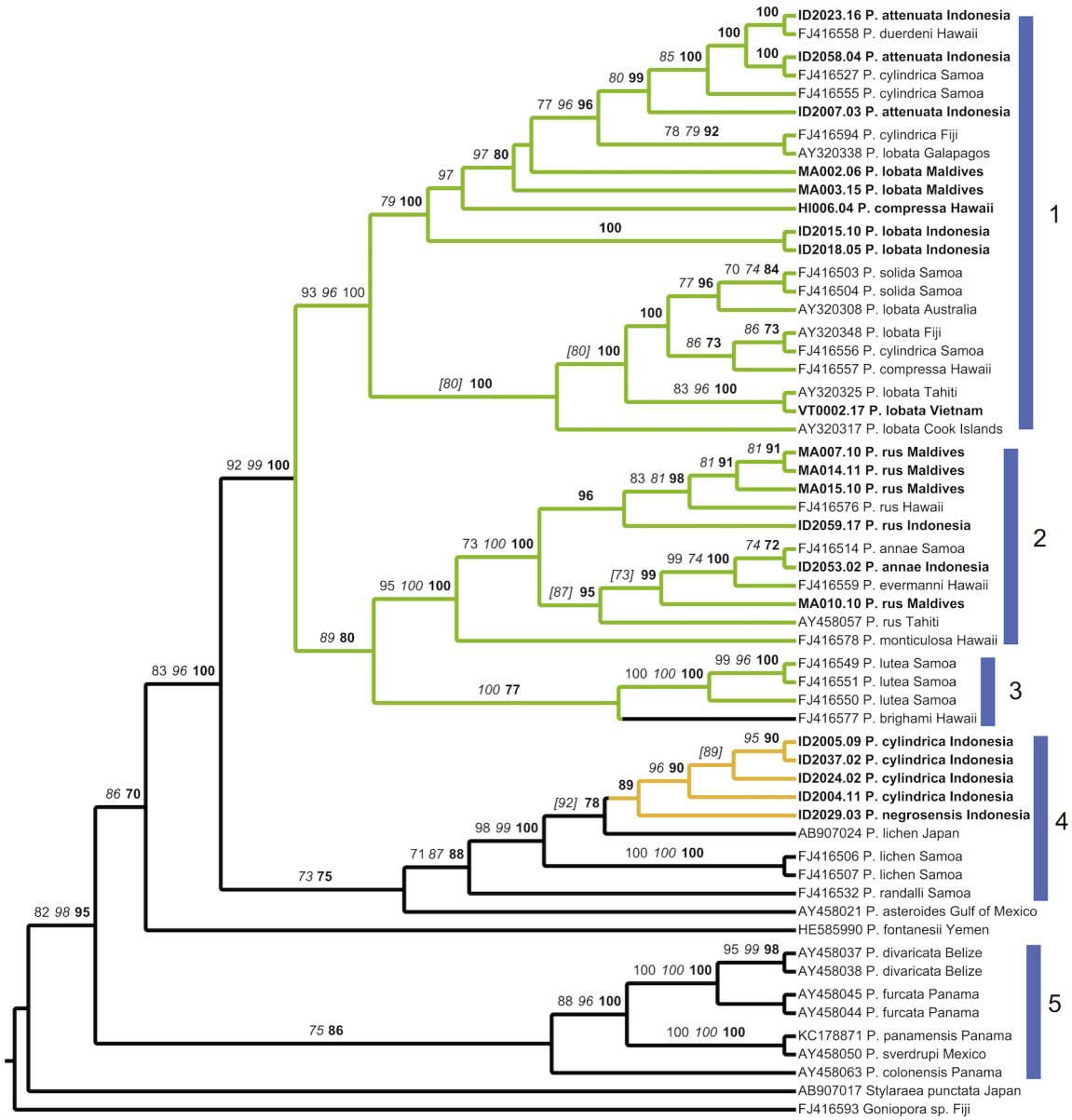


Figure 1.3 Phylogenetic tree of 57 ITS sequences of *Porites* spp. and outgroups with support values >70 (PhyML bootstrap proportion, MrBayes posterior probability = italics, and FastTree support values = bold). Bolded tip labels are samples of coral hosts of *Coralliophila radula* and *C. violacea* from this study. Tip labels are the sample code or GenBank accession number followed by the species name and collection location. Branches are colored by lineage (Green = 1, Gold = 2), only including species sampled in this study. Blue bars indicate various clades with strong support values.

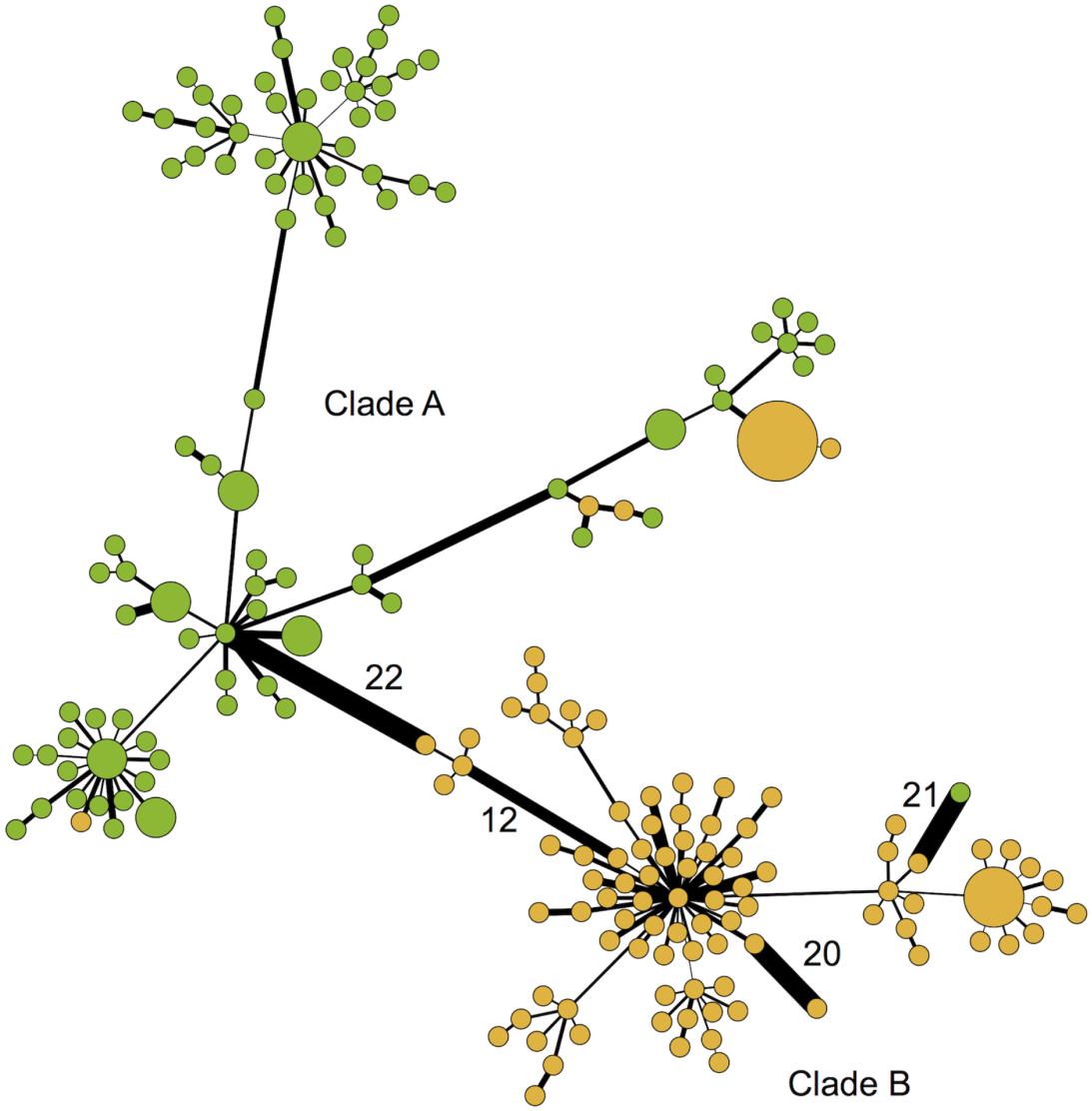
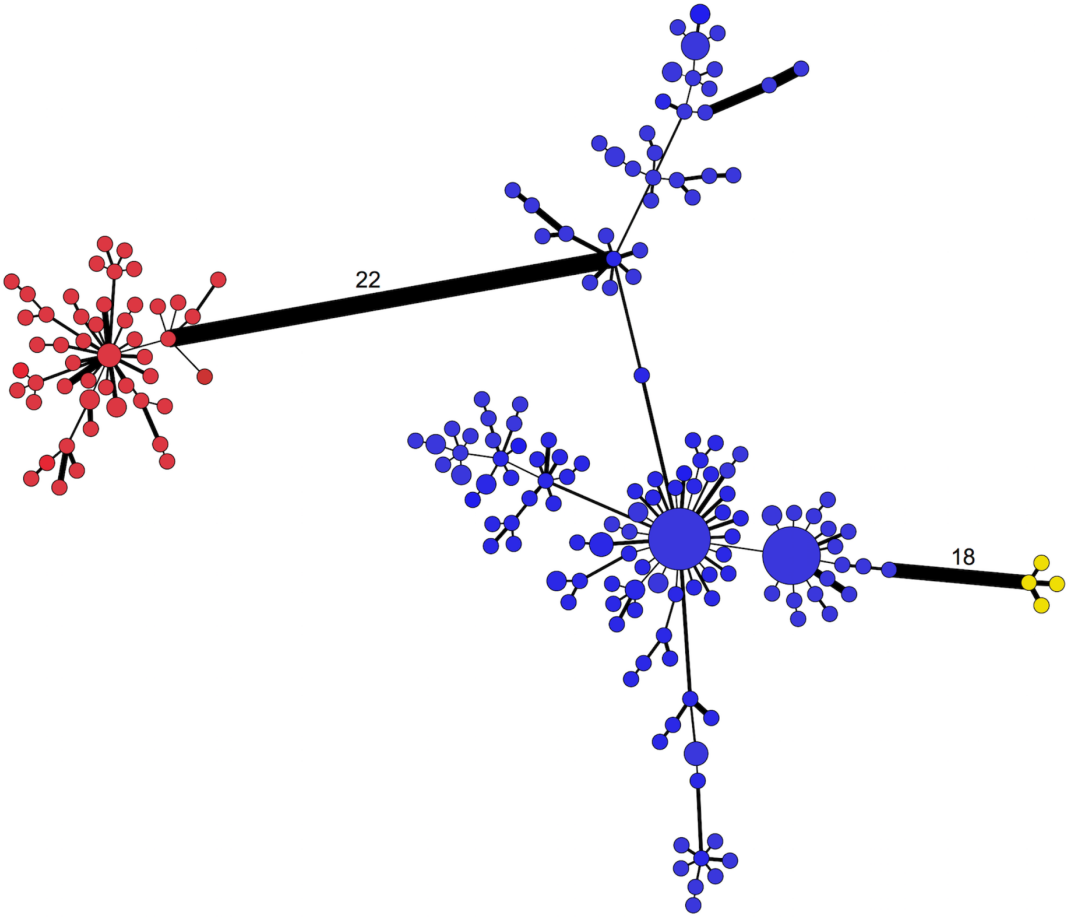


Figure 1.4 Sympatric coral host lineages (colors same as in Figure 1.3) plotted onto a minimum spanning tree of 188 haplotypes from 200 *Coralliophila violacea*. Circles are sized proportional to the frequency of haplotypes. Haplogroups separated by more than 20 steps are indicated with numbers.

a)



b)

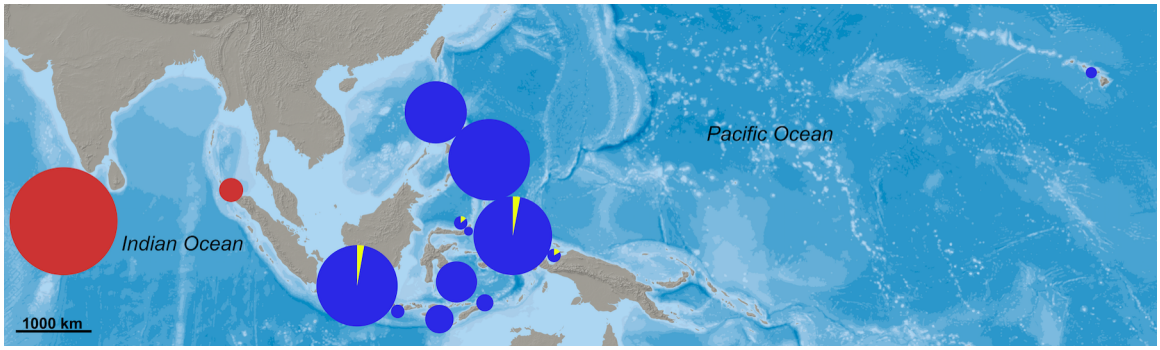


Figure 1.5 a) Minimum spanning tree of COI haplotypes of *Coralliophila radula.* The size of circles corresponds to haplotype frequency. Haplogroups with 18 or more mutational steps between them are colored. **b)** Map showing the geographic distribution of haplogroups. The size of the circles corresponds to the number of individuals sampled at each locality.

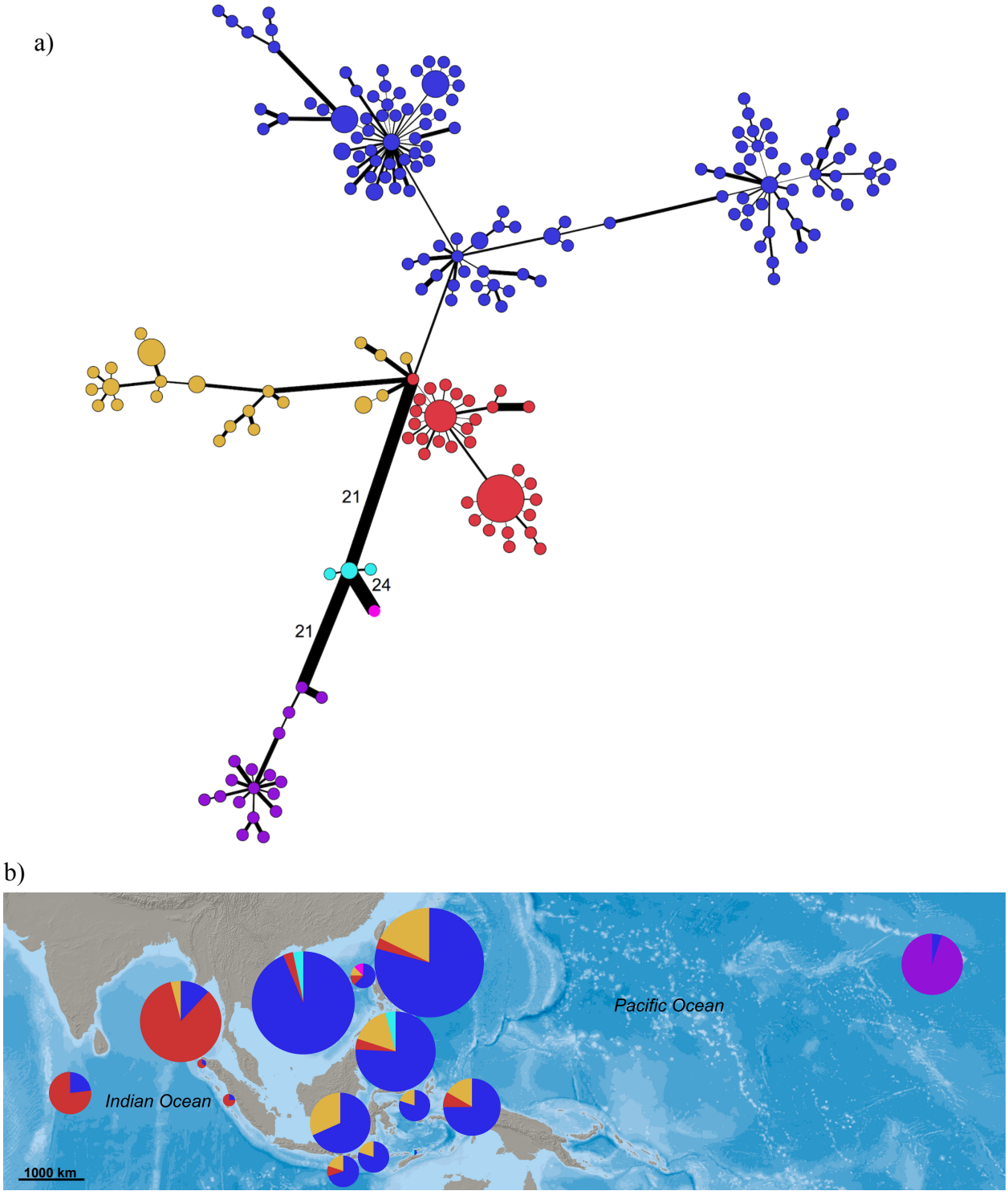


Figure 1.6 *Coralliophila violacea* collected from *Porites* lineage 1 only. **a)** Minimum spanning tree of 204 COI haplotypes from 234 snails. The size of circles corresponds to the number of individuals with that haplotype. Haplotypes are colored by groups, with 21 or more mutational steps between them, or groups of haplotypes dominating a geographic area. **b)** Map showing the geographic distribution of haplogroups. The size of the circles corresponds to the number of individuals sampled at each locality.

1.8 REFERENCES

Avise J.C. (2000) *Phylogeography: the history and formation of species.* Harvard University Press, Cambridge.

Barber P.H., Cheng S.H., Erdmann M.V., Tenggardjaja K., & Ambariyanto (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. *Phylogeography and population genetics in Crustacea. Crustean Issues* (ed. by C. Held, S. Koenemann, and C.D. Schubart), pp. 129–156. CRC Press, Boca Raton.

Barber P.H., Palumbi S.R., Erdmann M.V., & Moosa M.K. (2002) Sharp genetic breaks among populations of Haptosquilla pulchella (Stomatopoda) indicate limits to larval transport: patterns, causes, and consequences. *Molecular Ecology*, **11**, 659–674.

Barraclough T.G. & Vogler A.P. (2000) Detecting the geographical pattern of speciation from species-level phylogenies. *The American Naturalist*, **155**, 419–434.

Bastian M., Heymann S., & Jacomy M. (2009) Gephi: An open source software for exploring and manipulating networks.

Baums I.B., Boulay J.N., Polato N.R., & Hellberg M.E. (2012) No gene flow across the Eastern Pacific Barrier in the reef-building coral *Porites lobata*. *Molecular Ecology*, **21**, 5418–5433.

Baums I.B., Miller M.W., & Szmant M. (2003) Ecology of a corallivorous gastropod, *Coralliophila abbreviata*, on two scleractinian hosts. I: Population structure of snails and corals. *Marine Biology*, **142**, 1083–1091.

Bird C.E., Fernandez-Silva I., Skillings D.J., & Toonen R.J. (2012) Sympatric speciation in the post 'Modern Synthesis' era of evolutionary biology. *Evolutionary Biology*, **39**, 158–180.

Bird C.E., Karl S.A., Peter E.S., & Toonen R.J. (2011) Detecting and measuring genetic differentiation. *Crustacean Issues*, **19**, 39–81.

Bowen B.W., Rocha L.A., Toonen R.J., & Karl S.A. (2013) The origins of tropical marine biodiversity. *Trends in Ecology & Evolution*, **1666**, 1–8.

Briggs J.C. (1999) Modes of speciation: marine Indo-West Pacific. *Bulletin of Marine Science*, **65**, 645–656.

Briggs J.C. (2006) Proximate sources of marine biodiversity. *Journal of Biogeography*, **33**, 1–10.

Briggs J.C. & Bowen B.W. (2013) Marine shelf habitat: biogeography and evolution. *Journal of Biogeography*, **40**, 1–13.

Bush G.L. (1969) Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera, Tephritidae). *Evolution*, **23**, 237–251.

Carpenter K.E., Barber P.H., Crandall E.D., Ablan-Lagman M.C.A., Mahardika G.N., Manjaji-Matsumoto B.M., Juinio-Meñez M.A., Santos M.D., Starger C.J., & Toha A.H.A. (2011) Comparative phylogeography of the Coral Triangle and implications for marine management. *Journal of Marine Biology*, **2011**, 1–14.

Chen M.-H., Soong K., & Tsai M.-L. (2004) Host effect on size structure and timing of sex change in the coral-inhabiting snail *Coralliophila violacea*. *Marine Biology*, **144**, 287–293.

Connell J. (2012) Population ecology of reef-building corals. *Biology and Geology of Coral Reefs Volume II: Biology 1* (ed. by O. Jones and R. Endean), Academic Press, INC.

Crandall E.D., Frey M.A., Grosberg R.K., & Barber P.H. (2008) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology*, **17**, 611–626.

Crandall E.D., Sbrocco E.J., DeBoer T.S., Barber P.H., & Carpenter K.E. (2011) Expansion dating: calibrating molecular clocks in marine species from expansions onto the Sunda Shelf following the last glacial maximum. *Molecular Biology and Evolution*, **1**, 1–13.

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

DeBoer T.S., Subia M.D., Ambariyanto, Erdmann M.V., Kovitvongsa K., & Barber P.H. (2008) Phylogeography and limited genetic connectivity in the endangered boring giant clam across the Coral Triangle. *Conservation Biology*, **22**, 1255–1266.

Drès M. & Mallet J. (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B*, **357**, 471–492.

Duffy J.E. (1996) Resource-associated population subdivision in a symbiotic coral-reef shrimp. *Evolution*, **50**, 360–373.

Excoffier E. & Lischer H. (2011) ARLEQUIN: An integrated software package for population genetics data analysis. *What's New*, 1–174.

Faucci A., Toonen R.J., & Hadfield M.G. (2007) Host shift and speciation in a coral-feeding nudibranch. *Proceedings of the Royal Society of London. Series B*, **274**, 111–119.

Folmer O., Black M., Hoeh W., Lutz R., & Vrijenhoek R. (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fordyce J.A. (2010) Host shifts and evolutionary radiations of butterflies. *Proceedings of the Royal Society B*, **277**, 3735–43.

Forsman Z., Wellington G.M., Fox G.E., & Toonen R.J. (2015) Clues to unraveling the coral species problem: distinguishing species from geographic variation in *Porites* across the Pacific with molecular markers and microskeletal traits. *PeerJ*, **3**, e751.

Forsman Z.H., Barshis D.J., Hunter C.L., & Toonen R.J. (2009) Shape-shifting corals: molecular markers show morphology is evolutionarily plastic in *Porites*. *Evolutionary Biology*, **9**, 45–54.

Fujioka Y. & Yamazato K. (1983) Host selection of some Okinawan coral associated gastropods belonging to the genera *Drupella*, *Coralliophila* and *Quoyula*. *Galaxea*, **2**, 59–73.

Gaither M.R., Jones S.A., Kelley C., Newman S.J., Sorenson L., & Bowen B.W. (2011) High connectivity in the deepwater snapper *Pristipomoides filamentosus* (Lutjanidae) across the Indo-Pacific with isolation of the Hawaiian Archipelago. *PLoS ONE*, **6**, e28913.

Gaither M.R. & Rocha L.A. (2013) Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. *Journal of Biogeography*, **40**, 1638–1648.

Gittenberger A. & Gittenberger E. (2011) Cryptic, adaptive radiation of endoparasitic snails: sibling species of *Leptoconchus* (Gastropoda: Coralliophilidae) in corals. *Organisms Diversity & Evolution*, **11**, 21–41.

Hatfield T. & Schluter D. (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, **53**, 866–873.

Henry L.A. & Hart M. (2005) Regeneration from injury and resource allocation in sponges and corals - a review. *International Review of Hydrobiology*, **90**, 125–158.

Hurt C., Silliman K., Anker A., & Knowlton N. (2013) Ecological speciation in anemone-associated snapping shrimps (*Alpheus armatus* species complex). *Molecular Ecology*, **22**, 4532–4548.

Iacchei M., Gaither M.R., Bowen B.W., & Toonen R.J. (2016) Testing dispersal limits in the sea: range-wide phylogeography of the pronghorn spiny lobster *Panulirus penicillatus*. *Journal of Biogeography*, **43**, http://dx.doi.org/10.1111/jbi.12689.

Johnston L., Miller M.W., & Baums I.B. (2012) Assessment of host-associated genetic differentiation among phenotypically divergent populations of a coral-eating gastropod across the Caribbean. *PloS ONE*, **7**, e47630.

Kearse M., Moir R., Wilson A., Stones-Havas S., Cheung M., Sturrock S., Buxton S., Cooper A., Markowitz S., Duran C., Thierer T., Ashton B., Meintjes P., & Drummond A. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–9.

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

Kochzius M. & Nuryanto A. (2008) Strong genetic population structure in the boring giant clam, *Tridacna crocea*, across the Indo-Malay Archipelago: implications related to evolutionary processes and connectivity. *Molecular Ecology*, **17**, 3775–3787.

Kool J.T., Paris C.B., Barber P.H., & Cowen R.K. (2011) Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. *Global Ecology and Biogeography*, **20**, 695–706.

Krug P.J. (2011) Patterns of speciation in marine gastropods: a review of the phylogenetic evidence for localized radiations in the sea. *American Malacological Bulletin*, **29**, 169–186.

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

Li J. & O'Foighil D. (2012) Host-specific morphologies but no host races in the commensal bivalve *Neaeromya rugifera*. *Invertebrate Biology*, **131**, 197–203.

Litsios G., Sims C.A., Wüest R.O., Pearman P.B., Zimmermann N.E., & Salamin N. (2012) Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, **12**, 212.

Ludt W.B. & Rocha L.A. (2015) Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, **42**, 25–38.

Marko P.B. (2002) Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Molecular Biology and Evolution*, **19**, 2005–2021.

Matsubayashi K.W., Ohshima I., & Nosil P. (2010) Ecological speciation in phytophagous insects. *Entomologia Experimentalis et Applicata*, **134**, 1–27.

Miglietta M.P., Faucci A., & Santini F. (2011) Speciation in the sea: overview of the symposium and discussion of future directions. *Integrative and Comparative Biology*, 1–7.

Montano S., Maggioni D., Arrigoni R., Seveso D., Puce S., & Galli P. (2015) The hidden diversity of zanclea associated with scleractinians revealed by molecular data. *PLoS ONE*, **10**, 1–20.

Mullen L.M. & Hoekstra H.E. (2008) Natural selection along an environmental gradient: a classic cline in mouse pigmentation. *Evolution*, **62**, 1555–1570.

Munday P.L., Herwerden L.V., & Dudgeon C.L. (2004) Evidence for sympatric speciation by host shift in the sea. *Current Biology*, **14**, 1498–1504.

Nosil P., Gompert Z., Farkas T.E., Comeault A.A., Feder J.L., Buerkle C.A., & Parchman T.L. (2012) Genomic consequences of multiple speciation processes in a stick insect. *Proceedings of the Royal Society B*, **279**, 5058–5065.

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

Oliverio M., Barco A., Richter A., & Modica M. V (2009) The coralliophiline (Gastropoda: Muricidae) radiation: Repeated colonizations of the deep sea? *Nautilus*, **123**, 113–120.

Oren U., Brickner I., & Loya Y. (1998) Prudent sessile feeding by the corallivore snail *Coralliophila violacea* on coral energy sinks. *Proceedings of the Royal Society of London. Series B*, **265**, 2043–2050.

Palumbi S.R. (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547–572.

Peccoud J., Ollivier A., Plantegenest M., & Simon J.C. (2009) A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proceedings of the National Academy of Sciences*, **106**, 7495–7500.

Polato N.R., Concepcion G.T., Toonen R.J., & Baums I.B. (2010) Isolation by distance across the Hawaiian Archipelago in the reef-building coral *Porites lobata*. *Molecular Ecology*, **19**, 4661–4677.

Prunier R., Holsinger K.E., & Carlson J.E. (2012) The effect of historical legacy on adaptation: do closely related species respond to the environment in the same way? *Journal of Evolutionary Biology*, **25**, 1636–1649.

Puebla O. (2009) Ecological speciation in marine v. freshwater fishes. *Journal of Fish Biology*, **75**, 960–996.

Ritson-Williams R., Shjegstad S.M., & Paul V.J. (2009) Larval metamorphosis of *Phestilla* spp. in response to waterborne cues from corals. *Journal of Experimental Marine Biology and Ecology*, **375**, 84–88.

Rocha L.A. (2003) Patterns of distribution and preocess of speciation in Brazilian reef fishes. *Journal of Biogeography*, **30**, 1161–1171.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Hohna S., Larget B., Liu L., Suchard M.A., & Huelsenbeck J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Schluter D. & Conte G.L. (2009) Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, **106 Suppl**, 9955–9962.

Skillings D.J., Bird C.E., & Toonen R.J. (2011) Gateways to Hawai‘i: genetic population structure of the tropical sea cucumber *Holothuria atra*. *Journal of Marine Biology*, **2011**, 1–16.

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

Soong K. & Chen J.L. (1991) Population structure and sex-change in the coral-inhabiting snail *Coralliophila violacea* at Hsiao-Liuchiu, Taiwan. *Marine Biology*, **111**, 81–86.

Sotka E.E. (2005) Local adaptation in host use among marine invertebrates. *Ecology Letters*, **8**, 448–459.

Sotka E.E., Wares J.P., & Hay M.E. (2003) Geographic and genetic variation in feeding preference for chemically defended seaweeds. *Evolution*, **57**, 2262–2276.

Stöver B.C. & Müller K.F. (2010) TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC bioinformatics*, **11**, 7.

Thibert-Plante X. & Hendry A.P. (2010) When can ecological speciation be detected with neutral loci? *Molecular Ecology*, **19**, 2301–2314.

Tornabene L., Ahmadia G.N., Berumen M.L., Smith D.J., Jompa J., & Pezold F. (2013) Evolution of microhabitat association and morphology in a diverse group of cryptobenthic coral reef fishes (Teleostei: Gobiidae: *Eviota*). *Molecular Phylogenetics and Evolution*, **66**, 391–400.

Tsang L.M., Chan B.K.K., Shih F.-L., Chu K.H., & Allen Chen C. (2009) Host-associated speciation in the coral barnacle *Wanella milleporae* (Cirripedia: Pyrgomatidae) inhabiting the *Millepora* coral. *Molecular Ecology*, **18**, 1463–1475.

Villesen P. (2007) FaBox: An online toolbox for FASTA sequences. *Molecular Ecology Notes*, **7**, 965–968.

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

Waldrop E., Hobbs J.-P.A., Randall J.E., DiBattista J.D., Rocha L.A., Kosaki R.K., Berumen M.L., & Bowen B.W. (2016) Phylogeography, population structure and evolution of coral-eating butterflyfishes (Family Chaetodontidae, genus *Chaetodon*, subgenus *Corallochaetodon*). *Journal of Biogeography*, **43**, 1–14.

Walsh P.S., Metzger D.A., & Higuchi R. (2013) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, **54**, 506–513.

Wang C.Y., Liu H.Y., Shao C.L., Wang Y.N., Li L., & Guan H.S. (2008) Research progress on chemical defensive substances from soft corals and gorgonians. *Acta Ecologica Sinica*, **28**, 2320–2328.

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

Yamashiro H., Oku H., Higa H., Chinen I., & Sakai K. (1999) Composition of lipids, fatty acids and sterols in Okinawan corals. *Comparative Biochemistry and Physiology Part B*, **122**, 397–407.

CHAPTER 2:

DIVERGENCE WITH GENE FLOW BETWEEN HOST-RACES IN A CORALLIVOROUS SNAIL

2.1 ABSTRACT

Our study of mtDNA in the corallivorous snail *Coralliophila violacea* presented in Chapter 1, showed that sympatric populations living on different coral hosts are genetically distinct. As a result, we hypothesized that the observed divergence was occurring with gene flow via directional selection on different hosts. To test this hypothesis, we generated 2,718 single nucleotide polymorphism (SNP) genotypes, from 51 snails collected from two *Porites* coral species, using type II restriction-site associated DNA sequencing (2b-RAD). Our results show evidence of unidirectional migration of snails from *Porites lobata* to *P. cylindrica*, as well as hybridization between the host-associated populations, supporting the hypothesis of divergence with gene flow. Outlier tests detected loci under divergent selection between snails collected from different corals. We identified 74 loci as outliers under directional selection, including a nuclear hormone receptor gene (HR96) involved in the control of xenobiotic detoxification pathway gene expression, possibly allowing snails to neutralize coral-specific toxins. Ecological divergence with gene flow, among sympatric populations of snails on different coral hosts, contributes to lineage diversification in the mega-diverse Coral Triangle.

2.2 INTRODUCTION

Since the Modern Synthesis, evolutionary biologists have largely viewed allopatry as the principal mode of speciation (Mayr, 1963; Bird et al., 2012), while alternatives such as speciation in sympatry were regarded as exceptionally rare and extremely controversial (Coyne & Or 2004). However, recent studies are increasingly demonstrating that sympatric speciation via divergent selection, a process known as *ecological speciation*, is more common than previously thought (Rundle & Nosil, 2005; Mallet et al., 2009; Schluter, 2009; Bowen et al., 2013).

Ecological speciation is the evolution of reproductive isolation among populations by natural selection acting in opposing directions in different ecological niches or environments (Schluter, 2001; Rundle & Nosil, 2005), rather than by physical isolation. Here, selection is driven by individuals interacting with their environment, or with other organisms via predation, competition, or symbiosis. Researchers have documented evidence for ecological speciation across a wide variety of terrestrial organisms (for review see Schluter, 2009) including monkey flowers (Case & Willis, 2008), indigobirds (Sorenson et al., 2003), and anoles (Thorpe et al., 2010), as well as aquatic animals such as sticklebacks (Hatfield & Schluter, 1999), cichlids (Terai et al., 2006), mosquitofish (Langerhans & Gifford, 2009) and killer whales (Foote et al., 2009). However, ecological speciation is most commonly observed in herbivorous insects (Matsubayashi et al., 2011), largely because of the prevalence of strong associations between insects and the plants upon which they feed.

Because of the direct impacts on growth and fitness, parasite-host relationships are among the strongest interspecific associations in nature, and it is believed that these

associations provide the opportunity for rapid divergence (Drès & Mallet, 2002; Sotka, 2005). In particular, divergence can be driven by host-expansion, where a new host is added to the list of existing host species, and/or host-switching, where an old host is abandoned for a new host species. If assortative mating occurs within populations utilizing separate hosts, then divergent selection and reproductive isolation can occur. As such, host-shifting is an established mechanism for ecological speciation (Drès & Mallet, 2002).

While well-documented in terrestrial and freshwater ecosystems, the importance of ecological speciation in marine ecosystems is as yet unclear. However, there are good reasons to believe that ecological speciation could be common in the sea. First, absolute physical barriers are exceedingly rare (Rocha et al., 2005; Rocha & Bowen, 2008; Ludt & Rocha, 2015). As a result, speciation must proceed with varying levels of gene flow, and this process may be aided by divergent selection (Palumbi, 1994). Second, the strong interspecific interactions believed to promote ecological speciation in terrestrial species (e.g. host-parasite, mutualisms), are extremely common in certain marine ecosystems. For example, reef-building corals have a strong symbiotic relationship with endosymbiotic zooxanthellae that live within coral polyp tissues. Tight ecological associations have also been documented between corals and a wide variety of invertebrate taxa (Zann, 2002), including ~870 known species of sponges, copepods, barnacles, crabs, shrimp, worms, bivalves, nudibranchs, and snails (reviewed by Stella et al., 2011). These symbiotic relationships create the potential for host shifting and the development of host races that could culminate in sympatric speciation. Indeed, recent work using traditional genetic methods shows the potential for ecological speciation in a

variety of marine taxa (for review see Miglietta et al., 2011; Bowen et al., 2013), including amphipods on macroalgae (Sotka, 2005), coral-dwelling barnacles (Tsang et al., 2009), coral-eating nudibranchs (Faucci et al., 2007), parasitic snails (Reijnen et al., 2010; Gittenberger & Gittenberger, 2011), and coral-dwelling gobies (Munday et al., 2004). However, studies that explicitly look for *genomic* signatures of ecological divergence in sympatric populations have yet to be conducted on parasitic marine taxa.

Defined by the presence of over 500 species of reef-building corals (Veron et al., 2011), the Coral Triangle is the most biodiverse marine environment on Earth (Cowman & Bellwood, 2011). Spanning the waters of Indonesia, the Philippines, Timor Leste, Malaysia, Papua New Guinea and the Solomon Islands, and covering around 50,000 km^2, an area larger than the Great Barrier Reef, these coral reefs are home to thousands of unique species of corals, fishes and other creatures, many of which are unknown to science (Briggs, 2003). Extensive research exists on the evolution of this biodiversity hotspot, but most of this work has focused exclusively on identifying and understanding mechanisms of allopatric divergence (see Barber et al., 2011; Carpenter et al., 2011 for reviews). The enormous amount of marine biodiversity in this region, combined with the prevalence of strong species-species interactions on coral reefs, makes it likely that ecological speciation is also occurring.

In Chapter 1, we demonstrated that allopatric divergence is occurring among populations of the corallivorous snail, *Coralliophila violacea,* spanning the Sunda Shelf (Fig. 2.1). This pattern is commonly observed in marine taxa such as giant clams (DeBoer et al., 2014), reef fish (Ackiss et al., 2013) and seahorses (Lourie et al., 2005), and is typically attributed to low sea level stands during the Pleistocene that exposed the

continental shelf, although ocean currents may also cause, or reinforce, isolation (Kool et al., 2011). Physical isolation is predicted to result in the divergence of neutral markers, as genetic drift and natural selection act independently in each population in the absence of gene flow. However, such neutral processes cannot explain the striking genetic divergence we found within sympatric populations of *C. violacea*—snails living on different coral host species on the same reef seen in Chapter 1. Strong preferences for coral host in adults (unpublished data S. Simmonds), combined with clear patterns of assortative genetic variation, suggest that selective processes must be involved, and divergent selection leaves clear signatures within the genome. This conclusion is reinforced by the absence of genetic structure between *C. violacea* collected from the same species of coral from geographically extensive reefs east of the Sunda Shelf, namely the hyperdiverse Coral Triangle region (Chapter 1).

Genetic drift and migration should have approximately equal effects on all parts of the genome that are selectively neutral (Nielsen, 2005). However, natural selection is expected to affect non-neutral parts of the genome, as well as hitchhiker loci, to a greater extent (Smith & Haigh, 1974). Neutral loci should, therefore, show similar levels of differentiation between populations on different hosts (Via, 2009). In contrast, frequencies of loci under selection (outlier loci) or linked loci should either be unusually high, or unusually low, in host-associated populations, depending on the type of selection occurring (Beaumont & Nichols, 1996). If divergent selection has taken place, then the allele being selected for should increase in one population compared to the other.

The goal of this study is to examine the potential role of ecological divergence in generating biodiversity in the megadiverse Coral Triangle. Specifically, we use single

nucleotide polymorphisms (SNPs) from thousands of loci across the genome of *C. violacea* to: 1) determine the direction and amount of gene flow between sympatric populations of host-associated snail populations, 2) identify outlier loci under putative selection between hosts, and 3) annotate possible functions of linked genes that might be necessary for adaptation to each host.

2.3 MATERIALS AND METHODS

2.3.1 Collection of snail samples

In Chapter 1 we showed that two major genetic groups of *C. violacea* occur sympatrically on the same reefs, but on two separate lineages of *Porites* corals, each consisting of a suite of different species. We subsampled snails from six sympatric populations of coral host lineages spanning the Coral Triangle (Table 1, Fig 2.1) to: 1) test whether the divergence observed in mitochondrial sequences of *Coralliophila violacea* was present in loci across the genome; and 2) to determine if migration and gene flow was occurring between "ecotypes". For the current study, we chose snails from the most abundant coral species in each lineage to maximize the number of samples, but reduce potentially confounding effects of differences among host within the same lineage.

2.3.2 Creation of RAD tag libraries

To obtain DNA for SNP analyses, we extracted genomic DNA from 20 mg of foot tissue using a Qiagen DNeasy Blood and Tissue Kit (Qiagen) according to manufacturer's instructions, except in the last step when we cleaned DNA eluted with 100 µl of molecular grade H_2O rather than AE buffer. We estimated initial DNA concentrations

using a NanoDrop and visualized DNA quality on a 1% agarose gel stained with SYBR Green. We used only high-quality DNA with a bright high molecular weight band and minimal smearing. We dried DNA extractions using a SpeedVac on medium heat and reconstituted using molecular grade H_2O to a final uniform 250 ng/µl DNA concentration.

To create reduced representation libraries to survey SNP variation, we prepared 2b-RAD libraries following published protocols (Wang et al., 2012) as updated by Dr. Eli Meyer (http://people.oregonstate.edu/~meyere/tools.html). AlfI restriction enzyme digest reduced representation (1/16th) libraries were labeled with individual barcodes and subjected to 18–20 PCR amplification cycles. Products were electrophoresed on a 2% agarose gel in 1× TBE buffer and run at 150 V for 90 minutes. Target bands (165-bp) were visualized with SYBR SAFE dye and excised from the gel. The excised band was then purified using a QIAquick gel purification kit (Qiagen). A final cleaning step used Ampure XP beads (Beckman-Coulter).

To obtain DNA sequences, prepared libraries were sent to the QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC, Berkeley for quality checks (qPCR, BioAnalyzer) and sequencing. We multiplexed 10–20 snails per lane in 5 lanes of a 50 bp single end run on the Illumina HiSeq2000 platform.

2.3.3 RADseq data processing

All raw reads were truncated to the insert size (36 bp), filtered for quality (PHRED scores >20), and empty constructs discarded to prepare raw sequence data for SNP identification. To process the data, we used custom scripts available on GitHub https://github.com/z0on/2bRAD_denovo. STACKS is a program commonly used to

process RADseq data but it cannot be used with 2b-RAD sequences because they have double strands and STACKS considers reverse-complements as separate loci. Therefore, we used a pipeline to emulate the steps of STACKS, while taking advantage of the fact that 2b-RAD sequences both strands. In a step similar to making "stacks", unique tag sequences (minimum sequencing depth 5×) were counted, and the number in reverse-complement orientation were recorded. These stacks were merged into one table. Then all sequences were clustered in CD-HIT using a 91% similarity threshold. We defined the most abundant sequence in the cluster as the reference, and then filtered a locus-annotated table from the previous two steps, excluding reads below 5× depth or exhibiting strand bias. The orientation of the resulting clustered sequences was flipped to match the most abundant tag in a cluster.

To call genotypes (as population-wide RAD-tag haplotypes), we applied mild allele filters (10× total depth, allele bias and strand bias), with the additional requirement that alleles appear in at least two individuals. We then applied locus filters allowing maximum 50% heterozygotes at a locus, no more than two alleles, genotyped in 30% of samples and polymorphic. Finally, we removed loci with too many heterozygotes (75%) and missing genotypes (70%). The final set of SNPs was then thinned to one per tag (the one with the highest minor allele frequency) for F_{ST} and STRUCTURE analysis to remove linked loci.

2.3.4 Individual sample filtering steps

We filtered out individuals ($N = 11$) with low genotyping rates, indicating low DNA quality, by taking the $\log_{10}$ of the number of sites genotyped per individual, and removing any individuals that were outside one standard deviation (SD) of the mean. We also

removed individuals ($N = 5$) with high homozygosity (+/- 2 SD of mean F inbreeding coefficient) indicating potential contamination. The remaining 51 individuals were used in analyses of population genetic structure. The final data file was in VCF format and converted to other formats using PGDSpider v2.0.8.0 (Lischer & Excoffier, 2012).

2.3.5 Genetic structure

To test whether the patterns observed in a mitochondrial locus were present in loci genome-wide, we inferred the population genetic structure of the full RADseq dataset of 2,718 loci and 51 individuals using two methods. First, we ran the Bayesian model-based clustering method STRUCTURE (Pritchard et al., 2000) using a burn-in period of 20,000 followed by 50,000 MCMC replicates for $K = 1–12$, and 10 runs for each K. We used the admixture model, with allele frequencies correlated among populations. The results from STRUCTURE were analyzed in CLUMPAK v1.1 (Kopelman et al., 2015) to select for the best K and display the results graphically. Then, using the best K in CLUMPAK to estimate gene flow between populations, we identified migrants (movement of individuals between hosts) and admixed individuals (gene flow) between host-associated populations in STRUCTURE.

2.3.6 Outlier analyses

To test for evidence of natural selection in relation to coral host, we compared SNPs between populations of snails on different hosts, pooled across six localities, with two datasets: 1) including all individuals, and 2) excluding migrants and admixed individuals that we identified using STRUCTURE. First, we performed an outlier loci analysis using BayeScan v2.1 (Foll & Gaggiotti, 2008) with a burn-in of 50,000, a thinning interval 10,

a sample size 5,000, 100,000 iterations, and 20 pilot runs of 5,000 each. We examined two false discovery rates (0.10 and 0.05). We then used a second method to detect loci under selection (FDIST2) as implemented in ARLEQUIN (Excoffier & Lischer, 2011). We ran 100 demes per group and 50 groups for 50,000 simulations. This model compares a simulated neutral distribution of F_{ST} to the observed distribution and identifies outliers. Loci with significant F_{ST} *p*-values (<0.01) were considered to be under selection (Excoffier & Lischer, 2010).

2.3.7 Candidate gene identification and annotation

To annotate the functions of genes linked to outlier loci, we aligned sequences containing SNP outlier loci to nucleotide collections (nr/nt) available on the NCBI website, using the BLASTN algorithm at two different taxonomic levels: 1) Mollusca (taxid:6447), and 2) Lophotrochozoa (taxid:1206795). We adjusted parameters (expected threshold 10, word size 7, no low complexity filter, no mask for look up table) to accommodate short read sequences. We only examined hits with a high query coverage (>80%) and identified and annotated any associated genes using NCBI.

2.4 RESULTS

After removing empty constructs and filtering for quality, the average number of unique reads per individual was 5,710,091 at minimum 5× depth. We sequenced and genotyped 17,676 high-quality RADseq loci with ≥25× coverage, in 67 individuals, from two different coral host species, at six locations across the Coral Triangle. This dataset was then filtered for 30% maximum missing data per locus, leaving 5,999 loci, and then thinned to one SNP per loci to

remove any physically linked SNPs for STRUCTURE and F_{ST} analyses, leaving 2,718 SNPs. Next, we removed 16 individuals that had either low DNA quality (missing data ≥ +1SD from the mean) or potential contamination issues (inbreeding coefficient ≥ +2SD from the mean), leaving 51 individuals.

2.4.1 Genetic structure

Genetic differentiation between sympatric snail populations on different coral hosts was moderate across all loci (Fig. 2.2, mean $F_{ST} = 0.047$, weighted $F_{ST} = 0.090$, Weir & Cockerham, 1984). Using the full dataset of SNPs, we observed significant population structure between host-associated populations of snails (Fig. 2.3). CLUMPAK analysis of the STRUCTURE results indicated $K = 2$ as the best K value (Fig 2.4). At $K = 2$, the majority (88%) of all snails grouped by their coral host. Grouping by host was stronger in snails collected from *P. lobata* (97%) than from *P. cylindrica* (79%). Higher *K*s were not more informative in relation to geography. Neutral loci (outlier loci identified in BayeScan removed, FDR = 0.10), and outlier loci, show identical patterns of population structure in STRUCTURE as the full dataset of SNPs (Fig 2.3).

2.4.2 Migration and admixture

We inferred the ancestry of individuals in STRUCTURE, using their host as a prior. This analysis revealed some migration and admixing between sympatric populations of *C. violacea,* despite residing on different hosts (*Porites lobata* and *P. cylindrica*). However, migration rates varied among snails living on the two different hosts. Migration rates also varied among geographically distinct populations of snails occupying the same host.

The direction of migration between snails living on the different coral hosts (*P. cylindrica* and *P. lobata*) was strongly asymmetric. In total, 19% of the snails collected from *P. cylindrica* had *P. lobata* genetic ancestry, while no snails with *P. cylindrica* ancestry were ever found on *P. lobata* (Table 2.3). Despite this strong asymmetry, evidence of migration from one host to the other was only found in three of the six sampled locations (Table 2.3). Although the two lineages of *C. violacea* are diverging on different hosts, the presence of snails with *P. lobata* genotypes living on a *P. cylindrica* hosts indicates the potential for ongoing gene flow; indeed, individuals of admixed ancestry were found. The percentage of admixed snails, and thus gene flow, was equal across coral hosts (8% of snails were admixed individuals; Table 2.3). Admixed individuals were only found at locations where migration was also observed (Dumaguete and Pulau Mengyatan; Table 2.3). After excluding migrants and admixed individuals, the mean F_{ST} across all loci increased from 0.047 to 0.075 and the weighted F_{ST} from 0.090 to 0.150.

2.4.3 Host-specific directional selection

Using STRUCTURE, we identified 9/51 individuals that were either migrants from one coral host population to the other, or of admixed ancestry (Table 2.2). In analyses for detecting host-specific selection, we used two different datasets: 1) including all individuals; and 2) excluding migrants and admixed individuals. We searched for loci under selection using two methods, and with two sets of individuals.

Our first method involved a Bayesian model, BayeScan (Foll & Gaggiotti, 2008). Using the default false discovery rate (FDR) of 10%, we identified six loci as outliers (pairwise F_{ST} = 0.255–0.354) in the dataset with all snails (Fig. 2.5a). After excluding all

admixed and migrant individuals, the number of outlier loci only increased to eight (pairwise F_{ST} = 0.385–0.526; Fig. 2.5b). All outlier loci had positive alpha values, indicating they are under directional selection between snails on different coral hosts.

In the second method, FDIST2, we used the infinite island model of migration to identify 51 outlier loci (pairwise F_{ST} = 0.177–0.729; mean F_{ST} = 0.492; Fig. 2.6a) in the dataset with all snails. After removing migrants and admixed individuals, the number of outliers increased to 65 with higher F_{ST} values (pairwise F_{ST} = 0.320–0.925; mean F_{ST} = 0.620; Fig. 2.6b). A total of 43/73 outlier loci were shared between the two datasets; 8 were unique to the all-individual dataset, and 22 were unique to the dataset that excluded migrants and admixed individuals.

2.4.4 Mapping and annotation of outlier loci

The majority (55%) of outlier loci we identified did not successfully align to any other mollusc or lophotrochozoa nucleotides currently available in the NCBI database (Table 2.4). However, we were able to align the remaining 33/73 (45%) of outlier loci to DNA sequences from a variety of molluscs and a diversity of other taxa, including four marine snails, a freshwater snail, the California sea hare, the giant owl limpet, an octopus, a brachiopod and a leech (Table 2.4). Of these loci, three mapped to a non-coding region, and 13 to genes of unknown function. The remaining 17 loci mapped to genes regions with predicted functions. The two most common gene ontology function terms were protein binding and metal ion binding but annotated genes had various putative functions, including one (tag 28347, HR96 gene) for xenobiotic detoxification (Lindblom & Dodd, 2006; Richter & Fidler, 2014; Table 2.4). At this tag, there were two alleles, in almost equal frequency (43%, 57%) in *P. lobata*-associated populations of snails, and nearly fixed (97%) for one allele in *P. cylindrica*-associated populations of snails.

2.5 DISCUSSION

In this study, we obtained genome-wide data (2,718 SNPs) from six sympatric populations of *C. violacea* living on different coral hosts in the Coral Triangle. Model-based clustering revealed two clearly differentiated clusters that were largely concordant with coral host. This result is consistent with what we found in mitochondrial DNA analyses (Chapter 1). However, the estimate of F_{ST} we found in our genome-wide analysis (0.047) was considerably lower than in our mtDNA analysis ($\Phi_{CT} = 0.561$ v. $F_{ST} = 0.047$). This lower value suggests intermediate gene flow between the distinct host-races (Nm>10). $F_{ST} = 0.047$ is also similar in magnitude to other cases of sympatric host-associated divergence, for example in pea aphids (SNPs, $F_{ST} = 0.062$; Smadja et al., 2012). However, while lower F_{ST} values certainly indicate considerable gene flow, it can be difficult to determine whether this gene flow is ongoing, or historical because they are a per-generation estimate of genetic exchange (Hedgecock et al., 2007). By using STRUCTURE, we have been able to shed some light on this issue. Our results clearly indicate the presence of hybrid genotypes, and we also see evidence for the unidirectional migration of snails from *Porites lobata* to P. *cylindrica*. Taken together, these findings show that the divergence among host-races of *C. violacea* is occurring in the face of *ongoing* gene flow. Strong natural selection must, therefore, be contributing to the partitioning of *C. violacea* races by coral host.

2.5.1 Divergence with gene flow

In parasites, such as *C. violacea*, divergence with gene flow likely happens through two main mechanisms of premating isolation (Nosil et al., 2005). First is host preference, where adults prefer to lay eggs on, or larvae/juveniles prefer to recruit to,

their natal host. Mating then takes place solely on that host. Second is host adaptation, where selection acts against immigrants from another host via immigrant inviability (Nosil et al., 2005).

Our study suggests that both mechanisms may be occurring in *C. violacea*. For instance, although the majority of snails partitioned into two genetically distinct groups living on different coral hosts, genome-wide SNP data also identified several migrants and individuals of admixed ancestry. Interestingly, all migrants were individuals that genetically sorted with *P. lobata*, but were living on *P. cylindrica*. Additionally, only higher-order hybrids (e.g. F_2, F_3) were observed on *P. lobata*, suggesting that gene flow and hybridization between host-races is unidirectional. This pattern could result from two possible scenarios. First, larvae from *P. cylindrica* snails prefer their natal host over *P. lobata*, or do not respond to chemical settlement cues from *P. lobata*. And the fact that there are twice as many coral species ($N = 8$) in *Porites* lineage 1 (to which *P. lobata* belongs) than in *Porites* lineage 2 (to which *P. cylindrica* belongs), suggests that snails living on *Porites* lineage 1 corals may be less specific in terms of their settlement cues, resulting in the occasional dis-assortative larval settlement. An alternative, but not mutually exclusive explanation, is that larvae from *P. cylindrica* snails may settle on *P. lobata,* but are less likely to survive and reproduce leading to immigrant inviablility (Nosil et al., 2005). Outlier loci are genes whose frequencies can only be explained by selective forces. Genes that are beneficial to snails living on *P. cylindrica* are likely less helpful on *P. lobata*. If this is the case, we should see some indications of a selective sweep in the derived population from the standing genetic variation of the ancestral population. Indeed, we observed some outlier loci (e.g. HR96, detoxification gene) that

were in equal proportions in *P. lobata*, but were at near fixation in *P. cylindrica* (97%), indicating a selective sweep at that locus for a single variant.

Still, regardless of whether the misalignment of snails and coral hosts results from pre- or post-recruitment processes, the fact that the vast majority of snails sort by host coral in the face of hybridization and gene flow indicates that natural selection must be relatively strong to counteract gene flow of Nm>10 (Funk et al., 2011). Moreover, the high fidelity of the snails occupying *P. cylindrica* and lower fidelity of snails occupying *P. lobata*, combined with selective sweeps in *P. cylindrica,* suggest that snails parasitizing *P. lobata* are the ancestral population. This conjecture is consistent with the observation that specialist species often evolve from generalist ancestors (Nosil, 2002), likely because specialization constrains further evolution by reducing genetic variation (Moran, 1988). If it is generally true that specialists evolve from generalists (Kawecki, 1996, 1998), then host specialization could be an major mechanism of divergence within the Coral Triangle (Briggs, 2005); increased diversity should raise niche partitioning, leading to more opportunities for host specialization (Janz et al., 2006).

2.5.2 Candidate genes involved in adaptation to host

Outlier loci provide insights into the targets of natural selection (Storz, 2005), and are a useful starting point for determining how selection may be acting on populations diverging on different hosts. Our analysis revealed 74 putative gene regions with F_{ST} values significantly higher than neutral expectations, suggesting that they are likely under selection and could be involved in adaptation to coral hosts or linked to such genes via hitchhiking. Due to a lack of genomic resources for *C. violacea*, we were unable to identify the majority of these 74 gene regions. Only 17 mapped to a gene region with a

predicted function. There is no *a priori* information on the types of genes involved in molluscs adaption to different hosts. However, a useful comparison can be found in ectoparasitic phloem-feeding insects adapting to different host plants (Oren et al., 1998). Gene categories under selection in insect-plant interactions include those involved in sensing hosts (Simon et al., 2015), those that protect insects against plant defenses and facilitate feeding, and those that code for digestive and detoxifying enzymes to neutralize plant toxins (e.g. metal ion binding) (Simon et al., 2015). Recent experimental evidence suggests genes with metal ion binding functions are repeatedly under selection in stick insects adapting to different host plants (Soria-Carrasco et al., 2014). Indeed, four of the *C. violacea* candidate genes we identified in outlier tests are involved in metal ion binding (KTM2D, ODHGY, GLX-1, SMAP 1). Very little is known about how corals and their algal symbionts chemically defend themselves against or react to parasites and predators. Corals or their algal symbionts (*Symbiodinium* spp.) could potentially produce metabolites or toxins. In fact, *Symbiodinium*-specific toxins called Zooxanthellatoxins have been characterized (Gordon & Leggat, 2010), but it is not known whether these toxins are upregulated in response to parasites or predators. Given that insects diverging on different hosts show evidence of selection on metal ion binding genes involved in detoxification, a similar process may be occurring in *C. violacea*. If true, this would be the first evidence that similar processes shape insect and mollusc host-parasite associations.

Additional evidence for detoxification playing a role in host divergence comes from another gene region, HR96. This gene is a nuclear hormone receptor, involved in the control of xenobiotic detoxification pathway gene expression (Richter & Fidler, 2014)

and might allow snails to neutralize coral-specific toxins. It is possible that control of detoxification pathway gene expression adapts in response to different chemicals that snails are exposed to in their diet. Interestingly, HR96 was nearly at fixation in *P. cylindrica* (97%), indicating a selective sweep at that locus for a single variant. This result combined with the four metal ion binding gene regions indicates that there must be strong differences in detoxification in the different host-races of *C. violacea*. Therefore, it is reasonable to hypothesize that host specificity may be driven by adaptation to host-specific toxins. Mismatches between snail metabolic abilities and coral hosts could explain the strong asymmetry in snails being found on an atypical coral host.

The other six predicted genes we identified by outlier tests are involved in protein binding activity (KTM2D, AAK1, TRIP11, ADGRL, BBS9, PCM1). It is unclear why selection may be driving diversification of these genes in snails from different coral hosts. Future work would benefit from a fully annotated genome of *Coralliophila* to map to as a reference to help us understand this. A full genome would let us to examine the genomic architecture of divergence with gene flow and quantitative trait loci; allowing us to better pinpoint regions of the genome under selection, and the specific functions of genes involved in adapting to different hosts.

2.6 CONCLUSIONS

John Briggs originally proposed the idea that sympatric speciation is an important mechanism contributing to the Coral Triangle's biodiversity, as well as the successful export of species formed under intense competition within the region (Briggs, 1999, 2005). To support his hypothesis he pointed to multiple cases of sympatric sibling species

with distributions centered on the Coral Triangle, where the older of the two species has a wide range, while the younger has a much more restricted range limited to the Coral Triangle (Briggs, 1999). Our study provides the first genomic evidence to support his assertion that ecological divergence with gene flow could be promoting sympatric speciation and generating of biodiversity in the Coral Triangle. In addition, the spatial pattern of *C. violacea* sympatric host-races also matches the pattern Briggs described, with the ancestral *P. lobata* host-race having a broad geographic distribution, and the derived *P. cylindrica* host-race restricted to the Coral Triangle. Future research on speciation in the ocean should include the study of sympatric sibling species or ecotypes that occupy different niches, most of which are undiscovered cryptic biodiversity.

2.7 TABLES AND FIGURES

Table 2.1 *Coralliophila violacea* collection localities, coral host and number of samples.

	Coral host	
Locality	*P. lobata*	*P. cylindrica*
1. Pemuteran	-	7
2. Nusa Penida	11	9
3. Pulau Mengyatan	5	3
4. Lembeh	7	1
5. Bunaken	8	6
6. Dumaguete	2	8
Total *N*	**33**	**34**

Table 2.2 *Coralliophila violacea*. Full RADseq dataset 2,718 loci all individuals, split out by coral host. STRUCTURE results for $K = 2$.

		Inferred cluster	
Locality	**Coral host**	*P. cylindrica*	*P. lobata*
1. Pemuteran	*P. cylindrica*	98%	2%
1. Pemuteran	*P. cylindrica*	100%	0%
1. Pemuteran	*P. cylindrica*	100%	0%
1. Pemuteran	*P. cylindrica*	100%	0%
1. Pemuteran	*P. cylindrica*	100%	0%
1. Pemuteran	*P. cylindrica*	100%	0%
2. Nusa Penida	*P. cylindrica*	97%	3%
2. Nusa Penida	*P. cylindrica*	100%	0%
2. Nusa Penida	*P. cylindrica*	100%	0%
2. Nusa Penida	*P. cylindrica*	98%	2%
2. Nusa Penida	*P. cylindrica*	98%	2%
3. Pulau Mengyatan	*P. cylindrica*	0%	100%
3. Pulau Mengyatan	*P. cylindrica*	100%	0%
4. Lembeh	*P. cylindrica*	10%	90%
5. Bunaken	*P. cylindrica*	100%	0%
5. Bunaken	*P. cylindrica*	97%	3%
5. Bunaken	*P. cylindrica*	100%	0%
5. Bunaken	*P. cylindrica*	100%	0%
5. Bunaken	*P. cylindrica*	99%	1%
5. Bunaken	*P. cylindrica*	100%	0%
6. Dumaguete	*P. cylindrica*	0%	100%
6. Dumaguete	*P. cylindrica*	25%	75%
6. Dumaguete	*P. cylindrica*	45%	55%
6. Dumaguete	*P. cylindrica*	0%	100%
6. Dumaguete	*P. cylindrica*	100%	0%
6. Dumaguete	*P. cylindrica*	77%	23%
Location	**Coral host**	*P. cylindrica*	*P. lobata*
2. Nusa Penida	*P. lobata*	0%	100%
2. Nusa Penida	*P. lobata*	2%	98%
2. Nusa Penida	*P. lobata*	1%	99%
2. Nusa Penida	*P. lobata*	0%	100%
2. Nusa Penida	*P. lobata*	0%	100%
2. Nusa Penida	*P. lobata*	1%	99%
2. Nusa Penida	*P. lobata*	0%	100%
2. Nusa Penida	*P. lobata*	3%	97%
2. Nusa Penida	*P. lobata*	0%	100%
3. Pulau Mengyatan	*P. lobata*	1%	99%
3. Pulau Mengyatan	*P. lobata*	42%	58%
3. Pulau Mengyatan	*P. lobata*	0%	100%

4. Lembeh	*P. lobata*	0%	100%
4. Lembeh	*P. lobata*	0%	100%
4. Lembeh	*P. lobata*	0%	100%
4. Lembeh	*P. lobata*	0%	100%
4. Lembeh	*P. lobata*	0%	100%
4. Lembeh	*P. lobata*	0%	100%
5. Bunaken	*P. lobata*	0%	100%
5. Bunaken	*P. lobata*	0%	100%
5. Bunaken	*P. lobata*	0%	100%
5. Bunaken	*P. lobata*	1%	99%
5. Bunaken	*P. lobata*	0%	100%
6. Dumaguete	*P. lobata*	19%	81%
6. Dumaguete	*P. lobata*	0%	100%

Table 2.3 *Coralliophila violacea*. Again splitting out by coral host, we identified individuals classified as migrant* and admixed** in STRUCTURE.

		Probability of ancestry				
Locality	**Coral host**	*P. cylindrica*	*P. lobata*	Parent	Grandparent	**Classification**
1. Pemuteran	*P. cylindrica*	100%	0%	0%	0%	
1. Pemuteran	*P. cylindrica*	100%	0%	0%	0%	
1. Pemuteran	*P. cylindrica*	100%	0%	0%	0%	
1. Pemuteran	*P. cylindrica*	100%	0%	0%	0%	
1. Pemuteran	*P. cylindrica*	100%	0%	0%	0%	
1. Pemuteran	*P. cylindrica*	100%	0%	0%	0%	
2. Nusa Penida	*P. cylindrica*	100%	0%	0%	0%	
2. Nusa Penida	*P. cylindrica*	100%	0%	0%	0%	
2. Nusa Penida	*P. cylindrica*	100%	0%	0%	0%	
2. Nusa Penida	*P. cylindrica*	100%	0%	0%	0%	
2. Nusa Penida	*P. cylindrica*	100%	0%	0%	0%	
3. Pulau Mengyatan	*P. cylindrica*	0%	100%	0%	0%	*
3. Pulau Mengyatan	*P. cylindrica*	100%	0%	0%	0%	
4. Lembeh	*P. cylindrica*	0%	100%	0%	0%	*
5. Bunaken	*P. cylindrica*	100%	0%	0%	0%	
5. Bunaken	*P. cylindrica*	100%	0%	0%	0%	
5. Bunaken	*P. cylindrica*	100%	0%	0%	0%	
5. Bunaken	*P. cylindrica*	100%	0%	0%	0%	
5. Bunaken	*P. cylindrica*	100%	0%	0%	0%	
5. Bunaken	*P. cylindrica*	100%	0%	0%	0%	
6. Dumaguete	*P. cylindrica*	0%	100%	0%	0%	*
6. Dumaguete	*P. cylindrica*	0%	100%	0%	0%	*
6. Dumaguete	*P. cylindrica*	0%	8%	22%	70%	**
6. Dumaguete	*P. cylindrica*	0%	100%	0%	0%	*
6. Dumaguete	*P. cylindrica*	100%	0%	0%	0%	
6. Dumaguete	*P. cylindrica*	0%	0%	0%	100%	**

		Probability of ancestry				
Location	**Coral host**	*P. cylindrica*	*P. lobata*	Parent	Grandparent	**Classification**
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
2. Nusa Penida	*P. lobata*	0%	100%	0%	0%	
3. Pulau Mengyatan	*P. lobata*	0%	100%	0%	0%	
3. Pulau Mengyatan	*P. lobata*	0%	0%	0%	100%	**

Locality	Coral host	Probability of ancestry				Classification
		P. cylindrica	*P. lobata*	Parent	Grandparent	
3. Pulau Mengyatan	*P. lobata*	0%	100%	0%	0%	
4. Lembeh	*P. lobata*	0%	100%	0%	0%	
4. Lembeh	*P. lobata*	0%	100%	0%	0%	
4. Lembeh	*P. lobata*	0%	100%	0%	0%	
4. Lembeh	*P. lobata*	0%	100%	0%	0%	
4. Lembeh	*P. lobata*	0%	100%	0%	0%	
4. Lembeh	*P. lobata*	0%	100%	0%	0%	
5. Bunaken	*P. lobata*	0%	100%	0%	0%	
5. Bunaken	*P. lobata*	0%	100%	0%	0%	
5. Bunaken	*P. lobata*	0%	100%	0%	0%	
5. Bunaken	*P. lobata*	0%	100%	0%	0%	
5. Bunaken	*P. lobata*	0%	100%	0%	0%	
6. Dumaguete	*P. lobata*	0%	0%	0%	100%	**
6. Dumaguete	*P. lobata*	0%	100%	0%	0%	

Table 2.4 Outlier loci sequences from *Coralliophila violacea*, BLAST hits and functional annotations.

Dataset	Tag	Sequence	Obs. F_{ST}	Organism	Description	Score	E-value	Identity	Gene name	GO functions
All ind.	19628	GGCTATGGGTTTGCAAGGGA GTGCACTCTGCAATCA	0.702	*Helobdella robusta*	protein	31.9	1.9	80%		
All ind.	14249	AGACAAATTGCCGCACACACA TGCAGACAAAACACA	0.633	*Aplysia californica*	histone-lysine N-methyltransferase 2D-like	37.4	0.019	90%	KTM2D	Protein binding, DNA binding, zinc ion binding, methyltransferase activity
All ind.	39884	GGGTTGGCTGTAGCAACCTG CTGCCCCCAAAACCTT	0.718	*Lingula anatina*	uncharacterized	30.1	6.5	79%		
All ind.	37258	GATGATCCTGCAGCAGTGTAC TGCCTCTCTCTCTCT	0.650		KIF1-binding protein	35.6	0.066	84%	KIF1BP	Protein binding, kinesin, binding
All ind.	52997	CCAGGGATCAGCAGTCTCCT GCCACTGTTCCACAAG	0.659	*Aplysia californica*	mRNA for hemocyanin (HC gene)	33.7	0.23	84%	ODHGY	oxidoreductase activity, oxygen transporter activity, metal ion binding, copper binding
All ind.	38182	CGACGGCTAGTGGCAATGCTT TGCAATCGAACATCA	0.572	*Lottia gigantea*	protein	31.9	0.8	83%		
All ind.	24158	GGCCTGATCACTGCAGGATCT TGCTGGTATTTGTCA	0.570	*Biomphalaria glabrata*	uncharacterized	31.9	0.8	82%		
All ind.	17358	CAGAATGTTCATGCAGTCCCA TGCCATGTCTCAACT	0.550	*Astralium milloni*	cytochrome oxidase subunit I (COI)	31.9	0.8	83%		
All ind.	20062	CACCATGTCTATGCACGTGCA TGCAGACACTGGGCA	0.593	*Biomphalaria glabrata*	protein PTHB1-like, transcript variant X2	30.1	2.8	81%	BBS9	protein binding
All ind.	28305	TGCTTGCAACATGCACGCATA TGCACACCACAAACT	0.525	*Babylonia lutosa*	microsatellite sequence	30.1	2.8	84%		
All ind.	16452	AGTGACTGGAGAGCACTTGTT TGCGGCCTATGTTCC	0.427	*Littorina saxatilis*	DNA sequence	41	0.002	88%		
All ind.	11006	CGCAGAAGGAAGGCAAGCAG ATGCCTAATAATCGCT	0.195	*Lottia gigantea*	protein	31.9	0.8	84%		

Dataset	Tag	Sequence	Obs. F_{ST}	Organism	Description	Score	E-value	Identity	Gene name	GO functions
All ind.	32708	TGTGATACTCTTGCACTTTACTGCAAAGGCCATGTT	0.462	*Octopus bimaculoides*	AP2-associated protein kinase 1-like	35.6	0.11	85%	AAK1	protein binding, protein kinase activity, notch activity
All ind.	28347	AGAAAAAGAGGCAGAGAAAGATATGGGAGAAGAACA	0.429	*Aplysia californica*	nuclear hormone receptor HR96-like	39.2	0.005	83%	HR96	xenobiotic detoxification
All ind.	10755	GGTGTGAAATTGGCAGGCAAATGCCTTACTCATCCT	0.471	*Lottia gigantea*	protein	30.1	2.8	83%		
All ind.	10161	CACCCCCTCTATGCAACAATATGCACGTCCCCCTCT	0.478	*Aplysia californica*	pericentriolar material 1 protein-like	28.3	9.7	80%	PCM1	protein binding
All ind.	34705	AGCAGTCTCACTGCAGTTTTCTGCACTGCATAAACT	0.374	*Thais clavigera*	microsatellite sequence	35.6	0.066	86%		
All ind.	33550	TGAGGAAACACAGCATTAGTTTGCAAATTTATTTCT	0.468	*Lingula anatina*	nucleolar pre-ribosomal-associated protein 1-like	33.7	0.54	82%	URB1	poly(A) RNA binding
All ind.	24087	TGCATATTGTGTGCAGTGCCTTGCAGAGTATATGCC	0.404	*Aplysia californica*	latrophilin-1-like	30.1	2.8	83%	ADGRL1	protein binding, transmembrane signaling receptor activity, G-protein coupled receptor activity, latrotoxin receptor activity, carbohydrate binding
All ind.	27266	TGCAATGAAAACACATAAAAACACCTGTGTGCACTC	0.334	*Lottia gigantea*	protein	31.9	0.8	83%		
All ind.	15079	GGCTGAGCAGAGGCAGACGGCTGCGGAGCAGGAGGA	0.407	*Aplysia californica*	thyroid receptor-interacting protein 11-like	31.9	0.8	83%	TRIP11	protein binding, transcription coactivator activity, structural constituent of ribosome
No migrant / hybrid	42043	CGCAATCGTATTGCAAAATTGTGCAATTGCTCCACT	0.748	*Aplysia californica*	uncharacterized	33.7	0.39	84%		

Dataset	Tag	Sequence	Obs. F_{ST}	Organism	Description	Score	E-value	Identity	Gene name	GO functions
No migrant / hybrid	18108	CACATCCATCTCGCATAGTTC TGCTGATCCAGAGCA	0.486	*Aplysia californica*	elongation of very long chain fatty acids protein 2-like	30.1	2.8	81%	Elovl6	transferase activity
No migrant / hybrid	22586	AGAGACAGAGTTGCATCCCTT TGCGTCGCACTCACC	0.651	*Octopus bimaculoides*	uncharacterized	30.1	4.7	78%		
No migrant / hybrid	32951	TACCTTGGGTATGCAACCCGA TGCCAAGACCAAGAT	0.476	*Lottia gigantea*	protein	30.1	4.7	81%	uncharacterized protein	zinc ion binding
No migrant / hybrid	17181	AGCACACAGCACGCACGTGTT TGCACACCAAGAGCA	0.404	*Babylonia formosae habei*	microsatellite sequence	31.9	1.4	86%		
No migrant / hybrid	13296	AGAAAATTCTTGGCACTGTGC TGCTATTGCTTATCA	0.407	*Crassostrea gigas*	stromal membrane-associated protein 1-like	31.9	1.4	81%	SMAP 1	GTPase activator activity, metal ion binding, zinc ion binding, clathrin binding,
No migrant / hybrid	31609	CGAACAGATGTGGCAAAAGAC TGCTGCCTTGGACCA	0.676	*Octopus bimaculoides*	C4orf29-like	30.1	4.7	83%	C4orf 29	
No migrant / hybrid	11613	GGTCCGTGGCTTGCACAGGG ATGCAATGCAATGTCT	0.514	*Helobdella robusta*	protein	30.1	6.5	83%		
No migrant / hybrid	16737	TGTGTTGTGTGTGCAGGTTCA TGCAGCTGATTGGTG	0.440	*Lingula anatina*	transmembrane protein 26-like	30.1	6.5	82%	TMEM26	integral component of the membrane
No migrant / hybrid	17800	TGTGCTTCCTTGGCAGAACCC TGCAAAAATAATCTG	0.424	*Octopus bimaculoides*	lactoylglutathione lyase-like	31.9	1.4	81%		lactoylglutathione lyase activity, metal ion binding,
No migrant / hybrid	10122	TGAGTTCAGCATGCAATCTAG TGCTGCTGGTAGTCC	-0.040	*Lottia gigantea*	protein	31.9	0.8	82%		
No migrant / hybrid	58394	AGGCACACAAATGCAAACACA TGCACGCTGTGCACG	-0.041	*Aplysia californica*	CCR4-NOT transcription complex subunit 4-like	33.7	0.39	85%	CNOT6	RNA binding, metal ion binding, exoribonuclease activity, poly(A)-specific

Dataset	Tag	Sequence	Obs. F_{ST}	Organism	Description	Score	E-value	Identity	Gene name	GO functions
										ribonuclease activity
No migrant / hybrid	31557	CGGAGGTTTGTAGCAGAGCC TTGCCTGCCATAGTCT	0.624	*Aplysia californica*	neurogenic protein mastermind-like				MAM	transcription coactivator activity
No migrant / hybrid	16929	GGGTAATCCAAAGCAACTCAG TGCCTTACCCCCCCT	0.373	*Helobdella robusta*	protein					

C. violacea locus28347: AGAAAAAGAGGCAGAGAAAGATATGGGAGAAGAACA
T
Aplysia californica AGAATGAGAGGGAGAGAAAGAGAGGGGAGAGGAACA

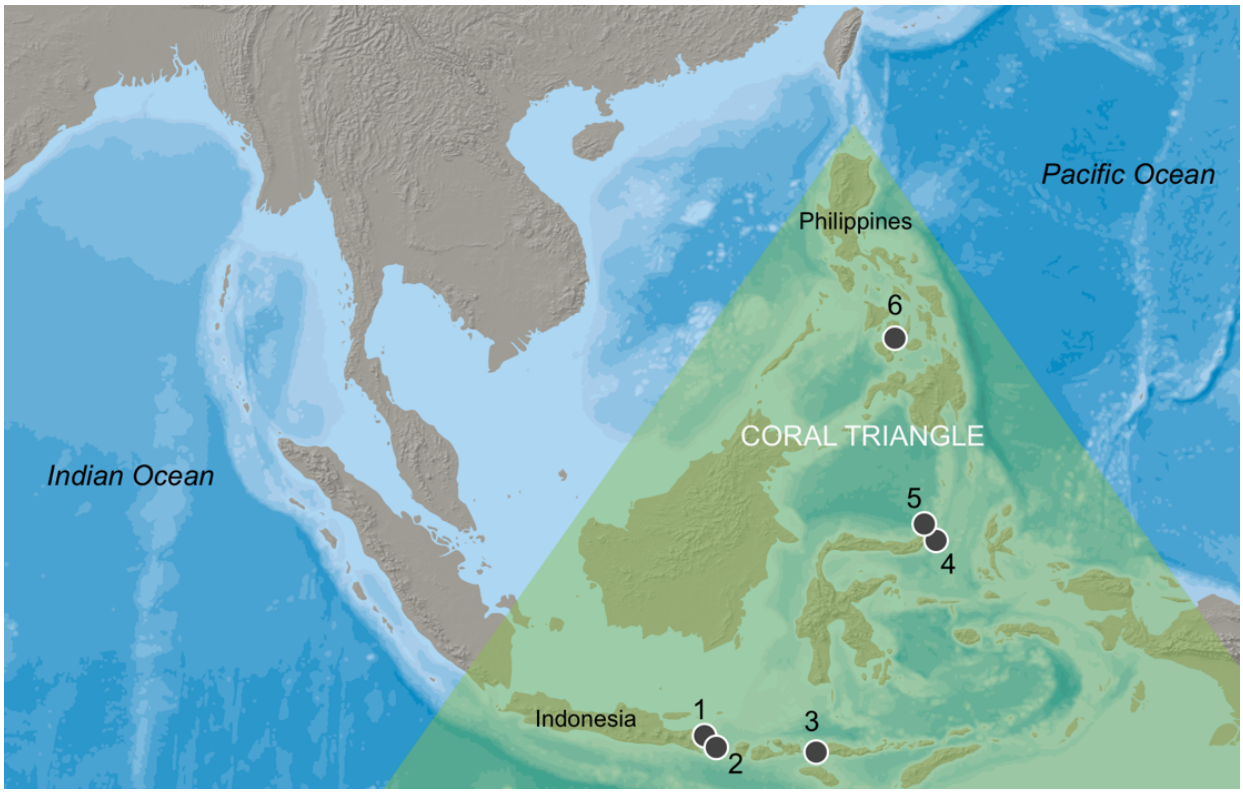


Figure 2.1 Collection localities for *Coralliophila violacea* from coral host species *Porites lobata* and *P. cylindrica*. 1. Pemuteran, 2. Nusa Penida, 3. Pulau Mengyatan, 4. Lembeh, 5. Bunaken, 6. Dumaguete. Raster map made with Natural Earth.

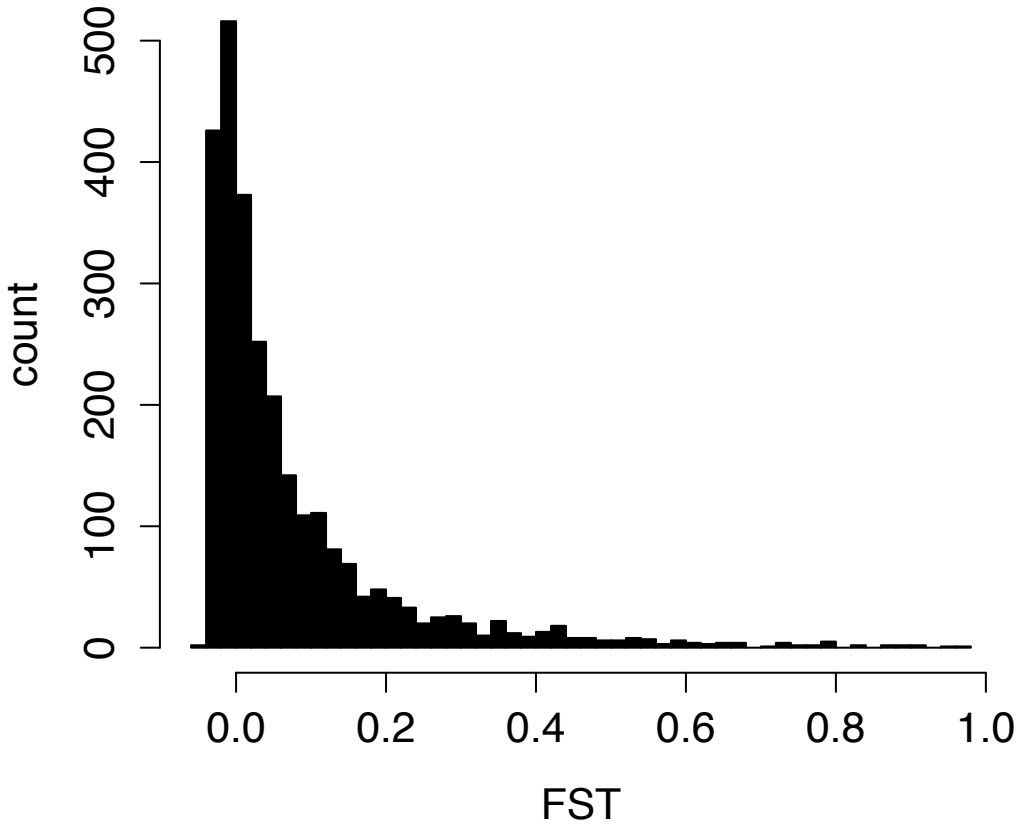


Figure 2.2 Histogram of variation in F_{ST} between sympatric populations of *Coralliophila violacea* on two different coral hosts across all SNPs. Excluding migrants and admixed individuals, F_{ST} calculated using FDIST in ARLEQUIN. Arrow shows the mean F_{ST} value (0.075).

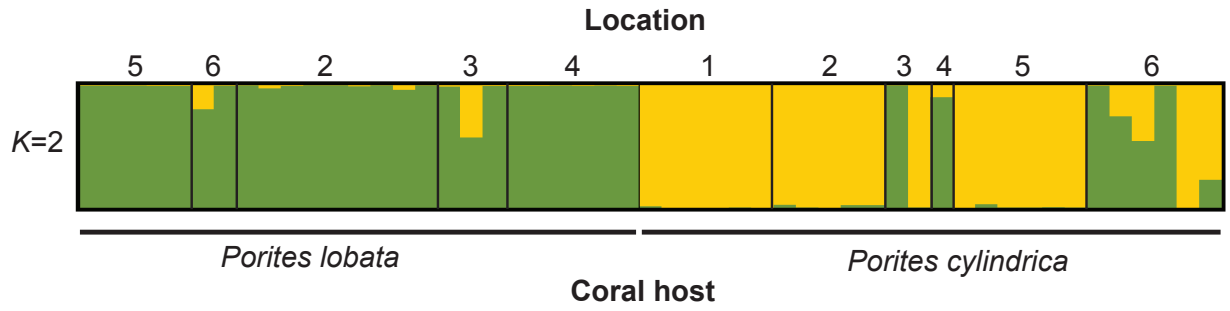


Figure 2.3 Full RADseq dataset (2,718 loci) of all *Coralliophila violacea* individuals. Bar plot of Bayesian assignment probability from STRUCTURE for $K = 2$. Each vertical bar corresponds to an individual. The proportion of each bar represents an individual's assignment probability to cluster one (green) or two (gold), shown grouped by coral host and then by location as numbered in Fig. 2.1.

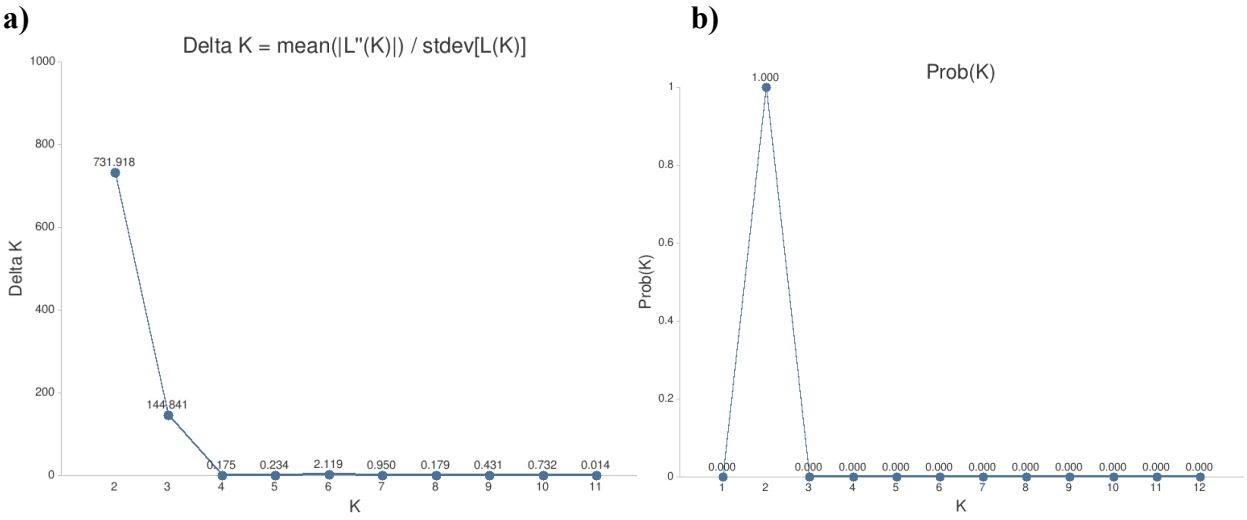
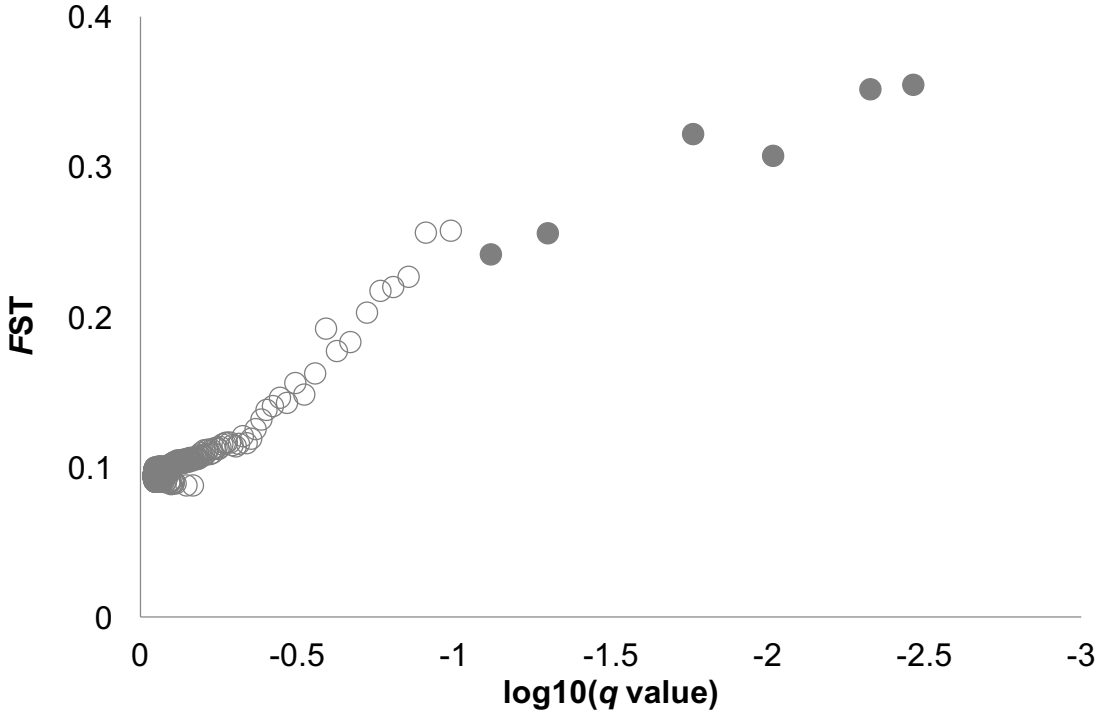


Figure 2.4 Estimates of the best *K* using **a)** the Evanno method (Evanno et al., 2005), and **b)** the method from STRUCTURE identifying the *K* with the highest Pr(K=k). Both methods indicate *K* = 2 is best.

a)



b)

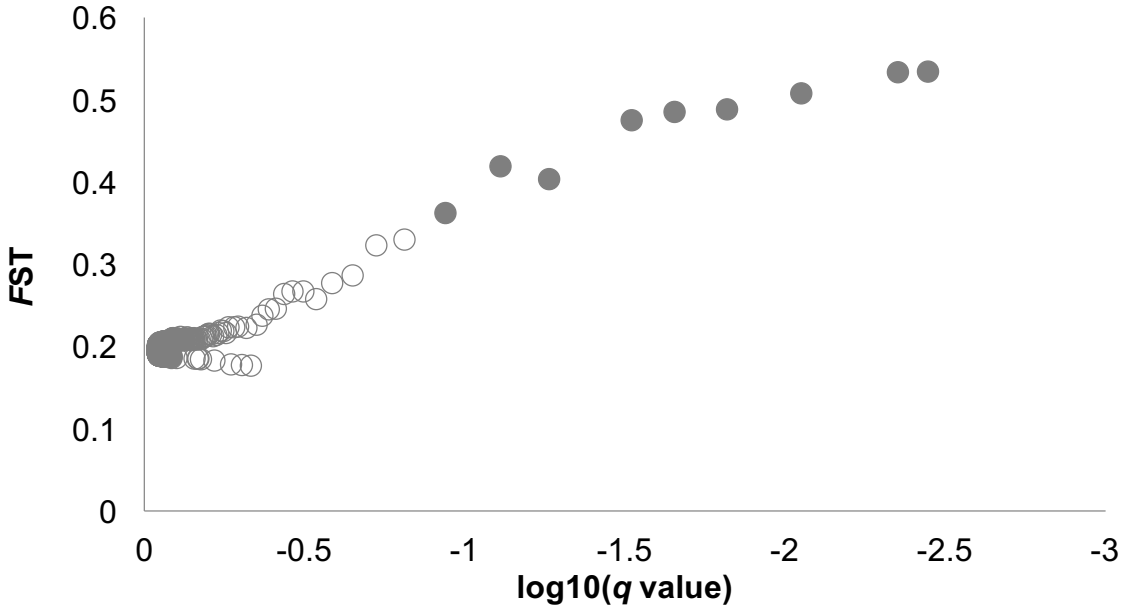
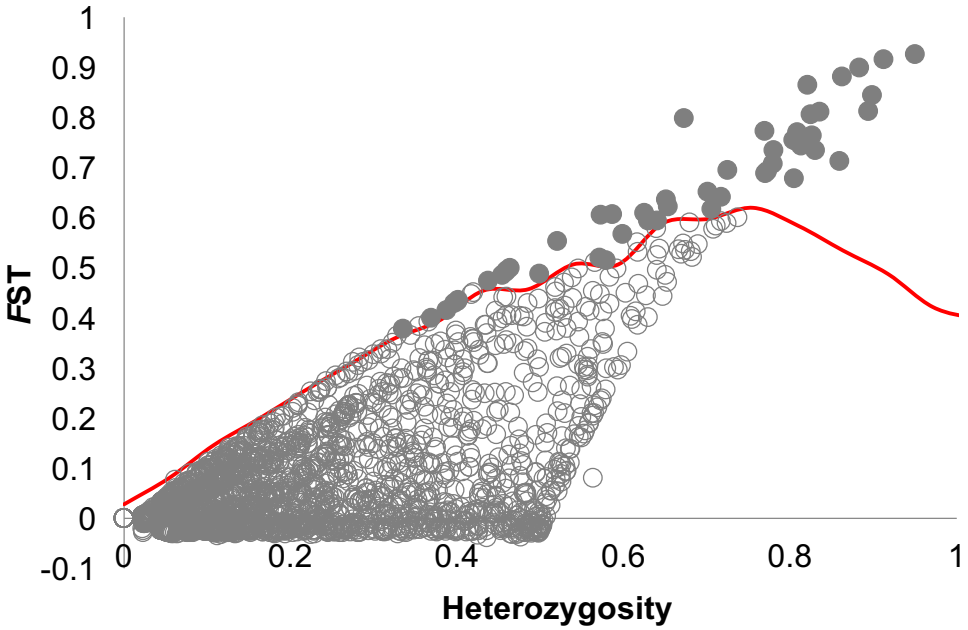


Figure 2.5 Results from BayeScan analysis of full RADseq dataset (2,718 loci) from *Coralliophila violacea*. Filled grey dots are F_{ST} outlier loci. **a)** all individuals, 6 outlier loci identified FDR = 0.10, **b)** excluding migrants and admixed individuals, 8 outlier loci identified FDR = 0.10.

a)



b)

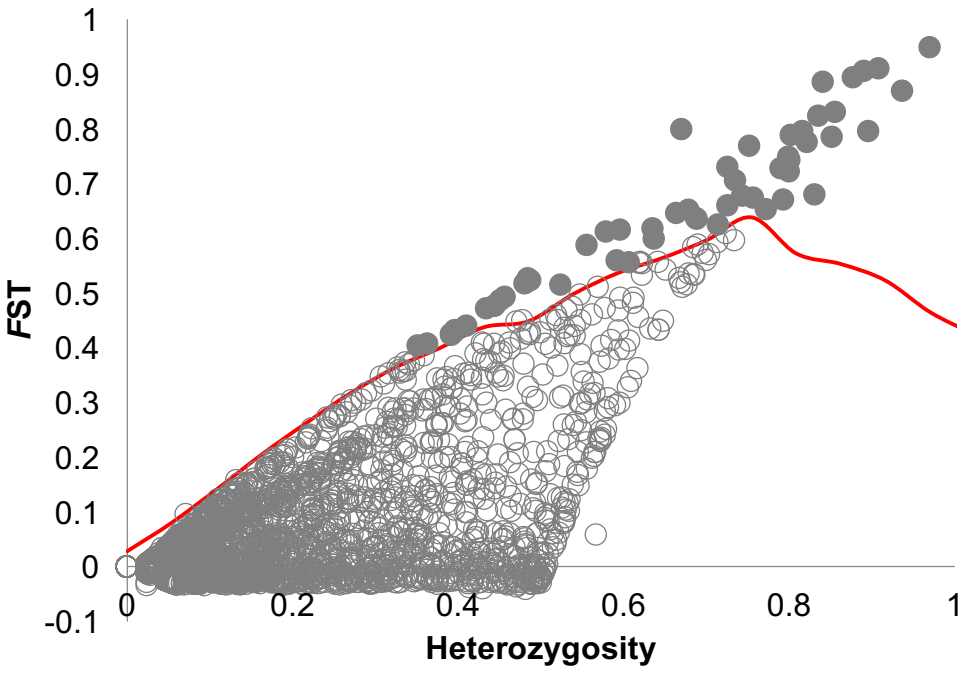


Figure 2.6 Results from FDIST2 analysis implemented in ARELQUIN using the hierarchical island model of migration. Full RADseq dataset (2,718 loci) from *Coralliophila violacea*. Filled grey dots are F_{ST} outlier loci. **a)** all individuals, 51 outliers, **b)** excluding migrants and admixed individuals, 65 outliers.

2.8 REFERENCES

Ackiss A.S., Pardede S., Crandall E.D., Ablan-Lagman M.C.A., Romena N., Barber P.H., & Carpenter K.E. (2013) Pronounced genetic structure in a highly mobile coral reef fish, *Caesio cuning*, in the Coral Triangle. *Marine Ecology Progress Series*, **480**, 185–197.

Barber P.H., Cheng S.H., Erdmann M.V., Tenggardjaja K., & Ambariyanto (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. *Phylogeography and population genetics in Crustacea. Crustean Issues* (ed. by C. Held, S. Koenemann, and C.D. Schubart), pp. 129–156. CRC Press, Boca Raton.

Beaumont M. & Nichols R. (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B*, **263**, 1619–1626.

Bird C.E., Fernandez-Silva I., Skillings D.J., & Toonen R.J. (2012) Sympatric speciation in the post 'Modern Synthesis' era of evolutionary biology. *Evolutionary Biology*, **39**, 158–180.

Bowen B.W., Rocha L.A., Toonen R.J., & Karl S.A. (2013) The origins of tropical marine biodiversity. *Trends in Ecology & Evolution*, **1666**, 1–8.

Briggs J.C. (1999) Modes of speciation: marine Indo-West Pacific. *Bulletin of Marine Science*, **65**, 645–656.

Briggs J.C. (2003) Marine centres of origin as evolutionary engines. *Journal of Biogeography*, **30**, 1–18.

Briggs J.C. (2005) The marine East Indies: diversity and speciation. *Journal of Biogeography*, **32**, 1517–1522.

Carpenter K.E., Barber P.H., Crandall E.D., Ablan-Lagman M.C.A., Mahardika G.N., Manjaji-Matsumoto B.M., Juinio-Meñez M.A., Santos M.D., Starger C.J., & Toha A.H.A. (2011) Comparative phylogeography of the Coral Triangle and implications for marine management. *Journal of Marine Biology*, **2011**, 1–14.

Case A.L. & Willis J.H. (2008) Hybrid male sterility in *Mimulus* (Phrymaceae) is associated with a geographically restricted mitochondrial rearrangement. *Evolution*, **62**, 1026–1039.

Cowman P.F. & Bellwood D.R. (2011) Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. *Journal of Evolutionary Biology*, **24**, 2543–2562.

DeBoer T.S., Naguit M.R.A., Erdmann M.V., Ablan-Lagman M.C.A., Ambariyanto, Carpenter K.E., Toha A.H.A., & Barber P.H. (2014) Concordance between phylogeographic and biogeographic boundaries in the Coral Triangle: Conservation implications based on comparative analyses of multiple giant clam species. *Bulletin of Marine Science*, **90**, 277–300.

Drès M. & Mallet J. (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B*, **357**, 471–492.

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

Excoffier E. & Lischer H. (2011) ARLEQUIN: An integrated software package for population genetics data analysis. *What's New*, 1–174.

Excoffier L. & Lischer H.E.L. (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Faucci A., Toonen R.J., & Hadfield M.G. (2007) Host shift and speciation in a coral-feeding nudibranch. *Proceedings of the Royal Society of London. Series B*, **274**, 111–119.

Foll M. & Gaggiotti O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–93.

Foote A.D., Newton J., Piertney S.B., Willerslev E., & Gilbert M.T.P. (2009) Ecological, morphological and genetic divergence of sympatric North Atlantic killer whale populations. *Molecular Ecology*, **18**, 5207–5217.

Funk D.J., Egan S.P., & Nosil P. (2011) Isolation by adaptation in *Neochlamisus* leaf beetles: host-related selection promotes neutral genomic divergence. *Molecular Ecology*, 4671–4682.

Gittenberger A. & Gittenberger E. (2011) Cryptic, adaptive radiation of endoparasitic snails: sibling species of *Leptoconchus* (Gastropoda: Coralliophilidae) in corals. *Organisms Diversity & Evolution*, **11**, 21–41.

Gordon B.R. & Leggat W. (2010) *Symbiodinium* - Invertebrate symbioses and the role of metabolomics. *Marine Drugs*, **8**, 2546–2568.

Hatfield T. & Schluter D. (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, **53**, 866–873.

Hedgecock D., Barber P.H., & Edmands S. (2007) Genetic approaches to measuring connectivity. *Oceanography*, **20**, 70–79.

Janz N., Nylin S., & Wahlberg N. (2006) Diversity begets diversity: host expansions and the diversification of plant-feeding insects. *BMC Evolutionary Biology*, **6**, doi:10.1186/1471–2148–6–4.

Kawecki T. (1998) Red Queen meets Santa Rosalia: Arms races and the evolution of host specialization in organisms with parasitic lifestyles. *The American Naturalist*, **152**, 635–651.

Kawecki T.J. (1996) Sympatric speciation driven by beneficial mutations. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1515–1520.

Kool J.T., Paris C.B., Barber P.H., & Cowen R.K. (2011) Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. *Global Ecology and Biogeography*, **20**, 695–706.

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

Langerhans R.B. & Gifford M.E. (2009) Divergent selection, not life-history plasticity via food limitation, drives morphological divergence between predator regimes in *Gambusia hubbsi*. *Evolution*, **63**, 561–567.

Lindblom T.H. & Dodd A.K. (2006) Xenobiotic detoxification in the nematode *Caenorhabditis elegans*. *Journal of Experimental Zoology Part A: Comparative Experimental Biology*, **305**, 720–730.

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

Lourie S.A., Green D.M., & Vincent A.C.J. (2005) Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: *Hippocampus*). *Molecular Ecology*, **14**, 1073–1094.

Ludt W.B. & Rocha L.A. (2015) Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, **42**, 25–38.

Mallet J., Meyer A., Nosil P., & Feder J.L. (2009) Space, sympatry and speciation. *Journal of Evolutionary Biology*, **22**, 2332–2341.

Matsubayashi K.W., Kahono S., & Katakuraà H. (2011) Divergent host plant specialization as the critical driving force in speciation between populations of a phytophagous ladybird beetle. *Journal of Evolutionary Biology*, **24**, 1421–1432.

Mayr E. (1963) *Animal species and evolution.* Belknap Press of Harvard University Press, Cambridge, MA.

Miglietta M.P., Faucci A., & Santini F. (2011) Speciation in the sea: overview of the symposium and discussion of future directions. *Integrative and Comparative Biology*, 1–7.

Moran N.A. (1988) The evolution of host-plant alternation in aphids: evidence for specialization as a dead end. *The American Naturalist*, **5**, 681–706.

Munday P.L., Herwerden L.V., & Dudgeon C.L. (2004) Evidence for sympatric speciation by host shift in the sea. *Current Biology*, **14**, 1498–1504.

Nielsen R. (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.

Nosil P. (2002) Transition rates between specialization and generalization in phytophagous insects. *Evolution*, **56**, 1701–1706.

Nosil P., Vines T.H., & Funk D.J. (2005) Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution*, **59**, 705.

Oren U., Brickner I., & Loya Y. (1998) Prudent sessile feeding by the corallivore snail *Coralliophila violacea* on coral energy sinks. *Proceedings of the Royal Society of London. Series B*, **265**, 2043–2050.

Palumbi S.R. (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547–572.

Pritchard J.K., Stephens M., & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Reijnen B.T., Hoeksema B.W., & Gittenberger E. (2010) Host specificity and phylogenetic relationships among Atlantic Ovulidae (Mollusca: Gastropoda). *Contributions to Zoology*, **79**, 69–78.

Richter I. & Fidler A.E. (2014) Marine invertebrate xenobiotic-activated nuclear receptors: their application as sensor elements in high-throughput bioassays for marine bioactive compounds. *Marine Drugs*, **12**, 5590–5618.

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

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

Rundle H.D. & Nosil P. (2005) Ecological speciation. *Ecology Letters*, **8**, 336–352.

Schluter D. (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.

Schluter D. (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Simon J.-C., D'Alencon E., Guy E., Jacquin-Joly E., Jaquiery J., Nouhaud P., Peccoud J., Sugio A., & Streiff R. (2015) Genomics of adaptation to host-plants in herbivorous insects. *Briefings in Functional Genomics*, **14**, 413–423.

Smadja C.M., Canbäck B., Vitalis R., Gautier M., Ferrari J., Zhou J.-J., & Butlin R.K. (2012) Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant specialization and speciation in the pea aphid. *Evolution*, **66**, 2723–2738.

Smith J.M. & Haigh J. (1974) The hitch-hiking effect of a favourable gene. *Genetical Research*, **23**, 23–35.

Sorenson M.D., Sefc K.M., & Payne R.B. (2003) Speciation by host switch in brood parasitic indigobirds. *Nature*, **424**, 928–931.

Soria-Carrasco V., Gompert Z., Comeault A.A., Farkas T.E., Parchman T.L., Johnston J.S., Buerkle C.A., Feder J.L., Bast J., Schwander T., Egan S.P., Crespi B.J., & Nosil P. (2014) Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, **344**, 738–42.

Sotka E.E. (2005) Local adaptation in host use among marine invertebrates. *Ecology Letters*, **8**, 448–459.

Stella J.S., Pratchett M.S., Hutchings P.A., & Jones G.P. (2011) Coral-associated invertebrates: diversity, ecological importance and vulnerability to disturbance. *Oceanography and Marine Biology: An Annual Review*, **49**, 43–104.

Storz J.F. (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, **14**, 671–688.

Terai Y., Seehausen O., Sasaki T., Takahashi K., Mizoiri S., Sugawara T., Sato T., Watanabe M., Konijnendijk N., Mrosso H.D.J., Tachida H., Imai H., Shichida Y., & Okada N. (2006) Divergent selection on opsins drives incipient speciation in Lake Victoria cichlids. *PLoS ONE*, **4**, e433.

Thorpe R.S., Surget-Groba Y., & Johansson H. (2010) Genetic tests for ecological and allopatric speciation in anoles on an island archipelago. *PLoS Genetics*, **6**, 1–12.

Tsang L.M., Chan B.K.K., Shih F.-L., Chu K.H., & Allen Chen C. (2009) Host-associated speciation in the coral barnacle *Wanella milleporae* (Cirripedia: Pyrgomatidae) inhabiting the *Millepora* coral. *Molecular Ecology*, **18**, 1463–1475.

Veron J.C.E.N., DeVantier L.M., Turak E., Green A.L., Kininmonth S., Stafford-Smith M., & Peterson N. (2011) The Coral Triangle. *Coral Reefs: An Ecosystem in Transition* (ed. by Z. Dubinsky and N. Stambler), pp. 47–55. Springer Netherlands.

Via S. (2009) Natural selection in action during speciation. *Proceedings of the National Academy of Sciences*, **106**, 9939–9946.

Wang S., Meyer E., McKay J.K., & Matz M.V. (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, **9**, 808–810.

Weir B. & Cockerham C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Zann L.P. (2002) A review of macrosymbiosis in the coral reef ecosystem. *International Journal of Parasitology*, **17**, 399–405.

CHAPTER 3:

GENOME-WIDE SIGNATURES OF POPULATION STRUCTURE AND LOCAL ADAPTATION IN A CORALLIVOROUS SNAIL, *CORALLIOPHILA VIOLACEA*

3.1 ABSTRACT

The objective of this study was to differentiate between neutral and adaptive genetic variation in a highly abundant and broadly distributed coral reef species. The corallivorous snail *Coralliophila violacea* inhabits reefs throughout the tropical Indo-Pacific. We collected *C. violacea* from colonies of their coral host *Porites* at ten localities spanning a significant portion of their geographic range. Four genetic partitions were concordant with regions previously observed in mtDNA in Chapter 1 (the Indian Ocean, the Coral Triangle and Hawai'i), with the addition of Vietnam, and varying levels of isolation and admixture. The Vietnam and Hawai'i populations were each strongly isolated, showing insignificant gene flow from other localities. In contrast, localities in the Coral Triangle were highly connected to each other, with inputs from surrounding waters (Vietnam, Indian Ocean). Finally, sites in the Indian Ocean had the highest levels of admixture we measured, indicating that this region acts as a sink for dispersal from other sources. Using BayeScan we found outlier loci among individual sites (FDR = 0.10, N = 72) and between regions (FDR = 0.10, N = 34), suggesting that some loci are putatively under divergent selection, or are linked to genes under selection. We also investigated possible local adaptation via genetic-environmental associations with five ocean climate variables. These association analyses showed that the two strongest drivers of local adaptation were the annual

range and the mean of sea surface temperature variation. Marginal populations (e.g. Hawai'i) drive these associations at the periphery of *C. violacea*'s range adapting to lower temperature. Our results show that local adaption to different environments likely reinforces neutral divergence, especially in peripheral populations.

3.2 INTRODUCTION

Marine speciation has perplexed evolutionary biologists for decades. Since the pioneering work of Scheltema (1971), which showed that larvae can travel extreme distances, many evolutionary biologist have noted the difficulty of explaining speciation in the ocean by allopatric models (Palumbi, 1994; Briggs, 2006; Rocha & Bowen, 2008; Puebla, 2009; Miglietta et al., 2011). However, this viewpoint changed with the advent of DNA sequencing and phylogeographic studies that uncovered pronounced geographic structure in the population genetics of marine species (e.g. Benzie & Williams, 1997; Williams & Benzie, 1998; Barber et al., 2000; Wares, 2002; Taylor & Hellberg, 2003), and identified potential dispersal barriers in the sea. Studies also demonstrated that larvae can fall short of their dispersal potential l(Knowlton & Keller, 1986), predictions from ocean currents frequently overestimate realized dispersal (Shanks, 2009), and self-recruitment can be much higher than expected (Jones et al. 1999, Swearer et al. 1999).

Over the past decade, researchers have used mitochondrial DNA (mtDNA) to document strong phylogeographic structure in a wide diversity of marine taxa (Kelly & Palumbi, 2010; Bowen et al., 2014). As a result of this growing body of literature, the pendulum has swung back toward allopatry as the primary mechanism of speciation, particularly among those focused on studying the biodiversity of the Indo-Pacific (for reviews see Gaither et al., 2010; Barber et al.,

2011; Carpenter et al., 2011; Bowen et al., 2013). These studies often report one or more of the following phylogeographic patterns: 1) population divergence between the Indian and Pacific Ocean basins; 2) differentiation of populations within the Coral Triangle, the global epicenter of marine biodiversity; or 3) differentiation of populations on the periphery of the Pacific (e.g. Hawai'i, Marquesas) and Indian oceans (e.g. the Red Sea). The divergence between the Indian and Pacific Oceans is typically ascribed to the exposure of the Sunda and Sahul continental shelves (Fig. 3.1) when Plio-Pleistocene glaciations lowered sea levels by 115–130m (Voris, 2000). These shelves created terrestrial landmass barriers that constricted the waterways of the Indonesian and Philippine Archipelago, and are implicated in reducing gene flow between the Indian and Pacific oceans in a wide diversity of marine taxa including seahorses (Lourie & Vincent, 2004; Lourie et al., 2005); soldierfish (Craig et al., 2007); anemonefish (Timm & Kochzius, 2008; Dohna et al., 2015); damselfish (Drew & Barber, 2009; Liu et al., 2014; Raynal et al., 2014); groupers (Gaither et al., 2011a); fusiliers (Ackiss et al., 2013); limpets (Kirkendale & Meyer, 2004); snails (Reid et al., 2006; Crandall et al., 2008a); seastars (Kochzius et al., 2009); and giant clams (DeBoer et al., 2008, 2014; Kochzius & Nuryanto, 2008; Nuryanto & Kochzius, 2009). Within the Coral Triangle, studies suggest that oceanographic features such as the Halmahera Eddy (Fig. 3.1) can limit gene flow by constraining larval exchange, a hypothesis supported by phylogeographic studies (Barber et al., 2006, 2011; DeBoer et al., 2008; Ackiss et al., 2013), as well as biophysical connectivity models (Kool et al., 2011; Treml et al., 2015). Lastly, explanations of diversification on the periphery of species' ranges have focused largely on physical isolation (Briggs, 2005), either because great expanses of open ocean between habitats act a filter to dispersal (Kobayashi, 2006), or due to the cumulative effects of isolation-by-distance (Wright, 1943; Slatkin, 1993).

While it is clear that a range of processes contributes to isolation and divergence in the sea, purely allopatric models still face challenges. For example, while Plio-Pleistocene sea levels did constrict the waterways of Indonesia and the Philippines (Voris, 2000; Ludt & Rocha, 2015: Fig 3.1), the Makassar Straight, Maluku and Banda Seas—the major pathways for the waters of the Indonesian Throughflow—have depths that exceed 3,000 m. Thus, even during periods of extremely low sea levels, the major pathway for water movement and dispersal between the Pacific and Indian Oceans remained open, forcing authors to invoke other processes such as increased cold water upwelling as a mechanism to limit dispersal (Fleminger, 1986). Moreover, periods of isolation were punctuated by tens of thousands of years of oceanic conditions that were more similar to today. These conditions would have resulted in gene flow and population expansion, potentially erasing the signal of historical isolation (e.g. Crandall et al., 2008a, 2008b). Similarly, ocean currents in the region are temporally variable, both seasonally (Shinoda et al., 2012) and across epochs (Kuhnt et al., 2004). Therefore, while oceanographic features may promote isolation during some periods, they can be reversed in others.

These incomplete or ephemeral dispersal barriers, combined with the immense population sizes and extensive geographic ranges of marine taxa, suggest that other processes, like natural selection, must play a role in generating biodiversity in the ocean (Sanford & Kelly, 2011). As such, authors are increasingly advocating for a more inclusive model of speciation that combines both neutral and adaptive processes (Bowen et al., 2013; Horne, 2014). Since natural selection limits realized connectivity between populations (Burgess et al., 2012), local adaptation in response to differences in climate or habitat can reinforce patterns of neutral divergence driven by gene flow, genetic drift, or mutations (Gavrilets, 2003). Even species with high productivity and long planktonic duration times can have relatively low effective gene flow if selection favors

local progeny over those recruiting from different environments. Given the combination of demographic and geographic features of many marine species, natural selection can dominate over genetic drift if the selection coefficient is greater than the inverse of the effective population size (Slatkin, 1993).

There is growing empirical evidence to support the notion that selection may play a larger role in the diversification of marine taxa than previously thought. For example, studies on several Indo-Pacific fauna show that environmental and habitat heterogeneity (Briggs, 2006; Rocha & Bowen, 2008; Longo & Bernardi, 2015; Saenz-Agudelo et al., 2015), or competition among species (Briggs, 1992; Bowen et al., 2013), can drive diversification or reinforce nascent allopatric divergence. Similarly, within the hyper-diverse waters of the Coral Triangle, recent work suggests that divergent selection between habitats or hosts contributes to adaptive variation and speciation (Meyer et al., 2005; Reid et al., 2006; Sbrocco, 2012; Cheng, 2015; Tornabene et al., 2015). Particularily within peripheral areas of a species ranges, environmental conditions may be at or near the limits of that species' physiological threshold (Johannesson & Andre, 2006; Kawecki, 2008; Gaither et al., 2010; DiBattista et al., 2016), creating distinct selective environments.

Until recently, limits to DNA sequencing and a lack of high-resolution marine environmental databases have hampered our ability to test specifically the role of environmental variation in shaping patterns of divergence in the sea. Recently, however, studies using high-throughput genomic tools have clearly demonstrated that natural selection can drive genetic divergence of peripheral marine populations (e.g. Gaither et al., 2015; Saenz-Agudelo et al., 2015). These sequencing advances, combined with the establishment of global, high-resolution marine environmental databases (Sbrocco & Barber, 2013), mean we are entering an era where

we can more directly assess the role of natural selection driving or reinforcing diversification in the ocean.

The marine snail *Coralliophila violacea* (Kiener, 1836) is an ectoparasite, specializing on corals in the family Poritidae. *C. violacea* inhabits shallow-water coral reefs over an extremely broad geographic range (Demond, 1957) across the tropical Indo-Pacific from the Red Sea to the Eastern Pacific. They are obligate corallivores, living, feeding, and reproducing exclusively on their coral hosts. Previous research on *C. violacea* demonstrated striking genetic differentiation of snails living on different coral host lineages, despite ongoing gene flow (Chapters 1, 2). The snail ecomorphs specializing on coral taxa in *Porites* lineage 1 are broadly distributed and exhibit phylogeographic structure both in the core of their geographic range (i.e. across the Sunda Shelf, Fig. 3.1) and in peripheral populations (i.e. Hawai'i), and can potentially disperse long distances. As such, it is likely that variation in ocean climate variables across the range of *C. violacea* is reinforcing or accelerating population divergence via natural selection, similar to Gaither et al., (2015).

In this Chapter, we test the relative roles of neutral and adaptive processes in shaping population divergence in the sea by combining genome-wide surveys of genetic variation in *C. violacea* with data from marine environmental databases. Specifically, we test for divergence across known phylogeographic provinces within the Coral Triangle, and for divergence among peripheral populations in the Indian and Pacific Oceans, and the South China Sea. We then compare patterns of neutral and non-neutral variation to geography and environmental variables, to assess their relative roles in shaping population divergence.

3.3 MATERIALS AND METHODS

3.3.1 Sample collection

To test for patterns of divergence related to geography and local adaptation in populations at the center and the periphery of the range of *C. violacea*, we collected snails on snorkel or scuba at ten localities (Table 3.1, Fig. 3.2). These localities spanned a significant portion of the snail's geographic range, including either side of the Sunda Shelf Barrier, an area where phylogeographic structure is commonly observed (Barber et al., 2011), and also known areas of peripheral isolation (i.e. Hawai'i). We only used samples of snails collected from coral colonies from the *Porites* lineage 1 (*P. lobata* and *P. compressa*, Chapter 1), and which had been previously DNA barcoded. This selectivity was necessary because *Porites* corals are notoriously difficult to identify *in situ* due to their morphological plasticity and small corallites (Forsman et al., 2015). Also, genetically similar colonies can have vastly different morphologies and *vice versa* (Forsman et al., 2009, 2015). In total, we collected snails from 1–3 colonies at each locality for a total of 71 snails from 32 coral colonies. A portion of each snail's foot tissue was preserved in 95% ethanol and stored at room temperature for DNA analysis.

3.3.2 Creation of RADseq libraries

For each snail collected, we extracted genomic DNA from 20 mg of foot tissue using a Qiagen DNeasy Blood and Tissue Kit (Qiagen). We followed the manufacturer's instructions, except in the last step when we cleaned DNA eluted with 100 µl of molecular grade H_2O rather than AE buffer. We estimated initial DNA concentrations using a NanoDrop and visualized DNA quality on a 1% agarose gel stained with SYBR

Green. We used only high-quality DNA with a bright high molecular weight band and minimal smearing. We dried DNA extractions using a SpeedVac on medium heat and reconstituted using molecular grade H_2O to a final uniform 250 ng/µl DNA concentration.

To create reduced representation libraries to survey SNP variation, we prepared 2bRAD libraries following the protocol of Wang et al., (2012) as modified by Dr. Eli Meyer (http://people.oregonstate.edu/~meyere/tools.html). AlfI restriction enzyme digest reduced representation ($1/16^{th}$) libraries were labeled with individual barcodes and subjected to 18–20 PCR amplification cycles. We electrophoresed products on a 2% agarose gel in 1× TBE buffer and run at 150 V for 90 minutes. Target bands (165 bp) were visualized with SYBR SAFE dye and excised from the gel. The excised band was then purified using a QIAquick gel purification kit (Qiagen). For a final cleaning step, we used Ampure XP beads (Beckman-Coulter). The QB3 Vincent J. Coates Genomics Sequencing Laboratory at UC, Berkeley conducted quality checks (qPCR, BioAnalyzer) and sequencing of the resulting libraries, multiplexing 10–20 snails per lane in 5 lanes of a 50 bp single-end run on the Illumina HiSeq2000 platform.

3.3.3 RADseq data processing

To prepare raw sequence data for SNP identification, we truncated all raw reads to the insert size (36 bp), filtered them for quality (PHRED scores >20), and discarded empty constructs. To process the data, we used custom scripts available on GitHub https://github.com/z0on/2bRAD_denovo. STACKS is a program commonly used to process RADseq data but it cannot be used with 2bRAD sequences because they have double strands and STACKS considers reverse-complements as separate loci. Therefore,

we used a pipeline to emulate the steps of STACKS, while taking advantage of the fact that 2bRAD sequences both strands. In a step similar to making "stacks", unique tag sequences (minimum sequencing depth 5×) were counted, the number in reverse-complement orientation recorded and merged into one table. Then all sequences were clustered in CD-HIT using a 91% similarity threshold. We defined the most abundant sequence in the cluster as the reference, and we then filtered a locus-annotated table from the previous two steps, excluding reads below 5× depth or exhibiting strand bias. The orientation of the resulting clustered sequences was then flipped to match the most abundant tag in a cluster.

To call genotypes (as population-wide RADtag haplotypes), we applied mild allele filters (10× total depth, allele bias and strand bias), with the additional requirement that alleles appear in at least two individuals. We then applied locus filters, allowing maximum 50% heterozygotes at a locus, no more than two alleles, genotyped in 30% of samples and polymorphic. Finally, loci with too many heterozygotes (75%) and missing genotypes (70%) were removed. The final set of SNPs was then thinned to one per tag (the one with the highest minor allele frequency) for STRUCTURE analysis and gene-environment association tests to remove linked loci that might be in linkage disequilibrium.

3.3.4 Individual sample filtering steps

From the first 71 individuals, we filtered out those with low genotyping rates ($N = 5$) indicating poor DNA quality, by taking the $\log_{10}$ of the number of sites genotyped per individual, and removing individuals ≥ 2 standard deviations (SD) of the mean. We also removed individuals ($N = 4$) with high homozygosity (+/– 2 SD of the mean F inbreeding coefficient) indicating

potential contamination. We used the remaining 63 individuals in all analyses. The final data file was in VCF format and we converted it to other data formats using PGDSpider v2.1.0.1 (Lischer & Excoffier, 2012).

3.3.5 Genetic diversity

We calculated basic population genetic summary statistics using all filtered loci for each locality with ARLEQUIN v 3.5 (Excoffier & Lischer, 2011). For each site, we estimated the number and frequency of polymorphic loci, observed and expected heterozygosity, and genetic diversity (averaged across all loci).

3.3.6 F_{ST} outlier loci test

We tested for loci potentially under natural selection by identifying outlier F_{ST} values using BayeScan v2.1 (Foll & Gaggiotti, 2008). We ran BayeScan with the data structured in two ways to look for outliers at different spatial scales, 1) comparing individual sites, and 2) comparing sites grouped into regions (Table 3.1). We ran the data with a burn-in of 50,000, a thinning interval 10, a sample size 5,000, 100,000 iterations, and 20 pilot runs of 5,000, each examining two different false discovery rates (0.10 and 0.05).

3.3.7 Population genetic structure

To infer the population genetic structure and individual admixture proportions, we used the Bayesian model-based clustering method STRUCTURE (Pritchard et al., 2000). We examined three datasets consisting of 63 individuals: 1) full RADseq dataset using all loci; 2) using only neutral loci as determined by BayeScan; and 3) outlier loci with a false discovery rate of 10% (FDR = 0.10, BayeScan). We ran STRUCTURE with a burn-in period of 20,000 followed by 50,000 MCMC replicates for K = 1–10, and ten runs for each K. We used the admixture model,

with allele frequencies correlated among populations. We selected the optimal value of *K* using the Δ*K* statistic described in (Evanno et al., 2005), and summarized the results graphically using the program CLUMPAK v1.1 (Kopelman et al., 2015). To test for the significance of genetic partitions identified by STRUCTURE, we ran analyses of molecular variance (AMOVA), both with and without regional groupings (Table 3.1). We determined significance by 100,000 random replicates in ARLEQUIN.

3.3.8 Genetic-environment association test

Ocean climate variables that impact species distribution (Briggs, 2006) may also structure population genetics within species (Sanford & Kelly, 2011). We, therefore, examined differences in allele frequencies associated with environmental variables using a Bayesian framework (Bayenv 2.0; Coop et al., 2010; Günther & Coop, 2013) that accounts for demographic history. We obtained ocean environmental variables from the MARSPEC database (Sbrocco & Barber, 2013). Because of strong correlations among environmental variables (Sbrocco, 2012; Sbrocco & Barber, 2013), we selected only five variables in the MARSPEC database for analysis: 1) temperature of the warmest ice-free month (biogeo15), 2) mean annual sea surface temperature (annual sea surface temperature; biogeo13), 3) annual range in sea surface temperature (biogeo16). 4) mean annual sea surface salinity (biogeo08), and 5) annual range in sea surface salinity (biogeo11). Temperature variables are of particular importance to this system because reefs in locations with high temperatures are most likely to be impacted by coral bleaching events (Hoegh-Guldberg, 1999). Coral mortality from bleaching could directly impact *C. violacea* living on those hosts, as well as indirectly by limiting the number and species of hosts available to *C. violacea* in affected reefs.

To create maps of each of the five environmental variables across the study region, we projected the annual mean and range of sea surface salinity (SSS) and temperature (SST), plus mean warmest monthly temperature at ~5 km resolution onto an equidistant cylindrical world using the '*raster'* package in R. Then, we extracted the five climate variable data to each point locality and divided by a scaling factor of 100 using custom R scripts. Next, we estimated a covariance matrix with standardized environmental variables for each sampling locality, as suggested by the authors of Bayenv (Coop et al., 2010; Günther & Coop, 2013), and 3,186 loci that were polymorphic among all sites, for 100,000 iterations, outputting the results every 500 iterations. We used the last printed covariance matrix for all further analyses. To estimate the Bayes Factor (BF) for each SNP with each ocean climate variable, we ran Bayenv for 100,000 Markov chain Monte Carlo (MCMC) iterations. For each ocean climate variable, SNPs with $\log_{10}$ BF > 1 were considered to give substantial-to-strong support for candidate loci, based on criteria from Jefferys (1961).

3.3.9 Candidate gene annotation

To annotate the function of genes linked to outlier loci and candidate loci, we aligned sequences containing SNP outlier loci to nucleotide collections (nr/nt) available on the NCBI website using the BLASTN algorithm at two different taxonomic levels, 1) Mollusca (taxid:6447) and 2) Lophotrochozoa (taxid:1206795). We adjusted parameters (expected threshold 10, word size 7, no low complexity filter, no mask for look up table) to accommodate short read sequences. We only examined hits with a high percent query coverage (>85%) and used NCBI to identify and annotate any associated genes.

3.4 RESULTS

The average number of unique reads per individual was 6.6 million at minimum 5× depth after removing empty constructs and filtering for quality. We sequenced and genotyped 46,148 high-quality RADseq loci with ≥25× coverage, in 71 individuals at ten locations. We filtered the dataset for 30% maximum missing data per locus, leaving 7,862 loci, and then thinned to one SNP per loci to remove any physically linked SNPs for STRUCTURE and F_{ST} analyses, leaving 3,188 SNPs. Next, we removed nine individuals that had either low DNA quality (missing data ≥ +1SD from the mean) or potential contamination issues (inbreeding coefficient ≥ +2SD from the mean), leaving 63 individuals.

3.4.1 Genetic diversity

The mean frequency of polymorphic loci across localities was 47%. Hon Mun in Vietnam had the highest frequency of polymorphic loci (81%), most likely because it had the most sequenced individuals (Table 3.2). Observed heterozygosity (H_O = 0.196–0.570) and expected heterozygosity (H_E = 0.225–0.586) varied across localities, but H_O was consistently lower than the H_E at every locality sampled (Table 3.2). The mean of the gene diversity over all loci across sites was 0.122, and gene diversity varied across localities (0.089–0.162; Table 3.2). Gene diversity was highest in Dumaguete (0.162) and lowest in Bunaken (0.095) (Table 3.2).

3.4.2 Gene flow and genetic structure

Results from STRUCTURE using both the full dataset of 3,188 loci or just 3,116 neutral loci, confirmed that K = 4 was the best-supported number of population partitions as determined using the methods described by Evanno et al. (2005). The four genetic partitions corresponded to distinct biogeographic regions: a) Indian Ocean (1. Vav'varu, 2. Pulau Weh); b) Vietnam (3. Hon

Mun); c) the Coral Triangle (4. Dumaguete, 5. Bunaken, 6. Lembeh, 7. Pulau Mengyatan, 8. Nusa Penida and 9. Manokwari); and d) Hawai'i (10. Ka'a'awa) (Fig 3.3). Levels of admixture among these four regions were variable. Hawai'i and Vietnam showed minimal admixture from other localities. In contrast, sites in the Indian Ocean had the highest levels of admixture, with numerous individuals exhibiting a mix of alleles from the Indian Ocean and populations to the east. Specifically, Pulau Weh exhibited substantial admixture with the Coral Triangle, and Vav'varu in the Maldives had significant admixture from both Vietnam and the Coral Triangle. Levels of admixture varied across sites in the Coral Triangle. Pulau Mengyatan and Manokwari had the highest levels, while Dumaguete, Bunaken, Lembeh and Nusa Penida had the least; admixture in these populations was largely from the Indian Ocean and Vietnam. Results were virtually identical whether we used all loci or just neutral loci.

Consistent with STRUCTURE, AMOVA showed significant population structure across all RADseq loci (overall F_{ST} = 0.071, p >0.001). Hon Mun in Vietnam was isolated from all other localities (pairwise F_{ST} = 0.063–0.147; Table 3.3) except Pulau Mengyatan in southern Indonesia. Ka'a'awa in Hawai'i was strongly divergent from four other localities (pairwise F_{ST} = 0.142–0.235; Table 3.3). Similarly, Pulau Weh in the Indian Ocean was different from all localities (pairwise F_{ST} = 0.021–0.168; Table 3.3) except for Vav'varu in the Maldives, Dumaguete in the Philippines and Ka'a'awa in Hawai'i. AMOVA analyses grouping sites into four biogeographic regions (Indian Ocean, South China Sea, Coral Triangle, Hawai'i) (Table 3.1, Fig. 3.4) increased the overall F_{ST} to 0.098 (p >0.001), and the F_{CT} was 0.109 (p >0.001). However, regional variation only accounted for 10% of the overall variation, and the majority (90%) was within populations (Table 3.4).

3.4.3 Outlier tests

With a false discovery rate (FDR) of 10%, BayeScan revealed 72 outlier loci putatively under directional selection among all localities (FDR = 0.10, 72; FDR = 0.05, 57), and 34 outlier loci among regions (Indian Ocean, Vietnam, Coral Triangle, Hawai'i; FDR = 0.10, 34; FDR = 0.05, 25) (Fig 3.4). Considering only these 72 outlier loci, STRUCTURE indicated that $K = 2$ was the best-supported number of genetic partitions (Fig. 3.3). In contrast to results from all loci or just neutral loci, the two genetic partitions identified using only outlier loci do not correspond to individual sites. Instead, two groups of sites were found 1) Hon Mun, Vietnam and Ka'a'awa, Hawai'i, two locations separated by ~9,800 km of open ocean, and 2) all other sites (Fig. 3.3).

3.4.4 Ocean climate variables and associations with allele frequency differences

Results from MARSPEC show significant variation in the five environmental variables across the sampled region. The temperature of the warmest ice-free month varied by more than 3 ˚C across the sampled sites, with the highest temperatures at sites in the Indian Ocean (~29.4 ˚C), and lowest in Hawai'i (26.4 ˚C) and Vietnam (28.20 ˚C) (Table 3.1. Fig 3.5a). The mean annual sea surface temperature (SST) also varied substantially across sampling localities by ~4 ˚C (Table 3.1, Fig. 3.5a). However, in this case, Manokwari (28.90 ˚C) and Pulau Weh (28.59 ˚C) at the edges of Indonesia were the warmest, while again Hawai'i (24.99 ˚C) and Vietnam (26.45 ˚C) were the coldest localities (Table 3.1, Fig. 3.5a). SST at some sites was very stable, with only small annual ranges: Manokwari (0.78 ˚C) and Bunaken (1.05 ˚C). At other locations, SST varied more widely (e.g. Dumaguete (3.16 ˚C) and Vietnam (4.22 ˚C); Table 3.1, Fig. 3.5b).

Sea surface salinity (SSS) also showed variation across sampling localities (Table 3.1, Fig. 3.5c). Sites in the two oceanic archipelagos, Vav'varu, Maldives (34.93 psu) and Ka'a'awa, Hawai'i (34.92 psu), had the highest values while salinity at sites in the Coral Triangle was

consistently lower (33.89–34.09 psu) (Table 3.1, Fig. 3.5c). Localities also experienced annual variation in salinity with inputs from freshwater runoff and precipitation (Fig. 3.5d). Sites in the Indian Ocean, Vav'varu (SSS range = 1.64 psu) and Pulau Weh (SSS range = 1.43 psu) had the most variable salinity while Ka'a'awa (SSS range = 0.21 psu) in Hawai'i was the most stable (Table 3.1, Fig. 3.5d).

Using Bayenv, we identified a total of 88 SNPs for which allele frequency differences correlated with one or more of the five ocean climate variables (Fig. 3.6). The highest number of these 88 SNPs were associated with annual range of temperature ($N = 38$), followed by the mean annual temperature ($N = 31$), the temperature of the warmest ice-free month ($N = 21$), annual range of salinity ($N = 12$) and mean annual salinity ($N = 10$) (Fig. 3.6). SNPs associated with temperature had the highest Bayes Factors (Fig. 3.6). There was much overlap of SNPs identified as correlating with the temperature variables (mean, range and warmest month SST). 22 of these SNPs were associated with more than one variable ($N = 2$ with three variables, $N = 20$ with two variables). Mean SST and temperature of the warmest month were the most highly correlated, with 15 SNPs shared between them, followed by mean and range of SST ($N = 5$ SNPs). There was no overlap between the 88 candidate loci associated with environmental variables and the 72 outlier loci.

3.4.5 Mapping and annotation of outliers and candidate loci

The majority (77%) of candidate loci associated with environmental variables did not map to any other mollusc or lophotrochozoan sequences available in the NCBI database (Table SI 3.1). However, we were able to align the other 23% of candidate loci to DNA sequences from a variety of organisms including an octopus (*Octopus bimaculoides*), an oyster (*Crassostrea gigas*), an abalone (*Haliotis discus*), a sea hare (*Aplysia californica*),

and several snails (*Rapana venosa*, *Biophalaria glabrata*, *Conus episcopatus*). Of these loci, two mapped to putatively neutral markers (microsatellites), and the remaining 18 mapped to protein-coding regions. Seven of these coded for uncharacterized proteins and the other eleven were characterized with annotated functions (Table SI 3.1). In addition, BayeScan identified 72 outlier loci among sites and we were only able to align a small proportion (19%) of these to nucleotides in the NCBI database from various of organisms including an oyster (*Crassostrea gigas*), a sea hare (*Aplysia californica*), a limpet (*Lottia gigantea*) and two freshwater snails (*Bellamya aeruginosa*, *Biophalaria glabrata*) and two marine snails (*Conus episcopatus, Coralliophila abbreviata*) (Table SI 3.1). The top three outlier loci, in terms of % identity of sequences and the lowest E-value, aligned with a microsatellite from a congener (*Coralliophila abbreviata*), a conotoxin from a cone snail (*Conus episcopatus*), and a gene (CAPLA) coding for a protein involved in behavioral plasticity in the California seahare (*Aplysia californica*). Interestingly, despite no overlap between the 72 outlier loci identified by BayeScan and the 88 candidate loci identified by Bayenv, both sets contained loci that mapped to genes coding for conotoxins (Table SI 3.1).

3.5 DISCUSSION

Our results reveal a complex mixture of genetic patterns from *Coralliophila violacea* that appear to be shaped by both geography and environmental variables. Genome-wide data from >3,000 SNP loci identified four genetic partitions concordant with traditionally recognized biogeographic regions (Indian Ocean, Vietnam, Coral Triangle and Hawai'i) (Spalding et al., 2007). These findings are concordant with those from our mitochondrial DNA studies of *C.*

violacea (Chapter 1) as well as other marine phylogeography studies from this region (see Barber et al., 2011; Carpenter et al., 2011 for reviews). Repeated patterns across multiple taxa suggest a common origin (Avise 2000), most likely broad-scale physical isolation of population. However, examination of outlier loci and the correlation between allelic variation and environmental variables also indicate the importance of natural selection in shaping population genetics. More than 2% of loci were outliers among sites, similar to results from Gaither et al. (2015), showing evidence of divergence that is significantly greater than can be explained by neutrality alone (Foll & Gaggiotti, 2006, 2008). Also, nearly 3% of loci were candidate loci, having strong correlations with five ocean environmental variables. Hawai'i and Vietnam had similar climate profiles and were differentiated at both neutral and outlier loci, suggesting that climate differences, in particular, sea surface temperature, may be reinforcing neutral geographic structure, and leading to differentiation of peripheral populations. These results, combined with the observation of ecological differentiation of *C. violacea* on different coral hosts from Chapters 1 and 2, indicate that diversification in marine environments is more complex and varied than currently believed, and likely includes selective and neutral processes

3.5.1 Potential role of natural selection in shaping genetic patterns

While neutral loci showed a clear signal of physical limits to gene flow (below), all five environmental variables showed some gene associations (changes in allele frequency). In total, nearly 3% of loci showed evidence of environmental associations, with the range and mean of sea surface temperature (SST) having the strongest associations. SST affects the growth and survival rates of gastropods, including marine snails (for review see Sanford & Kelly, 2011). Thus, it is perhaps not surprising that there are a substantial number of candidate loci associated with temperature. Results also showed an interesting pattern of associations driven by

temperature of the coldest sites (Hon Mun, Vietnam, and Ka'a'awa, Hawai'i). The Hawaiian Islands have environmental conditions considered to be marginal for the coral host's (*Porites*) growth and survival (Polato et al., 2010). Of the locations we sampled, Hawai'i and Vietnam had the coldest temperatures and Vietnam was the most variable throughout the year. The differences in environmental variables between peripheral populations and central populations in the Coral Triangle may enable selection to drive populations apart. For example, Reid et al., (2006) found that environmental and ecological factors (i.e. continental vs. oceanic habitat and primary productivity) shaped the genetics of intertidal snails in the Indo-Pacific. Interestingly, while all of the four regions sampled have unique environmental signatures when examining the candidate loci under environmental influence, our results only differentiate between populations in two clusters: Vietnam and Hawai'i cluster together to the exclusion of Indonesia, the Philippine and the Maldives.

If environmental variation is driving natural selection, it is reasonable to expect some degree of overlap between the outlier loci identified among sites with BayeScan and those identified by Bayenv. BayeScan identified 72 outlier loci among sites, and Bayenv identified 81 loci, but there was no overlap between these two sets of loci. That being said, given that the total number of loci identified in both analyses is relatively small, the lack of overlap may not be a complete surprise. Moreover, while temperature and salinity can certainly exert selection pressure on marine populations (e.g. Limborg et al., 2012; Defaveri et al., 2013), these are not the only variables that do so. For example, Gaither et al. (2015) explained divergent selection at a visual pigment gene locus in a coral reef fish as a function of the high turbidity and low light conditions in the Marquesas.

A small proportion of loci aligned to available sequences in the NCBI database. Only 11/88 loci from Bayenv aligned to nucleotides in protein-coding regions with predicted functions. This situation highlights the lack of genomic resources currently available for non-model organisms, hindering identification of the likely functions of the candidate loci. Also, none of the genes identified seemed to have obvious roles in relation to their associated climate variables. However, two groups of genes are of interest for future research. First, conotoxin genes showed up in both sets of loci. While neurotoxins have not been studied in *Coralliophila*, venoms have been identified in other Neogastropoda (Fange, 2012) and could be important in facilitating feeding and living on corals to deactivate nematocysts. Second, another gene of interest (CHS5) is predicted to be involved in chitin biosynthesis; a substance found mollusc shells and radulae. Chitin is also used in the epidermis and stomach lining of nudibranchs that feed on Cnidarians as protection from nematocysts (Rainer et al., 2007). Coral-eating snails such as *Coralliophila* and *Drupella* could deploy chitin in a similar fashion.

3.5.2 Populations at the periphery

Hawaiian populations of *C. violacea* showed strong genetic differentiation from populations in the Coral Triangle, Indian Ocean, and the Vietnam. Many studies report strong genetic differentiation of Hawaiian populations of taxa with Indo-Pacific distributions, invoking various mechanisms for this divergence, but most commonly geographic isolation (Leray et al., 2010; Dibattista et al., 2012; Fernandez-Silva et al., 2015; Ahti et al., 2016; Waldrop et al., 2016), and population bottlenecks (Szabó et al., 2014). The Hawaiian populations of *Porites lobata,* the primary host of *C. violacea,* are similarly diverged (Polato et al., 2010; Baums et al., 2012), suggesting the action of a common diversification process. Given that strong differences were observed in the neutral loci, as well as the full data set, there are clearly limits to gene flow in *C.*

violacea, resulting in genetic drift and population divergence. This pattern may stem from the relative isolation of Hawai'i from other coral reefs. The Hawaiian Islands are the most isolated archipelago in the world, 860 km to the northeast of the nearest Pacific island (Johnston Atoll) and ~7,000 km from the Coral Triangle in the core of *C. violacea*'s range. However, another explanation for high F_{ST} values between the Hawaiian archipelago and other sampled populations is isolation-by-distance. Being a good fit to Wright's Island Model (Wright, 1931), many marine species exhibit isolation-by-distance (Wright, 1943; Slatkin, 1993), with several partial filters to gene flow adding up to large genetic differences among populations at the periphery. Additional sampling will be required to test this possibility, but there was no evidence of isolation-by-distance within Coral Triangle populations, suggesting that geographic isolation may be the best explanation. However, given the limited number of samples from Hawai'i, these results should be considered preliminary.

The second example of differentiation on the periphery is in the Vietnam population. In contrast to Hawai'i though, populations in Vietnam are much less geographically isolated. These reefs are only 10s to 100s of kms from other reef systems in the South China Sea and, compared with the Hawai'i population, are relatively close to sites in the Coral Triangle (e.g. 1,700 km east of Dumaguete). These strong differentiations in neutral loci we observed in Vietnam populations of *C. violacea* most likely results from a combination of factors. First, the South China Sea was partially enclosed during Plio-Pleistocene low sea levels stands (Voris, 2000; Ludt & Rocha, 2015; Fig. 3.1), resulting in isolation and genetic differentiation of other marine taxa in this region (Barber et al., 2000). Second, ocean circulation patterns limit water movement from the Coral Triangle into the South China Sea (Kool et al., 2011), creating a strong barrier dividing populations north and south across the middle of the South China Sea (Treml et al., 2015). In

fact, ocean circulation patterns have been invoked to explain genetic structure of *C. violacea* populations from Taiwan and Taiping, a small island in the South China Sea (Lin & Liu, 2008). While both sea level and oceanography could promote differentiation, it is important to note that the environment of the South China Sea is also very different, impacting the frequencies of nearly 3% of the loci sampled. As such, physical isolation could be augmented or reinforced by environmental variation.

Interestingly, populations in the Indian Ocean (Vav'varu and Pulau Weh) were less differentiated than Hawai'i or Vietnam. Although STRUCTURE results had a clearly distinct Indian Ocean signature, these populations also had a substantial amount of admixture from the Coral Triangle and the South China Sea. Given that water flows from the Pacific to the Indian Ocean through the Indonesian archipelago and the South China Sea (Gordon & Fine, 1996; Gordon et al., 2003), it is not surprising to see admixture in the Indian Ocean. In fact, this pattern is observed in a wide variety of taxa (e.g. Williams & Benzie, 1998; Drew & Barber, 2009). However, the environmental conditions in this region are also different than other areas of the sampled range, suggesting that selection might act as it does in other parts of the range of *C. violacea*. Thus, it is possible that gene flow from the Coral Triangle and the South China Sea or other un-sampled areas limit the signal of environmental selection in the Maldives, despite the relative isolation of these reef ecosystems.

3.5.3 Populations at the core

In contrast to peripheral populations, populations within the Coral Triangle at the core of *C. violacea*'s range, showed no significant evidence of divergence. Populations in the Philippines (Dumaguete), eastern (Manokwari) and central Indonesia (Bunaken, Lembeh, Pulau Mengyatan, Nusa Penida) were all connected to each other by gene flow, echoing findings for other molluscs

in the region (Kirkendale & Meyer, 2004; DeBoer et al., 2008, 2014; Kochzius & Nuryanto, 2008; Nuryanto & Kochzius, 2009). High connectivity among these populations is expected due to the strong currents of the Indonesian Throughflow and is predicted by coupled biophysical models of larval dispersal (Kool et al., 2011, Treml et al., 2015). The only isolation experienced by Coral Triangle populations is from populations on the periphery: Vietnam in the South China Sea, the Indian Ocean, and Hawai'i.

However, high connectivity is not expected among all populations in the Coral Triangle. Specifically, studies of multiple taxa including giant clams (DeBoer et al., 2008, 2014; Kochzius & Nuryanto, 2008); mantis shrimp (Barber et al., 2006, 2011); echinoderms (Crandall et al., 2008b); and fish (Timm & Kochzius, 2008; Ackiss et al., 2013; Jackson et al., 2014) show isolation of populations spanning the Maluku Sea. This pattern has typically been attributed to the effects of the Halmahera Eddy (Barber et al., 2006; Fig. 3.1) and is predicted by coupled bio-physical models (Kool et al., 2011; Treml et al., 2015). However, Treml et al. 2015 found that only 8.3–10.6% of simulated taxa were filtered by the Halmahera Eddy. The strength of dispersal barriers in the region was largely determined by life history traits such as reproductive output, the timing of spawning and length of larval dispersal phase (Treml et al., 2015). Therefore, differences in the life history traits could explain why some species respond to the Halmahera Eddy while others are unaffected (Treml et al., 2015). For instance, *C. violacea* has traits that would make it easier to cross seasonally variable dispersal barriers. These snails have high fecundity (Lin & Liu, 1995), brood larvae most of the year (Soong & Chen, 1991; Lin & Liu, 1995), and likely have a long dispersal duration (Taylor, 1975). In contrast, giant clams have a relatively short planktonic larval duration (9–10 days, Crawford et al., 1986), which constrains their dispersal potential. Population structure in *Tridacna* may also be influenced by

environmental conditions that impact their symbiotic communities of *Symbiodinium* (DeBoer et al., 2011).

3.6 CONCLUSIONS

While allopatric divergence is clearly an important process in shaping the evolutionary history of marine taxa, physical limits to dispersal and gene flow do not operate in isolation. Our results show that local adaption to different environments likely reinforces neutral divergence, especially in peripheral populations. These results add to the growing recognition that a range of factors impact diversification in coral reef species (Hoeksema, 2007; Floeter et al., 2008; Rocha & Bowen, 2008; Bowen et al., 2013; Fernandez-Silva et al., 2015; Ludt & Rocha, 2015). The intense focus on studying neutral process was partially driven by limitations in DNA sequencing. However, recent advances in high throughput genome-wide sequencing will facilitate the exploration of neutral and adaptive variation in concert. Combined with the new availability of marine environmental databases, we are on the verge of developing a more inclusive view of processes driving genetic divergence in marine populations, hopefully leading to a paradigm shift in how we approach the study of speciation in the sea.

3.7 TABLES AND FIGURES

Table 3.1 Sampling localities for *Coralliophila violacea* collected from coral hosts *Porites lobata* and *P. compressa* from *Porites* lineage 1, with corresponding environmental variables from MARSPEC database for each locality used in Bayenv2 analysis. Coordinates are in decimal degrees. Locality numbers correspond to those in Figure 1. Regions were used for AMOVA analyses. Sea surface salinity (SSS) in psu (practical salinity units). Sea surface (SST) in ˚C. *N* = no. of individuals.

Locality	Country	Region	Latitude	Longitude	Warmest mo. SST	SST	SST range	SSS	SSS range
1. Vav'varu	Maldives	Indian Ocean	5.419	73.358	29.42	28.37	1.76	34.93	1.64
2. Pulau Weh	Indonesia	Indian Ocean	5.887	95.348	29.39	28.59	1.40	33.29	1.43
3. Hon Mun	Vietnam	Vietnam	12.170	109.308	28.20	26.45	4.22	33.34	0.77
4. Dumaguete	Philippines	Coral Triangle	9.332	123.312	28.54	27.61	3.16	33.96	0.92
5. Bunaken	Indonesia	Coral Triangle	1.612	124.783	28.74	28.34	1.05	33.95	0.79
6. Lembeh	Indonesia	Coral Triangle	1.479	125.251	28.81	27.87	2.94	33.95	0.85
7. Pulau Mengyatan	Indonesia	Coral Triangle	-8.557	119.685	28.47	27.51	1.95	34.09	0.68
8. Nusa Penida	Indonesia	Coral Triangle	-8.675	115.513	28.81	27.60	2.12	33.89	1.19
9. Manokwari	Indonesia	Coral Triangle	-0.888	134.085	29.16	28.90	0.78	34.09	1.03
10. Ka'a'awa	USA	Hawai'i	21.584	-157.887	26.43	24.99	2.81	34.92	0.21

Table 3.2 Standard diversity indices and summary statistics for all 3,188 loci obtained from *Coralliophila violacea*. *N* = no. of individuals, no. of polymorphic loci per locality, frequency of polymorphic loci, observed heterozygosity (H_O), expected heterozygosity (H_E) and gene diversity over all loci.

Locality	*N*	No. poly. loci	Freq. poly. loci	H_o	H_E	Gene diversity
1. Vav'varu	4	1293	41%	0.345	0.449	0.104
2. Pulau Weh	3	1051	33%	0.459	0.544	0.103
3. Hon Mun	19	2598	81%	0.196	0.225	0.159
4. Dumaguete	2	1022	32%	0.570	0.586	0.162
5. Bunaken	7	1577	49%	0.281	0.342	0.095
6. Lembeh	6	1515	48%	0.301	0.341	0.114
7. Pulau Mengyatan	4	1143	36%	0.360	0.463	0.089
8. Nusa Penida	9	1881	59%	0.271	0.296	0.115
9. Manokwari	5	1680	53%	0.274	0.357	0.149
10. Ka'a'wa	4	1229	39%	0.353	0.397	0.132

Table 3.3 Genetic distances of pairwise populations of *Coralliophila violacea* for all 3,188 SNPs. Regions follow labelling in Table 1.1. Significant pairwise F_{ST} values at p >0.01 (α=0.05, corrected for multiple tests using the B-Y method, Narum 2006) are shaded.

	Indian Ocean		**Vietnam**	**Coral Triangle**					**Hawai'i**
Locality	1	2	3	4	5	6	7	8	9	10
1. Vav'varu	0									
2. Pulau Weh	-0.026	0								
3. Hon Mun	0.014	0.034	0							
4. Dumaguete	0.063	0.021	0.122	0						
5. Bunaken	0.068	0.072	0.067	0.011	0					
6. Lembeh	0.085	0.077	0.109	0.008	0.007	0				
7. Pulau Mengyatan	0.022	0.033	-0.004	-0.025	-0.005	0.002	0			
8. Nusa Penida	0.059	0.052	0.098	0.013	0.000	0.011	-0.020	0		
9. Manokwari	-0.005	-0.022	0.063	0.020	-0.011	0.024	-0.050	0.024	0	
10. Ka'a'awa	0.157	0.168	0.147	0.217	0.221	0.235	0.142	0.223	0.164	0

Table 3.4 Hierarchical AMOVA of *Coralliophila violacea* populations sampled across all 3,188 loci with localities grouped by region as indicated in Table 1. Significant *F*-statistic values at p >0.001 are in bold face.

Source of variation	***F*-statistics**	**P-values**	**% var.**
Among regions	$\boldsymbol{F_{CT} = 0.109}$	<0.001	10.93
Among populations	F_{SC} = -0.012	0.389	-1.09
Within populations	$\boldsymbol{F_{ST} = 0.098}$	<0.001	90.16

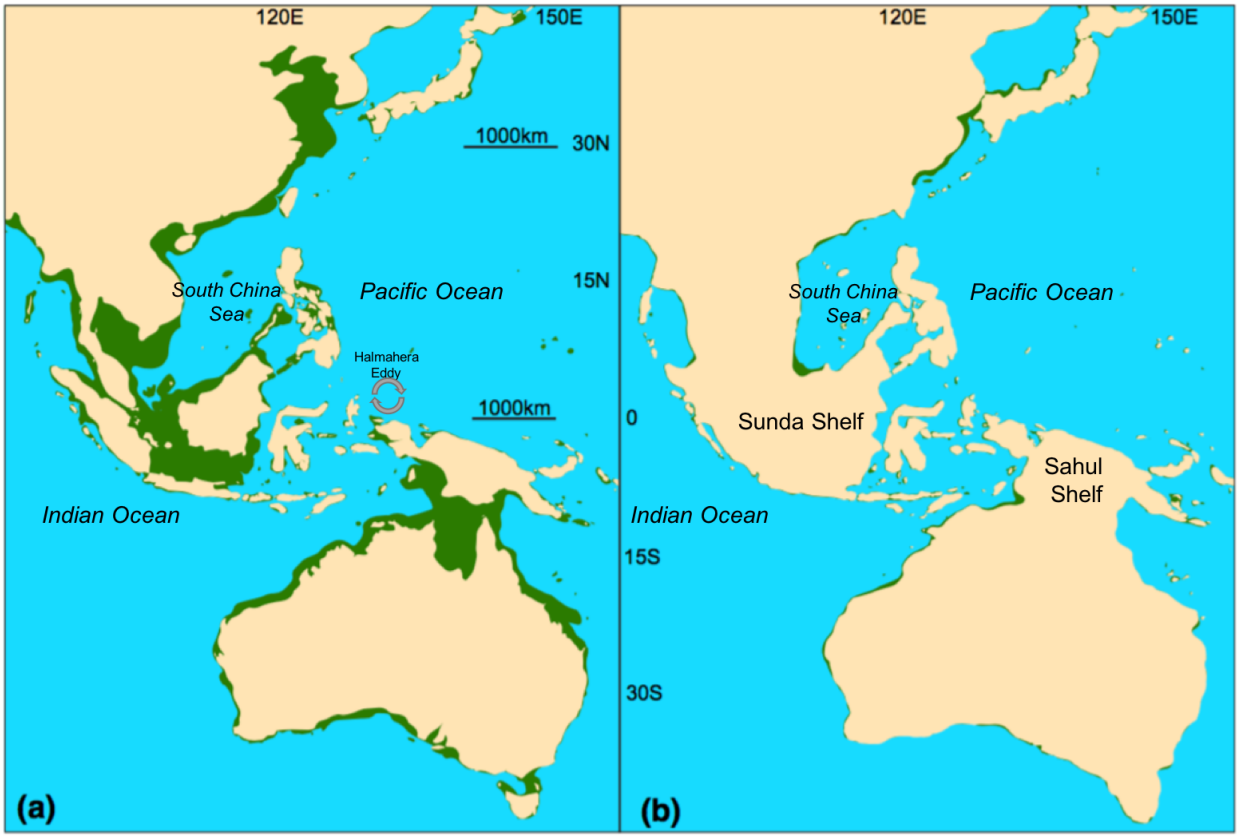


Figure 3.1. Maps showing the exposed landmasses (tan) and habitable shelf habitat (green) for present sea levels **a)**, and sea levels at the Last Glacial Maximum **b)**, in the Western Pacific and Eastern Indian Oceans as well as the South China Sea. The approximate location of the Halmahera Eddy is also indicated. Author modified the figure from Ludt & Rocha (2015).

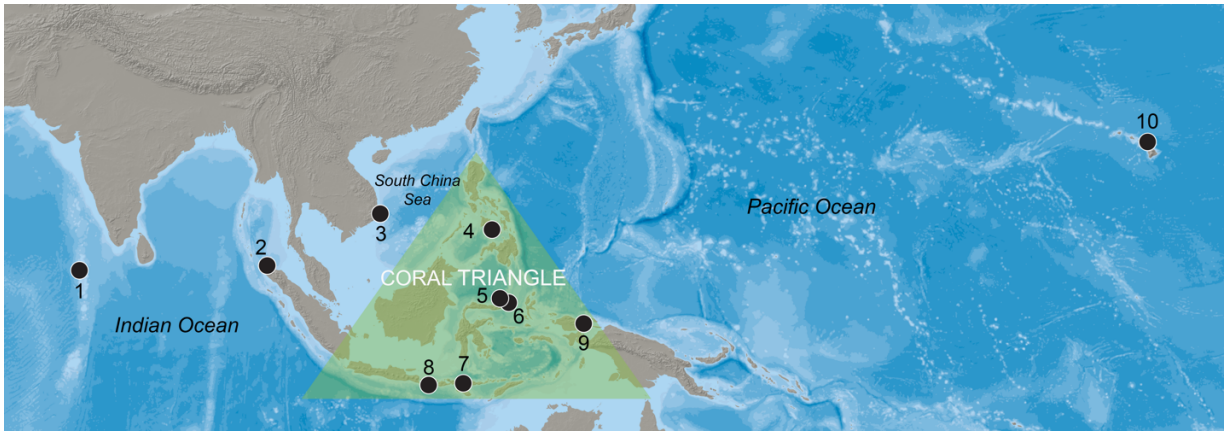


Figure 3.2 Map of *Coralliophila violacea* collection localities across the Indo-Pacific region. Sampled populations include: 1. Vav'varu. 2. Pulau Weh 3. Hon Mun 4. Dumaguete 5. Bunaken 6. Lembeh 7 Pulau Mengyatan 8. Nusa Penida 9. Manokwari 10. Ka'a'awa. Raster map made with Natural Earth.

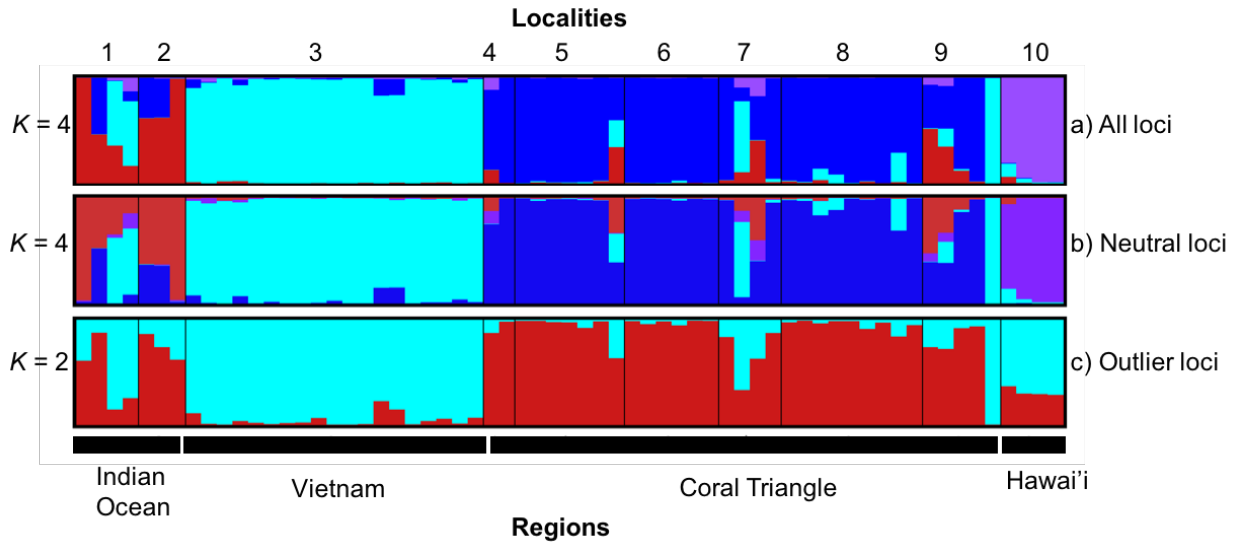


Figure 3.3 Patterns of genetic structure in *Coralliophila violacea* in the Indo-Pacific region. CLUMPAK-averaged STRUCTURE plots for 10 independent runs, clustered and average using CLUMPAK. Each bar represents an individual and the color is the proportion of assignment to each genetic partition. Localities plotted from left to right and numbered above, and black bars indicate regions as in Table 3.1 **a)** all 3,188 loci at $K = 4$, **b)** 3,116 neutral loci at $K = 4$, **c)** 72 outlier loci at $K = 2$.

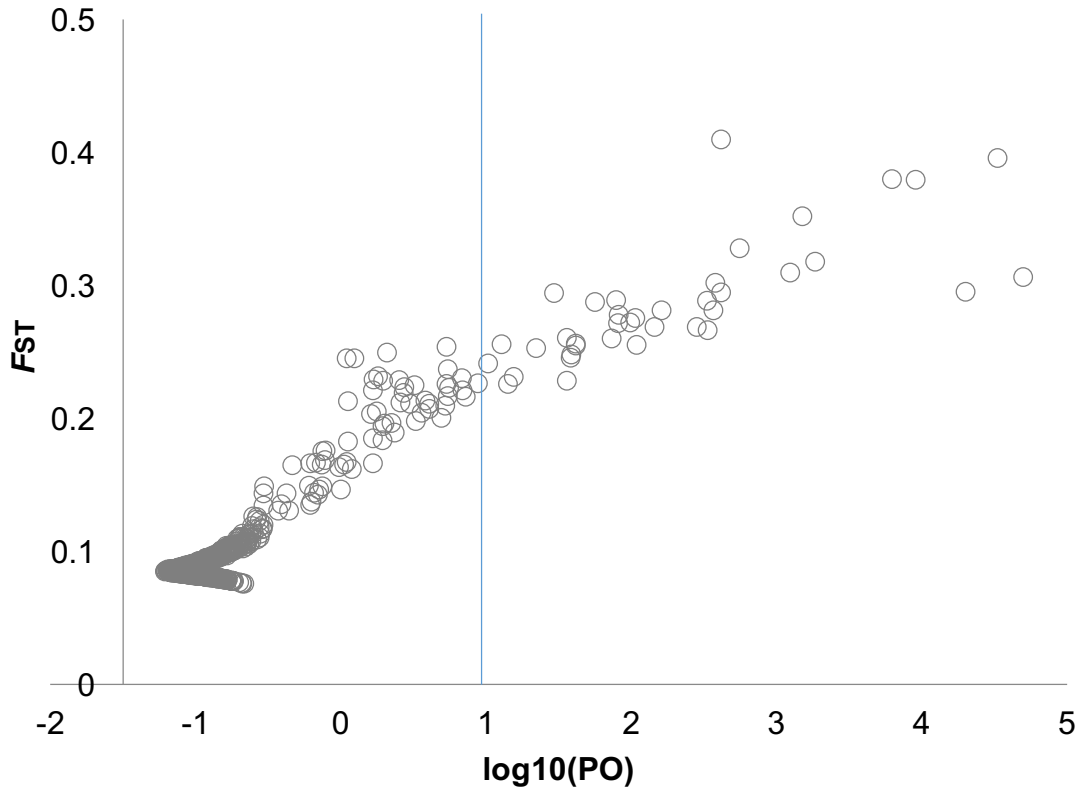


Figure 3.4 Outlier loci among sampled populations of *Coralliophila violacea as* identified by BayeScan. Points to the right of the blue line are outliers.

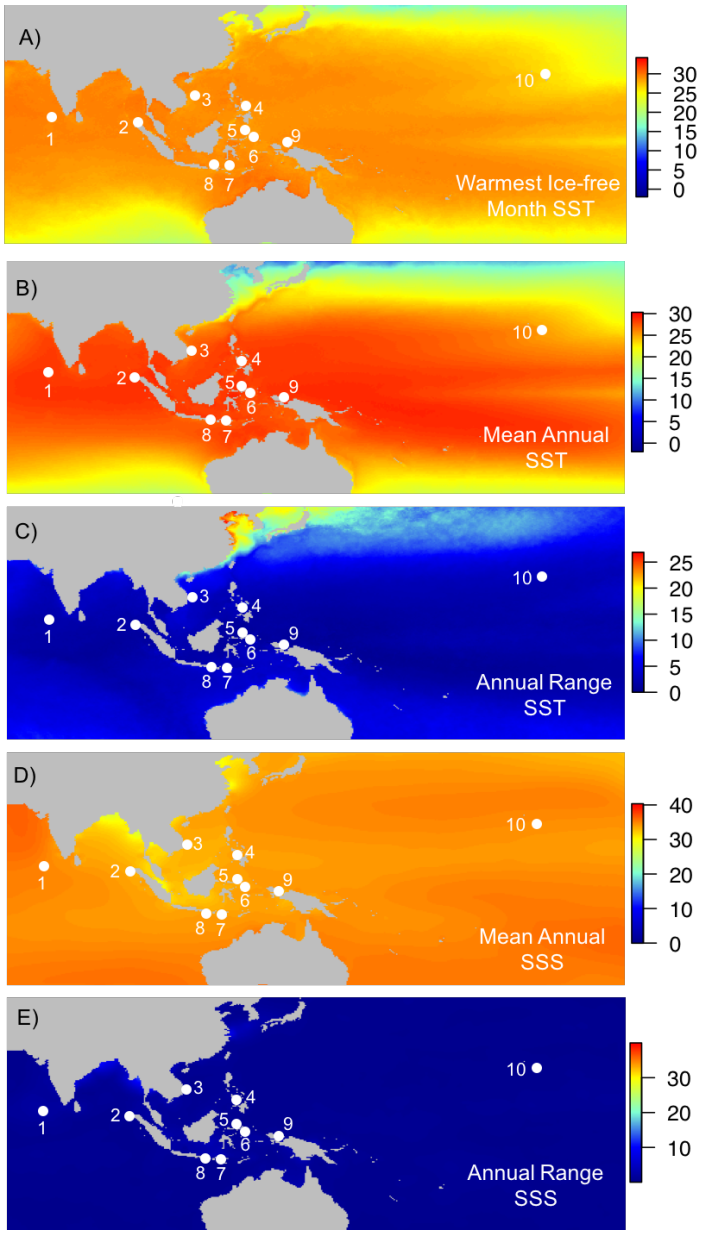


Figure 3.5 Ocean climate variables **a)** temperature of the warmest ice-free month, **b)** annual mean, and **c)** range of sea surface temperature (SST) in ˚C, **d)** annual mean, and **e)** range of sea surface salinity (SSS) in psu. Data source MARSPEC (Sbrocco & Barber, 2013).

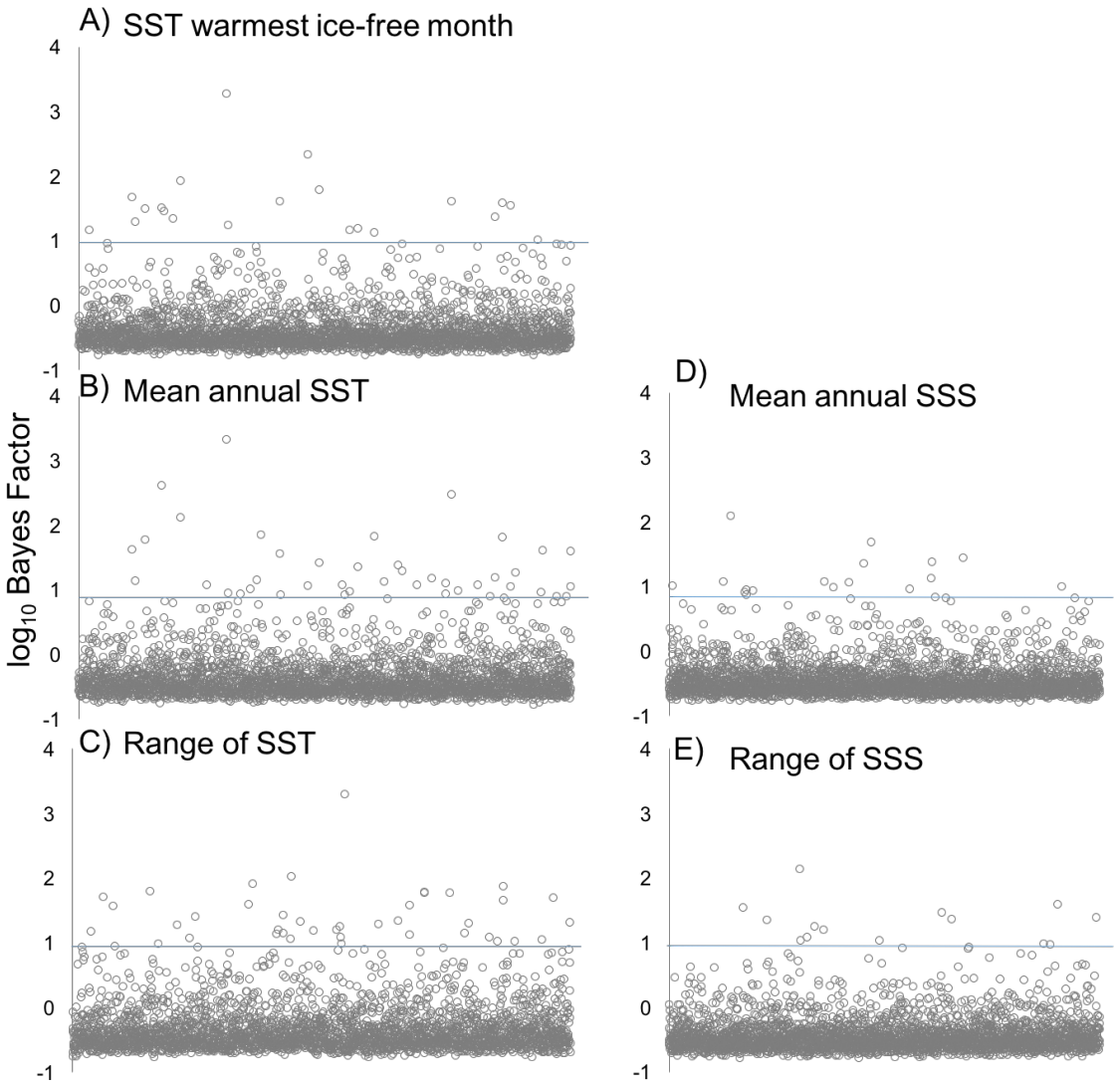


Figure 3.6 Bayenv association of allele frequencies for all polymorphic loci (3,186) with each ocean climate variable. SNPs are ordered by tagID. Blue line indicates significance level above which loci are considered to have a strong association with climate variable (Jeffereys 1961).

APPENDIX I

Table SI 3.1 Annotations for BayeScan outlier loci among sites and Bayenv candidate loci that mapped to DNA sequences in the NCBI database.

BAYESCAN

Tag	Sequence	Organism	Description	Score	E-value	Identity	Gene	GO functions
27523	CGTATGCCGTCTTCTGCTTG AAAAAAAAAAAAATAC	*Conus episcopatus*	conotoxin gene	55.4	7.00E-08	91%		conotoxin
29969	AGGACCTGCTACGCAAGCTG CTGCAGGTGGACCTCA	*Aplysia californica*	CAPL-A	53.6	2.00E-07	94%	CAPLA	ATP binding, protein serine/threonine kinase activity, behavior plasticity, learning and memory
19005	GGAGAGTGTGATGCAGCCAT CTGCAGTGTCTTTGCA	*Coralliophila abbreviata*	microsatellite	37.4	0.019	97%		
18735	TGAGATGATCTTGCAACCAC ATGCCTCCATCCCTCT	*Aplysia californica*	striatin-interacting protein 1-like	35.6	0.066	94%	STRIP1	protein binding, regulation of cell morphogenesis, cortical actin cytoskeleton organization
19841	TGCTTCTGTGGTGCAGGCCA GTGCAAGACCTCAGTC	*Biomphalaria glabrata*	zinc finger ZZ-type and EF-hand domain-containing protein 1-like	31.9	0.8	91%	ZZEF1	calcium and zinc ion binding
30344	CACTGTGCCAAGGCATGGTA TTGCAAGGGGGAGCCC	*Aplysia californica*	E3 ubiquitin-protein ligase DTX4-like	30.1	2.8	86%	DTX4 E3	ligase activity, zinc ion binding
36204	AGATCAACAGGGGCATCTCT TTGCTCGGTAAAGTCC	*Aplysia californica*	protein FAM193A-like	30.1	2.8	86%	FAM193A	uncharacterized protein
32125	AGTAATATGAGTGCATTTAGT TGCAGGAGTCGTGTG	*Aplysia californica*	epidermal retinol dehydrogenase 2-like	30.1	2.8	86%	SDR16C5	detection of light stimulus involved in visual perception, retinol dehydrogenase activity, keratinocyte proliferation, retinal and retinol metabolism
26718	AGCAGGACAATGCCATACCT TGAAGCAGGACAATGC	*Biomphalaria glabrata*	chitin biosynthesis protein CHS5-like	28.3	9.7	94%	CHS5	chitin biosynthesis
36113	GACAGCCATCTCGCATAGTT CTGCTGTTCCAGAGCA	*Crassostrea gigas*	complete sequence	33.7	0.39	86%		
53661	AGTCAAAGCATTGCAGATGT CTGCTATCAAACTTCT	*Lottia gigantea*	hypothetical protein	33.7	0.23	91%		
40104	TATATAGAACAAGCAGCAAG ATGCACAAGAGAAGCT	*Lottia gigantea*	hypothetical protein	31.9	0.8	88%		
36681	TAGTCATCAACAGCATTACA CTGCCATCATTACATG	*Biomphalaria glabrata*	uncharacterized	30.1	2.8	86%		
48982	GGCTGATGGCTTGCATTATT CTGCTTTATACTTCCT	*Bellamya aeruginosa*	microsatellite	30.1	2.8	86%		
30094	CACACACCAGAGGCAGGCA CGTGCACGAGCATACCT	*Biomphalaria glabrata*	uncharacterized	30.1	2.8	94%		

BAYEVN

Tag	Sequence	Organism	Description	Score	E-value	Identity	Gene	GO functions
31366	AGATCGGAAGAGCACATGTC TGAACTCCAGTCACGA	*Conus episcopatus*	conotoxin	53.6	4.00E-07	88%		conotoxin
32340	TGTGTTTTTCTTGCATCGCAA TGCATTGTCAATGAG	*Aplysia californica*	rho GTPase-activating protein 17-like	37.4	0.032	91%	ARHGAP17	GTPase activator activity, signal transduction
14160	TGCTCTGAAGGCGCATTGTT TTGCATTTAATTGACC	*Biomphalaria glabrata*	ubiquitin carboxyl-terminal hydrolase 44-like	33.7	0.39	86%	USP44	cell division, protein deubiquitination, mitotic nuclear division
31520	CGCACGAACTACGCACACTC ATGCATACAAATCATC	*Octopus bimaculoides*	little elongation complex subunit 2-like	33.7	0.39	88%	ICE2	snRNA transcription regulation
11426	CGTTTAGTTTTAGCAGGAGC ATGCTTGTCAAAGATA	*Haliotis discus hannai*	microsatellite	33.7	0.39	86%		
63145	CAAGTTCCAATAGCATTTCA CTGCTTGATGATGCCC	*Octopus bimaculoides*	uncharacterized	33.7	0.39	91%		
35977	TGTGGATGCTTGGCAACTCG GTGCGACAGAGAATCT	*Aplysia californica*	uncharacterized	33.7	0.39	94%		
26516	GGCGTTCTGTATGCACTTTC TTGCTTCAGTTCTCCT	*Biomphalaria glabrata*	coiled-coil domain-containing protein 94-like	31.9	1.4	88%	Ccdc94	protein binding
30745	GGGCTCTGTCTCGCAATACC GTGCTCAGATAAAGTG	*Octopus bimaculoides*	A-kinase anchor protein 10, mitochondrial-like	31.9	1.4	88%	AKAP10	blood coagulation, signal transduction, protein localization
19751	AGGTCTTCCCTGGCACTGAT ATGCCAAGCATCCACA	*Octopus bimaculoides*	glutamate receptor 2-like	31.9	1.4	88%	GRIA2	glutamate receptor activity, learning, memory and behavior
27006	GGTGGGGTCGCTGCAATTCT TTGCGGATCAGGTCTG	*Crassostrea gigas*	uncharacterized	31.9	1.4	88%		
17278	AGGTCATCTTGTGCATAGAC CTGCCAGTGCCTTACT	*Rapana venosa*	microsatellite	31.9	1.4	94%		
13973	TGTTCACTGACGGCAAAAGT ATGCAAAATGAGGTCA	*Biomphalaria glabrata*	DNA mismatch repair protein Msh2-like	30.1	4.8	94%	MSH2	embryonic development, meiotic mismatch repair, DNA repair, oxidative phosphorylation, immune response
23643	TGTTTTGGGGGAGCATGTCC CTGCACACATACCCCT	*Crassostrea gigas*	calcium-transporting ATPase type 2C member 1-like	30.1	4.8	86%	ATP2C1	calcium and manganese ion transport, metabolic process, signal transduction
33972	TGGGACAGGATTGCAACTTC TTGCCTCAACAGATCT	*Crassostrea gigas*	ankyrin repeat domain-containing protein 24-like	30.1	4.8	94%	ANKRD24	protein binding
33093	TAAGGAATGACAGCAAACCA ATGCTACCTCTCCTCT	*Octopus bimaculoides*	exonuclease 1-like	30.1	4.8	100%	EXO1	immune response, DNA repair, mismatch repair, double-strand break repair
31348	TGGGAATGTAAAGCAAGATA CTGCTTTTTGTGCGCT	*Biomphalaria glabrata*	uncharacterized	30.1	4.8	86%		
33978	AGTTTGAATGGGGCAGTCAC GTGCTAAGGTGGTGTC	*Octopus bimaculoides*	uncharacterized	30.1	4.8	91%		
15186	CACCCTCTCTATGCAATGAC ATGCAAGCCCCCCTCT	*Aplysia californica*	uncharacterized	30.1	4.8	86%		

3.8 REFERENCES

Ackiss A.S., Pardede S., Crandall E.D., Ablan-Lagman M.C.A., Romena N., Barber P.H., & Carpenter K.E. (2013) Pronounced genetic structure in a highly mobile coral reef fish, *Caesio cuning*, in the Coral Triangle. *Marine Ecology Progress Series*, **480**, 185–197.

Ahti P.A., Coleman R.R., DiBattista J.D., Berumen M.L., Rocha L.A., & Bowen B.W. (2016) Phylogeography of Indo-Pacific reef fishes: sister wrasses *Coris gaimard* and *C. cuvieri* in the Red Sea, Indian Ocean and Pacific Ocean. *Journal of Biogeography*, doi:10.1111/jbi.12712.

Barber P.H., Cheng S.H., Erdmann M.V., Tenggardjaja K., & Ambariyanto (2011) Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. *Phylogeography and population genetics in Crustacea. Crustean Issues* (ed. by C. Held, S. Koenemann, and C.D. Schubart), pp. 129–156. CRC Press, Boca Raton.

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

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

Baums I.B., Boulay J.N., Polato N.R., & Hellberg M.E. (2012) No gene flow across the Eastern Pacific Barrier in the reef-building coral *Porites lobata*. *Molecular Ecology*, **21**, 5418–5433.

Benzie J.A.H. & Williams S.T. (1997) Genetic structure of giant clam (*Tridacna maxima*) populations in the West Pacific is not consistent with dispersal by present-day ocean currents. *Evolution*, **51**, 768–783.

Bowen B.W., Rocha L.A., Toonen R.J., & Karl S.A. (2013) The origins of tropical marine biodiversity. *Trends in Ecology & Evolution*, **1666**, 1–8.

Bowen B.W., Shanker K., Yasuda N., Celia M., Malay M.C.M.D., von der Heyden S., Paulay G., Rocha L.A., Selkoe K.A., Barber P.H., Williams S.T., Lessios H.A., Crandall E.D., Bernardi G., Meyer C.P., Carpenter K.E., & Toonen R.J. (2014) Phylogeography unplugged: comparative surveys in the genomic era. *Bulletin of Marine Science*, **90**, 13–46.

Briggs J.C. (1992) The marine East Indies: centre of origin? *Global Ecology and Biogeography Letters*, **2**, 149–156.

Briggs J.C. (2005) The marine East Indies: diversity and speciation. *Journal of Biogeography*, **32**, 1517–1522.

Briggs J.C. (2006) Proximate sources of marine biodiversity. *Journal of Biogeography*, **33**, 1–10.

Burgess S.C., Treml E.A., & Marshall D.J. (2012) How do dispersal cost and habitat selection influence realized population connectivity? *Ecology*, **93**, 1–32.

Carpenter K.E., Barber P.H., Crandall E.D., Ablan-Lagman M.C.A., Mahardika G.N., Manjaji-Matsumoto B.M., Juinio-Meñez M.A., Santos M.D., Starger C.J., & Toha A.H.A. (2011) Comparative phylogeography of the Coral Triangle and implications for marine management. *Journal of Marine Biology*, **2011**, 1–14.

Cheng S.H. (2015) *Evolution and population genomics of Loliginid squids.* Ph.D. Dissertation, University of California, Los Angeles.

Coop G., Witonsky D., Di Rienzo A., & Pritchard J.K. (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**, 1411–1423.

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

Crandall E.D., Frey M.A., Grosberg R.K., & Barber P.H. (2008a) Contrasting demographic history and phylogeographical patterns in two Indo-Pacific gastropods. *Molecular Ecology*, **17**, 611–626.

Crandall E.D., Jones M.E., Muñoz M.M., Akinronbi B., Erdmann M.V., & Barber P.H. (2008b) Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. *Molecular Ecology*, **17**, 5276–5290.

Crawford C.M., Nash W.J., & Lucas J.S. (1986) Spawning induction, and larval and juvenile rearing of the giant clam, *Tridacna gigas*. *Aquaculture*, **58**, 281–295.

DeBoer T.S., Baker A.C., Erdmann M.V., Ambariyanto, Jones P.R., & Barber P.H. (2011) Patterns of Symbiodinium distribution in three giant clam species across the biodiverse Bird's Head region of Indonesia. *Marine Ecology Progress Series*, **444**, 117–132.

DeBoer T.S., Naguit M.R.A., Erdmann M.V., Ablan-Lagman M.C.A., Ambariyanto, Carpenter K.E., Toha A.H.A., & Barber P.H. (2014) Concordance between phylogeographic and biogeographic boundaries in the Coral Triangle: Conservation implications based on comparative analyses of multiple giant clam species. *Bulletin of Marine Science*, **90**, 277–300.

DeBoer T.S., Subia M.D., Ambariyanto, Erdmann M.V., Kovitvongsa K., & Barber P.H. (2008) Phylogeography and limited genetic connectivity in the endangered boring giant clam across the Coral Triangle. *Conservation Biology*, **22**, 1255–1266.

Defaveri J., Jonsson P.R., & Merilä J. (2013) Heterogeneous genomic differentiation in marine threespine sticklebacks: adaptation along an environmental gradient. *Evolution*, **67**, 2530–2546.

Demond J. (1957) Micronesian reef-associated gastropods. *Pacific Science*, **XI**, 275–341.

DiBattista J.D., Roberts M., Bouwmeester J., Bowen B.W., Coker D., Lozano-Cortés D., Choat J.H., Gaither M.R., Hobbs J.-P.A., Khalil M.T., Kochzius M., Myers R.F., Paulay G.,

Robitzch V., Rocha L.A., Saenz-Agudelo P., Salas E., Sinclair-Tailor T.H., Toonen R.J., Weastneat M., Williams S.T., & Berumen M.L. (2016) A review of contemporary patterns of endemism in the Red Sea. *Journal of Biogeography*, **43**, 423–439.

Dibattista J.D., Rocha L.A., Craig M.T., Feldheim K.A., & Bowen B.W. (2012) Phylogeography of two closely related Indo-Pacific butterflyfishes reveals divergent evolutionary histories and discordant results from mtDNA and microsatellites. *Journal of Heredity*, **103**, 617–629.

Dohna T.A., Timm J., Hamid L., & Kochzius M. (2015) Limited connectivity and a phylogeographic break characterize populations of the pink anemonefish, *Amphiprion perideraion*, in the Indo-Malay Archipelago: inferences from a mitochondrial and microsatellite loci. *Ecology and Evolution*, **5**, 1717–1733.

Drew J. & Barber P.H. (2009) Sequential cladogenesis of the reef fish *Pomacentrus moluccensis* (Pomacentridae) supports the peripheral origin of marine biodiversity in the Indo-Australian archipelago. *Molecular Phylogenetics and Evolution*, **53**, 335–339.

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

Excoffier E. & Lischer H. (2011) ARLEQUIN: An integrated software package for population genetics data analysis. *What's New*, 1–174.

Fange R. (2012) Venoms and venom glands of marine molluscs. *Toxins, drugs and pollutants in marine animals* (ed. by L. Bolis, J. Zadunaisky, and R. Gilles), pp. 47–62. Springer, Berlin.

Fernandez-Silva I., Randall J.E., Coleman R.R., Dibattista J.D., Rocha L.A., Reimer J.D., Meyer C.G., & Bowen B.W. (2015) Yellow tails in the Red Sea: Phylogeography of the Indo-Pacific goatfish *Mulloidichthys flavolineatus* reveals isolation in peripheral provinces and cryptic evolutionary lineages. *Journal of Biogeography*, **42**, 2402–2413.

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

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

Foll M. & Gaggiotti O. (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–91.

Foll M. & Gaggiotti O. (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, **180**, 977–93.

Forsman Z., Wellington G.M., Fox G.E., & Toonen R.J. (2015) Clues to unraveling the coral species problem: distinguishing species from geographic variation in *Porites* across the Pacific with molecular markers and microskeletal traits. *PeerJ*, **3**, e751.

Forsman Z.H., Barshis D.J., Hunter C.L., & Toonen R.J. (2009) Shape-shifting corals: molecular markers show morphology is evolutionarily plastic in *Porites*. *Evolutionary Biology*, **9**, 45–54.

Gaither M.R., Bernal M.A., Coleman R.R., Bowen B.W., Jones S.A., Simison W.B., & Rocha L.A. (2015) Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Molecular Ecology*, **24**, 1543–1557.

Gaither M.R., Bowen B.W., Bordenave T., Rocha L.A., Newman S.J., Gomez J.A., van Herwerden L., & Craig M.T. (2011) Phylogeography of the reef fish *Cephalopholis argus* (Epinephelidae) indicates Pleistocene isolation across the Indo-Pacific Barrier with contemporary overlap in the Coral Triangle. *Evolutionary Biology*, **11**, 189–203.

Gaither M.R., Toonen R.J., Robertson D.R., Planes S., & Bowen B.W. (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.

Gavrilets S. (2003) Perspective: models of speciation: what have we learned in 40 years? *Evolution*, **57**, 2197–215.

Gordon A.L. & Fine R.A. (1996) Pathways of water between the Pacific and Indian oceans in the Indonesian seas. *Nature*, **379**, 146–149.

Gordon A.L., Susanto R.D., & Vranes K. (2003) Cool Indonesian throughflow as a consequence of restricted surface layer flow. *Science*, **425**, 21–25.

Günther T. & Coop G. (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, **195**, 205–220.

Hoegh-Guldberg O. (1999) Climate change, coral bleaching and the future of the world's coral reefs. *Marine and Freshwater Research*, **50**, 839–866.

Hoeksema B. (2007) Delineation of the Indo-Malayan centre of maximum marine biodiversity: the Coral Triangle. *Biogeography, Time, and Place: Distributions, Barriers, and Islands* (ed. by W. Renema), pp. 117–178. Springer, Netherlands.

Horne J.B. (2014) Thinking outside the barrier: neutral and adaptive divergence in Indo-Pacific coral reef faunas. *Evolutionary Ecology*, doi: 10.1007/s10682–014–9724–9.

Jackson A., Ambariyanto, Erdmann M., Toha A., Stevens L., & Barber P. (2014) Phylogeography of commercial tuna and mackerel in the Indonesian Archipelago. *Bulletin of Marine Science*, **90**, 471–492.

Jefferys H. (1961) *Theory of Probability.* Oxford University Press, United Kingdom.

Johannesson K. & Andre C. (2006) Life on the margin: Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, **15**, 2013–2029.

Kawecki T.J. (2008) Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics*, **39**, 321–342.

Kelly R.P. & Palumbi S.R. (2010) Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS ONE*, **5**, e8594.

Kirkendale L.A. & Meyer C.P. (2004) Phylogeography of the *Patelloida profunda* group (Gastropoda: Lottidae): diversification in a dispersal-driven marine system. *Molecular Ecology*, **13**, 2749–2762.

Knowlton N. & Keller B.D. (1986) Larvae which fall far short of their potential: highly localized recruitment in an alpheid shrimp with extended larval development. *Bulletin of Marine Science*, **39**, 213–223.

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

Kochzius M. & Nuryanto A. (2008) Strong genetic population structure in the boring giant clam, *Tridacna crocea*, across the Indo-Malay Archipelago: implications related to evolutionary processes and connectivity. *Molecular Ecology*, **17**, 3775–3787.

Kochzius M., Seidel C., & Hauschild J. (2009) Genetic population structures of the blue starfish *Linckia laevigata* and its gastropod ectoparasite *Thyca crystallina*. *Marine Ecology Progress Series*, **396**, 211–219.

Kool J.T., Paris C.B., Barber P.H., & Cowen R.K. (2011) Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. *Global Ecology and Biogeography*, **20**, 695–706.

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

Kuhnt W., Holbourn A., & Hall R. (2004) Neogene history of the Indonesian Throughflow. *Geophysical Monograph Series*, **149**, 299–320.

Leray M., Beldade R., Holbrook S.J., Schmitt R.J., Planes S., & Bernardi G. (2010) Allopatric divergence and speciation in coral reef fish: the three-spot dascyllus, *Dascyllus trimaculatus*, species complex. *Evolution*, **64**, 1218–30.

Limborg M.T., Helyar S.J., De Bruyn M., Taylor M.I., Nielsen E.E., Ogden R., Carvalho G.R., & Bekkevold D. (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, **21**, 3686–703.

Lin T. & Liu P. (1995) Fecundity of female coral-inhabiting snails, *Coralliophila violacea* (Gastropoda: Coralliophilidae). *Veliger*, **38**, 319–322.

Lin T.-Y. & Liu L.-L. (2008) Low levels of genetic differentiation among populations of the coral-inhabiting snail *Coralliophila violacea* (Gastropoda: Coralliophilidae) in regions of the Kurishio and South China Sea. *Zoological Studies*, **47**, 17–24.

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

Liu S.-Y.V., Chang F.-T., Borsa P., Chen W.-J., & Dai C.-F. (2014) Phylogeography of the humbug damselfish, *Dascyllus aruanus* (Linnaeus, 1758): evidence of Indo-Pacific vicariance and genetic differentiation of peripheral populations. *Biological Journal of the Linnean Society*, **113**, 931–942.

Longo G. & Bernardi G. (2015) The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. *Molecular Phylogenetics and Evolution*, **88**, 55–63.

Lourie S. & Vincent A. (2004) A marine fish follows Wallace's Line: the phylogeography of the three-spot seahorse (*Hippocampus trimaculatus*, Syngnathidae, Teleostei) in Southeast Asia. *Journal of Biogeography*, **31**, 1975–1985.

Lourie S.A., Green D.M., & Vincent A.C.J. (2005) Dispersal, habitat differences, and comparative phylogeography of Southeast Asian seahorses (Syngnathidae: *Hippocampus*). *Molecular Ecology*, **14**, 1073–1094.

Ludt W.B. & Rocha L.A. (2015) Shifting seas: the impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, **42**, 25–38.

Meyer C.P., Geller J.B., & Paulay G. (2005) Fine scale endemism on coral reefs: archipelagic differentiation in turbinid gastropods. *Evolution*, **59**, 113–125.

Miglietta M.P., Faucci A., & Santini F. (2011) Speciation in the sea: overview of the symposium and discussion of future directions. *Integrative and Comparative Biology*, 1–7.

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

Palumbi S.R. (1994) Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics*, **25**, 547–572.

Polato N.R., Concepcion G.T., Toonen R.J., & Baums I.B. (2010) Isolation by distance across the Hawaiian Archipelago in the reef-building coral *Porites lobata*. *Molecular Ecology*, **19**, 4661–4677.

Pritchard J.K., Stephens M., & Donnelly P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Puebla O. (2009) Ecological speciation in marine v. freshwater fishes. *Journal of Fish Biology*, **75**, 960–996.

Rainer M., Hild S., Walther P., Ploss K., Boland W., & Tomaschko K.-H. (2007) Granular chitin in the epidermis of nudibranch molluscs. *Biological Bulletin*, 307–315.

Raynal J.M., Crandall E.D., Barber P.H., Mahardika G.N., Lagman M.C., & Carpenter K.E. (2014) Basin isolation and oceanographic features influencing lineage divergence in the humbug damselfish (*Dascyllus aruanus*) in the Coral Triangle. *Bulletin of Marine Science*, **90**, 513–532.

Reid D.G., Lal K., Mackenzie-Dodds J., Kaligis F., Littlewood D.T.J., & Williams S.T. (2006) Comparative phylogeography and species boundaries in *Echinolittorina* snails in the central Indo-West Pacific. *Journal of Biogeography*, **33**, 990–1006.

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

Saenz-Agudelo P., DiBattista J.D., Piatek M.J., Gaither M.R., Harrison H.B., Nanninga G.B., & Berumen M.L. (2015) Seascape genetics along environmental gradients in the Arabian Peninsula: insights from ddRAD sequencing of anemonefishes. *Molecular Ecology*, **24**, doi: 10.1111/mec.13471.

Sanford E. & Kelly M.W. (2011) Local adaptation in marine invertebrates. *Annual Review of Marine Science*, **3**, 509–535.

Sbrocco E.J. (2012) *A seascape genetics approach to exploring the phylogeographic response of marine fishes to late quaternary climate change.* Ph.D. Dissertation, Boston University.

Sbrocco E.J. & Barber P.H. (2013) MARSPEC: ocean climate layers for marine spatial ecology. *Ecology*, **94**, 979.

Scheltema R.S. (1971) Larval dispersal as a means of genetic exchange between geographically separated populations of shallow-water benthic marine gastropods. *Biological Bulletin*, **140**, 284.

Shanks A.L. (2009) Pelagic larval duration and dispersal distance revisited. *Biological Bulletin*, **216**, 373–85.

Shinoda T., Han W., Metzger E.J., & Hurlburt H.E. (2012) Seasonal variation of the Indonesian Throughflow in Makassar Strait. *Journal of Physical Oceanography*, **42**, 1099–1123.

Slatkin M. (1993) Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, **47**, 264–279.

Soong K. & Chen J.L. (1991) Population structure and sex-change in the coral-inhabiting snail *Coralliophila violacea* at Hsiao-Liuchiu, Taiwan. *Marine Biology*, **111**, 81–86.

Szabó Z., Snelgrove B., Craig M.T., Rocha L.A., & Bowen B.W. (2014) Phylogeography of the manybar goatfish, *Parupeneus multifasciatus*, reveals isolation of the Hawaiian Archipelago and a cryptic species in the Marquesas Islands. *Bulletin of Marine Science*, **90**, 493–512.

Taylor J. (1975) *Planktonic prosobranch veligers of Kaneohe Bay.* Ph.D. Dissertation, University of Hawaii.

Taylor M.S. & Hellberg M.E. (2003) Genetic evidence for local retention of pelagic larvae in a Caribbean reef fish. *Science*, **299**, 107–109.

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

Tornabene L., Valdez S., Erdmann M., & Pezold F. (2015) Support for a 'Center of Origin' in the Coral Triangle: Cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: Eviota). *Molecular Phylogenetics and Evolution*, **82**, 200–210.

Treml E.A., Roberts J., Halpin P.N., Possingham H.P., & Riginos C. (2015) The emergent geography of biophysical dispersal barriers across the Indo-West Pacific. *Diversity and Distributions*, **21**, 465–476.

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

Waldrop E., Hobbs J.-P.A., Randall J.E., DiBattista J.D., Rocha L.A., Kosaki R.K., Berumen M.L., & Bowen B.W. (2016) Phylogeography, population structure and evolution of coral-eating butterflyfishes (Family Chaetodontidae, genus *Chaetodon*, subgenus *Corallochaetodon*). *Journal of Biogeography*, **43**, 1–14.

Wang S., Meyer E., McKay J.K., & Matz M.V. (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, **9**, 808–810.

Wares J.P. (2002) Community genetics in the Northwestern Atlantic intertidal. *Molecular Ecology*, **11**, 1131–44.

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

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

Wright S. (1943) Isolation by distance. *Genetics*, **28**, 114–38.