Title
Genomic signatures of host-associated divergence and adaptation in a coral-eating snail, Coralliophila violacea (Kiener, 1836).

Permalink
https://escholarship.org/uc/item/2vg5526v

Journal
Ecology and evolution, 10(4)

ISSN
2045-7758

Authors
Simmonds, Sara E
Fritts-Penniman, Allison L
Cheng, Samantha H
et al.

Publication Date
2020-02-05

DOI
10.1002/ece3.5977

Peer reviewed
Genomic signatures of host-associated divergence and adaptation in a coral-eating snail, *Coralliophila violacea* (Kiener, 1836)

Sara E. Simmonds¹ | Allison L. Fritts-Penniman¹ | Samantha H. Cheng¹,² | Gusti Ngurah Mahardika³ | Paul H. Barber¹

¹Department of Ecology and Evolutionary Biology, University of California Los Angeles, Los Angeles, CA, USA
²Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY, USA
³Animal Biomedical and Molecular Biology Laboratory, Faculty of Veterinary Medicine, Udayana University Bali, Denpasar, Indonesia

**Abstract**

The fluid nature of the ocean, combined with planktonic dispersal of marine larvae, lowers physical barriers to gene flow. However, divergence can still occur despite gene flow if strong selection acts on populations occupying different ecological niches. Here, we examined the population genomics of an ectoparasitic snail, *Coralliophila violacea* (Kiener 1836), that specializes on *Porites* corals in the Indo-Pacific. Previous genetic analyses revealed two sympatric lineages associated with different coral hosts. In this study, we examined the mechanisms promoting and maintaining the snails’ adaptation to their coral hosts. Genome-wide single nucleotide polymorphism (SNP) data from type II restriction site-associated DNA (2b-RAD) sequencing revealed two differentiated clusters of *C. violacea* that were largely concordant with previous genetic results. However, the presence of some admixed genotypes indicates gene flow from one lineage to the other. Combined, these results suggest that differentiation between host-associated lineages of *C. violacea* is occurring in the face of ongoing gene flow, requiring strong selection. Indeed, 2.7% of all SNP loci were outlier loci (73/2,718), indicative of divergence with gene flow, driven by adaptation of each *C. violacea* lineage to their specific coral hosts.

**KEYWORDS**

adaptation, coral reefs, ecological divergence, gastropods, population genomics, RAD-seq

1 | INTRODUCTION

While ecological speciation has been documented for almost three decades across a wide variety of organisms on land (Case & Willis, 2008; Feder et al., 1994; Jiggins, 2008; Martin et al., 2013; Schluter, 2009; Seehausen et al., 2008; Sorenson, Sefc, & Payne, 2003; Thorpe, Surget-Groba, & Johansson, 2010; Waser & Campbell, 2004) and in freshwater (Hatfield & Schluter, 1999; Langerhans, Gifford, & Joseph, 2007; Puebla, 2009; Seehausen et al., 2008; Seehausen & Wagner, 2014), ecological speciation in the ocean was thought to be rare, and only recently has that viewpoint begun to change (Bird, Fernandez-Silva, Skillings, & Toonen, 2012; Bird, Holland, Bowen, & ToBo Laboratory, 2013; Foote & Morin,
There are a number of reasons for this reassessment. First, absolute physical barriers in the sea are exceedingly rare (Ludt & Rocha, 2015; Rocha & Bowen, 2008; Rocha et al., 2005). As a result, speciation must often proceed with varying levels of gene flow and aided by divergent selection in different habitats or hosts (Palumbi, 1994). Second, the strong interspecific interactions that can promote ecological speciation in terrestrial species (e.g., host–parasite, mutualisms) are also common in certain marine ecosystems (Blackall, Wilson, & van Oppen, 2015; Stella, Jones, & Pratchett, 2010). For example, reef-building corals have tight ecological associations with a wide variety of invertebrate taxa (Zann, 1987), including ~900 named species of sponges, copepods, barnacles, crabs, shrimp, worms, bivalves, nudibranchs, and snails (reviewed by Stella et al., 2010). This wide array of symbiotic relationships creates tremendous potential for host shifting and the development of host specificity that can lead to sympatric speciation.

While encouraging, there are gaps in our knowledge that with the expansion of genomic technologies, we are now in a position to begin to fill. Detecting signatures of natural selection in populations where there is likely ongoing gene flow is now possible using genome-wide data, lending insight into the mechanisms of ecological speciation (Bernal, Gaither, Simison, & Rocha, 2017; Campbell, Poelstra, & Yoder, 2018; Puebla, Bermingham, & McMillan, 2014; Westram et al., 2018). To date, however, no studies examining the genomic signatures of ecological divergence in marine host–parasite systems have been conducted.

The ~6 million km² Coral Triangle region is home to over 500 species of reef-building corals (Veron et al., 2011) and thousands of unique species of fishes and invertebrates (Barber & Boyce, 2006; Briggs, 2003), making it the global center of marine biodiversity (Cowman & Bellwood, 2011; Hoeksema, 2007). Most of the literature examining the evolution of this biodiversity hotspot has focused on allopatric processes such as divergence across geological and oceanographic features such as the Sunda Shelf or Halmahera Eddy during Pleistocene low sea levels stands (for reviews, see Barber, Cheng, Erdmann, Tenggardjaja, & Ambaryanto 2011; Carpenter et al., 2011; Gaither & Rocha, 2013). Allopatric divergence is clearly an important factor in the biodiversity of the Coral Triangle. However, the extraordinary diversity in this region, combined with the prevalence of strong species–species interactions on coral reefs, makes it likely that ecological speciation also contributes to the evolution of biodiversity in this hotspot.

The marine snail, Coralliophila violacea (Figure 1), is an obligate ectoparasite, living, feeding, and reproducing exclusively on corals in Poritidae, a highly abundant and diverse coral family (Kitahara, Cairns, Stolarski, Blair, & Miller, 2010), which is found in shallow reefs across the tropical Indo-Pacific. The snails attach themselves to their host, form feeding aggregations, and drain energy from their host as it tries to repair damaged tissues (Oren, Brickner, &...
Loya, 1998). They are sequential hermaphrodites, a common trait of parasitic mollusks (Heller, 1993), and breed with conspecifics on their host coral colony. Two genetically distinct lineages of *C. violacea* occur sympatrically on reefs of the Coral Triangle, but each lineage occupies one of two groups of *Porites* corals, suggesting ecological divergence (Simmonds et al., 2018). A lack of evidence of genetic structure within each lineage of *C. violacea* inside the Coral Triangle precludes physical isolation as an explanation for the observed divergence. Host specificity commonly results from preferential larval settlement (Ritson-Williams, Shjegstad, & Paul, 2003, 2007, 2009). This genetic evidence combined with observations of adult preference for specific coral hosts (unpubl. data S. Simmonds) strongly suggests ecological divergence driven by host association.

To determine where diverging populations of *C. violacea* lie on the continuum of the speciation process (i.e., host-associated lineages, sibling species or good species), it is important to examine patterns of realized gene flow between the divergent coral host-associated lineages. Effective contemporary gene flow should result in linkage disequilibria between host-associated marker loci in populations utilizing different hosts. However, if lower rates of gene flow (<1% per generation) are found, then populations should be considered incipient species (Drès & Mallet, 2002; Malausa et al., 2007).

**TABLE 1**  *Coralliophila violacea* collection locations, latitude, longitude, coral host species, and number of samples collected

| Location          | Country       | Province          | Latitude  | Longitude  | *Porites lobata* | *Porites cylindrica* |
|-------------------|---------------|-------------------|-----------|------------|-------------------|----------------------|
| 1. Pemuteran      | Indonesia     | Bali              | −8.1400   | 114.6540   |                   |                      |
| 2. Nusa Penida    | Indonesia     | Bali              | −8.6750   | 115.5130   | 11                 | 10                   |
| 3. Pulau Mengyatan| Indonesia     | East Nusa Tenggara| −8.5570   | 119.6850   | 4                  | 3                    |
| 4. Lembeh          | Indonesia     | North Sulawesi    | 1.4790    | 125.2510   | 7                  | 1                    |
| 5. Bunaken         | Indonesia     | North Sulawesi    | 1.6120    | 124.7830   | 9                  | 6                    |
| 6. Dumaguete      | Philippines   | Negros Oriental   | 9.3220    | 123.3120   | 2                  | 7                    |
| Total N           |               |                   |           |            | 33                 | 34                   |

**FIGURE 2** Collection locations 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. Map made with vector and raster map data available at naturalearthdata.com
Genomic tests of selection are key to distinguishing between these possibilities. If divergence among C. violacea lineages results purely from neutral processes, genetic drift and migration should have approximately equal effects on all parts of the genome (Nielsen, 2005), and frequencies of neutral loci should show similar levels of differentiation (Via, 2009). However, if divergent selection is driving diversification of C. violacea lineages, there should be clear signatures of divergent selection (Feder et al., 1994; Nosil, Funk, & Ortiz-Barrientos, 2009), because natural selection affects non-neutral parts of the genome, as well as linked loci, to a greater extent (Smith & Haigh, 1974). As such, 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).

In this study, we use genome-wide single nucleotide polymorphisms (SNPs) to investigate the possibility of ecological divergence with gene flow in populations of a corallivorous gastropod, C. violacea, from the Coral Triangle. Specifically, we (a) test for reduced gene flow between sympatric lineages of host-associated snails, (b) identify outlier loci under putative selection between hosts, and (c) annotate possible functions of linked genes that might be necessary for adaptation to hosts.

2 | MATERIALS AND METHODS

2.1 | Sample collection

We collected snails on snorkel during 2011–2013 from six sympatric populations of two lineages of C. violacea representing unique parasite–host groups (Table 1, Figure 2, Appendix S1). We chose snails from the most abundant Porites species from each group (P. lobata, P. cylindrica, Dana, 1846, Figure 1) to maximize the number of samples and reduce potentially confounding effects of differences among hosts within the same group. To further reduce confounding effects resulting from taxonomic complexity within P. lobata (Forsman, Barshis, Hunter, & Toonen, 2009; Prada et al., 2014), we used coral host species identifications from Simmonds et al. (2018) that were confirmed through RAD-seq data.

2.2 | Creation of RAD libraries

We extracted genomic DNA from 20 mg of foot tissue from 67 individual C. violacea (34 from P. cylindrica and 33 from P. lobata; Table 1) using a DNeasy® Blood and Tissue Kit (QIAGEN), following manufacturer’s instructions, save for elution of DNA with molecular grade H₂O rather than AE buffer. We estimated initial DNA concentrations using a NanoDrop™ 2000 Spectrophotometer (Thermo Scientific™) and visualized DNA quality on a 1% agarose gel stained with SYBR® Safe DNA Gel Stain (Invitrogen™). We used only high-quality DNA with a bright high molecular weight band and minimal smearing. We dried DNA extractions using a SpeedVac™ (Thermo Scientific™) on medium heat and reconstituted using molecular grade H₂O to a final uniform 250 ng/µl DNA concentration.

We created reduced representation libraries to survey SNP variation following published protocols (Wang, Meyer, McKay, & Matz, 2012) as updated by Dr. Eli Meyer (http://people.oregonstate.edu/~meyer/docs/Preparing2bRAD.pdf). Alfl restriction enzyme digest reduced representation (1/16th) libraries were labeled with individual barcodes and subjected to 18–20 PCR amplification cycles. The number of PCR cycles varied based on the optimal number determined in the test-scale PCR to find the minimum number of cycles to produce a visible product at 166 bp. We electrophoresed products on a 2% agarose gel in 1 x TBE buffer and ran at 150 V for 90 min, visualized target bands (165 bp), and excised them from the gel. Then, we purified the excised bands using a QIAquick® Gel Extraction Kit (QIAGEN). A final cleaning step used Agencourt® AMPure® XP beads (Beckman Coulter). QB3 Genomics at the University of California, Berkeley performed quality checks (qPCR, BioAnalyzer) and sequencing, multiplexing 10–20 snails per lane in 5 lanes of a 50 bp Single-End run on an Illumina HiSeq 2000 sequencer.

2.3 | RAD-seq data processing

To prepare raw sequence data for SNP identification, we truncated all raw sequence reads to the insert size (36 bp), filtered for quality (PHRED scores >20), and discarded empty constructs. We then processed the resulting data using custom scripts written by Misha Matz, available on the 2bRAD GitHub site (https://github.com/z00n/2bRAD_denovo). First, we counted unique tag sequences (minimum sequencing depth 5x) and the number of sequences in reverse-complement orientation and then merged these tags into one table. Then, we clustered all sequences in CD-HIT (Fu, Niu, Zhu, Wu, & Li, 2012) using a 91% similarity threshold. Next, we defined the most abundant sequence in the cluster as a reference sequence and then filtered a locus-annotated table from the previous two steps, excluding reads below 5x depth and those exhibiting strand bias. Lastly, we flipped the orientation of the resulting clustered sequences to match the most abundant tag in a cluster.

To call genotypes (as population-wide RAD-tag haplotypes), we used GATK (McKenna et al., 2010) and applied mild allele filters (10× total depth, allele bias cutoff 10, and strand bias cutoff 10), with the additional requirement that alleles appear in at least two individuals. We then applied locus filters allowing a maximum of 50% heterozygotes at a locus, no more than two alleles, genotyped in 30% of samples and polymorphic. Finally, we removed loci with the fraction of heterozygotes >75% (potential lumped paralogs) and missing >70% of genotypes. The final set of SNPs was then thinned to one per tag (that with the highest minor allele frequency) for F̂₅, and STRUCTURE analysis to remove linked loci.
2.4 | Individual sample filtering steps

To maximize the quality of the final dataset, we further 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 used VCFtools (Danecek et al., 2011) to estimate inbreeding coefficients and removed individuals (N = 5) with inbreeding coefficients outside the normal range (±2 SD of mean F) indicating possible low coverage sequencing or lumped paralogs (https://github.com/z0on/2bRAD_denovo). 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 et al., 2012).

2.5 | 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 RAD-seq dataset (2,718 loci), outlier loci only (73 loci), and neutral loci only (2,645 loci), from 51 individuals using two methods. First, we ran the Bayesian model-based clustering method STRUCTURE (Pritchard, Stephens, & Donnelly, 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 then analyzed in CLUMPAK v1.1 (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015) to select for the best K and graphically display the results.

2.6 | Outlier analyses

To test for evidence of natural selection in relation to coral host, we compared SNPs between lineages of snails on different hosts, pooled across six localities, with two datasets: (a) including all individuals and (b) 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 prior of 10, a sample size of 5,000, and 100,000 iterations, using a burn-in of 50,000, and 20 pilot runs of 5,000 each. To explore the impact of misleading data, we employed a 10% false discovery rate.

To further explore outlier loci, we used a second method to detect loci under selection (FDIST2) as implemented in ARLEQUIN (Excoffier & Lischer, 2010). 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).

2.7 | Candidate gene identification and annotation

To annotate the putative functions of genes linked to outlier loci, we aligned sequences containing SNP outlier loci to nucleotide collections (nr/nt) available on the NCBI website, in Blast2GO 5 Basic version (October 7, 2019) using the BLASTn algorithm (Altschul et al., 1997) with a taxonomic filter for Mollusca (taxid:6447). We adjusted parameters (expected threshold 10, word size 7, no low complexity filter, no mask for look-up table only) to accommodate short read sequences. We only examined hits with a high query coverage (>80%). Then, we identified and annotated any associated genes using NCBI and GeneCards®.

3 | RESULTS

After removing empty constructs and filtering for quality, we obtained an average of 5,710,091 unique sequence reads per individual at a minimum 5× depth. In total, we sequenced and genotyped 17,676 high-quality RAD-seq loci with ≥25× coverage, in 67 snails collected from two different coral host species, at six locations. After filtering for 30% maximum missing data per locus, this total decreased to 5,999 loci and then to 2,718 SNPs following thinning to one SNP per loci to remove any physically linked SNPs for STRUCTURE and F_{ST} analyses. 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.

3.1 | Genetic structure

Tests of genetic differentiation between sympatric snail lineages on different coral hosts revealed moderate but significant structure (mean F_{ST} = 0.047, weighted F_{ST} = 0.090 (Weir & Cockerham, 1984)), between host-associated lineages of snails (Figure 3). CLUMPAK

FIGURE 3  Histogram of variation in F_{ST} between lineages of Coralliophila violacea on two different coral hosts (Porites lobata and P. cylindrica) across all SNPs, excluding migrants and admixed individuals. F_{ST} calculated using FDIST in ARLEQUIN. Red line indicates the mean F_{ST} value (0.075)
analysis of the STRUCTURE results indicated $K = 2$ as the best $K$ value (Appendix S2). At $K = 2$, the majority (88%) of all snails grouped by their coral host (Figure 4). Grouping by host was stronger in snails collected from P. lobata (79%) than from P. cylindrica (79%). Neutral loci (2,645) and outlier loci only (73) showed similar patterns of population structure in STRUCTURE to the full dataset of SNPs (Appendix S3).

3.2 | Migration and admixture

Inferring the ancestry of individuals in STRUCTURE, using host as a prior, revealed strong differences among C. violacea living on different coral hosts (P. lobata and P. cylindrica, Figure 4), despite some migration and admixing between sympatric lineages. Moreover, migration rates were strongly asymmetric between snails living on these two hosts. In total, 19% (5 of 26 samples) of the snails collected from P. cylindrica had P. lobata genetic ancestry, while no snails (0 of 25 samples) with P. cylindrica ancestry were ever found on P. lobata (Appendix S4 and S5). Admixed individuals were only found at locations where migration was also observed (Dumaguete and Pulau Mengyat; Appendix S5). After excluding migrants and admixed individuals, the number of outlier loci only increased to eight (pairwise $F_{ST} = 0.419–0.543$, mean $F_{ST} = 0.480$, Figure 5b, Table 2). Four of these outlier loci (tag21753, tag28478, tag39884, and tag25141) had $\log_{10}(PO)> 1$ giving substantial-to-strong support as candidate loci, based on criteria from (Jeffreys, 1961). 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 $F_{ST} = 0.492$, Figure 5c) 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 $F_{ST} = 0.620$, Figure 5d) indicating directional selection, resulting in a combined total of 73 outlier loci across the two methods and datasets. Of these 73, a total of 43 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 (Table 2). Three outlier loci (tag28478, tag21753, and tag39884) were common among all datasets and methods (Table 2).

3.4 | Mapping and annotation of outlier loci

The majority (78%) of putative outlier loci did not align to any other mollusk sequences currently available in the NCBI database (11/2019, Table 2). Sixteen outlier loci DNA sequences aligned with a variety of mollusks including four gastropods (Aplysia californica, Littorina saxatilis, Lottia gigantea, and Pomacea canaliculata), three bivalves (Mizuhopecten yessoensis, Crassostrea gigas, and C. virginica), and two cephalopods (Octopus bimaculoides and O. vulgaris) (Table 2). Of these loci, 7 mapped to hypothetical or uncharacterized proteins. The remaining 9 loci mapped to gene regions with predicted functions. The annotated genes had various associated gene ontology terms including lipid metabolism, metal-ion binding, methyltransferase activity, immune response, chromatin binding, DNA binding, and serine/threonine-protein kinase. The top two hits (lowest e-values) were a neurotransmitter gene (tag15079, SLC6A7 gene) that plays a role in gastropod feeding behavior (Miller, 2019), and a hormone receptor gene (tag28347, HR96 gene) involved in the regulation of xenobiotic detoxification (Lindblom & Dodd, 2006; Richter & Fidler, 2014). At tag28347, there were two alleles that occurred
in almost equal frequency (43%, 57%) in the *P. lobata*-associated lineage of snails but were nearly fixed (97%) for one allele in the *P. cylindrica*-associated lineage of snails. Another gene of interest (tag13930, DRPR gene) codes for receptors involved in larval locomotory behavior (Freeman, Delrow, Kim, Johnson, & Doe, 2003).

**4 | DISCUSSION**

Genome-wide SNP data from six sympatric populations of *C. violacea* revealed two clearly differentiated clusters that were largely concordant with coral host, consistent with results from mitochondrial DNA (Simmonds et al., 2018). As with insects (Jean & Jean-Christophe, 2010; Simon et al., 2015), this genome-wide differentiation supports the conclusion of ecological divergence based on host association and adds to a small but growing literature on ecological divergence in marine environments (Fritts-Penniman et al., 2020; Potkamp & Fransen, 2019; Titus, Blischak, & Daly, 2019).

While SNP data reveal significant divergence between host-specific lineages of *C. violacea*, divergence was substantially lower in genome-wide SNPs compared to mtDNA ($F_{ST} = 0.047$ vs. $\Phi_{CT} = 0.561$).

This result may partially be a function of the smaller effective population size of the mitochondrial genome (Palumbi, Cipriano, & Hare, 2001). However, lower divergence values also suggest intermediate levels of gene flow between distinct host-associated lineages ($N_{m}>10$), values that are similar to other cases of sympatric host-associated divergence (e.g., Gouin et al., 2017; Peccoud, Ollivier, Plantegenest, & Simon, 2009; Smadja et al., 2012). Divergence with gene flow is further supported by the presence of admixed genotypes and unidirectional gene flow from one host lineage to the other. Moreover, considerable detection of outlier loci under directional selection (2.7% of all SNP loci; 73/2,718) strongly suggests that selection by coral host is likely contributing to the partitioning of *C. violacea* lineages.

**4.1 | Divergence with gene flow**

In parasitic species such as *C. violacea*, divergence with gene flow likely occurs through two mechanisms of premating isolation (Nosil, Vines, & Funk, 2005). The first is host preference for egg laying and/or recruitment to their host (either individual or species). Divergence occurs when mating takes place solely on that host, eventually
Table 2  Outlier loci analysis from *Coralliophila violacea* found on different coral hosts (*Porites lobata, P. cylindrica*), BLAST hits, and functional annotations

| Dataset     | Method  | FST     | log10(PO) | Tag ID | DNA sequence                   | Organism                        | Description                      | Score  | Coverage (%) | E-value   | Identity (%) | Gene symbol | GO terms                                | Predicted function |
|-------------|---------|---------|-----------|--------|--------------------------------|---------------------------------|----------------------------------|--------|---------------|-----------|--------------|-------------|------------------------------------------|-------------------|
| all ind.    | FDIST2  | 0.716   |           | 21753  | AGTCTCCTCTGGA                   | Mizuhopecten yessoensis         | Prosaposin-like                  | 35.6   | 80%           | 0.23      | 86%          | PSAPL1      | Lipid metabolic process                  | NA                |
| all ind.    | Bayescan| 0.354   |           | 2.465   | CACTGAGCTGCCAGCTCCACA           |                                 |                                  |        |               |           |             |             |                                          |                   |
| all ind.    | FDIST2  | 0.665   |           | 28478  | CATCCCCCTCTAT                   | Mizuhopecten yessoensis         | Prosaposin-like                  | 35.6   | 80%           | 0.23      | 86%          | PSAPL1      | Lipid metabolic process                  | NA                |
| all ind.    | Bayescan| 0.241   |           | 0.585   | GAACAGTGATAGC                   | Mizuhopecten yessoensis         | Prosaposin-like                  | 35.6   | 80%           | 0.23      | 86%          | PSAPL1      | Lipid metabolic process                  | NA                |

(Continues)
| Dataset | Method  | FST  | log10(P/PO) | Tag ID | DNA sequence | Organism | Description | Score | Coverage (%) | Identity (%) | E-value  | Gene symbol | GO terms | Predicted function |
|---------|---------|------|-------------|--------|--------------|----------|-------------|-------|--------------|-------------|----------|-------------|-----------|-------------------|
| all ind. | FDIST2  | 0.623 | 0.633       | 14249  | AGACGATGCG   | Aplysia  | Histone—lysine N-methyltransferase 2D-like | 38.3  | 80%          | 90%         | 0.066    | AGACGATGCG   | NA        | NA                |
| no mig./ adm. | FDIST2  | 0.796 | GAAGAGAGTTT | 19628  | GCACACACATGC | Aplysia  | Histone–lysine N-methyltransferase activity | Transcription co-activator activity | 36.5  | 100%        | 90%        | 0.23     | KMT2D       | NA        | Hypothetical protein |
| all ind. | FDIST2  | 0.54  | 0.54        | 36127  | TGTAGAGCTTTCTGCCTTC | Aplysia  | Metal ion binding activity | DNA binding | 36.5  | 100%        | 84%        | 0.461   | AGACGATGCG   | NA        | NA                |

**Table 2 (Continued)**
### TABLE 2 (Continued)

| Outlier analysis | BLAST search results | Gene ontology |
|------------------|-----------------------|---------------|
| Dataset Method   | FST log10(PO) Tag ID  | DNA sequence  | Organism | Score | Coverage (%) | E-value | Identity (%) | Gene symbol | GO terms | Predicted function |
| all ind. FDIST2  | 0.515 24247           | AGTTGCGGCAGG  | GCAGAAGACTGC | ATGACAGATCCC |
| no mig./adm.     | 0.784 24247           | GCAGAAGACTGC | ATGACAGATCCC |
| all ind. FDIST2  | 0.572 38182           | CGAACGGCTAATGG | CAATGCTTTGCAG ATGCAACATCA |
| no mig./adm.     | 0.775 38182           | Lotia gigantea | Hypothetical protein | 32.6 | 83% | 2.6 | 83% |
| all ind. FDIST2  | 0.55 17358            | CAGATGTTCTAGT | CATGACCATGCC | ATGTCACACT |
| no mig./adm.     | 0.768 17358           | Mizuhopecten yesoensis | Uncharacterized | 37.4 | 83% | 0.066 | 87% |
| all ind. FDIST2  | 0.541 38553           | AGCACACGACATG | CATTTCGTTGCC | TGAGAAATGCC |
| no mig./adm.     | 0.742 38553           | | | |
| all ind. FDIST2  | 0.485 33555           | AGGGCTTCATCAG | CATCCCATGCCG | TCTCAAGAAACA |
| no mig./adm.     | 0.735 33555           | | | |
| all ind. FDIST2  | 0.518 22329           | TGCTAACACAAG | CATAGTATTGCGA | CATATAACCG |
| no mig./adm.     | 0.729 22329           | Crassostrea gigas | Uncharacterized | 38.3 | 91% | 0.066 | 85% |
| all ind. FDIST2  | 0.536 21872           | CGACTCGGCAATG | CATCCCTTTGCT | GCCCTCTTTTC |
| no mig./adm.     | 0.727 21872           | | | |
| all ind. FDIST2  | 0.456 39420           | TGTTGGCTATGG | CAGCTTTGTTGC | TACAAACAGAAT |
| no mig./adm.     | 0.721 39420           | | | |
| all ind. FDIST2  | 0.468 33550           | TGAGGAAACACA | GCATTAGTTGC | AAATTAATTTTC |
| no mig./adm.     | 0.705 33550           | | | |
| all ind. FDIST2  | 0.415 30176           | AGGCCCTTTATG | GCAAAACAGCTG | CAACATACCTGCGA |
| no mig./adm.     | 0.679 30176           | | | |
| all ind. FDIST2  | 0.526 32580           | CACCGTTATCTTG | CAAACACATGCCG | ACGCCTGAACACT |
| no mig./adm.     | 0.673 32580           | | | |
| Dataset | Method | FST   | log10(PO) | Tag ID  | DNA sequence       | Organism | Description                                              | Score | Coverage (%) | E-value | Identity (%) | Gene symbol | GO terms                         | Predicted function          |
|---------|--------|-------|-----------|---------|-------------------|----------|----------------------------------------------------------|-------|--------------|---------|--------------|-------------|---------------------------------|-----------------------------|
| all ind. | FDIST2 | 0.525 |           | 28305   | TGGTTCAACATG      | Pomacea canaliculata | PR domain zinc finger protein 8-like | 30.1  | 86%          | 9.9     | 81%          | PRDM8       | Metal-ion binding               | NA                          |
| no mig./ adm. | FDIST2 | 0.67  |           |         | CAGCCATATGCA      |                       |                                            |       |              |         |             | PRDM8       | Histone methyltransferase activity | Chromatin binding           |
| all ind. | FDIST2 | 0.471 |           | 10755   | GGTGTGAAATTTG    | Octopus bimaculoides | AP2-associated protein kinase 1-like | 35.6  | 91%          | 0.23    | 85%          | AAK1        | Kinase, serine/threonine-protein kinase, transferase | NA                          |
| no mig./ adm. | FDIST2 | 0.659 |           |         | CAGCCAAATGCG      |                       |                                            |       |              |         |             | AAK1        | DNA binding, ATP binding, endocytosis | NA                          |
| all ind. | FDIST2 | 0.498 |           | 24085   | GGATAAAAGCCCG     | Aplysia californica  | Nuclear hormone receptor HR96-like | 39.2  | 100%         | 0.019   | 83%          | HR96        | Metal-ion binding               | Xenobiotic detoxification   |
| no mig./ adm. | FDIST2 | 0.652 |           |         | CACCAAAATGCG      |                       |                                            |       |              |         |             | HR96        | DNA binding, Receptor            |                             |
| all ind. | FDIST2 | 0.462 |           | 32708   | TGTGATCTCTTTG    | Octopus bimaculoides | PR domain zinc finger protein 8-like | 30.1  | 86%          | 9.9     | 81%          | PRDM8       | Metal-ion binding               | NA                          |
| no mig./ adm. | FDIST2 | 0.646 |           |         | ACTTTACTGCAA      |                       |                                            |       |              |         |             | PRDM8       | Histone methyltransferase activity | Chromatin binding           |
| all ind. | FDIST2 | 0.57  |           | 24158   | GGCCCATGTG       | Octopus bimaculoides | AP2-associated protein kinase 1-like | 35.6  | 91%          | 0.23    | 85%          | AAK1        | Kinase, serine/threonine-protein kinase, transferase | NA                          |
| no mig./ adm. | FDIST2 | 0.634 |           |         | GAGCCATGTG       |                       |                                            |       |              |         |             | AAK1        | DNA binding, ATP binding, endocytosis | NA                          |
| all ind. | FDIST2 | 0.429 |           | 28347   | AGAAAAGAGGC       | Aplysia californica  | Nuclear hormone receptor HR96-like | 39.2  | 100%         | 0.019   | 83%          | HR96        | Metal-ion binding               | Xenobiotic detoxification   |
| no mig./ adm. | FDIST2 | 0.617 |           |         | AGAGAAAGATAT     |                       |                                            |       |              |         |             | HR96        | DNA binding, Receptor            |                             |
| all ind. | FDIST2 | 0.417 |           | 37421   | AACTCCAAATCG      | Aplysia californica  | Nuclear hormone receptor HR96-like | 39.2  | 100%         | 0.019   | 83%          | HR96        | Metal-ion binding               | Xenobiotic detoxification   |
| no mig./ adm. | FDIST2 | 0.614 |           |         | CATTGTTTGCT       |                       |                                            |       |              |         |             | HR96        | DNA binding, Receptor            |                             |
| all ind. | FDIST2 | 0.463 |           | 22275   | TGCAATTGCGAAG     | Aplysia californica  | Nuclear hormone receptor HR96-like | 39.2  | 100%         | 0.019   | 83%          | HR96        | Metal-ion binding               | Xenobiotic detoxification   |
| no mig./ adm. | FDIST2 | 0.611 |           |         | CAAATGTGCTG      |                       |                                            |       |              |         |             | HR96        | DNA binding, Receptor            |                             |
| all ind. | FDIST2 | 0.404 |           | 24087   | TGCTATATGCTG      | Aplysia californica  | Nuclear hormone receptor HR96-like | 39.2  | 100%         | 0.019   | 83%          | HR96        | Metal-ion binding               | Xenobiotic detoxification   |
| no mig./ adm. | FDIST2 | 0.599 |           |         | AGTCCTCTGCA       |                       |                                            |       |              |         |             | HR96        | DNA binding, Receptor            |                             |

(Continues)
| Dataset          | Method | FST  | log10(PO) | Tag ID | DNA sequence | Organism          | Description                                  | Score | Coverage (%) | E-value | Identity (%) | Gene symbol | GO terms            | Predicted function                  |
|------------------|--------|------|-----------|--------|--------------|------------------|----------------------------------------------|-------|---------------|---------|---------------|-------------|-------------------|--------------------------------------|
| all ind.         | FDIST2 | 0.427|           | 16452  | AGTGAATGGAGAG | Littorina saxatilis | NA                                           | 41    | 88%           | 0.005  | 88%           |            |                   | Uncharacterized                      |
| all ind.         | FDIST2 | 0.432|           | 27928  | CACCTTGTTCGCC | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| all ind.         | FDIST2 | 0.458|           | 48048  | GCCATGTTCC   | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| all ind.         | FDIST2 | 0.414|           | 17029  | TGGTTTACCTTG | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| all ind.         | FDIST2 | 0.374|           | 34705  | AGCAGTCTCAGT | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| all ind.         | FDIST2 | 0.34  |           | 20904  | CAAAGACCTGCT | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| all ind.         | FDIST2 | 0.372|           | 20142  | AGATTCCATGCCA | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| all ind.         | FDIST2 | 0.388|           | 21098  | TGGGAAGATGTTG | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| all ind.         | FDIST2 | 0.334|           | 27266  | TGCAATGAAAAA | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| all ind.         | FDIST2 | 0.407|           | 15079  | GGCTGACGGAGAG | Pomacea canaliculata | Sodium-dependent proline transporter-like | 43.7  | 86%           | 0.002  | 90%           | SLC6A7      | Neurotransmitter activity          | Gastropod feeding behavior             |
| all ind.         | FDIST2 | 0.451|           |        | GCCGACGGCTGCG | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| no mig./ adm.    | FDIST2 | 0.748|           | 42043  | CGCAATCGTATTC | NA               | TGGAACAGCAGC   | AGTGAATGGAGAG |      | 88%           | 0.005  | 88%           |            |                   |                                      |
| Dataset     | Method | FST | Tag ID  | DNA sequence          | Organism        | Description                  | Score | Coverage (%) | E-value | Identity (%) | Gene symbol | GO terms                              | Predicted function                      |
|-------------|--------|-----|---------|-----------------------|-----------------|------------------------------|-------|--------------|---------|--------------|-------------|----------------------------------------|-------------------------------------------|
| no mig./adm. | FDIST2 | 0.676 | 31609   | GCAACAGATGTTG         | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.651 | 22586   | AGAGACAGATGG          | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.636 | 22561   | TGTGTGTGGTTT          | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.624 | 31557   | GGGGTTTGTAGC          | A. thaliana      | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.595 | 21042   | AGGCGTGGGCC          | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.555 | 33474   | TGACACTAGTCAG         | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.514 | 11613   | GGTGCGTGGCAC         | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.492 | 15069   | TGAACATGGG         | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.486 | 18109   | CACATCCATCGCA         | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.478 | 27744   | GAAGGTACCAAGC         | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.476 | 32951   | TACGTGCTGAG         | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| no mig./adm. | FDIST2 | 0.448 | 33996   | CACGTGCGTCAG         | C. elegans       | Uncharacterized              | NA    | NA           | NA      | NA           | NA          | NA                                     | NA                                        |
| Dataset        | Method  | FST  | log10(PO) | Tag ID | DNA sequence                          | Organism                        | Description         | Score (%) | Coverage (%) | E-value | Identity (%) | Gene symbol | GO terms                        | Predicted function |
|---------------|---------|------|-----------|--------|--------------------------------------|--------------------------------|---------------------|-----------|---------------|---------|--------------|------------|--------------------------------|-------------------|
| no mig./adm.  | FDIST2  | 0.44 |           | 16737  | TGTTGTGTGTGTGC                       | AGGTTCATGCAGCT                 | GATGGTGTG           |           |               |         |              |            |                                |                    |
| no mig./adm.  | FDIST2  | 0.431|           | 13930  | AGGTGAAATAAGACAGA                    | ATGAAATGCTGGGTG                 | GATGGTGTG           |           |               |         |              |            |                                |                    |
| no mig./adm.  | FDIST2  | 0.428|           | 34999  | GGATCTGCTCTGCA                       | AAAGCTTGCGGG                     | CTGGTCTCTG           |           |               |         |              |            |                                |                    |
| no mig./adm.  | FDIST2  | 0.424|           | 27749  | TGAGACGTTAACGCA                       | TACGGCTGCTTGT                    | AAGTAGGCC            |           |               |         |              |            |                                |                    |
| no mig./adm.  | FDIST2  | 0.424|           | 17800  | TGTGCTTCTTGGGC                       | AGAACCCTGCAAA                   | AATAATCTG            |           |               |         |              |            |                                |                    |
| no mig./adm.  | FDIST2  | 0.407|           | 13296  | AGAAAATCTCTGGGCA                      | CTTGTCCTGTAFT                   | GCTTTATCA            |           |               |         |              |            |                                |                    |
| no mig./adm.  | FDIST2  | 0.404|           | 17181  | AGCACACAGCACGCA                       | CGGTGGTGACAC                    | CAAGAGGCC            |           |               |         |              |            |                                |                    |
| no mig./adm.  | FDIST2  | 0.373|           | 16929  | GGGAATCCAAAGCA                        | ACTCAGTGCCTTAC                  | CCCCCCT              |           |               |         |              |            |                                |                    |
| all ind.      | FDIST2  | -0.033|          | 23096  | CACCCCTCTATGCA                       | AAGTCAATGCAAGG                  | CTGCCTCCT            |           |               |         |              |            |                                |                    |
| all ind.      | Bayescan | 0.638|           | 21172  | GGTACTAAAAAGCA                       | ACCGTATCCGAAT                   | CTGCTCA              |           |               |         |              |            |                                |                    |
| all ind.      | FDIST2  | 0.593|           | 20062  | CACCATGTCATGTC                       | AGTGCATGAGG                     | ACACATGCC            |           |               |         |              |            |                                |                    |
| all ind.      | FDIST2  | 0.491|           | 38482  | AAGGCCACACAGGCC                       | ACACAGATGCCCA                   | TCTTACTCA            |           |               |         |              |            |                                |                    |
| Dataset | Method   | FST | log10(PO) | Tag ID | DNA sequence                  | Organism               | Description | Score | Coverage (%) | E-value | Identity (%) | Gene symbol | GO terms | Predicted function |
|---------|----------|-----|-----------|--------|--------------------------------|------------------------|--------------|-------|--------------|---------|--------------|-------------|----------|-------------------|
|         | FDIST2   | 0.417 |           | 32340  | GAGTTGCTCAAGGC               | AAAATTCTGCAGA         | AAGGAAACA    |       |              |         |              |             |          |                   |
|         | FDIST2   | 0.366 |           | 33003  | TGAGGCTATTTGC                | ATGCAGGCTCTA          | GATCTCTTCC    |       |              |         |              |             |          |                   |
|         | FDIST2   | 0.323 |           | 9230   | TGCAGCTTTTGGCA               | TTCTTTGGCAAT          | CGAAGGCT      |       |              |         |              |             |          |                   |
|         | FDIST2   | 0.225 |           | 19533  | TGCTGATTTCTGGCA              | TACTGTTGCCATGCTTGCT   | TTCAAGACT     |       |              |         |              |             |          |                   |
|         | FDIST2   | 0.195 |           | 11006  | GCAGAAAGGAAGG                | CAAGCAAGATGCCTT       | AATAATGCCCT   |       |              |         |              |             |          |                   |

Note: Only the results that met cutoff statistics are shown.
Abbreviations: adm., admixed; ind., individuals; mig., migrants.
leading to speciation (Funk, Filchak, & Feder, 2002; Hawthorne & Via, 2001). Second is host adaptation, where selection acts against immigrants from another host via immigrant inviability (Nosil, 2007; Nosil et al., 2005; Porter & Benkman, 2017). Our study suggests that both mechanisms may be occurring in C. violacea. All migrants were individuals that genetically sorted to the lineage associated with P. lobata but were instead living on P. cylindrica. Additionally, only admixed individuals were observed on P. lobata. This pattern suggests that gene flow and admixture between host-associated lineages are unidirectional—from lobata to cylindrica. Such unidirectional gene flow could result from two possible scenarios, either the failure of larvae to recruit, or the failure of recruited larvae to survive.

Larval recruitment processes could promote asymmetrical gene flow if the lineage associated with P. cylindrica strongly prefers P. cylindrica as a host over P. lobata or does not respond to chemical settlement cues from P. lobata, preventing the recruitment of P. cylindrica-associated larvae to P. lobata. In addition, larvae from P. lobata would need to be less selective in their recruitment, occa-
sionally landing on P. cylindrica rather than P. lobata. Such a mechanism makes sense, given that there are twice as many coral species (N = 8) in the clade of Porites to which P. lobata belongs, than in that to which P. cylindrica belongs.

An alternative, but not mutually exclusive explanation is that asymmetry in gene flow and admixture could result from postset-
tlement processes. For example, if larvae from P. cylindrica-associated individuals settle on P. lobata, but are less likely to survive and reproduce, this could lead to immigrant inviability (Ingley & Johnson, 2016; Nosil et al., 2005; Richards & Ortiz-Barrientos, 2016) and asymmetry in admixture. Under such a scenario, genes beneficial to snails living on P. cylindrica would likely be less helpful on P. lobata and we should see some indication of a selective sweep in the de-

erved lineage with respect to the standing genetic variation of the ancestral lineage (Przeworski, Coop, & Wall, 2005). Indeed, results showed some outlier loci (e.g., HR96, detoxification gene) that were in equal proportions in P. lobata (43%, 57%) but were at near fixa-
tion in P. cylindrica (97%), indicating a soft sweep on standing genetic variation at that locus.

Regardless of whether the limited misalignment of snails and coral hosts results from pre- or postrecruitment processes, the fact that the vast majority of snails sort by host coral in the face of hy-
bridization and gene flow indicates that natural selection must be relatively strong to counteract gene flow of Nm>10 (Funk, Egan, & Nosil, 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 lineage. This conjecture is consistent with the observation that specialist species often evolve from general-
ist 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 important mech-
anism of divergence within the Coral Triangle (Briggs, 2005) as increased diversity should raise niche partitioning, leading to more opportunities for host specialization (Janz, Nylin, & Wahlberg, 2006).

### 4.2 Candidate genes involved in adaptation to host

Outlier loci can provide insights into the targets of natural selec-
tion (Storz, 2005) and are a useful starting point for determining how selection may be acting on lineages diverging on different hosts. Our analysis revealed 73 putative gene regions with FST values signi-

cificantly 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 (Via, 2012).

There is no a priori information on the types of genes involved in the adaptation of mollusks to different hosts and, due to a lack of genomic resources for C. violacea, only 9 of 73 outlier loci mapped to gene regions with predicted functions. However, a useful compar-
ison can be found in ectoparasitic phloem-feeding insects adapting to different host plants (Oren et al., 1998). Genes under selection in these insect–plant interactions include genes involved in sensing hosts, that protect insects against plant defenses and facilitate feed-
ing, and that code for digestive and detoxifying enzymes to neutral-
ize plant toxins (e.g., metal-ion binding, Simon et al., 2015).

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, KLH1, PRDM8, and HR96). Very little is known about how corals and their algal symbionts chemically defend themselves against or react to parasites and predators. Symbiodinium species do produce toxins—Zooxanthellatoxins—(Gordon & Leggat, 2010), but it is unknown whether these toxins are upregulated in response to parasites or predators.

Additional evidence for detoxification playing a role in host di-
vergence comes from HR96, a nuclear hormone receptor involved in xenobiotic detoxification (Richter & Fidler, 2014). Interestingly, HR96 was nearly fixed for one allele in C. violacea from P. cylindrica (97%) but was at 50% in C. violacea from P. lobata, which indicates a selective sweep at that locus. This result, combined with the four metal-ion binding gene regions, suggests that there may be important differences in host-associated detoxification processes in the different C. violacea lineages. If adaptation to host-specific toxins drives host specificity, mismatches between snail metabolic abilities and coral hosts could explain the strong asymmetry in snails being found on an atypical coral host.

While the above results suggest a putative detoxification role for some outlier loci, two other genes with predicted functions, a neurotransmitter (SLC6A7) important for gastropod feeding behav-
ior (Miller, 2019) and a transmembrane receptor (DRPR) involved in larval locomotory behavior, indicate a possible role of behavior in adaptation (Freeman et al., 2003). Notably, this is only the first ge-
"
are crucial for adaptations to coral hosts. Future work would benefit from a fully annotated genome of C. violacea that would allow us to examine the genomic architecture of divergence with gene flow and quantitative trait loci. In turn, this would allow us to better pinpoint regions of the genome under selection, and the specific functions of genes involved in adapting to different hosts.

4.3 | Ecological divergence in the sea

John Briggs originally proposed the idea of sympatric speciation as an important diversification mechanism within the Coral Triangle (i.e., “Center of Origin” hypothesis), as well as in the 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 generating biodiversity in the Coral Triangle. In addition, spatial patterning of C. violacea sympatric host lineages also matches the pattern Briggs described, with the ancestral P. lobata host lineage having a broad geographic distribution, and the derived P. cylindrica host lineage restricted to the Coral Triangle (Simmonds et al., 2018).

As the global epicenter of marine biodiversity, there is a large and diverse literature on the processes shaping the Coral Triangle (Barber, Erdmann, & Palumbi, 2006; Bowen et al., 2013; Carpenter et al., 2011; Gaither et al., 2011; Hoeksema, 2007; Kochzius & Nuryanto, 2008; Tornabene, Valdez, Erdmann, & Pezold, 2015). While there is ongoing debate (Evans, McKenna, Simpson, Tournois, & Genner, 2016; Huang, Goldberg, Chou, & Roy, 2018; Di Martino, Jackson, Taylor, & Johnson, 2018; Matias & Riginos, 2018), there is clearly a multiplicity of processes driving diversification in this region (Barber & Meyer, 2015). Given the results of this study, it is important to expand our thinking beyond models that focus solely on allopatry to advance our understanding of marine speciation and origins of the Coral Triangle biodiversity hotspot.

ACKNOWLEDGMENTS

We are grateful to Dr. Elia Meyer and Dr. Misha Matz for the use of their scripts to process and analyze 2b-RAD sequence data. We acknowledge the Indonesian government, including the Indonesian Ministry of Research and Technology (RISTEK), Indonesian Institute of Sciences (LIPI), Nature Conservation Agency (BKSDA), and the National Marine Park offices of Bunaken and Wakatobi for their support. We are also grateful to the Indonesian Biodiversity Research Center at Udayana University and the Institute for Environmental and Marine Sciences at Silliman University for institutional support. This work was funded by three National Science Foundation programs (OISE-0730256, OISE-1243541, and OCE-0349177) and a grant from the United States Agency for International Development (497-A-00-10-00008-00). The UCLA Department of Ecology and Evolutionary Biology, Lemelson Foundation, Conchologists of America, and Sigma Xi gave additional funding to S. Simmonds. Sampling was covered under research permits obtained in the Philippines (Department of Agriculture—Bureau of Fisheries and Aquatic Resources) and Indonesia (RISTEK 2011, 198/SIP/FRP/SMNI/2012, 187/SIP/FRP/SM/VI/2013).

AUTHOR CONTRIBUTIONS

SES conceived of and designed the study, SES, AFP, and SHC collected samples, prepared libraries, and analyzed genomic data. All authors worked on and approved of the manuscript.

DATA AVAILABILITY STATEMENT

Raw single-end Illumina HiSeq 2000 reads and RAD-seq loci datasets are archived on Dryad: https://doi.org/10.5068/D1995V.

ORCID

Sara E. Simmonds https://orcid.org/0000-0003-3438-3559

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25(17), 3389–3402. https://doi.org/10.1093/nar/25.17.3389
Barber, P., & Boyce, S. L. (2006). Estimating diversity of Indo-Pacific coral reef stomatopods through DNA barcoding of stomatopod larvae. Proceedings. Biological Sciences/the Royal Society, 273(1597), 2053–2061. https://doi.org/10.1098/rspb.2006.3540
Barber, P. H., Erdmann, M. V., & Tengardjaja, K. (2011). Evolution and conservation of marine biodiversity in the Coral Triangle: Insights from stomatopod Crustacea. In C. Held, S. Koenemann & C.D. Schubart (Eds.), Phylogeography and population genetics in Crustacea. Crustean Issues (pp. 264–277). Boca Raton, FL:CRC Press.
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(9), 1825–1839. https://doi.org/10.1111/j.0011-0428.2006.00026.x
Barber, P. H., & Meyer, C. P. (2015). Pluralism explains diversity in the Coral Triangle. In C. Mora (Ed.), Ecology of fishes on coral reefs (pp. 258–263). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781131605412.032
Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. Proceedings. Biological Sciences/the Royal Society, 263(1377), 1619–1626. https://doi.org/10.1098/rspb.1996.0237
Bernal, M. A., Gaither, M. R., Simison, W. B., & Rocha, L. A. (2017). Introgression and selection shaped the evolutionary history of sympatric sister-species of coral reef fishes (Genus: Haemulon). Molecular Ecology, 26(2), 639–652. https://doi.org/10.1111/mec.13937
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(2), 158–180. https://doi.org/10.1007/s11692-012-9183-6
Bird, C. E., Holland, B. S., Bowen, B. W., & Toonen, R. J. (2011). Diversification of sympatric broadcast-spawning limpets (Cellana spp.) within the Hawaiian archipelago. Molecular Ecology, 20(10), 2128–2141. https://doi.org/10.1111/j.1365-294X.2011.05081.x
Blackall, L. L., Wilson, B., & van Oppen, M. J. H. (2015). Coral-the world’s most diverse symbiotic ecosystem. Molecular Ecology, 24(21), 5330–5347. https://doi.org/10.1111/mec.13400
Biogeography, time, and place: Distributions, barriers, and islands (pp. 117–178). Dordrecht, Netherlands: Springer.

Huang, D., Goldberg, E. E., Chou, L. M., & Roy, K. (2018). The origin and evolution of coral species richness in a marine biodiversity hotspot. Evolution, 72(2), 288–302. https://doi.org/10.1111/evol.13402

Hurt, C., Silliman, K., Anker, A., & Knowlton, N. (2013). Ecological speciation in anemone-associated snipping shrimps (Alpheus armatus species complex). Molecular Ecology, 22(17), 4532–4548. https://doi.org/10.1111/mec.12398

Ingley, S. J., & Johnson, J. B. (2016). Divergent natural selection promotes immigrant inviability at early and late stages of evolutionary divergence. Evolution, 70(3), 600–616. https://doi.org/10.1111/evo.12872

Ingram, T. (2010). Speciation along a depth gradient in a marine adaptive radiation. Proceedings of the Royal Society B: Biological Sciences, 278(1705), 613–618. https://doi.org/10.1098/rspb.2010.1127

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

Jean, P., & Jean-Christophe, S. (2010). The pea aphid complex as a model of ecological speciation. Ecological Entomology, 35, 119–130. https://doi.org/10.1111/j.1365-2311.2009.01147.x

Jeffreys, H. (1961). Theory of Probability (3rd). Oxford, UK: Clarendon Press.

Jiggins, C. D. (2008). Ecological speciation in mimetic butterflies. BioScience, 58(6), 541–548. https://doi.org/10.1641/B580610

Kawecki, T. J. (1996). Sympatric speciation driven by beneficial mutations. Proceedings of the Royal Society of London. Series B: Biological Sciences, 263(1376), 1515-1520. https://doi.org/10.1098/rspb.1996.0221

Kawecki, T. J. (1998). Red queen meets Santa Rosalia: Arms races and the evolution of host specialization in organisms with parasitic lifestyles. The American Naturalist, 152(4), 635–651. https://doi.org/10.1086/286195

Kitahara, M. V., Cairns, S. D., Stolarski, J., Blair, D., & Miller, D. J. (2010). A comprehensive phylogenetic analysis of the Scleractinia (Cnidaria, Anthozoa) based on mitochondrial CO1 sequence data. PLoS ONE, 5(7), e11490. https://doi.org/10.1371/journal.pone.0011490

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(17), 3775–3787. https://doi.org/10.1111/j.1365-294X.2008.03803.x

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A comprehensive computational approach to population structure analysis: The three-times rule. Review of Genetics, 39, 179–197. https://doi.org/10.1007/s10992-008-9226-y

Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Kitahara, M. V., Cairns, S. D., Stolarski, J., Blair, D., & Miller, D. J. (2010). A comprehensive phylogenetic analysis of the Scleractinia (Cnidaria, Anthozoa) based on mitochondrial CO1 sequence data. PLoS ONE, 5(7), e11490. https://doi.org/10.1371/journal.pone.0011490

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(17), 3775–3787. https://doi.org/10.1111/j.1365-294X.2008.03803.x

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Peccoud, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A comprehensive computational approach to population structure analysis: The three-times rule. Review of Genetics, 39, 179–197. https://doi.org/10.1007/s10992-008-9226-y

Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Nicholson, R. (2005). Molecular signatures of natural selection. Annual Review of Genetics, 39, 197–218. https://doi.org/10.1146/annurev.genet.39.073003.112420

Nosil, P. (2002). Transition rates between specialization and generalization in phytophagous insects. Evolution, 56(8), 1701–1706. https://doi.org/10.1111/j.0014-3820.2002.tb01482.x

Nosil, P. (2007). Divergent host plant adaptation and reproductive isolation between ecotypes of Timema cristinae walking sticks. The American Naturalist, 169(2), 151–162.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. Molecular Ecology, 18(3), 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Nosil, P., Vines, T. H., & Funk, D. J. (2005). Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution, 59(4), 705–719. https://doi.org/10.1554/04-0428

Oren, U., Brickner, I., & Loya, Y. (1998). Prudent sessile feeding by the coral reef algal sponge, Cassidulina philiea violacea on coral energy sinks. Proceedings of the Royal Society B: Biological Sciences, 265(1410), 2043–2050. https://doi.org/10.1098/rspb.1998.0538

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. Annual Review of Ecology and Systematics, 25(1), 547–572. https://doi.org/10.1146/annurev.es.25.110194.002555

Palumbi, S. R., Cipriano, F., & Hare, M. P. (2001). Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. Evolution, 55(5), 859–868. https://doi.org/10.1554/0014-3820(2001)055<0859:PGCGFM>2.0.CO;2

Peccedo, J., Ollivier, A., Plantegenest, M., & Simon, J.-C. (2009). A continual of genetic divergence from sympatric host races to species in the pea aphid complex. Proceedings of the National Academy of Sciences of the United States of America, 106(18), 7495–7500. https://doi.org/10.1073/pnas.081117106
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(8), 808–810. https://doi.org/10.1038/nmeth.2023

Waser, N. M., & Campbell, D. R. (2004). Ecological speciation in flowering plants. In U. Dieckmann, M. Doebeli, J. Metz & D. Tautz (Eds.), *Adaptive speciation (Cambridge Studies in Adaptive Dynamics)* (pp. 264–277). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9781139342179.015

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. https://doi.org/10.2307/2408641

Westram, A. M., Rafajlović, M., Chaube, P., Faria, R., Larsson, T., Panova, M., ... Butlin, R. (2018). Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. *Evolution Letters*, 2(4), 297–309. https://doi.org/10.1002/evl3.74

Zann, L. P. (1987). A review of macrosymbiosis in the coral reef ecosystem. *International Journal for Parasitology*, 17(2), 399–405. https://doi.org/10.1016/0020-7519(87)90115-9

**SUPPORTING INFORMATION**

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Simmonds SE, Fritts-Penniman AL, Cheng SH, Mahardika GN, Barber PH. Genomic signatures of host-associated divergence and adaptation in a coral-eating snail, *Coralliophila violacea* (Kiener, 1836). *Ecol Evol*. 2020;10:1817–1837. https://doi.org/10.1002/ece3.5977