Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Previously unidentified genetic structure revealed for the sponge Suberites diversicolor: implications for sponge phylogeography and population genetics

Diede Maas[1], Stefan Prost[2,3], Christiaan de Leeuw[1], Ke Bi[4], Lydia Smith[4], Purwanto Purwanto[5], Ludi Aji[6], Ricardo Tapilatu[7], Rosemary Gillespie[4], and Leontine Becking[1]

[1]Wageningen University & Research
[2]University of California, Berkeley
[3]Senckenberg Natural History Museum
[4]University of California Berkeley
[5]Coral Triangle Center
[6]Lembaga Ilmu Pengetahuan Indonesia
[7]State University of Papua

August 23, 2020

Abstract

The relative influence of geography, oceanography and environment on gene flow within sessile marine species remains an open question. Detecting subtle genetic differentiation at small scales relevant to marine protected areas is challenging in benthic populations due to large effective population sizes, general lack of resolution in genetic markers, potential microbial associations, and because barriers to dispersal often remain elusive. We genotyped the sponge species Suberites diversicolor using double digest restriction-site associated DNA sequencing (4,826 Single Nucleotide Polymorphisms, SNPs), compared it to same individuals using single markers (COI and ITS), and used previously published data on the associated microbial communities from a subset of the same locations. Studying S. diversicolor from marine lakes at different spatial scales (1-1,400 km), along a gradient of connection to the surrounding sea, and with different environmental regimes, we did not detect strong effects of geographic distance, permeability of seascape barriers or local environments in shaping population genetic structure. All markers detected two major lineages and geographic clustering over a large spatial scale. However, with the SNP dataset we provide new evidence of strong population structure even at scales <10km (average FST = 0.56), where previously none was detected. A lack of congruence between host population structure and microbial community patterns of S. diversicolor from the same locations was observed, suggesting they are on different eco-evolutionary tracks. Our results call for a reassessment of poorly dispersing benthic organisms that were previously assumed to be highly connected based on low resolution markers.

Introduction

The spatial and temporal processes that generate and maintain marine biodiversity are not fully understood (Bowen *et al.* , 2013; Orsini*et al.* , 2013; De Meester *et al.* , 2016; Costello and Chaudhary, 2017). Marine populations display diverse patterns of genetic structure, such as isolation-by-distance (Wright, 1943; Chaves-Fonnegra*et al.* , 2015; Pérez-Portela, Noyer and Becerro, 2015), regional clustering (Selkoe *et al.* , 2014; Brown, Davis and Leys, 2017; Riesgo *et al.* , 2019), isolation-by-environment (Orsini *et al.* , 2013; Giles *et al.* , 2015), as well as patterns that are not clearly linked to spatial or environmental structuring (Cornwell*et al.* , 2016; Miller *et al.* , 2018; Taboada *et al.* , 2018). However, barriers to dispersal and isolating mechanisms over small spatial scales, such as the range of marine protected areas, remain elusive especially for sessile

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

marine organisms with a dispersive larval stage (Liggins, Treml and Riginos, 2013). Sponges, integral but often underappreciated assets of benthic communities (Diaz and Rützler, 2001; Bell, 2008; Bell *et al.* , 2013; De Goeij *et al.* , 2013; Dunn, Leys and Haddock, 2015; Webster and Thomas, 2016), are generally considered to be poor dispersers as their larvae have limited swimming capacity and are short-lived (Maldonado, 2006). Sponges are therefore excellent candidates to investigate marine population genetic structure on small scales. However, despite the recognized relevance of sponges in benthic ecosystems, results are ambiguous on sponge genetic diversity, degrees of gene flow between populations and drivers of divergence.

Different studies have explored how geography, oceanography and environmental factors may influence gene flow within sponge species. Due to their restricted dispersal, the spatial scale of sponge gene flow should be limited. The majority of studies investigating genetic structure in sponges have revealed species complexes with divergence among morphologically cryptic lineages (Oppen, Wörheide and Takabayashi, 2002; Wörheide, Solé-Cava and Hooper, 2005; Uriz and Turon, 2012; Pérez-Portela and Riesgo, 2018). However, studies investigating within-lineage divergence are scarce. This may be a result of many most studies using relatively small amounts of genetic data and a single type of genetic marker (Selkoe *et al.* , 2016; Timm, 2020). An increase in number of molecular markers is expected to advance inferences on demography and structure (Felsenstein, 2004; Allendorf, Hohenlohe and Luikart, 2010; Kelley *et al.* , 2016; Pérez-Portela and Riesgo, 2018). Molecular markers commonly used to assess sponge phylogeography and population structure include mitochondrial markers (mtDNA) such as Cyochrome c oxidase I (*COI*) and ATP6, and nuclear markers such as introns, internal transcribed spacers (*ITS*) and microsatellites (Oppen, Wörheide and Takabayashi, 2002; Wörheide, Solé-Cava and Hooper, 2005; Uriz and Turon, 2012; Pérez-Portela and Riesgo, 2018). Though widely used in phylogeographic and population genetic studies (Avise, 2000, 2009), mitochondrial markers exhibit low mutation rates in sponges (Wörheide, Solé-Cava and Hooper, 2005; Huang *et al.* , 2008). As a result, the majority of studies using mtDNA find panmixia among sponge populations across broad geographic ranges (e.g. Duran, Pascual and Turon, 2004; Whalan *et al.* , 2008; De Bakker *et al.* , 2016; Ekins *et al.* , 2016). *ITS* markers can show more structure (Bentlage and Wörheide, 2007; Becking *et al.* , 2013; Ekins*et al.* , 2016), but generally at larger spatial scales and are hampered by intra-genomic polymorphisms (Frankham, Briscoe and Ballou, 2002). Microsatellites could be reliable and sufficiently variable to detect population structure, yet are time-consuming to design *de novo* for each species (Frankham, Briscoe and Ballou, 2002; Pérez-Portela and Riesgo, 2018), and generally relatively few markers have been used per study (<20), again limiting the molecular marker panel. Furthermore, microsatellites can be confounded by homogenizing forces of evolution, making them less effective in detecting genetic divergence (Oppen, Wörheide and Takabayashi, 2002). Hence, there is a need for increasing genetic resolution in order to reassess assumptions of panmixia within sponge populations at fine spatial scales.

Recently, there has been an increase in the use of reduced representation genomic methods and Single Nucleotide Polymorphisms (SNPs) for population studies on non-model organisms (Baird *et al.* , 2008; Peterson *et al.* , 2012; Puritz *et al.* , 2014; Catchen *et al.* , 2017). Genome-wide SNP data increases the number of loci compared to traditional mitochondrial or nuclear markers and a larger marker panel is expected to small scaled population structure when compared to single marker studies. The effect of an increased marker panel has been shown for example in mussels (Becking *et al.* , 2016; Maas *et al.* , 2018; de Leeuw *et al.* , 2020), and fish (Bradbury *et al.* , 2015; Lemopoulos *et al.* , 2019; D'Aloia *et al.* , 2020; Sunde *et al.* , 2020). However, high resolution studies on sponges are lagging behind (Pérez-Portela and Riesgo, 2018), with notable exception of Brown *et al.* (2014), Brown *et al.* (2017) and Leiva *et al.* (2019). Using the SNPs generated by Brown *et al.* (2014), Brown et al. (2017) genotyped 67 SNPs for the deep-sea glass sponge *Aphrocallistes vastus* and found high differentiation between geographic regions (average F_{ST}= 0.25), but no structure at distances <275km, indicating connectivity at this scale. Leiva et al. (2019) observed panmixia for the Antarctic sponge *Dendrilla antarctica* over 900km when analyzing 389 neutral SNPs. However, 140 SNPs under putative positive selection did show genetic differentiation (global F_{ST} = 0.20) over 100km. Potentially the number of SNPs used in these studies are still too low to detect small-scaled population divergence. Using RADseq techniques such as ddRAD (Peterson*et al.* , 2012) may increase the number of retained SNPs to thousands and provide the necessary resolution.

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Another challenge to unveiling sponge population genetic patterns is that sponges are considered true holobionts, associations between the host and its microbes (Webster and Thomas, 2016), and may not evolve as single units. Due to the propensity of sponges to harbor dense communities of microbes, there is a potential of including associated microbial material in extractions, therefore clouding host specific patterns. Studies into the sponge holobiont have suggested that microbial communities are highly specific to sponge host identity (Easson and Thacker, 2014; Reveillaud *et al.* , 2014), and communities to be stable across gradients in geography (Taylor *et al.* , 2005), time (Hardoim and Costa, 2014), and, for tropical reef sponges, depth (Steinert *et al.* , 2016). However, these expectations do not always hold true (Swierts, Cleary and de Voogd, 2018; Cleary *et al.* , 2019; Easson *et al.* , 2020; Ferreira*et al.* , 2020). For example, Easson *et al.* (2020) concluded that microbe community structure is influenced by the interplay of geographic, environmental and host factors, with a potential effect of even small population-level genetic structure. Since patterns of microbial diversity can differ from sponge host diversity (Noyer and Becerro, 2012), it is important to understand how microbial community patterns are related to sponge host population genetics.

Islands, and other insular systems, provide ideal models to test factors that underlie population structure since they are well-defined and are of lower complexity than open areas (Warren *et al.* , 2015). Marine lakes are insular systems of bodies of seawater surrounded completely by land that maintaining a connection with the surrounding sea through caves or porous rock (Holthuis, 1973; Hamner, Gilmer and Hamner, 1982; Dawson *et al.* , 2009; Becking *et al.* , 2011). Clusters of marine lakes are present in the Caribbean, Vietnam, Palau, and Indonesia, particularly in East Kalimantan and in West Papua (Dawson*et al.* , 2009; Becking *et al.* , 2011; Becking, de Leeuw and Vogler, 2015). Marine lakes were formed *de novo* when depressions in karstic rock were filled with sea water after the Last Glacial Maximum (approximately 20,000 years ago) (Tomascik and Mah, 1994), and house clearly defined populations (Gotoh *et al.* , 2011; Itescu, 2018). Sponges are usually well-represented in marine lakes, having high diversity and abundance (Azzini *et al.* , 2007; Becking *et al.* , 2011, 2013; Cleary *et al.* , 2013). Having originated roughly at the same time marine lakes represent relatively controlled biotopes where each lake can be seen as an independent replicate of eco-evolutionary dynamics over time.

Marine lakes have been used before to study population genetic and microbial community patterns studies (Becking *et al.* , 2013; Cleary *et al.* , 2013; Cleary, Polónia and de Voogd, 2018; Ferreira*et al.* , 2020). The sponge *Suberites diversicolor*(Porifera, Demospongiae, Suberitidae, Becking and Lim, 2009) has been found to occur extensively in marine lakes, and also in brackish coastal areas (Becking and Lim, 2009; Cleary *et al.* , 2013). Using*COI* and *ITS* genetic markers, Becking *et al.* (2013) studied the phylogeography of *S. diversicolor* from multiple marine lakes and lagoon populations in the Indo-Pacific. They identified two distinct genetic lineages and regional structuring yet did not observe subtle levels of structuring at smaller spatial scales. The lack of structure could be explained by recurrent gene flow among lakes, or by lack of resolution of genetic markers used by Becking *et al.*(2013), as they recovered a low number of haplotypes (4 for *ITS*and 3 for *COI*). Analyzing the microbes of the same *S. diversicolor* populations, Cleary *et al.* (2013) found that the associated microbial community did not differ among sponges sampled from marine lakes and open sea habitats within one region. Between broad geographic regions (>1,400km) the associated microbial communities were significantly different (Ferreira *et al.* , 2020). Clearly, there is a need to further elucidate population genetic patterns and see how host and microbe patterns contrast.

Studying *a priori* defined sponge populations from nine marine lakes and two lagoon locations in Indonesia (Berau, East-Kalimantan and Raja Ampat, West-Papua) and Australia, we aim to assess the population structure in *S. diversicolor* and associated drivers. Selecting marine lakes on different spatial scales (1-1,400km), along a gradient of connection to the surrounding sea and with different environmental regimes allows the opportunity to assess effects of geographic distance, permeability of barriers and local environments in shaping genetic structure. In order to assess the effect of level of genetic resolution, we compared results of our genome-wide sequencing strategy (double-digest restriction-site associated DNA sequencing, (ddRAD, Peterson et al., 2012)) to previously published results on the same individuals using single markers (*COI* and *ITS*) (Becking*et al.* , 2013). Furthermore, we compare host population structure with previously published structure from associated microbial communities in *S. diversicolor* (Cleary *et al.* , 2013; Ferreira *et al.* , 2020)

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

to assess whether sponge host and microbes are on similar evolutionary tracts.

Material and methods

Sample collection and lake profiling

Tissue samples (~1cm^3) were collected from 168 individuals of *Suberites diversicolor* (Fig 1, Table 1). One lagoon was sampled in Darwin, Australia (DAR), one lagoon and three marine lakes were sampled in Berau (Bay, B.1, B.2 and B.3), and six marine lakes were sampled in Raja Ampat (P.27, P.30, P.32, P.1, P.4 and P.5). Of these locations, nine overlap with the sponge phylogeography study of Becking et al. (2013) and five with the microbial community analysis of Ferreira et al. (2020) (Supplemental Table 1 for corresponding lake codes between the three studies). Samples were collected between 1-5m depth while snorkeling. Some lakes had very low densities of *S. diversicolor* , therefore sample sizes were lower (see Becking *et al.,* 2013 Table 2 for densities). In the field, tissue samples were immediately preserved in 99% ethanol or RNAlater after excision at 0-4°C (4-8 weeks), and upon returning to the laboratory stored in a -20°C freezer until further use.

Lake characterization was performed concordant with a protocol published in Maas et al. (2018). In brief, lake area (m^2) was approximated using Google Earth Pro (v. 7.3.2), maximum depth was measured using a handheld sonar system (Hawkeye), and water parameters (temperature (°C) and salinity (ppt) were measured with an YSI Professional Plus multimeter at 10 locations per lake at 1m intervals from the surface to 5m depth. To define connection to the surrounding sea we measured maximum tidal amplitude simultaneously in the lake and the sea using Hobo water-level loggers. The ratio of maximum tidal amplitude in meters of the lake compared to the sea was used as a proxy to determine as the degree of physical connection between the lake and sea (conform to calculations in Maas et al., 2018).

DNA extraction, library preparation and sequencing

DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen), with the only modification from manufacturer instructions being an extended lysis time (overnight). DNA quality and quantity were assessed using 1.5% agarose gels and Qubit dsDNA HS assays. Next, ddRAD libraries were prepared following the protocol of Maas *et al* ., (2018), adapted from the original protocol of Peterson et al., (2012). We refer to the extensive protocol included in Maas *et al.* , (2018) for details, but describe here how we adapted it for *S. diversicolor* . In brief, genomic DNA (600ng) was double-digested using enzymes SphI-HF (rare-cutting) and MlucI (frequent-cutting) (See Supplementary Information S1 for example of a successful enzyme digestion). Size distribution of the fragments was assessed with the BioAnalyzer High Sensitivity Chip (Agilent). We used the spreadsheet provided in Peterson*et al.* , (2012) "Locus count from Bioanalyzer % in region" to calculate the number of fragments to be expected assuming a genome size of 600Mb (common for sponges (Jeffery, Jardine and Ryan Gregory, 2013) and various size selections of RAD fragments. This number can subsequently be used to calculate the expected coverage when generating a known amount (Gb) of sequencing data. Custom-made sample-specific barcodes were ligated to the fragments to allow for the pooling of 21 samples per library, resulting in 8 libraries in total. The Sage Science Pippin Prep was used to size-select adapter-ligated fragments of length 500-575bp (indicating an insert size of 425-500bp). A trial was run for 8, 10 and 12 polymerase chain reaction cycles (PCR) reactions. In the end ten PCR cycles were chosen as a balance between DNA output and PCR duplication and were run on each library for enrichment and ligation of Illumina indices unique to each library pool. Quality and quantity of libraries throughout the process were checked using BioAnalyzer High Sensitivity chips (Agilent, Supplemental Information S2 for an example). Libraries were pooled at equimolar volumes and 150bp single-end sequenced on Illumina HiSeq 2500 at the Vincent J. Coates Genomic Sequencing Facility at UC Berkeley.

Reference assembly, bioinformatic filtering and genotype calling

Custom perl scripts were used for processing the resulting sequences (RADTOOLKIT v. 0.13.10, available in Dryad Digital Repository). Raw fastq reads were demultiplexed using a maximum of one mismatch and removed if expected cut sites were not found. Resulting demultiplexed reads were trimmed of Illumina adapter

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

contaminations and low-quality reads using cutadapt (Martin, 2011) and Trimmomatic (Bolger, Lohse and Usadel, 2014). Cleaned reads were clustered with CD-HIT (Li and Godzik, 2006; Fu *et al.* , 2012), with a minimum support per cluster set at three reads, and representative sequences retained for each cluster. RepeatMasker (http://repeatmasker.org/) was used to mask putative repetitive elements, low complexity regions and short repeats using 'Mytilidae' as a database (Smit, Hubley and Green, 2014). Loci were discarded if >60% of nucleotides per loci were Ns. The resulting RAD loci were combined for all individuals, and a reference was built from loci shared by at least 70% of individuals.

Cleaned sequence reads for each individual were aligned to the *de novo* generated reference separately using Novoalign (http://www.novocraft.com), and only uniquely mapping reads were retained. Picard (http://www.picard.sourceforge.net) was used to add read groups, SAMtools (Li *et al.* , 2009) to generate a BAM file per individual, and GATK (McKenna *et al.* , 2010) to perform realignment. SAMtools and BCFtools were used to generate a VCF file. Only monoallelic and biallelic sites were retained. Single Nucleotide Polymorphisms (SNPs) and invariant sites were masked around 10bp of an indel. Sites were removed if the depth was outside 1st and 99th percentile of the overall coverage. Another custom perl script (SNPcleaner, github.com/tplinderoth/ngsQC/tree/master/snpCleaner (Bi*et al.* , 2013, 2019)) was implemented for further filtering of SNPs.

Calling SNPs and genotypes based on allele counts may be highly uncertain if coverage is low (Johnson and Slatkin, 2008; Lynch, 2008), which subsequently may bias downstream analyses. Therefore, we compared results from genotype calls and genotype likelihoods. Genotype likelihoods were generated via an empirical Bayesian framework via Analysis of Next-Generation Sequencing Data (ANGSD) (Korneliussen, Albrechtsen & Nielsen, 2014). We set genotype posterior probabilities of 0.95 as a threshold in ANGSD to output high-confidence genotypes for analyses performed in GENODIVE requiring genotype calls (Meirmans and Van Tienderen, 2004). For downstream analyses based on either genotype likelihoods and genotype calls we tested the effect of coverage (3X and 10X) and missing data included (max. 30%, 10%, 5% and 1% allowed missing data).

Microbial filtering and pattern comparison

We screened for loci from putative microbes in three different ways. First, potential bacterial, viral and human sequence contamination were removed via Blasting to reference sequences from GenBank following Maas et al., (2018) (see their Supplemental Table 1 for Genbank data used). Next, we ran Kraken (Wood and Salzberg, 2014), a fast sequence classifier to BLAST (Altschul *et al.* , 1990) our loci against bacterial databases with default settings. Finally, we used BlobTools (Laetsch and Blaxter, 2017) to taxonomically partition reads and cut off loci with >55% GC content, as we expect sponge microbes to have higher GC content than sponge hosts (Horn *et al.* , 2016). The identified microbial loci were filtered out using a custom made perl script (Bi *et al.* , 2013).

Population genetic patterns of the sponge host were contrasted to sponge microbial community patterns from five populations as studied by Ferreira *et al.* , (2020) (B.1, B.2, B.3, P.4, P.5). Two datasets from filtered 16s amplicon metabarcoding were downloaded from Ferreira*et al.* (2020): the abundance of microbial genera (24 genera total), and the presence/absence of the 35 most abundant operational taxonomic units (OTUs). We compared three levels of variation among the host genetic dataset and associated microbial community dataset: (1) among genetic lineages of the host sponge (Lineage A and B, only Lineage B, and one sub-lineage within Lineage B, as defined by Becking *et al.* (2013)), (2) among two regions >1,400km apart (Berau and Raja Ampat), and (3) among lakes within the same region (<250km). We tested whether microbial community patterns were related to sponge host population structure by running Mantel tests (Legendre and Legendre, 2012) between the Bray-Curtis dissimilarity matrix of the microbial communities and the genetic distance (F_{ST}) matrix of sponge host.

Population genetic analyses

To assess phylogeographic structure, a genetic distance matrix was computed using ngsDist using genotype likelihoods (Vieira *et al.* , 2016). Following recommendations of RAxML (Stamatakis, 2014), a bootstrapped

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

neighbor-joining tree matrix was computed from 1,000 possible trees and visualized as a phylogenetic tree using FASTME (Lefort, Desper and Gascuel, 2015) and FigTree v.1.4.2 (Rambaut, 2009).

We estimated the within-population genetic diversity of Lineage B using two diversity measures, namely expected heterozygosity and nucleotide diversity. We calculated expected heterozygosity (H_e) using GENODIVE (Meirmans and Van Tienderen, 2004), and overall heterozygosity and nucleotide diversity (π) using ANGSD (Nei, 1987).

Population structure was identified using three methods. First, we performed a neighbor-joining network (NeighborNet) analysis using Splitstree (Huson, 1998; Huson and Bryant, 2006). Splitstree does not force a tree-like structure onto the data and thus can verify the extent to which the data conform to a hierarchical tree structure. Next, we ran a Principal Component Analysis (PCA) based on a covariance matrix computed by ngsTools on genotype likelihoods (Fumagalli, Vieira and Linderoth, 2014) and via GENODIVE using genotype calls. As an unsupervised clustering method, PCA estimates population genetic structure in an unbiased way. Finally, we explored admixture patterns using ngsAdmix (Skotte, Korneliussen, 2013). Ancestry of populations was explored through calculating admixture proportions per individual and varying the estimated number of ancestral populations (K). The most likely K was determined by running 10 replicate runs of each respective K, calculating the log likelihood value of each, and choosing the value of K where an addition of an ancestral group did not result in a higher likelihood (Evanno, Regnaut and Goudet, 2005).

Normalized population differentiation was calculated using high confidence genotype calls in GENODIVE. Normalized fixation index (F'_{ST}) was calculated to eliminate the effect of within-population diversity (Meirmans and Hedrick, 2011). Finally, we constructed a migration network using Nei's G_{ST} with the threshold at 0.4 using the DiveRsity package in R (Keenan *et al.* , 2013), as demonstrated by Sundqvist *et al.* (2016).

Spatial and environmental association

Next, we used Mantel tests (Mantel, 1967; Legendre and Legendre, 2012) to test significance of correlations between genetic, geographic, environmental and connection distance matrices. For genetic distances, we used normalized pairwise genetic differentiation ($F'_{ST}/(1\text{-}F'_{ST})$ and p-distance. Geographic distance was calculated as minimum pairwise distances in meters between lakes using lake coordinates as input for the*geosphere* package in R. Using averages of 1-5m measurements of the water quality, environmental distance was calculated as pairwise Euclidean distances between locations. Connection distance was calculated following the equations of Maas et al., (2018). Mantel's tests were run with 10,000 permutations using *vegan* in R. We verified the absence of autocorrelation between geographic, connection and environmental distances using Mantel's tests. Finally, Spearman correlation tests were performed for within-population diversity indices with temperature, salinity, connection and lake area. Correlations of r [?] 0.5 were considered strong, and alpha was set to 0.05.

Results

Lake characterization and read statistics

The physical and environmental profiles of the two lagoons and nine marine lakes are provided in Table 1. In general, we observed higher temperatures (30.8°C ±1.22°C) and lower salinities (27.3ppt ±2.7ppt) in lakes than in lagoons (29°C and 33.5ppt). Connection to the surrounding sea varied among lakes, with highly connected to highly isolated lakes based on tidal amplitudes. For instance, lake P.4 was found to have the highest connection with tidal amplitude representing 80% of that of the surrounding sea, while lake P.1 was most isolated, with tidal amplitude only being 7% of the surrounding sea.

After sequencing and demultiplexing we obtained 1,127,497,643 reads from 168 sponges. On average, we obtained 7,673,269 reads per individual. Individuals with less than 2,000,000 reads were removed from subsequent analyses. Based on the calculation table from Peterson *et al.*(2012) and on an estimated genome size of 600Mb and a size selection of 425-500bp, we expected to retain 27,300 RADtags. However, our *de novo* reference retained only 14,442 tags when keeping RADtags with at least 3X coverage and tags present in at least 70% of individuals. Kraken and Blobtools identified 13 out of the 14,442 RADtags as possible bacterial

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

contamination. The RADtags mapped to *Synechococcus* sp., a Cyanobacteria genus, and were removed from the data set.

After filtering we retained 125 sponges with 973,697,804 reads in total, with coverage ranging from 3.1 - 82.2X (average 24.0X). In total, 23,742 SNPs were called over all tags, and after selecting one SNP per tag we retained 4,826 SNPs for subsequent analyses. Depending on the filtering options (genotype calls or genotype likelihoods, coverage 3X or 10X, included missing data 30%, 10%, 5% or 1%) the number of SNPs varied from 56 to 4,826 (Supplemental Table 2).

Lineage divergence

The phylogenetic tree based on pairwise genetic distances showed two divergent lineages (Fig. 2, Supplemental Table 3). These lineages are concordant with Lineage A and B as defined in Becking et al. (2013). Lineage A was only represented by individuals of B.3. The remaining populations fell under Lineage B. For this lineage, two sub-clades could be identified: B-I representing regions in Australia and Berau, and B-II representing the region Raja Ampat. For the Australia/Berau clade, lagoon populations Bay and DAR appeared to be ancestral to the lake populations. Within Raja Ampat, lake P.4 was most distant from the other lakes.

The number of genetic markers retained strongly differed depending on the inclusion/exclusion of Lineages. When including both lineages, 541 SNPs were retained. When only including Lineage B, the number of SNPs increased almost 9-fold to 4,826. All subsequent analyses were run for Lineage B for 105 individual sponges.

Within-lineage population genetics

Within Lineage B, genetic patterns remained highly similar for all filters (but see supplemental figures and tables for differences). As conclusions remained the same, all further reported analyses were performed filtering on 3X coverage and max. 30% missing data, as this retained the most SNPs.

Population genetic diversity varied among lakes (Table 1, Supplemental Table 4). The highest genetic diversity was consistently found for the lagoon populations Bay and DAR, as seen for nucleotide diversity (π) (0.0101 and 0.0095, respectively), and for the expected heterozygosity (H_e) (0.157 and 0.117, respectively). Lowest genetic diversity was observed in populations P.1 ($\pi = 0.0036$, $H_e = 0.054$) and P.27 ($\pi = 0.0037$, $H_e = 0.038$). Population B.3 also showed low heterozygosity ($H_e = 0.034$), but relatively high nucleotide diversity ($\pi = 0.0074$). However, this may be an artefact of low sample size. When estimating heterozygosity from genotype likelihoods via ANGSD, we found the lowest heterozygosity for the populations P.5 (0.019) and P.27 (0.021).

The samples clustered per lake and lagoon location (Fig. 3, Supplemental Fig. 1, Supplemental Fig. 2). The first four Principal Components (PCs) in the Principal Component Analysis (PCA) explained 80.5% of total variation (Fig. 3A). PC1, explaining 45.6% of the variation, separated populations by geographic region, with the Raja Ampat lakes being distinct from the lakes in Berau. PC2, explaining 24.4% of variation, separated lake MIS01 from the other lakes. PC3 and PC4 (explaining 10.5% in total) further separated lagoon DAR and lakes P.5, and to a lesser extend P.1 and P.30. In the PC1 versus PC2 plot the lagoon populations (Bay and DAR) clustered towards the center of the graph, indicating them to be ancestral. For Bay, this continued for the PC3 versus PC4 plot, but not for DAR. Lakes P.27 and P.32 remained closely associated.

The Admixture analysis further supported the pattern of clustering per lakes (Fig. 3B). Convergence of likelihood values indicated the number of ancestral populations to be K = 9 (Supplemental Fig. 3, 4). When putative number of populations was set to 9, all populations were separated apart from B.2, which consisted of a mix of Bay and B.1 genetic lineages. Some admixture of B.1 genetic diversity into Bay and DAR populations was observed, indicating some genetic connection between these populations. Setting K at 7 or 8 indicated some admixture between P.30 and P.5 (K=8) or among P.27 and P.32 with Bay being a mixture of other populations (K=7). Setting K at 10 separated all populations.

Findings form the phylogenetic network were consistent with patterns found for PCA and Admixture plots

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

(Fig. 3C, Supplemental Fig. 5). The network showed a high fit (fit = 99.2) and small degree of reticulation (d = 0.153), thus indicating a tree-like structure. The lagoon populations Bay and DAR showed higher reticulation than the marine lake populations, indicating higher intra-population diversity.

Pairwise fixation indices (F'_{ST}) showed high levels of genetic structuring (0.629 ±0.133) (Fig. 3D). The F'_{ST}ranged from 0.182 between Bay and B.2 to 0.778 between P.30 and P.32 (Supplemental Fig. 6, Supplemental Table 5). All pairwise comparisons were significant, except for the comparison between P.32 and B.2, potentially due to sample size (n = 4 and 2, respectively. The migration network among lakes indicated strongest relative bidirectional migration among lakes in Berau (Fig. 1D). Lagoon population Bay was linked to some degree to all other populations (relative fraction 0.4-1). Within Raja Ampat, bidirectional migration above the threshold of 0.4 was observed between P.5 and four other lakes (P.30, P.32, P.1, and P.4). There was low connectivity among lakes P.27, P.30, P.32 and P.1 in Raja Ampat (>0.4).

Association to drivers

Within-population genetic diversity (nucleotide diversity π and heterozygosity H_e) was not influenced by lake area (π: Spearman's rho = 0.03, $p = 0.95$, H_e: rho = -0.06,$p = 0.88$), connection (π: rho = 0.43, $p = 0.25$, H_e: rho = 0.53, $p = 0.15$), salinity (π: rho = 0.37, $p = 0.33$, H_e: rho = 0.24, $p = 0.53$) (Supplemental Table 6, Supplemental Figure 7). However, there appeared to be a trend towards higher nucleotide diversity with lower water temperatures (rho = -0.61, $p = 0.08$), but not for heterozygosity (rho = -0.20, $p = 0.60$).

Mantel tests indicated no correlation between the geographic and genetic distance matrices over all filter options (Figure 4A, Supplemental Table 7, r = 0.007, $p = 0.504$). Finding no correlation refutes the isolation-by-distance hypothesis and indicates other factors might explain the distribution of *S. diversicolor* genetic diversity. However, the genetic distance matrix also did not correlate with matrices of environmental distance (r = 0.002, $p = 0.503$, Figure 4B) or connection distance (r = 0.041, $p = 0.441$, Figure 4C).

Comparison host population structure to microbial community patterns

Genetic structure from sponge hosts (Fig 3A) was contrasted to the microbial community dataset of *S. diversicolor* from a subset of the same lakes, collected by Ferreira *et al* 2020 (Supplemental Figure 8). The microbial community did not appear to cluster between lineages (Lineage A versus Lineage B), not between subclades within lineage B. As for geographic regions, a clear distinction of broad geographic regions (Berau versus Raja Ampat, >1,400km) could be observed in microbial community patterns, concordant to sponge host patterns. However, no clear clustering was seen for microbial communities on smaller spatial scales (<250km). Mantel tests indicated no significant correlation between host genetic distances (F_{ST}) and microbial community dissimilarities based on genus abundance (Mantel r = 0.15, $p = 0.37$) or on OTUs (r = 0.01,$p = 0.46$) (Figure 4D).

Discussion

A major objective of marine molecular ecology is to obtain accurate estimates of subtle genetic structure, as it can inform efforts to identify units of management and design effective marine protected areas (Kelley *et al.* , 2016; Selkoe *et al.* , 2016). By comparing sponge populations in Indo-Pacific marine lakes and lagoons at different spatial scales, varying in degree of connection to the sea and differing in local environmental conditions, we were able to study fine-scaled genetic structure and the drivers of genetic diversity and differentiation of marine populations. Using a reduced representation genomic approach, we confirmed broad-scale patterns of structure identified in a prior single-marker study and provided new evidence of small-scaled structure for sessile species with a short dispersive larval stage. Furthermore, we found no associations between sponge host population patterns and previously studied microbial community patterns, suggesting that sponge host and associated microbes may be on different eco-evolutionary tracts. Below, we discuss our findings on population structure for marine lake sponges, the possible drivers of diversity, microbe versus host patterns and finally the implications for future phylogeographic and population genetic studies on sponges.

Population genetics of sponges in marine lakes

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

The use of thousands of RADseq-based SNPs provided the resolution necessary to reveal genetic patterns of *Suberites diversicolor*that had not previously been captured at finer spatial scales. We observed clear clustering for the marine lake locations per lake. The lagoon populations Bay and to a lesser extend DAR showed to hold a basal position in the phylogenetic tree and in PCAs. They also showed links to most other populations in the migration network. The presence of ancestral polymorphisms in the marine lake populations could explain this pattern. The observation of finding more structure when using higher numbers of genetic markers has been shown in other marine organisms as well (Bradbury *et al.* , 2015; Maas *et al.* , 2018; Lemopoulos *et al.* , 2019; D'Aloia *et al.* , 2020; Sunde*et al.* , 2020; Timm, 2020). In a comparison among three high-throughput genotyping approaches, the RADseq generated markers were found to be the most sensitive and robust in detecting fine-scaled structure (D'Aloia *et al.* , 2020). The discrepancy in observed genetic structure based on a higher number of markers as compared to single markers is important in interpreting results from other studies for sessile marine organisms using low resolution markers.

While traditional, low-resolution markers have been useful in exposing morphologically cryptic sponge species, they have often failed to detect within-species diversity (as reviewed in Oppen et al., 2002; Pérez-Portela & Riesgo, 2018; Uriz & Turon, 2012; Wörheide et al., 2005). Using the high resolution of RAdseq generated markers allowed us to see clear clustering per lake even on very small spatial scales 1-10km. The scale at which we find strong structure is smaller compared to studies using microsatellites in the sponges *Crambe crambe*(Duran *et al.* , 2004), *Scopalina lophyropoda* (Blanquer and Uriz, 2010), *Spongia lamella* (Noyer and Becerro, 2012; Pérez-Portela, Noyer and Becerro, 2015), *Stylissa carteri* (Giles*et al.* , 2015), *Cliona delitrix* (Chaves-Fonnegra *et al.* , 2015), *Xestospongia muta* (Richards *et al.* , 2016),*Paraleucilla magna* (Guardiola, Frotscher and Uriz, 2016)*, Plenaster cragi* (Taboada *et al.* , 2018) and *Petrosia ficiformis* (Riesgo *et al.* , 2019). Even studies using higher resolution markers also little structure at small spatial scales, with Brown *et al.* (2017) detecting little structuring for*Aphrocallistes vastus* in British Colombia at scales <275km and Leiva *et al.* (2019) finding panmixia at scales >900km for *Dendrilla antarctica* . It could be that these are highly connected populations, possibly through rafting or sperm-mediated gene flow (Maldonado, 2006; DeBiasse, Nelson and Hellberg, 2014). Yet it is also possible that the number of SNPs from Brown *et al.* (2017) and Leiva *et al.* (2019) (67 and 529, respectively) was too low to detect subtle structure at small scales. Alternatively, the filtering strategy of these studies possibly was not rigorous enough to eliminate sufficient or all microbial contamination, possibly clouding patterns.

We assessed the effects of several drivers of population diversity and structure. First, we tested to what extent marine genetic differentiation conforms to the decay of population similarity with geographical distance resulting in a pattern of isolation-by-distance (Wright, 1943) using only Lineage B. We found strong population structure with clustering per lake, yet no pattern of isolation-by-distance was observed. This is remarkable, since we sampled at geographical distances of 1km - 1,400km. We also did not detect a pattern of isolation-by-environment, despite the great environmental variability among lakes (temperature: 29 - 32.4 degC, salinity: 24 - 33.4 ppt). Previous studies using a low number of markers did find a pattern of isolation-by-distance for sponges (Duran *et al.* , 2004; Blanquer and Uriz, 2010; Noyer and Becerro, 2012; Perez-Portela, Noyer and Becerro, 2015), which is usually expected for species with restricted dispersal abilities (Worheide, Sole-Cava and Hooper, 2005; Maldonado, 2006). Other studies report an influence of oceanographic currents (Chaves-Fonnegra *et al.* , 2015; Richards *et al.* , 2016; Riesgo *et al.* , 2019), or environmental heterogeneity (temperature and productivity) (Giles *et al.* , 2015) on sponges. Our results indicate that mechanisms other than only dispersal limitation by geographical distance or local environments are important in structuring *S. diversicolor* populations. In addition, the permeability of the landscape barrier surrounding the marine lakes, determining the degree of water flowing in and out of the lakes, did not seem to influence the population structure or diversity. Perhaps*S. diversicolor* populations are truly isolated per lake as their low dispersal ability restricts effective gene flow. Then, populations can become differentiated through genetic drift or via local adaptation to environmental parameters that we have not recorded (Frankham, Briscoe and Ballou, 2002). Alternatively, founder effects and subsequent priority effects could explain the pattern (Orsini *et al.* , 2013; Fukami, 2015; De Meester *et al.* , 2016).

Priority effects were previously discussed as potential drivers of structure in marine lake organisms by

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Maas *et al.* (2018) and de Leeuw *et al.* (2020). Depending on spatial scale Maas *et al.* (2018) found an effect of geographic distance and connectivity influencing mussel population structure. They argued that despite founder events stochastically driving alleles to fixation in small populations, ongoing dispersal would overwhelm this effect (Mayr, 1963; Waters, Fraser and Hewitt, 2013). Mussels have extensive pelagic larval duration periods, and Maas *et al.* (2018) hence argued that priority effects mediated by local adaptation could facilitate the observed patterns of population structure (Orsini *et al.* , 2013; Fukami, 2015; De Meester *et al.* , 2016). Sponges, in contrast, generally have poor dispersal abilities (Maldonado, 2006). As the current study does not find an effect of connection to the sea in structuring populations, stochastic fixation of alleles due to genetic drift may be the cause of each population being distinct. Including more lakes with replicates of local environments and/or connection to the sea may further elucidate drivers of sponge differentiation in fragmented habitats.

Microbes vs. host patterns

Suberites diversicolor is categorized as a Low Microbial Abundance (LMA) sponge (Weisz, Lindquist and Martens, 2008; Cleary*et al.* , 2013). A recent study into prokaryotic communities of*S. diversicolor* (5 lakes overlap with the current study) showed a pattern of distinct broad geographic groups (>1,400km) but no distinction between host genetic lineages and only some clustering per lake (Ferreira *et al.* , 2020). When comparing the microbial patterns observed by Ferreira *et al* (2020) to genetic structure of *S. diversicolor* from the current study, interestingly we find no relationship apart from the broad geographic distinction (>1,400km). This is consistent with a study by Noyer and Becerro, (2012) that did not find correlations between host genetic and microbial diversity for the sponge *Spongia lamella* in the Mediterranean (<500km). Perhaps microbial communities are evolving separately from their host for some sponges, such as *S. diversicolor* . Or perhaps the microbial community was not measured extensively enough since there are known differences between the relatively stable core microbiome and the environmentally variable microbiome (Pita *et al.* , 2018).

Changing oceans may shift symbioses of sponge holobiont (Fan *et al.* , 2013). It is already known that sponge-associated microbiomes can respond to temperature (Webster, Cobb and Negri, 2008), although dependent on the extent of the heat stress (Simister *et al.* , 2012), and pH (Cleary *et al.* , 2013; Morrow, Fiore and Lesser, 2016; Coelho *et al.* , 2018). Responses of microbes may have effects on host persistence and viability. It is expected that the microbiome can respond more quickly than the host to changing environments due to shorter generation times (Reshef *et al.* , 2006; Pita *et al.* , 2018). However, here, the opposite seems to be the case, where the sponge host is adjusting to specific marine lake environments (or genetic drift) while microbes appear to remain stable along the marine lake gradient. Moving on from 16S-amplicon sequencing to whole bacterial genome or gene expression analyses (Liu *et al.* , 2012) will allow for a better understanding of microbial community structure and function, and could provide the necessary depth to move forward in exploring how sponges function as holobionts.

Implications for sponge phylogeography and population genetic studies

The RADseq strategy was effective in detecting two major genetic lineages (Lineage A and B) (Becking *et al.* , 2013). When combining both lineages significantly less markers were recovered than when analyzing lineages separately. Based on our filters requiring a read depth of at least 3X and loci having to be present in at least 70% of the individuals, we retained 541 SNPs when including both lineages, compared to 4,826 SNPs when analyzing only Lineage B. This is more than a 90% loss of common markers and indicates the resolution of RADseq generated markers can be less effective when one (unknowingly) includes multiple lineages. Given that there is a prevalence in morphologically cryptic species in sponges (e.g. Becking, 2013; Swierts *et al.* , 2013; Morrow and Cardenas, 2015), it may be advised to first verify broad genetic lineages using common single markers before starting an extensive sponge population genetic study implementing high resolution markers. Perhaps the low number of SNPs recovered in the previous two studies on sponges (Brown, Davis and Leys, 2017; Leiva *et al.* , 2019) was caused by including different lineages. Further, our adjustments to the existing low-cost protocol of Peterson *et al.*(2012) with a step-by-step protocol presented in Maas *et al.*(2018) can help to retrieve extensive data for non-model marine organisms in general and

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

tropical sponges in particular, thus benefitting future studies. Our strategy for bioinformatically filtering out possible microbial contamination proved effective since we did not detect congruence between host population structure and microbial community structure. We further showed that reduced representation genome sequencing can work for DNA that was extracted for other purposes and stored for long times in a -20degC freezer before sequencing, or suboptimal removal of contamination before sequencing. Recent developments with capture based methods such as hyRAD (Suchan *et al.* , 2016) can further exploit the potential of older DNA extractions. This gives hope to the wealth of knowledge to be gained from extractions from past sponge studies across the world.

Assessing genetic connectivity between populations is crucial to determine the scale of marine reserves (Richards *et al.* , 2016). Moving on from studying few genetic markers to (reduced representation) genome sequencing provides the potential to look for genetic basis of adaptation (Catchen *et al.* , 2017), a major goal in molecular ecology. As the oceans are changing, it is imperative how sponge host and its associated microbe community will respond (Pita *et al.* , 2018). Understanding within- and between population diversity, demography and connectivity serves to facilitate conservation management decisions. It is important to ensure connectivity between marine populations where necessary, while also allowing crucial local adaptation, in view of projected climate change and habitat fragmentation (IPCC, 2019).

Acknowledgements

The following people provided help with logistics during fieldwork and/or labwork: . B. Alvarez, M. Ammer, Bahruddin, , H. Breeuwer, L. Dong, M. Erdmann, B. Hoeksema, P. Kuperus, S.C. Lim, S. Menken, A. Miners, A. Oherenan, K. Peijnenburg, W. Renema, Suharsono, R. Tapilatu, Y. Tuti, B. Voetdijk, N.J. de Voogd, the staff of Nabucco Island Dive Resort, Derawan Dive Resort, Misool Eco resort, Papua Diving. We are grateful for useful discussions with Dr. P. Meirmans. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation grants S10RR029668 and S10RR027303. The Netherlands Organisation for Scientific Research provided funding to Leontine E. Becking through the grants RUBICON #825.12.007 and VENI #863.14.020. Fieldwork in Indonesia was made possible through additional financial support of National Geographic Society Waitt Grant, De Beukelaar-van der Hucht Stichting, World Wildlife Fund, Netherlands-INNO Fund, the Schure-Beijerinck-Popping Fund of the Royal Dutch Academy of Science (KNAW), Conservation International Ecosystem Based Management program (funded by the David and Lucile Packard Foundation), the Treub- Maatschappij Fund, the Leiden University Fund (LUF)/Slingelands, Singapore Airlines, the A.M. Buitendijk Fund and the J.J. ter Pelkwijk Fund (Naturalis Biodiveristy Center). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are grateful to the Indonesian Institute of Sciences (LIPI) and the Indonesian State Ministry of Research and Technology (RISTEK) for providing the research permits (0094/FRP/SM/V/2009;098/SIP/FRP/SM/V/2011;098/SIP/FRP/SM/V/20113246/FRP/SM/VII/2012;#3B/TKPIPA/E5/Dit.KI/III/2016).

Data accessibility

Data is accessible via the Dryad Digital Repository (*to be uploaded*). Data consists of aligned reads in bam format for all 125 individuals sequenced. An accompanying text file assigns individuals. Also, the de-novo reference file in fasta format is included. Finally, the RADTOOLKIT v. 0.13.10 is stored here.

Author contributions

D.L. Maas performed research, analyzed data and wrote the paper. **S. Prost** analyzed data, wrote the paper and provided feedback. **C.A. de Leeuw** analyzed data and provided feedback.**K. Bi** contributed analytical tools, analyzed data and provided feedback. **L. Smith** contributed reagents and lab support and provided feedback. **Purwanto** contributed logistical support and provided feedback. **L.P. Aji** contributed logistical support and provided feedback. **R.F. Tapilatu** contributed logistical support and provided feedback. **R. Gillespie** contributed to research design and provided feedback. **L.E. Becking** designed research, performed research, wrote the paper and provided feedback.

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Tables and Figures

Table 1: Overview of sampling in marine lakes and lagoon locations. Recorded are location, site codes, number of individuals sampled per site and number of individuals retained after filtering, physiographic, environmental and genetic parameters.

Code	Location	#Samples Total	# Samples After filter	Area (m^2)	Depth (m)	Fraction (lake/sea)	Connection	Temperature (°C)	Salinity (ppt)	Nucleotid diveristy (π)
DAR	Australia	8	7	45,640			Open			0.0095
Bay	Berau	5	5				Open	29	33.5	0.0101
B.1	Berau: Maratua	29	26	140,000	17	0.51	Medium	29.5	27	0.0050
B.2	Berau: Tanah Banban	4	2	231,500		0.38	Low	29.5	26	0.0074
B.3	Berau: Kakaban	32	20	4,900,000	12	0.11	Low	30	23.5	
P.27	Papua: Wayag	8	5	22,000	2		Medium	29.5	31	0.0037
P.30	Papua: Wayag	9	8	13,000	4.1	0.75	Medium	32.4	28.9	0.0045
P.32	Papua: Wayag	7	4	6,100	5.5	0.45	Medium	31.2	30.7	0.0053
P.1	Papua: Gam	20	11	88,530	19	0.07	Low	32.3	24	0.0036
P.4	Papua: Misool	26	19	13,750	20.4	0.8	High	31.7	25.9	0.0047
P.5	Papua: Misool	20	18	3,700	4.8	0.26	Low	31.5	28.9	0.0060

Hosted file

`image1.emf` available at https://authorea.com/users/352398/articles/476687-previously-unidentified-genetic-structure-revealed-for-the-sponge-suberites-diversicolor-implications-for-sponge-phylogeography-and-population-genetics

Figure 1: Sampling sites of *Suberites diversicolor* from nine marine lakes and two lagoon locations and associated relative migration networks. (A) Overview of Indonesia including two geographic regions sampled: Berau and Raja Ampat. Also shows location of Australian lagoon location (DAR). (B) Berau, East-Kalimantan with locations of three marine lakes (B.1, B.2, B.3) and one lagoon (Bay). (C) Raja Ampat, West-Papua with locations of six marine lakes (P.27, P.30, P.32, P.1, P.4, P.5). (D) Relative migration network of of individual lakes and lagoons. Fractions of relative migration are displayed. (E) Specimen of*S. diversicolor* , photograph by L.E. Becking.

Hosted file

`image2.emf` available at https://authorea.com/users/352398/articles/476687-previously-unidentified-genetic-structure-revealed-for-the-sponge-suberites-diversicolor-implications-for-sponge-phylogeography-and-population-genetics

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Figure 2: Neighbor-Joining tree based on pairwise genetic distances of*Suberites diversicolor* populations. Bootstrap support values are displayed based on 1000 bootstraps. Each branch represents one individual. Colors and codes correspond to Fig. 1 and Table 1.

Hosted file

`image3.emf` available at https://authorea.com/users/352398/articles/476687-previously-unidentified-genetic-structure-revealed-for-the-sponge-suberites-diversicolor-implications-for-sponge-phylogeography-and-population-genetics

Figure 3: Analyses of population genetic structure for *Siberites diversicolor* populations. (A) Principal Component analysis (PCA) based on pairwise covariance. Each dot represents one individual. (B) Bayesian admixture analysis for most likely putative ancestral populations (K = 9) based on genotype likelihoods via ngsAdmix. Each bar represents one individual. (C) Neighbor-Joining Network with equal angles computed in Splitstree based on pairwise genetic distances. (D) Visualization of normalized F'_{ST} in a multidimensional scaling plot (values in Supplemental Table 4). Colors and codes correspond to Fig. 1 and Table 1.

Hosted file

`image4.emf` available at https://authorea.com/users/352398/articles/476687-previously-unidentified-genetic-structure-revealed-for-the-sponge-suberites-diversicolor-implications-for-sponge-phylogeography-and-population-genetics

Figure 4: Correlational tests (Mantel) between genetic distance matrix (F'ST) versus A) drivers of diversity geographic distance, environmental distance and connection distance, and B) Microbial community matrices based on Bray Curtis dissimilarities for genera abundance and OTU presence/absence.

References

Allendorf, F. W., Hohenlohe, P. A. and Luikart, G. (2010) 'Genomics and the future of conservation genetics', *Nature Reviews Genetics* . Nature Publishing Group, 11(10), pp. 697–709. doi: 10.1038/nrg2844.

Altschul, S. F. *et al.* (1990) 'Basic Local Alignment Search Tool', *Journal of Molecular Biology* , 215(3), pp. 403–410. Available at: https://doi.org/10.1016/S0022-2836(05)80360-2.

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

Avise, J. C. (2009) 'Phylogeography: Retrospect and prospect',*Journal of Biogeography* , 36(1), pp. 3–15. doi: 10.1111/j.1365-2699.2008.02032.x.

Azzini, F. *et al.* (2007) 'Sponges of the marine karst lakes and of the coast of the islands of Ha Long Bay (North Vietnam)', in*Porifera Research: Biodiversity, Innovation and Sustainability* , pp. 157–164.

Baird, N. A. *et al.* (2008) 'Rapid SNP discovery and genetic mapping using sequenced RAD markers', *PLoS ONE* , 3(10), pp. 1–7. doi: 10.1371/journal.pone.0003376.

De Bakker, D. M. *et al.* (2016) 'Population genetic structure, abundance, and health status of two dominant benthic species in the saba bank national park, caribbean netherlands:Montastraea cavernosa and xestospongia muta', *PLoS ONE* , 11(5), pp. 1–20. doi: 10.1371/journal.pone.0155969.

Becking, L. E. *et al.* (2011) 'Recently discovered landlocked basins in Indonesia reveal high habitat diversity in anchialine systems', *Hydrobiologia* , 677(1), pp. 89–105. doi: 10.1007/s10750-011-0742-0.

Becking, L. E. *et al.* (2013) 'Phylogeography of the Sponge Suberites diversicolor in Indonesia: Insights into the Evolution of Marine Lake Populations', *PLoS ONE* , 8(10). doi: 10.1371/journal.pone.0075996.

Becking, L. E. (2013) 'Revision of the genus Placospongia (Porifera, Demospongiae, Hadromerida, Placospongiidae) in the Indo-West Pacific',*ZooKeys* , 298, pp. 39–76. doi: 10.3897/zookeys.298.1913.

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Becking, L. E. *et al.* (2016) 'Highly divergent mussel lineages in isolated Indonesian marine lakes', *PeerJ* , 4, p. e2496. doi: 10.7717/peerj.2496.

Becking, L. E., de Leeuw, C. and Vogler, C. (2015) 'Newly discovered "jellyfish lakes" in Misool, Raja Ampat, Papua, Indonesia',*Marine Biodiversity* , 45(4), pp. 597–598. doi: 10.1007/s12526-014-0268-6.

Becking, L. E. and Lim, S. (2009) 'A new Suberites (Demospongiae: Hadromerida: Suberitidae) from the tropical Indo-West Pacific',*Zoologische Mededelingen* , 83(29), pp. 853–862. Available at: http://dpc.uba.uva.nl/cgi/t/text/get-pdf?idno=m8304a29;c=zoomed.

Bell, J. J. (2008) 'The functional roles of marine sponges',*Estuarine, Coastal and Shelf Science* , 79(3), pp. 341–353. doi: 10.1016/j.ecss.2008.05.002.

Bell, J. J. *et al.* (2013) 'Could some coral reefs become sponge reefs as our climate changes?', *Global Change Biology* , 19(9), pp. 2613–2624. doi: 10.1111/gcb.12212.

Bentlage, B. and Wörheide, G. (2007) 'Low genetic structuring among Pericharax heteroraphis (Porifera: Calcarea) populations from the Great Barrier Reef (Australia), revealed by analysis of nrDNA and nuclear intron sequences', *Coral Reefs* , 26(4), pp. 807–816. doi: 10.1007/s00338-007-0267-1.

Bi, K. *et al.* (2013) 'Unlocking the vault: Next-generation museum population genomics', *Molecular Ecology* , 22(24), pp. 6018–6032. doi: 10.1111/mec.12516.

Bi, K. *et al.* (2019) 'Temporal genomic contrasts reveal heterogeneous evolutionary responses within and among montane chipmunk species during recent climate change', *PLoS Genetics* , 15, p. e1008119. doi: 10.5281/Abstract.

Blanquer, A. and Uriz, M. J. (2010) 'Population genetics at three spatial scales of a rare sponge living in fragmented habitats',*BMC Evolutionary Biology* , 10(1). doi: 10.1186/1471-2148-10-13.

Bolger, A. M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: A flexible trimmer for Illumina sequence data', *Bioinformatics* , 30(15), pp. 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bowen, B. W. *et al.* (2013) 'The origins of tropical marine biodiversity', *Trends in Ecology and Evolution* . Elsevier Ltd, 28(6), pp. 359–366. doi: 10.1016/j.tree.2013.01.018.

Bradbury, I. R. *et al.* (2015) 'Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure', *Molecular Ecology* , 24(20), pp. 5130–5144. doi: 10.1111/mec.13395.

Brown, R. R., Davis, C. S. and Leys, S. P. (2014) 'SNP discovery in a reef-forming glass sponge, Aphrocallistes vastus, using the Ion Torrent next generation sequencing platform', *Conservation Genetics Resources* , 6(1), pp. 49–51. doi: 10.1007/s12686-013-0073-8.

Brown, R. R., Davis, C. S. and Leys, S. P. (2017) 'Clones or clans: the genetic structure of a deep-sea sponge, Aphrocallistes vastus, in unique sponge reefs of British Columbia, Canada', *Molecular Ecology* , 26(4), pp. 1045–1059. doi: 10.1111/mec.13982.

Catchen, J. M. *et al.* (2017) 'Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations', *Molecular Ecology Resources* , 17(3), pp. 362–365. doi: 10.1111/1755-0998.12669.

Chaves-Fonnegra, A. *et al.* (2015) 'Population structure and dispersal of the coral-excavating sponge Cliona delitrix',*Molecular Ecology* , 24(7), pp. 1447–1466. doi: 10.1111/mec.13134.

Cleary, D. F. R. *et al.* (2013) 'Habitat- and host-related variation in sponge bacterial symbiont communities in Indonesian waters', *FEMS Microbiology Ecology* , 85(3), pp. 465–482. doi: 10.1111/1574-6941.12135.

Cleary, D. F. R. *et al.* (2019) 'The sponge microbiome within the greater coral reef microbial metacommunity', *Nature Communications* . Springer US, 10(1), pp. 1–12. doi: 10.1038/s41467-019-09537-8.

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Cleary, D. F. R., Polónia, A. R. M. and de Voogd, N. J. (2018) 'Bacterial Communities Inhabiting the Sponge Biemna fortis, Sediment and Water in Marine Lakes and the Open Sea', *Microbial Ecology* . Microbial Ecology, 76(3), pp. 610–624. doi: 10.1007/s00248-018-1156-6.

Coelho, F. J. R. C. *et al.* (2018) 'Sponge Prokaryote Communities in Taiwanese Coral Reef and Shallow Hydrothermal Vent Ecosystems',*Microbial Ecology* . Microbial Ecology, 75(1), pp. 239–254. doi: 10.1007/s00248-017-1023-x.

Cornwell, B. H. *et al.* (2016) 'Chaotic genetic patchiness without sweepstakes reproduction in the shore crab Hemigrapsus oregonensis',*Marine Ecology Progress Series* , 548, pp. 139–152. doi: 10.3354/meps11628.

Costello, M. J. and Chaudhary, C. (2017) 'Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation', *Current Biology* . Elsevier, 27(11), pp. R511–R527. doi: 10.1016/j.cub.2017.04.060.

D'Aloia, C. C. *et al.* (2020) 'Unraveling hierarchical genetic structure in a marine metapopulation: A comparison of three high-throughput genotyping approaches', *Molecular Ecology* , 29(12), pp. 2189–2203. doi: 10.1111/mec.15405.

Dawson, M. N. *et al.* (2009) 'Marine Lakes', in Gillespie, R. and Clague, D. A. (eds) *Encyclopedia of Islands* . University of California Press, pp. 603–607.

DeBiasse, M. B., Nelson, B. J. and Hellberg, M. E. (2014) 'Evaluating summary statistics used to test for incomplete lineage sorting: Mito-nuclear discordance in the reef sponge Callyspongia vaginalis',*Molecular Ecology* , 23(1), pp. 225–238. doi: 10.1111/mec.12584.

Diaz, M. C. and Rützler, K. (2001) 'Sponges: An essential component of Caribbean coral reefs', *Bulletin of Marine Science* , 69(2), pp. 535–546.

Dunn, C. W., Leys, S. P. and Haddock, S. H. D. (2015) 'The hidden biology of sponges and ctenophores', *Trends in Ecology and Evolution* . Elsevier Ltd, 30(5), pp. 282–291. doi: 10.1016/j.tree.2015.03.003.

Duran, S. *et al.* (2004) 'Strong population structure in the marine sponge Crambe crambe (Poecilosclerida) as revealed by microsatellite markers', *Molecular Ecology* , 13(3), pp. 511–522. doi: 10.1046/j.1365-294X.2004.02080.x.

Duran, S., Pascual, M. and Turon, X. (2004) 'Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge Crambe crambe (Poecilosclerida)', *Marine Biology* , 144(1), pp. 31–35. doi: 10.1007/s00227-003-1178-5.

Easson, C. G. *et al.* (2020) 'Host population genetics and biogeography structure the microbiome of the sponge Cliona delitrix',*Ecology and Evolution* , (October 2019), pp. 2007–2020. doi: 10.1002/ece3.6033.

Easson, C. G. and Thacker, R. W. (2014) 'Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges', *Frontiers in Microbiology* , 5(OCT), pp. 1–11. doi: 10.3389/fmicb.2014.00532.

Ekins, M. *et al.* (2016) 'Staying well connected-Lithistid sponges on seamounts', *Journal of the Marine Biological Association of the United Kingdom* , 96(2), pp. 437–451. doi: 10.1017/S0025315415000831.

Evanno, G., Regnaut, S. and Goudet, J. (2005) 'Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study', *Molecular Ecology* , 14(8), pp. 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x.

Fan, L. *et al.* (2013) 'Marine microbial symbiosis heats up: The phylogenetic and functional response of a sponge holobiont to thermal stress', *ISME Journal* . Nature Publishing Group, 7(5), pp. 991–1002. doi: 10.1038/ismej.2012.165.

Felsenstein, J. (2004) *Inferring phylogenies* . Sunderland, MA: Sinauer associates.

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Ferreira, M. *et al.* (2020) 'Geographical location and habitat predict variation in prokaryotic community composition of Suberites diversicolor', *Annals of Microbiology* . Annals of Microbiology, 70(14), pp. 1–12.

Frankham, R., Briscoe, D. A. and Ballou, J. D. (2002) *Introduction to conservation genetics* . Cambridge University Press.

Fu, L. *et al.* (2012) 'CD-HIT: Accelerated for clustering the next-generation sequencing data', *Bioinformatics* , 28(23), pp. 3150–3152. doi: 10.1093/bioinformatics/bts565.

Fukami, T. (2015) 'Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects', *Annu. Rev. Ecol. Evol. Syst.* , 46, pp. 1–23. doi: 10.1146/annurev-ecolsys-110411-160340.

Fumagalli, M., Vieira, F. G. and Linderoth, T. (2014) 'ngsTools : methods for population genetics analyses from Next-Generation Sequencing data (c) The Authors 2014 . Published by Oxford University Press on behalf of the Japanese', *Bioinformatics* , pp. 1–2. doi: 10.1093/bioinformatics/btu041.

Giles, E. C. *et al.* (2015) 'Exploring seascape genetics and kinship in the reef sponge Stylissa carteri in the Red Sea',*Ecology and Evolution* , 5(13), pp. 2487–2502. doi: 10.1002/ece3.1511.

De Goeij, J. M. *et al.* (2013) 'Surviving in a marine desert: The sponge loop retains resources within coral reefs', *Science* , 342(6154), pp. 108–110. doi: 10.1126/science.1241981.

Gotoh, R. O. *et al.* (2011) 'Population genetic structure of the striped silverside, Atherinomorus endrachtensis (Atherinidae, Atheriniformes, Teleostei), inhabiting marine lakes and adjacent lagoons in Palau: marine lakes are "Islands" for marine species', *Genes & Genetic Systems* , 86(5), pp. 325–337. doi: 10.1266/ggs.86.325.

Guardiola, M., Frotscher, J. and Uriz, M. J. (2016) 'High genetic diversity, phenotypic plasticity, and invasive potential of a recently introduced calcareous sponge, fast spreading across the Atlanto-Mediterranean basin', *Marine Biology* . Springer Berlin Heidelberg, 163(5), pp. 1–16. doi: 10.1007/s00227-016-2862-6.

Hamner, W. M., Gilmer, R. W. and Hamner, P. P. (1982) 'The physical, chemical, and biological characteristics of a stratified, saline, sulfide lake in Palau', *Limnology and Oceanography* , 27(5), pp. 896–909. doi: 10.4319/lo.1982.27.5.0896.

Hardoim, C. C. P. and Costa, R. (2014) 'Temporal dynamics of prokaryotic communities in the marine sponge Sarcotragus spinosulus',*Molecular Ecology* , 23(12), pp. 3097–3112. doi: 10.1111/mec.12789.

Holthuis, L. B. (1973) 'Caridean shrimps found in land-locked salt-water pools at four Indo-West Pacific localities (Sinai Peninsula, Funafuti Atoll, Maui and Hawaii Islands), with the description of one new genus and four new species', *Zoologische Verhandelingen* , 128, pp. 1–48.

Horn, H. *et al.* (2016) 'An Enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes', *Frontiers in Microbiology* , 7(NOV), pp. 1–15. doi: 10.3389/fmicb.2016.01751.

Huang, D. *et al.* (2008) 'Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding', *Journal of Molecular Evolution* , 66(2), pp. 167–174. doi: 10.1007/s00239-008-9069-5.

Huson, D. H. (1998) 'SplitsTree: analyzing and visualizing evolutionary data.', *Bioinformatics (Oxford, England)* , 14(1), pp. 68–73. doi: btb043 [pii].

Huson, D. H. and Bryant, D. (2006) 'Application of phylogenetic networks in evolutionary studies', *Molecular Biology and Evolution* , 23(2), pp. 254–267. doi: 10.1093/molbev/msj030.

IPCC (2019) *IPCC Special Report on the Ocean and Cryosphere in a Chcanging Climate* .

Itescu, Y. (2018) 'Are island-like systems biologically similar to islands? A review of the evidence', *Ecography* , (1967), pp. 1298–1314. doi: 10.1111/ecog.03951.

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Jeffery, N. W., Jardine, C. B. and Ryan Gregory, T. (2013) 'A first exploration of genome size diversity in sponges', *Genome* , 56(8), pp. 451–456. doi: 10.1139/gen-2012-0122.

Johnson, P. L. F. and Slatkin, M. (2008) 'Accounting for bias from sequencing error in population genetic estimates', *Molecular Biology and Evolution* , 25(1), pp. 199–206. doi: 10.1093/molbev/msm239.

Keenan, K. *et al.* (2013) 'DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors', *Methods in Ecology and Evolution* , 4(8), pp. 782–788. doi: 10.1111/2041-210X.12067.

Kelley, J. L. *et al.* (2016) 'The life aquatic: Advances in marine vertebrate genomics', *Nature Reviews Genetics* . Nature Publishing Group, 17(9), pp. 523–534. doi: 10.1038/nrg.2016.66.

Laetsch, D. R. and Blaxter, M. L. (2017) 'BlobTools: Interrogation of genome assemblies', *F1000Research* , 6, p. 1287. doi: 10.12688/f1000research.12232.1.

de Leeuw, C. A. *et al.* (2020) 'First come, first served: Possible role for priority effects in marine populations under different degrees of dispersal potential', *Journal of Biogeography* . doi: 10.1111/jbi.13873.

Lefort, V., Desper, R. and Gascuel, O. (2015) 'FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program', *Molecular Biology and Evolution* , 32(10), pp. 2798–2800. doi: 10.1093/molbev/msv150.

Legendre, J. V. and Legendre, L. F. (2012) *Numerical Ecology* . Vol. 24. Elsevier, Amsterdam, Netherlands.

Leiva, C. *et al.* (2019) 'Population substructure and signals of divergent adaptive selection despite admixture in the sponge Dendrilla antarctica from shallow waters surrounding the Antarctic Peninsula',*Molecular Ecology* , (January), pp. 3151–3170. doi: 10.1111/mec.15135.

Lemopoulos, A. *et al.* (2019) 'Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation', *Ecology and Evolution* , 9(4), pp. 2106–2120. doi: 10.1002/ece3.4905.

Li, H. *et al.* (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics* , 25(16), pp. 2078–2079. doi: 10.1093/bioinformatics/btp352.

Li, W. and Godzik, A. (2006) 'Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences',*Bioinformatics* , 22(13), pp. 1658–1659. doi: 10.1093/bioinformatics/btl158.

Liggins, L., Treml, E. A. and Riginos, C. (2013) 'Taking the Plunge: An Introduction to Undertaking Seascape Genetic Studies and using Biophysical Models', *Geography Compass* , 7(3), pp. 173–196. doi: 10.1111/gec3.12031.

Liu, M. *et al.* (2012) 'Metaproteogenomic analysis of a community of sponge symbionts', *ISME Journal* . Nature Publishing Group, 6(8), pp. 1515–1525. doi: 10.1038/ismej.2012.1.

Lynch, M. (2008) 'Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects', *Molecular Biology and Evolution* , 25(11), pp. 2409–2419. doi: 10.1093/molbev/msn185.

Maas, D. L. *et al.* (2018) 'Rapid divergence of mussel populations despite incomplete barriers to dispersal', *Molecular Ecology* , 27(7), pp. 1556–1571. doi: 10.1111/mec.14556.

Maldonado, M. (2006) 'The ecology of the sponge larva', *Canadian Journal of Zoology* , 84(2), pp. 175–194. doi: 10.1139/z05-177.

Mantel, N. (1967) 'The detection of disease clustering and a generalized regression approach', *Cancer Research* , 27(2), pp. 209–220.

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Martin, M. (2011) 'Cutadapt removes adapter sequences from high-throughput sequencing reads', *EMBnet.journal* , 17(1), pp. 10–12. doi: 10.14806/ej.17.1.200.

Mayr, E. (1963) *Animal species and evolution* . Cambridge, Massachusetts: Harvard Univeristy Press.

McKenna, A. *et al.* (2010) 'The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data',*Genome Research* , 20(9), pp. 1297–1303.

De Meester, L. *et al.* (2016) 'Evolving Perspectives on Monopolization and Priority Effects', *Trends in Ecology and Evolution* . Elsevier Ltd, 31(2), pp. 136–146. doi: 10.1016/j.tree.2015.12.009.

Meirmans, P. G. and Hedrick, P. W. (2011) 'Assessing population structure: FST and related measures', *Molecular Ecology Resources* , 11(1), pp. 5–18. doi: 10.1111/j.1755-0998.2010.02927.x.

Meirmans, P. G. and Van Tienderen, P. H. (2004) 'GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms', *Molecular Ecology Notes* , 4(4), pp. 792–794. doi: 10.1111/j.1471-8286.2004.00770.x.

Miller, K. J. *et al.* (2018) 'Complex genetic structure revealed in the circum-Antarctic broadcast spawning sea urchin Sterechinus neumayeri', *Marine Ecology Progress Series* , 601, pp. 153–166. doi: 10.3354/meps12648.

Morrow, C. and Cardenas, P. (2015) 'Proposal for a revised classification of the Demospongiae (Porifera)', *Frontiers in Zoology* . ???, 12(1), pp. 1–27. doi: 10.1186/s12983-015-0099-8.

Morrow, K. M., Fiore, C. L. and Lesser, M. P. (2016) 'Environmental drivers of microbial community shifts in the giant barrel sponge, X estospongia muta, over a shallow to mesophotic depth gradient.',*Environmental Microbiology* , 18(6), pp. 2025–2038.

Nei, M. (1987) *Molecular evolutionary genetics* . Columbia University Press.

Noyer, C. and Becerro, M. A. (2012) 'Relationship between genetic, chemical, and bacterial diversity in the Atlanto-Mediterranean bath sponge Spongia lamella', *Hydrobiologia* , 687(1), pp. 85–99. doi: 10.1007/s10750-011-0884-0.

Oppen, M. J. H., Worheide, G. and Takabayashi, M. (2002) 'Nuclear markers in evolutionary and population genetic studies of scleractinian corals and sponges', *Proceedings 9th International Coral Reef Symposium* , 1. doi: 10.1371/journal.pone.0025628.

Orsini, L. *et al.* (2013) 'Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization', *Molecular Ecology* , 22(24), pp. 5983–5999. doi: 10.1111/mec.12561.

Perez-Portela, R., Noyer, C. and Becerro, M. A. (2015) 'Genetic structure and diversity of the endangered bath sponge Spongia lamella',*Aquatic Conservation: Marine and Freshwater Ecosystems* , 25(3), pp. 365–379. doi: 10.1002/aqc.2423.

Perez-Portela, R. and Riesgo, A. (2018) 'Population Genomics of Early-Splitting Lineages of Metazoans', in Rajora, P. (ed.)*Population genomics* . Cham, Switzerland: Springer, pp. 103–137. doi: 10.1007/13836_-2018_13.

Peterson, B. K. *et al.* (2012) 'Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species', *PLoS ONE* , 7(5). doi: 10.1371/journal.pone.0037135.

Pita, L. *et al.* (2018) 'The sponge holobiont in a changing ocean: from microbes to ecosystems', *Microbiome* . Microbiome, 6(1), p. 46. doi: 10.1186/s40168-018-0428-1.

Puritz, J. B. *et al.* (2014) 'Demystifying the RAD fad',*Molecular Ecology* , 23(24), pp. 5937–5942. doi: 10.1111/mec.12965.

Rambaut, A. (2009) 'FigTree.'

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Reshef, L. *et al.* (2006) 'The Coral Probiotic Hypothesis',*Environmental Microbiology* , 8(12), pp. 2068–2073. doi: 10.1111/j.1462-2920.2006.01148.x.

Reveillaud, J. *et al.* (2014) 'Host-specificity among abundant and rare taxa in the sponge microbiome', *ISME Journal* , 8(6), pp. 1198–1209. doi: 10.1038/ismej.2013.227.

Richards, V. P. *et al.* (2016) 'Patterns of population structure and dispersal in the long-lived "redwood" of the coral reef, the giant barrel sponge (Xestospongia muta)', *Coral Reefs* . Springer Berlin Heidelberg, 35(3), pp. 1097–1107. doi: 10.1007/s00338-016-1435-y.

Riesgo, A. *et al.* (2019) 'Genetic diversity, connectivity and gene flow along the distribution of the emblematic Atlanto-Mediterranean sponge Petrosia ficiformis (Haplosclerida, Demospongiae)', *BMC Evolutionary Biology* . BMC Evolutionary Biology, 19(1), pp. 1–18. doi: 10.1186/s12862-018-1343-6.

Selkoe, K. A. *et al.* (2014) 'Emergent patterns of population genetic structure for a coral reef community', *Molecular Ecology* , 23(12), pp. 3064–3079. doi: 10.1111/mec.12804.

Selkoe, K. A. *et al.* (2016) 'A decade of seascape genetics: Contributions to basic and applied marine connectivity', *Marine Ecology Progress Series* , 554, pp. 1–19. doi: 10.3354/meps11792.

Simister, R. L. *et al.* (2012) 'Sponge-Microbe Associations Survive High Nutrients and Temperatures', *PLoS ONE* , 7(12), pp. 21–23. doi: 10.1371/journal.pone.0052220.

Smit, A. F. A., Hubley, R. and Green, P. (2014) *RepeatMasker Open-4.0. 2013-2015.*

Stamatakis, A. (2014) 'RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies', *Bioinformatics* , 30(9), pp. 1312–1313. doi: 10.1093/bioinformatics/btu033.

Steinert, G. *et al.* (2016) 'In four shallow and mesophotic tropical reef sponges from Guam the microbial community largely depends on host identity', *PeerJ* , 2016(4), pp. 1–25. doi: 10.7717/peerj.1936.

Suchan, T. *et al.* (2016) 'Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens', *PLoS ONE* , 11(3), pp. 1–22. doi: 10.1371/journal.pone.0151651.

Sunde, J. *et al.* (2020) 'Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (Esox lucius) and a Synthesis of Previous Studies',*Frontiers in Genetics* , 11(March), pp. 1–17. doi: 10.3389/fgene.2020.00218.

Sundqvist, L. *et al.* (2016) 'Directional genetic differentiation and relative migration', *Ecology and Evolution* , 6(11), pp. 3461–3475. doi: 10.1002/ece3.2096.

Swierts, T. *et al.* (2013) 'Lock, Stock and Two Different Barrels: Comparing the Genetic Composition of Morphotypes of the Indo-Pacific Sponge Xestospongia testudinaria', *PLoS ONE* , 8(9), pp. 1–12. doi: 10.1371/journal.pone.0074396.

Swierts, T., Cleary, D. F. R. and de Voogd, N. J. (2018) 'Prokaryotic communities of Indo-Pacific giant barrel sponges are more strongly influenced by geography than host phylogeny', *FEMS microbiology ecology* . Oxford University Press, 94(12), pp. 1–12. doi: 10.1093/femsec/fiy194.

Taboada, S. *et al.* (2018) 'Implications of population connectivity studies for the design of marine protected areas in the deep sea: An example of a demosponge from the Clarion-Clipperton Zone',*Molecular Ecology* , 27(23), pp. 4657–4679. doi: 10.1111/mec.14888.

Taylor, M. W. *et al.* (2005) 'Biogeography of bacteria associated with the marine sponge Cymbastela concentrica', *Environmental Microbiology* , 7(3), pp. 419–433. doi: 10.1111/j.1462-2920.2004.00711.x.

Timm, L. E. (2020) 'A fair fight between molecular marker types in a seascape genetics setting', *Molecular ecology* , 29(12), pp. 2133–2136. doi: 10.1111/mec.15473.

Posted on Authorea 23 Aug 2020 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.159818467.76147162 — This a preprint and has not been peer reviewed. Data may be preliminary.

Tomascik, T. and Mah, A. J. (1994) 'The ecology of Halimeda Lagoon: An anchialine Lagoon of a raised atoll, Kakaban Island, Eas kalimantan, Indonesia', *Tropical Biodiversoty* , 2(3), pp. 385–399.

Uriz, M. J. and Turon, X. (2012) *Sponge Ecology in the Molecular Era* . 1st edn, *Advances in Marine Biology* . 1st edn. Elsevier Ltd. doi: 10.1016/B978-0-12-387787-1.00006-4.

Vieira, F. G. *et al.* (2016) 'Improving the estimation of genetic distances from Next-Generation Sequencing data', *Biological Journal of the Linnean Society* , 117(1), pp. 139–149. doi: 10.1111/bij.12511.

Warren, B. H. *et al.* (2015) 'Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson',*Ecology Letters* , 18(2), pp. 200–217. doi: 10.1111/ele.12398.

Waters, J. M., Fraser, C. I. and Hewitt, G. M. (2013) 'Founder takes all: Density-dependent processes structure biodiversity', *Trends in Ecology and Evolution* . Elsevier Ltd, 28(2), pp. 78–85. doi: 10.1016/j.tree.2012.08.024.

Webster, N. S., Cobb, R. E. and Negri, A. P. (2008) 'Temperature thresholds for bacterial symbiosis with a sponge', *ISME Journal* , 2(8), pp. 830–842. doi: 10.1038/ismej.2008.42.

Webster, N. S. and Thomas, T. (2016) 'The sponge hologenome',*mBio* , 7(2), pp. 1–14. doi: 10.1128/mBio.00135-16.

Weisz, J. B., Lindquist, N. and Martens, C. S. (2008) 'Do associated microbial abundances impact marine demosponge pumping rates and tissue densities?', *Oecologia* , 155(2), pp. 367–376. doi: 10.1007/s00442-007-0910-0.

Whalan, S. *et al.* (2008) 'Low genetic variability within and among populations of the brooding sponge rhopaloeides odorabile on the central great barrier reef', *Aquatic Biology* , 3(2), pp. 111–119. doi: 10.3354/ab00066.

Wood, D. E. and Salzberg, S. L. (2014) 'Kraken: Ultrafast metagenomic sequence classification using exact alignments', *Genome Biology* , 15(3). doi: 10.1186/gb-2014-15-3-r46.

Worheide, G., Sole-Cava, A. M. and Hooper, J. N. A. (2005) 'Biodiversity, molecular ecology and phylogeography of marine sponges: Patterns, implications and outlooks', *Integrative and Comparative Biology* , 45(2), pp. 377–385. doi: 10.1093/icb/45.2.377.

Wright, S. (1943) 'Isolation by Distance', *Genetics* , 28(2), pp. 114–138. doi: Article.