UC Berkeley
UC Berkeley Previously Published Works

Title

Sponge diversification in marine lakes: Implications for phylogeography and population genomic studies on sponges

Permalink

Journal

ISSN

Authors

Maas, Diede L
Prost, Stefan
de Leeuw, Christiaan A
et al.

Publication Date

DOI

RESEARCH ARTICLE

Ecology and Evolution
Open Access
WILEY

Sponge diversification in marine lakes: Implications for phylogeography and population genomic studies on sponges

Diede L. Maas[1] [ORCID] | Stefan Prost[2,3] | Christiaan A. de Leeuw[1] | Ke Bi[4,5] |
Lydia L. Smith[4] | Purwanto Purwanto[6] | Ludi P. Aji[1,7] | Ricardo F. Tapilatu[8,†] |
Rosemary G. Gillespie[9] | Leontine E. Becking[9,10]

[1]Marine Animal Ecology, Wageningen University & Research, Wageningen, The Netherlands

[2]LOEWE Centre for Translational Biodiversity Genomics, Senckenberg Natural History Museum, Frankfurt am Main, Germany

[3]South African National Biodiversity Institute, National Zoological Gardens of South Africa, Pretoria, South Africa

[4]Museum of Vertebrate Zoology, University of California Berkeley, Berkeley, California, USA

[5]Computational Genomics Resource Laboratory, California Institute for Quantitative Biosciences, University of California Berkeley, Berkeley, California, USA

[6]Coral Triangle Center, Bali, Indonesia

[7]Research Centre for Oceanography, Indonesian Institute of Sciences, Lembaga Ilmu Pengetahuan Indonesia, Jakarta, Indonesia

[8]Marine Science and Fisheries Departments and Research Center of Pacific Marine Resources, State University of Papua, Manokwari, Indonesia

[9]Department of Environmental Science, Policy and Management, University of California Berkeley, Berkeley, California, USA

[10]Aquaculture and Fisheries, Naturalis Biodiversity Center, Wageningen University & Research, Wageningen, The Netherlands

Correspondence
Diede L. Maas, Marine Animal Ecology, Wageningen University & Research, Droevendaalsesteeg 1, Wageningen 6708PB, The Netherlands.
Email: diede.maas@wur.nl

Leontine E. Becking, Department of Environmental Science, Policy and Management, University of California Berkeley, Berkeley, California, USA.
Email: lisa.becking@wur.nl

Abstract

The relative influence of geography, currents, and environment on gene flow within sessile marine species remains an open question. Detecting subtle genetic differentiation at small scales is challenging in benthic populations due to large effective population sizes, general lack of resolution in genetic markers, and because barriers to dispersal often remain elusive. Marine lakes can circumvent confounding factors by providing discrete and replicated ecosystems. Using high-resolution double digest restriction-site-associated DNA sequencing (4826 Single Nucleotide Polymorphisms, SNPs), we genotyped populations of the sponge *Suberites diversicolor* ($n = 125$) to test the relative importance of spatial scales (1–1400 km), local environmental conditions, and permeability of seascape barriers in shaping population genomic structure. With the SNP dataset, we show strong intralineage population structure, even at scales <10 km (average $F_{ST} = 0.63$), which was not detected previously using single markers. Most variation was explained by differentiation between populations (AMOVA: 48.8%) with signatures of population size declines and bottlenecks per lake. Although the populations were strongly structured, we did not detect significant effects of geographic distance, local environments, or degree of connection to the sea on population structure, suggesting mechanisms such as founder events with subsequent

†Deceased

priority effects may be at play. We show that the inclusion of morphologically cryptic lineages that can be detected with the COI marker can reduce the obtained SNP set by around 90%. Future work on sponge genomics should confirm that only one lineage is included. Our results call for a reassessment of poorly dispersing benthic organisms that were previously assumed to be highly connected based on low-resolution markers.

1 | INTRODUCTION

Isolating mechanisms causing population genomic structuring still remain elusive for many marine organisms (Liggins et al., 2013; Peijnenburg & Goetze, 2013; Selkoe et al., 2016). Examples of such mechanisms are dispersal limitation due to geographic distance resulting in isolation-by-distance patterns (Chaves-Fonnegra et al., 2015; Pérez-Portela et al., 2015; Wright, 1943), and dispersal/establishment limitation due to ecologically heterogeneous habitats resulting in a pattern of isolation-by-environment (Giles et al., 2015; Nosil et al., 2009; Orsini et al., 2013; Rundle & Nosil, 2005). In some cases, patterns of genomic structure in marine organisms may not be clearly linked to geographic or environmental influences (Cornwell et al., 2016; Miller et al., 2018; Taboada et al., 2018). Here, other explanations may include other barriers to dispersal, such as ocean currents, resulting in an isolation-by-resistance pattern (McRae, 2006), or processes involving historical contingency (Fukami, 2015). Recent work on benthic invertebrates has shown strong population genomic structure, providing evidence against the previously widely held belief that marine organisms are panmictic due to few barriers to dispersal (e.g., Bierne et al., 2016; Marshall et al., 2010; Van Wyngaarden et al., 2017). Understanding population structure at smaller scales is relevant as marine sessile invertebrates can show ecological dynamics and incipient speciation at these scales (<500 km, Bernatchez et al., 2019; Van Wyngaarden et al., 2017; Xuereb et al., 2018). It is also the scale at which human activities can cause direct change, for example, through habitat change and degradation, and through restoration activities and marine spatial planning.

In marine invertebrates, which encompass the largest diversity in the marine animal kingdom (Chen, 2021), there are still outstanding questions on the processes that lead to the small-scale structuring of population genomic diversity. A key issue is that for many non-model marine invertebrate species, there is still a dearth of high-resolution genomic approaches that allow detection of small-scale population genomic structure, diversity, and demographic histories (Oleksiak & Rajora, 2020). Sponges are a prime example in this regard. Sponges are integral yet often underappreciated assets of

benthic communities (Bell, 2008; De Goeij et al., 2013; Webster & Thomas, 2016). Given that sponges are generally considered to be poor dispersers as their larvae have limited swimming capacity and are short-lived (Maldonado, 2006), patterns of strong genetic divergence over relatively small geographic ranges would be expected. However, studies so far typically find connectivity at scales of 100s–1000s of kilometers (e.g., De Bakker et al., 2016; Taboada et al., 2018). Studies that have sought to understand the processes shaping genetic structure in sponges have revealed species complexes with divergence among morphologically cryptic lineages (Pérez-Portela & Riesgo, 2020; Uriz & Turon, 2012; van Oppen et al., 2002), yet seem to show little genetic diversity and signatures of panmixia within species encompassing large geographic areas in shallow (de Bakker et al., 2016; Whalan et al., 2008) and in deep sea areas (Ekins et al., 2016; Taboada et al., 2018). The findings of high connectivity within species may indicate that the dispersal ability of sponges is inherently greater than that of other marine organisms, with potential explanations including rafting, asexual budding, and sperm-mediated gene flow (DeBiasse et al., 2014; Maldonado & Uriz, 1999; Wörheide et al., 2008). However, this runs counter to sponges generally being considered to be poor dispersers with short-lived larval stages with limited swimming capacity (Maldonado, 2006).

The lack of structure that has been found in sponge populations may be the result of the dearth of high-resolution genomic studies (Oleksiak & Rajora, 2020; Pérez-Portela & Riesgo, 2020). The majority of studies on sponge phylogeography and population structure have deployed mitochondrial markers (mtDNA) such as Cytochrome *c* oxidase I (*COI*) and ATP6, and nuclear markers such as introns, internal transcribed spacers (*ITS*) (Wörheide et al., 2008) and microsatellites (Pérez-Portela & Riesgo, 2020; Uriz & Turon, 2012; Van Oppen et al., 2002). Although widely used in phylogeographic and population genetic studies (Avise, 2000, 2009), mitochondrial markers exhibit low mutation rates in sponges (Huang et al., 2008; Wörheide et al., 2005), resulting in lower diversity. In contrast, studies using *ITS* markers have shown more structure (Becking et al., 2013; Bentlage & Wörheide, 2007;

Ekins et al., 2016), but generally at large spatial scales (~1000 km). Furthermore, *ITS* markers can be hampered by intragenomic polymorphisms (Frankham et al., 2002; Wörheide et al., 2004), clouding patterns of true population structure. Finally, microsatellites could be reliable and sufficiently variable to detect population structure (Taboada et al., 2018), yet are time-consuming to design de novo for each species (Frankham et al., 2002; Pérez-Portela & Riesgo, 2020), can be confounded by homogenizing forces of evolution (Van Oppen et al., 2002), and generally relatively few are used per study (<20). An increase in the number of molecular markers is expected to advance inferences on structure and demography (Allendorf et al., 2010; Kelley et al., 2016; Pérez-Portela & Riesgo, 2020), allowing researchers to reassess assumptions of panmixia within sponge lineages.

Recently, there has been an increase in the use of reduced representation genomic methods and Single Nucleotide Polymorphisms (SNPs) for population genomic studies on nonmodel organisms (e.g., Baird et al., 2008; Catchen et al., 2017; Peterson et al., 2012; Puritz et al., 2014). The additional power of an increased marker panel has been demonstrated in for example mussels (Becking et al., 2016; de Leeuw et al., 2020; Maas et al., 2018), fish (Bradbury et al., 2015; D'Aloia et al., 2020; Lemopoulos et al., 2019; Sunde et al., 2020), and sponges (Taboada et al., 2022). However, high-resolution studies on sponges are lagging behind (Pérez-Portela & Riesgo, 2020), with the notable exception by Brown et al. (2014), Brown et al. (2017), Leiva et al. (2019), and Taboada et al. (2022). Using restriction site-associated DNA sequencing (RADseq), techniques such as ddRAD (double digest RADseq, Peterson et al., 2012), such as in Leiva et al. (2019) and Taboada et al. (2022), may increase the number of retained SNPs to thousands and provide the necessary resolution. However, most of these studies still used a limited marker panel and showed no differentiation at scales below 100 km. In this study, we used data with >4000 SNPs to investigate population structuring and demography in a shallow water sessile marine invertebrate in relation to geographic distance, permeability of barriers, and environmental variables using marine lake ecosystems.

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 bodies of seawater surrounded completely by land that maintain connection with the surrounding sea through caves or porous rock (Becking et al., 2011; Dawson et al., 2009; Hamner et al., 1982; Holthuis, 1973). The extent of the connection of marine lakes to the sea ranges from being highly connected where seawater moves in and out of the lake through caves, to highly isolated where seawater has to travel through porous rock. Marine lakes are common in Vietnam, Palau, and Indonesia, particularly in East Kalimantan and in West Papua (Becking et al., 2011, 2015; Dawson et al., 2009). Having originated roughly at the same time (after the Last Glacial Maximum approximately 8000–10,000 years ago; Sathiamurthy & Voris, 2006; Tomascik & Mah, 1994), marine lakes represent relatively controlled biotopes where each lake can be seen as an independent replicate of eco-evolutionary dynamics.

Although marine lakes have barriers to connectivity and may represent local environments distinct from open marine systems, they could still be representative of heterogenous coasts that are prevalent in the Indo-Pacific. Furthermore, studying marine lakes can help in better understanding drivers of genomic structure such as founder effects, and bottlenecks and expansion events in marine populations.

Sponges are usually well-represented in marine lakes, both in diversity and in biomass (Azzini et al., 2007; Becking et al., 2011, 2013; Cleary et al., 2013). The sponge *Suberites diversicolor* (Porifera, Demospongiae, Suberitidae, Becking & Lim, 2009) has been found to occur in Indonesian marine lakes and brackish coastal areas (Becking & Lim, 2009; Cleary et al., 2013). Using *COI* and *ITS* genetic markers, Becking et al. (2013) studied its phylogeography from multiple marine lakes and lagoon populations in the Indo-Pacific. They identified two distinct genetic lineages (Lineage A and B) and regional structuring, yet did not observe structure 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*). Given the high genomic structuring observed in co-distributed species from marine lakes (de Leeuw et al., 2020; Gotoh et al., 2011; Maas et al., 2018, 2020), we expect that the markers used did not provide sufficient resolution to detect signals.

Selecting nine marine lakes and two lagoon locations in the Indo-Pacific at different spatial scales (1–1400 km), with different environmental conditions, and along a gradient of connection to the surrounding sea, we assessed the relative influence of these drivers on genomic structure of *S. diversicolor*. Furthermore, we (a) optimized laboratory and bioinformatic filtering methods for low-coverage RAD-generated data using old stock DNA extractions to assess effects of filtering, (b) compared phylogeographic and population genetic structure of *COI* and *ITS* markers from Becking et al. (2013) to RAD-generated markers, and (c) assessed the influence of phylogenetic level on the number of markers retained. We expect populations to be highly structured at small spatial scales, reflecting the life history of benthic sponges with short larval durations, with this structure being detectable via the higher resolution gained via using a RAD approach. Next, if the structure is linked to dispersal potential due to geographic distance, we expect to find isolation-by-distance patterns, where gene flow decreases with increasing distance. In contrast, if the environment strongly influences population structure, we expect to find patterns of isolation-by-environment, where marine lakes similar in local environmental conditions should resemble each other regardless of the extent of geographic distance. If the extent of connection to the surrounding sea influences dispersal potential, we expect to observe an isolation-by-resistance pattern. Here, populations in highly connected lakes would show signals of genomic connectivity among each other and the lagoon populations, whereas isolated marine lakes would be particularly distinct, reflecting low dispersal from the sea. Finally, populations in isolated marine lakes should show evidence for strong genetic bottlenecks, reflecting the ontogeny of the lakes and subsequent low immigration from

the sea populations, while highly connected lakes should not, due to high water exchange.

2 | MATERIALS AND METHODS

2.1 | Sample collection and lake profiling

Marine lakes are not common and occur predominantly in Indonesia, Vietnam, and Palau (Dawson et al., 2009). The sponge *Suberites diversicolor* is found in many marine lakes, but some of the sampled lakes had very low densities; therefore, sample sizes were smaller (see Becking et al., 2013 table 2 for densities). As *S. diversicolor* is not frequently found outside of marine lakes, we took an opportunistic sampling strategy. Tissue samples (~1 cm^3) were collected from 168 individuals of *Suberites diversicolor* (Figure 1 and Table 1). One lagoon was sampled in Darwin, Australia (Sea Australia), one lagoon in East-Kalimantan (Sea Indonesia). In East-Kalimantan, we

additionally sampled three marine lakes (Kalimantan1, Kalimantan2, and Kalimantan3), and six marine lakes in West-Papua (Papua27, Papua30, Papua32, Papua1, Papua4, and Papua5). As many locations have no official names, we used a coding system consistent with de Leeuw et al. (2020) and Maas et al. (2018, 2020). Of these locations, nine overlap with the sponge phylogeography study of Becking et al. (2013) (Table S1 for corresponding lake codes), two additional marine lakes were sampled for this study (Papua5 and Papua27). Samples were collected between 1 and 5 m depth while snorkeling, collecting individuals at least 25 m apart. In the field, tissue samples were cleaned of any debris or metazoan symbionts visible to the naked eye, and 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 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 a YSI Professional Plus multimeter at 10
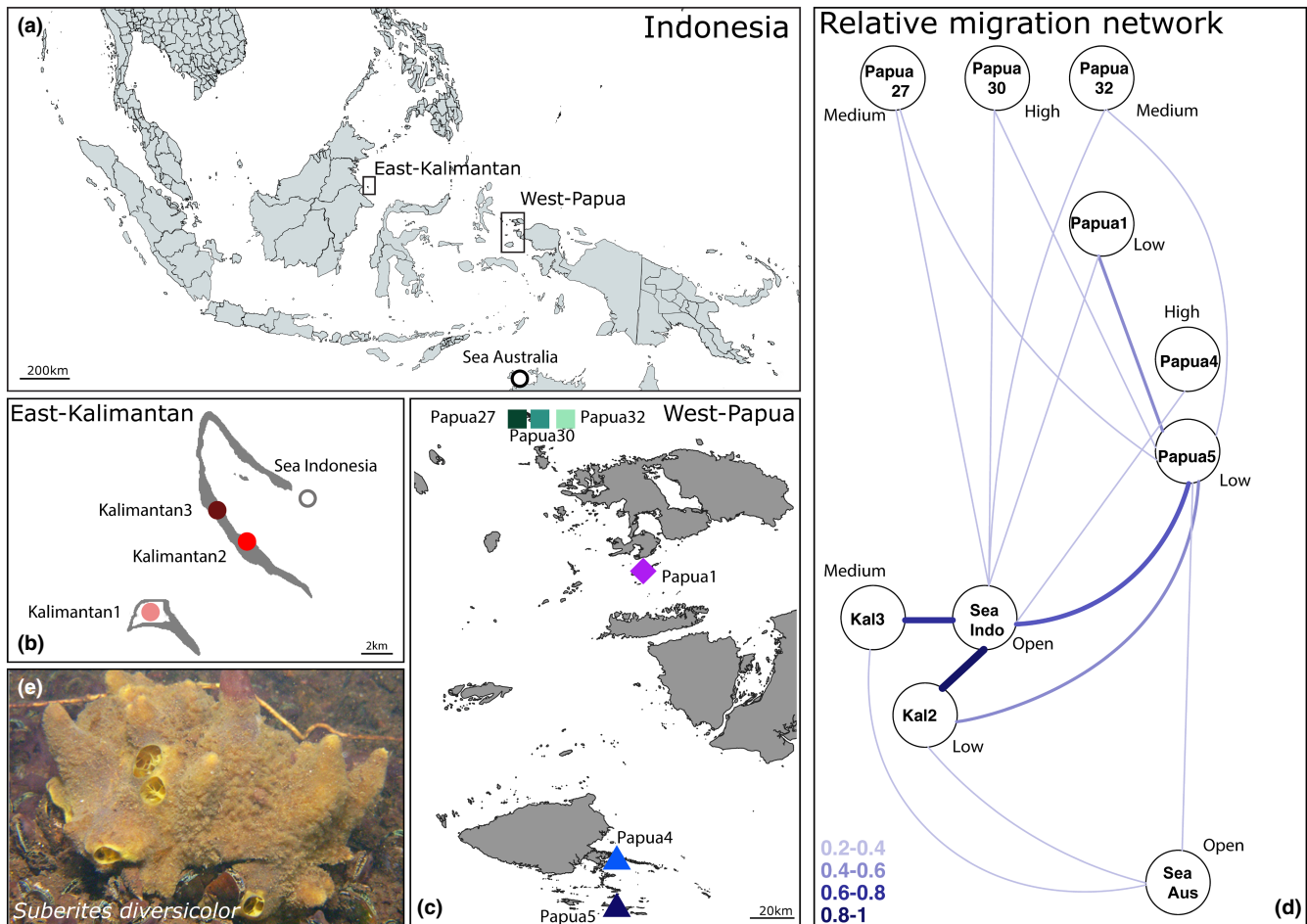

FIGURE 1 Sampling sites of *Suberites diversicolor* from nine marine lakes and two lagoon locations and associated relative migration networks. (a) Overview of Indonesia including two geographic regions sampled: Berau, East-Kalimantan and Raja Ampat, West-Papua. It also shows the location of the Australian lagoon sampled (Sea Australia). (b) Zoom of Berau, with locations of three marine lakes (Kalimantan1, Kalimantan2, and Kalimantan3) and one lagoon (Sea Indonesia). (c) Zoom of Raja Ampat, with locations of six marine lakes (Papua27, Papua30, Papua32, Papua1, Papua4, and Papua5). (d) Relative migration network including only samples from Lineage B based on 4826 SNPs and run with 1000 bootstraps. Fractions of relative migration are displayed and categories of level of connection to the surrounding sea are indicated. (e) Specimen of *S. diversicolor*, photograph by L.E. Becking.

TABLE 1 Overview of sampling in marine lakes and lagoon locations.

Code	Location	Connection category	Fraction tidal amplitude	Area (×1000m²)	Max. depth (m)	Average temperature	Average salinity	Number specimens passed QC	Nucleotide diversity	Heterozygosity (He)
Sea Australia	Australia	Open						7	0.0095	0.117
Indonesia	Region: Island									
Sea Indonesia	Berau	Open				29.0	33.5	5	0.0101	0.157
Kalimantan1	Berau: Kakaban	Low	0.1	4900	12	30.0	23.5	20		
Kalimantan2	Berau: Tanah Banban	Low	0.4	232	17	29.5	26.0	2	0.0074	0.034
Kalimantan3	Berau: Maratua	Medium	0.5	140	2	29.5	27.0	26	0.0050	0.081
Papua27	Papua: Wayag	Medium		22	4	29.5	31.0	5	0.0037	0.038
Papua30	Papua: Wayag	High	0.8	13	6	32.4	28.9	8	0.0045	0.052
Papua32	Papua: Wayag	Medium	0.5	6		31.2	30.7	4	0.0053	0.059
Papua1	Papua: Gam	Low	0.1	89	19	32.3	24.0	11	0.0036	0.054
Papua4	Papua: Misool	High	0.8	14	20	31.7	25.9	19	0.0047	0.080
Papua5	Papua: Misool	Low	0.3	4	5	31.5	28.9	18	0.0060	0.095

Note: Recorded are site codes, location, physiographic, environmental, and genomic parameters, and number of individuals retained after filtering. Explanations of how physiographic parameters (lake area, depth and connection) as well as how the local water quality was measured (temperature and salinity) can be found in the methods section. In the calculation of population genomic parameters (nucleotide diversity and heterozygosity) only localities containing samples from Lineage B were considered.

locations per lake at 1 m intervals from the surface to 5 m depth. To define the connection to the surrounding sea, we measured maximum tidal amplitude simultaneously in the lake and the sea using Hobo water-level loggers (Onset HOBO U20L). The degree of water exchange between the marine lakes and the adjacent sea was assessed by placing a water-level logger inside the marine lake and one directly outside in the surrounding sea during a 48 h period. Pressure (Pa) was converted to depth (m) using the Hoboware Pro 3.7.16 software. The fraction of tidal amplitude of the lake compared with the sea was then calculated by $(Lake_{max})/(Sea_{max})$, where max stands for the maximum water level in either the lake or the sea. A maximum value of 1 would indicate limited (or no) obstruction to water flow in and out of the lake, and the minimum value of 0 would mean there is no water exchange at all. We categorized the level of connection to the surrounding sea as low (≤0.4), medium (0.5–0.7), and high (0.8–1). The sea locations were categorized as open.

2.2 | 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, double digest restriction site-associated DNA (ddRAD) libraries were prepared following the protocol of Peterson et al. (2012). The adapted protocol used by the current study can be found in the Appendix S2. In brief, genomic DNA (600 ng) was double-digested using enzymes SphI-HF (rare-cutting) and MlucI (frequent-cutting) (See Appendix 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 publicly available from Peterson et al. (2012) "Locus count from Bioanalyzer % in region" to calculate the number of fragments to be expected assuming a common genome size for sponges of ~300 Mb (Jeffery et al., 2013; Srivastava et al., 2010), 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 eight libraries in total. The Sage Science Pippin Prep was used to size-select adapter-ligated fragments of length 500–575 bp (indicating an insert size of 425–500 bp). A trial was run for 8, 10, and 12 polymerase chain reaction (PCR) cycles. In the end, 10 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. The quality and quantity of libraries throughout the process were checked using BioAnalyzer High Sensitivity chips (Agilent, Appendix S1 for an example). Libraries were pooled at equimolar volumes and 150 bp single-end sequenced on Illumina HiSeq 2500 at the Vincent J. Coates Genomic Sequencing Facility at UC Berkeley.

2.3 | Reference assembly, bioinformatic filtering, and genotype calling

Custom perl scripts were used for processing the resulting sequences (RADTOOLKIT v. 0.13.10, made available Supplemental Information). 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 contaminations and low-quality reads using cutadapt v1.15 (Martin, 2011) and Trimmomatic (Bolger et al., 2014). Cleaned reads were clustered with CD-HIT v4.6.1 (Fu et al., 2012; Li & Godzik, 2006), with a minimum support per cluster set at three reads, and representative sequences retained for each cluster. RepeatMasker v4.0 (http://repeatmask er.org/) was used to mask putative repetitive elements, low complexity regions, and short repeats using "Suberitidae" as a database (Smit et al., 2013). Loci were discarded if >60% of nucleotides per loci were Ns. The resulting RAD loci were combined for all individuals, and a de novo reference was built from loci shared by at least 70% of individuals.

We screened for loci from putative microbes in 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 v1 (Wood & Salzberg, 2014), a fast sequence classifier to BLAST (Altschul et al., 1990) our loci against bacterial databases with default settings. Third, we used BlobTools (Laetsch & 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 (SNPcleaner, github.com/tplin deroth/ngsQC/tree/master/snpCleaner; Bi et al., 2013, 2019).

Cleaned sequence reads for each individual were aligned to the de novo generated reference separately using Novoalign v4.0 (http://www.novocraft.com), and only uniquely mapping reads were retained. Picard (www.picard.sourceforge.net) was used to add read groups, SAMtools v1.9 (Li et al., 2009) to generate a BAM file per individual, and GATK (McKenna et al., 2010) to perform realignment. SAMtools and BCFtools v1.2 were used to generate a VCF file. Single Nucleotide Polymorphisms (SNPs) and invariant sites were masked around 10 bp of an indel. Sites were removed if the depth was outside 1st and 99th percentile of the overall coverage. The custom perl script SNPcleaner was used for further filtering of SNPs. Ultimately, one random SNP per RADtag was retained for downstream analyses.

Calling SNPs and genotypes based on allele counts may be highly uncertain if coverage is low (Johnson & 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 v.0.930) (Korneliussen et al., 2014). We set genotype posterior probabilities of 0.95 as a threshold in ANGSD to output high-confidence genotypes for analyses performed in GENODIVE v3.0 requiring genotype calls (Meirmans & 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).

2.4 | Detection of major lineages

We reconstructed phylogeographic relationships among and within lineages via a maximum likelihood tree. The maximum likelihood approach was performed via genotype calling and the software IQ-Tree (Nguyen et al., 2015). First, we created consensus sequences in fasta format for all 125 individuals using ANGSD (Korneliussen et al., 2014), applying the options -doFasta 3 and -doCounts 1. Next, we concatenated the consensus sequences for all loci for each individual, resulting in a consensus sequence of 55 kb per individual, and carried out an alignment using MAFFT (Katoh & Standley, 2013). We then constructed the maximum likelihood phylogenetic tree using the IQ-Tree software with 1000 ultrafast bootstraps and an SH-like approximate likelihood test for 1000 replicates. The best fitting substitution model was inferred using the -m TEST function in IQ-Tree.

Next, we explored admixture patterns using ngsAdmix (Skotte et al, 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 the real *K* after which the likelihood plateaus or increases only slightly (Evanno et al., 2005). Differentiation among lineages was also assessed via an Analysis of Molecular Variance (AMOVA) with 1000 permutations using genotype calls (Excoffier et al., 1992). We used major lineages, two sublineages determined from the maximum likelihood tree, and populations as the nested levels. For the AMOVA, we used genotype calls allowing only 10% missing data to improve accuracy.

2.5 | Population genomic structure and migration network of lineage B

We assessed population structure and differentiation within Lineage B using four approaches. We ran a Principal Component Analysis (PCA) based on a covariance matrix computed by ngsTools on genotype likelihoods (Fumagalli et al., 2014) and via GENODIVE using genotype calls (Meirmans & Van Tienderen, 2004). As an unsupervised clustering method, PCA estimates population genomic structure without bias. Next, we performed a neighbor-joining network (NeighborNet) analysis using Splitstree using genotype likelihoods (Huson, 1998; Huson & 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. Then, between-population differentiation was assessed via 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 & Hedrick, 2011). Finally, a migration network was constructed using Nei's G_{ST} with the threshold at 0.2 and 1000 bootstraps using the DiveRsity package in R (Keenan et al., 2013), as demonstrated by Sundqvist et al. (2016).

2.6 | Genomic diversity and inference of demographic histories

We estimated the within-population genomic diversity using two diversity measures. We calculated expected heterozygosity (H_e) using GENODIVE, and overall heterozygosity and nucleotide diversity (π) using ANGSD (Nei, 1987). Population demographic histories were inferred via Stairway Plots using Site Frequency Spectra (SFS) (Liu & Fu, 2015). Stairway Plots offer an opportunity to infer demographic changes without requiring predefined models to test. Mutation rate is not known for sponges, but was calculated at 1.1×10^{-8} per generation via the regression coefficient by Lynch (2010) using the estimated genome size of 300 Mb. Generation time was set at 1 year. Stairway analyses were run for a subset of locations: two lagoon populations (Sea Australia and Sea Indonesia), two highly connected lakes (Papua4 and Papua30), and two isolated lakes (Papua1 and Papua5).

2.7 | Associations to geographic distance, environmental variables, and degree of connection

Finally, we explored spatial, environmental, and oceanographic associations to genomic structure. Mantel tests (Legendre & Legendre, 2012; Mantel, 1967) were used to test significance of correlations between genomic, geographic, environmental, and connection distance matrices. For genomic distances, we used normalized pairwise genomic differentiation ($F'_{ST}/(1-F'_{ST})$). 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 temperature (°C) and salinity (ppt) per lake, we performed a Principal Component Analysis (PCA) to calculate informative scores. The scores from all retained PCA axes were then used computing the environmental distance matrix using the function *dist* in R. Here, low distance values indicate lakes with a similar environment in terms of temperature and salinity, whereas high distance values indicate lakes with very dissimilar environments. Connection distance was calculated following the equations of Maas et al. (2018), so that a high distance value indicated a pairwise comparison of isolated lakes and a low distance value indicated a pairwise comparison of connected lakes, with all other comparisons falling in between, so that the resulting matrix would reflect resistance values. Mantel tests were run with 10,000 permutations using *vegan* in R. We verified the absence of autocorrelation between geographic, environmental and connection distances using Mantel's tests. Finally, we performed Spearman correlation tests among genomic diversity indices per lake with temperature, salinity, degree of connection and lake area. Correlations of $r \geq .5$ were considered strong, and alpha was set to 0.01.

3 | RESULTS

3.1 | Lake characterization

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 ± 1.2°C) and lower salinities (27.3 ppt ± 2.7 ppt) in lakes when compared to the open lagoon locations (29°C and 33.5 ppt). Connection to the surrounding sea varied among lakes, with highly connected to highly isolated lakes based on tidal amplitudes. For instance, lake Papua4 was found to have the highest connection with tidal amplitude representing 80% of that of the surrounding sea, indicating high water exchange with the sea, while lake Papua1 was most isolated, with the tidal amplitude only being <10% of the surrounding sea, and indicating limited exchange with the sea. To illustrate, the highly connected lakes had large enough channels or caves for fish to swim through, and we could see water flooding in during high tides. We categorized two lakes with high connection to the sea, three lakes with medium connection to the sea, and four isolated lakes with low connection to the adjacent sea (Table 1).

3.2 | Read statistics and filtering

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), an estimated genome size of 300 Mb, and a size selection of 425–500 bp, we expected to retain 13,652 RADtags. This was close to the actual retained loci, as the de novo reference retained 14,442 tags when keeping RADtags with at least 3× coverage and present in at least 70% of individuals. Kraken and Blobtools identified 15 out of the 14,442 RADtags containing possible bacterial contamination, which mapped to *Synechococcus* sp., a Cyanobacteria genus. These were removed from the data set (tags available in Supplemental Information).

After filtering, we retained 125 sponges with 973,697,804 reads in total, with coverage ranging from 3.1–82.2× (average 24.0×). The number of genomic markers retained strongly varied depending on the inclusion or exclusion of the two major lineages. When including both lineages 541 SNPs were retained, yet when only including Lineage B the number of SNPs increased almost ninefold to 4826. For Lineage B, in total 23,742 SNPs were called over all tags, but after selecting one SNP per tag we retained the 4826 SNPs for subsequent analyses. Also for Lineage B, depending on the filtering options (genotype calls or genotype likelihoods, coverage 3× or 10×, included missing data 30%, 10%, 5% or 1%) the number of SNPs varied from 56 to 4826 (Table S2).

3.3 | Detection of major lineages and sub-lineages

A maximum likelihood phylogenetic tree based on 541 SNPs showed two divergent lineages for 125 individuals (Figure 2a). These lineages are concordant with Lineage A and B as defined by Becking et al. (2013). Lineage A was only represented by individuals of Kalimantan1. The remaining populations in marine lakes fell under Lineage B. Within Lineage B, two sublineages could be seen representing on the one hand locations from East-Kalimantan, Australia and Papua4, and on the other the rest of the locations from West-Papua. Lagoon population Sea Indonesia was a sister group to both the East-Kalimantan and West-Papua sub-lineages.

Admixture analyses showed Kalimantan1 to be consistently distinct from all other populations throughout the range of ancestral populations K (Figure 2b). The most likely number of ancestral populations was found to be at $K=9$ (Figure S1), after which new groups did not add additional information. At $K=9$, all marine lakes are uniquely colored, except Papua27 and Papua32, which are grouped together (dark green), and Kalimantan2, which showed admixture from multiple groups. Furthermore, Sea Indonesia and Sea Australia also showed admixture from multiple groups.

When analyzing the complete dataset, the AMOVA showed most variation to be explained by the two major lineages (A and B) (Figure 2C, 62%). Within Lineage B, most variation was explained by differences among populations (48.8%). Variation among the two sublineages (East-Kalimantan, Australia and Papua4 on the one hand and the other populations of West-Papua on the other) was also found to significantly contribute to the variation, but only 7%. All subsequent analyses were run for Lineage B for 105 sponge individuals with 56–4826 SNPs, depending on the filtering option.

3.4 | Population genomic structure and migration network of lineage B

We tested different filters to assess the influence on the genomic patterns. Within Lineage B, patterns remained highly similar for all filters (see Figures S2–S4 and Tables S3–S6). Therefore, all further reported analyses were performed filtering on 3× coverage and max. 30% missing data, as this retained the most SNPs and thus would result in the highest resolution.

The sponge individuals clustered by lake and lagoon location in the Principal Component Analysis (PCA) (Figure 3a, Figures S2 and S3). The first four Principal Components (PCs) explained 80.5% of total variation (Figure 3a). PC1 explained 45.6% of the variation, separating populations by geographic region, with the lakes from West-Papua

being distinct from the lakes in East-Kalimantan. PC2 explained 24.4% of variation, separating lake Papua4 from the other lakes. The lagoon populations (Sea Australia and Sea Indonesia) clustered towards the center of the graph. PC3 and PC4 (explaining 10.5% in total) further separated Sea Australia and lakes Papua5, and to a lesser extent Papua1 and Papua30. Lakes Papua27 and Papua32 remained closely associated in both plots. High and medium connected lakes did not cluster together, nor with the open sea populations.

Findings from the Splitstree network were consistent with patterns found for the PCA (Figure 3b, Figure S4). The network showed a high fit (fit=99.2) and a small degree of reticulation ($d=0.153$), indicating a tree-like structure. The sea populations Sea Australia and Sea Indonesia showed higher reticulation than the marine lake populations, indicating potential introgression or hybridization events. All marine lake populations were diverged and showed no reticulation.
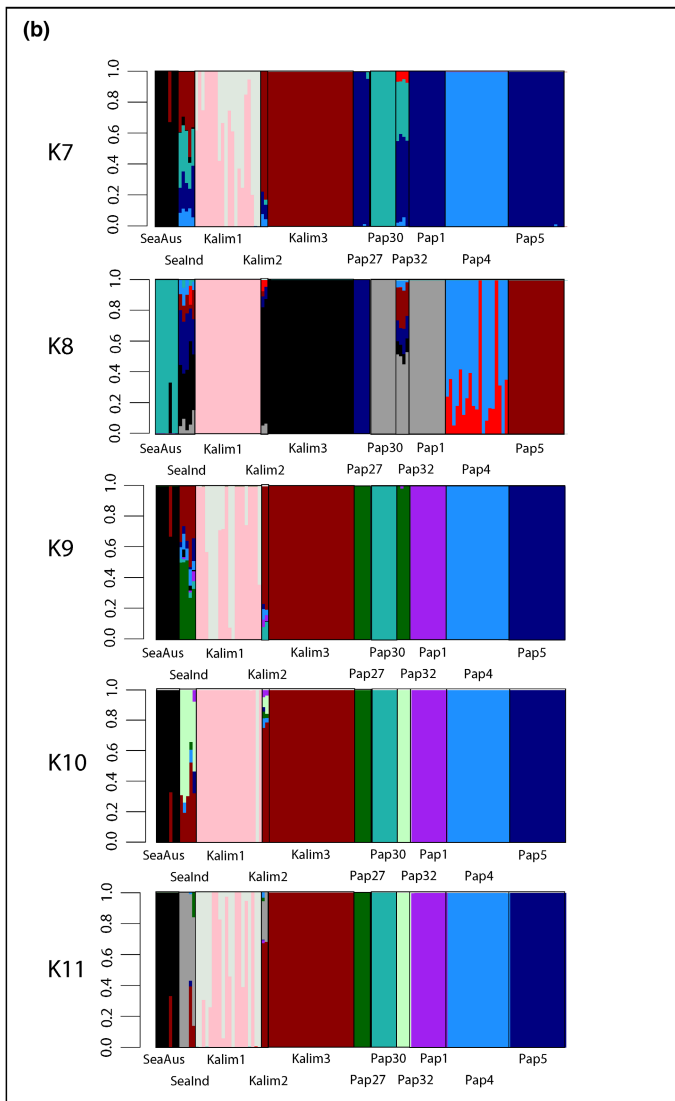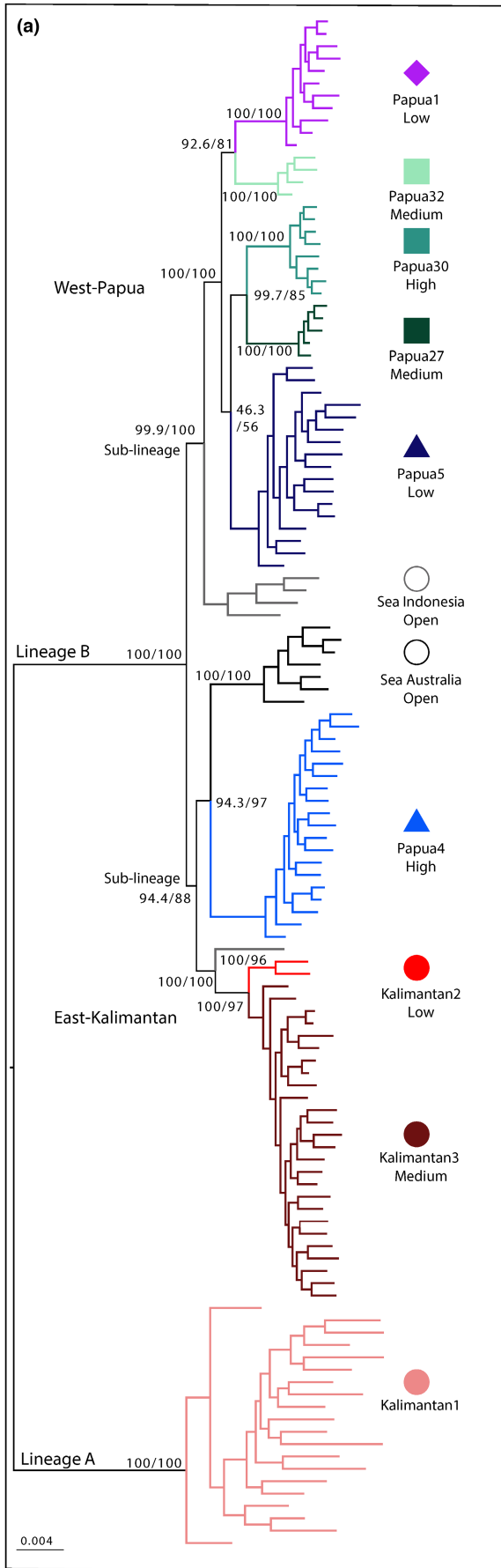
Pairwise fixation indices (F'_{ST}) showed high levels of genomic structuring (0.63 ± 0.13) (Figure 3C). When including Lineage A, the comparisons of populations to Kalimantan1 showed F'_{ST} values of >0.8 (Figure S5). Within Lineage B, the F'_{ST} values ranged from 0.18 between Sea Indonesia and Kalimantan2, to 0.78 between Papua30 and Papua32 (Table S3). All pairwise comparisons were significant, except for the comparison between Papua32 and Kalimantan2, likely due to sample size ($n=4$ and 2, respectively). Striking was that the F'_{ST} values among populations from highly connected lakes were equally strong as among highly and low connected lakes.

The migration network among lakes indicated the strongest relative bidirectional migration between marine lakes and the lagoon population in East-Kalimantan (Figure 1D). Lagoon population Sea Indonesia was linked to some degree to all other populations (relative bidirectional migration 0.2–1). Within West-Papua, bidirectional migration was observed between Papua5 and four other lakes (Papua27, Papua30, Papua32, and Papua1). Lagoon population Sea Australia showed links to Kalimantan2 and 3, and Papua5. There was no apparent strong bidirectional migration among the high and medium connected lakes, nor among those and the sea.

3.5 | Genomic diversity and demographic inferences per population

Population genomic diversity varied among lakes (Table 1, Table S4). The highest genomic diversity was consistently found for the lagoon populations Sea Australia and Sea Indonesia, 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). Low genomic diversity was consistently observed in populations Papua1 ($\pi=0.0036$,

FIGURE 2 Distinction of major genetic lineages A and B of sponge populations. (a) Maximum Likelihood IQTree using consensus sequences of 125 individuals based on 541 SNPs. Bootstrap values for major branches are displayed. (b) Bayesian admixture analysis for putative ancestral populations (K 7–11) based on genotype likelihoods via ngsAdmix. Each bar represents one individual. $K=9$ was indicated as the most likely ancestral populations. (c) Analysis of Molecular Variance (AMOVA) for Lineage A and B (top), and only Lineage B (bottom). Amount of variance explained in percentage, F-values and significance values are displayed. Colors and codes correspond to Figure 1 and Table 1.

(a)

West-Papua

100/100

92.6/81

100/100

100/100

99.7/85

100/100

100/100

46.3/56

99.9/100
Sub-lineage

Lineage B   100/100

94.3/97

Sub-lineage
94.4/88

100/100

100/100
100/97

East-Kalimantan

Lineage A   100/100

0.004

Papua1 Low

Papua32 Medium

Papua30 High

Papua27 Medium

Papua5 Low

Sea Indonesia Open

Sea Australia Open

Papua4 High

Kalimantan2 Low

Kalimantan3 Medium

Kalimantan1

(b)

K7

K8

K9

K10

K11

(c)

AMOVA: Infinite allele model, Lineage A + B, 10% missing data, 157 SNPs

Source of variation	%Variance	F-value	Std. Dev	p-value
Within individual	0.212	0.788	0.036	
Among individuals within populations	−0.012	−0.06	0.05	0.994
Among populations within lineages	0.180	0.473	0.042	0.001
Among lineages	0.620	0.620	0.065	0.001

AMOVA: Infinite allele model, Lineage B, 10% missing data, 2,106 SNPs

Source of variation	%Variance	F-value	Std. Dev	p-value
Within individual	0.487	0.513	0.012	
Among individuals within populations	−0.044	−0.099	0.012	1
Among populations within sub-lineages	0.488	0.524	0.009	0.001
Among sub-lineages	0.069	0.069	0.016	0.001

$H_e = 0.054$) and Papua27 ($\pi = 0.0037$, $H_e = 0.038$). However, for populations Kalimantan2 and Papua5, there were differences in estimates on diversity depending on which method was chosen. When estimating heterozygosity from genotype likelihoods via ANGSD, we found high diversity for Kalimantan2 (0.1200) and low diversity for Papua5 (0.0188), yet when estimating expected heterozygosity via GENODIVE we found low diversity for Kalimantan2 (0.034) and high diversity for Papua5 (0.095). For Kalimantan2, this may be an artifact of low sample size ($n = 2$), as GENODIVE takes sample size into account. These were the only two lakes in which such a remarkable difference between estimates was observed.

We assessed changes in effective population size for six locations using Stairway plots (removing populations with small sample sizes from the analysis; Figure 4). All locations showed a decrease in effective population size after the Last Glacial Maximum (approximately 20,000 years ago). For the lagoon population Sea Australia and the highly connected lakes Papua27 and Papua4, this bottleneck was followed by an expansion (although less pronounced for Papua4). For lagoon population Sea Indonesia and low connected lakes Papua1 and Papua5 no notable subsequent expansion was observed.

3.6 | Association to geographic distance, environmental variables, and degree of connection

Spearman correlation tests indicated that within-population genomic diversity (nucleotide diversity π and heterozygosity H_e) was not influenced by lake area (π: Spearman's rho $= 0.03$, $p = .95$, H_e: rho $= -0.06$, $p = .88$), degree of connection (π: rho $= 0.43$, $p = .25$, H_e: rho $= 0.53$, $p = .15$), salinity (π: rho $= 0.37$, $p = .33$, H_e: rho $= 0.24$, $p = .53$), or temperature (π: rho $= -0.61$, $p = .08$, H_e: rho $= -0.20$, $p = .60$) (Table S5, Figure S6).

Mantel tests, furthermore, indicated no correlation between the geographic and genomic distance matrices ($r = 0.007$, $p = .50$) (Figure S7A). This was consistent when repeating the analysis for all filter options (Table S6). Finding no correlation refutes the isolation-by-distance hypothesis and indicates other factors might explain the distribution of *S. diversicolor* genomic diversity. However, the genomic distance matrix also did not correlate with matrices of environmental distance ($r = 0.002$, $p = .503$, Figure S7B), or connection distance ($r = 0.041$, $p = .441$, Figure S7C). Therefore, we did not find evidence for isolation-by-environment or isolation-by-resistance patterns.
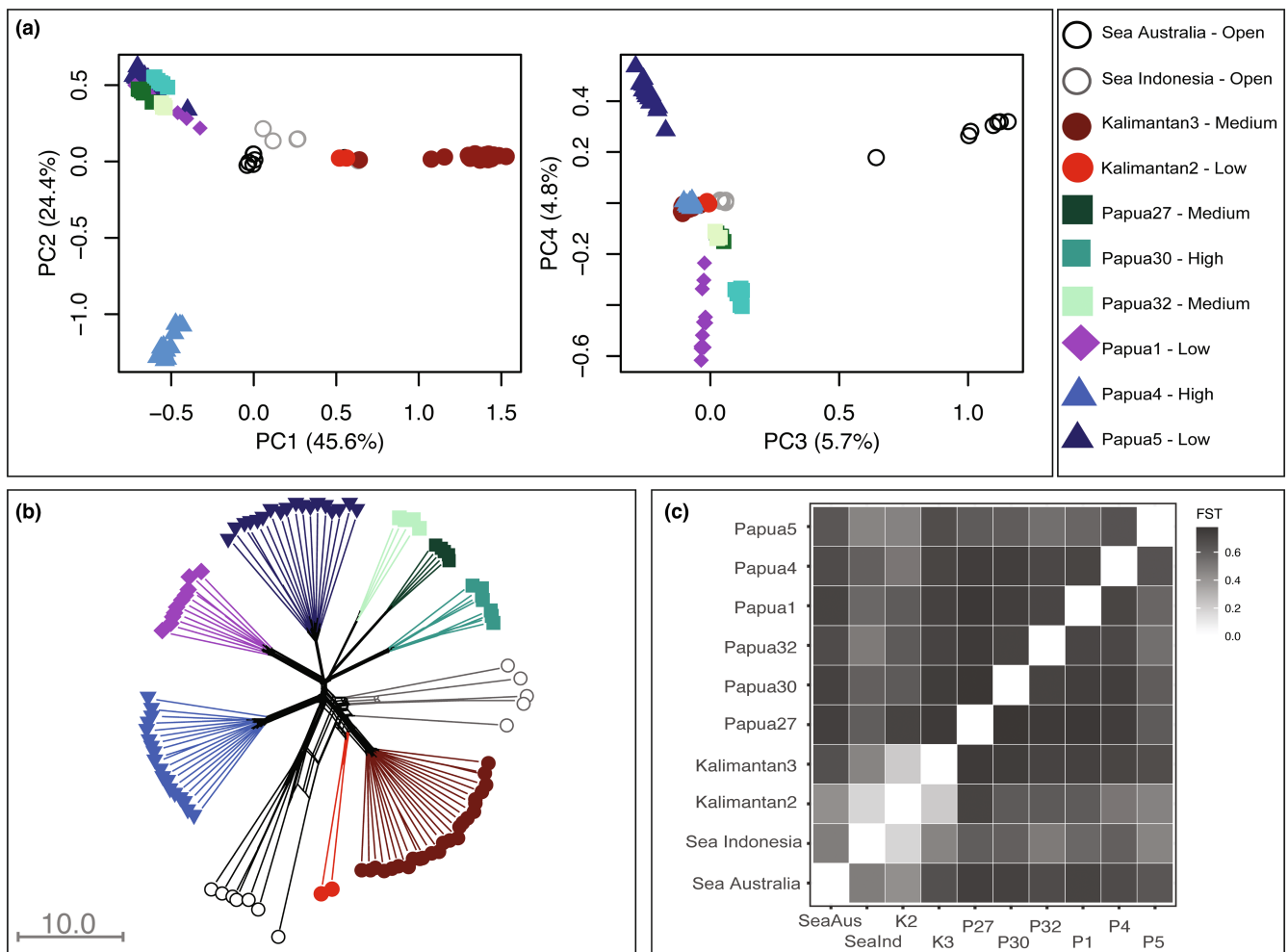

FIGURE 3 Population genomic structure analyses for Lineage B from sponge populations. (a) Principal Component Analysis (PCA) based on pairwise covariance. Each dot represents one individual. (b) Neighbor-Joining Network with equal angles computed in Splitstree based on pairwise genomic distances. (c) Heatmap of normalized F'_{ST} (values in Table S5). Colors and codes correspond to Figure 1 and Table 1.

4 | DISCUSSION

The existence of multiple independently derived populations in marine lakes provides an opportunity for fundamental research into the role of isolation in population divergence in marine taxa. By comparing sponge populations in the Indo-Pacific from marine lakes and lagoons at different spatial scales, environmental conditions, and degree of connection to the sea, we detected high levels of genomic differentiation across the studied geographic area and provided new evidence of small scale structure for sessile species with a short dispersive larval stage. The structure of the populations could, however, not be explained by geographic distance, local environments or degree of connection to the sea alone. Our work exemplifies that a higher resolution of markers retained with reduced representation genome sequencing can elucidate small-scale population genomic structure which are not visible with single marker studies. In general, we showed that old extractions yielding low-coverage data from nonmodel organisms can readily be used in population genetic and genomic analyses to study small-scale population differentiation. Below, we discuss our findings and potential implications for future phylogeographic and population genomic studies on sponges.

4.1 | RADseq reveals small scale structure in Suberites diversicolor

The sponge populations show strong population structure and high levels of genomic differentiation across the study area. These results are in concordance with our hypothesis based on the life history traits of benthic sponges with a predominantly sessile life cycle and short larval durations (Maldonado, 2006). Restriction site-associated DNA sequencing proved suitable to retrieve major genetic lineages in the sponge *S. diversicolor*, which had been split in Lineage A and B by Becking et al. (2013), using the Cytochrome Oxydase I (*COI*) and the Internal Transcribed Spacer 1&2 (*ITS*) nuclear markers. Investigation of morphological traits based on spicules and analyses of *COI* and *ITS* markers determined that Lineage B is likely one species, yet did not show any spatial intraspecific variation across populations (Becking et al., 2013). Diving deeper into Lineage B, the thousands of SNPs that the RADseq approach of the current study provided revealed small-scale (<10 km) genomic patterns for *S. diversicolor* that had not previously been shown. We observed high intralineage genomic differentiation (F_{ST} range 0.18–0.78), with admixture PCA and Splitstree analyses showing clear clustering by marine lake.

Finding more structure when using higher numbers of genomic markers has previously been shown for other marine organisms (e.g., D'Aloia et al., 2020; Lemopoulos et al., 2019; Maas et al., 2018; Sunde et al., 2020; Timm, 2020). Thus, increasing the marker panel can result in the reassessment of previously assumed panmictic populations. Particularly in the highly diverse Indo-Pacific, population genomic studies are increasingly revealing high population structure (Hernawan et al., 2017; Lal et al., 2017; Vu et al., 2020). The signature

of strong population differentiation observed for the marine lake sponges using high-resolution data now better matches those of co-distributed species in marine lakes (Dawson & Hamner, 2005; de Leeuw et al., 2020; Maas et al., 2018). Population genomic studies point toward the role of both historical and contemporary processes in establishing the current population genomic structure. Yet for marine invertebrates, especially sponges, the underlying mechanisms driving the structure are poorly understood (Oleksiak & Rajora, 2020; Pérez-Portela & Riesgo, 2020).

4.2 | Influence of geographic distance, environment and connection on genomic structure and diversity

The results highlighted the effects of several drivers of sponge population diversity and structure. First, the analysis of genomic differentiation showed that population similarity among sponges did not decay with increasing geographic distance. While populations were strongly clustered per marine lake, with a clear regional split between populations from East-Kalimantan and West-Papua, we did not observe a pattern of isolation-by-distance. This is remarkable, as we sampled at a wide range of geographical distances (1–1400 km). Previous studies using a low number of markers did find a pattern of isolation-by-distance for sponges in oceanic locations (Blanquer & Uriz, 2010; Duran et al., 2004; Noyer & Becerro, 2012; Pérez-Portela et al., 2014). Therefore, we would have expected the ocean populations seeding the lake populations to show isolation-by-distance patterns, which would then be reflected in the lakes even if they had no ongoing gene flow among the individual lakes. However, as we could only sampled two lagoon populations, it is possible that we did not sample the actual seeding populations. Secondly, we also did not detect a pattern of isolation-by-environment (Nosil et al., 2009; Orsini et al., 2013; Rundle & Nosil, 2005), despite the great environmental disparity among lakes (temperature: 29–32.4°C, salinity: 24–33.4 ppt). This finding contrasts another study on sponges reporting an influence of environmental heterogeneity (temperature and productivity; Giles et al., 2015) on population structure. As there is clearly a "lake-effect," we may not have measured the environmental variables that are relevant for the structuring. Finally, we did not find evidence for higher gene flow among the highly connected lakes as compared to isolated lakes, disputing the presence of an isolation-by-resistance pattern. Other studies did find evidence for dispersal limitations through oceanographic currents (Chaves-Fonnegra et al., 2015; Richards et al., 2016; Riesgo et al., 2019).

Our results indicate that mechanisms other than dispersal limitation or divergence due to differences in local environments are important in structuring *S. diversicolor* populations. Perhaps *S. diversicolor* populations are isolated in the lake environment, irrespective of the degree of connection of the lake to the adjacent sea, where the low dispersal ability of sponges restricts effective gene flow. Populations can then become differentiated through genetic drift or via local adaptation to environmental parameters not yet recorded
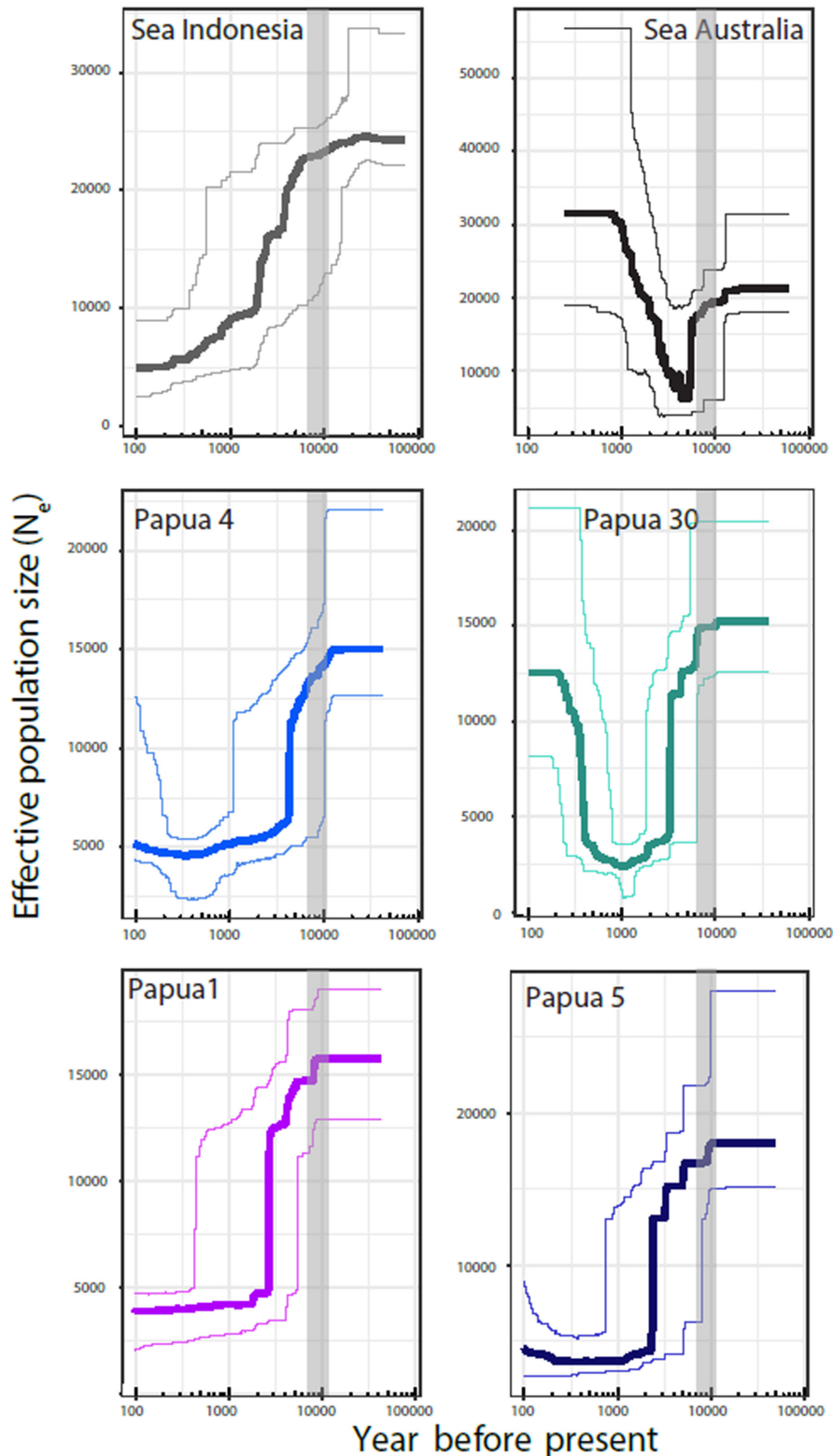
FIGURE 4 Demographic history inferences on sponge populations within Lineage B. Demographic histories of six locations are displayed: lagoon populations Sea Indonesia and Sea Australia (top), highly connected marine lakes Papua 4 and Papua30 (middle), and isolated lakes Papua1 and Papua5 (bottom). Mean (dark) and 12.5/87.5% confidence intervals are displayed. The gray box indicates putative filling of marine lakes after the Last Glacial Maximum (8000–10,000 years ago).

(Frankham et al., 2002). The observation of a severe population bottlenecks with subsequent recovery in Papua30, and population declines with little-to-no expansion in Papua1, Papua4 and Papua5, suggests a potential effect of founder events, which in turn leads to increased drift effects. Founder effects and subsequent priority effects could explain the pattern of strong population structure (De Meester et al., 2016; Fukami, 2015; Orsini et al., 2013).

Priority effects were previously discussed as potential drivers of structure in marine lake organisms by Maas et al. (2018) and de Leeuw et al. (2020). Depending on the spatial scale that was analyzed, Maas et al. (2018) found an effect of geographic distance and connectivity influencing mussel population structure. They argued that while founder events can stochastically drive alleles to fixation in small populations, ongoing dispersal would overwhelm this effect

(Mayr, 1963; Waters et al., 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 (De Meester et al., 2016; Fukami, 2015; Orsini et al., 2013). 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 amplified by founder events may more likely be the cause of each population being distinct. Future studies should further attempt to decouple genetic drift from migration, for instance by using a rigorous modeling approach as provided by fastsimcoal (Excoffier & Foll, 2011), or equivalent.

4.3 | Implications for sponge phylogeography and population genomic studies

The RADseq strategy effectively detected two major genetic lineages (Lineage A and B) (similar to Becking et al., 2013). When combining both lineages, significantly fewer markers were recovered than when analyzing the lineages separately. Based on our filtering strategy, we retained 541 SNPs when including both lineages, compared with 4826 SNPs when analyzing only Lineage B. This reflects a more than 90% loss of common markers and indicates that the resolution of RADseq generated markers can be less effective when multiple lineages are (unknowingly) included. Given that morphologically cryptic species are prevalent in sponges (e.g., Becking, 2013; Morrow & Cárdenas, 2015; Swierts et al., 2013), it may be advised to first verify broad genetic lineages using traditional single markers before starting an extensive sponge population genomic study implementing high-resolution markers.

While traditionally single markers have been successful in identifying sponge species complexes, they have often lacked the resolution for studies on intraspecific population genetic diversity (as reviewed in Pérez-Portela & Riesgo, 2020; Uriz & Turon, 2012; Van Oppen et al., 2002; Wörheide et al., 2005), with notable exceptions (Klautau et al., 1999; Wörheide et al., 2002, 2008). Using the high resolution provided by RADseq generated markers allowed us to detect clear clustering by lake even on very small spatial scales: 1–10 km. The scale at which we find strong structure is smaller compared with recent studies using microsatellites in the sponges *Xestospongia muta* (Richards et al., 2016), *Paraleucilla magna* (Guardiola et al., 2016), *Plenaster cragi* (Taboada et al., 2018), and *Petrosia ficiformis* (Riesgo et al., 2019). Previous studies using SNPs at higher resolution also revealed little structure at small spatial scales, with Brown et al. (2017) detecting little structuring for *Aphrocallistes vastus* in British Colombia at scales of <275 km and Leiva et al. (2019) finding panmixia at scales <900 km for *Dendrilla antarctica*. It could be that these are indeed highly connected populations, possibly through rafting or sperm-mediated gene flow (DeBiasse et al., 2014; Maldonado, 2006). 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. However, even when using a larger SNP dataset, Taboada et al. (2022) found high connectivity on a geographic range of 2500 km for the deep-sea sponge *Phakellia ventilabrum*, suggesting findings of connectivity can be ecosystem- and species-specific.

A challenge in high-resolution genomic studies is presented by potential effects on downstream analyses by different bioinformatic filtering strategies. We rigorously tested the effects of several bioinformatic genotype calling and filtering procedures one can opt for when working with RADseq data. We tested the effect of coverage (3× or 10×) and missing data included (max. 30%, 10%, 5% or 1%). Additionally, we tested the effects of using either genotype calls used by GENODIVE (Meirmans & Van Tienderen, 2004), or genotype likelihoods used by ANGSD (Korneliussen et al., 2014). While the stringency of the filter had severe effects on the number of SNPs included in the analyses (4826 in the most relaxed option versus 56 in the strictest option, Table S2), the downstream analyses showed consistent results over all filtering options (Figures S2–S4, Tables S3–S5). The patterns of highest versus lowest genomic nucleotide diversity, heterozygosity and expected heterozygosity (Table S3), highest versus lowest genetic differentiation (F'_{ST}, Table S4), and associations of diversity with environmental predictors (Table S5) remained remarkably consistent. Hence, when patterns are as strong as we find for the marine lakes in the Indo-Pacific, it matters little how many SNPs are used or how strictly coverage or amount of missing data is selected for. Although when using very few SNPs, the population structure observed in the PCAs did begin to dissociate (Figures S2 and S3). Other studies using RADseq data found evidence for both a high effect (Marandel et al., 2020; Shafer et al., 2017) and a low effect (Tripp et al., 2017) of filtering on downstream analyses. In general, we recommend each study to perform a sensitivity analysis regarding filtering options and downstream results, as it may be different for each specific study system (Díaz-Arce & Rodríguez-Ezpeleta, 2019).

Another challenge to unveiling sponge population genomic patterns is the association between the host sponge and its microbial communities (Webster & Thomas, 2016). Since sponges have the propensity to harbor dense communities of microbes, there is a potential of host genomic material being contaminated by that of microbes, thus clouding host-specific patterns. As microbial biodiversity patterns can differ from sponge host diversity (Noyer & Becerro, 2012), it is important to filter these out. We screened for loci from putative microbes using different filtering strategies and feel confident that the results reflect host diversity. Furthermore, the lack of congruence between the host population structure and previously published microbial communities in *S. diversicolor* from the same locations (Cleary et al., 2013; Ferreira et al., 2020) confirmed that the rigorous bioinformatic filtering removed associated microbes. However, microbial contamination remains a point of attention for future studies on sponge population genomics using RADseq.

Finally, our adjustments to the existing low-cost protocol of Peterson et al. (2012) with a step-by-step protocol presented in the

supplement can help retrieve extensive data for non-model marine organisms in general and tropical sponges in particular, thus benefitting future studies. We further showed that reduced representation genome sequencing can work for DNA extracted for other purposes and stored for long times in a −20°C 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 studies across the world.

5 | CONCLUDING REMARKS

In conclusion, we observed high population genomic structure for the sponge _S. diversicolor_ even on small spatial scales of <10 km that was previously undetected using single markers. This finding reflects the life history of the sponge with short larval durations, but also highlights the importance of higher genomic resolution to detect structure. Based on this dataset, we could not link the observed strong genomic differentiation to any of the expected predictors (geographic distance, local environments or extent of connection to the sea). As we observed population bottlenecks for all marine lake locations, regardless of their extent of connection to the sea, other factors such as strong founder events coupled with priority effects could be at play. A major objective of marine molecular ecology is to obtain accurate estimates of genomic structure, as it can inform efforts to identify units of management and design effective marine protected areas (MPAs) (Kelley et al., 2016; Selkoe et al., 2016). Our results call for a reassessment of population connectivity of poorly dispersing organisms such as sessile benthic organisms, as previous assumptions of panmixia may be an artifact of low-resolution markers. In turn, these studies can then more accurately inform the determination of relevant spatial scales of MPA networks.

AUTHOR CONTRIBUTIONS
Diede L. Maas: Formal analysis (lead); investigation (equal); methodology (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Stefan Prost:** Formal analysis (supporting); methodology (supporting); software (equal); supervision (supporting); visualization (supporting); writing – review and editing (supporting). **Christiaan A. de Leeuw:** Data curation (supporting); investigation (supporting); methodology (supporting); writing – review and editing (supporting). **Ke Bi:** Formal analysis (supporting); methodology (supporting); resources (equal); software (lead); writing – review and editing (supporting). **Lydia L. Smith:** Formal analysis (supporting); methodology (supporting); supervision (supporting); writing – review and editing (supporting). **Purwanto Purwanto:** Resources (supporting); writing – review and editing (supporting). **Ludi P. Aji:** Resources (supporting); writing – review and editing (supporting). **Ricardo F. Tapilatu:** Resources (supporting); writing – review and editing (supporting). **Rosemary G. Gillespie:** Supervision (supporting); writing – review and editing (supporting). **Leontine E. Becking:** Conceptualization (lead); data curation (lead); formal analysis (supporting); funding acquisition (lead); investigation (lead); methodology (lead); supervision (lead); visualization (supporting); writing – original draft (equal); writing – review and editing (equal).

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
Data and customary scripts are accessible via the Dryad Digital Repository (https://doi.org/10.5061/dryad.5mkkwh79w). Data

consists of aligned reads in bam format for all 125 individuals sequenced, with an accompanying text file assigning the individuals and the de-novo reference file in fasta format. Additionally, the tags mapped to microbes are included. Finally, custom scripts RADTOOLKIT v. 0.13.10 and SNPcleaner224.pl are stored here.

PERMISSION TO REPRODUCE MATERIALS FROM OTHER SOURCES
None.

ORCID
Diede L. Maas https://orcid.org/0000-0002-3599-1364

REFERENCES
Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

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), 3–15.

Azzini, F., Calcinai, B., Cerrano, C., Bavestrello, G., & Pansini, M. (2007). Sponges of the marine karst lakes and of the coast of the islands of Ha Long Bay (North Vietnam). In M. R. Custodia, G. Lobo-Hajdu, E. Hajdu, & G. Muricy (Eds.), *Porifera research: Biodiversity innovation and sustainability* (pp. 157–164). Museu Nacional, Universidade Federal do Rio de Janeiro.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, *3*(10), e3376.

Becking, L. E. (2013). Revision of the genus Placospongia (Porifera, Demospongiae, Hadromerida, Placospongiidae) in the indo-West Pacific. *ZooKeys*, *298*, 39–76.

Becking, L. E., de Leeuw, C., & Vogler, C. (2015). Newly discovered "jellyfish lakes" in Misool, Raja Ampat, Papua, Indonesia. *Marine Biodiversity*, *45*, 597–598.

Becking, L. E., de Leeuw, C. A., Knegt, B., Maas, D. L., De Voogd, N. J., Suyatna, I., & Peijnenburg, K. T. (2016). Highly divergent mussel lineages in isolated Indonesian marine lakes. *PeerJ*, *4*, e2496.

Becking, L. E., Erpenbeck, D., Peijnenburg, K. T., & de Voogd, N. J. (2013). Phylogeography of the sponge Suberites diversicolor in Indonesia: Insights into the evolution of marine lake populations. *PLoS One*, *8*(10), e75996.

Becking, L. E., & Lim, S. C. (2009). A new Suberites (Demospongiae: Hadromerida: Suberitidae) from the tropical indo-West Pacific. *Zoologische Mededelingen*, *83*(29), 853–862.

Becking, L. E., Renema, W., Santodomingo, N. K., Hoeksema, B. W., Tuti, Y., & de Voogd, N. J. (2011). Recently discovered landlocked basins in Indonesia reveal high habitat diversity in anchialine systems. *Hydrobiologia*, *677*, 89–105.

Bell, J. J. (2008). The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science*, *79*(3), 341–353.

Bentlage, B., & 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), 807–816.

Bernatchez, S., Xuereb, A., Laporte, M., Benestan, L., Steeves, R., Laflamme, M., Bernatchez, L., & Mallet, M. A. (2019). Seascape genomics of eastern oyster (*Crassostrea virginica*) along the Atlantic coast of Canada. *Evolutionary Applications*, *12*(3), 587–609.

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., Moritz, C., & Good, J. M. (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS Genetics*, *15*(5), e1008119.

Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2013). Unlocking the vault: Next-generation museum population genomics. *Molecular Ecology*, *22*(24), 6018–6032.

Bierne, N., Bonhomme, F., & Arnaud-Haond, S. (2016). Editorial dedicated population genomics for the silent world: The specific questions of marine population genetics. *Current Zoology*, *62*(6), 545–550.

Blanquer, A., & Uriz, M. J. (2010). Population genetics at three spatial scales of a rare sponge living in fragmented habitats. *BMC Evolutionary Biology*, *10*, 1–9.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., & Verspoor, E. (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), 5130–5144.

Brown, R. R., Davis, C. S., & 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*, 49–51.

Brown, R. R., Davis, C. S., & 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), 1045–1059.

Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F. W. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources*, *17*(3), 362–365.

Chaves-Fonnegra, A., Feldheim, K. A., Secord, J., & Lopez, J. V. (2015). Population structure and dispersal of the coral-excavating sponge *Cliona delitrix*. *Molecular Ecology*, *24*(7), 1447–1466.

Chen, E. Y. S. (2021). Often overlooked: Understanding and meeting the current challenges of marine invertebrate conservation. *Frontiers in Marine Science*, *8*, 690704.

Cleary, D. F., Becking, L. E., Voogd, N. J. D., Pires, A. C., Polónia, A. R., Egas, C., & Gomes, N. C. (2013). Habitat-and host-related variation in sponge bacterial symbiont communities in Indonesian waters. *FEMS Microbiology Ecology*, *85*(3), 465–482.

Cornwell, B. H., Fisher, J. L., Morgan, S. G., & Neigel, J. E. (2016). Chaotic genetic patchiness without sweepstakes reproduction in the shore crab *Hemigrapsus oregonensis*. *Marine Ecology Progress Series*, *548*, 139–152.

D'Aloia, C. C., Andrés, J. A., Bogdanowicz, S. M., McCune, A. R., Harrison, R. G., & Buston, P. M. (2020). Unraveling hierarchical genetic structure in a marine metapopulation: A comparison of three high-throughput genotyping approaches. *Molecular Ecology*, *29*(12), 2189–2203.

Dawson, M. N., & Hamner, W. M. (2005). Rapid evolutionary radiation of marine zooplankton in peripheral environments. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(26), 9235–9240.

Dawson, M. N., Martin, L. E., Bell, L. J., & Patris, S. (2009). Marine lakes. *Encyclopedia of islands*, 603–607.

de Bakker, D. M., Meesters, E. H., van Bleijswijk, J. D., Luttikhuizen, P. C., Breeuwer, H. J., & Becking, L. E. (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), e0155969.

De Goeij, J. M., Van Oevelen, D., Vermeij, M. J., Osinga, R., Middelburg, J. J., De Goeij, A. F., & Admiraal, W. (2013). Surviving in a marine desert: The sponge loop retains resources within coral reefs. *Science*, *342*(6154), 108–110.

de Leeuw, C. A., Peijnenburg, K. T., Gillespie, R. G., Maas, D. L., Hanzawa, N., Tuti, Y., Toha, A. H. A., Aji, L. P., & Becking, L. E. (2020). First come, first served: Possible role for priority effects in marine populations under different degrees of dispersal potential. *Journal of Biogeography*, *47*(8), 1649–1662.

De Meester, L., Vanoverbeke, J., Kilsdonk, L. J., & Urban, M. C. (2016). Evolving perspectives on monopolization and priority effects. *Trends in Ecology & Evolution*, *31*(2), 136–146.

DeBiasse, M. B., Nelson, B. J., & Hellberg, M. E. (2014). Evaluating summary statistics used to test for incomplete lineage sorting: Mitonuclear discordance in the reef sponge C allyspongia vaginalis. *Molecular Ecology*, *23*(1), 225–238.

Díaz-Arce, N., & Rodríguez-Ezpeleta, N. (2019). Selecting RAD-Seq data analysis parameters for population genetics: The more the better? *Frontiers in Genetics*, *10*, 533.

Duran, S., Pascual, M., Estoup, A., & Turon, X. (2004). Strong population structure in the marine sponge *Crambe crambe* (Poecilosclerida) as revealed by microsatellite markers. *Molecular Ecology*, *13*(3), 511–522.

Ekins, M., Erpenbeck, D., Wörheide, G., & Hooper, J. N. (2016). Staying well connected–Lithistid sponges on seamounts. *Journal of the Marine Biological Association of the United Kingdom*, *96*(2), 437–451.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620.

Excoffier, L., & Foll, M. (2011). fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics*, *27*(9), 1332–1334.

Excoffier, L., Smouse, P. E., & Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491.

Ferreira, M. R. S., Cleary, D. F. R., Coelho, F. J. R. C., Gomes, N. C. M., Huang, Y. M., Polónia, A. R. M., & de Voogd, N. J. (2020). Geographical location and habitat predict variation in prokaryotic community composition of Suberites diversicolor. *Annals of Microbiology*, *70*, 1–12.

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

Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152.

Fukami, T. (2015). Historical contingency in community assembly: Integrating niches, species pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 1–23.

Fumagalli, M., Vieira, F. G., Linderoth, T., & Nielsen, R. (2014). ngsTools: Methods for population genetics analyses from next-generation sequencing data. *Bioinformatics*, *30*(10), 1486–1487.

Giles, E. C., Saenz-Agudelo, P., Hussey, N. E., Ravasi, T., & Berumen, M. L. (2015). Exploring seascape genetics and kinship in the reef sponge S tylissa carteri in the R ed S Ea. *Ecology and Evolution*, *5*(13), 2487–2502.

Gotoh, R. O., Chiba, S. N., Goto, T. V., Tamate, H. B., & Hanzawa, N. (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), 325–337.

Guardiola, M., Frotscher, J., & 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*, *163*, 1–16.

Hamner, W. M., Gilmer, R. W., & Hamner, P. P. (1982). The physical, chemical, and biological characteristics of a stratified, saline, sulfide lake in Palau 1. *Limnology and Oceanography*, *27*(5), 896–909.

Hernawan, U. E., van Dijk, K. J., Kendrick, G. A., Feng, M., Biffin, E., Lavery, P. S., & McMahon, K. (2017). Historical processes and contemporary ocean currents drive genetic structure in the seagrass T halassia hemprichii in the indo-Australian archipelago. *Molecular Ecology*, *26*(4), 1008–1021.

Holthuis, L. B. (1973). Caridean shrimps found in land-locked saltwater 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*(1), 1–48.

Horn, H., Slaby, B. M., Jahn, M. T., Bayer, K., Moitinho-Silva, L., Förster, F., Abdelmohsen, U. R., & Hentschel, U. (2016). An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Frontiers in Microbiology*, *7*, 1751.

Huang, D., Meier, R., Todd, P. A., & Chou, L. M. (2008). Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. *Journal of Molecular Evolution*, *66*, 167–174.

Huson, D. H. (1998). SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics (Oxford, England)*, *14*(1), 68–73.

Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267.

Jeffery, N. W., Jardine, C. B., & Gregory, T. R. (2013). A first exploration of genome size diversity in sponges. *Genome*, *56*(8), 451–456.

Johnson, P. L., & Slatkin, M. (2008). Accounting for bias from sequencing error in population genetic estimates. *Molecular Biology and Evolution*, *25*(1), 199–206.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (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), 782–788.

Kelley, J. L., Brown, A. P., Therkildsen, N. O., & Foote, A. D. (2016). The life aquatic: Advances in marine vertebrate genomics. *Nature Reviews Genetics*, *17*(9), 523–534.

Klautau, M., Russo, C. A., Lazoski, C., Boury-Esnault, N., Thorpe, J. P., & Solé-Cava, A. M. (1999). Does cosmopolitanism result from overconservative systematics? A case study using the marine sponge Chondrilla nucula. *Evolution*, *53*(5), 1414–1422.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 1–13.

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*(1287), 1287.

Lal, M. M., Southgate, P. C., Jerry, D. R., Bosserelle, C., & Zenger, K. R. (2017). Swept away: Ocean currents and seascape features influence genetic structure across the 18,000 km indo-Pacific distribution of a marine invertebrate, the black-lip pearl oyster Pinctada margaritifera. *BMC Genomics*, *18*, 1–21.

Legendre, J. V., & Legendre, L. F. (2012). *Numerical ecology*. Elsevier.

Leiva, C., Taboada, S., Kenny, N. J., Combosch, D., Giribet, G., Jombart, T., & Riesgo, A. (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*, *28*(13), 3151–3170.

Lemopoulos, A., Prokkola, J. M., Uusi-Heikkilä, S., Vasemägi, A., Huusko, A., Hyvärinen, P., Koljonen, M. L., Koskiniemi, J., & Vainikka, A.

(2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness—Implications for brown trout conservation. *Ecology and Evolution*, 9(4), 2106–2120.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.

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

Liggins, L., Treml, E. A., & Riginos, C. (2013). Taking the plunge: An introduction to undertaking seascape genetic studies and using biophysical models. *Geography Compass*, 7(3), 173–196.

Liu, X., & Fu, Y. X. (2015). Exploring population size changes using SNP frequency spectra. *Nature Genetics*, 47(5), 555–559.

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

Lynch, M. (2010). Evolution of the mutation rate. *Trends in Genetics*, 26(8), 345–352.

Maas, D. L., Capriati, A., Ahmad, A., Erdmann, M. V., Lamers, M., de Leeuw, C. A., Prins, L., Purwanto, P. A. P., Tapilatu, R. F., & Becking, L. E. (2020). Recognizing peripheral ecosystems in marine protected areas: A case study of golden jellyfish lakes in Raja Ampat, Indonesia. *Marine Pollution Bulletin*, 151, 110700.

Maas, D. L., Prost, S., Bi, K., Smith, L. L., Armstrong, E. E., Aji, L. P., Toha, A. H. A., Gillespie, R. G., & Becking, L. E. (2018). Rapid divergence of mussel populations despite incomplete barriers to dispersal. *Molecular Ecology*, 27(7), 1556–1571.

Maldonado, M. (2006). The ecology of the sponge larva. *Canadian Journal of Zoology*, 84(2), 175–194.

Maldonado, M., & Uriz, M. J. (1999). Sexual propagation by sponge fragments. *Nature*, 398(6727), 476.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2_Part_1), 209–220.

Marandel, F., Charrier, G., Lamy, J. B., Le Cam, S., Lorance, P., & Trenkel, V. M. (2020). Estimating effective population size using RADseq: Effects of SNP selection and sample size. *Ecology and Evolution*, 10(4), 1929–1937.

Marshall, D. J., Monro, K., Bode, M., Keough, M. J., & Swearer, S. (2010). Phenotype–environment mismatches reduce connectivity in the sea. *Ecology Letters*, 13(1), 128–140.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1), 10–12.

Mayr, E. (1963). *Animal species and evolution*. Harvard University Press.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.

McRae, B. H. (2006). Isolation by resistance. *Evolution*, 60(8), 1551–1561.

Meirmans, P. G., & Hedrick, P. W. (2011). Assessing population structure: FST and related measures. *Molecular Ecology Resources*, 11(1), 5–18.

Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, 4(4), 792–794.

Miller, K. J., Baird, H. P., Van Oosterom, J., Mondon, J., & King, C. K. (2018). Complex genetic structure revealed in the circum-Antarctic broadcast spawning sea urchin Sterechinus neumayeri. *Marine Ecology Progress Series*, 601, 153–166.

Morrow, C., & Cárdenas, P. (2015). Proposal for a revised classification of the Demospongiae (Porifera). *Frontiers in Zoology*, 12, 1–27.

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

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268–274.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. A. N. I. E. L. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402.

Noyer, C., & Becerro, M. A. (2012). Relationship between genetic, chemical, and bacterial diversity in the Atlanto-Mediterranean bath sponge Spongia lamella. *Hydrobiologia*, 687(1), 85–99.

Oleksiak, M. F., & Rajora, O. P. (Eds.). (2020). *Population genomics: Marine organisms*. Springer.

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J., & De Meester, L. (2013). Drivers of population genetic differentiation in the wild: Isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, 22(24), 5983–5999.

Peijnenburg, K. T., & Goetze, E. (2013). High evolutionary potential of marine zooplankton. *Ecology and Evolution*, 3(8), 2765–2781.

Pérez-Portela, R., Noyer, C., & Becerro, M. (2014). Genetic structure and diversity of the endangered bath sponge Spongia lamella. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 25, 365–379.

Pérez-Portela, R., & Riesgo, A. (2020). Population genomics of early-splitting lineages of metazoans. In M. F. Oleksiak & O. P. Rajora (Eds.), *Population genomics: Marine organisms* (pp. 103–137). Springer.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7(5), e37135.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014). Demystifying the RAD fad. *Molecular Ecology*, 23, 5937–5942.

Richards, V. P., Bernard, A. M., Feldheim, K. A., & Shivji, M. S. (2016). Patterns of population structure and dispersal in the long-lived "redwood" of the coral reef, the giant barrel sponge (*Xestospongia muta*). *Coral Reefs*, 35, 1097–1107.

Riesgo, A., Taboada, S., Pérez-Portela, R., Melis, P., Xavier, J. R., Blasco, G., & López-Legentil, S. (2019). Genetic diversity, connectivity and gene flow along the distribution of the emblematic Atlanto-Mediterranean sponge *Petrosia ficiformis* (Haplosclerida, Demospongiae). *BMC Evolutionary Biology*, 19(1), 1–18.

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, 8(3), 336–352.

Sathiamurthy, E., & Voris, H. K. (2006). Maps of Holocene Sea level transgression and submerged lakes on the Sunda shelf. *Tropical Natural History*, 2, 1–44.

Selkoe, K. A., Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., von der Heyden, S., & Toonen, R. J. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. *Marine Ecology Progress Series*, 554, 1–19.

Shafer, A. B., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution*, 8(8), 907–917.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3), 693–702.

Smit, A., Hubley, R., & Green, P. (2013). *RepeatMasker 4.0*. Institute for Systems Biology.

Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E., Mitros, T., Richards, G. S., Conaco, C., Dacre, M., Hellsten, U., Larroux, C., Putnam, N. H., Stanke, M., Adamska, M., Darling, A., Degnan, S. M., Oakley, T. H., Plachetzki, D. C., Zhai, Y., ... Rokhsar, D. S. (2010). The Amphimedon queenslandica genome and the evolution of animal complexity. *Nature*, 466(7307), 720–726.

Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., Pajkovic, M., Ronikier, M., & Alvarez, N. (2016). Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS One*, *11*(3), e0151651.

Sunde, J., Yıldırım, Y., Tibblin, P., & Forsman, A. (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*, 218.

Sundqvist, L., Keenan, K., Zackrisson, M., Prodöhl, P., & Kleinhans, D. (2016). Directional genetic differentiation and relative migration. *Ecology and Evolution*, *6*(11), 3461–3475.

Swierts, T., Peijnenburg, K. T., de Leeuw, C., Cleary, D. F., Hörnlein, C., Setiawan, E., Wörheide, G., Erpenbeck, D., & de Voogd, N. J. (2013). Lock, stock and two different barrels: Comparing the genetic composition of morphotypes of the indo-Pacific sponge Xestospongia testudinaria. *PLoS One*, *8*(9), e74396.

Taboada, S., Riesgo, A., Wiklund, H., Paterson, G. L., Koutsouveli, V., Santodomingo, N., Dale, A. C., Smith, C. R., Jones, D. O. B., Dahlgren, T. G., & Glover, A. G. (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), 4657–4679.

Taboada, S., Ríos, P., Mitchell, A., Cranston, A., Busch, K., Tonzo, V., Cárdenas, P., Sánchez, F., Leiva, C., Koutsouveli, V., Cristobo, J., Xavier, J. R., Hentschel, U., Rapp, H.-T., Morrow, C., Drewery, J., Romero, P. E., Arias, M. B., Whiting, C., & Riesgo, A. (2022). Genetic diversity, gene flow and hybridization in fan-shaped sponges (*Phakellia* spp.) in the north-East Atlantic deep sea. *Deep Sea Research Part I: Oceanographic Research Papers*, *181*, 103685.

Timm, L. E. (2020). A fair fight between molecular marker types in a seascape genetics setting. *Molecular Ecology*, *29*, 2133–2136.

Tomascik, T., & Mah, A. J. (1994). An anchialine lagoon of a raised atoll, Kakaban Island, East Kalimantan, Indonesia. *Tropical Biodiversity*, *2*(3), 385.

Tripp, E. A., Tsai, Y. H. E., Zhuang, Y., & Dexter, K. G. (2017). RAD seq dataset with 90% missing data fully resolves recent radiation of Petalidium (Acanthaceae) in the ultra-arid deserts of Namibia. *Ecology and Evolution*, *7*(19), 7920–7936.

Uriz, M. J., & Turon, X. (2012). Sponge ecology in the molecular era. *Advances in Marine Biology*, *61*, 345–410.

Van Oppen, M. J. H., Wörheide, G., & Takabayashi, M. (2002, October). Nuclear markers in evolutionary and population genetic studies of scleractinian corals and sponges. In proc 9th Int coral reef Symp (Vol. 1, pp. 131–138).

Van Wyngaarden, M., Snelgrove, P. V., DiBacco, C., Hamilton, L. C., Rodríguez-Ezpeleta, N., Jeffery, N. W., Stanley, R. R. E., & Bradbury, I. R. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, *Placopecten magellanicus*, using RAD seq-derived SNP s. *Evolutionary Applications*, *10*(1), 102–117.

Vu, N. T., Zenger, K. R., Guppy, J. L., Sellars, M. J., Silva, C. N., Kjeldsen, S. R., & Jerry, D. R. (2020). Fine-scale population structure and evidence for local adaptation in Australian giant black tiger shrimp (*Penaeus monodon*) using SNP analysis. *BMC Genomics*, *21*(1), 1–18.

Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., Conti, E., Cornuault, J., Fernández-Palacios, J. M., Hengl, T., Norder, S. J., Rijsdijk, K. F., Sanmartín, I., Strasberg, D., Triantis, K. A., …

Thébaud, C. (2015). Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecology Letters*, *18*(2), 200–217.

Waters, J. M., Fraser, C. I., & Hewitt, G. M. (2013). Founder takes all: Density-dependent processes structure biodiversity. *Trends in Ecology & Evolution*, *28*(2), 78–85.

Webster, N. S., & Thomas, T. (2016). The sponge hologenome. *MBio*, *7*(2), e00135-16.

Whalan, S., de Nys, R., Smith-Keune, C., Evans, B. S., Battershill, C., & Jerry, D. R. (2008). Low genetic variability within and among populations of the brooding sponge Rhopaloeides odorabile on the central great barrier reef. *Aquatic Biology*, *3*(2), 111–119.

Wood, D. E., & Salzberg, S. L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, *15*(3), 1–12.

Wörheide, G., Epp, L. S., & Macis, L. (2008). Deep genetic divergences among indo-Pacific populations of the coral reef sponge Leucetta chagosensis (Leucettidae): Founder effects, vicariance, or both? *BMC Evolutionary Biology*, *8*, 1–18.

Wörheide, G., Hooper, J. N., & Degnan, B. M. (2002). Phylogeography of western Pacific Leucetta 'chagosensis' (Porifera: Calcarea) from ribosomal DNA sequences: Implications for population history and conservation of the great barrier reef world heritage area (Australia). *Molecular Ecology*, *11*(9), 1753–1768.

Wörheide, G., Nichols, S. A., & Goldberg, J. (2004). Intragenomic variation of the rDNA internal transcribed spacers in sponges (phylum Porifera): Implications for phylogenetic studies. *Molecular Phylogenetics and Evolution*, *33*(3), 816–830.

Wörheide, G., Solé-Cava, A. M., & Hooper, J. N. (2005). Biodiversity, molecular ecology and phylogeography of marine sponges: Patterns, implications and outlooks. *Integrative and Comparative Biology*, *45*(2), 377–385.

Wright, S. (1943). Isolation by distance. *Genetics*, *28*(2), 114–138.

Xuereb, A., Benestan, L., Normandeau, É., Daigle, R. M., Curtis, J. M., Bernatchez, L., & Fortin, M. J. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RAD seq, in a highly dispersive marine invertebrate (*Parastichopus californicus*). *Molecular Ecology*, *27*(10), 2347–2364.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.