Islands of Sea

The evolution, ecology and conservation of marine lake invertebrates

Diede Louise Maas

Islands of Sea

The evolution, ecology and conservation of marine lake invertebrates

Diede Louise Maas

Thesis committee

Promotor

Prof. Dr Albertinka J. Murk

Professor of Marine Animal Ecology

Wageningen University & Research

Co-promotors

Dr Leontine E. Becking

Assistant professor, Marine Animal Ecology

Wageningen University & Research

Dr Stefan Prost

Researcher, Natural History Museum

Central Research Laboratories, Vienna, Austria

Other members

Prof. Dr Frank van Langevelde, Wageningen University & Research

Dr Katja T.C.A. Peijnenburg, University of Amsterdam

Dr Hawis Madduppa, University of Bogor, Indonesia

Dr Alejandro M. García, CNR-IRSA: Water Research Institute

This research was conducted under the auspices of the Graduate School Wageningen Institute of Animal Sciences

Islands of Sea

The evolution, ecology and conservation of marine lake invertebrates

Diede Louise Maas

Thesis

Submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 2 June 2021
at 1:30 p.m. in the Aula.

Diede Louise Maas
Islands of Sea - The evolution, ecology and conservation of marine lake invertebrates, 240 pages

PhD thesis, Wageningen University, Wageningen, the Netherlands (2021)
With references, with summaries in English, Dutch and Bahasa Indonesia

ISBN: 978-94-6395-743-4
DOI: 10.18174/543115

Table of contents

Chapter 1 **7**
General Introduction

Chapter 2 **23**
Islands of sea: Biophysical assessments of marine lakes in Raja Ampat, Indonesia

Chapter 3 **57**
Recognizing peripheral ecosystems in marine protected areas: A case study of golden jellyfish lakes in Raja Ampat, Indonesia

Chapter 4 **85**
New evidence for genetic structure in the sponge *Suberites diversicolor*: Implications for sponge phylogeography and population genetics

Chapter 5 **115**
Rapid divergence of mussel populations despite incomplete barriers to dispersal

Chapter 6 **147**
Genetic divergence does not follow environmental acclimatization in marine lake mussels

Chapter 7 **179**
General Discussion

References **197**

Summary/Samenvatting/Ringkasan **213**

Appendices **229**

Chapter 1

General Introduction

A major question of ecologists and evolutionary biologists is how biodiversity arises and how it is maintained. Biodiversity is multifaceted, and genetic diversity within populations is one of its components besides species and ecosystem diversity (Vellend and Geber, 2005). Studies of microevolution focus on intra-specific variation within populations which eventually might lead to the macroevolutionary process of speciation (Hendry et al., 2009). Genetic diversity within populations is usually measured via allele frequencies, or the number of times a variant of a gene occurs in a population. Allele frequencies can change due to mutation, selection, gene flow (migration between populations) and genetic drift (stochastic loss or gain of alleles). The fields of population genetics and genomics aim to improve our understanding of population differentiation (Charlesworth, 2010; Luikart et al., 2018), by questioning how spatial and environmental factors influence microevolutionary processes. Microevolutionary studies in marine systems are lagging behind terrestrial counterparts for several reasons (Selkoe et al., 2008, 2016), first and foremost being the difficulty to access many marine areas. Furthermore, there is a long-standing assumption that marine populations show low rates of genetic differentiation due to the lack of clear physical barriers, assumed high dispersal potential, and associated large effective population sizes (Palumbi, 1994; Ward, 1994; Bowen et al., 2013). While this assumption has both been confirmed and rejected for different organisms (Bierne et al., 2016), in any case the open nature of the marine realm makes defining marine populations *a priori* challenging. Still, recent studies are showing more population structure in marine ecosystems and at finer spatial scales (1-100kms) than expected by predicted larval dispersal distances (Hauser and Carvalho, 2008; Marshall et al., 2010; Van Wyngaarden et al., 2016; Bernatchez et al., 2018). Further elucidating marine population genetic patterns on small scales and shedding light on what drives marine population connectivity and adaptation is particularly pressing in a changing world. This thesis aims to disentangle relative importance of neutral (geographic and dispersal barriers) and selective (local adaptation) processes on marine population genomic variation using the unique opportunity presented by marine lakes which offer replicated, independent natural laboratories of evolution and ecology.

Seascape genomics and predictors of population structure

The terms population genetics and population genomics are broadly synonymous in the sense they both study genetic diversity within and between populations (Charlesworth, 2010; Luikart et al., 2018). Traditionally, population genetics uses single or few genetic markers (or genes) such as mitochondrial DNA or ribosomal RNA. The advantages of these genes are there are a lot of copies per cell and they are expected to be neutrally evolving (Lynch, 1997). Together, these genes have often been used in resolving phylogenetic patterns (Avise, 2000). Population genomics takes a broader perspective and uses numerous loci to study how evolutionary processes affect variation genome-wide. The transition to the higher marker panels of genomics has been made possible by various advances in high-throughput sequencing and associated computing power (Metzker, 2010; Luikart et al., 2018). The advancement allows for the genotyping of hundreds to thousands of loci in populations. The

higher resolution provides improved, unbiased insights into within-population diversity and between-population differentiation, as well as connectivity especially on fine spatial scales compared to single marker approaches (Lemopoulos et al., 2019; D'Aloia et al., 2020; Timm, 2020). Large neutral marker panels can be used to quantify changes in effective population size (demography), and infer migration patterns (Oleksiak and Rajora, 2020a). While for some seascape genetic patterns it has been shown that high-resolution markers confirm previous findings from traditional markers (Timm, 2020), there are examples where the larger marker panel was able to indicate discontinuations in population connectivity and finer demographic signatures not previously shown (Bradbury et al., 2015; Cammen et al., 2016; Crawford and Oleksiak, 2016). Furthermore, and perhaps even more importantly, the larger set of loci available in population genomics increases the chance of revealing loci under putative selection that can provide insights into adaptive variation and organism-environment interactions (Crawford and Oleksiak, 2016). Particularly for marine organisms, which can be difficult to track, the advancements of population genomics provide new opportunities to get a better understanding of connectivity and adaptation.

Seascape genomics is part of the broader population genomics field, but focusses on understanding how biotic and abiotic factors underlie patterns of genetic variation within and between marine populations (Selkoe et al., 2008, 2016; Riginos et al., 2016; Liggins et al., 2020). The ultimate goal is to detect genetic-environment associations. Here, spatial and environmental features can influence neutral genomic diversity mainly through genetic drift and gene flow, where adaptive genomic structure is influenced by selection and local adaptation (Liggins et al., 2020). Population differentiation can be a first phase in the speciation continuum (Hendry et al., 2009). Particularly in the marine realm, where obvious barriers to dispersal are often lacking, processes that can mediate differentiation in the face of ongoing gene flow are thought to be important (Puebla, 2009; Feder et al., 2012). Still, a form of isolation is generally required for differentiation to occur, be it geographical isolation (i.e., allopatry or parapatry) or environmental isolation (i.e., ecological speciation in sympatry). This thesis considers four main modes of isolation, which predictions are often explicitly tested in seascape genomic studies: isolation-by-distance, isolation-by-resistance, isolation-by-environment and effects of historical contingency (including priority effects) (Wright, 1943; McRae, 2006; Orsini et al., 2013; Wang and Bradburd, 2014; Fukami, 2015; De Meester et al., 2016). All theories have a mixed, panmictic population over the spatial scale tested as their null hypothesis. The theories are not mutually exclusive and may have different relative importance depending on the spatial and temporal scale (Orsini et al., 2013). Still, they may provide critical insights into the eco-evolutionary processes that shape diversity.

Predictions on the relationships between genetic variation and spatial and environmental factors differ between these theories (Figure 1). An isolation-by-distance pattern arises when there is a reduction in the movement of individuals over increasing geographic distance (Wright, 1943). Even in marine organisms, where many have at least one pelagic life stage,

propagules become increasingly diluted with increasing geographic distances, lowering the chance of gene flow over large distances (Johannesson, 1988). Similar to isolation-by-distance, the theory of isolation-by-resistance bases its predictions on dispersal limitations of propagules or adult individuals (McRae, 2006). Instead of geographic distance limiting dispersal, here the difficulty of individuals traversing terrain is a central part of this model. The most extreme example is a landmass as a complete barrier, but more complex cases

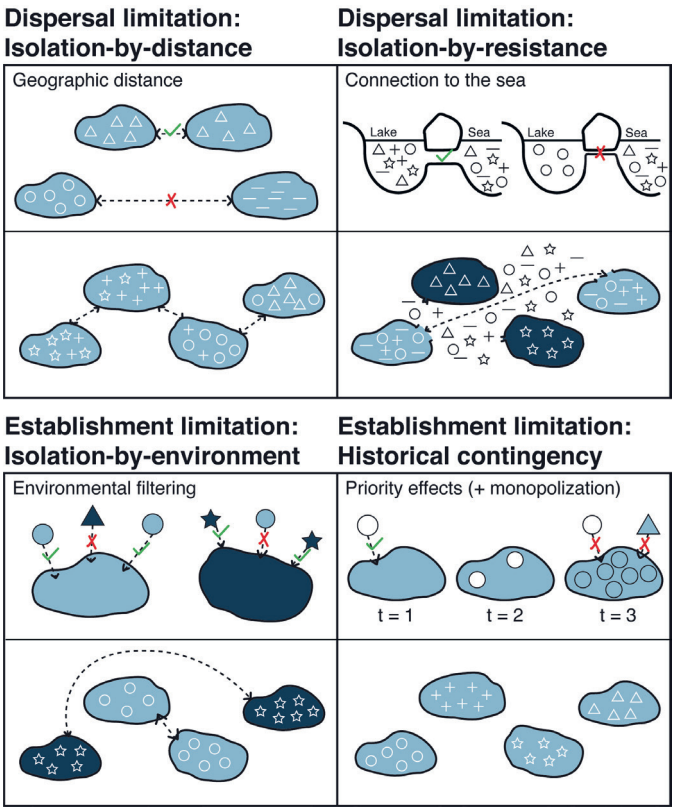


Figure 1: Schematic overview of four modes of isolation: isolation-by-distance, isolation-by-resistance, isolation-by-environment and historical contingency (priority effects and monopolization). Per box, the top picture represents the process, and the bottom picture represents connection between populations. Different symbols indicate different species or genotypes.

constitute permeable barriers that only reduce gene flow (Treml et al., 2015; Liggins et al., 2020). Isolation-by-environment predicts that gene flow between environmentally divergent habitats is limited and as a result habitats that share environmental characteristics should support populations that are genetically similar (Wang and Bradburd, 2014). Isolation-by-environment patterns are predicted by ecological speciation: limitations in gene flow as a result of divergent selection with an ecological basis (Rundle and Nosil, 2005; Schluter, 2009). Examples can be biased dispersal, selection against migrants and/or reduced hybrid fitness. Lastly, historical contingency predictions are based on the order of colonization (Orsini et al., 2013; Fukami, 2015). These priority effects may have an ecological component

and/or an evolutionary component (De Meester et al., 2016), and are an example of eco-evolutionary dynamics (Legrand et al., 2017). The ecological component states that a simple numerical advantage of the first colonizer may suppress successful establishment of subsequent immigrating individuals. The initial numerical effects may then be fixed via the evolutionary component of local adaptation, enhancing competitive abilities.

The theories differ in how they predict genomic variation to be distributed over spatial and environmental components. Therefore, they aid studies to uncover the eco-evolutionary processes that underlie biogeographic distribution of genomic variation. Still, the caveat remains that defining populations *a priori* is often not possible in seascape genomic studies, potentially leading to circular reasoning. Besides, environmental factors are often spatially and temporally variable, clouding the ultimate causes of divergence (Legendre, 1993; Lee and Mitchell-Olds, 2011). Finding clear, testable marine systems has the potential to overcome this issue.

Insular systems and the Island Biogeography Theory

Islands and island-like systems have a long-standing history in evolutionary and ecological studies since they provide a clearly defined spatial, temporal and environmental context and as such alleviate confounding factors (MacArthur and Wilson, 1967; Warren et al., 2015; Dawson, 2016; Whittaker et al., 2017; Itescu, 2018). Most well-known evolutionary theories are based on studies on 'true' islands: pieces of land surrounded by water (Darwin, 1872; MacArthur and Wilson, 1967; Itescu, 2018). Critical characteristics that define islands and island-like systems are that they are fragmented, their geographic boundaries can be defined and are limited, and between units there is relatively low or no connection, i.e., they are temporally and/or spatially isolated (Itescu, 2018). All these characteristics together make islands 'natural laboratories' that can function as model systems.

The Island Biogeography Theory by MacArthur and Wilson (1967) was formulated as an opportunity for ecologists to advance beyond solely descriptive studies by providing clear, testable predictions on the number of species a system could hold. Under the assumption of neutrality (i.e. species at a given trophic position being equal in birth/death rates and dispersal (Hubbell, 2001)) the theory argues that processes of immigration and extinction on an island should vary predictably with island size and degree of island isolation. Larger islands likely can sustain larger populations, thus decrease the risk of extinction while also being an easier target for immigration (species-area relationship) (MacArthur and Wilson, 1967; Gillespie and Clague, 2009; Warren et al., 2015). Similarly, islands that are closer to the mainland receive a higher immigration load than islands that are further away (species-isolation or species-distance relationship). Higher immigration also influences the rescue effect (Brown and Kodric-Brown, 1977), which lowers extinction risk via an influx of new colonizers and thus genetic variation. The ability for immigration and extinction to reach equilibrium also depends on island age. While inspiring numerous researchers, the original Island Biogeography Theory notoriously lacked an evolutionary component (Whittaker et al.,

2017). However, this component has been added in the form of cladogenesis, or the formation of new species on an island (Warren et al., 2015; Whittaker et al., 2017). It can be expected that more isolated islands allow for more time for speciation to occur before homogenizing effects of gene flow by new immigrants (Feder et al., 2012; Whittaker et al., 2017). Genomic population differentiation lies at the basis of speciation, so it can be expected that the evolutionary component should also include the formation of unique genetic diversity on an island, as well as population differentiation between island and mainland populations. All in all, islands are useful in defining testable predictions on development of species (or genetic diversity (Vellend, 2003)) communities.

The Island Biogeography Theory was formulated using true islands as model systems, yet other island-like systems can vary from true islands in several ways. Where for true islands the separating matrix is water (be it fresh or seawater), for other island-like systems this can be more variable (Itescu, 2018). For example, elevation may pose a barrier for species that is difficult to traverse. Of course, some barriers may be more permeable for some species than for others. The permeability of the surrounding matrix, and thereby the degree of isolation of an island-like system, can influence biotic and abiotic conditions within the system (Dawson, 2016). Furthermore, defining an adequate 'mainland' for island-like systems can be less straightforward as for true islands, since the separating matrix may also contain the species pool that can colonize the system (Itescu, 2018). The origin of islands may also differ from true islands. Where true islands are formed *de novo*, emerging anew in an area, fragmented islands originate by splitting off from already existing habitat (Gillespie and Clague, 2009). Despite their differences, it is still interesting to extend biogeographical studies to include island-like systems as they provide opportunities to investigate consequences of insularity (Warren et al., 2015; Patiño et al., 2017). While there are many equivalents to oceanic islands, such as hydrothermal vents, sky islands, and urban green spaces (Itescu, 2018), attention towards marine island-like systems has only been increasing relatively recently (Dawson, 2016).

Study system: marine lakes

Marine equivalents of island-like systems are anchialine systems: isolated bodies of seawater with varying connection to the surrounding sea (Holthuis, 1973). Examples of anchialine systems are caves, cenotes, pools or lakes (Becking et al., 2011; Weese et al., 2012; Dawson, 2016; Calderón-Gutiérrez et al., 2018). Most population genetic studies in anchialine systems find patterns of strong geographic structuring on small scales, yet most establish distinct genetic lineages potentially indicating cryptic species instead of within-species structure (Supplemental Table 1). Additionally, many anchialine systems, such as caves, represent environments highly dissimilar from the surrounding sea, and as such may limit the generalization of findings on marine processes. Marine lakes are present at sea level and consist of bodies of seawater completely surrounded by land but with subterranean connections to the sea via caves or porous rock (Figure 2) (Holthuis, 1973; Hamner, 1982; Tomascik and Mah, 1994; Hamner and Hamner, 1998; Dawson et al., 2009). Marine lakes

are examples of *de novo* formed insular systems, as rising sea water levels filled natural depressions in karstic islands after the Last Glacial Maximum (Tomascik and Mah, 1994; Dawson et al., 2009). Depending on lake depth, marine lakes are estimated to be around 6,000-12,000 years old (Sathiamurthy and Voris, 2006). The extent of the connection to the sea can be observed through the presence of tidal movements, which are typically delayed and dampened (Dawson et al., 2009; Dawson, 2016). Beyond this physical isolation, lakes also have different local environments, such variable salinity, temperature and dissolved oxygen content, which can add a physiological component to isolation. Currently, around 200 marine lakes are known to science globally, with hotspots of multiple marine lakes in the Indo-Pacific (Dawson et al., 2009).

Indonesia harbors a multitude of clustered marine lakes, in which historical context and climate conditions are similar and lakes draw from the same species pool (Becking et al., 2011, 2015; de Leeuw et al., 2020). As such, marine lakes provide independent replicates of ecology and evolution, and can be used to test the predictions made by various theories on the distribution of (genetic) diversity. Two regions containing hotspots of marine lakes in Indonesia are Berau (East-Kalimantan) and Raja Ampat, (West Papua) (Figure 2). The islands in Berau are part of a Marine Protected Area (MPA) (Becking et al., 2013a). Berau has a tropical rainforest climate with, apart from increasing winds in December and March, no clear distinction between rainy and dry seasons (Tomascik and Mah, 1994; Becking et al., 2013a). Two islands in Berau contain marine lakes that have been studied so far: Maratua and Kakaban. Raja Ampat is located right at the center of the Coral Triangle, famous for its extremely high marine biodiversity (Hoeksema, 2007). Situated on the equator, Raja Ampat has a tropical climate with high yearly precipitation (2,500-4,500mm). Seasonal change is indicated by change of direction of monsoons including persistent winds. Raja Ampat consists of a plethora of karstic islands with complex coastlines (Becking et al., 2011; Mangubhai et al., 2012).

Population genetic studies focusing on these two regions include work on sponges (Becking and Lim, 2009; Becking et al., 2013a, 2013b) and mussels (Becking et al., 2016; de Leeuw et al., 2020). Using single genetic markers, these studies have found clear distinctions between major genetic lineages for both taxa (Becking et al., 2013b, 2016; de Leeuw et al., 2020), potentially representing (cryptic) species. In the case of the mussel (*Brachidontes* sp.), the genetic divergence between lineages was corroborated by shell morphology. Interestingly, each marine lake contained only one of the genetic lineages. Studying within-lineage diversity, for sponge *Suberites diversicolor* little genetic structure was observed (Becking et al., 2013b), whereas the mussels showed intra-lineage structure accompanied by signatures of population bottlenecks (Becking et al., 2016; de Leeuw et al., 2020). The studies suggest stochastic processes such as priority effects accompanied by local adaptation may be at the basis of such short-term divergence on small spatial scales.

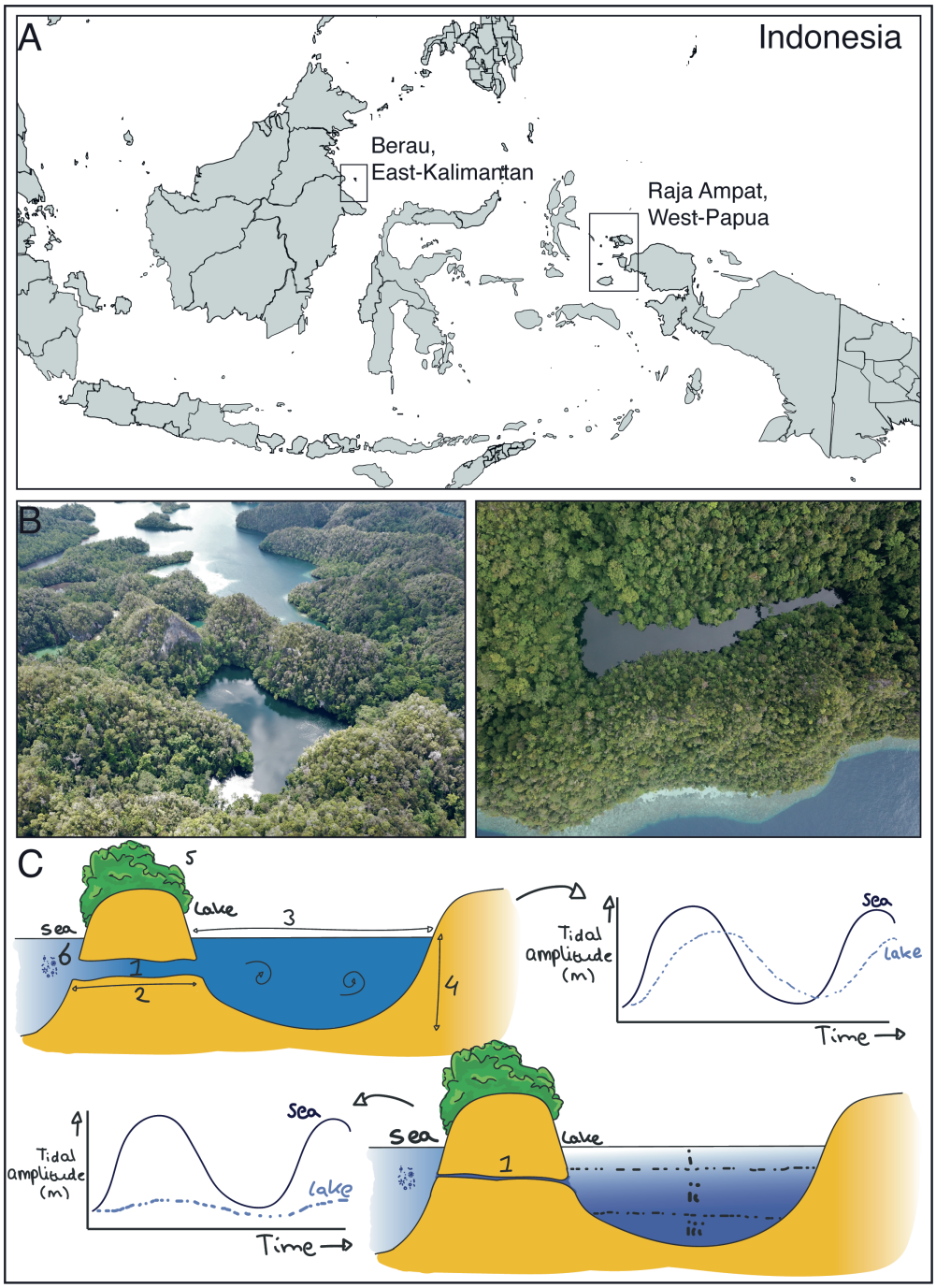


Figure 2: Study area and overview of marine lakes. (A) Map of Indonesia indicating two major study areas of this thesis: Berau in East-Kalimantan and Raja Ampat in West-Papua. (B) Overview pictures of two marine lakes. (C) Schematic overview of two marine lakes differing in degree of isolation and physical and environmental processes that could influence communities within lakes: 1) degree of connection, 2) distance to sea, 3) lake area, 4) lake depth, 5) terrestrial influence, 6) source pool in sea. In the more isolated lake stratification can occur with i) a brackish top-layer, ii) a strong chemocline and iii) potentially an anoxic layer. Two graphs indicating connection through tidal amplitude are displayed.

Knowledge gap and aim of thesis

A remaining question is what processes underlie marine invertebrate genetic structure on small spatial scales (1-100s km). A factor confounding such research in the marine realm is the unknown temporal and spatial variability in gene flow and reproductive success that makes quantifying discrete populations difficult. Marine lakes provide a unique opportunity to delve deeper and increase understanding of how small-scale processes of microevolution can determine population structure. This enables the elucidation of barriers to gene flow and connectivity and supports the prediction of genetic responses to changing environments. Specifically, marine lakes allow for the testing of predictions of genetic patterns both within and between populations. Predictions of genetic diversity within populations follow from the Island Biogeography Theory and include species-area and species-distance relationships, as well as effects of island age and evolution. Predictions of genetic differentiation between populations include expectations of isolation-by-distance, isolation-by-environment, isolation-by-resistance, and historical contingency.

This thesis aims to elucidate small scale marine invertebrate microevolution using clearly defined marine lakes in Indonesia. The four main objectives are:

1. To set environmental and biological baselines in marine lakes of Raja Ampat.
2. To assess how marine lake benthic communities and populations of three invertebrates conform to expectations from Island Biogeography Theory.
3. To test whether marine invertebrate populations conform to the null hypothesis of panmixia or if they show levels of genetic differentiation indicative of evolutionary independent replicates.
4. To quantify the relative importance of neutral (geographic and physical barriers to dispersal) and selective (local environments) factors shaping observed patterns of genetic structure.

The findings could provide insights important to support marine population conservation, as successful management of marine species depends on the understanding of their connectivity (Carr et al., 2017).

Study organisms

This thesis focuses on three marine invertebrates: a jellyfish (*Mastigias papua*, Lesson 1830), a sponge (*Suberites diversicolor*, Becking and Lim 2009) and a mussel (*Brachidontes* sp., Swainson 1840). Golden jellyfish, or *Mastigias papua* (Scyphozoa: Rhizostomae) are famous for their occurrence in the so-called Jellyfish Lake in Palau (local name Ongeim'l Tketau), where tourists can swim together with thousands of these non-stinging jellyfish (Dawson et al., 2001). Similar to corals, *M. papua* specimens can harbor photosynthetic algae in their tissues which provide at least part of their energy demand. Jellyfish have a complicated life cycle starting from benthic polyps (scyphistomae) that can reproduce asexually and depending on water parameters later strobilate into free swimming sexually

reproducing adults (Sugiura, 1965). Originally occurring in the Indo-Pacific Ocean, the jellyfish are known to colonize marine lakes. The colonization events are often accompanied by shifts in morphology, behavior and genetic signatures (Dawson, 2005a; Dawson and Hamner, 2005; Swift et al., 2016). The sponge *Suberites diversicolor* (Porifera: Demospongiae) occurs extensively in marine lakes of Indonesia (Becking et al., 2013a, 2013b). Sponges are an integral but often neglected assets of benthic communities, which with their extensive filtering capacities perform essential roles in benthic-pelagic coupling (Bell, 2008; De Goeij et al., 2013). Sponges can reproduce sexually or asexually via budding or gemmules (Ayling, 1980). Finally, the mussel *Brachidontes* sp. also occurs in high densities in marine lakes and has been investigated for morphological and genetic divergence (Goto et al., 2011; Becking et al., 2016; de Leeuw et al., 2020). Mussels are ecosystem engineers as they create or modify their environment (Jones et al., 1994), providing three-dimensional structure for other species and thus increasing opportunities for diversity (Gutiérrez et al., 2003). *Brachidontes* sp. mussels are broadcast spawners and have a dispersive larval stage that can last up to four weeks (Monteiro-Ribas et al., 2006). Mussels of this genus are known to tolerate wide ranges of temperature and salinity (Sarà et al., 2008), which makes them an excellent first colonizer of new habitats.

Approach

To tackle the objectives, this thesis will use genetic markers (both a traditional single marker and a reduced representation genomic strategy) and assess ecological and morphological aspects of diversity (Figure 3).

Population genetics has a long history of documenting how spatial and environmental factors structure genetic variation (Wright, 1943; Avise, 2000). Traditionally, single genetic markers were used that only cover a small portion of an organisms' genome, such as individual nuclear and mitochondrial genes (Avise, 2000, 2009) (Figure 3A). Through technological and bioinformatic advances, it now becomes increasingly affordable to assess significantly larger portions of the genome (Allendorf et al., 2010; Kelley et al., 2016), giving rise to population genomic approaches. The increase in resolution allows researchers to reassess findings from traditional single marker studies at finer spatial scales (Lemopoulos et al., 2019; D'Aloia et al., 2020; Timm, 2020). For non-model organisms, like most marine organisms, the increasing use of reduced representation genomic methods allows the genotyping of hundreds to hundreds of thousands Single Nucleotide Polymorphisms (SNPs) for population genomic studies (Baird et al., 2008; Peterson et al., 2012; Puritz et al., 2014; Catchen et al., 2017). In reduced representation sequencing approaches parts of a genome are sequenced (0.1-1%, typically), so that it allows the genotyping of the same loci in many (hundreds) individuals (Oleksiak and Rajora, 2020a). Restriction-site associated sequencing (RADseq (Baird et al., 2008; Andrews et al., 2016)) uses enzymes to cut the DNA of multiple individuals at predictable sites and together with shearing and size selection allows for the genotyping of thousands of genetic markers (or loci). Double-digest RADseq (ddRAD (Peterson et al., 2012)) uses two restriction enzymes, thus eliminating the need for random

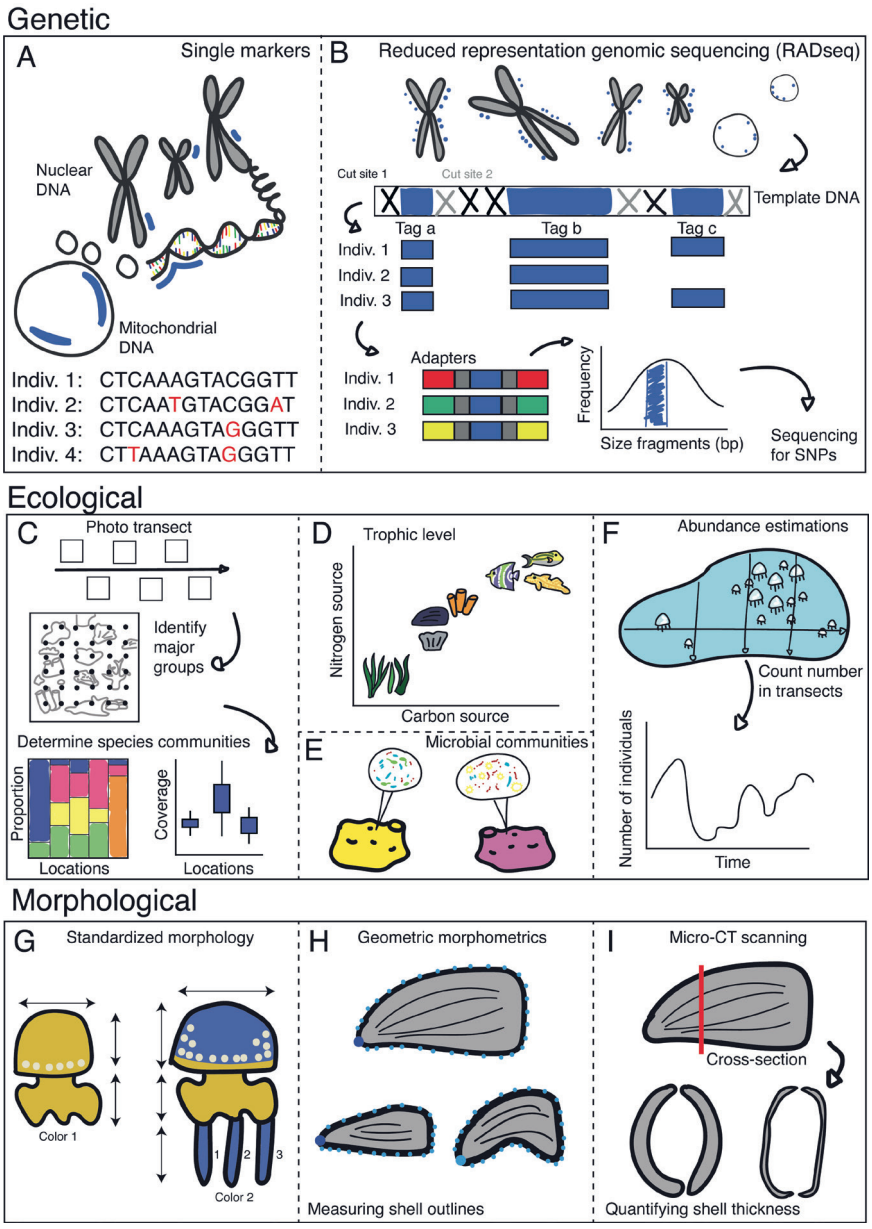


Figure 3: Genetic, ecological and morphological approaches taken by this thesis. (A) Single genetic marker studies sequence known single genes and identify haplotypes with variants. (B) Reduced representation sequencing sequence anonymous variation over the entire genome by using restriction enzymes to cut the DNA, adding unique adapters, size selecting resulting fragments and then sequencing the tags for Single Nucleotide Polymorphisms (SNPs). (C) Benthic species communities assessed via photo transects and identifying major groups that give information on the presence of the groups across different locations. (D) Trophic level determined by stable isotopes of carbon and nitrogen, which give information on the food source and placement in the food web. (E) Microbial communities determined by identifying microbial species and abundances over different host individuals. (F) Individual abundance in lakes counted in standardized transects and repeated over time to search for patterns. (G) Standardized morphological measurements such as length, width, coloration and counts of attributes. (H) Geometric morphometrics assesses differences in shape by the use of (semi)landmarks. (I) Micro-CT scanning allows for an inside view of organisms for example to quantify shell thickness.

shearing and increasing reproducibility over individuals in a cost-effective way (Figure 3B). The method of ddRAD was chosen for this thesis.

The strong genetic focus of this thesis is complemented with ecological and morphological tools. Ecological assessments were done in several ways. Community compositions within lakes were assessed by performing photo-transects to estimate coverage of major benthic groups (e.g., corals, sponges, molluscs, macroalgae, cyanobacterial mats, tunicates, polychaetes and other invertebrate groups) (Figure 3C). Quantifying coverage of benthic communities is important since how organisms adapt to abiotic conditions can be influenced by their biotic environment and the resulting ecological relationships such as competition and/or predation. Next, interactions within the food web and the availability of different food sources can be influenced by local lake environments and can be investigated via stable isotope analysis (Katzenberg, 2000; Nagelkerken et al., 2020) (Figure 3D). This thesis performed stable isotope analysis for the mussel to investigate shifts of trophic position. Furthermore, it increasingly becomes clear that internal (microbial) communities can affect functioning of organisms (Webster and Thomas, 2016; Gonzalez-Zapata et al., 2018) (Figure 3E). Shifting microbial communities may be a first response to local environments, and this thesis assessed associated microbes from mussel gill tissue to infer links between microbial communities, local environments and host genetic structure. Finally, population dynamics, such as the changes in abundance over time, may be influenced by local environments (Figure 3F). For the jellyfish, qualitative and quantitative measurements of changes in population size were done over a range of eleven years. Beyond ecological assessments, individual morphology may also determine how organisms cope with their abiotic and biotic environment. This thesis assessed morphological variation via standardized morphological measurements (Figure 3G), geometric morphometrics (Slice, 2007) (Figure 3H), and via micro-CT scanning (Ngan-Tillard and Huisman, 2017) (Figure 3I). Morphology was determined for the jellyfish and the mussel in this thesis.

At the species level, ecological and/or morphological tools can be combined to define phenotypic diversity. Phenotypes are a good proxy for how the environment affects the organism of study. Phenotypic diversity can follow rapidly from acclimatization to the local environment (Mariani et al., 2012), and in turn can affect distribution of genetic variation in eco-evolutionary dynamics (Miner et al., 2005; Fierst, 2011). An association between genetic and phenotypic divergence can provide compelling evidence for incipient ecological speciation. This thesis explicitly tests for these associations in the mussel.

Thesis outline

Chapter 2 presents a comprehensive overview of marine lakes in Raja Ampat, Indonesia. It aims to set environmental and biological baselines and focusses on the potential of marine lakes to advance ecological and evolutionary theory in the marine realm. Water quality profiling and benthic transects were performed, and diversity assessments were made for

major benthic groups, qualitative estimates of diversity of important groups were done, and presence/absence of key species was indicated.

Chapter 3 studies the evolutionary, ecological and social context of 'jellyfish lakes', marine lakes containing high densities of the golden jellyfish *Mastigias papua*. The study aims to provide a scientific basis for incorporating jellyfish lakes into conservation management plans. The interdisciplinary approach constituted of assessing genetic structure, morphological variation, abundance assessments and stakeholder interviews.

Chapter 4 assesses the importance of neutral and selective isolating processes in structuring populations of the sponge *Suberites diversicolor*. Furthermore, it assesses how an increase in genetic markers affects conclusions made on population genetic structure for the sponge. It aims to compare results from a previous sponge genetic structure based on single markers to the current reduced representation genomic approach.

Chapter 5 relates the genetic population structure of the mussel *Brachidontes* sp. to environmental variables in marine lakes. The aim is to test relative contribution of isolation-by-distance, isolation-by-environment and historical contingency effects in the formation of population genetic patterns.

Chapter 6 brings together *Brachidontes* sp. genomic and phenotypic datasets to explicitly test if we can find signatures of ecological speciation in marine lake populations. It contrasts genetic, morphological, microbial community and trophic niche space variation from ocean and lake locations. It also provides a first insight into potential outlier loci.

Finally, **chapter 7** summarizes the findings from chapter 2-6. It evaluates marine lakes as island-like systems, discusses how patterns observed from marine lakes fit into seascape genomic studies, and considers signatures of incipient speciation. It reflects on marine connectivity and conservation and suggests there is scientific basis to delimit marine lakes as unique management units within conservation management plans. Future perspectives of using marine lakes in eco-evolutionary studies are addressed.

Supplemental Table 1: Review on anchialine system population genetic patterns. Geographic location, anchialine system, species taxon, genetic marker used, genetic pattern, spatial scale at which genetic divergence was observed and the reference is indicated. Abbreviations mtDNA, rDNA and nDNA stand for mitochondrial, ribosomal and nuclear DNA, respectively. COI = cytochrome C oxidase subunit I, COIII = cytochrome C oxidase subunit III, ITS = internal transcribed spacer, cyt b = cytochrome b, H3 = histone H3. IBD = isolation-by-distance.

Location	Anchialine system	Taxon	Marker(s)	Genetic pattern	Spatial scale of divergence	Reference
Bahamas	Cave	Annelid *Pelagomacellicephala illiffei*	mtDNA COI rDNA 16S, 18S	Five divergent genetic groups, no oceanic dispersal.	>115km	(Gonzalez et al., 2017)
Canary Islands	Cave	Lobster *Munidopsis polymorpha*	mtDNA COI 8 microsatellites	Panmixia, low diversity.	-	(Cabezas et al., 2012)
Hawaii	Pool	Atyid hrimp *Holocaridina rubra*	mtDNA COI	Two divergent lineages, significant levels of structure.	>30km	(Santos, 2006)
Hawaii	Pool	Atyid shrimp *Holocaridina rubra*	mtDNA COI	Eight divergent lineages, 13 genetic groups. No exchange between groups.	10km	(Craft et al., 2008)
Hawaii	Pool	Anchialine shrimp *Metabenaeus lohena*	mtDNA COI	Panmixia	-	(Russ et al., 2010)
Hawaii	Pool	Atyid shrimp *Holocaridina rubra*	mtDNA COI	Genetic divergence related to island age	1-100km	(Santos and Weese, 2011)
Indonesia	Lake	Sponge *Suberites diversicolor*	mtDNA COI nDNA ITS	Two divergent lineages.	<10km	(Becking et al., 2013b)
Indonesia	Lake	Mussel *Brachidontes* sp.	mtDNA COI rDNA 18S, 28S	Four divergent lineages.	2-6km	(Becking et al., 2016)
Indonesia	Lake	Mussel *Brachidontes* sp.	mtDNA COI rDNA 18S, 28S	Six divergent lineages with IBD pattern.	<5km	(de Leeuw et al., 2020)
Japan	Pool	Atyid shrimp *Caridina rubella*	mtDNA COI	Genetic structure, two divergent lineages.	<20m to >10km	(Weese et al., 2012)
Japan	Pool	Caridean shrimp *Antecaridina lauensis, Halocaridines trigonophthalma, Metabaeus minutus*	mtDNA COI rDNA 16S	2-3 divergent lineages per species.	>200km	(Weese et al., 2013)
Mexico	Cave	Shrimp *Creaseria morleyi*	mtDNA COI rDNA 16S	Two divergent lineages with IBD.	10-100km	(Botello and Alvarez, 2010)
Palau	Lake	Jellyfish *Mastigias papua*	mtDNA COI	High genetic structure among lake	1-50km	(Dawson and Hamner, 2005)
Palau	Lake	Fish *Sphaeramia orbicularis*	mtDNA control region	Divergence between lakes	1-50km	(Gotoh et al., 2009)
Palau	Lake	Mussel *Brachidontes* sp.	mtDNA COI rDNA 18S	Divergence between morphological groups.	1-50km	(Goto et al., 2011)
Palau	Lake	Fish *Atherinomorus endrachtensis*	mtDNA cyt b, control region	High genetic structure among lakes.	1-50km	(Gotoh et al., 2011)
Palau	Lake	Jellyfish *Mastigias papua*	mtDNA COI, COIII rDNA 16S, H3, 28S	High genetic structure among lakes.	1-50km	(Swift et al., 2016)
Philippines	Cave	Gastropod *Neritilia cavernicola*	mtDNA COI	Panmixia.	-	(Kano and Kase, 2004)
Spain	Cave	Amphipod *Metacrangonyx longipes*	mtDNA COI rDNA 16S	Five divergent lineages.	>10km	(Bauzà-Ribot et al., 2011)

Chapter 2

Islands of sea: Biophysical assessments of marine lakes in Raja Ampat

Diede L. Maas, Christiaan A. de Leeuw, Awaludinnoer Ahmad, Ludi. P. Aji, Agustin Capriati, Purwanto, Leontine E. Becking

Submitted

Abstract

Islands are ideal model systems for ecological and evolutionary studies and have been essential in the development of biogeographical theories. Marine lakes may present an analogue of island systems in the marine realm, with marine water surrounded by a terrestrial matrix but with a subterranean connection to the sea. Raja Ampat in Indonesia has a vast karstic landscape and is a hotspot for marine lakes. It is part of the global center of marine biodiversity. We aimed to discover marine lakes in Raja Ampat, describe their biophysical characteristics, and assess how marine lakes conform to expectations of the Island Biogeography Theory. We described 32 marine lakes, 17 new to science, where we measured environmental and physical parameters (temperature, salinity, pH and dissolved oxygen, lake depth, surface area, perimeter, shortest distance to sea, and proxies for terrestrial influence and connection to the surrounding sea) and assessed diversity indicators of benthic community structure based on photo-transects, presence-absence and qualitative categories of target species groups. The average temperature per lake ranged from 30.0° to 36.8°C. Connection to the surrounding sea also differed per lake: the relative tidal amplitude ranged from 5% to 89% compared to the sea. The biological diversity differed significantly among lakes. Corals were present in lakes up until 32°C. Based on the biophysical characteristics, three categories of lakes were identified: lagoon-like lakes, highly isolated lakes with distinct local environments and depauperate communities, and intermediate lakes. Temperature was significantly related to the variation in composition and cover of benthic groups ($F(1,12)=3.06$, $p = 0.005$). Furthermore, connection to the surrounding sea ($R^2 = 0.52$, $p = 0.005$) and temperature ($R^2 = 0.39$, $p = 0.02$) were found to be important predictors of qualitative diversity distributions. Richness of major benthic groups significantly increased with increasing connection to the sea and decreasing water temperatures. We found indications for species-isolation relationships, but not for species-area relationships. Our study sets environmental and biological baselines and emphasizes the utility of marine lakes in Raja Ampat in eco-evolutionary studies, and their value as a collection of unique microhabitats. We recommend all marine lakes should be incorporated into specific conservation management plans to protect the pristine nature of these 'islands of sea'.

Keywords: Marine lakes, anchialine ecosystems, benthic communities, Island Biogeography

Introduction

The utility of island systems to advance ecological and evolutionary theories has been recognized since Darwin (Darwin, 1872). In the 1960s, this utility was further explored by MacArthur and Wilson when they formulated their Island Biogeography Theory (MacArthur and Wilson, 1963, 1967). They proposed that on an island, immigration and extinction maintain a dynamic equilibrium depending on island characteristics such as size or distance from the mainland. Recently, Island Biogeography Theory has undergone a renaissance with the rise of big data analyses (Helmus et al., 2014; Fernandez-Palacios et al., 2015; Warren et al., 2015; Santos et al., 2016; Graham et al., 2017; Whittaker et al., 2017). Islands have discrete boundaries and, when found in an archipelago context, can serve as independent replicates of ecology and evolution. Therefore, islands are suitable entities to test hypotheses such as species-area and species-isolation relationships, and the role of geographic and environmental factors in the assembly and differentiation of populations and communities (Warren et al., 2015).

Beyond oceanic islands, island-like systems include freshwater lakes (Arnott, 2009), sky islands (Wiens et al., 2019), and habitat islands (Fattorini et al., 2018), among others (Itescu, 2018). Classically, marine equivalents to islands are thought to be lacking due to the seemingly boundless nature of marine systems (Hedgecock, 1986; Palumbi, 1994; Bowen et al., 2013). As a result, studies into how biogeographic processes influence marine communities have been lagging behind terrestrial counterparts (Selkoe et al., 2016; but see Hachich et al., 2015; Pinheiro et al., 2017). Anchialine systems (*sensu* Holthuis, 1973) are salt to brackish water bodies which fluctuate with the tide but without a surface connection to the sea. Recently, anchialine systems such as caves and cenotes are receiving increased attention from studies investigating patterns of community richness and speciation (Weese et al., 2012; Mercado-Salas et al., 2013; Gonzalez et al., 2017; Calderón-Gutiérrez et al., 2018). They generally found communities and populations distinct from oceanic counterparts. Subterranean systems such as anchialine caves, however, represent environments highly dissimilar from surface habitats, with limited to no light, and therefore may not reflect general marine processes. In contrast, marine lakes, anchialine systems open to the air (Holthuis, 1973; Dawson et al., 2009), have been suggested to be suitable marine insular systems to serve as an equivalent to terrestrial islands, being 'islands of sea' (Dawson and Hamner, 2005; Becking et al., 2016; Dawson, 2016; Maas et al., 2018; de Leeuw et al., 2020).

Marine lakes are ideal systems to study ecological and evolutionary dynamics of marine organisms. Marine lakes are completely surrounded by a terrestrial matrix, but maintain a subterranean connection to the surrounding sea (Holthuis, 1973; Dawson et al., 2009). Most lakes originated *de novo* around 12,000 years ago, when seawater levels rose after the Last Glacial Maximum (Tomascik and Mah, 1994; Sathiamurthy and Voris, 2006). Depressions in karstic rock filled with seawater, entering through subterranean fissures, channels and caves. The nature of these underground connections varies per lake, and determines the

present-day connection to the sea, the source of potential immigrants. Approximately 200 marine lakes are known worldwide, with notable aggregations in the Caribbean, Palau and Indonesia (Hamner and Hamner, 1998; Dawson et al., 2009; Becking et al., 2011, 2015). Studies investigating marine lakes have found high levels of endemism and population differentiation (Gotoh et al., 2009, 2011; Goto et al., 2011; Becking et al., 2013b, 2016; Swift et al., 2016; Maas et al., 2018), and even suggest incipient speciation (Dawson and Hamner, 2005). Becking et al. (2011) discovered a large assembly of marine lakes in Raja Ampat, Indonesia, providing many replicated 'natural experiments' of community assembly and evolution on a relatively small spatial scale.

Raja Ampat is situated in the Bird's Head Seascape in West Papua, part of the Coral Triangle: the global center of tropical marine biodiversity (Hoeksema, 2007; Mangubhai et al., 2012). The Coral Triangle contains over 75% of reef-building coral species, and species richness peaks here for multiple restricted-range taxa (Roberts et al., 2002; Asaad et al., 2018). The underlying processes giving rise to this unparalleled marine biodiversity are still under debate, but proposed mechanisms include glacial cycles of sea level change, complex oceanic currents, and habitat heterogeneity (Hoeksema, 2007). Insular systems such as marine lakes may also contribute to the diversity by generating species and/or genetic diversity, which would support the hypothesis of the Coral Triangle being a 'Center of Origin' (Tornabene et al., 2015). So far, fifteen marine lakes in Raja Ampat have been reported in scientific literature, ranging from descriptive studies (Becking et al., 2011; Purba et al., 2018), to population genetic studies (Becking et al., 2013b, 2016; Maas et al., 2018, 2020; de Leeuw et al., 2020). These studies show that marine lakes harbor unique populations and communities distinct from each other and from the surrounding reefs. However, the marine lakes of Raja Ampat are currently not explicitly included in the Marine Protected Areas (MPAs) encompassing them (Agostini et al., 2012; Maas et al., 2020). While all marine lakes are geographically located in one of the nine MPAs in Raja Ampat, only one management plan explicitly mentions 'marine lake subzones' in Fakfak (DKP Papua Barat., 2018). In order to facilitate the incorporation of marine lakes into conservation management plans, thorough environmental and biological baselines need to be set.

Here, we present a comprehensive overview of marine lakes in the Raja Ampat region and demonstrate their potential to study ecological and evolutionary processes in marine systems. From 2016 to 2018, we environmentally profiled marine lakes and performed benthic coverage and diversity assessments to set environmental and biological baselines. Using a benthic community dataset, we investigated how the marine lake setting conforms to expectations arising from Island Biogeography Theory (MacArthur and Wilson, 1967; Warren et al., 2015). We tested which geographic, environmental and physical parameters best explained the variation in benthic coverage and diversity. We assumed that lakes with environmental conditions dissimilar from the sea, i.e., higher temperature and lower salinity, would have depauperate communities due to environmental filtering. We also tested whether lake area and distance to the sea was related to diversity and richness, following the

hypotheses of the species-area and species-isolation relationships. Ultimately, this study aims to indicate the uniqueness of marine lakes and to provide baselines for conservation purposes.

Materials and methods

Study area and localization marine lakes

Raja Ampat, West Papua is located within the center of the Coral Triangle and has been designated as a priority area for conservation (Mangubhai et al., 2012; Asaad et al., 2018). Sea surface temperatures in Raja Ampat average at 29°C, and yearly precipitation ranges from 2,500-4,500mm. Monsoons with strong winds from the northwest in November-March and from the southeast in May-October are indicators of seasonal change. Previous studies discovered several anchialine systems in Raja Ampat (Becking et al., 2011, 2016; Swift et al., 2016; Maas et al., 2018, 2020). We have included previously discovered marine lakes (15 lakes in total) and located 17 additional lakes via Google Earth satellite images, and by flying a Drifter water-airplane around the area (Fig. 1).

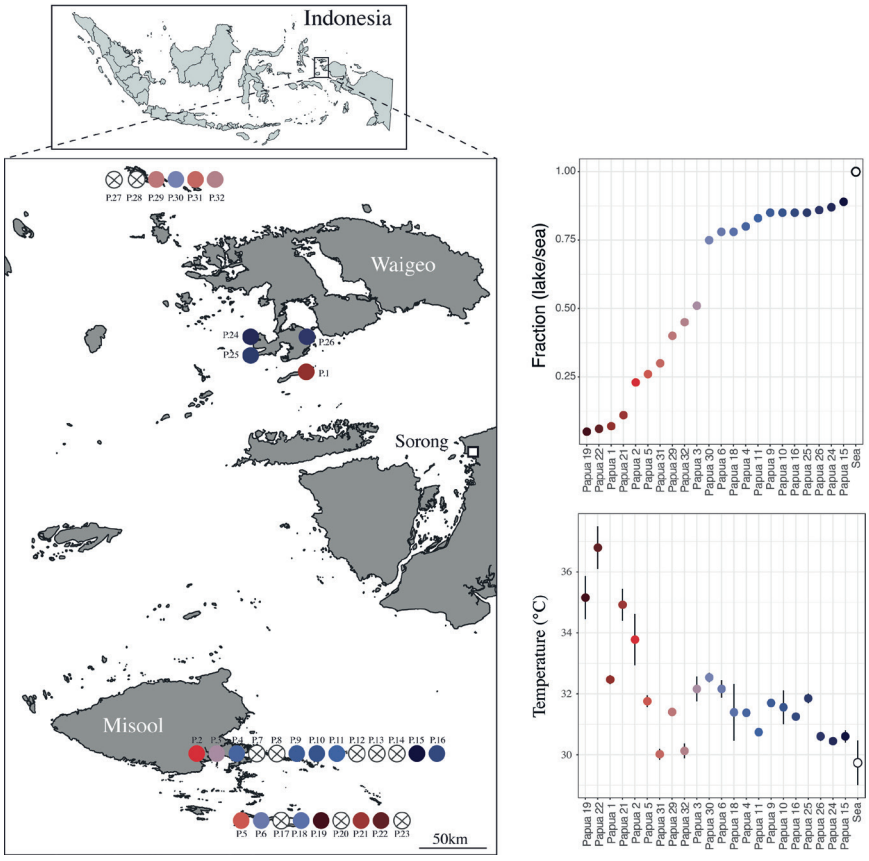


Figure 1: Overview of marine lake locations in Raja Ampat, West Papua, Indonesia. Color gradient reflects gradient in connection to the surrounding sea. Fraction (difference in maximal tidal amplitude of the lake divided by the sea) is displayed for all lakes. Average lake temperature is also displayed for all lakes, ordered by degree of connection. Lakes with a cross were not environmentally profiled.

Lake characterization

We visited marine lakes for three consecutive years (May and November 2016, May 2017, and May 2018), and performed environmental (Table 1A) and biological (Table 1B) profiling. Lake coordinates were logged using a Global Positioning System (Garmin GPSmap 64S). To measure the degree of connection of each lake to the sea we logged tidal amplitude within the lake and the directly adjacent sea using Hobo water-level loggers. Degree of connection was defined as a fraction resulting from dividing the maximum tidal amplitude (m) of the lakes by that of the sea, providing a theoretical minimum of 0 and maximum of 1. Lakes were designated as being stratified (with a clear thermo- and/or halocline) or non-stratified (where water was well-mixed). Depth was measured using a handheld sonar system (H22PX) at approximately 100 points per lake taken each 1-5m. Lake area (m^2) and perimeter (m) were calculated via Google Earth, and their ratio was used as a proxy for the terrestrial influence of the surrounding forests and/or mangroves. We expect lakes with a higher perimeter to surface ratio to be influenced more by the influx of terrestrial nutrients, which may impact food web dynamics and thus the species community as a whole. We also indicated whether or not mangrove trees were present along the perimeter of the lake. Distance to sea (m) was measured as the shortest straight-line distance from the lake outline to the adjacent sea. Using a YSI Professional Plus multimeter, we logged temperature (°C), salinity (ppt), pH and dissolved oxygen (mg/L) at 1m intervals from the surface to the sediment. Within each lake at least 10 sites were measured.

Biological assessments

Two assessments of biological diversity were made, quantitative coverage assessments of benthic groups and quick-scan diversity assessments based on categories of target groups and presence/absence of key species. In May of 2016, we performed horizontal photo-transects in 16 lakes along a 15m transect line placed at 1m below the low tide mark to quantify coverage of benthic communities. Three photo transects were performed at evenly spaced locations around the lake perimeter. Every $25cm^2$ above and below the transect line was photographed without overlap, resulting in 120 photos per transect and 360 photos per lake. A subset of 180 photos per lake were analyzed. Major benthic groups were scored using a Coral Point Count with Excel extension (CPCe v4.1) (Kohler and Gill, 2006), including ten biotic groups and substrate. We included corals, bivalves, gastropods, sponges, benthic cyanobacteria, macroalgae, crustose coralline algae, polychaetes, tunicates and echinoderms as benthic biota. A uniform grid was placed with 25 points per photo, resulting in 72,000 identified points in total for all 16 lakes, representing coverage of the major benthic groups. Coverage was summarized in boxplots.

We qualitatively indicated diversity of sponges and molluscs in classes since these groups contribute the highest diversity and biomass in marine lakes (Table 1B). We indicated five classes of species diversity from 'Very low' (sponges <5, molluscs <3 species), 'Low' (sponges 5-10, molluscs 3-5 species), 'Medium' (sponges 10-15, molluscs 6-10 species), 'High' (sponges 15-20, molluscs 11-15 species), and 'Very high' (sponges >20, molluscs >15

species). We also scored the diversity and size of (reef) fish in three classes: 0-3 species <10cm (only gobies and silversides), 3-10 species <30cm, >10 species with no size limit. Furthermore, we indicated the presence/absence of several characteristic groups: (hard) coral, dominant macroalgae (including the species where possible), two sponges (*Darwinella* aff. *gardineri* (Topsent, 1905) and *Suberites diversicolor* (Becking and Lim, 2009)), three jellyfish (*Mastigias papua* (Lesson, 1830), *Aurelia* sp. (Linneaus, 1758), and *Cassiopea* sp. (Haeckel, 1880)), two shrimp species (*Antecardina lauensis* (Edmondson, 1935), *Parhippolyte uvaea* (Borradaile, 1900)), reef fish and gobies (Gobiidae). Finally, we indicated if a disturbance by human activity was seen.

Statistical analyses

To infer variation in the distribution of biota in marine lakes we ran multivariate analyses. For the coverage of major benthic groups, we ran a canonical correspondence analysis (CCA) on Hellinger-transformed average coverage of the three transects performed per lake. We displayed all transects per lake separately in a non-metric multidimensional scaling analysis (nMDS) implemented by the vegan package in R (Oksanen et al., 2016). For the qualitative estimations of the diversity of molluscs, sponges and fish communities we ran an nMDS, and used the function *envfit* implemented in vegan to fit environmental factors.

We tested the effects of environmental, physical and geographic factors on coverage and diversity of benthic groups via Mantel tests (Mantel, 1967; Slatkin, 1993). For this purpose, we computed distance matrices for the benthic coverage data (Bray-Curtis dissimilarities) and qualitative diversity estimates (Euclidean distances). We related these two measures of biotic distance to geographic distance and environmental distance. Geographic distance was calculated as the minimum pairwise distance in meters between coordinates from the center of each lake using the function *distm* from the geosphere R package. We defined the environment by taking averages of temperature, salinity, pH per lake at 1m depth, since this was the depth of the transects. The variation in water quality variables was decomposed using a Principal Component Analysis (PCA) including temperature, salinity, pH and proxy terrestrial influence. To define environmental distance the first two PCA axes were used to calculate Euclidean distances between lakes (together explaining 82.1% of total environmental variation). Environmental distances were also calculated separately for temperature, salinity and pH measurements.

Additionally, diversity indices were calculated for the major benthic groups. Shannon diversity was calculated with the *diversity* function in the *vegan* R package. We used the number of benthic groups present in each lake as a measure of richness. To test assumptions of Island Biogeography Theory, Shannon diversity and richness measures were correlated to lake area, distance to sea, and depth, as well as to the water quality parameters.

All statistical analyses were performed in R v.3.5.0 (R Core Team, 2018). Multi-dimensional scaling was performed using the *vegan* package for nMDS, CCA and *envfit* analyses and the *stats* package for PCA analyses. Mantel tests for correlation between distance matrices were performed using the *vegan* package with 999 permutations. Since the data was not normally distributed and variances were not homogeneous, Spearman correlations were performed. Significance of results were indicated with an alpha value of < 0.05, unless stated otherwise. The package *ggplot2* was used for data visualization.

Results

In total 32 marine lakes were located during this study (Supplemental Fig. 1), of which 17 were new to science. Environmental and biotic surveys were done for a subset of 22 lakes (Fig. 1, Table 1A and B). Most lakes are located in southeast Misool (22 lakes), with others in Gam (4) and Wayag (6) (Fig. 1). Six lakes (Papua 1, Papua 2, Papua 4, Papua 5, Papua 10 and Papua 26) were impacted by anthropogenic activity in the form of jellyfish tourism (Maas et al., 2020), proximity to a village, or aquaculture. All other lakes were considered undisturbed.

Environmental profiles and description of lakes

We made depth profiles of water quality parameters for the subset lakes that could be environmentally profiled (Supplemental Fig. 2). Based on averages from 1-5m depth we observed a temperature gradient of 30.0°C (±0.18°C, Papua 31) to 36.8°C (±0.70°C, Papua 22), and connection values ranging from 0.05 - 0.89 (Fig. 1, Table 1A). Three lakes, Papua 19, 21 and 22, showed extremely high temperatures and low salinities compared to the sea, and low connection to the sea (Table 1A). Spearman tests showed correlations for some combinations of environmental and physical lake parameters (Supplemental Fig. 3). Most notably, temperature and salinity (rho = -0.79, $p < 0.0001$), temperature and connection to the sea (fraction, rho = -0.67, $p = 0.001$*), and salinity and connection to the sea (rho = 0.75, $p = 0.0001$*) were highly correlated. Even when excluding the three extreme lakes these associations remained (rho = -0.71, -0.55 and 0.67, and $p = 0.001$*, 0.018* and 0.003*, respectively).

We will describe in brief four representative lakes, one from the extreme cluster (Papua 22), one more lagoon-like lake (Papua 11), and two intermediate lakes (Papua 1 and Papua 32) (Fig. 2). Descriptions of all lakes can be found in Supplemental Information A.

Table 1A: Physical and environmental characteristics of marine lakes in Raja Ampat, Indonesia. Ordered by connection, first for lakes that were environmentally profiled, then for other lakes.

Location	Island	Connection to the sea (fraction)	Stratification	Depth (m)	Area (m^2)	Circum-ference (m)	Proxy terrestrial influence	Mangrove	Distance to sea (m)	Temperature (°C)	Salinity (ppt)	pH	Dissolved oxygen (mg/L)
Sea										30.4	30.1	8.3	5.8
Papua 15	North Misool	High (0.89)	Not stratified	33.9	10300	410	0.0398	No	98	30.5	30.2	8	3.14
Papua 24	Gam	High (0.87)	Not stratified	6	4200	260	0.0619	No	55	30.3	30.7	7.9	2.96
Papua 26	Gam	High (0.86)	Not stratified	4.5	16700	734	0.0440	Yes	80	30.4	29.9	8	3.12
Papua 25	Gam	High (0.85)	Not stratified	8.3	21500	790	0.0367	Yes	75	31.7	29.6	7.8	2.88
Papua 16	North Misool	High (0.85)	Not stratified	19	21100	867	0.0411	No	37	31.2	30	8.1	4.34
Papua10	North Misool	High (0.85)	Not stratified	3	15000	580	0.0387	Yes	88	30.9	28.1	7.8	-
Papua 9	North Misool	High (0.85)	Not stratified	-	33600	735	0.0219	-	111	31.7	25.5	7.8	4.88
Papua 11	North Misool	High (0.83)	Not stratified	8.9	27300	1090	0.0399	No	25	30.7	28.3	8.1	3.24
Papua 4	North Misool	High (0.8)	Stratified	20.4	13750	612	0.0445	No	55	31.7	25.9	175	5.15
Papua 18	North Misool	Medium (0.78)	Not stratified	4.6	7000	550	0.0786	Yes	101	31.5	28.4	7.76	1.34
Papua 6	South Misool	Medium (0.78)	Stratified	12.4	2950	206	0.0698	No	28	31.9	28.3	7.9	3.05
Papua 30	Wayag	Medium (0.75)	Not stratified	4.1	13000	470	0.0362	Yes	35	32.4	28.9	7.6	3.00
Papua 3	North Misool	Medium (0.51)	Stratified	7.5	20800	760	0.0365	Yes	145	32.6	30.8	7.9	4.34
Papua 32	Wayag	Medium (0.45)	Not stratified	5.5	6100	316	0.0518	Yes	75	31.2	30.7	7.3	1.69
Papua 29	Wayag	Medium (0.4)	Not stratified	2	15700	693	0.0441	Yes	50	31.4	29.5	7.5	1.99
Papua 31	Wayag	Low (0.3)	Stratified	7.1	6200	274	0.0442	Yes	275	30.1	29.2	7.2	0.63
Papua 5	South Misool	Low (0.26)	Not stratified	4.8	3700	300	0.0811	No	75	31.5	28.9	8.1	3.32
Papua 2	North Misool	Low (0.23)	Stratified	7.3	12200	600	0.0492	Yes	87	33.6	25.2	7.8	2.63
Papua 21	South Misool	Low (0.11)	Stratified	13	18950	846	0.0446	Yes	80	35.9	23.9	7.9	3.25
Papua 1	Gam	Low (0.07)	Stratified	19	88500	1666	0.0188	Yes	128	32.3	24	7.6	3.10
Papua 22	North Misool	Low (0.06)	Stratified	12.3	23100	729	0.0316	No	83	35.6	16.3	8.2	4.41
Papua 19	North Misool	Low (0.05)	Stratified	21.5	61050	1125	0.0184	No	90	35.4	19.6	8.2	4.54
Papua 12	North Misool	High	Not stratified	-	7500	463	0.0617	-	55	-	-	-	-
Papua 13	North Misool	High	Not stratified	4	1400	174	0.1243	-	75	-	31.3	7.4	-
Papua 14	North Misool	High	Not stratified	10	3900	267	0.0685	No	48	-	-	-	-
Papua 17	South Misool	High	Not stratified	-	6500	314	0.0483	Yes	20	31.7	31.7	7.2	4.02
Papua 27	Wayag	Medium	Not stratified	2	22000	615	0.0280	Yes	270	29.5	31	7.5	-
Papua 7	North Misool	Low	Not stratified	-	9700	529	0.0545	No	175	-	14.9	7.2	-
Papua 8	North Misool	Low	Not stratified	2.4	2100	449	0.2138	No	40	-	-	-	-
Papua 20	South Misool	Low	Stratified	-	22400	660	0.0295	No	260	35	-	-	-
Papua 23	South Misool	Low	Not stratified	10	6050	348	0.0689	No	56	30	-	-	-
Papua 28	Wayag	Low	Not stratified	1.5	2050	207	0.1010	Yes	100	-	-	-	-

Table 1B: Biological characteristics of marine lakes in Raja Ampat, Indonesia. Ordered by connection, first for lakes that were environmentally profiled, then for other lakes. Abbreviated species names: Suberites diversicolor, Mastigias papua, Parhippolyte uvaea, Antecardina lauensis.

Loc.	Sponge diversity	Mollusc diversity	Fish class	Hard coral	Dominant algae	*S. diversicolor*	*Darwinella sp.*	*M. papua*	*Aurelia sp.*	*Cassiopea sp.*	*P. uvaea*	*A. lauensis*	Reef fish	Gobies	Human influence
P 15	High	Very high	Large pops	Yes	None	No	No	No	No	No	No	No	Yes	Yes	Undisturbed
P 24	High	High	Large pops	Yes	All	No	No	No	No	No	No	No	Yes	No	Undisturbed
P 26	Very high	High	Large pops	Yes	None	No	No	No	No	No	No	No	Yes	No	Fish cages
P 25	Very high	Very High	Large pops	Yes	None	No	No	No	No	No	No	No	Yes	No	Undisturbed
P 16	Very high	Very high	Large pops	Yes	None	No	No	No	No	No	No	No	Yes	Yes	Undisturbed
P10	Medium	-	Large pops	No	Halimeda	No	No	No	No	Yes	No	No	Yes	Yes	Fish cages
P 9	-	-	-	-	-	-	-	Yes	Yes	-	-	-	-	-	Undisturbed
P 11	Very high	Very high	Large pops	Yes	None	No	No	No	No	No	No	No	Yes	Yes	Undisturbed
P 4	Medium	Medium	Small pops	No	Cladophora	Yes	No	Yes	Yes	No	No	No	No	Yes	Tourism, aquaculture
P 18	Low	Medium	Large pops	Yes	None	No	No	No	No	Yes	No	No	Few	Yes	Undisturbed
P 6	Medium	Medium	Medium pops	Yes	Cladophora	No	Yes	Yes	Yes	No	No	No	Few	Yes	Fish cages
P 30	-	-	Small pops	No	Caulerpa	Yes	No	Yes	No	Yes	Yes	Yes	No	No	Undisturbed
P 3	High	Medium	Large pops	No	Halimeda	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Undisturbed
P 32	-	-	-	No	All	Yes	Yes	No	No	No	Yes	No	No	Yes	Undisturbed
P 29	-	-	-	No	Caulerpa	No	Yes	No	No	No	Yes	No	No	Yes	Undisturbed
P 31	-	-	-	No	-	No	No	No	No	No	Yes	Yes	No	No	Undisturbed
P 5	Medium	High	Medium pops	No	Caulerpa	Yes	Yes	Yes	No	No	No	No	Few	Yes	Tourism, aquaculture
P 2	Very high	Medium	Medium pops	No	Halimeda	No	Yes	Yes	Yes	Yes	No	No	Few	Yes	Tourism, fish cages
P 21	Low	Very low	Medium pops	No	None	No	No	No	No	No	No	No	No	Yes	Undisturbed
P 1	Medium	Low	Small pops	No	None	Yes	Yes	No	No	No	Yes	No	No	Yes	Introduced turtles, village
P 22	Very low	Very low	Small pops	No	None	No	No	No	No	No	No	No	No	Yes	Undisturbed
P 19	Very low	Very low	Small pops	No	None	No	No	No	No	No	No	No	No	Yes	Undisturbed
P 12	Very high	-	-	-	-	No	No	No	No	No	No	No	-	-	Undisturbed
P 13	Very high	-	-	Yes	-	No	No	No	No	No	No	No	Yes	-	Undisturbed
P 14	-	-	-	Yes	None	No	No	No	No	No	No	No	Yes	Yes	Undisturbed
P 17	Very high	-	-	Yes	Cladophora	No	No	No	No	No	No	No	-	-	Undisturbed
P 27	Low	-	-	No	None	Yes	No	No	No	No	No	No	Yes	Yes	Undisturbed
P 7	Very low	-	-	No	None	No	No	No	No	No	No	No	-	-	Undisturbed
P 8	Very low	Very low	-	No	None	No	No	No	No	No	No	No	-	-	Undisturbed
P 20	Very low	Very low	-	No	None	No	No	No	No	No	No	No	No	Yes	Undisturbed
P 23	Very low	Low	-	No	Cladophora	No	No	No	No	No	No	Yes	No	No	Undisturbed
P 28	Very low	-	-	No	None	No	No	No	No	No	Yes	Yes	No	No	Undisturbed

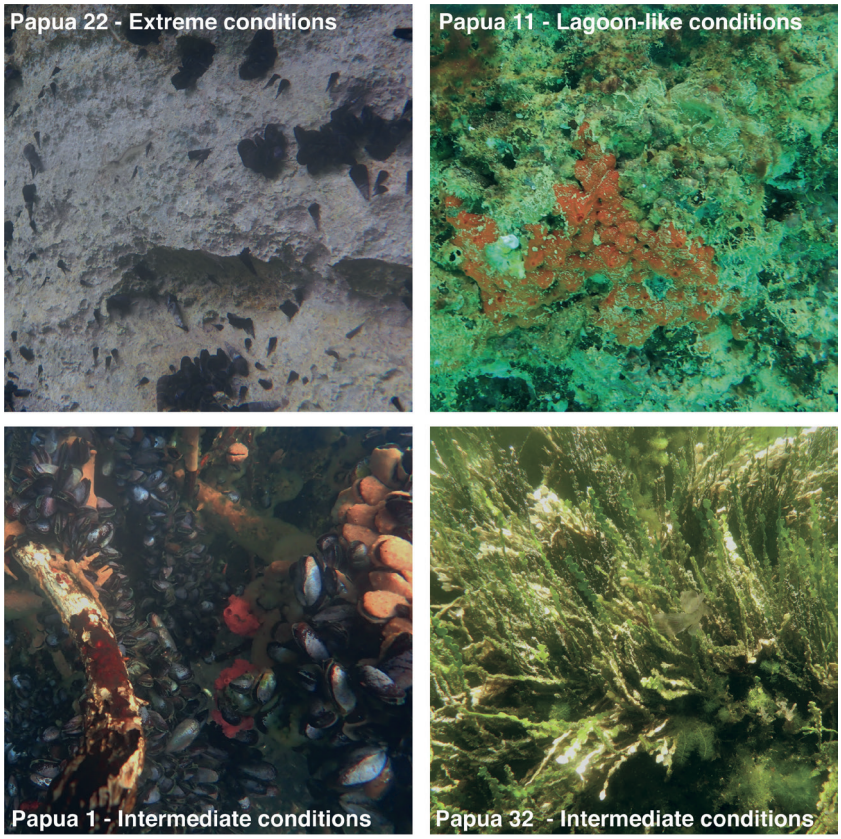


Figure 2: Representative photos of four marine lakes. Papua 22 as a highly isolated lake, Papua 11 as a lagoon-like lake, and Papua 1 and 32 as intermediate lakes.

Papua 22 is a highly stratified lake located on the island of Warakaraket in the archipelago of southeast Misool (Fig. 1). The marine lake is 12.5m deep, 250m long, 170m wide, has a surface area of 23,000m^2 and the lake circumference is 730m. The closest distance to the surrounding sea is 83m. Tidal influence is very limited in the lake, with tidal amplitude in the lake being only one-tenth of the tidal amplitude of the surrounding sea. No apparent tunnels or caves are visible. A shallow pool is partially separated from the main basin by an outcrop of karstic rock in the north of the lake. Papua 22 is the warmest known marine lake in Raja Ampat, with a recorded maximum of 42°C. Water column characteristics include a strong halocline at approximately 1.5m depth (Supplemental Fig. 2). A colder (31.5°C), brackish (salinity: 12ppt) water layer of about 1.5m lays on top of a warmer (39°C), more saline (19ppt) layer. Temperature decreases slightly to 36°C at the bottom of the lake, and salinity increases slightly to 22ppt. The lake is likely heliothermal, resulting in extreme temperature differences between water layers. In heliothermal lakes, the low density surface layer acts as a one-way mirror to solar radiation, preventing convection of heated higher salinity water from deeper layers to the surface and thus preventing the subsequent cooling via evaporation (Sonnenfeld and Hudec, 1980). The whole lake is oxygenated, ranging from 6.3mg/L at the surface, to 4.5mg/L at the bottom. Two benthic gobiidae species are present

(*Exyrias puntang* (Bleeker, 1851) and *Acentrogobius janthinopterus* (Bleeker, 1852)), as well as one free swimming species (*Mugilogobius* sp. (Smitt, 1900)). None of the fish were observed to venture below the strong thermocline. The benthic community is highly depauperate, composed of 2 sponge species, 1 gastropod and 2 bivalves (*Brachidontes* sp. (Swainson, 1840) and one endobenthic *Bastissa* species (Prime, 1862)). All molluscs found below the thermocline were dead, except the gastropod which seemed to tolerate the extreme temperatures.

Papua 11 is a highly mixed, lagoon-like lake located near the main island of Misool (Fig. 1). The lake is 8.9m deep, 540m long, 150-350m wide, has a surface area of 27,300m^2, and a perimeter of 1100m. The closest distance to the sea is 25m. There is a high connection to the surrounding sea via a large cave located in the north of the lake, and tidal amplitudes are only 20% less than the surrounding sea. The lake contains some small islands of karstic rock and has a large terrestrial cave at the surface in the eastern part. Water characteristics remain very stable throughout the water column (Supplemental Fig. 2). Temperature ranges from 30.1-30.6°C from surface to the bottom, oxygen from 4.1mg/L to 4.2mg/L, with only salinity being notably lower at the surface (25.4ppt) compared with the bottom (28.7ppt). At the time of survey, assemblages of reef fish were present and as well as a single green turtle (*Chelonia mydas* (Brongniart, 1800)). Likely, these large organisms are able to pass through the northern cave to the surrounding lagoon. The benthic community resembles that of nearby reef flats, harboring hard corals and high sponge and mollusc diversity.

Papua 1 is an arrowhead-shaped stratified lake located in Gam, on the island of Kri (Fig. 1). The lake is 19m deep, 460m long, 200m wide, has a surface area of 88,000m^2, and a circumference of 1700m. Shortest distance to the sea is 130m. No obvious connections to the sea were observed, and tidal amplitude was highly dampened as compared to the sea (7% of the sea). The lake contains one karstic island overgrown by trees. Water column characteristics include a halocline at approximately 7m depth. Temperature ranges from 32.1 to 33°C above the halocline and decreases to 28°C at the bottom. Salinity ranges from 21.0 to 24.1ppt, but steeply increases to 28.2ppt below the halocline. The lake is well oxygenated above the halocline (3.5mg/L on average), but rapidly becomes anoxic at around 8m depth until the bottom (Supplemental Fig. 2). The benthic community is dominated by approximately 23 species of sponge, and mussel beds of *Brachidontes* sp. For the fish community, only one species of Gobiidae was present (*Acentrogobius janthinopterus*), as well as striped silverside (*Atherinomorus endtrachtensis* (Quoy and Gaimard, 1825)). Two turtles were present at the time of surveying. The lake is closely located to a village, and the turtles were likely placed there by villagers. There is a wooden path constructed along one side of the lake, connecting the northern shore of the island to the southern side where the village is located.

Papua 32 is a small, non-stratified lake located on the island of Urani in Wayag (Fig. 1). The lake is 5.5m deep, 125m long, 55m wide, has a surface area of 6,000m^2, and a

circumference of 315m. Shortest distance to the sea is 75m. The tidal amplitude of this lake is approximately half of that of the surrounding sea. The lake has mangroves at its perimeter. An underwater cave is present with a high density of sponges near the entrance. Throughout the water column, temperature ranges from 29.4°C to 31.2°C, salinity from 29.1ppt to 31ppt, and oxygen from 1.2mg/L to 3.4mg/L (Supplemental Fig. 2). Benthic communities of the lake are characterized by large assemblages of algae (*Cladophora* sp. (Kütz, 1843)), covering the entire wall and bottom of the lake. In between algae mats, high diversity of sponge and medium diversity of mollusc species are present. A pelagic species of Gobiidae was present (unidentified), as well as large numbers of large shrimp (*Parhippolyte uvaea*).

Benthic lake communities and their predictors

Ordination methods were employed to show clustering of benthic group coverage and qualitative diversity measures per lake. A non-metric Multidimensional Scaling analysis (nMDS) of coverage of major benthic groups was performed for all transects separately (Supplemental Fig. 4). Not all transects clustered close together per lake (e.g., Papua 6 or 19), indicating that there were differences between different sites within the lakes. For averages of transect data per lake, more connected lakes clustered towards higher abundances of corals and CCA, while more isolated lakes tended to have higher coverage of bivalves (Fig. 3A, Supplemental Fig. 5). The nMDS of qualitative diversity estimates of mollusc, sponge and fish communities also showed a cluster of highly connected, lagoon-like lakes, while other lakes were more spread out (Fig. 3B).

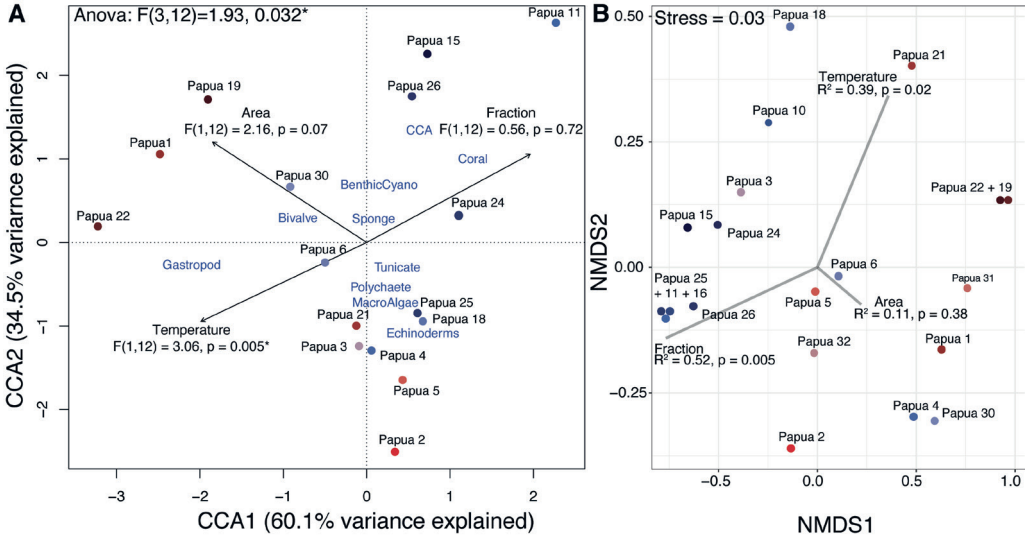


Figure 3: Ordinations of biological diversity assessments. A: Canonical correspondence analysis of coverage data from major benthic groups. B: non-Metric Multidimensional Scaling plot of qualitative diversity assessments of sponge, mollusc and fish diversity. Colors indicate gradient in connection to the surrounding sea.

Next, we tested which predictors significantly explained the (dis)similarity between lakes using ordination analyses and Mantel tests. For the coverage of major benthic groups, the canonical correspondence analysis showed temperature to be a significant predictor in variation among lakes (Anova: $F(1,12) = 3.06$, $p = 0.005$) (Fig 3A). In contrast, Mantel tests correlating Bray-Curtis dissimilarity of benthic group coverage to geographic and environmental distance measures did not indicate any significance (Supplemental Table 1). For the Mantel tests, an adjusted p-value to correct for running multiple tests was applied and a p-value of less than 0.008 (0.05/6) was required in order for significance to assigned. The fitting of environmental factors on the nMDS ordination showed connection to the surrounding sea ($R^2 = 0.52$, $p = 0.005$) and temperature ($R^2 = 0.39$, $p = 0.02$) to be important predictors of qualitative diversity estimates. This time, the environmental correlation was corroborated by Mantel tests for environmental distance as a whole ($r = 0.40$, $p = 0.001$) and specifically for temperature and salinity distances ($r = 0.32$, $p = 0.002$, and $r = 0.4$, $p = 0.001$, respectively) (Supplemental Table 1).

Drivers of diversity measures

Finally, we explored whether major benthic group diversity and richness were correlated to specific predictor variables from Island Biogeographic Theory. Shannon diversity score was highest in lakes Papua 3, 4, 15 and 21 (average score = 1.56), being almost 60% higher than Shannon diversity in lakes with the lowest scores (Papua 1, 2, and 11, average score = 0.62) (Fig. 4). Remarkably, only one lake (Papua 15) harbored all 10 benthic groups within the transects, while gastropods were missing from the next lake with high richness (Papua 11). Lowest richness was found for lakes Papua 1 and 22, harboring representatives of only 6 of the benthic groups.

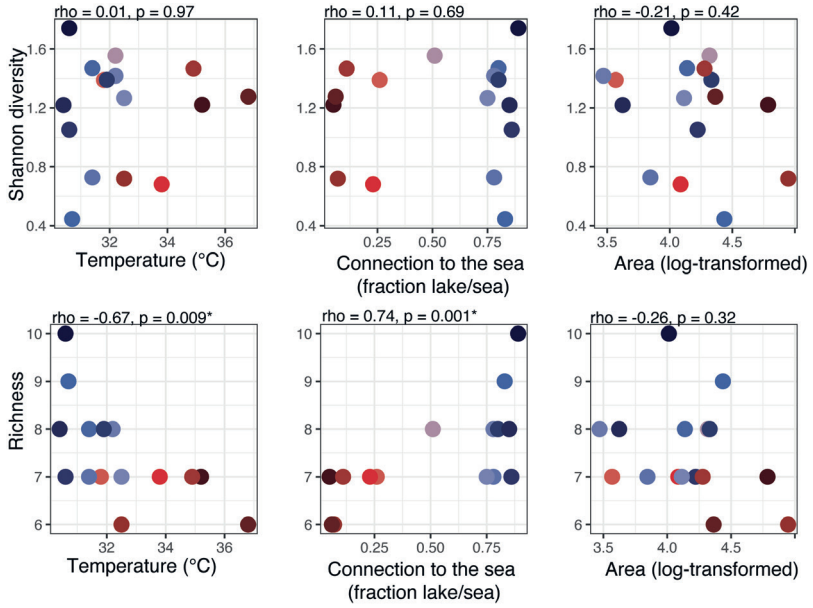


Figure 4: Spearman correlations between Shannon diversity (top) and richness (bottom) of major benthic groups and temperature, connection to the surrounding sea and lake area. Colors indicate gradient in connection to the surrounding sea.

Shannon diversity was not significantly correlated to any of the environmental factors (Fig. 4, Supplemental Table 2). Conversely, richness of major benthic groups in marine lakes did show a significant correlation with connection to the surrounding sea. We corrected the p-value for the effects of multiple testing, and a p-value smaller than 0.0055 was required to assign significance. Richness significantly increased with increasing connection to the sea (rho = 0.74, p = 0.001*) (Fig. 4, Supplemental Table 2). Other hypothesized predictors of richness such as area did not show significant correlations (rho = -0.26, p = 0.32) (Fig. 4, Supplemental Table 2).

Discussion

We visited 32 marine lakes in Raja Ampat, Indonesia, 17 of which new to science. From these 32 lakes, 20 could be surveyed extensively, and we established environmental and ecological baselines. The lakes span a temperature gradient of 30.0-36.8°C and vary highly in connection to the surrounding sea (ranging from tidal amplitudes of 5% to 89% compared to the sea). Two distinct clusters were observed when categorizing the lakes: 1) lagoon-like lakes with a pronounced connection to the sea, environmental conditions similar to the surrounding sea and communities resembling those of lagoons (Papua 11, 16, 24, 25 and 26), and 2) extreme condition lakes, with a minimal connection to the sea, extremely high temperatures and low salinities, and depauperate benthic communities (Papua 19, 21, 22). The other 11 lakes did not fall in either of these clusters and were moderately isolated, with a variety of environmental conditions. Connection to the surrounding sea and water temperature were found to be most correlated to community structure, qualitative diversity estimates and major benthic group richness.

Drivers of benthic community structure

General assumptions of Island Biogeography Theory were not completely corroborated by patterns found in the marine lakes of Raja Ampat. Island Biogeography Theory puts forward several hypotheses about the assembly of communities (MacArthur and Wilson, 1963, 1967; Losos and Ricklefs, 2010; Warren et al., 2015). Firstly, increased island area is expected to result in a larger number of species as extinction rate is expected to be lower on larger islands. Secondly, with increased distance to the source population, species number is hypothesized to be reduced due to lower immigration rates. Finally, communities need time to establish, and as a result, older islands are expected to have more diverse communities than young islands. Additionally, it can be expected that species communities on islands are more similar if islands are geographically closer to each other, since they sample from the same species pool (Fukami, 2015). Contrasting to these assumptions, we found benthic group richness not to be related to lake area (species-area hypothesis), distance to the sea (species-isolation hypothesis) or to depth as a proxy for age. Assumed terrestrial nutrient availability also did not seem to play an important role in determining benthic group diversity. Additionally, we found similarity of marine lake benthic communities not to be explained by geographic distance.

The discrepancy between general expectations of IBT and results from marine lakes of Raja Ampat could have the following explanations. It is important to note that the marine lakes included in this study are of limited size (average surface area: 16,500m^2). A study into species-area relationships in freshwater lakes found that in lakes under of 1,000km^2, species richness did not correlate with lake area (Wagner et al., 2014). Hence, the lakes of Raja Ampat may be too small to be able to detect an effect of lake size. To study effects of age, lake depth was used as a proxy for time since isolation, since deeper lakes were filled earlier with sea water than shallower lakes (Dawson et al., 2009). Consistent with a study into genetic differentiation of marine lake jellyfish populations (Dawson and Hamner, 2005), we did not observe a relationship between community diversity/richness and time since isolation, indicating that other drivers are more important. Ratio of lake circumference and area can be a proxy for the influx of terrestrial nutrients, which may affect food webs and thus species communities, but the availability of nutrients is likely determined by a multitude of complex interactions as found for freshwater lakes (Scheffer, 1997). Hence, future studies should include additional measurements of nutrient level, as food availability is likely important in the establishment success of species (Losos and Ricklefs, 2010). Finally, distance to sea may not be a true equivalent of 'distance to source pool' defined for oceanic islands, since other factors also play a role, such as type of connection (caves versus fissures) and presence of an anoxic layer. Alternatively, effects of the biogeographical parameters discussed on benthic group diversity and richness may simply be overwhelmed by other drivers.

Geographic proximity did not explain benthic community similarity, hence each lake likely harbors a distinct subset of the species pool, likely established through a combination of stochastic historical contingency and environmental filtering effects (Fukami, 2015). We did observe strong relationships of both benthic community structure and richness with degree of connection to the surrounding sea and water quality parameters (temperature and salinity). Connection to the surrounding sea may be a better proxy defining 'distance to the source pool' than geographic distance between lake and sea. Therefore, the associations between benthic community coverage and diversity and connection may indicate a species-isolation relationship. An ecological study in anchialine caves also found connection type to be most important in determining benthic community richness, both directly through regular influx of new propagules, but also indirectly through increased nutrient availability (Calderón-Gutiérrez et al., 2018). While degree of connection has obvious implications for the permeability of lakes by propagules, the local environment of temperature, salinity, pH and dissolved oxygen may pose a subsequent filter to the establishment or persistence of species (Belyea and Lancaster, 1999). Environmental filtering may therefore also explain the presence/absence of certain taxa by selectively culling species that do not tolerate local conditions. This would explain why in marine lakes with more extreme environments we find fewer benthic groups, presumably because not all organisms can tolerate the conditions of high temperatures and low salinities. However, interactive biotic effects such as competition

or predation may also be important in determining community assembly (Kraft et al., 2015; Cadotte and Tucker, 2017). As this was not explored in the current study, we cannot confidently distinguish biotic and abiotic factors. While we could not disentangle the effects of connection and water quality because of high correlations, we did observe some lakes that had contrasting environments (e.g., Papua 31 and 32 showing low connection and low temperature, Papua 6, 25 and 30 showing high connection and high temperature Fig. 1). Discovering and including more marine lakes with such contrasting environments will allow assessing the effects of connection and water quality separately.

Future studies into community assembly of marine lakes of Raja Ampat should continue from the baselines set in this study. There is a need to go beyond defining broad benthic groups, and instead look at actual species. In the current study, many lakes harbored a high number of major benthic groups, but from personal observations we know lakes can differ extensively in the number of species per benthic group. This could also explain why we found no relationships of environmental and geographic predictor variables with Shannon diversity calculated for major benthic groups. Monitoring studies of marine lakes therefore should look beyond the level of broad groups as important relationships may be missed. Furthermore, the influence of lake characteristics such as geographic distance, connection to the sea and local environments on population genetic differentiation has already been touched upon by Maas et al. (2018), but would benefit from increased number of replicates. Since population differentiation is the first step in the speciation continuum (Lowry and Gould, 2016), using in particular the highly isolated marine lakes may help explore patterns of incipient speciation in the marine realm. Furthermore, studying how food webs change with different lake environments may elucidate biotic effects on community assembly.

Utility of marine lakes as island-like systems and implications for conservation

Marine lakes function as marine analogues of oceanic islands in key aspects (Warren et al., 2015), making them suitable systems to study as island-like systems. Marine lakes have clear boundaries, are variably isolated, are replicated on small spatial scales, and have high endemism (Dawson and Hamner, 2005; Gotoh et al., 2009, 2011; Goto et al., 2011; Becking et al., 2013a, 2016; Swift et al., 2016; Maas et al., 2018, 2020). In addition, marine lakes have originated *de novo* with the rising of sea water level after the Last Glacial Maximum (Tomascik and Mah, 1994; Dawson et al., 2009), and thus their age can be approximated. Furthermore, as the lakes occur as a aggregations of multiple lakes on a small spatial scale, they can be seen as archipelagoes of replicated systems. Eco-evolutionary work greatly benefits from such replicated settings (i.e. Galapagos islands (Darwin, 1872; Almén et al., 2016), Hawaii (Gillespie, 2004, 2016), and work on the threespine stickleback fish (Berner et al., 2009; Roesti et al., 2014)). Despite old assumptions of marine species encountering few barriers to dispersal and therefore showing little differentiation (Hedgecock, 1986; Palumbi, 1994; Avise, 2004), semi-isolated habitats such as marine lakes likely exist on various spatial and temporal scales, allowing differential community assembly and diversification.

Raja Ampat, located in the center of the Coral Triangle, is an area of extreme marine species richness (Hoeksema, 2007; Mangubhai et al., 2012; Asaad et al., 2018). Multiple recent population genetic studies find high population differentiation on small spatial scales in this area (Barber et al., 2006; Carpenter et al., 2011; Starger et al., 2015). It is still not clear what processes generated and maintain the unparalleled biodiversity of the Coral Triangle (Hoeksema, 2007). Isolated systems such as marine lakes or ocean basins likely have been present throughout glacial cycles and could be refuges for relict species and/or cradles of diversity (Bowen et al., 2013; Tornabene et al., 2015). Hence, high marine biodiversity could be generated and maintained by such habitats, and as such, they warrant effective conservation. As explored by Maas et al. (2020), most marine lakes in Raja Ampat geographically fall within Marine Protected Areas (MPAs). However, the zonation of Raja Ampat's MPA network excludes marine lakes from management (Ministerial regulation No. 36, 2014). The current study provides further incentive to view all marine lakes as specific unique habitats, which should be included under a specific zonation type under current MPA planning.

Marine lakes as future oceans

Besides quantitative correlations of benthic community structure, diversity and richness, striking observations were made for the three types of marine lakes (lagoon-like, intermediate and extreme). Communities are rich, albeit very different in composition, for lagoon-like lakes and more isolated lakes. Even semi-isolated lakes (tidal amplitude 50-75% that of the surrounding sea), with associated higher temperatures and lower salinities, harbor rich communities of molluscs and sponges. Particularly the bivalve *Brachidontes* sp. is able to form dense mussel beds in these types of lakes, potentially owing to its large tolerance of water quality conditions (Sarà et al., 2008), and potentially due to absence of predators. From personal observations we know that in lakes containing large fish, dense mussel beds are absent, which would support the hypothesis on predation being important. Sponges may also benefit from release from competition by corals, which are typically less temperature and salinity tolerant (Bell et al., 2013).

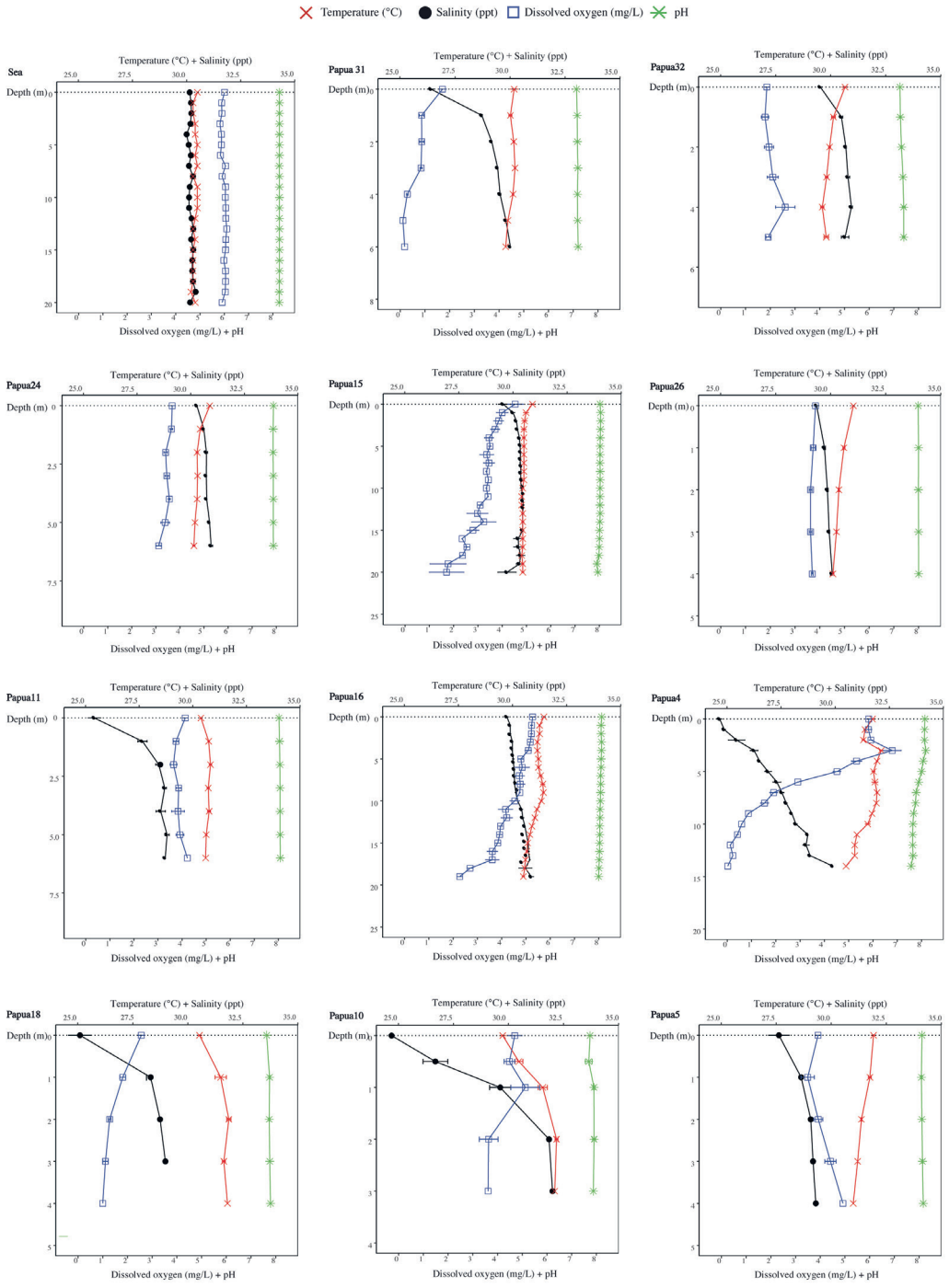
A stark contrast is provided by lakes with 'extreme' environments. These lakes have temperatures >34°C and salinities <20ppt, as well as an anoxic layer. The benthic communities in these lakes are extremely depauperate, only harboring a few species of sponges and molluscs. As climate change progresses, in the most extreme IPCC scenario (RCP8.5), sea surface temperature for the world's oceans will increase by 2.58°C (5-95% range: 2.34-2.82°C) by the year 2100 (IPCC, 2014), and waters will become less saline due to the melting of the ice caps. Furthermore, changes in temperature and salinity will facilitate stratification in coastal waters, resulting in increased anoxic events. Since certain marine lakes already display such situations, and therefore act like 'future oceans', marine lakes present a unique model system to study the effects of climate change on marine populations.

Acknowledgements

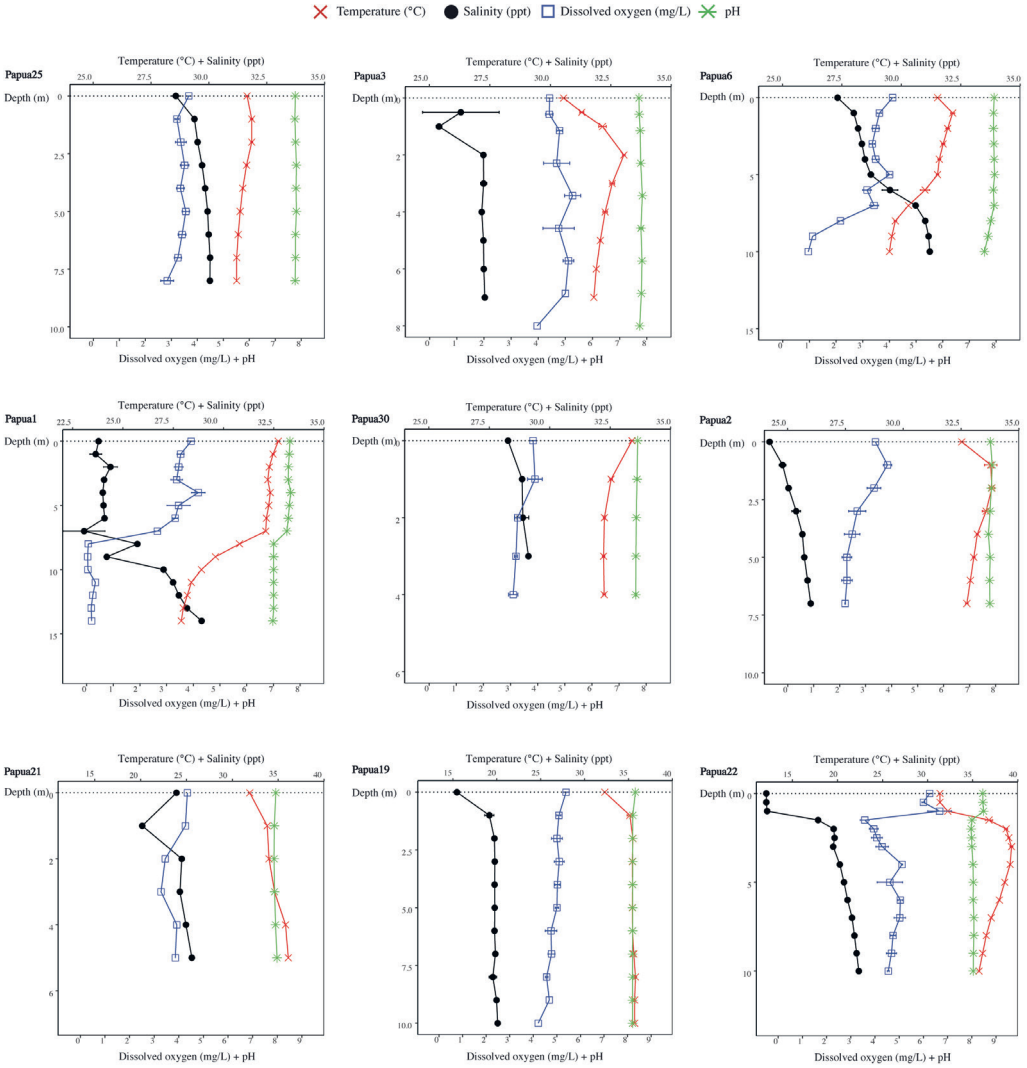
For logistical support during our sampling periods, we would like to thank the Nature Conservancy, Conservancy International, Papua Diving, Misool Eco Resort, and Baseftin. We would like to thank Maria Meijer, Luca Prins, Stephanie Martinez and Merethe Zalm for their contributions towards the field work. This study was funded through grant VENI# 863.14.020 provided by the Netherlands Organization for Scientific Research. Field work was made possible through the National Geographic Early Career Grant, Waitt grant and Young Explorers grant, the Rufford Grant, The Schure-Beijerinck-Popping grant of the Royal Durch Academy of Science, and the Treub-Maatschappij fund. We are grateful to the Indonesian Institute of Sciences (LIPI) and the Indonesian State Ministry of Research and Technology and Higher Education (KEMENRISTEK-DIKTI) for providing the following research permits: 098/SIP/FRP/SM/V/ 2011, 3B/TKPIPA/E5/Dit.KI/III/ 2016, 30/VI-P2SP2/UNIPA/ 2017, 56/SIP/FRP/E5/Dit.KI/III/ 2018).

Supplemental Figures

Supplemental Figure 1: Overview picture of all marine lakes, ordered by geographic location (Fig. 1).

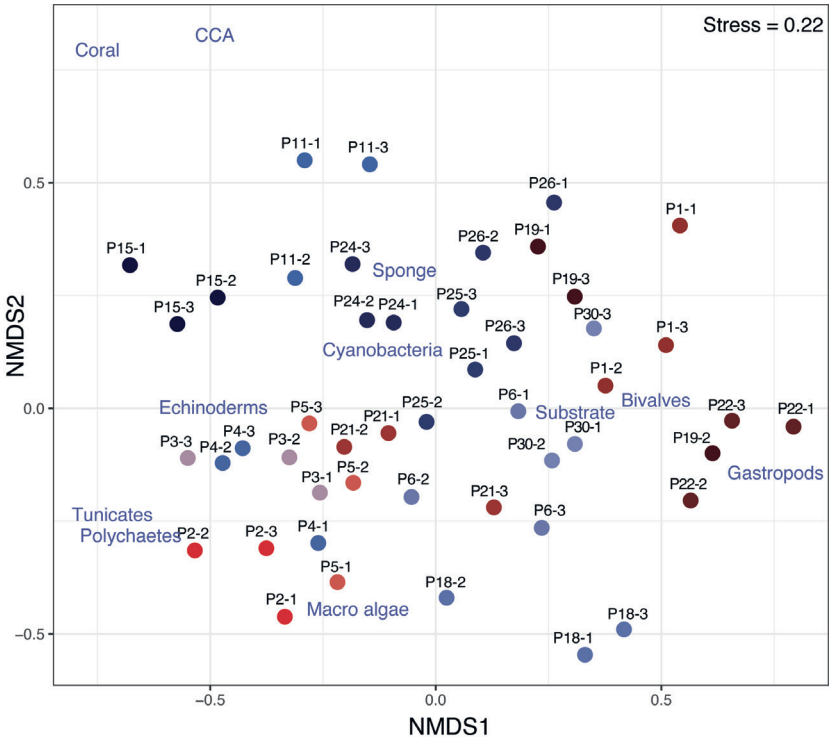


Supplemental Figure 2: Depth profiles of ocean location and marine lakes. Temperature, salinity, pH and dissolved oxygen are displayed from surface of the lake until the sediment.

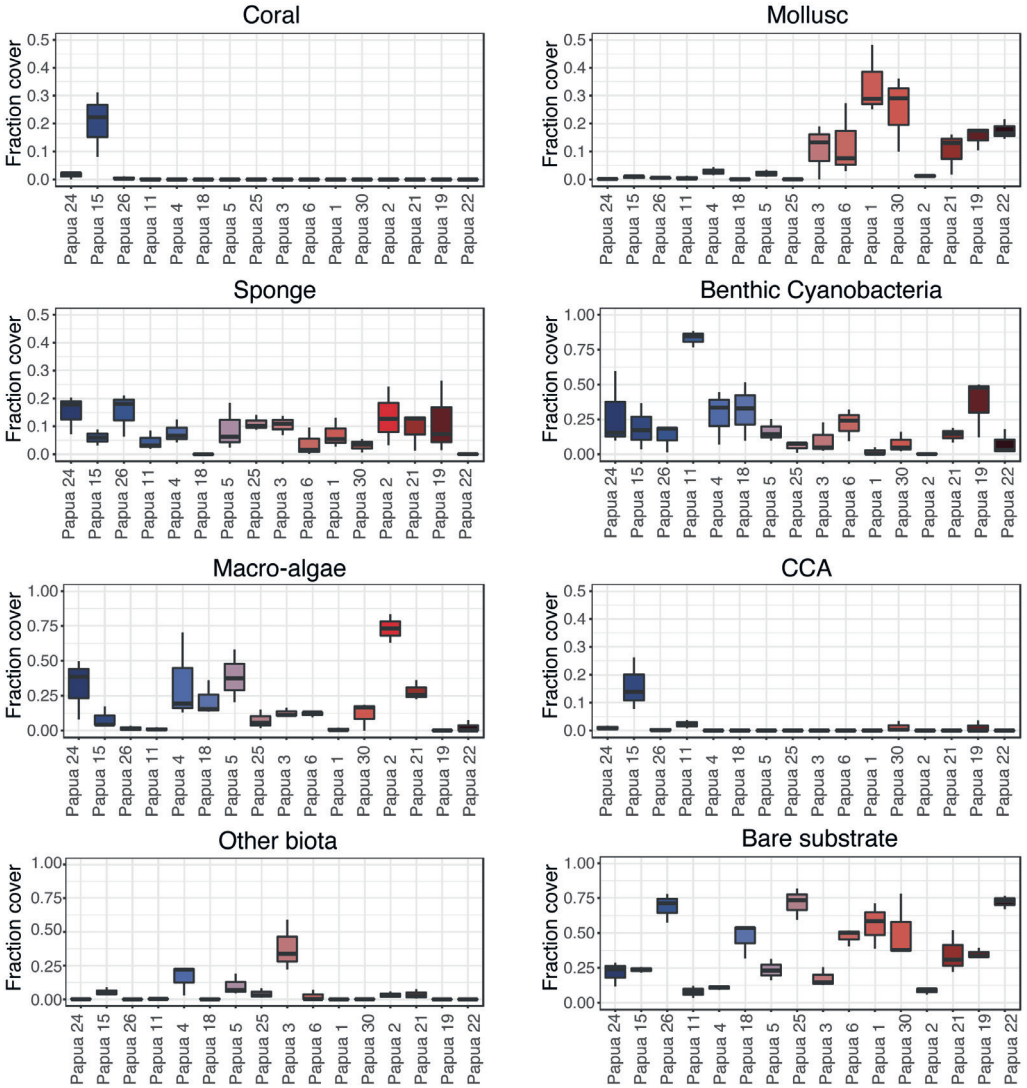


Supplemental Figure 2: Continued.

Supplemental Figure 3: Spearman correlations of biophysical parameters. Circles are drawn when there was a significant correlation. Size of the circle indicates strength of the correlation.



Supplemental Figure 4: Non-metric Multidimensional Scaling plot of coverage of major benthic groups for each individual transect. Color indicates gradient in connection to the surrounding sea.



Supplemental Figure 5: Boxplots of coverage of major benthic groups. Color indicates gradient in connection to the surrounding sea.

Supplemental Tables

Supplemental Table 1: Mantel tests of Bray-Curtis dissimilarities of coverage data of major benthic groups and Euclidean distances of qualitative diversity assessments versus geographic and environmental distance matrices.

Predictor	Coverage	Diversity
Geographic distance	r = 0.05, p = 0.36	r = 0.05, p = 0.24
Environmental distance	r = 0.32, p = 0.02	r = 0.40, p = 0.001
Temperature distance	r = 0.30, p = 0.02	r = 0.32, p = 0.002
Salinity distance	r = 0.26, p = 0.04	r = 0.39, p = 0.001
pH distance	r = 0.14, p = 0.12	r = 0.07, p = 0.18
Proxy terrestrial distance	r = 0.09, p = 0.26	r = 0.07, p = 0.18

Supplemental Table 2: Spearman correlations of Shannon diversity and richness of major benthic groups versus physical and environmental factors.

Predictor	Shannon diversity	Richness
Temperature	rho = 0.01, p = 0.97	rho = -0.63, p = 0.009
Salinity	rho = 0.25, p = 0.35	rho = 0.58, p = 0.02
pH	rho = 0.18, p = 0.52	rho = 0.22, p = 0.41
Proxy Terrestrial Influence	rho = 0.01, p = 0.96	rho = 0.22, p = 0.41
Connection	rho = 0.11, p = 0.69	rho = 0.74, p = 0.001*
Area	rho = -0.22, p = 0.42	rho = -0.21, p = 0.43
Depth	rho = 0.39, p = 0.14	rho = 0.21, p = 0.43
Distance to sea	rho = 0.05, p = 0.85	rho = -0.35, p = 0.18
Stratification	Chisq = 0.04, p = 0.83	Chisq = 1.51, p = 0.22

Supplemental Information A: Lake descriptions

WAYAG

Papua 27

Papua 27 is more an anchialine pool than a true marine lake and located in Wayag. The lake is 2m deep, 200m long, 100m wide, has a surface area of 22,000m^2, and a circumference of 615m. The lake is located very far from the surrounding sea (270m). Connection was not measured quantitatively but appeared to be medium. The lake is surrounded by jungle and patches of mangroves. Environmental profiling was not done, but preliminary measurements indicated a temperature of 29.5°C, a salinity of 31ppt. Ecological surveys were not done, but presence of red shrimp was noted.

Papua 28

Papua 28 is a small lake or perhaps anchialine pool located in Wayag. The lake is 1.5m deep, 53m long, 27m wide, has a surface area of 2,000m^2, and a circumference of 207m. Shortest distance to the sea is 100m. Connection is estimated to be low, but we only very shortly visited this lake, so environmental profiling and ecological surveys were not done. It was noted, however, that the benthic community held a very low diversity of sponge species but did contain two species of shrimp.

Papua 29

Papua 29 is located in Wayag. The lake consists of two basins connected via a shallow passage surrounded by mangroves. The first basin is very shallow and characterized by a muddy sediment. We did not visit the second basin. The lake is 2m deep, 223m long, 110m wide, has a surface area of 15,700m^2, and a circumference of 690m. Shortest distance to the sea is 50m. Connection to the surrounding sea was medium (tidal amplitude about 40% of surrounding sea), although no visible cave or tunnel was present. Extensive environmental profiling was not performed, but from preliminary measurements in the first basin temperature was 31.4°C, salinity 29.5ppt, and oxygen 2.0mg/L. A strange sensation of burning on our skin was felt when swimming in the lake. Ecological surveyance was not performed, but an absence of mussels and a presence of red shrimp was noted.

Papua 30

Papua 30 is a round lake located in Wayag, in northern Raja Ampat. The lake is 4.1m deep, 130m long, 100m wide, has a surface area of 13,000m^2, and a circumference of 470m. The lake is located very close to the surrounding sea (35m shortest distance). Although no conspicuous connection to the sea was noticed, the lake had a medium connection to the surrounding sea (tidal amplitude 75% of that of the sea). The lake is surrounded by rock and

some patches of mangroves. The lake is not stratified, and temperature (32.7°C ±0.5), salinity (28.7ppt ±0.4) and oxygen (3.6mg/L ±0.6) are stable throughout the water column. The benthic community is characterized by dense population of upside-down jellyfish (*Cassiopea* sp). The major macro algae were *Caulerpa* sp. Also, aggregations of *Brachidontes* sp. mussels were present. No fish were observed, with Gobiidae being notably absent. One large sea turtle was present, indicating that a tunnel must exist, or that it was placed there by humans. Large, oceanic-like individuals of *Mastigias papua* were present.

Papua 31

Papua 31 is located in the western part of Wayag on the island of Bag. The lake is 7m deep, 85m long, 70m wide, has a surface area of 6,200m^2, and a circumference of 274m. Distance to the sea is very large (275m), which probably explains the low connection of the lake to the surrounding sea (tidal amplitude 30% of that of the sea). The lake has a large patch of mangroves on one side of the perimeter, and an outcrop of karstic rock on the other. A halocline was present at 1m depth, and salinity varied from 26.4ppt at the surface to 29.4ppt at the bottom. Temperature was remarkably stable, from 30.2°C to 30.1°C at the bottom, but oxygen ranged from 1.5mg/L at the surface to becoming almost anoxic (0.2mg/L) at the bottom. The benthic community held dense mussel beds of *Brachidontes* sp., and two species of shrimp were present.

Papua 32

Papua 32 is a small, non-stratified lake located on the island of Urani in Wayag. The lake is 5.5m deep, 125m long, 55m wide, has a surface area of 6,000m^2, and a circumference of 315m. Shortest distance to the sea is 75m. The tidal amplitude of this lake is approximately half of that of the surrounding sea. The lake has mangroves at its perimeter. An underwater cave is present with a high density of sponges near the entrance. Throughout the water column, temperature ranges from 29.4°C to 31.2°C, salinity from 29.1ppt to 31ppt, and oxygen from 1.2mg/L to 3.4mg/L. Benthic communities of the lake are characterized by large assemblages of macroalgae (*Cladophora* sp. (Kütz, 1843)), covering the entire wall and bottom of the lake. In between algae, high diversity of sponge and medium diversity of mollusc species are present. A pelagic species of Gobiidae was present (unidentified), as well as large numbers of large shrimp (*Parhippolyte uvaea*).

GAM

Papua 24

Papua 24 is a lagoon-like, small lake located at the western part of Gam island. The lake is 6m deep, 90m long, 77m wide, has a surface area of 4,200m^2, and a circumference of 260m. The lake is located very close to the sea (55m). The lake is highly connected to the surrounding sea through a large cave (estimated tidal amplitude 87% of the sea). (30.4°C ±0.3), salinity (30.6ppt ±0.2), and oxygen (3.5mg/L ±0.3) and highly stable throughout the water column. The benthic community has high diversities of both sponges and molluscs, no conspicuous aggregations of bivalves, and presence of both oysters and mussels. Many reef fish are present.

Papua 25

The lagoon-like lake Papua 25 is located at the western side of the main island of Gam. The lake is 8.3m deep, 180m long, 90-160m wide, has a surface area of 21,500m^2, and a circumference of 790m. Shortest distance to the sea is 75m. The lake is highly connected to the surrounding sea through a large cave. Estimated tidal amplitude 85% of that of the sea. There are patches of mangroves surrounding the lake, and the lake bottom consists of

patches of karstic rock and soft sediment. Again, temperature (31.6°C ±0.3), salinity (29.5ppt ±0.5), and oxygen (3.4mg/L ±0.4) are highly stable, similar to Papua 26 albeit 1 degree warmer. The benthic community is similar to that of Papua 26 in that it has very high diversity of both sponges and mollusc, no conspicuous aggregations of bivalves, and presence of both oysters and mussels. Fish communities are diverse with many reef fish present.

Papua 26

Papua 26 is a lagoon-like lake located at the eastern part of the main island of Gam. The lake is 4.5m deep, 190m long, 95m wide, has a surface area of 16,700m^2, and a circumference of 734m. Shortest distance to the sea is 80m. The lake is highly connected to the surrounding sea (tidal amplitude 86% of that of the sea) through a large cave. The lake is not stratified, and temperature (30.5°C ±0.3), salinity (29.7ppt ±0.3), and oxygen (3.7mg/L ±0.2) are highly stable throughout the water column. Both sponge and mollusc diversity are (very) high in the benthic community of Papua 26. Although conspicuous aggregations of bivalves are not present, both mussels (*Brachidontes* sp.) and oysters could be in the karstic rocks and among the mangrove roots. Fish communities consist of reef fish. The lake is made more accessible by local people through a rope ladder. Fish cages were present in the lake during our surveys.

Papua 1

Papua 1 is an arrowhead-shaped stratified lake located in Gam, on the island of Kri. The lake is 19m deep, 460m long, 200m wide, has a surface area of 88,000m^2, and a circumference of 1700m. Shortest distance to the sea is 130m. No obvious connections to the sea were observed, and tidal amplitude was highly dampened as compared to the sea (7% of the sea). The lake contains one karstic island overgrown by trees. Water column characteristics include a halocline at approximately 7m depth. Temperature ranges from 32.1 to 33°C above the halocline and decreases to 28°C at the bottom. Salinity ranges from 21.0 to 24.1ppt, but steeply increases to 28.2ppt below the halocline. The lake is well oxygenated above the halocline (3.5mg/L on average), but rapidly becomes anoxic at around 8m depth until the bottom. The benthic community is dominated by approximately 23 species of sponge, and mussel beds of *Brachidontes* sp. For the fish community, only one species of Gobiidae was present (*Acentrogobius janthinopterus*), as well as striped silverside (*Atherinomorus endtrachtensis* (Quoy and Gaimard, 1825)). Two turtles were present at the time of surveying. The lake is closely located to a village, and the turtles were likely placed there by villagers. There is a wooden path constructed along one side of the lake, connecting the northern shore of the island to the southern side where the village is located.

MISOOL

Papua 9

Papua 9 is located on the main island of Misool, at the northeast side. Depth was not determined, but the lake is 250m long, 115m wide, has a surface area of 33,600m^2, and a circumference of 735m. Shortest distance to the sea is 111m. Connection appeared to be high, with a large cave being present. Extensive profiles of water quality parameters were not done, but preliminary measurements indicated a temperature of 31.7°C, salinity of 25.5ppt and dissolved oxygen content of 4.9mg/L. The lake seemed not to be stratified. Reef fish like groupers and snappers were present in this lake, as well as a few individuals of golden jellyfish (*Mastigias papua*) and many small moon jellyfish (*Aurelia* sp.).

Papua 2
Papua 2 is a slightly stratified lake located near the main island of Misool at the northeast part. The lake is 7.3m deep, 180m long, 70m wide, has a surface area of 12,200m^2, and a circumference of 600m. Shortest distance to the sea is 87m. There is no visible cave, and tidal amplitude is about 20% compared to the surrounding sea. The lake consists of two main basins connected through a narrow (~5m) passage. Mangrove trees are fringing the perimeter of the lake. A slight pattern of stratification was observed for this lake, but no presence of an anoxic layer. Temperature ranged from 32.5°C at the surface to 33°C at the bottom, salinity from 24.2 to 26ppt, and oxygen from 3.4 to 1.2mg/L. The benthic community consisted of a high diversity of sponge species, medium diversity of molluscs including mussel beds of *Brachidontes* sp., and *Halimeda* sp. macro algae. Fish were present, namely species of *Gobiidae* and some reef fish. Most remarkably, large assemblages of golden jellyfish (*Mastigias papua*) were present. Human activities were evident here, in the form of eco-tourism for the jellyfish, but also evidence of aquaculture activities such as fish cages was found.

Papua 10
Located directly adjacent to Papua 2 is the lagoon-like lake of Papua 10. The lake is 3m deep, 100m long, 90m wide, has a surface area of 15,000m^2, and a circumference of 580m. Shortest distance to the sea is 88m. A large cave is present, through which there is a high degree of connection to the surrounding sea (although not measured quantitatively). Temperature was measured as 30.9°C, salinity as 28.1 ppt and pH as 7.8. Benthic community is characterized by presence of hard coral, a medium diversity of coral, and *Halimeda* macro algae. Reef fish were present in this lake, as well as a few large individuals of golden jellyfish *Mastigias papua*. Like its neighboring lake, Papua 10 contained fish cages for aquaculture.

Papua 3
Papua 3 is located near the main island of Misool in the northeast. The lake is 7.5m deep, 180m long, 130m wide, has a surface area of 156,000m^2, and a circumference of 760m. Shortest distance to the sea is 145m. A large, shallow cave is present but likely does not connect the lake directly to the sea, but instead to a small body of water situated in between the lake and the sea. Tidal amplitude is about half that of the sea. The lake has patches of mangroves. Temperature fluctuated slightly from the surface to the bottom: average, 31.7°C ±0.9. Salinity was notably lower at the surface (21.9ppt) than throughout the rest of the water column (23.5-27.8ppt). The lake was well oxygenated, having an average dissolved oxygen of 4.7mg/L (±0.7). Almost all benthic area was covered by dense agglomerations of mussels (*Brachidontes* sp.) or tubeworms. Additionally, sponge diversity is high and there is a medium number of mollusc species. Fluctuating populations of jellyfish (*Mastigias papua* and *Aurelia* sp.) are present.

Papua 11
Papua 11 is a highly mixed, lagoon-like lake located near the main island of Misool. The lake is 8.9m deep, 540m long, 150-350m wide, has a surface area of 27,300m^2, and a perimeter of 1100m. The closest distance to the sea is 25m. There is a high connection to the surrounding sea via a large cave located in the north of the lake, and tidal amplitudes are only 20% less than the surrounding sea. The lake contains some small islands of karstic rock and has a large terrestrial cave at the surface in the eastern part. Water characteristics remain very stable throughout the water column. Temperature ranges from 30.1-30.6°C from surface to the bottom, oxygen from 4.1mg/L to 4.2mg/L, with only salinity being notably lower at the surface (25.4ppt) compared with the bottom (28.7ppt). At the time of survey,

assemblages of reef fish were present and as well as a single green turtle (*Chelonia mydas* (Brongniart, 1800)). Likely, these large organisms are able to pass through the northern cave to the surrounding lagoon. The benthic community resembles that of nearby reef flats, harboring hard corals and high sponge and mollusc diversity.

Papua 7
Papua 7 is located near the main island of Misool in the northeast. The lake was not measured for depth, but is 185m long, 80m wide, has a surface area of 9,700m^2, and a circumference of 530m. Shortest distance to the sea is 175m. Connection was not quantified but seemed low, although the lake appeared not to be stratified. We could not perform environmental profiling of the entire water column, but initial measurements indicated a salinity of 14.9ppt. We did not perform extensive benthic surveys, but sponge diversity seemed to be very low. Mussel beds were present.

Papua 12
Papua 12 is an elongated lake in northeast Misool. We did not determine depth for this lake, but it is 190m long, 80m wide, has a surface area of 7,500m^2, and a circumference of 463m. Shortest distance to the sea is 55m, and connection seemed to be very high, although we did not exactly quantify it. We could not environmentally profile this lake, but sponge diversity seemed to be high.

Papua 8
Papua 8 is a small, shallow lake located in northeast Misool. The lake is 2.4m deep, 95m long, 80m wide, has a surface area of 2,100m^2, and a circumference of 449m. Shortest distance to the sea is 40m. Despite the short distance to the sea, connectivity is estimated to be low, although no quantitative measurements were performed. Due to its shallow depth, the lake did not appear to be stratified. Environmental profiling was not performed. Ecological surveys were not done, but benthic sponge and mollusc community were estimated to be very low in diversity. Mussel beds were present.

Papua 13
Papua 13 is a small lake located on one of the many islands of northeast Misool. The lake is 4m deep, 63m long, 33m wide, has a surface area of 1,400m^2, and a circumference of 174m. Shortest distance to the sea is 145m. Connection estimated to be high, but was not quantified. The lake appeared to be not stratified. Extensive environmental profiling was not done, but a preliminary measurement indicated a salinity of 31.3ppt. Ecological surveys were not done, but sponge diversity appeared to be high. Reef fish were present.

Papua 4
Papua 4 is located in northeast Misool on Lenmakana island. The lake is 20m deep, 190m long, 90m wide, has a surface area of 13,750m^2, and a circumference of 612m. Shortest distance to the sea is 55m. Although no large connection was found during our survey, the lake has a high tidal amplitude (80% of that of the adjacent sea), suggesting a large connection. However, the lake is strongly stratified, with an anoxic layer containing purple sulphur bacteria at around 5-7m depth. Temperature is relatively stable throughout the water column (average 31.6°C ±0.4), but salinity is notably lower above 25ppt and below the halocline, increasing to 29.8ppt. Oxygen above the halocline is 5.8mg/L, but steeply decreasing to becoming anoxic below the halocline. The benthic community above the anoxic layer contains mussel beds of *Brachidontes* sp., and medium diversity of molluscs and sponges. Some species of Gobiidae are present, as well as catfish. However, we

suspect the latter to be introduced. The lake contains dense populations of golden jellyfish (*Mastigias papua*) and therefore is an increasingly popular tourist destination.

Papua 14
Adjacent to Papua 4 is lake Papua 14. The lake is approximately 10m deep, 97m long, 40m wide, has a surface area of 4,000m^2, and a circumference of 267m. Shortest distance to the sea is 50m. We were unable to environmentally profile this lake, but it seemed to have a high connectivity to the surrounding sea. There is no visible connection to the sea, but there is to a blue hole. Species of Gobiidae were present at the side near Papua 4, and mating aggregations of nudibranchs were observed when we surveyed the lake.

Papua 15
Papua 15 is a lagoon-like lake located on one of the outer islands of northeast Misool. The lake is 34m deep, 150m long, 85m wide, has a surface area of 10,300m^2, and a circumference of 410m. Shortest distance to the sea is 98m. The lake is highly connected to the surrounding sea (tidal amplitude of lake almost 90% of that of the sea). To access the lake, one first has to pass through a shallow lagoon which connects the lake to the sea through a cave. The lake is not stratified, and temperature, salinity and oxygen remain stable throughout the water column (30.5°C ±0.2, 31.1ppt ±0.7, 3.2mg/L ±0.8, respectively). On the rocky sediment the benthic community consists of high sponge diversity and very high mollusc diversity. While no conspicuous aggregations of either were present, both mussels (*Septifer* sp.) and oysters were present. Reef fish were present as well as Gobiidae. Fluctuating populations of golden jellyfish (*Mastigias papua*) were also present.

Papua 16
Another lagoon-like lake, Papua 16 is a very large located at the very outskirts of northeast Misool. The lake is 19m deep, 380m long, 95m wide, has a surface area of 401,000m^2, and a circumference of 876m. Shortest distance to the sea is only 37m. Connection to the sea was very high (85% of the sea). Similar to lake Papua 15 water quality parameters were stable throughout the water column: temperature 30.1°C (±0.3), salinity 30.2ppt (±0.3), oxygen being slightly more variable 2.1-5.4mg/L. The benthic community was also very similar to Papua 15, with very high diversity of sponges and mollusc being present, including individuals of both mussels (*Septifer* sp.) and oysters, although not in obvious aggregations. Reef fish were present as well as Gobiidae.

Papua 23
Papua 23 is a lake located on the island of Yellit in southeast Misool. The lake is 10m deep, 97m long, 95m wide, has a surface area of 5,000m^2, and a circumference of 350m. Shortest distance to the sea is 56m. Connection appeared to be low. The lake is surrounded by very high outcrops of karstic rock. Environmental profiling was not done, but one measurement of temperature indicted the lake was 30°C. The benthic community appeared to be low in sponge and mollusc diversity, and no aggregation of bivalves were observed. *Cladophora* sp. appeared to be the major macroalga. No large fish communities were present.

Papua 5
Papua 5 is a shallow lake closely located to a marine lake converted to an aquaculture pond on the island of Karawapop in southeast Misool. The lake is 5m deep, 75m long, 60m wide, has a surface area of 3,700m^2, and a circumference of 300m. Shortest distance to the sea is 75m. No large tunnels or caves are visible, and it has a low tidal connection to the surrounding sea (tidal amplitude less than a third than the surrounding sea). Being very shallow, the water column is not stratified despite the low connection. Temperature ranges

from 31-32.6°C, salinity from 27.3-29.4ppt and oxygen from 2.9-5.44mg/L throughout the water column. The benthic community consists of a large diversity of molluscs, including *Brachidontes* sp. mussel beds, and a medium diversity of sponges. Conspicuous are the large numbers of sea cucumbers (Holothuroidea). The lake previously contained dense populations of golden jellyfish (*Mastigias papua*), and the owners of the island have started building a jetty into the lake. However, since a few years the jellyfish population disappeared, and the building came to a stop.

Papua 17
Located at the western tip of Kalig island, the round lake Papua 17 is located in southeast Misool. Depth could not be determined, but the lake is 100m long, 92m wide, has a surface area of 6,500m^2, and a circumference of 314m. Shortest distance to the sea is only 20m. The lake has an estimated large connection to the surrounding sea via a large cave. Mangroves fringe the lake. Extensive environmental profiling was not done, but the lake appeared to be not stratified. Preliminary measurements of temperature, salinity and oxygen yielded values of 31.7°C, 31.7ppt, and 4.0mg/L. Ecological surveying was not done, but sponge diversity appeared to be very high. There were only oysters present regarding bivalves. The main macroalgae was *Cladophora* sp.

Papua 6
Located on the same island as Papua 17 (Kalig), Papua 6 is another small, round lake in southeast Misool. The lake is 12.4m deep, 65m long, 60m wide, has a surface area of 37,000m^2, and a circumference of 200m. The lake is located very close to the surrounding sea, shortest distance being 28m. The lake has a medium connection to the surrounding sea through a cave (tidal amplitude 78% of surrounding sea). Despite the large connection, the lake shows a halocline starting at 7m depth. Temperature above the halocline is 31.8°C and decreases to 29.9°C below the halocline. Salinity goes from 28.2ppt above the halocline to 31.1ppt below. Oxygen decreases from 3.6mg/L above the halocline to 1.0mg/L below, but no anoxic layer is present. Sponge and mollusc diversity is medium, and both oysters and *Brachidontes* sp. mussels are present. The mussels form dense mussel beds. The main macroalga is *Cladophora* sp. Fluctuating populations of golden jellyfish (*Mastigias papua*) are present, as well as some reef fish and gobies. Anthropogenic influence was apparent through the presence of fish cages.

Papua 18
Located slightly to the west of Papua 6 on the island of Kalig, Papua 18 is a very elongated lake in southeast. The lake is 4.6m deep, 220m long, 30m wide, has a surface area of 7,000m^2, and a circumference of 550m. Adjacent to the lake is a small lagoon which connects it to the surrounding sea. Shortest distance to the lagoon is 100m. Although there is no apparent channel, the lake has clear water flow at the side of the lagoon, despite the considerable distance to this lagoon. The lake has a medium connection to the surrounding sea, tidal amplitude being about 78% of tidal amplitude of the sea. The lake is not stratified, temperature (31.2°C ±0.8), salinity (27.2ppt ±2.0) and oxygen (1.8mg/L ±0.7) being relatively stable throughout the water column. The benthic community is characterized by a low diversity of sponges and a medium diversity of molluscs, including the bivalve *Brachidontes* sp. There are multiple species of large reef fish present (such as groupers and napoleon wrasse). The community of reef fish was remarkably similar between two surveying years, indicating that perhaps only juvenile specimens can successfully cross into the lake. There is one species of burrowing Gobiidae.

Papua 19
The largest lake of Misool, Papua 19, is located on the island of Warakaket in southeast Misool. It is one of four lakes closely clustered on Warakaket island. The lake is 31.5m deep, 470m long, 160m wide, has a surface area of 61,000m^2, and a circumference of 1125m. Shortest distance to the sea is 90m. The lake is surrounded by karstic rock with no mangroves present. The bottom of the lake has much bare substrate. The lake is highly isolated from the surrounding sea, with tidal amplitudes being only 5% of that of the surrounding sea. The low connection results in stratification, and generally lower salinities and higher temperatures than the surrounding sea. A very brackish, colder water (32.3°C, 15.5ppt) layer lays on top of a warmer, more saline layer (35.5°C, 19.7ppt) that extends until at least 20m depth as that was the length of the cable of our multimeter. The lake was oxygenated throughout the 20m (5.0mg/L). A chemocline is probably present at 12m depth, below which visibility is dramatically reduced and macrobenthos are absent. The benthic community was very depauperate, with very low diversities of sponge (4 species) and mollusc species (3 species), although mussel beds of *Brachidontes* sp. were present. Except for a species of Gobiidae, and one unknown species, no fish were present. The fish were notably bigger than the fish in the other three lakes on Warakaket.

Papua 20
Papua 20 is directly adjacent to Papua 19, but more to the west, on Warakaket island in southeast Misool. Depth was not determined, but the lake is 211m long, 122m wide, has a surface area of 22,400m^2, and a circumference of 660m. Shortest distance to the sea is 260m. Connection was estimated to be very low. Environmental profiling was not done, but the lake was warm (35°C) and a thermocline was present. The benthic community was depauperate, with only 2 species of sponge and 2 species of mollusc. Three fish species were observed, including Gobiidae.

Papua 21
More west and to the north to Papua 20 is Papua 21, on Warakaket island in southeast Misool. The lake is 13m deep, 230m long, 90m wide, has a surface area of 19,000m^2, and a circumference of 850m. Shortest distance to the sea is 80m. Although this lake is slightly more connected to surrounding sea as compared with Papua 19 and Papua 22, it still is very isolated (tidal amplitude 10% of surrounding sea). The lake is surrounded by mangroves. No clear halocline is present, but water parameters still notably differ throughout the water column (temperature average 34.4°C ±1.5, average salinity 23.9ppt ±1.9, and average oxygen 3.9mg/L ±0.4). The benthic community is notably richer than those of Papua 19 and Papua 22, but still low diversities of mollusc and sponges are present, albeit in higher abundance. Between visits in May and November, turbidity decreased, and visibility increased. A few fish species are present, including Gobiidae and grey mullet.

Papua 22
Papua 22 is a highly stratified lake located on the island of Warakaraket in the archipelago of southeast Misool. The marine lake is 12.5m deep, 250m long, 170m wide, has a surface area of 23,000m^2 and the lakes circumference is 730m. The closest distance to the surrounding sea is 83m. Tidal influence is very limited in the lake, with tidal amplitude in the lake being only one-tenth of the tidal amplitude of the surrounding sea. No apparent tunnels or caves are visible. A shallow pool is partially separated from the main basin by an outcrop of karstic rock in the north of the lake. Papua 22 is the warmest known marine lake in Raja Ampat, with a recorded maximum of 42°C. Water column characteristics include a strong halocline at approximately 1.5m depth. A colder (31.5°C), brackish (salinity: 12ppt) water layer of about 1.5m lays on top of a warmer (39°C), more saline (19ppt) layer. Temperature

decreases slightly to 36°C at the bottom of the lake, and salinity increases slightly to 22ppt. The lake is likely heliothermal, resulting in extreme temperature differences between water layers. In heliothermal lakes, the low density surface layer acts as a one-way mirror to solar radiation, preventing convection of heated higher salinity water from deeper layers to the surface and thus preventing the subsequent cooling via evaporation (Sonnenfeld and Hudec, 1980). The whole lake is oxygenated, ranging from 6.3mg/L at the surface, to 4.5mg/L at the bottom. Two benthic gobiidae species are present (*Exyrias puntang* (Bleeker, 1851) and *Acentrogobius janthinopterus* (Bleeker, 1852)), as well as one free swimming species (*Mugilogobius* sp. (Smitt, 1900)). None of the fish were observed to venture below the strong thermocline. The benthic community is highly depauperate, composed of 2 sponge species, 1 gastropod and 2 bivalves (*Brachidontes* sp. (Swainson, 1840) and one endobenthic *Bastissa* species (Prime, 1862)). All molluscs found below the thermocline were dead, except the gastropod which seemed to tolerate the extreme temperatures.

Chapter 3

Recognizing peripheral ecosystems in marine protected areas: A case study of golden jellyfish lakes in Raja Ampat, Indonesia

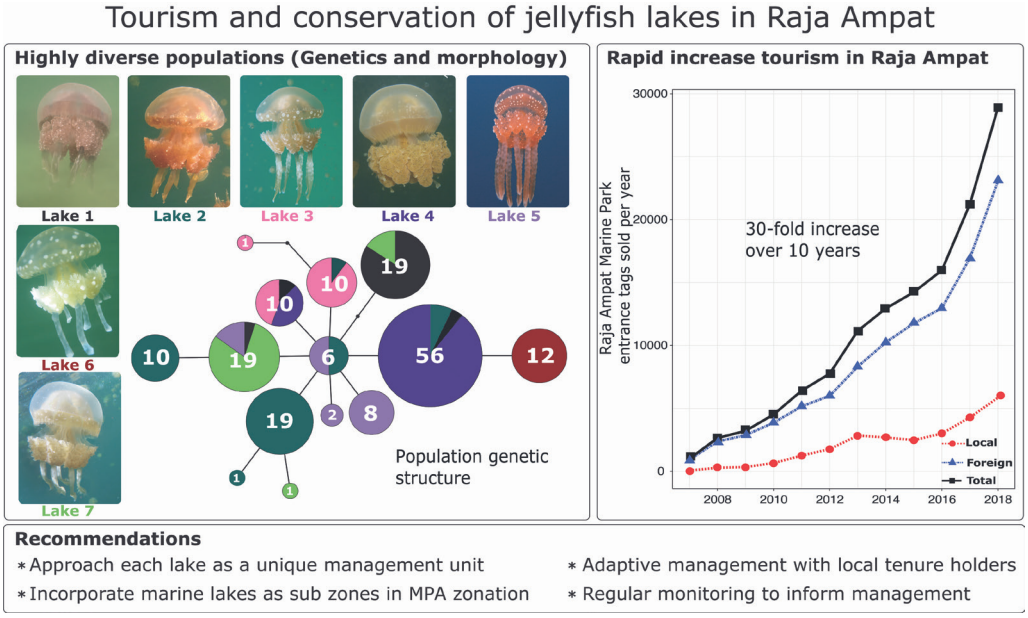
Diede L. Maas, Agustin Capriati, Awaludinnoer Ahmad, Mark. V. Erdmann, Machiel Lamers, Christiaan A. de Leeuw, Luca Prins, Purwanto, Amanda P. Putri, Ricardo F. Tapilatu, Leontine E. Becking

Marine Pollution Bulletin ***2020****, 151*

Abstract

Peripheral marine ecosystems can harbor endemic diversity and attract tourism attention, yet are generally not included in conservation management plans due to their remoteness or inland positioning. A case study in Raja Ampat of seven landlocked marine lakes containing golden jellyfish (*Mastigias* spp.) was conducted to address the lack of fundamental insights into evolutionary, ecological and social contexts of these ecosystems. An interdisciplinary approach was taken towards identifying the jellyfish lakes as distinct management units in order to incorporate them into existing Marine Protected Areas. *Mastigias papua* populations showed strong genetic (ϕ_{ST}: 0.30 - 0.86) and morphological ($F = 28.62$, $p = 0.001$) structure among lakes, with putative new subspecies. Risks arising from rapid increase in tourism to Raja Ampat (30-fold since 2007) and lake tenure disputes warrant restrictions on jellyfish lake use. Recommendations are provided for adaptive management and science-based conservation policies for jellyfish lakes across Indonesia.

Keywords: *Mastigias papua*, population genetics, morphometrics, tourism, anchialine ecosystems, customary tenure



Introduction

Peripheral marine ecosystems, such as marine lakes, anchialine caves, and cenotes attract significant attention from tourism due to their propensity to harbor endemic diversity and relatively pristine environments (Dawson et al., 2001; Mercado-Salas et al., 2013; Becking et al., 2015; Gerovasileiou et al., 2016; Lopez-Maldonado and Berkes, 2017; Martínez et al., 2019; Masonjones et al., 2019). Tourism interest can have socio-economic benefits for local communities, can lead to an awareness towards management of the ecosystem and prevent destructive exploitation (Pretty and Smith, 2004; Trave et al., 2017). However, if tourism is not managed effectively, recreational use of peripheral marine ecosystems may have devastating effects, particularly in combination with other stressors such as climate change (Cerrano et al., 2006; Hoegh-Guldberg and Bruno, 2010; Hughes et al., 2017; Trave et al., 2017). Prime examples of peripheral ecosystems that are tourism magnets are "jellyfish lakes", landlocked marine lakes containing dense populations of golden jellyfish, *Mastigias* spp. (Scyphozoa: Rhizostomae, Lesson 1830). Of the approximately 200 marine lakes known worldwide (Holthuis, 1973; Dawson et al., 2009), 22 located in Indonesia, Palau and Vietnam contain jellyfish (Hamner and Hamner, 1998; Dawson and Hamner, 2005; Becking et al., 2015; Swift et al., 2016). Ongeim'l Tketau in Palau, probably the most famous marine lake in the world, attracts thousands of tourists yearly (Dawson et al., 2001). Recently discovered jellyfish lakes in Indonesia are also rapidly becoming tourist attractions (Becking et al., 2015). However, conservation management plans are lacking for peripheral ecosystems in general (Mercado-Salas et al., 2013; Martínez et al., 2019; Masonjones et al., 2019), and for jellyfish lakes in Indonesia in particular. Indonesian jellyfish lakes can serve as a case study to formulate an approach towards effectively incorporating peripheral systems into conservation management plans.

The development of effective conservation programs for jellyfish lakes in Indonesia is hampered by incomplete knowledge of local jellyfish evolution and ecology and the unique social context of the lakes. Designating marine lakes as individual management units (MUs) could facilitate their incorporation into governmental management plans (Moritz, 1994; Palsbøll et al., 2007). To be considered an MU *sensu* Moritz (1994), biological populations should show "significant divergence at nuclear or mitochondrial loci, regardless of the phylogenetic distinctiveness of alleles", preferably in concordance with non-molecular traits. Marine lakes in Palau are reported to harbor isolated jellyfish populations of unique subspecies with adapted morphology, suggested to be indicative of incipient speciation (Dawson and Hamner, 2005; Swift et al., 2016). The adapted morphology is described to be 'derived' from the ancestral oceanic ecotype (Dawson and Hamner, 2005; Swift et al., 2016). The oceanic ecotype displays eight large terminal clubs (>0.5 times bell diameter), white spots and blue pigmentation, whereas the derived ecotype has fewer and smaller (~0.2 times bell diameter) terminal clubs and no spots or blue pigmentation. The shift from the oceanic towards the derived ecotype in marine lakes is hypothesized to be facilitated by higher turbidity and lower predator presence in isolated lakes (Dawson, 2005a; Swift et al., 2016). Demonstrating concordant genetic differentiation and morphological adaptation in

Indonesian jellyfish lake populations would provide scientific incentive to view marine lakes as distinct MUs.

Vulnerability of jellyfish lake populations to external stressors has been reported for the marine lake Ongeim'l Tketau in Palau, which suffered severe crashes in jellyfish abundance in 1998, and again in 2016 (Dawson et al., 2001; Howard, 2016). The crashes were associated with changes in lake temperature and salinity following El Niño Southern Oscillation (ENSO) events (Dawson et al., 2001; Martin et al., 2006). In addition, concerns about the impact of sunscreen pollution on jellyfish proliferation have been raised (Bell et al., 2017). Jellyfish are known for boom-bust dynamics (Pitt and Lucas, 2014; Dawson et al., 2015). Here, a 'boom' of rapid increase in medusa abundance occurs via strobilation of benthic polyps (scyphistoma) related to potentially seasonal patterns of water temperature increase (Sugiura, 1965; Collins, 2002; Holst et al., 2007), followed by steep declines after a few months due to age or starvation resulting in a 'bust' (Pitt and Lucas, 2014). However, in marine lakes of Palau and Indonesia environmental conditions are considered to be relatively stable throughout the year, with no clear seasonal patterns (Hamner et al., 1982; Muscatine et al., 1986). Therefore, strobilation is expected to occur continuously throughout the year, which would preclude boom-bust dynamics. While some natural fluctuation in jellyfish abundance has been recorded for jellyfish lakes in Palau (Dawson et al., 2015), the reported population crashes suggest severe external perturbation. Lessons learned from Palau can aid in the formulation of conservation management plans for jellyfish lakes in Indonesia.

There are thirteen jellyfish lakes currently known in Indonesia, located in Berau (East Kalimantan province, 3 lakes), Togean (Central Sulawesi, 1 lake), and Raja Ampat (West Papua, 9 lakes) (Becking et al., 2015; Swift et al., 2016, current study). West Papua lies within one of twelve ecoregions of Indonesia (Spalding et al., 2007), and has by far the most marine lakes documented of any region in Indonesia. West Papua has been the focus of significant Marine Protected Area (MPA) development over the past two decades (Mangubhai et al., 2012), and has become the first "conservation province" in Indonesia (Act No. 522.5/123/6, 2015). As of 2013, Raja Ampat contained seven multiple-use MPAs forming a network of over 1 million hectares (Gaman et al., 2012; Grantham et al., 2013). Two new protected areas (North Misool and Fam MPAs) have been added in the past three years, bringing the total protected marine area to just over 2 million hectares. Development of MPA zonation plans in West Papua is strongly based on the customary tenure rights of local communities, which allow them to control utilization of surrounding natural resources (Kartasapoetra et al., 1985; McLeod et al., 2009). In Indonesia, the special autonomy of West Papua regarding customary tenure rights is recognized nationally (Act No. 21, 2001), and also by local communities and tourism stakeholders. Customary tenure therefore has important bearing on the current and future management of the jellyfish lakes.

Local tenure holders provided significant input on recommendation for the management of jellyfish lakes during the first iteration of the Raja Ampat MPA management plan (Mustaghfirin et al., 2012). However, it was subsequently decided by the Ministry of Marine Affairs and Fisheries (MMAF) to change the classification of Raja Ampat's MPA network from a "coastal and small islands park" to a "marine tourism park" meant that only areas below the high tide mark are explicitly managed under the Raja Ampat MPA zonation system (MMAF Ministerial regulation No. 36, 2014). Jellyfish lakes therefore currently are not incorporated in the MPA management plans as they only govern areas up to the high tide mark. Nonetheless, jellyfish lakes in Raja Ampat are actively being promoted as tourist attractions, with jellyfish depicted on the tourism entrance tag in 2019, and advertisement of lakes through homestays. The lack of explicit conservation and tourism management regulations for jellyfish lakes may have implications for the persistence of these fragile ecosystems.

We have conducted an interdisciplinary study with the aim to provide scientific basis to incorporate jellyfish lakes, unique peripheral ecosystems, into conservation management plans. Our objectives were to: 1) determine the phylogenetic position of Raja Ampat jellyfish lake populations in an Indo-Pacific context, 2) quantify variation in jellyfish population genetics, morphometrics and abundance between lakes and over time, and 3) assess lake use, customary tenure and perceived threats. To address these objectives we quantitatively sampled seven jellyfish lakes, held stakeholder interviews, and made direct observations on the use and threats of the lakes throughout Raja Ampat over eleven years. Ultimately, this study provides practical recommendations to aid management agencies in Raja Ampat, and more broadly throughout Indonesia, in implementing biologically sound, science-based conservation policies for peripheral ecosystems.

Materials and Methods

Lake profiling and sample collection

We characterized the environment of seven jellyfish lakes in Raja Ampat following the approach of Becking et al. (2011) and Maas et al. (2018), recording temperature (°C), salinity (ppt), pH, depth and connection to sea (Fig. 1, Supplemental Table 1). The lakes were coded following Maas et al. (2018) and de Leeuw et al. (2020). Three of the lakes have local names: Tomolol (Papua 2), Lenmakana (Papua 4), and Karawapop (Papua 5). Two additional lakes in Raja Ampat, Gam mentioned by Swift et al. (2016) (IDWPDAG and IDWPDCG) did not contain jellyfish during our study period. We made direct observations on jellyfish presence, lake use and tourism increase between 2009-2019. We sampled 133 *Mastigias* specimens for genetics analyses, and photographed 338 specimens for morphometric analyses between 2013-2018 (Supplemental Table 1).

DNA extraction and sequencing

Tissue samples were immediately preserved in the field in 96% ethanol or RNAlater and kept at room temperature or at 4°C throughout the field work period. Upon returning to the

lab samples were stored in -20°C. DNA was extracted using the Qiagen DNeasy Blood & Tissue kit (Qiagen, Germantown, MD, USA). Polymerase chain reaction (PCRs) were performed to amplify mitochondrial cytochrome *c* oxidase subunit I (COI) using dgLCO/HCO primers from Meyer (2003) and LCOjf and MpHCO from Dawson and Hamner (2005). We followed a protocol as defined in Becking et al. (2016). Briefly, PCR reactions were performed in 25μL volumes containing 12.5μL Onetaq Mastermix (2x), 8.5μL ddH_2O, 2μL BSA, 0.5μL of both primers (10μM), and 1μL template DNA (10ng/μL). We used denaturing, annealing and extension temperatures of 94°C, 45°C and 72°C, respectively. Amplicons were purified and Sanger sequenced at Macrogen, Inc (the Netherlands).

Forward and reverse sequences were aligned and assembled in Geneious (version 10.2.5) using the Muscle algorithm (Edgar, 2004) and MEGA7 (Kumar et al., 2016). The assemblies were visually inspected and primer-ends were trimmed, resulting in a final fragment size of 542bp. All sequences had high similarity to *M. papua*, as verified by BLASTn (Altschul et al., 1997). Two datasets were compiled: 1) Only sequences from this study, and 2) including additional samples from Palau, East Kalimantan, Papua New Guinea, Japan, Vietnam (GenBank accession numbers: KU900912-KU901464, (Swift et al., 2016)), and Puerto Rico (GenBank accession numbers: JN215543–JN21551, (Bayha and Graham, 2011)). The sequences are deposited in GenBank (MN107549-MN107722).

Phylogenetics and population genetics

Using dataset 2, a maximum likelihood phylogenetic tree was constructed and bootstrapped 1,000 times in MEGA7 (Kumar et al., 2016), with *M. andersoni* as the outgroup. The Akaike Information Criterion (AIC), calculated in jModelTest v.2.1.5 (Darriba et al., 2012), indicated HKY + I as the best fitting evolutionary model (Hasegawa et al., 1985). The tree was visualized with FigTree v1.4.2. Pairwise sequence divergence between populations (p-distances) were computed in MEGA7. Haplotype networks were constructed using HaploViewer (Salzburger et al., 2011).

Using dataset 1, population genetic structure and diversity was estimated. Nucleotide (π) (Tajima, 1983), haplotype diversity (*h*) (Nei, 1987), and number of private haplotypes per lake were calculated via the package *pegas* (Paradis, 2010), implemented in R v. 3.5.0 (R Core Team, 2018). Population structure was estimated between lakes and years via ϕ_{ST} using Arlequin with 1,000 permutations to estimate significance (Excoffier and Lischer, 2010).

Morphological measurements

To assesss morphological variation, standardized photographs of jellyfish along the longitudinal axis of specimens with a color chart and scale were taken using a GoPro HERO3 or Olympus TG-3 between 2016-2018. Images were analyzed in ImageJ (version 1.5oi). Morphological measurements were corrected for bell diameter to adjust for size effects and rescaled to values between zero and one, following Dawson (2005a). Of 13 morphological

characteristics measured, eight were concordant with Dawson (2005a). The additional 4 variables included total length, bell length, total length of oral arms and width of terminal clubs, and were measured to explore morphology in more depth. We measured continuous and categorical variables. Continuous variables included: total length, bell length, bell width (*f*9 in Dawson, 2005a), total length of oral arms, length of unwinged portion of oral arms (*f*11), length of winged portion of oral arms (*f*12), length of terminal clubs (*f*15), and width of terminal clubs. Categorical variables included: bell color (*f*1, categories: orange/brown, white and blue), oral arm color (*f*4, categories same as *f*1), terminal club color (*f*5, categories same as *f*1), shape of terminal clubs (*f*13, categories: Thick end, pointy, straight, wide), and presence/absence of white spots on the bell (*f*6). Morphological characteristics were summarized in bar plots. Kruskal Wallis tests with post-hoc Dunn's tests were performed to assess significant differences among groups, after assumptions for data normality and homogeneity of variances were found to be violated.

Morphological variation between lakes and years was displayed using Principal Component Analysis (PCA) for quantitative data and a Multiple Factor Analysis (MFA) to include categorical data. Multivariate homogeneity of group dispersions were tested to verify the assumption of homogeneous variance distributions. Permutational multivariate analyses of variance using Euclidean distances were performed to assess significant differences between groups (lakes and years). All analyses were performed in R, using *stats*, *car*, *FSA* and *vegan* (Oksanen et al., 2016) packages for statistical tests and *ggplot2* (Wickham, 2016) for visualization.

Abundance estimations

Quantitative assessments of jellyfish abundance were conducted once every year in May of 2016-2018 for lakes containing jellyfish most consistently: Papua 2, Papua 3, Papua 4 and Papua 15, resulting in 12 assessments in total. A non-invasive method was used by swimming transects with a frame (39 x 53.5 cm) with a fixed GoPro HERO3 video camera. Five to eight transects were conducted horizontally with the frame held approximately 50 cm below the surface. One transect was conducted along the longest axis of the lake, and, depending on lake size, four to seven evenly spaced along its perpendicular. Additionally, three vertical transects were performed to estimate the maximum depth of jellyfish presence. All transects were swum between 9:00-10:00AM. Transect videos were processed by counting the number of jellyfish passing the frame every 10 seconds using a hand counter. Abundance was calculated via the following formula:

$$Abundance\ transect_i = \left(\frac{Transect_i * Frame\ size}{Lake\ area * Max.depth\ jellyfish\ present}\right) * \#\ jellyfish\ transect_i$$

Finally, the average of all transects was used as the final estimation of total abundance in the lake. The jellyfish density was graphically displayed in density maps using ESRI ArcGIS (10.4.1), via Inverse Distance Weight (IDW) interpolations. Additionally, we assessed

presence/absence and categories of abundance (tens, hundreds, thousands, hundred-thousands) based on own observations and of collaborators from 2009-2019.

Tourist numbers and interviews

To assess trends in tourism visitation in Raja Ampat over time, we obtained the number of tourism entrance tags issued by the Raja Ampat Tourism Department from 2007-2018. While these permit tags are valid for the whole of Raja Ampat, we assumed that an increase in the region also meant a general increase in visitation to Misool and Wayag, and subsequently to the jellyfish lakes. The magnitude of tourist visitation to lakes from 2009-2018 was further assessed by 1) means of access to the areas (Misool and Wayag), 2) number of homestays, and 3) estimated number of visitors by interviewees.

There is no formal registration on the use and tenure of the jellyfish lakes. Therefore, we conducted semi-structured interviews (Longhurst, 2003), to obtain information from stakeholders involved, in order to collect contextual information based on general discussion items. Specifically, we asked about lake use, tenure and perceived threats. Stakeholder interviews were conducted in 2017. In total, 28 in-depth interviews were conducted with government officials, homestay owners, liveaboard operators, and locals claiming tenure of the lakes (Supplemental Table 2). All interviews were anonymous and recorded with prior permission from the interviewees. Perceived threats that were identified by more than two respondents were reported.

Results

Phylogenetic reconstruction and genetic variation

A total of 174 COI sequences of 542bp were analyzed from seven jellyfish lakes from Raja Ampat (Supplemental Table 1). We obtained 14 haplotypes with 15 variable sites, and six haplotypes were shared by at least two lakes (Fig. 2). The remaining eight private haplotypes were found in Papua 2 (3 haplotypes), Papua 3 (1), Papua 15 (2), Papua 5 (1) and Papua 6 (1).

The samples from Raja Ampat fell within the clade of *Mastigias papua*, while samples from jellyfish lakes in Berau, Indonesia are putatively classified as *M. albipunctatus* (Fig. 2) (Souza et al., 2018). Jellyfish populations from Raja Ampat formed a distinct subclade, including some Palau locations. Pairwise sequence divergence among Raja Ampat populations ranged from 0.20% (Papua 4 vs. Papua 5) to 0.70% (Papua 5 vs. Papua 30) (Supplemental Fig. 1). Variable sites are displayed in Supplemental Table 3, and show lakes Papua 2 and Papua 15 to contain the same haplotype as subspecies *M. papua etpisoni* as defined by Dawson (2005a). Other subspecies defined by Dawson (2005a) were not observed.

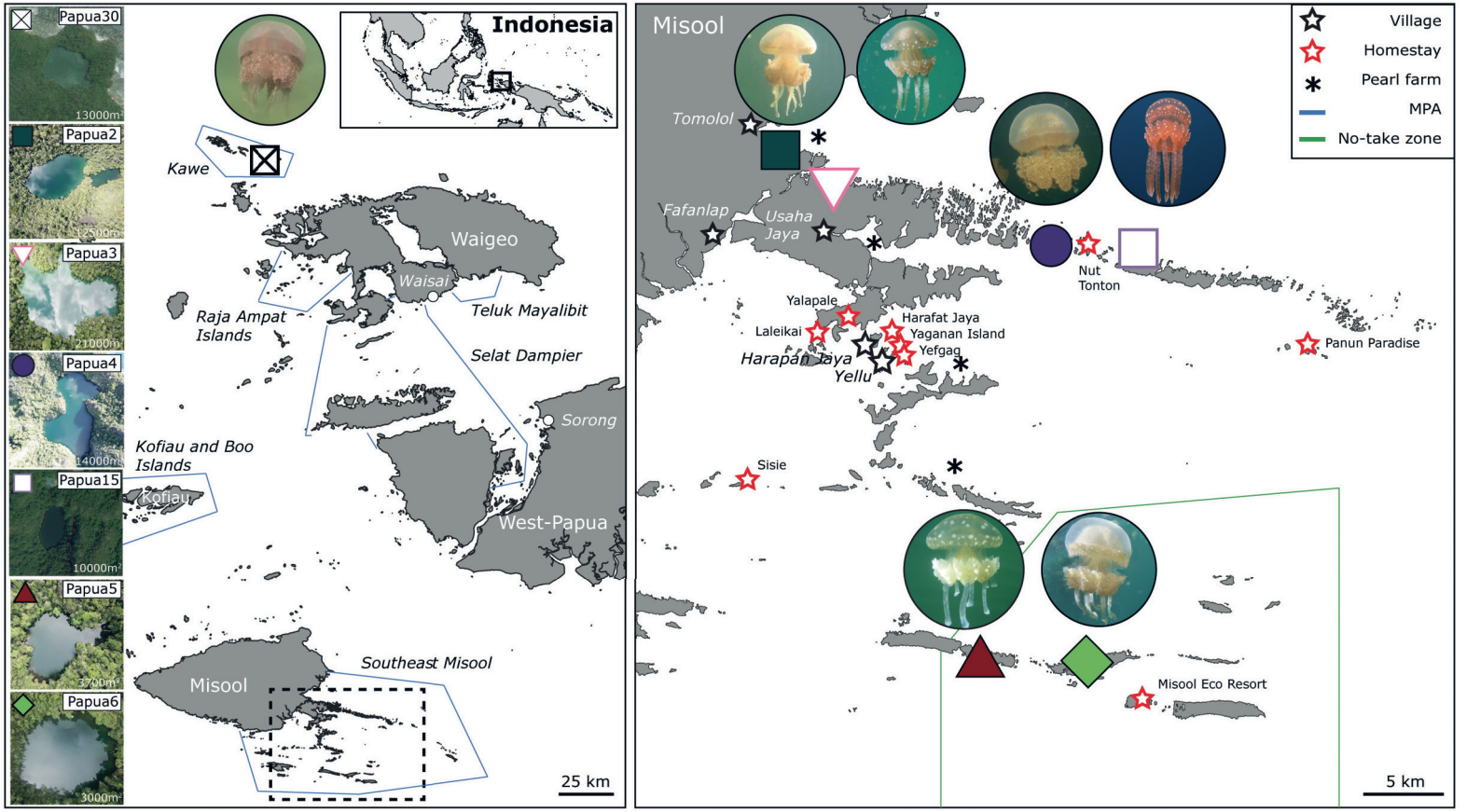


Figure 1: Sample localities of Mastigias papua *populations in Raja Ampat, West Papua, Indonesia. Jellyfish were sampled from seven marine lakes (Papua 30, Papua 2, Papua 3, Papua 4, Papua 15, Papua 5 and Papua 6). Left panel shows overview of Raja Ampat Regency, lake outlines, area, and location of Papua 30. Right panel shows zoomed-in map of Misool with locations of Papua 2, Papua 3, Papua 4, Papua 15, Papua 5 and Papua 6.*

Next, we estimated within-lake genetic diversity and genetic structure among lakes. Highest genetic diversity was found within lake Papua 2 (nucleotide diversity (π) = 0.003, haplotype diversity (h) = 0.723), and lowest in Papua 4 (π = 0.001, h = 0.233), and Papua 5, which contained only a single haplotype (π = 0, h = 0) (Supplemental Table 1). We observed significant genetic structure for all pairwise comparisons of jellyfish lakes with ϕ_{ST} ranging from 0.31 to 0.86 (average ϕ_{ST} = 0.63) (Supplemental Table 4, Supplemental Fig. 2). Comparisons between years within lakes were not significant with ϕ_{ST} ranging from 0.05 to 0.09.

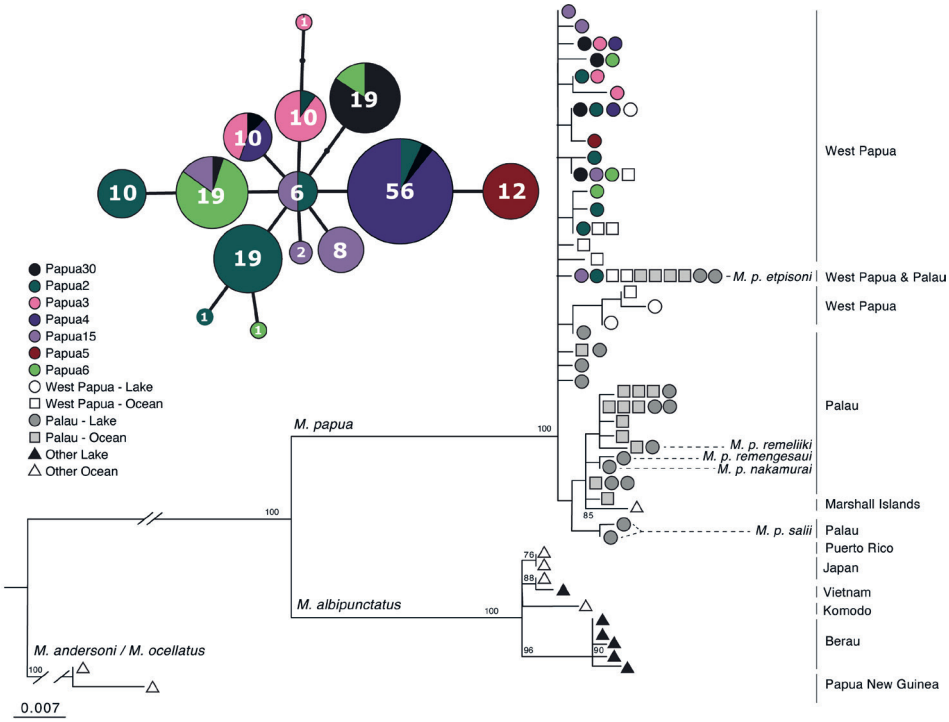


Figure 2: Phylogenetic relatedness of Mastigias spp. from West Papua and other Indo-Pacific locations. Maximum Likelihood Tree (HKY + I model) and haplotype network based on partial Cytochrome Oxidase I marker (542bp) of Mastigias spp. sampled from marine lake and ocean locations. Only maximum likelihood support of >70% are indicated, based on 1000 bootstraps. Samples from current study are colored according to Fig. 1. Samples were included from Swift, et al. 2016 (accession numbers GenBank: KU900912 - KU901464) and Bayha and Graham, 2011 (accession numbers GenBank: JN215543 - JN21551). Scale bar indicates substitutions/site. Each circle within the haplotype network represents a haplotype, with size representing number of individuals possessing the haplotype (number specified in circle), coloring according to locations in Fig. 1. Lines connecting haplotypes represent single base substitutions between haplotypes, missing haplotypes are indicated by dots.

Morphological variation

The PCA of quantitative morphometrics showed clustering of morphologies among lakes at the first two Principal Components explaining 61.2% of morphological variation (Fig. 3A). When including categorical morphological measurements, this pattern remained consistent (Supplemental Fig. 3). Multivariate analysis of variance indicated morphometrics were significantly different among lakes ($F(4,333) = 25.96$, $p < 0.001$), with all pairwise comparisons of lakes being significantly different ($p < 0.01$), apart from Papua 3 and Papua 15 ($p = 0.28$) (Supplemental Table 5). Jellyfish from Papua 3 and Papua 15 displayed the

oceanic ecotype as defined by Dawson and Hamner (2005) (terminal club length >0.5 times bell diameter, spots and blue pigmentation). Jellyfish from Papua 4 2016-2017 clearly showed the derived ecotype (loss of spots and pigmentation, terminal club length <0.2 times bell diameter). Remarkably, Papua 4 in 2018 displayed long terminal clubs (0.47 times bell diameter) and some spots, approaching the oceanic ecotype. Papua 2 and Papua 30 showed intermediary states. In 2016-2018, jellyfish were smallest in Papua 4 and Papua 15 (average bell width: 4.5 cm and 3.3 cm, respectively), and largest in Papua 30 (average: 15.0 cm) (Supplemental Fig. 4A). After correcting for size, differences among groups were most prominently shown by differential unwinged and winged oral arm length (Supplemental Fig. 4D, E), and terminal club length and width (Supplemental Fig. 4F, G).

There was no temporal variation in jellyfish morphology between years 2017-2018 for Papua 2 ($F(1,61) = 0.539$, $p = 0.38$) or Papua 3 ($F(1,74) = 4.785$, $p = 0.48$) (Fig 3B). Contrastingly, in Papua 4 a significant morphological shift was found between years 2016, 2017 and 2018 ($F(2,117) = 40.926$, $p < 0.001$) (Fig 3B). Here, jellyfish sampled in 2016 and 2017 were significantly different in morphometrics from all other groups ($p = 0.036$), whereas jellyfish sampled from 2018 were similar to Papua 2 and Papua 3.

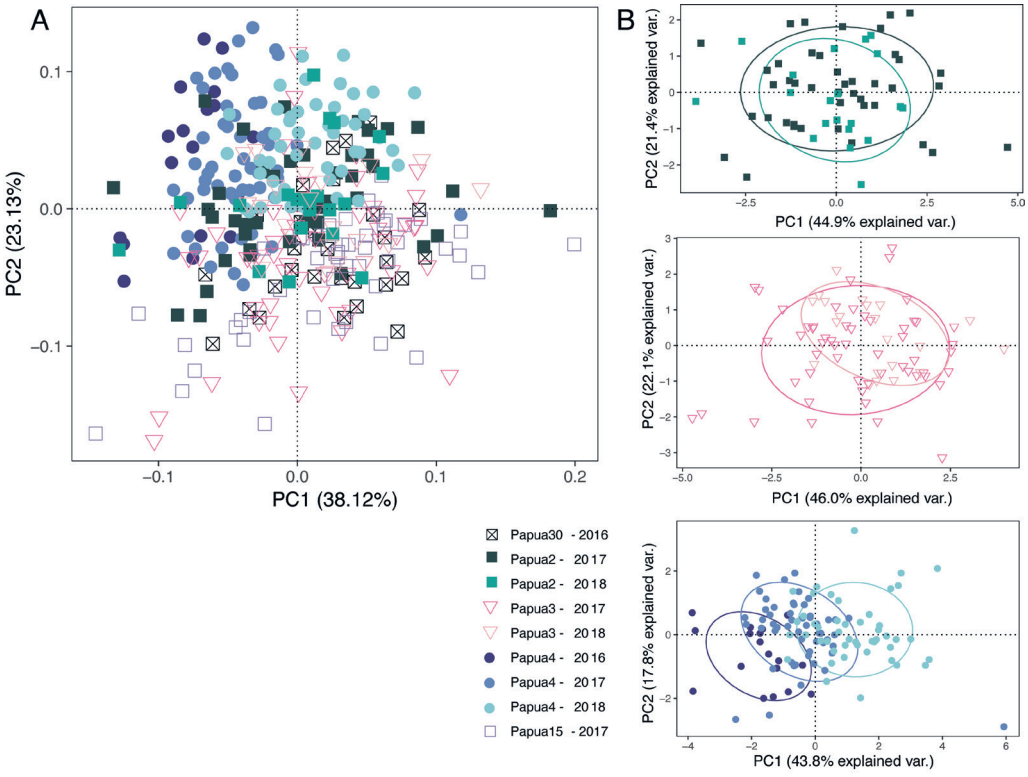
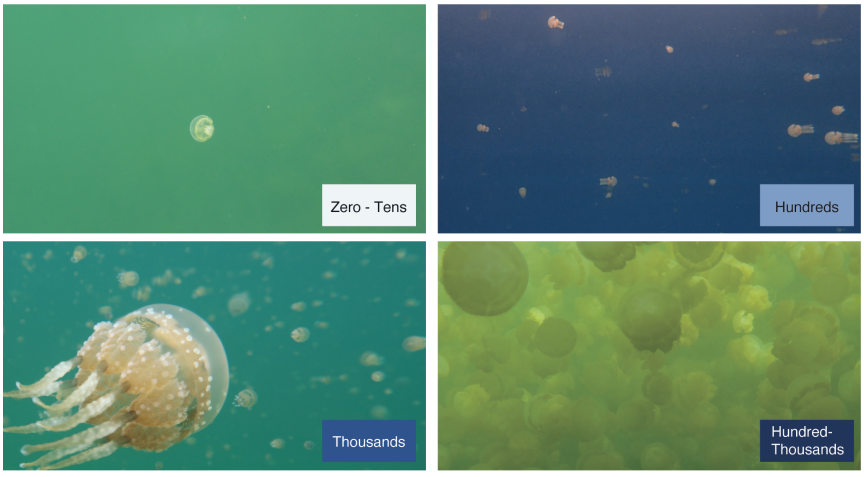


Figure 3: Spatial and temporal variation in quantitative morphometrics of Mastigias papua from marine lakes sampled in Raja Ampat. A) Principal Component Analysis (PCA) of all years sampled (2016-2018). In total, 61.25% of morphological variation is explained by first two Principal Components. B) PCAs subset of lakes. Separate plots for Papua2, Papua3 and Papua4 are displayed for different sampling years. Ellipses indicate samples falling within the standard deviation. Colors and codes correspond with Fig. 1.

Variation in abundance

Jellyfish abundance fluctuated in all lakes over eleven years (2009-2019) based on categorical observations (Fig. 4). Fluctuations, including complete disappearance of jellyfish, were present before the onset of tourist visitation in 2012. No consistent annual temporal patterns could be discerned since all lakes experienced periods with absences of jellyfish, while jellyfish were present in consecutive years. Periods of jellyfish absence were not synchronised among lakes. Quantitative assessment of jellyfish abundance between years 2017-2018 showed a decrease for all lakes (Fig. 4, Supplemental Fig. 5). The decrease was especially notable in Papua 4, where the total estimated number of jellyfish in the lake was 527,000 in 2017 and plummeted to 94,500 in 2018 (Fig. 4).



	2009 June	2011 May	2013 Sept	2014 July	2016 May	2016 Nov	2017 May	2017 Nov	2018 May	2019 Jan
Papua30										
Papua2					200,000		65,000		27,000	
Papua3					200,000		900,000		57,000	
Papua4					52,800		527,000		94,500	
Papua15							3,000		950	
Papua5										
Papua6										

Lake Papua4

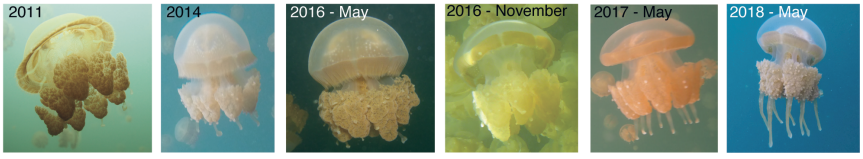


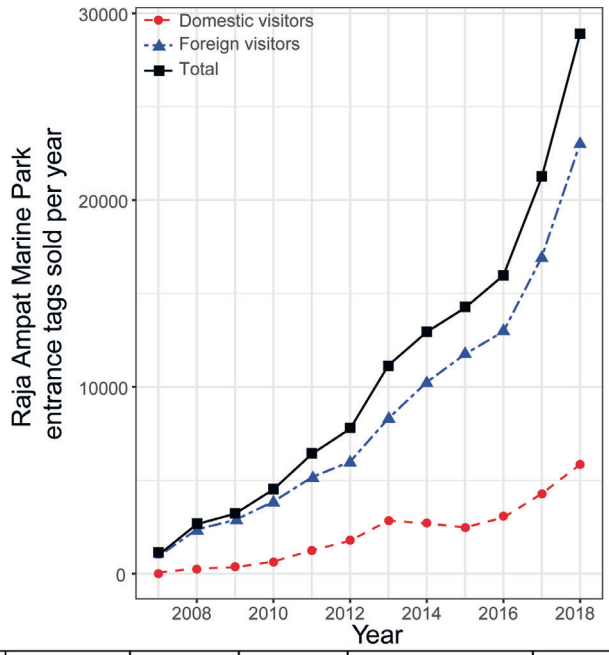
Figure 4: Presence-absence and abundance data of Mastigias papua in seven marine lakes in Raja Ampat. Colors of boxes in top picture refer to abundance categories used in the table and pictures represent reference abundance categories. Numbers in table represent quantitative assessments based on video analyses. The main tourist

attraction, Lake Papua4 (local name Lenmakana), is highlighted to show the morphological shift in jellyfish from this lake. Grey values indicate no assessment was done.

Tourism increase and lake tenure, use and perceived threats

In 2007, a total of 998 Raja Ampat tourist entrance tags were sold (domestic and foreign), which increased to 28,896 tags in 2018 (Fig. 5A). Access to Misool (location of most marine lakes) only became organised since 2013 with regular public ferry services, which currently run four times per week from Sorong (Fig. 5A). Jellyfish lakes are mainly visited by tourists for snorkelling and photography, although use is not spread evenly between lakes (Fig. 5B). The main operators of jellyfish lake visits are local guides, homestays, and liveaboard dive vessels. In 2018, there were eleven homestays in southeast Misool (Fig. 1). Homestays and liveaboards in southeast Misool offer regular trips to three jellyfish lakes: Papua 2 (local name Tomolol), Papua 4 (local name Lenmakana) and Papua 5 (local name Karawapop), of which Papua 4 is visited the most frequently (Fig. 5B). Transport to Wayag (location of Papua 30) is only possible via chartered tourism vessels. Regulations governing the Wayag MPA do not allow for homestay development, and visitation to this lake is limited to occasional liveaboard guests and private charters from Sorong.

A



B

Lake	#Operators	Visits / month	Est. visitors / year	Tenure	Perceived threats
Papua30				Selpele and Saleo	Low threat
Papua2 (Tomolol)	2	2	900	Tomolol and Usaha Jaya	Tourism, invasive species, waste
Papua3				Fafanlap and Usaha Jaya	Invasive species
Papua4 (Lenmakana)	14	50	1500	Fafanlap and Usaha Jaya	Tourism, invasive species, waste
Papua15				Tomolol and Usaha Jaya	Low threat
Papua5 (Karawapop)	3	<1	100	Yellu	Aquaculture, invasive species
Papua6				Yellu	Low threat

Figure 5: Tourism impact in Raja Ampat, Indonesia. A) Tourism increase in Raja Ampat, based on number of tourist entrance tags sold per year from 2007 - 2018 (Raja Ampat Tourism Department). Lines are drawn for domestic

visitors (red circles), foreign visitors (blue triangles) and the total (black squares). B) Estimates of tourist visits, tenure of jellyfish lakes and perceived threats based on interviews with stakeholders in 2017.

The interviews in 2017 indicated that families from six villages were mentioned as customary owners of jellyfish lakes in Misool (Fig. 5B). In 2011, before tourists visited Misool regularly, there were only three villages/families that claimed customary tenure. Some owners collect an entrance fee from visitors to jellyfish lakes, and fees are sometimes also asked by nearby pearl farms. Interviewees identified a range of threats including tenure disputes, conversion of lakes into aquaculture ponds, unregulated tourism, mosquito repellent and sunscreen pollution, lack of garbage handling, building impermanent jetties, and introduced species such as turtles, catfish, and sea anemones (Fig. 5B). Encouragingly, several respondents indicated there is an informal Code of Conduct in place that currently most tourist guides adhere to. This Code of Conduct includes a no sunscreen, no fins and no diving policy for the lakes.

Discussion

This is the first study to combine evolutionary, ecological and social approaches to investigate jellyfish lakes, peripheral ecosystems that are increasingly important tourism assets across Indonesia. Focussing on jellyfish lakes in Raja Ampat, we found distinct evolutionary and ecological diversity, and indicate the need for specific conservation management plans to cope with the challenges of increasing, and largely unregulated, tourism as well as other stressors. Below, we elaborate on the variation in genetics, morphometrics and abundance of jellyfish, explore the social context of jellyfish lakes, and finally discuss management recommendations and broader implications for jellyfish lakes in Indonesia.

Genetic and morphological variation indicate isolated populations

Jellyfish populations in Raja Ampat appear to be strongly structured based on genetic and morphological differentiation (Fig. 2, 3, Supplemental Table 4). Five subspecies were described by Dawson (2005a) from marine lakes in Palau, of which only one was observed in the haplotypes from Raja Ampat (*M. papua etpisoni)*; all other haplotypes recorded in Raja Ampat were unique to the region and potentially represent new subspecies. Strong genetic and morphological structure has previously been shown for jellyfish and other marine lake taxa, such as fish and mussels (Dawson and Hamner, 2005; Gotoh et al., 2011; Becking et al., 2016; Swift et al., 2016; Arai et al., 2017; Maas et al., 2018). Despite finding similar degrees of genetic structure as previous studies, we observed fewer private haplotypes when compared to jellyfish populations from Palau (Dawson and Hamner, 2005; Swift et al., 2016). This may result from less strong *in situ* divergence or smaller populations, but could also indicate that marine lakes in Raja Ampat were colonized more recently.

Shifts from the oceanic towards the derived ecotype in marine lakes are thought to be unidirectional and result from local adaptation to lake conditions (Dawson and Hamner, 2005; Swift et al., 2016). Interestingly, we observed a reversed morphological shift in Papua4 over sampling years (2016-2018) from the derived ecotype towards the oceanic ecotype.

The shift may be the result of an influx of oceanic jellyfish, although this is not corroborated by genetic results. Alternatively, the shift represents a new bloom strobilated from resident polyps showing high morphological plasticity. A shift in morphology was also observed in Palau after the disappearance of jellyfish coinciding with the ENSO events of 1997-1998 (Dawson, 2005a). When the jellyfish reappeared they displayed a larger size and differently shaped terminal clubs, but eventually reverted to pre-ENSO derived morphologies. To our knowledge, the genotypes were not published, therefore it is unclear if jellyfish before and after the crash were of the same population. For jellyfish in Papua4, we observed no shift in haplotype diversity. Recorded morphological variation of jellyfish in marine lakes may be therefore the result of phenotypic plasticity and acclimatization to specific local environments rather than evolutionary adaptation. However, phenotypic plasticity may over time facilitate genetic adaptation (Waddington, 1942; Fierst, 2011; Valladares et al., 2014), and as each jellyfish lake environment is unique, protection specific to each lake should be put in place.

Population dynamics in marine lakes

Jellyfish abundance in marine lakes in Raja Ampat appears to fluctuate greatly, indicating that jellyfish dynamics in these lakes are more heterogeneous than previously observed for lakes in Palau and lake Kakaban in Indonesia (Muscatine et al., 1986; Tomascik and Mah, 1994; Hamner and Hamner, 1998). While jellyfish populations in Palau also show some fluctuation in abundance (Dawson et al., 2015), fluctuations in Raja Ampat appear to be more frequent and severe, with regular occurrences of complete disappearance of jellyfish. Jellyfish lakes in Raja Ampat are five-fold smaller than those in Palau or Kakaban, and have a lower observed total abundance of jellyfish (900,000 vs. ~7 million: Dawson et al., 2001; Cimino et al., 2018). This difference may be conducive to dramatic fluctuations. Larger lakes have more stable environments (Scheffer, 1997), resulting in more stable populations and a higher carrying capacity. Remarkably, even for the large lake Ongeim'l Tketau, severe jellyfish population crashes have been observed twice since 1998, related to temperature and salinity change following ENSO events (Dawson et al., 2001; Martin et al., 2006). In Palau, there is no quota system for the number of tourists allowed to visit the lake, which resulted in continued vistitation until the total disappearance of jellyfish (Dawson et al., 2001). Tourists possibly exacerbated the decline of jellyfish abundance in Palau or may have led to longer recovery time of populations. It is likely that smaller lakes in Raja Ampat are even less resilient to such external perturbations.

Beyond direct effects on medusae presence, effects of stressors on the benthic polyps likely lie at the basis of jellyfish dynamics. Polyp sensitivity to temperature, salinity, food availability or chemical pollution may influence strobilation patterns (Sugiura, 1965; Prieto et al., 2010; Chi et al., 2019). Furthermore, introduction of invasive species such as fish and sea anemones may result in increased predator presence or competition for space for the benthic polyps (Takao et al., 2014; Patris et al., 2019). Jellyfish dynamics are currently poorly understood (Pitt and Lucas, 2014; Dawson et al., 2015), and a precautionary approach should be taken to avoid detrimental effects of unregulated tourism. Carefully monitoring

population abundance in jellyfish lakes is imperative to look for signs of population decline and to better understand jellyfish dynamics in general.

Lake use, threats and management status

Tourism to Raja Ampat is increasing dramatically, from 998 visitors in 2007 to 28,896 in 2018, representing a 30-fold increase (Fig. 5). While tourism could promote attention to jellyfish lake conservation, currently it is mostly unregulated. There are several threats from tourism towards jellyfish populations in marine lakes. For example, main tourism operators in Raja Ampat tend to 'follow the jellyfish', where they simply switch to a different lake when their favored jellyfish lake does not contain jellyfish anymore. This could lead to a cascade of crashes of additional jellyfish populations. Another threat is the introduction of invasive species, intentionally (fish and turtles for aquaculture, animistic purposes or amusement) or unintentionally (e.g. sea anemones carried in on wetsuits (Patris et al., 2019)). Additionally, disputes in lake tenure rights complicate controlled use and management of the lakes. Full engagement of and participation by local tenure holders in determining the management context and local regulations for any marine ecosystem (including the jellyfish lakes) is imperative for ensuring successful marine conservation in West Papua (McLeod et al., 2009; Grantham et al., 2013). As many of the local tenure holders now have a direct economic stake in tourism development in Raja Ampat, they have a strong incentive for seeking improved management of the jellyfish lakes.

The decision of the Ministry of Marine Affairs and Fisheries (MMAF) to denominate the Raja Ampat MPA network as a "marine tourism park" meant that marine lakes are not included in MPA zonation and conservation (MMAF Ministerial Regulation No. 36, 2014). Fortunately, Indonesian law (No. 5/1990 and No. 31/2004) allows for revision of zonation plans for protected areas based on scientific evidence that indicates a need to better conserve a particular ecosystem. Based on the current and previous studies (Becking et al., 2011, 2015, 2016; Maas et al., 2018), all marine lakes (including jellyfish lakes) are marine ecosystems with marine species, which implies that any marine lake which falls within the boundaries of an MPA should be managed as part of that MPA. The current study provides a strong justification to revise the zonation of the Raja Ampat MPA network to extend protection to jellyfish lakes and include explicit regulations for their management within the Raja Ampat MPA zonation system. The increase in tourism pressure to Raja Ampat's jellyfish lakes and their apparent susceptibility to perturbations, combined with the growing importance of tourism income to local tenure holders, suggests an urgent need to revise the MPA zonation plan of Raja Ampat to ensure the long-term sustainability of tourism to these jellyfish lakes. Below, we emphasize specific policy recommendations that we believe are feasible to implement in the near future.

Recommendations conservation approach

Jellyfish lakes are important tourism assets across Indonesia, with known lakes in East Kalimantan, Central Sulawesi and West Papua. Tourists visit the lakes for the singular

experience of swimming among thousands of jellyfish. Jellyfish lakes of Raja Ampat harbor endemic genotypes and morphologies of jellyfish, including putative undescribed subspecies. Marine lakes in general have unique local environments and associated organisms (e.g. Becking et al., 2011, 2016; Gotoh et al., 2011; Swift et al., 2016; Maas et al., 2018), indicating they may provide refuge for ancestral diversity and harbor endemic biodiversity (Gibson et al., 2017). We therefore argue that jellyfish lakes should be viewed as individual management units. The rapid increase in tourism in Raja Ampat is likely already exerting significant stress on its jellyfish lakes, and increased access to remote areas may result in discovery of new jellyfish lakes. Adequate conservation policies need to be in place to efficiently protect known and potential new jellyfish lakes.

Our primary recommendations for improved jellyfish lake conservation in Raja Ampat are twofold. First, we suggest to revise the current Raja Ampat MPA zonation system to explicitly include marine lakes as a sub-zone type within the "Other Zone Types" category (Mustaghfirin et al., 2012). We suggest including all marine lakes in this zonation, as all marine lakes represent unique peripheral ecosystems, and it is moreover possible that marine lakes without current jellyfish populations may be colonized by jellyfish in the future. Designation of a specific marine lake sub-zone type will thereby allow for the formulation of specific regulations for this sub-zone. We suggest that regulations for the marine lake sub-zone should include prohibiting scuba diving, prohibiting sunscreen and mosquito repellent use, requiring the removal of fins and booties, mandatory rinsing of clothing/wetsuits before entering the lake to avoid species introduction, and building permanent jetties as a single-entry point. Formalizing the Jellyfish Lake Code of Conduct which is currently used by tourism operators within the management plan will facilitate the acceptance of these regulations. Additionally, numbers of tourists should be limited in any lake at a given time. The actual limits or carrying capacity for each lake will require a productive and respectful collaborative determination including researchers, local tenure holders and marine tourism operators. Furthermore, we recommend that at least some lakes be zoned as 'core zones', strictly off limits to any use other than scientific research, in order that they can serve as reference points for lakes that are managed for tourism. Designating such core zone lakes will again require negotation with local tenure holders, with a likely focus on designating those lakes that are most difficult to access and therefore not currently used by tourism as core zones.

Secondly, we strongly urge the establishment of a collaboration between indigenous communities, local tenure holders, the Raja Ampat MPA Management Authority and Raja Ampat Tourism Department, and conservation NGOs and researchers to develop a specific Raja Ampat Jellyfish Lake Management Plan. By including stakeholders of the indigenous communities, local socio-economic needs can be met in a manner that also ensures adequate conservation. The plan should detail regular monitoring and adaptive management of the lakes to ensure they stay healthy. Monitoring jellyfish abundance and water quality parameters should be done regularly to record trends in abundance fluctuations. A multi-

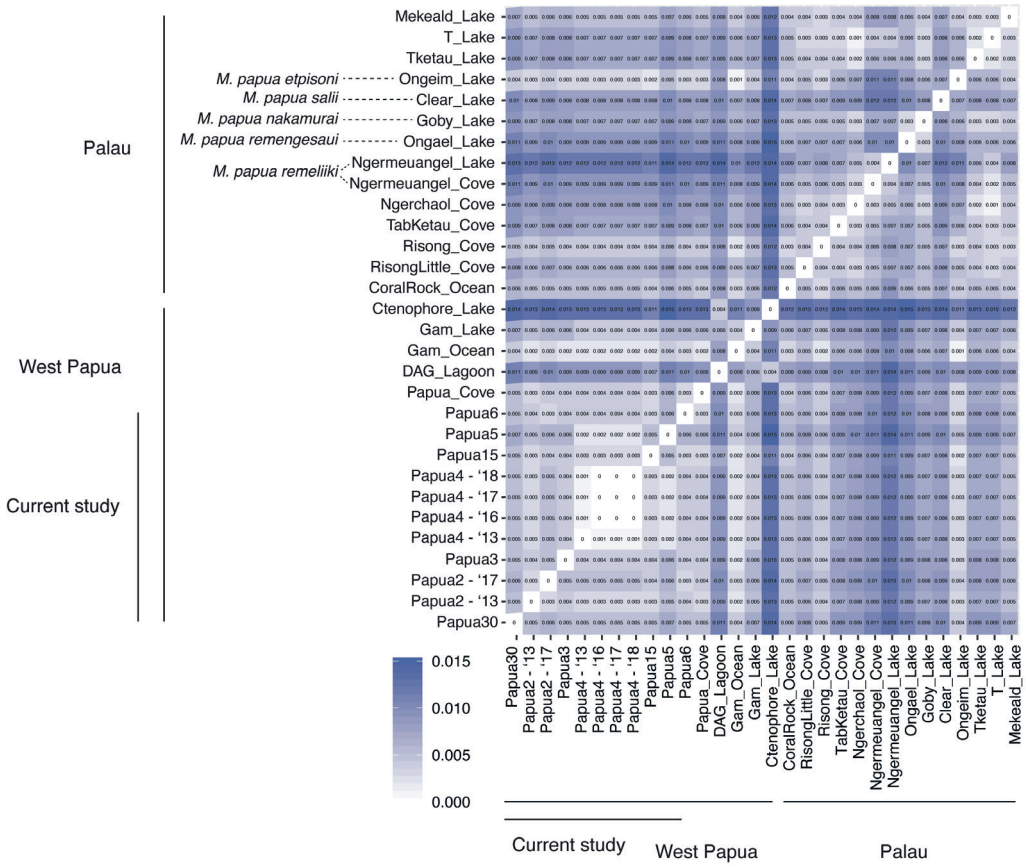
stakeholder working group could be set up including governmental agencies, local tenure holders, tourism operators, NGOs and experts. Such a working group was successfully established in Raja Ampat for the management of manta ray tourism (Kasmidi and Gunadharma, 2017), and jellyfish lakes could follow this model. The working group could meet at least twice yearly to review monitoring results and make any required decisions on tourism management, including potentially (temporary) closure of lakes when jellyfish populations begin showing signs of perturbation and decline. Ideally and as a precautionary measure, consideration should be given to a system of rotation of lakes being open to the public and others being closed. Importantly, we note that the recent designation of Raja Ampat as a National Geopark and the government's current bid to gain UNESCO Geopark status could provide a strong impetus for the development of a Raja Ampat Jellyfish Lake Management Plan, given the focus of the Geopark on unique karst habitats of Raja Ampat (of which the jellyfish lakes are certainly one).

We believe these practical applications are feasible at the local level of Raja Ampat and can moreover serve as an excellent model for other areas in Indonesia. Our approach can be followed by conservation managers generally when reconsidering the conservation status of vulnerable peripheral ecosystems, particularly where tourism is currently on the rise or expected to increase considerably in the future. Beyond jellyfish lakes and other marine lakes, effective protection for peripheral ecosystems is mostly insufficient worldwide (Mercado-Salas et al., 2013; Martínez et al., 2019; Masonjones et al., 2019). If not effectively managed, these rare ecosystems may be lost.

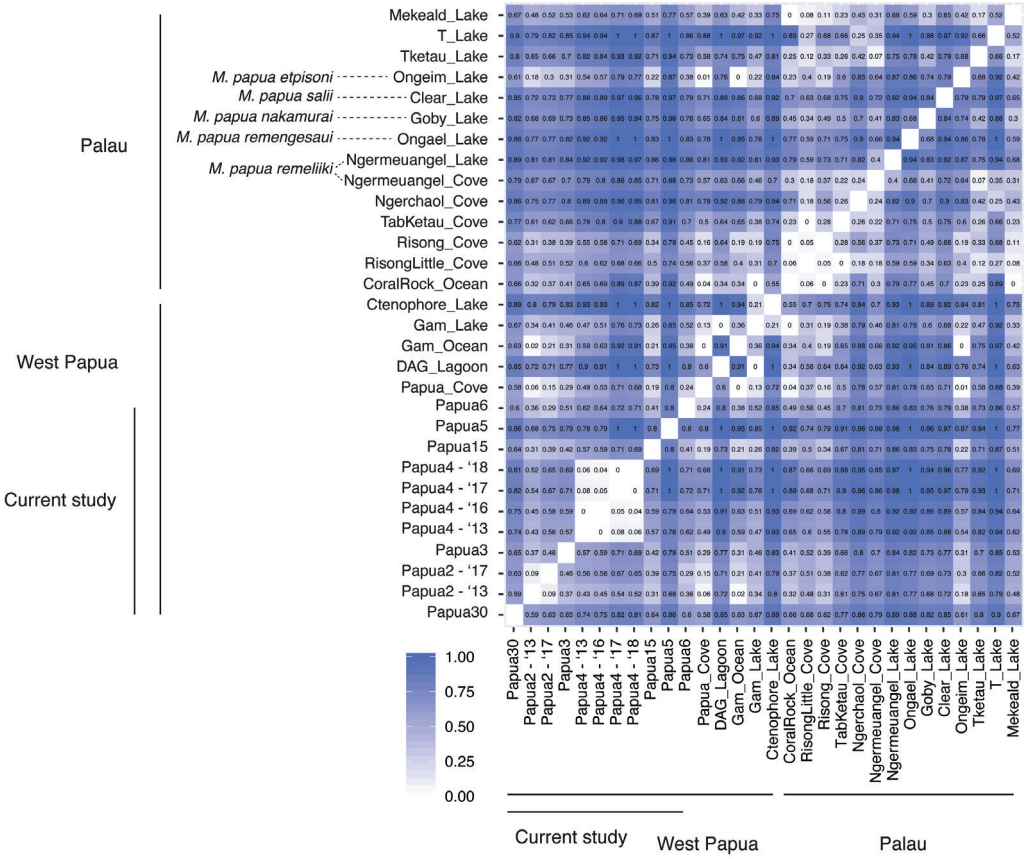
Acknowledgements
We would like to thank the following people who provided help with logistics and/or field work: M. Ammer, B. Hoeksema, A. Miners, T. Murk, G. Purba, Baseftin, and the staff of Misool Eco Resort and Papua Diving. We would like to thank M. Roest for doing preliminary analyses of jellyfish abundance and morphology of 2018.

This study was funded by the Netherlands Organisation for Scientific Research provided to LEB through grant VENI# 863.14.020. Field work in Indonesia was made possible through additional financial support of the National Geographic Waitt grant, National Geographic Young Explorers grant, Rufford Foundation, the Schure-Beijerinck-Popping fund of the Royal Dutch Academy of Science, and the Treub-Maatschappij fund. We are grateful to the Indonesian Institute of Sciences (LIPI) and the Indonesian State Ministry of Research and Technology and Higher Education (KEMENRISTEK-DIKTI) for providing research permits (098/SIP/FRP/SM/V/2011; #3B/TKPIPA/E5/Dit.KI/III/2016; 30/VI-P2SP2/UNIPA/2017; 56/SIP/FRP/E5/Dit.KI/III/2018).

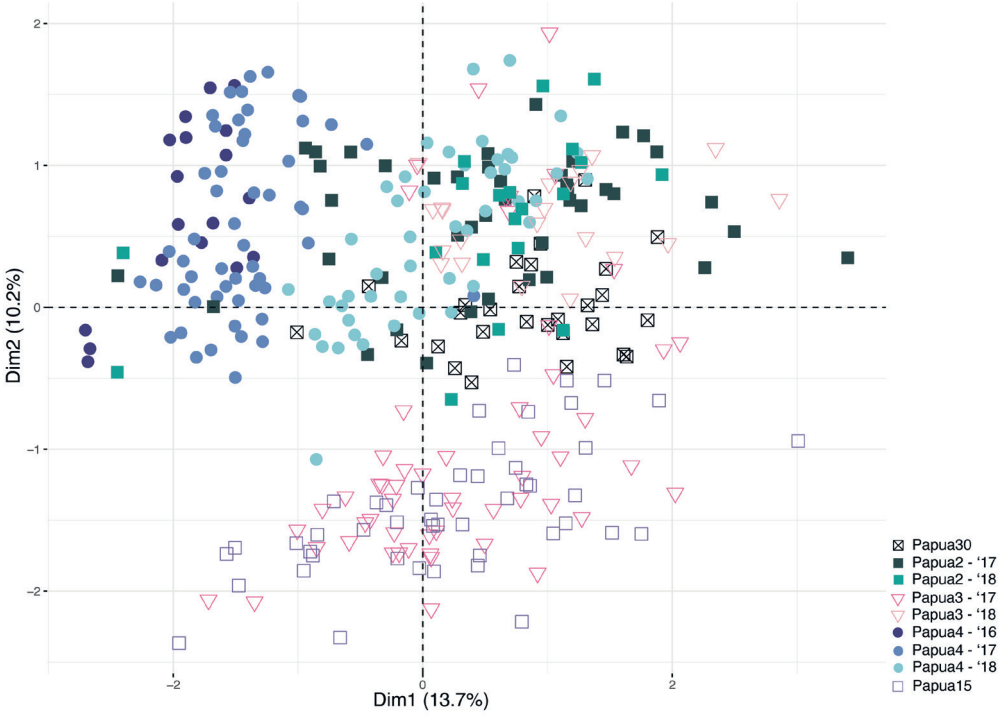
Supplemental Figures



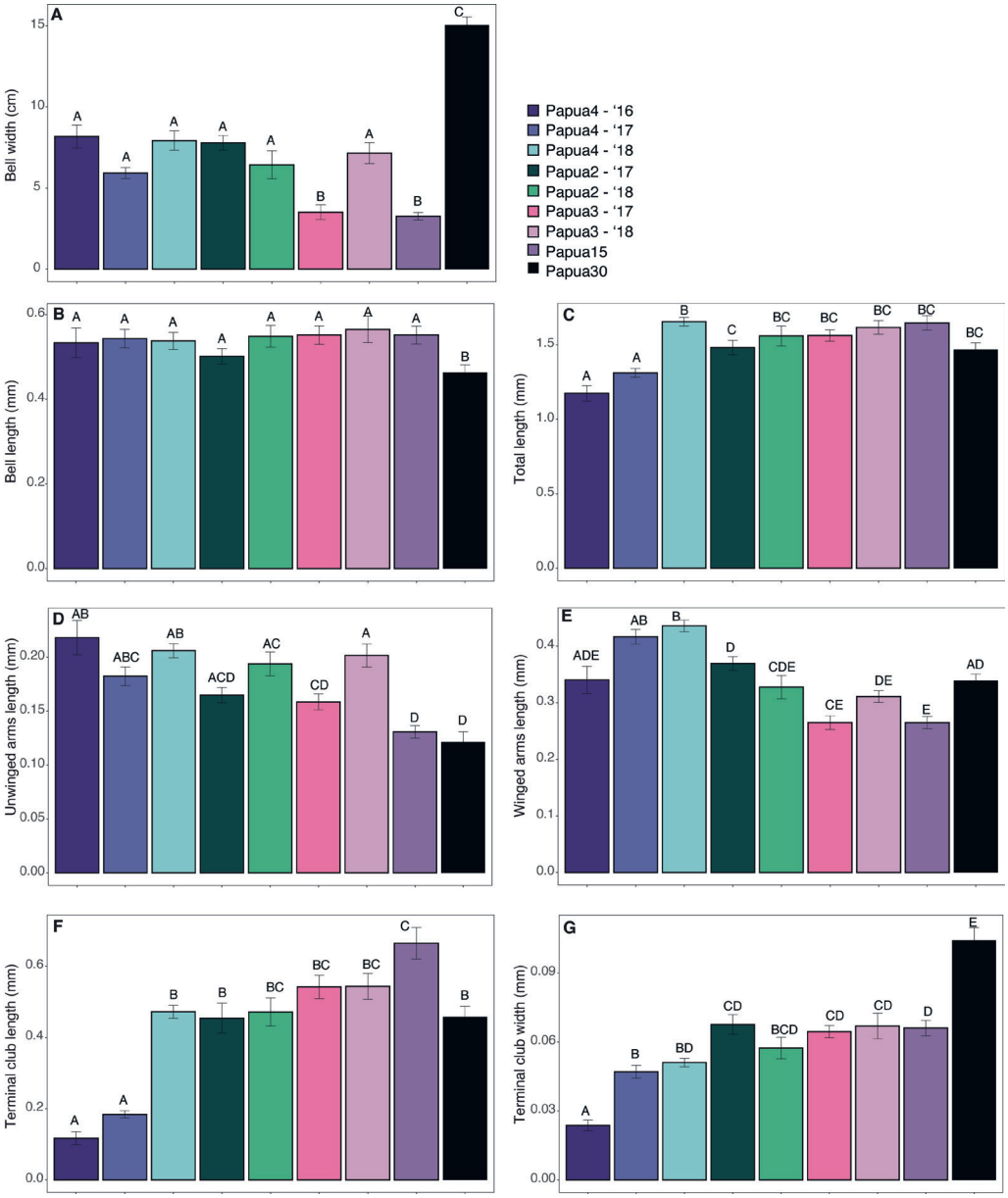
Supplemental Figure 1: Pairwise p-distances per population jellyfish from West Papua and Palau. Based on partial Cytochrome Oxidase I sequences of Mastigias papua sampled from marine lakes and ocean locations. Codes correspond to Fig. 1.



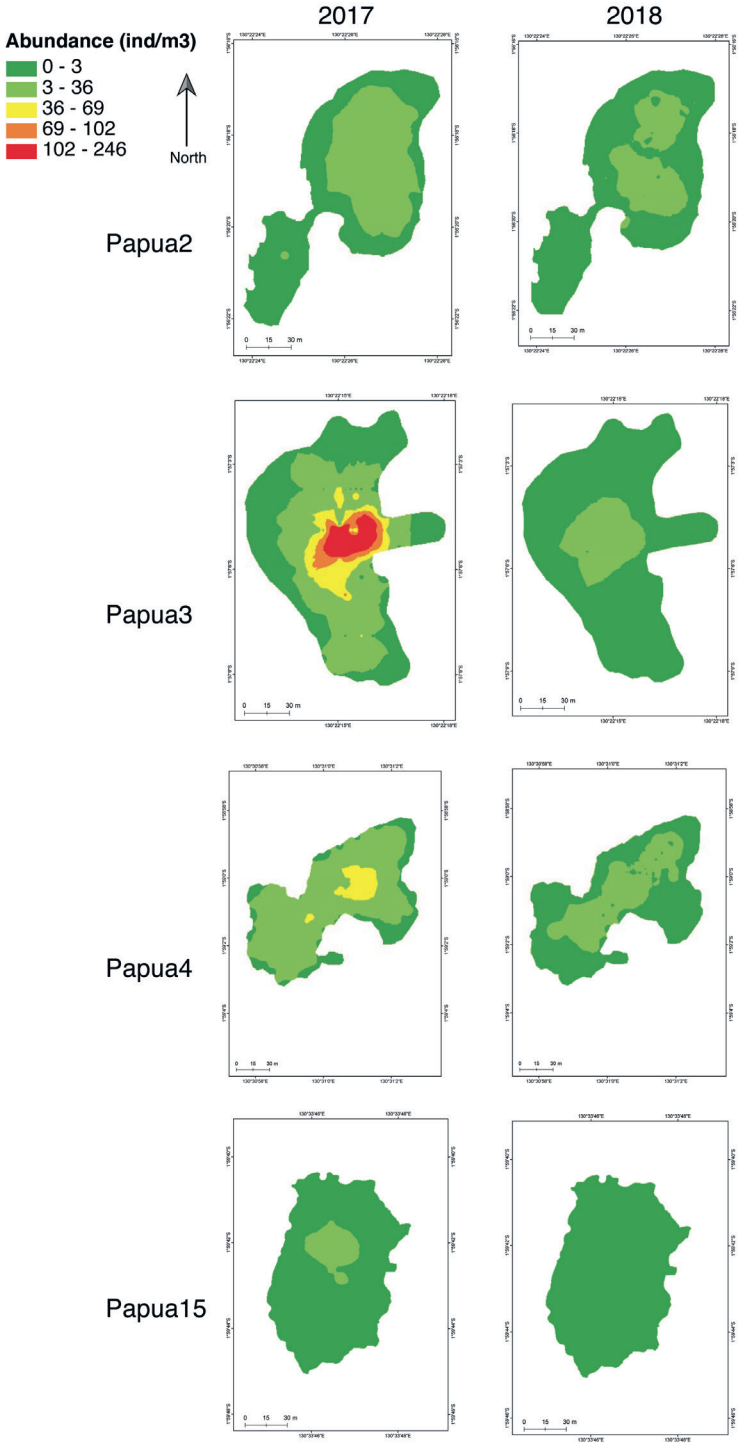
Supplemental Figure 2: Pairwise measurements of genetic differentiation (ϕ_{ST}) between locations from West Papua and Palau. Based on partial Cytochrome Oxidase I sequences of Mastigias papua sampled from marine lakes and ocean locations. Codes correspond to Fig. 1.



Supplemental Figure 3: Multiple Factor Analysis of quantitative and qualitative Mastigias papua morphometrics including all years sampled (2016-2018). In total, 24.4% of morphological variation is explained by first two dimensions. First dimension consists of 68% quantitative and 32% qualitative measurements, second dimension consists of 67% qualitative and 33% quantitative measurements. Colors and codes correspond with Fig. 1.



Supplemental Figure 4: Barplots of quantitative morphological characteristics of Mastigias papua. Included are A) bell width (f9 from Dawson 2005a, cm), and as fractions of the bell width: B) bell length, C) total length, D) unwinged arms length (f11), E) winged arms length (f12), F) terminal club length (f14), G) terminal club width. Significant differences amongst groups tested via Kruskal-Wallis and Dunn tests. Different letters indicate significant differences in morphology ($p << 0.001$). Codes correspond to Fig. 1.



Supplemental Figure 5: Density maps of Mastigias papua abundance recorded in four marine lakes in two consecutive years. Scale bar indicates the number of jellyfish counted in individual per m³. Lakes were sampled in May in both years, between 9:00 - 10:00 am.

Supplemental Table 1: Summary of lake and sample characteristics. Per marine lake the following was recorded: temperature (°C), salinity (ppt), pH, lake depth (m), connection to the sea (tidal amplitude lake divided by tidal amplitude sea), number of samples analyzed for genetics and morphometrics, nucleotide diversity (π, including standard deviation SD), haplotype diversity (h, including standard deviation) and Tajima's D. Environmental data averaged for the first 5m of each lake, including standard deviation. Connection was defined as the maximum tidal amplitude (m) of the lake divided by the maximum tidal amplitude of the sea.

Code	Environment			Physical		# Samples				Nucleotide diversity				Haplotype diversity				Tajima's D			
	Temp (°C)	Sal (ppt)	pH	Depth (m)	Conn (frac.)	Genetics/Morphometrics				π (%, SD)				*h* (SD)				* indicates significance ($p < 0.05$)			
						2013	2016	2017	2018	2013	2016	2017	2018	2013	2016	2017	2018	2013	2016	2017	2018
P 30	32.4 (0.2)	28.9 (0.3)	7.6 (0.05)	4.1	0.75	- / -	19 / 29	- / -	- / -	-	.162 (.13)	-	-	-	.298 (.13)	-	-	-	-1.20	-	-
P 2	33.6 (0.6)	25.2 (0.5)	7.8 (0.08)	7.3	0.23	22 / -	- / -	15 / 43	- / 20	.267 (.19)	-	.296 (.21)	-	.723 (.09)	-	.533 (.04)	-	-.38	-	2.18*	-
P 3	32.5 (0.5)	26.7 (0.9)	7.8 (0.07)	7.5	0.51	- / -	- / -	15 / 55	- / 21	-	-	.225 (.17)	-	-	-	.562 (.09)	-	-	-	-.03	-
P 4	31.7 (0.4)	25.9 (0.8)	8.1 (0.06)	20.4	0.80	14 / -	16 / 18	14 / 53	12 / 47	.098 (.10)	.086 (.08)	.000	.000	.264 (.14)	.233 (.13)	.000	.000	-.44	-.58	-	-
P 15	30.5 (0.1)	30.3 (0.1)	8.0 (0.02)	33.9	0.89	- / -	- / -	15 / 48	- / -	-	-	.190 (.15)	-	-	-	.686 (.10)	-	-	-	.34	-
P 5	31.5 (0.3)	28.9 (0.3)	8.1 (0.05)	4.8	0.26	12 / -	- / -	- / -	- / -	.000	-	-	-	.000	-	-	-	-	-	-	-
P 6	31.9 (0.3)	28.3 (0.3)	7.9 (0.03)	12.4	0.87	- / -	20 / -	- / -	- / -	-	.198 (.15)	-	-	-	.353 (.12)	-	-	-	-.73	-	-

Supplemental Table 2: List of informants and dates of interviews.

Informant's code	Informant's group	Affiliation	Interview Date
GOV1	Government	UPTD KKPD Raja Ampat	03-Nov-16
GOV2		Marine and Fisheries Department	14-Dec-16
GOV3		Culture and Tourism Department	08-Jun-17
GOV4		Marine and Fisheries Department	14-Jun-17
GOV5		Marine and Fisheries Department	14-Jun-17
GOV6		Culture and Tourism Department	13-Jun-17
VGO1	Village Government	Head of Harapan Jaya Village	08-Nov-16
VGO2		Secretary of Tomolol Village	10-Nov-16
VGO3		Head of Tomolol Village	11-Nov-16
VGO4		Head of Consultative Council of Usaha Jaya Village	12-Nov-16
VGO5		Head of Neighborhood Cluster of Fafanlap Village	16-Nov-16
ING1	NGO	Conservation International	03-Nov-16
ING2		The Nature Conservancy	02-Jan-17
ING3		The Nature Conservancy	15-Dec-16
ING4		Misool Baseftin	23-Nov-16
OCR1	Owner of Customary's Right	-	11-Nov-16
OCR2		-	12-Nov-16
OCR3		-	16-Nov-16
OCR4		-	16-Nov-16
OCR5		-	02-Dec-16
TOA1	Tourism Player (tour guide/homestay owner/dive center manager/resort owner/guest operation based in Misool)	Local Guide	07-Nov-16
TOA2		Dive Centre Manager of Harfat Jaya Homestay	08-Nov-16
TOA3		Local Guide	16-Nov-16
TOA4		Local Guide	16-Nov-16
TOA5		Local Guide	20-Nov-16
TOA6		Manager of Misool Eco-Resort	21-Nov-16
TOA7		Owner of Panun Paradise Homestay	02-Dec-16
TOA8		Guest Operation Manager of Misool Eco-Resort	21-Nov-16
TOA9		Local Guide of Nut Tonton Homestay	11-May-17
TOA10		Manager of Nut Tonton Homesta	11-May-17
TOA11		Manager of Harafat Jaya Homestay	12-May-17
TOA12		Local Guide	13-May-17
TOA13		Local Guide	13-May-17
TOA14		Local Guide	13-May-17
TOA15		Owner of Jannah Homestay	14-May-17
TOA16		Security of Pearl Company	16-May-17
TOA17		Manager of Sisie Homestay	17-May-17
TOA18		Owner of Ya'Lapale Homestay	17-May-17
TOA19		Owner of Yaganan Homestay	18-May-17
TOA20		Owner of Harafat Jaya Homestay	20-May-17
TOA21		Owner of Lalelkai Homestay	22-May-17
TOA22		Local Guide	19-May-17
LIA1	Liveaboards	Captain of Amanikan Yacht	17-Dec-16
LIA2		Captain of Full Moon Yacht	17-Dec-16
LIA3		Captain of El-Aleph Yacht	17-Dec-16
LIA4		Captain Tamukira	26-May-17
LIA5		Captain Waow	28-May-17
LIA6		Captain Pearl of Papua	29-May-17
LIA7		Cruise Director Pearl of Papua	03-Jun-17
LIA8		Captain Shakti	01-Jun-17
LIA9		Captain Assistant Safari 8	03-Jun-17
LIA10		Captain Mutiara Laut	16-Jun-17
LIA11		Captain and Cruise Director Mantamae	17-Jun-17
LIA12		Owner Kurabesi	22-May-17
COM1	Local Community	Local Community in Fafanlap	19-May-17

Supplemental Table 3: Variable positions of COI for Mastigias papua populations from marine lakes. Neotype of Mastigias papua as defined in Souza & Dawson (2018) contrasted to all other haplotypes. Subspecies as defined by Dawson (2005a) are also shown. Polymorphisms within populations are indicated by a slash.

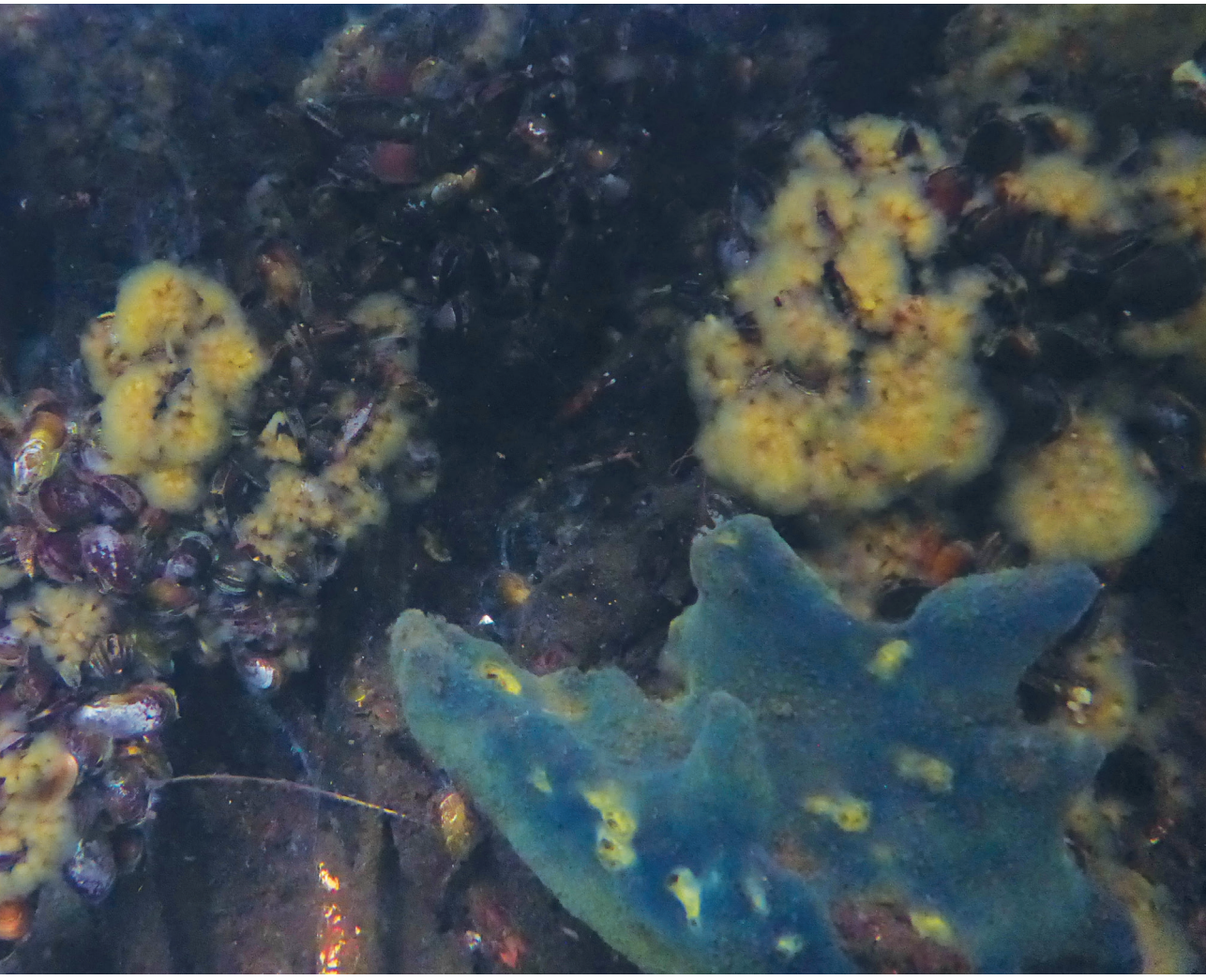
												1	1	1	2	2	2	2	2	3	3	4	4	4	4	5	5	5	5
			1	2	2	4	5	6	6	6	7	6	6	5	1	5	6	8	9	3	6	2	2	7	9	1	1	2	3
		4	6	1	5	6	4	3	6	9	2	2	5	9	9	2	4	8	2	9	4	0	5	1	0	0	9	5	5
Neotype specimen M. papua		G	A	C	T	T	A	C	C	C	T	A	A	A	G	T	G	G	G	T	C	C	G	T	C	A	T	T	T
remengesaui		.	.	.	.	.	G	.	.	.	.	.	G	.	.	C	.	.	.	.	.	.	.	.	T	T	.	.	.
etpisoni		.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
salii		.	.	.	.	.	G	T	.	C/T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C
nakamurai		.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	T	T	.	.	.
remeliiki		.	.	.	.	.	G	.	.	.	.	.	G	G	.	.	.	.	.	.	.	T	G/A	T/C	.	T	C	.	.
Hap. 1	Papua5	.	G	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
Hap. 4+5+7	Papua2	G/A	.	.	.	.	.	.	C/T	.	.	.	.	.	.	.	.	.	.	T/C	.	.	.	.	.	T	.	T/C	.
Hap. 9	Papua3	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	A	A	.	.	.	.	.	.	.	T	.	.	.
Hap. 10+12	Papua15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G/A	.	C/T	.	.	.	.	T	.	.	.
Hap. 13	Papua6	.	.	.	.	C	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
Hap. 2	Papua4, 5, 30	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
Hap. 3	Papua4, 3, 30	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
Hap. 6	Papua2, 3, *etpi.*	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
Hap. 8	Papua2, Papua3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.
Hap. 11	Papua15, 6, 30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	.	.	.
Hap. 14	Papua6, 30	.	.	.	.	.	.	.	.	.	C	.	.	.	A	.	.	.	.	.	.	.	.	.	.	T	.	.	.

Supplemental Table 4: Population genetic structure among populations of Mastigias papua from marine lakes in Raja Ampat. Pairwise comparisons of genetic differentiation (ϕ_{ST}) based on partial Cytochrome Oxidase I sequences (542bp). Asterisks (*) indicate significance (p-value < 0.001, based on 1,000 permutations). Year sampled provided after location code, and numbers between brackets indicate quantity individuals sampled. Pairwise comparisons show all lakes are significantly different from each other with no significant variation between sampling years within a lake.

	Papua30 2016 (19)	**Papua2 2013 (22)**	**Papua2 2017 (15)**	**Papua3 2017 (15)**	**Papua4 2013 (14)**	**Papua4 2016 (16)**	**Papua4 2017 (14)**	**Papua4 2018 (12)**	**Papua15 2016 (15)**	**Papua5 2013 (12)**
Papua 2 2013	0.590*									
Papua 2 2017	0.629*	0.094								
Papua 3 2017	0.649*	0.368*	0.457*							
Papua 4 2013	0.736*	0.427*	0.562*	0.569*						
Papua 4 2016	0.749*	0.449*	0.585*	0.594*	0.070					
Papua 4 2017	0.818*	0.536*	0.665*	0.707*	0.077	0.055				
Papua 4 2018	0.808*	0.519*	0.646*	0.690*	0.061	0.041	0.000			
Papua15 2016	0.640*	0.308*	0.385*	0.420*	0.566*	0.589*	0.707*	0.690*		
Papua 5 2013	0.858*	0.682*	0.747*	0.788*	0.781*	0.790*	1.000*	1.000*	0.799*	
Papua 6 2016 (20)	0.601*	0.360*	0.290*	0.511*	0.620*	0.638*	0.721*	0.707*	0.410*	0.795*

Supplemental Table 5: Significance among morphologies of marine lake medusae in Raja Ampat. Based on morphological characteristics f1,f4-9, and f11-14, as defined in Dawson, 2005. Tested by multivariate analysis of variance (adonis) in R. Adjusted p-values displayed below the diagonal and F model above the diagonal.

	Papua 30	**Papua 2**	**Papua 3**	**Papua 4**	**Papua 15**
Papua 30	-	11.82	14.82	39.13	11.54
Papua 2	0.01*	-	6.42	12.84	10.45
Papua 3	0.01*	0.01*	-	41.02	3.62
Papua 4	0.01*	0.01*	0.01*	-	46.35
Papua 15	0.01*	0.01*	0.28	0.01*	-



Chapter 4

New evidence for genetic structure in the sponge *Suberites diversicolor*: Implications for sponge phylogeography and population genetics

Diede L. Maas, Stefan Prost, Christiaan A. de Leeuw, Ke Bi, Lydia Smith, Purwanto, Ludi P. Aji, Ricardo F. Tapilatu, Rosemary G. Gillespie, Leontine E. Becking

Submitted

Abstract

The relative influence of geography, oceanography and environment on gene flow within sessile marine species remains an open question. Detecting subtle genetic differentiation is challenging in benthic populations due to large effective population sizes, elusive barriers to dispersal and general lack of resolution in genetic markers. Marine lakes can circumvent confounding factors in the open ocean by providing discrete and replicated model systems. Building on previous work on the marine lake sponge *Suberites diversicolor* using single marker (*COI* and *ITS*), we genotyped our samples using high resolution double digest restriction-site associated DNA sequencing (4,826 Single Nucleotide Polymorphisms, SNPs). We sampled *S. diversicolor* at different spatial scales (1-1,400 km), along a gradient of connection to the surrounding sea, and with different environmental regimes in order to test their relative importance on population genomic structure. While both low- and high-resolution genetic markers detected two major lineages and geographic clustering over large spatial scales, with the SNP dataset we provide new evidence of strong population structure within a lineage, even at scales <10km (average F_{ST} = 0.56). Within one lineage, most variation was explained by variation between populations (AMOVA: 48.8%). We observed signatures of population bottlenecks. We did not detect strong effects of geographic distance, permeability of seascape barriers or local environments in shaping population genomic structure, suggesting other drivers such as priority effects could play a role. Our results call for a reassessment of poorly dispersing benthic organisms that were previously assumed to be highly connected based on low resolution markers.

Keywords: RADseq, marine biodiversity, Porifera, genetic resolution, population genomics, seascape genetics

Introduction

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

Sponges, integral but often underappreciated assets of benthic communities (Bell, 2008; De Goeij et al., 2013; Webster and Thomas, 2016), are generally considered to be poor dispersers as their larvae have limited swimming capacity and are short-lived (Maldonado, 2006). With the assumption that sponges are poor dispersers, it could be expected that they show patterns of strong genetic divergence over relatively small geographic ranges. Sponges are therefore excellent candidates to investigate marine population genetic structure on small scales. The majority of studies investigating genetic structure in sponges have revealed species complexes with divergence among morphologically cryptic lineages (van Oppen et al., 2002; Uriz and Turon, 2012; Pérez-Portela and Riesgo, 2018), yet studies investigating within-lineage divergence are scarce.

That most sponge genetic studies remain at the species level instead of investigating structure within lineages may have to do with the low resolution of commonly used genetic markers (Selkoe et al., 2016; Timm, 2020). Molecular markers used to assess sponge phylogeography and population structure include mitochondrial markers (mtDNA) such as Cyochrome c oxidase I (*COI*) and ATP6, and nuclear markers such as introns, internal transcribed spacers (*ITS*) (Wörheide et al., 2008) and microsatellites (van Oppen et al., 2002; Uriz and Turon, 2012; Pérez-Portela and Riesgo, 2018). Though widely used in phylogeographic and population genetic studies (Avise, 2000, 2009), mitochondrial markers exhibit low mutation rates in sponges (Wörheide et al., 2005; Huang et al., 2008). As a result, the majority of studies using mtDNA find panmixia among sponge populations across broad geographic ranges (e.g. Duran, Pascual and Turon, 2004; Whalan *et al.*, 2008; De Bakker *et al.*, 2016; Ekins *et al.*, 2016). In contrast, *ITS* markers can show more structure (Bentlage and Wörheide, 2007; Becking et al., 2013b; Ekins et al., 2016), but generally at large spatial scales, and *ITS* markers are hampered by intra-genomic polymorphisms (Frankham et al., 2002). Finally, microsatellites could be reliable and sufficiently variable to detect population structure, yet are time-consuming to design *de novo* for each species (Frankham et al., 2002; Pérez-Portela and Riesgo, 2018), can be confounded by homogenizing forces of evolution (van Oppen et al., 2002), and generally relatively few markers have been used per study (<20). An increase in number of molecular markers is expected to advance inferences on demography and structure (Allendorf et al., 2010; Kelley et al., 2016; Pérez-Portela and

Riesgo, 2018), allowing researchers to reassess assumptions of panmixia within sponge lineages at fine spatial scales.

Recently, there has been an increase in the use of reduced representation genomic methods and Single Nucleotide Polymorphisms (SNPs) for population studies on non-model organisms (Baird et al., 2008; Peterson et al., 2012; Puritz et al., 2014; Catchen et al., 2017). The additional power of an increased marker panel has been demonstrated in for example mussels (Becking et al., 2016; Maas et al., 2018; de Leeuw et al., 2020), and fish (Bradbury et al., 2015; Lemopoulos et al., 2019; D'Aloia et al., 2020; Sunde et al., 2020). However, high resolution studies on sponges are lagging behind (Pérez-Portela and Riesgo, 2018), with notable exception of Brown et al. (2014), Brown et al. (2017) and Leiva et al. (2019). However, these studies still used a limited marker panel (<400 SNPs) and showed limited differentiation. Using restriction site-associated DNA (RADseq) sequencing techniques such as ddRAD (double digest RADseq, Peterson et al., 2012) may increase the number of retained SNPs to thousands and provide the necessary resolution.

Marine lakes provide ideal model systems for population genetic and genomic studies since they are well-defined and provide clear boundaries to define populations, similar to other island-like systems (Warren et al., 2015). Marine lakes are bodies of seawater surrounded completely by land, but which maintain a connection with the surrounding sea through caves or porous rock (Holthuis, 1973; Hamner et al., 1982; Dawson et al., 2009; Becking et al., 2011). Aggregations of marine lakes are present in the Caribbean, Vietnam, Palau, and Indonesia, particularly in East Kalimantan and in West Papua (Dawson et al., 2009; Becking et al., 2011, 2015). Marine lakes were formed *de novo* when depressions in karstic rock were filled with sea water after the Last Glacial Maximum (approximately 12,000 years ago) (Tomascik and Mah, 1994; Sathiamurthy and Voris, 2006), and house clearly defined populations (Gotoh et al., 2011; Itescu, 2018). Having originated roughly at the same time, marine lakes represent relatively controlled biotopes where each lake can be seen as an independent replicate of eco-evolutionary dynamics over time.

Sponges are usually well-represented in marine lakes, both in terms of diversity and biomass (Azzini et al., 2007; Becking et al., 2011, 2013b; Cleary et al., 2013). The sponge *Suberites diversicolor* (Porifera, Demospongiae, Suberitidae, Becking and Lim, 2009) has been found to occur extensively in Indonesian marine lakes and brackish coastal areas (Becking and Lim, 2009; Cleary et al., 2013). Using *COI* and *ITS* genetic markers, Becking et al. (2013) studied the phylogeography of *S. diversicolor* from multiple marine lakes and lagoon populations in the Indo-Pacific. They identified two distinct genetic lineages (Lineage A and B) and regional structuring, yet did not observe subtle levels of structuring at smaller spatial scales. The lack of structure could be explained by recurrent gene flow among lakes, or by lack of resolution of genetic markers used by Becking et al. (2013), as they recovered a low number of haplotypes (4 for *ITS* and 3 for *COI*). Clearly, there is a need to further elucidate population genetic patterns.

Studying *a priori* defined sponge populations from nine marine lakes and two lagoon locations in Indonesia (East-Kalimantan and West-Papua) and Australia, we aim to assess the population structure in *S. diversicolor* and associated drivers. Selecting marine lakes on different spatial scales (1-1,400km), along a gradient of connection to the surrounding sea and with different environmental regimes allows the opportunity to assess effects of geographic distance, permeability of barriers and local environments in shaping genetic structure. In order to assess the effect of level of genetic resolution, we compared results of our genome-wide sequencing strategy (double-digest restriction-site associated DNA sequencing, (ddRAD, Peterson et al., 2012)) to previously published results on the same individuals using single markers (*COI* and *ITS*) (Becking et al., 2013b). We expect with the higher resolution gained from using a RADseq approach we will be able to observe fine scale population structure and provide first insights into drivers underlying structure.

Material and methods

Sample collection and lake profiling

Tissue samples (~1cm^3) were collected from 168 individuals of *Suberites diversicolor* (Fig 1, Table 1). One lagoon was sampled in Darwin, Australia (DAR), one lagoon and three marine lakes were sampled in East-Kalimantan (Bay, Kalimantan 1, Kalimantan 2 and Kalimantan 3), and six marine lakes were sampled in West-Papua (Papua 27, Papua 30, Papua 32, Papua 1, Papua 4 and Papua 5). As the locations have no official names instead they received coding 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) (Supplemental Table 1 for corresponding lake codes). Samples were collected between 1-5m depth while snorkeling. Some lakes had very low densities of *S. diversicolor*, therefore sample sizes were lower (see Becking et al., 2013 Table 2 for densities). In the field, tissue samples were immediately preserved in 99% ethanol or RNAlater after excision at 0-4°C (4-8 weeks), and upon returning to the laboratory stored in a -20°C freezer until further use.

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

Table 1: Overview of sampling in marine lakes and lagoon locations. Recorded are location, site codes, number of individuals sampled per site and number of individuals retained after filtering, physiographic, environmental and genetic parameters. 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. the calculation of population genomic parameters (nucleotide diversity and heterozygosity) only localities containing samples from Lineage B were considered.

Lake_code	Location	# specimens Total set	# specimens After filtering	Lake_Area	Lake_depth	Connection fraction	Connection category	Temperature	Salinity	Nucleotide diversity	Heterozygosity (He)
DAR	Australia	8	7	45,640			open			0.0095	0.117
Bay	Berau	5	5				open	29	33.5	0.0101	0.157
Kalimantan 1	Berau: Kakaban	32	20	4,900,000	12	0.11	low	30	23.5		
Kalimantan 2	Berau: Tanah Banban	4	2	231,500		0.38	low	29.5	26	0.0074	0.034
Kalimantan 3	Berau: Maratua	29	26	140,000	17	0.51	medium	29.5	27	0.0050	0.081
Papua 27	Papua: Wayag	8	5	22,000	2		medium	29.5	31	0.0037	0.038
Papua 30	Papua: Wayag	9	8	13,000	4.1	0.75	medium	32.4	28.9	0.0045	0.052
Papua 32	Papua: Wayag	7	4	6,100	5.5	0.45	medium	31.2	30.7	0.0053	0.059
Papua 1	Papua: Gam	20	11	88,530	19	0.07	low	32.3	24	0.0036	0.054
Papua 4	Papua: Misool	26	19	13,750	20.4	0.8	high	31.7	25.9	0.0047	0.08
Papua 5	Papua: Misool	20	18	3,700	4.8	0.26	low	31.5	28.9	0.0060	0.095

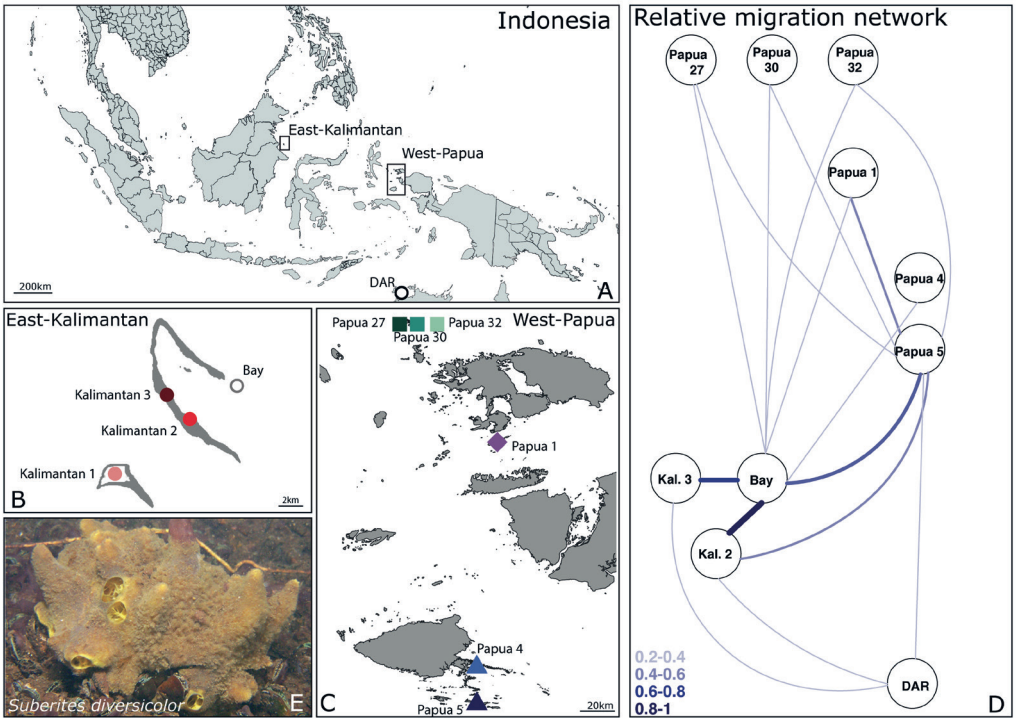


Figure 1: Sampling sites of Suberites diversicolor from nine marine lakes and two lagoon locations and associated relative migration networks. (A) Overview of Indonesia including two geographic regions sampled: Berau and Raja Ampat. Also shows location of Australian lagoon location (DAR). (B) Berau, East-Kalimantan with locations of three marine lakes (B.1, B.2, B.3) and one lagoon (Bay). (C) Raja Ampat, West-Papua with locations of six marine lakes (P.27, P.30, P.32, P.1, P.4, P.5). (D) Relative migration network including only samples from Lineage B run with 1000 bootstraps. Fractions of relative migration are displayed. (E) Specimen of S. diversicolor, photograph by L.E. Becking.

DNA extraction, library preparation and sequencing

DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen), with the only modification from manufacturer instructions being an extended lysis time (overnight). DNA quality and quantity were assessed using 1.5% agarose gels and Qubit dsDNA HS assays. Next, ddRAD libraries were prepared following the publicly available protocol published in Maas *et al.*, (2018), adapted from the original protocol of Peterson et al., (2012). We refer to the extensive protocol included in Maas et al., (2018) for details, but describe here how we adapted it for *S. diversicolor*. In brief, genomic DNA (600ng) was double-digested using enzymes SphI-HF (rare-cutting) and MluCI (frequent-cutting) (See Supplementary Information A 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 genome size of ~300Mb (common for sponges (Srivastava et al., 2010; Jeffery et al., 2013) and various size selections of RAD fragments. This number can subsequently be used to calculate the expected coverage when generating a known amount (Gb) of sequencing data. Custom-made sample-specific barcodes were ligated to the fragments to allow for the pooling of 21

samples per library, resulting in 8 libraries in total. The Sage Science Pippin Prep was used to size-select adapter-ligated fragments of length 500-575bp (indicating an insert size of 425-500bp). A trial was run for 8, 10 and 12 polymerase chain reaction cycles (PCR) reactions. In the end ten PCR cycles were chosen as a balance between DNA output and PCR duplication and were run on each library for enrichment and ligation of Illumina indices unique to each library pool. Quality and quantity of libraries throughout the process were checked using BioAnalyzer High Sensitivity chips (Agilent, Supplemental Information B for an example). Libraries were pooled at equimolar volumes and 150bp single-end sequenced on Illumina HiSeq 2500 at the Vincent J. Coates Genomic Sequencing Facility at UC Berkeley.

Reference assembly, bioinformatic filtering and genotype calling

Custom perl scripts were used for processing the resulting sequences (RADTOOLKIT v. 0.13.10, made available in the 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 (Li and Godzik, 2006; Fu et al., 2012), with a minimum support per cluster set at three reads, and representative sequences retained for each cluster. RepeatMasker v4.0 (http://repeatmasker.org/) was used to mask putative repetitive elements, low complexity regions and short repeats (Smit et al., 2014). Loci were discarded if >60% of nucleotides per loci were Ns. The resulting RAD loci were combined for all individuals, and a reference was built from loci shared by at least 70% of individuals.

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 and Salzberg, 2014), a fast sequence classifier to BLAST (Altschul et al., 1990) our loci against bacterial databases with default settings. Thirdly, we used BlobTools (Laetsch and Blaxter, 2017) to taxonomically partition reads and cut off loci with >55% GC content, as we expect sponge microbes to have higher GC content than sponge hosts (Horn et al., 2016). The identified microbial loci were filtered out using a custom made perl script (Bi et al., 2013).

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 10bp of an indel. Sites were removed if the depth was outside 1st and 99th

percentile of the overall coverage. Another custom perl script (SNPcleaner, github.com /tplinderoth/ ngsQC/ tree/ master/ snpCleaner (Bi et al., 2013, 2019)) was implemented for further filtering of SNPs. In the end, 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 and Slatkin, 2008; Lynch, 2008), which subsequently may bias downstream analyses. Therefore, we compared results from genotype calls and genotype likelihoods. Genotype likelihoods were generated via an empirical Bayesian framework via Analysis of Next-Generation Sequencing Data (ANGSD v.0.930) (Korneliussen, Albrechtsen & Nielsen, 2014). We set genotype posterior probabilities of 0.95 as a threshold in ANGSD to output high-confidence genotypes for analyses performed in GENODIVE v3.0 requiring genotype calls (Meirmans and Van Tienderen, 2004). For downstream analyses based on either genotype likelihoods and genotype calls we tested the effect of coverage (3X and 10X) and missing data included (max. 30%, 10%, 5% and 1% allowed missing data).

Detection of major lineages

We reconstructred phylogeographic structure among and within lineages via a maximum likelihood tree, a neighbor-joining tree. The maximum likelihood approach was done via genotype calling and the software IQ-Tree (Nguyen et al., 2015). First, we created consensus sequences in fasta format for all 124 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 55kb per individual, and carried out an alignment using MAFFT (Katoh and Standley, 2013). We then constructed the maximum likelihood phylogenetic tree using the IQ-Tree software with 1,000 ultrafast bootstraps and an SH-like approximate likelihood test for 1,000 replicates. The best fitting substitution model was inferred using the -m TEST function in IQ-Tree. The neighbor-joining tree was built via a genetic distance matrix computed using ngsDist using genotype likelihoods (Vieira et al., 2016). Following recommendations of RAxML (Stamatakis, 2014), a bootstrapped neighbor-joining tree matrix was computed from 1,000 possible trees and visualized as a phylogenetic tree using FASTME (Lefort et al., 2015) and FigTree v.1.4.2 (Rambaut, 2009).

Next, we explored admixture patterns using ngsAdmix (Skotte, Korneliussen, 2013). Ancestry of populations was explored through calculating admixture proportions per individual and varying the estimated number of ancestral populations (K). The most likely K was determined by running 10 replicate runs of each respective K, calculating the log likelihood value of each, and choosing the value of 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 1,000 permutations (Excoffier et al., 1992). We used major lineages, sub-lineages determined from the maximum likelihood tree, and populations as the nested levels.

Population genomics of Lineage B

Continuing analyses on populations within Lineage B, we estimated the within-population genomic diversity using two diversity measures. We calculated expected heterozygosity (H_e) using GENODIVE (Meirmans and Van Tienderen, 2004), and overall heterozygosity and nucleotide diversity (π) using ANGSD (Nei, 1987).

Next, we assessed population structure and differentiation. 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. As an unsupervised clustering method, PCA estimates population genetic structure in an unbiased way. Next, we performed a neighbor-joining network (NeighborNet) analysis using Splitstree (Huson, 1998; Huson and Bryant, 2006). Splitstree does not force a tree-like structure onto the data and thus can verify the extent to which the data conform to a hierarchical tree structure. 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 and Hedrick, 2011).

Population demographic histories were inferred via Stairway Plots using Site Frequency Spectra (SFS) (Liu and 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 300Mb. Generation time was set at 1 year. Stairway analyses were run for a subset of locations: two lagoon populations (Bay and DAR), two medium/high connected lakes (Kalimantan 3 and Papua 4), and two isolated lakes (Papua 4 and Papua 30). Lastly, 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).

Spatial and environmental association

Finally, we explored spatial and environmental associations to genetic structure. Mantel tests (Mantel, 1967; Legendre and Legendre, 2012) were used to test significance of correlations between genetic, geographic, environmental and connection distance matrices. For genetic distances, we used normalized pairwise genetic differentiation ($F'_{ST}/(1-F'_{ST})$. Geographic distance was calculated as minimum pairwise distances in meters between lakes using lake coordinates as input for the *geosphere* package in R. Environmental variables temperature (°C) and salinity (ppt) were measured from at least 10 sites per lake, and averages for 1-5m depth were calculated. Using these averages, we ran a Principal Component Analysis (PCA) to retain only informative scores. All PCA axes were then used computing the environmental distance matrix using the function *dist* in R. Connection distance was calculated following the equations of Maas et al., (2018), so that a high pairwise connection distance value indicated a comparison between isolated lakes and a low connection distance a comparison

between connected lakes. Mantel tests were run with 10,000 permutations using *vegan* in R. We verified the absence of autocorrelation between geographic, connection and environmental distances using Mantel's tests. Finally, Spearman correlation tests were performed for within-population diversity indices with temperature, salinity, connection and lake area. Correlations of $r \geq 0.5$ were considered strong, and alpha was set to 0.05.

Results

Lake characterization, read statistics and filtering

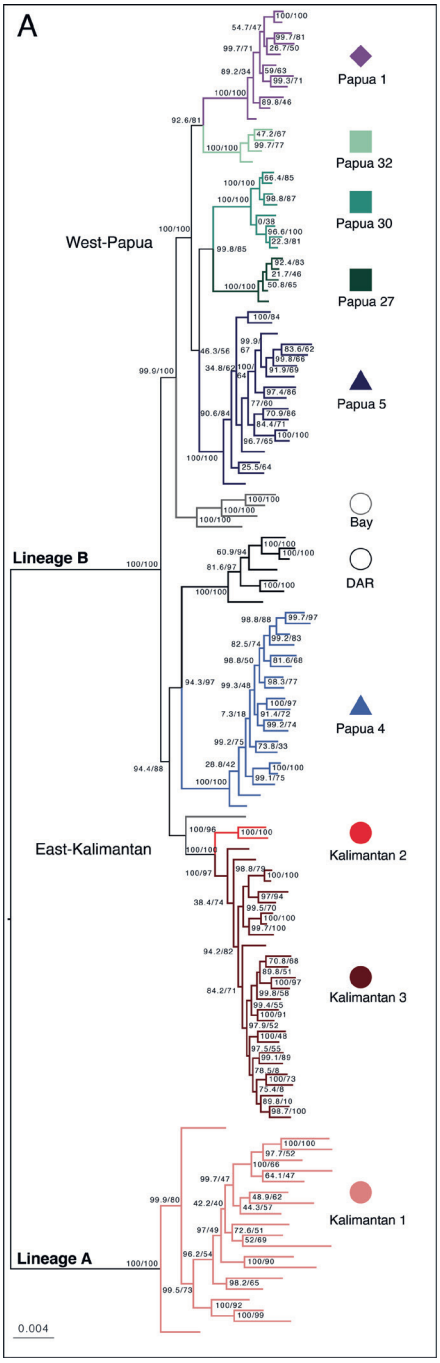
The physical and environmental profiles of the two lagoons and nine marine lakes are provided in Table 1. In general, we observed higher temperatures (30.8°C ±1.22°C) and lower salinities (27.3ppt ±2.7ppt) in lakes than in lagoons (29°C and 33.5ppt). Connection to the surrounding sea varied among lakes,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

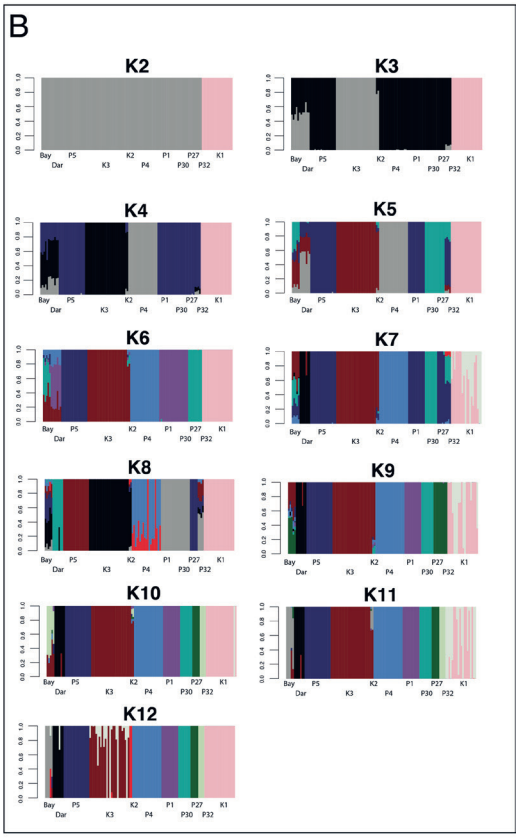
After sequencing and demultiplexing we obtained 1,127,497,643 reads from 168 sponges. On average, we obtained 7,673,269 reads per individual. Individuals with less than 2,000,000 reads were removed from subsequent analyses. Based on the calculation table from Peterson *et al.* (2012) and on an estimated genome size of 300Mb and a size selection of 425-500bp, we expected to retain 13,652 RADtags. This is close to the actual retained loci, as the *de novo* reference retained 14,442 tags when keeping RADtags with at least 3X coverage and present in at least 70% of individuals. Kraken and Blobtools identified 15 out of the 14,442 RADtags containing possible bacterial contamination. The RADtags mapped to *Synechococcus* sp., a Cyanobacteria genus, and were removed from the data set.

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

Major genetic lineages supported by SNP data

The phylogenetic tree based on maximum likelihoods showed two divergent lineages (Fig. 2A), corroborated by the Neighbor-Joining Tree (Supplemental Fig. 1, Supplemental Table 3). These lineages are concordant with Lineage A and B as defined in Becking et al. (2013). Lineage A was only represented by individuals of Kalimantan 1. The remaining populations fell under Lineage B. Within Lineage B, sub-lineages could be seen representing localities from East-Kalimantan, Australia and Papua 4, and the other populations from West-Papua. Lagoon population Bay was a sister group to both the East-Kalimantan and West-Papua clusters.





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

Figure 2: Distinction of major genetic lineages A and B of sponge populations. (A) Maximum Likelihood IQ-Tree using consensus sequences of 124 individuals. A clear split between Lineage A and B can be observed. (B) Bayesian admixture analysis for putative ancestral populations (K 1 - 12) 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 Fig. 1 and Table 1.

Admixture analyses showed Kalimantan 1 to be consistently distinct from all other populations throughout the range of ancestral populations K (Fig. 2B). The most likely number of groups was found to be at K = 9 (Supplemental Fig. 2), after which new groups did not add additional information. At K = 9, all locations are distinct except for Bay and TBB, which showed admixture of multiple groups, and Papua 27 and Papua 32 are clustered together.

When analyzing both lineages the AMOVA showed most variation to be explained by the difference in lineage (Fig. 2C, 62%). When only analyzing within Lineage B, most variation was explained by differences in populations and among individuals (both 49%). Variation among sub-lineages was also found to significantly contribute to the variation, but only 7%).

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

Population genomics of Lineage B

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

Population genetic diversity varied among lakes (Table 1, Supplemental Table 4). The highest genetic diversity was consistently found for the lagoon populations Bay and DAR, as seen for nucleotide diversity (π) (0.0101 and 0.0095, respectively), and for the expected heterozygosity (H_e) (0.157 and 0.117, respectively). Lowest genetic diversity was observed in populations Papua 1 (π = 0.0036, H_e = 0.054) and Papua 27 (π = 0.0037, H_e = 0.038). Population Kalimantan 2 also showed low heterozygosity (H_e = 0.034), but relatively high nucleotide diversity (π = 0.0074). However, this may be an artefact of low sample size. When estimating heterozygosity from genotype likelihoods via ANGSD, we found the lowest heterozygosity for the populations Papua 5 (0.019) and Papua 27 (0.021).

The samples clustered per lake and lagoon location (Fig. 3A, Supplemental Fig. 3 and 4). The first four Principal Components (PCs) in the Principal Component Analysis (PCA) explained 80.5% of total variation (Fig. 3A). PC1, explaining 45.6% of the variation, separated populations by geographic region, with the lakes from West-Papua being distinct from the lakes in East-Kalimantan. PC2, explaining 24.4% of variation, separated lake Papua 4 from the other lakes. PC3 and PC4 (explaining 10.5% in total) further separated lagoon DAR and lakes Papua 5, and to a lesser extend Papua 1 and Papua 30. In the PC1 versus PC2 plot the lagoon populations (Bay and DAR) clustered towards the center of the graph.

For Bay, this continued for the PC3 versus PC4 plot, but not for DAR. Lakes Papua 27 and Papua 32 remained closely associated.

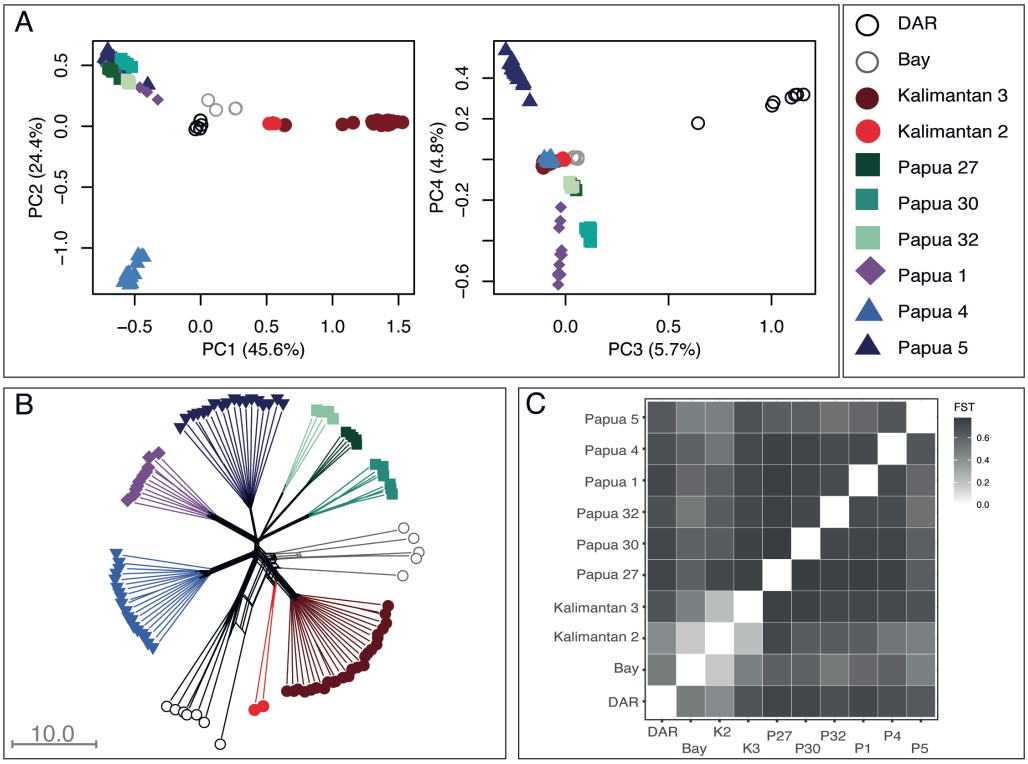


Figure 3: Population genomic structure analyses for Lineage B from sponge populations. (A) Principal Component analysis (PCA) based on pairwise covariance. Each dot represents one individual. (B) Neighbor-Joining Network with equal angles computed in Splitstree based on pairwise genetic distances. (C) Heatmap of normalized F'$_{ST}$ in a multidimensional scaling plot (values in Supplemental Table 5). Colors and codes correspond to Fig. 1 and Table 1.

Findings form the phylogenetic network were consistent with patterns found for PCA (Fig. 3B, Supplemental Fig. 5). The network showed a high fit (fit = 99.2) and small degree of reticulation (d = 0.153), indicating a tree-like structure. The lagoon populations Bay and DAR showed higher reticulation than the marine lake populations, indicating potential introgression or hybridization events.

Pairwise fixation indices (F'$_{ST}$) showed high levels of genetic structuring (0.629 ±0.133) (Fig. 3C). The F'$_{ST}$ ranged from 0.182 between Bay and Kalimantan 2 to 0.778 between Papua 30 and Papua 32 (Supplemental Table 5, Supplemental Fig. 6). All pairwise comparisons were significant, except for the comparison between Papua 32 and Kalimantan 2, potentially due to sample size (n = 4 and 2, respectively).

Stairway plots displayed changes in effective population size for six locations tested (Fig. 4). A decrease in effective population size was observed for lagoon population Bay. While the

lake Kalimantan 3 also showed a steady decline in population size, all other lakes showed a distinct signature of a population bottleneck. All changes in effective population size occurred approximately after sea levels rose to fill the lakes after the Last Glacial Maximum (~20,000 years ago).

The migration network among lakes indicated strongest relative bidirectional migration between marine lakes and the lagoon population in East-Kalimantan (Fig. 1D). Lagoon population Bay was in fact linked to some degree to all other populations (relative bidirectional migration 0.2-1). Within West-Papua, bidirectional migration was observed between Papua 5 and four other lakes (Papua 27, Papua 30, Papua 32, and Papua 1). Lagoon population DAR in Australia showed links to Kalimantan 2 and 3, and Papua 5.

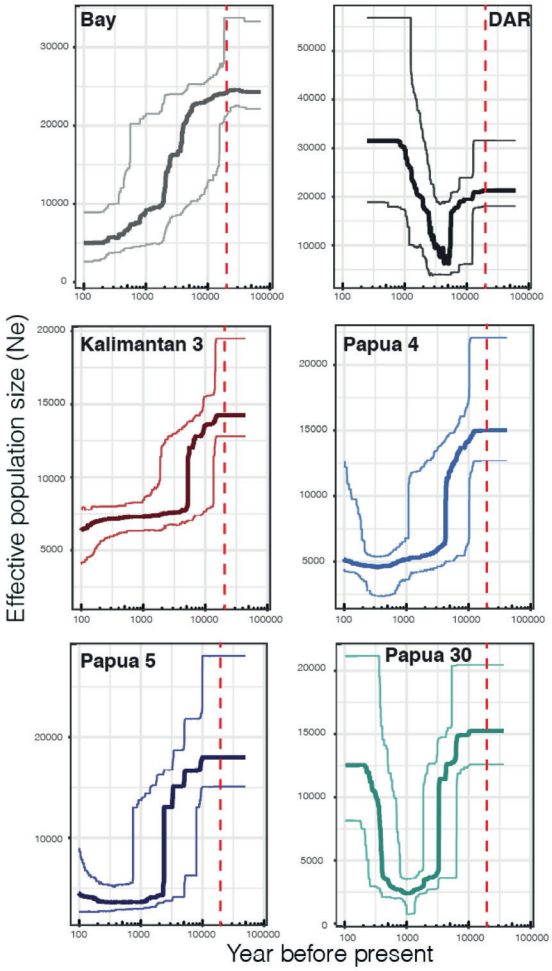


Figure 4: Demographic history inferences on sponge populations within Lineage B. Demographic histories of six locations are displayed: lagoon populations Bay and DAR (top), high/medium connected marine lakes Kalimantan 3 and Papua 4 (middle), and isolated lakes Papua 5 and Papua 30 (bottom). Mean (dark) and 12.5/87.5% confidence intervals are displayed. Red dashed line indicated putative rising of sea levels after Last Glacial Maximum.

Association to drivers

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

Mantel tests indicated no correlation between the geographic and genetic distance matrices over all filter options (Supplemental Fig. 8A, Supplemental Table 7, r = 0.007, p = 0.504). Finding no correlation refutes the isolation-by-distance hypothesis and indicates other factors might explain the distribution of *S. diversicolor* genetic diversity. However, the genetic distance matrix also did not correlate with matrices of environmental distance (r = 0.002, p = 0.503, Supplemental Fig. 8B) or connection distance (r = 0.041, p = 0.441, Supplemental Fig. 8C).

Discussion

A major objective of marine molecular ecology is to obtain accurate estimates of subtle genetic structure, as it can inform efforts to identify units of management and design effective marine protected areas (MPAs) (Kelley et al., 2016; Selkoe et al., 2016). Reviewing population connectivity of poorly dispersing organisms can inform the determination of relevant spatial scales of MPA networks. 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 were able to study fine-scaled genetic structure and the drivers of genetic diversity. Using a reduced representation genomic approach, we confirmed broad-scale patterns of structure identified in a prior single marker study and provided new evidence of strong small-scaled structure for sessile species with a short dispersive larval stage. Below, we discuss our findings on population structure for marine lake sponges, the possible drivers of diversity, and finally the implications for future phylogeographic and population genetic studies on sponges.

RADseq reveals fine scaled structure in Suberites diversicolor

Restriction site-associated DNA sequencing proved suitable to retrieve major genetic lineages (Lineage A and B, Becking et al., 2013) in the sponge *Suberites diversicolor*, albeit at a price of a 9-fold loss of testable common genetic markers. Maximum likelihood and Neighbour-joining trees showed a clear split in lineages previously determined by *COI* and *ITS* markers (Becking et al., 2013b), corroborated by an analysis of molecular variance (AMOVA). Delving deeper into Lineage B, the thousands of RADseq-based SNPs provided the resolution necessary to reveal fine-scaled (<10km) genetic patterns of *Suberites diversicolor* that had not previously been shown. The phylogenetic trees, admixture and PCA analyses all indicated clear clustering per marine lake, and some admixture for lagoon population Bay. A clear geographic split could be seen between localities from East-

Kalimantan to those from West-Papua in the Principal Component Analysis (PCA) and Splitstree analysis. Within West-Papua, marine lake Papua 4 showed a distinct signature from the other marine lakes. To a lesser extend this pattern was also observed for Papua 4 and DAR. Although we observed remarkable genetic differentiation (F_{ST} range 0.18-0.78) spicule analysis and *COI* and *ITS* markers determine the populations within Lineage B to be of the same species.

The observation of finding more structure when using higher numbers of genetic markers has been shown in other marine organisms as well (Maas et al., 2018; Lemopoulos et al., 2019; D'Aloia et al., 2020; Sunde et al., 2020; Timm, 2020). Particularly in the highly diverse Indo-Pacific, population genetic and genomic studies are increasingly revealing high genetic structure (Hernawan et al., 2017; Lal et al., 2017; Vu et al., 2020). The studies point towards the role of both historic and contemporary processes in establishing the current population genetic structure. To bring our understanding of processes influencing marine biodiversity further, it is important to understand the discrepancy in observed genetic structure based on higher resolution of markers as compared to single markers.

While traditionally low-resolution markers have allowed the exposure of sponge species while they were morphologically indistinguishable, they have often lacked detection within-species population genetic diversity (as reviewed in Oppen et al., 2002; Pérez-Portela & Riesgo, 2018; Uriz & Turon, 2012; Wörheide et al., 2005), with notable exceptions (Klautau et al., 1999; Wörheide et al., 2002, 2008). Using the high resolution of RADseq generated markers allowed us to see clear clustering per lake even on very small spatial scales: 1-10km. The scale at which we find strong structure is smaller compared to studies using microsatellites in the sponges *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), to name a few recent studies.

Even studies using higher resolution markers also reveal little structure at small spatial scales, with Brown et al. (2017) detecting little structuring for *Aphrocallistes vastus* in British Colombia at scales <275km and Leiva *et al.* (2019) finding panmixia at scales >900km for *Dendrilla antarctica*. It could be that these are highly connected populations, possibly through rafting or sperm-mediated gene flow (Maldonado, 2006; DeBiasse et al., 2014). Yet it is also possible that the number of SNPs from Brown et al. (2017) and Leiva et al. (2019) (67 and 529, respectively) was too low to detect subtle structure at small scales.

Underlying processes causing population genomic structure

We assessed the effects of several drivers of population diversity and structure. First, we tested to what extent marine genetic differentiation conforms to the decay of population similarity with geographical distance resulting in a pattern of isolation-by-distance (Wright, 1943) using only Lineage B. We found strong population structure with clustering per lake, and an apparent split between populations from East-Kalimantan and West-Papua, yet no

pattern of isolation-by-distance (IBD) was observed. This is remarkable, since we sampled at geographical distances of 1km - 1,400km. Even if individual lakes have no ongoing gene flow, we expect the ocean populations seeding the lake populations to show IBD patterns, which would then be reflected in the lakes. We also did not detect a pattern of isolation-by-environment, despite the great environmental variability among lakes (temperature: 29 - 32.4 °C, salinity: 24 - 33.4 ppt). Previous studies using a low number of markers did find a pattern of isolation-by-distance for sponges (Duran et al., 2004b; Blanquer and Uriz, 2010; Noyer and Becerro, 2012; Pérez-Portela et al., 2015), which is usually expected for species with restricted dispersal abilities (Wörheide et al., 2005; Maldonado, 2006). Other studies report an influence of oceanographic currents (Chaves-Fonnegra et al., 2015; Richards et al., 2016; Riesgo et al., 2019), or environmental heterogeneity (temperature and productivity) (Giles et al., 2015) on sponges.

Our results indicate that mechanisms other than only dispersal limitation by geographical distance or local environments alone are important in structuring *S. diversicolor* populations. In addition, the permeability of the landscape barrier surrounding the marine lakes, determining the degree of water flowing in and out of the lakes, did not seem to influence the population structure or diversity. Perhaps *S. diversicolor* populations are truly isolated per lake as their low dispersal ability restricts effective gene flow. Populations can then become differentiated through genetic drift or via local adaptation to environmental parameters not yet recorded (Frankham et al., 2002). The observation of severe population bottlenecks for marine lake populations Papua 4, Papua 30 and to a lesser extend Papua 5 suggests a potential effect of founder events. Population bottlenecks were observed regardless of level of connection, as even a lagoon population showed a bottleneck pattern. Founder effects and subsequent priority effects could explain the pattern of strong population structure (Orsini et al., 2013; Fukami, 2015; De Meester et al., 2016).

Priority effects were previously discussed as potential drivers of structure in marine lake organisms by Maas et al. (2018) and de Leeuw et al. (2020). Depending on spatial scale Maas et al. (2018) found an effect of geographic distance and connectivity influencing mussel population structure. They argued that despite founder events stochastically driving alleles to fixation in small populations, ongoing dispersal would overwhelm this effect (Mayr, 1963; Waters 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 (Orsini et al., 2013; Fukami, 2015; De Meester et al., 2016). Sponges, in contrast, generally have poor dispersal abilities (Maldonado, 2006). As the current study does not find an effect of connection to the sea in structuring populations, stochastic fixation of alleles due to genetic drift may be the cause of each population being distinct. Including more lakes with replicates of local environments and/or connection to the sea may further elucidate drivers of sponge differentiation in fragmented habitats.

Implications for sponge phylogeography and population genomic studies

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

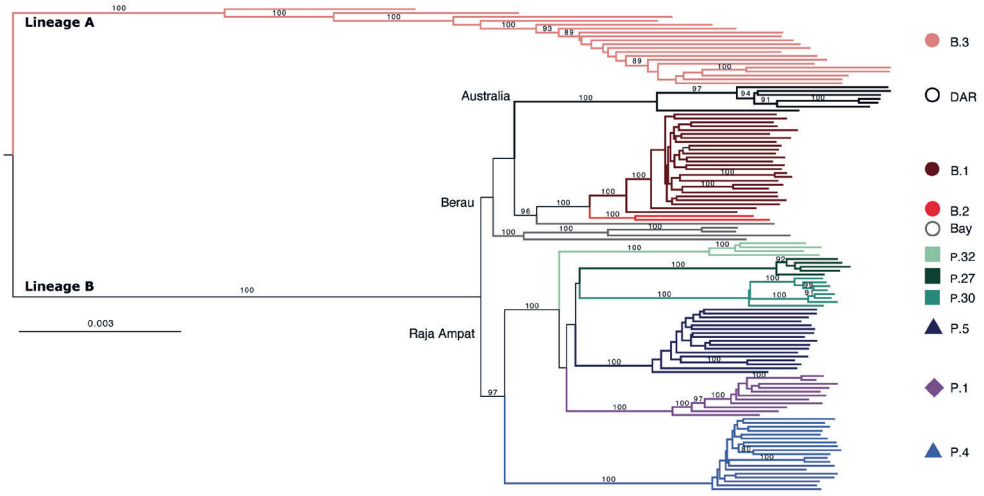
Further, our adjustments to the existing low-cost protocol of Peterson et al. (2012) with a step-by-step protocol presented in Maas et al. (2018) can help to retrieve extensive data for non-model marine organisms in general and tropical sponges in particular, thus benefitting future studies. We further showed that reduced representation genome sequencing can work for DNA that was extracted for other purposes and stored for long times in a -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 sponge studies across the world.

Acknowledgements

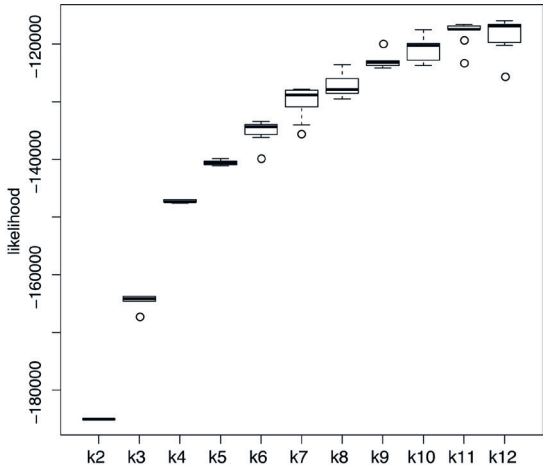
The following people provided help with logistics during fieldwork and/or labwork: B. Alvarez, M. Ammer, Bahruddin, , H. Breeuwer, L. Dong, M. Erdmann, B. Hoeksema, P. Kuperus, S.C. Lim, S. Menken, A. Miners, A. Oherenan, K. Peijnenburg, W. Renema, Suharsono, R. Tapilatu, Y. Tuti, B. Voetdijk, N.J. de Voogd, the staff of Nabucco Island Dive Resort, Derawan Dive Resort, Misool Eco resort, Papua Diving. We are grateful for useful discussions with Dr. P. Meirmans. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation grants S10RR029668 and S10RR027303. The Netherlands Organisation for Scientific Research provided funding to Leontine E. Becking through the grants RUBICON #825.12.007 and VENI #863.14.020. Fieldwork in Indonesia was made possible through additional financial support of National Geographic Society Waitt Grant, De Beukelaar-van der Hucht Stichting, World Wildlife Fund, Netherlands-INNO Fund, the Schure-Beijerinck-Popping Fund of the Royal Dutch Academy of Science (KNAW), Conservation International Ecosystem Based Management program (funded by the David and Lucile Packard Foundation), the Treub-Maatschappij Fund, the Leiden University Fund (LUF)/Slingelands, Singapore Airlines, the A.M. Buitendijk Fund and the J.J. ter Pelkwijk Fund (Naturalis Biodiveristy Center). The

funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are grateful to the Indonesian Institute of Sciences (LIPI) and the Indonesian Ministry of Research and Technology (RISTEK) for providing the research permits (0094/FRP/SM/V/2009;098/SIP / FRP/SM/V/2011;098/SIP/FRP/SM/V /20113246/FRP/SM/VII/2012;#3B/TKPIPA/E5/Dit.KI/III/2016).

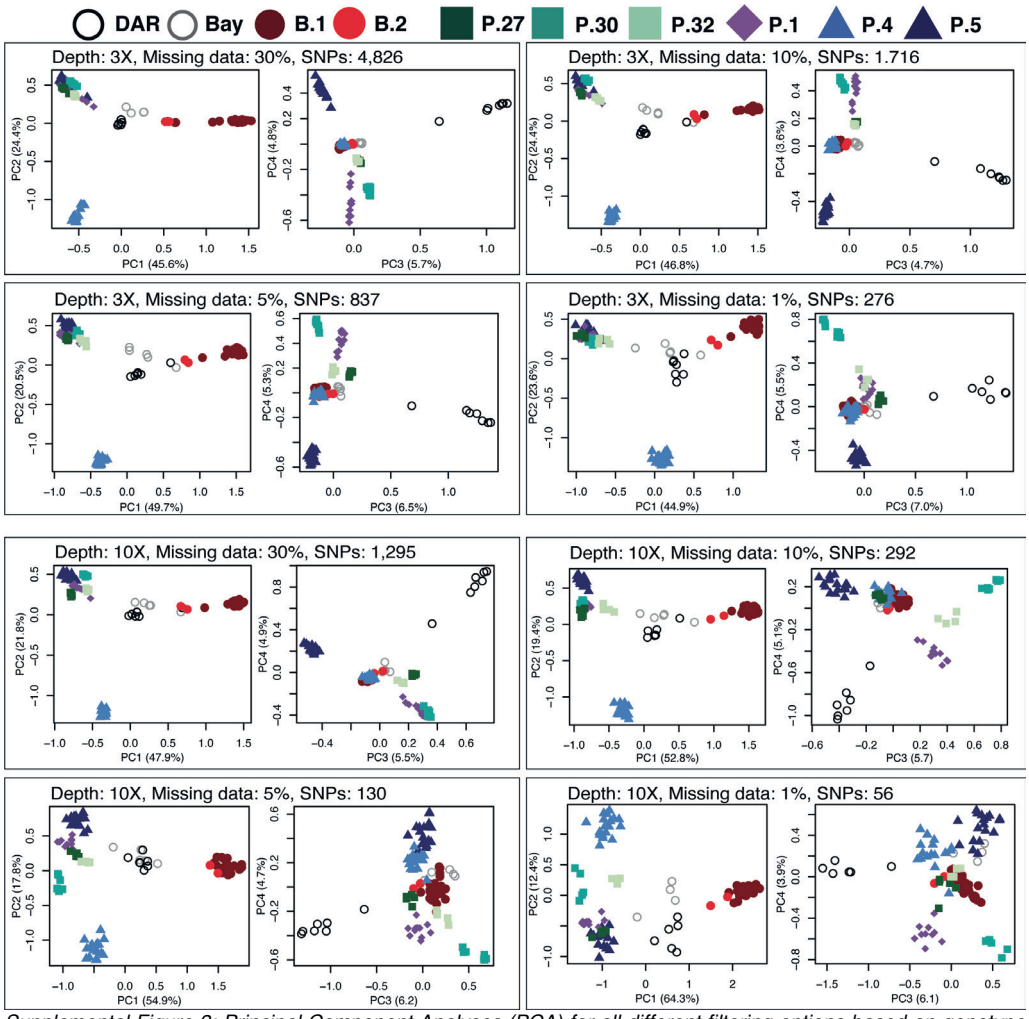
Supplemental Figures



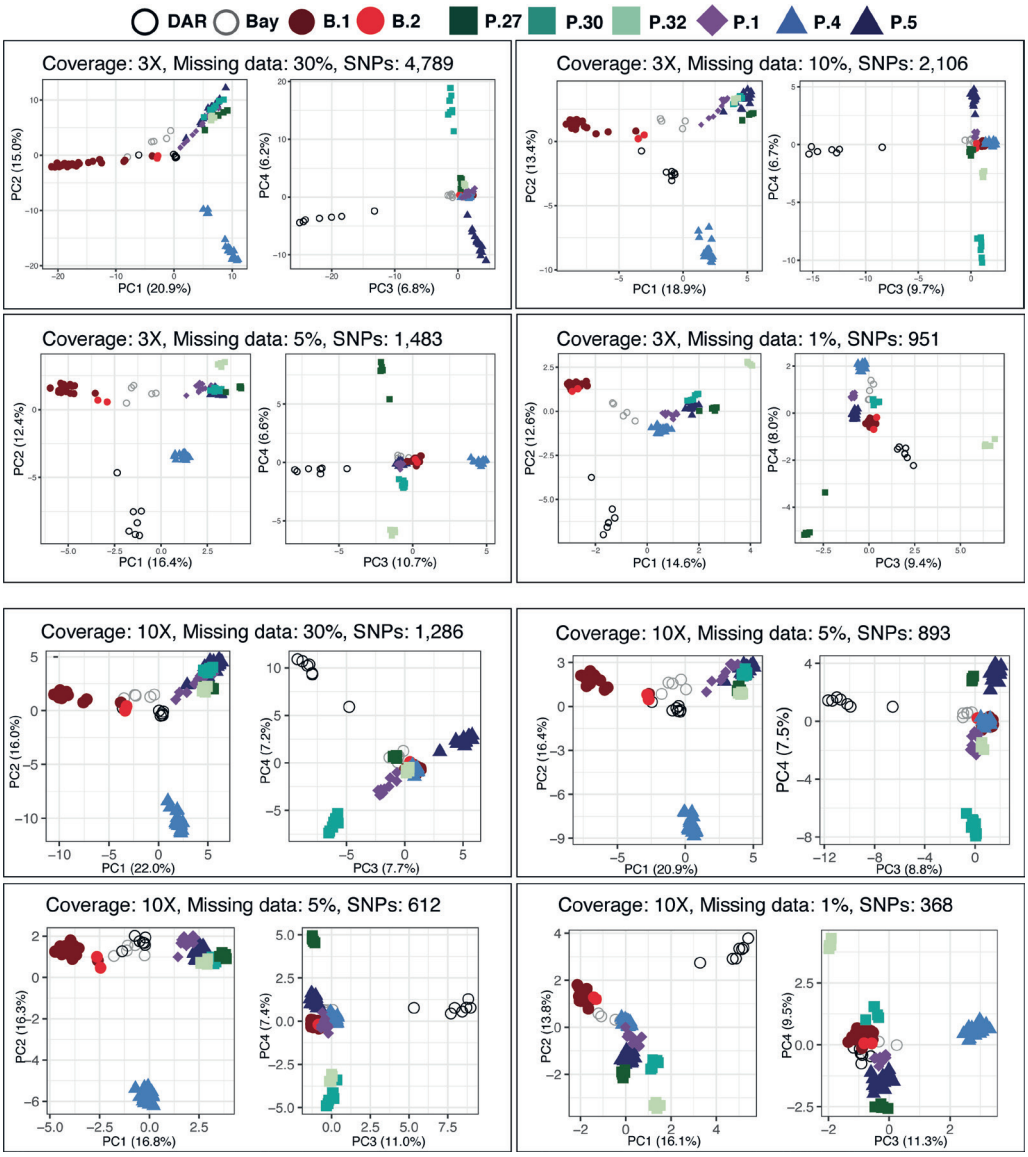
Supplemental Figure 1: Neighbor-Joining tree based on pairwise genetic distances of Suberites diversicolor *populations. Bootstrap support values are displayed and based on 1000 bootstraps. Each branch represents one individual. Colors and codes correspond to Fig. 1 and Table 1.*



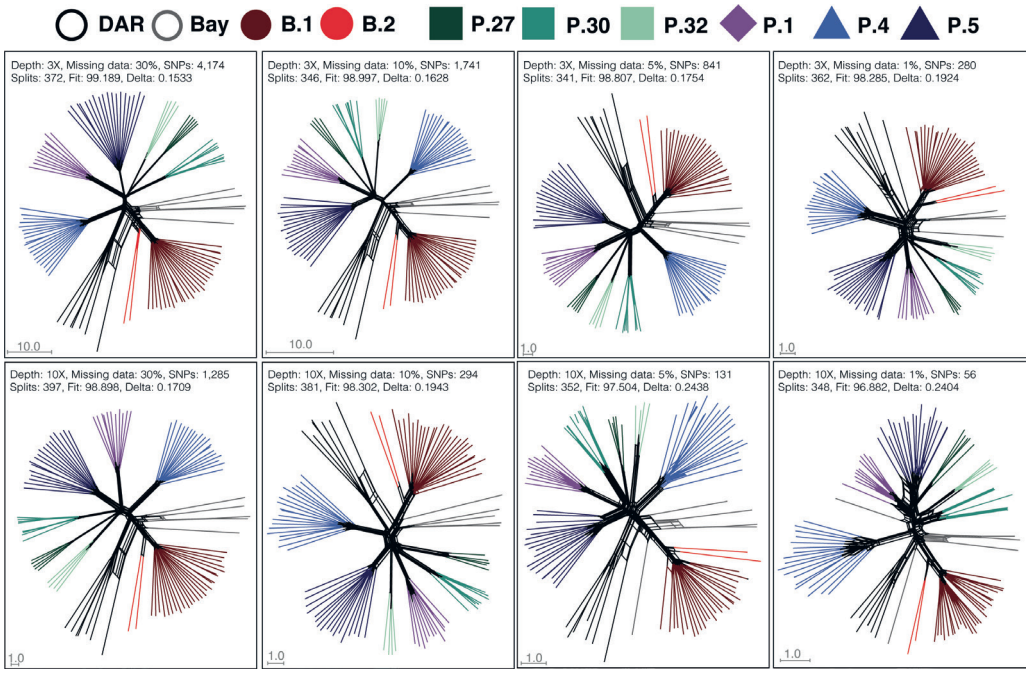
Supplemental Figure 2: Boxplot of likelihood values obtained from Admixture analyses. Values based on 10 replicate runs per putative ancestral population (K).



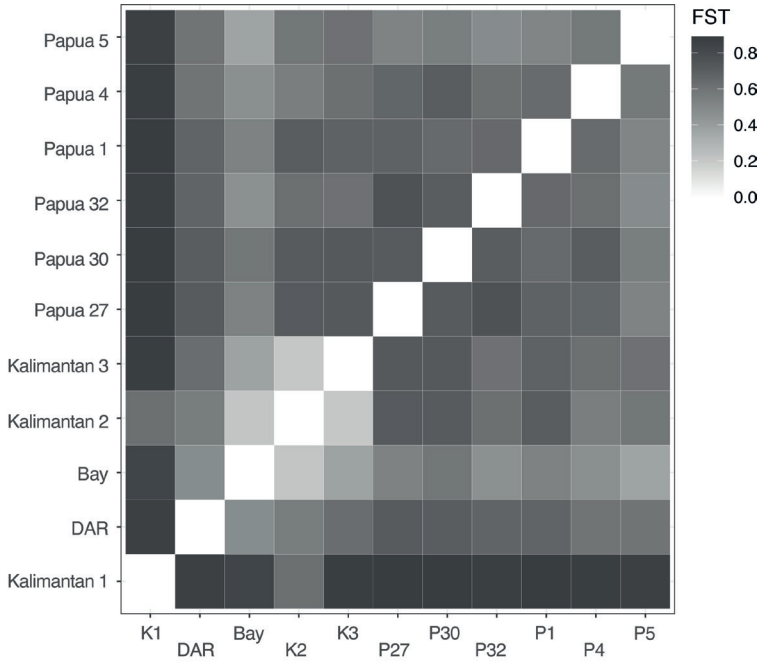
Supplemental Figure 3: Principal Component Analyses (PCA) for all different filtering options based on genotype likelihoods. Each dot represents one individual. Colors and codes correspond to Fig. 1.



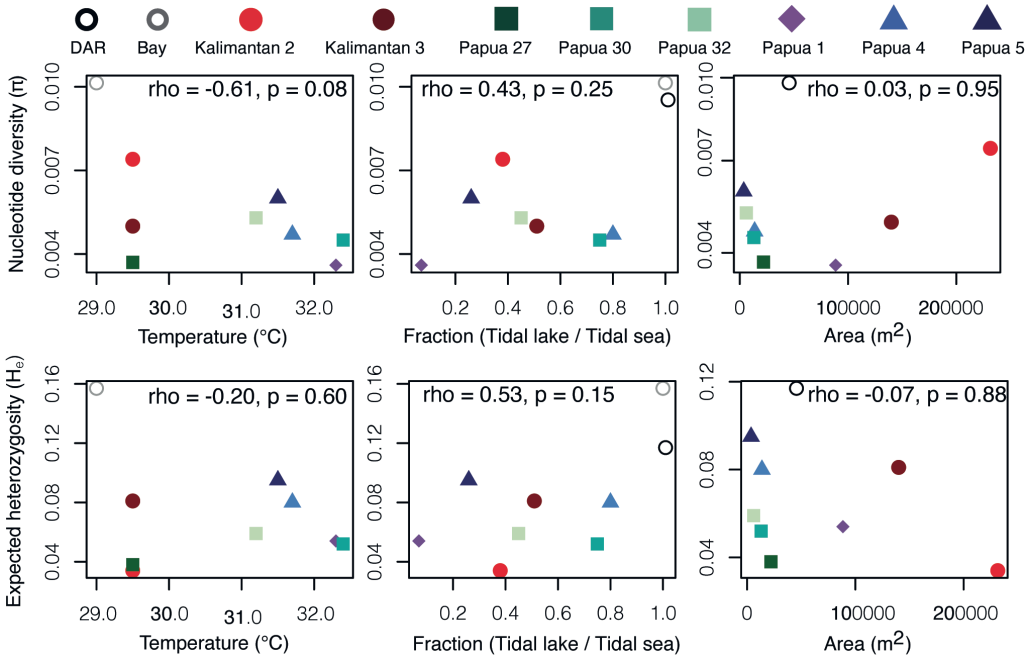
Supplemental Figure 4: Principal Component Analyses (PCA) for all different filtering options based on genotype calls. Each dot represents one individual. Colors and codes correspond to Fig. 1.



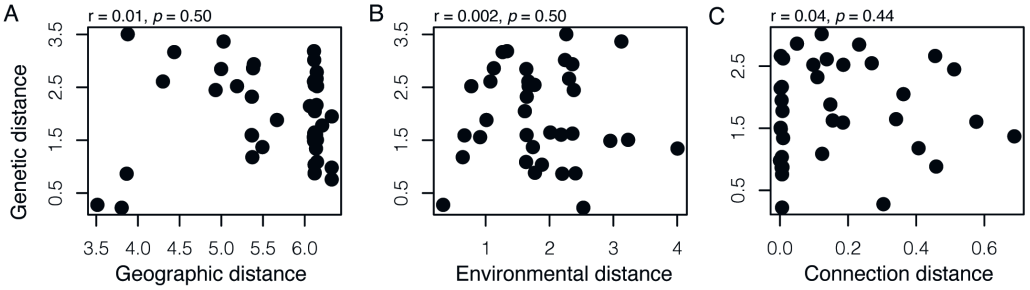
Supplemental Figure 5: Neighbor-Joining Network with equal angles computed in Splitstree based on pairwise genetic distances from different filtering options. Colors and codes correspond to Fig. 1.



Supplemental Figure 6: Visualization of normalized F'_{ST} in heatmap including both lineages.



Supplemental Figure 7: Visualizations of Spearman correlations of genetic diversity indices nucleotide diversity and heterozygosity to environmental factors. Correlations to temperature, connection to the surrounding sea and lake area are displayed. Colors and codes correspond to Fig. 1.



Supplemental Figure 8: Visualizations of Mantel tests based on genetic distance versus geographic, environmental and connection distance.

Supplemental Tables

Supplemental Table 1: Codes corresponding of current study and the study of Becking et al., (2013).

Code Current study	Code Becking *et al.* (2013)
DAR	DAR
Bay	BER
Kalimantan 1	KKB
Kalimantan 2	TBB
Kalimantan 3	HBL
Papua 27	
Papua 30	CAS
Papua 32	URA
Papua 1	RAJ
Papua 4	MIS
Papua 5	

Supplemental Table 2: Retained Single Nucleotide Polymorphisms (SNPs) for genotype calls and genotype likelihoods after various filtering options. Coverage was set at 3X or 10X, missing data allowed was 30%, 10%, 5% or 1%.

Filtering	SNPs Genotype likelihoods	SNPs Genotype calls
Cov3_Miss30	4,826	4,790
Cov3_Miss10	1,716	2,106
Cov3_Miss5	837	1,483
Cov3_Miss1	276	951
Cov10_Miss30	1,295	1,286
Cov10_Miss10	292	893
Cov10_Miss5	130	612
Cov10_Miss1	56	368

Supplemental Table 3: Pairwise genetic distances (p-distances) used for calculating the Neighbor-Joining tree of Supplemental Figure 1. Excel sheet available upon request: diede.maas@wur.nl

Supplemental Table 4: Estimates of genetic diversity (nucleotide diversity and heterozygosity) for Suberites diversicolor populations. Per site values are colord from low (red) to high (green) to visualize differences among genetic diversity index and filtering options.

Lake	Nucleotide Diversity 3X 30%	Nucleotide Diversity 3X 10%	Nucleotide Diversity 3X 5%	Nucleotide Diversity 3X 1%
DAR	0.0095	0.0090	0.0090	0.0088
Bay	0.0101	0.0097	0.0099	0.0095
Kalimantan 3	0.0050	0.0051	0.0055	0.0058
Kalimantan 2	0.0074	0.0072	0.0073	0.0080
Papua 27	0.0037	0.0040	0.0044	0.0052
Papua 30	0.0045	0.0047	0.0052	0.0069
Papua 32	0.0053	0.0054	0.0059	0.0071
Papua 1	0.0036	0.0039	0.0042	0.0045
Papua 4	0.0047	0.0048	0.0052	0.0061
Papua 5	0.0060	0.0060	0.0065	0.0065

Lake	Nucleotide Diversity 10X 30%	Nucleotide Diversity 10X 10%	Nucleotide Diversity 10X 5%	Nucleotide Diversity 10X 1%
DAR	0.0098	0.0093	0.0086	0.0080
Bay	0.0103	0.0095	0.0100	0.0100
Kalimantan 3	0.0055	0.0056	0.0052	0.0057
Kalimantan 2	0.0077	0.0072	0.0079	0.0080
Papua 27	0.0048	0.0051	0.0073	0.0086
Papua 30	0.0056	0.0061	0.0070	0.0091
Papua 32	0.0062	0.0074	0.0081	0.0087
Papua 1	0.0042	0.0043	0.0041	0.0037
Papua 4	0.0052	0.0063	0.0074	0.0088
Papua 5	0.0064	0.0059	0.0064	0.0060

Lake	Heterozygosity GL 3X 30%	Heterozygosity GL 3X 10%	Heterozygosity GL 3X 5%	Heterozygosity GL 3X 1%
DAR	0.1420	0.1504	0.1562	0.1320
Bay	0.1171	0.1281	0.1095	0.1140
Kalimantan 3	0.0295	0.0323	0.0342	0.0479
Kalimantan 2	0.1200	0.1145	0.1187	0.1212
Papua 27	0.0211	0.0228	0.0273	0.0372
Papua 30	0.0246	0.0233	0.0172	0.0240
Papua 32	0.0451	0.0470	0.0492	0.0419
Papua 1	0.0257	0.0226	0.0208	0.0221
Papua 4	0.0235	0.0246	0.0185	0.0295
Papua 5	0.0188	0.0120	0.0137	0.0091

Lake	Heterozygosity GL 10X 30%	Heterozygosity GL 10X 10%	Heterozygosity GL 10X 5%	Heterozygosity GL 10X 1%
DAR	0.1469	0.1247	0.1452	0.1541
Bay	0.1109	0.0955	0.0969	0.0893
Kalimantan 3	0.0306	0.0426	0.0504	0.0536
Kalimantan 2	0.1249	0.0970	0.1152	0.1117
Papua 27	0.0253	0.0242	0.0309	0.0544
Papua 30	0.0252	0.0249	0.0156	0.0381
Papua 32	0.0426	0.0525	0.0391	0.0715
Papua 1	0.0280	0.0212	0.0257	0.0357
Papua 4	0.0245	0.0162	0.0237	0.0537
Papua 5	0.0163	0.0194	0.0273	0.0080

Lake	Heterozygosity (He) GC 3x 30%	Heterozygosity (He) GC 3x 10%	Heterozygosity (He) GC 3x 5%	Heterozygosity (He) GC 3x 1%
DAR	0.117	0.105	0.089	0.077
Bay	0.157	0.119	0.117	0.125
Kalimantan 3	0.081	0.043	0.027	0.014
Kalimantan 2	0.034	0.024	0.023	0.019
Papua 27	0.038	0.032	0.028	0.024
Papua 30	0.052	0.037	0.029	0.019
Papua 32	0.059	0.05	0.048	0.049
Papua 1	0.054	0.03	0.021	0.01
Papua 4	0.08	0.043	0.03	0.015
Papua 5	0.095	0.043	0.027	0.011

Lake	Heterozygosity (He) GC 10x 30%	Heterozygosity (He) GC 10x 10%	Heterozygosity (He) GC 10x 5%	Heterozygosity (He) GC 10x 1%
DAR	0.12	0.11	0.103	0.098
Bay	0.143	0.13	0.112	0.104
Kalimantan 3	0.079	0.068	0.044	0.02
Kalimantan 2	0.044	0.038	0.035	0.026
Papua 27	0.039	0.032	0.028	0.022
Papua 30	0.055	0.047	0.038	0.026
Papua 32	0.062	0.055	0.047	0.047
Papua 1	0.057	0.047	0.036	0.022
Papua 4	0.075	0.06	0.047	0.028
Papua 5	0.092	0.071	0.049	0.03

Supplemental Table 5: Normalized F'_{ST} and associated significance based on different filtering options. Below diagonal normalized F'_{ST} is displayed, above diagonal p-values. Tables are colored according to F'_{ST} value from low (green) to high (red).

u3k70	DAR	Bay	Kalimantan 3	Kalimantan 2	Papua 30	Papua 27	Papua 32	Papua 1	Papua 4	Papua 5
DAR		0.002	0.001	0.034	0.001	0.001	0.007	0.001	0.001	0.001
Bay	0.495		0.001	*0.051*	0.001	0.014	0.01	0.002	0.001	0.001
Kalimantan 3	0.661	0.464		0.001	0.001	0.001	0.001	0.001	0.001	0.001
Kalimantan 2	0.431	0.182	0.217		0.024	0.048	*0.067*	0.01	0.003	0.003
Papua 30	0.727	0.598	0.751	0.619		0.002	0.003	0.001	0.001	0.001
Papua 27	0.736	0.609	0.761	0.723	0.778		0.01	0.001	0.001	0.001
Papua 32	0.684	0.509	0.718	0.622	0.723	0.76		0.002	0.001	0.001
Papua 1	0.724	0.573	0.727	0.616	0.74	0.771	0.71		0.001	0.001
Papua 4	0.682	0.601	0.716	0.521	0.741	0.746	0.699	0.716		0.001
Papua 5	0.64	0.466	0.672	0.469	0.614	0.615	0.541	0.578	0.653	

u10k70	DAR	Bay	Kalimantan 3	Kalimantan 2	Papua 30	Papua 27	Papua 32	Papua 1	Papua 4	Papua 5
DAR		0.001	0.001	0.035	0.001	0.002	0.003	0.001	0.001	0.001
Bay	0.494		0.001	0.048	0.003	0.012	0.007	0.001	0.001	0.001
Kalimantan 3	0.654	0.433		0.005	0.001	0.001	0.001	0.001	0.001	0.001
Kalimantan 2	0.506	0.211	0.183		0.021	0.045	*0.074*	0.01	0.004	0.003
Papua 30	0.725	0.611	0.738	0.718		0.001	0.004	0.001	0.001	0.001
Papua 27	0.724	0.618	0.749	0.762	0.785		0.007	0.002	0.001	0.002
Papua 32	0.688	0.548	0.716	0.674	0.727	0.772		0.001	0.001	0.001
Papua 1	0.708	0.585	0.72	0.641	0.713	0.765	0.717		0.001	0.001
Papua 4	0.682	0.596	0.701	0.597	0.747	0.746	0.706	0.721		0.001
Papua 5	0.646	0.492	0.675	0.497	0.624	0.617	0.553	0.591	0.664	

u3k90	DAR	Bay	Kalimantan 3	Kalimantan 2	Papua 30	Papua 27	Papua 32	Papua 1	Papua 4	Papua 5
DAR		0.003	0.001	0.033	0.001	0.004	0.007	0.001	0.001	0.001
Bay	0.45		0.001	0.048	0.002	0.011	0.007	0.003	0.001	0.001
Kalimantan 3	0.62	0.421		0.006	0.001	0.001	0.001	0.001	0.001	0.001
Kalimantan 2	0.4	0.162	0.191		0.019	0.037	*0.06*	0.009	0.008	0.007
Papua 30	0.645	0.503	0.652	0.595		0.002	0.002	0.001	0.001	0.001
Papua 27	0.663	0.547	0.716	0.691	0.737		0.011	0.002	0.001	0.001
Papua 32	0.622	0.47	0.677	0.609	0.667	0.743		0.002	0.001	0.001
Papua 1	0.643	0.489	0.608	0.595	0.665	0.747	0.699		0.001	0.001
Papua 4	0.617	0.49	0.572	0.482	0.644	0.693	0.664	0.612		0.001
Papua 5	0.616	0.436	0.573	0.502	0.568	0.636	0.587	0.508	0.546	

u10k90	DAR	Bay	Kalimantan 3	Kalimantan 2	Papua 30	Papua 27	Papua 32	Papua 1	Papua 4	Papua 5
DAR		0.003	0.001	0.024	0.001	0.003	0.006	0.001	0.001	0.001
Bay	0.475		0.001	*0.055*	0.001	0.006	0.008	0.001	0.001	0.001
Kalimantan 3	0.626	0.425		0.006	0.001	0.001	0.001	0.001	0.001	0.001
Kalimantan 2	0.474	0.218	0.139		0.017	0.05	*0.068*	0.01	0.011	0.005
Papua 30	0.7	0.574	0.694	0.69		0.002	0.001	0.001	0.001	0.001
Papua 27	0.7	0.559	0.707	0.753	0.78		0.009	0.001	0.001	0.001
Papua 32	0.679	0.545	0.702	0.674	0.705	0.766		0.001	0.001	0.001
Papua 1	0.685	0.534	0.661	0.599	0.69	0.74	0.707		0.001	0.001
Papua 4	0.667	0.564	0.63	0.543	0.715	0.723	0.698	0.696		0.001
Papua 5	0.638	0.451	0.616	0.487	0.591	0.6	0.579	0.551	0.624	

u3k95	DAR	Bay	Kalimantan 3	Kalimantan 2	Papua 30	Papua 27	Papua 32	Papua 1	Papua 4	Papua 5
DAR		0.001	0.001	0.026	0.001	0.004	0.009	0.001	0.001	0.001
Bay	0.424		0.001	*0.054*	0.001	0.004	0.009	0.001	0.001	0.001
Kalimantan 3	0.617	0.428		0.002	0.001	0.001	0.001	0.001	0.001	0.001
Kalimantan 2	0.365	0.15	0.194		0.022	*0.056*	*0.079*	0.011	0.005	0.007
Papua 30	0.593	0.445	0.612	0.554		0.001	0.005	0.001	0.001	0.001
Papua 27	0.629	0.498	0.723	0.677	0.702		0.009	0.001	0.001	0.001
Papua 32	0.616	0.459	0.709	0.592	0.657	0.736		0.002	0.001	0.001
Papua 1	0.602	0.416	0.556	0.556	0.589	0.722	0.699		0.001	0.001
Papua 4	0.588	0.468	0.525	0.468	0.574	0.691	0.686	0.541		0.001
Papua 5	0.602	0.428	0.528	0.506	0.524	0.659	0.651	0.43	0.478	

u10k95	DAR	Bay	Kalimantan 3	Kalimantan 2	Papua 30	Papua 27	Papua 32	Papua 1	Papua 4	Papua 5
DAR		0.003	0.001	0.026	0.001	0.003	0.006	0.001	0.001	0.001
Bay	0.473		0.001	0.049	0.001	0.009	0.01	0.004	0.001	0.001
Kalimantan 3	0.629	0.387		0.006	0.001	0.001	0.001	0.001	0.001	0.001
Kalimantan 2	0.445	0.19	0.155		0.02	*0.051*	*0.057*	0.014	0.004	0.008
Papua 30	0.647	0.532	0.645	0.66		0.002	0.003	0.001	0.001	0.001
Papua 27	0.691	0.559	0.716	0.745	0.753		0.008	0.002	0.001	0.001
Papua 32	0.644	0.532	0.693	0.678	0.686	0.783		0.001	0.001	0.001
Papua 1	0.628	0.446	0.567	0.536	0.61	0.713	0.696		0.001	0.001
Papua 4	0.664	0.54	0.58	0.53	0.651	0.72	0.693	0.636		0.001
Papua 5	0.616	0.423	0.52	0.433	0.534	0.603	0.608	0.439	0.561	

u3k99	DAR	Bay	Kalimantan 3	Kalimantan 2	Papua 30	Papua 27	Papua 32	Papua 1	Papua 4	Papua 5
DAR		0.002	0.001	0.031	0.001	0.004	0.001	0.001	0.001	0.001
Bay	0.31		0.001	*0.195*	0.001	0.008	0.014	0.001	0.001	0.001
Kalimantan 3	0.565	0.449		0.004	0.001	0.001	0.001	0.001	0.001	0.001
Kalimantan 2	0.272	0.098	0.122		0.033	0.036	*0.075*	0.019	0.006	0.008
Papua 30	0.486	0.355	0.574	0.551		0.001	0.004	0.001	0.001	0.001
Papua 27	0.521	0.412	0.721	0.646	0.663		0.004	0.001	0.001	0.001
Papua 32	0.545	0.414	0.751	0.573	0.63	0.69		0.003	0.002	0.001
Papua 1	0.491	0.378	0.535	0.604	0.537	0.71	0.728		0.001	0.001
Papua 4	0.489	0.41	0.488	0.487	0.504	0.688	0.716	0.434		0.001
Papua 5	0.551	0.437	0.538	0.612	0.526	0.696	0.731	0.507	0.444	

u10k99	DAR	Bay	Kalimantan 3	Kalimantan 2	Papua 30	Papua 27	Papua 32	Papua 1	Papua 4	Papua 5
DAR		0.002	0.001	0.025	0.001	0.001	0.003	0.001	0.001	0.001
Bay	0.432		0.001	*0.108*	0.002	0.01	0.01	0.001	0.001	0.001
Kalimantan 3	0.678	0.432		0.005	0.001	0.001	0.001	0.001	0.001	0.001
Kalimantan 2	0.402	0.136	0.185		0.026	0.043	*0.073*	0.007	0.004	0.005
Papua 30	0.563	0.445	0.658	0.625		0.003	0.003	0.001	0.001	0.001
Papua 27	0.615	0.487	0.727	0.692	0.711		0.009	0.001	0.001	0.001
Papua 32	0.617	0.515	0.782	0.672	0.673	0.763		0.001	0.001	0.001
Papua 1	0.548	0.328	0.521	0.463	0.545	0.665	0.72		0.001	0.001
Papua 4	0.605	0.442	0.532	0.479	0.58	0.669	0.728	0.479		0.001
Papua 5	0.595	0.387	0.507	0.409	0.53	0.575	0.683	0.36	0.468	

Supplemental Table 6: Spearman correlations of genetic diversity indices nucleotide diversity and heterozygosity to environmental factors. Outcomes for different filtering options are displayed.

u3k70		
Nucleotide diversity ~	Spearman's rho	*p*-value
Temperature	-0.610	0.081
Salinity	0.368	0.330
Connection	0.433	0.250
Area	0.033	0.948

u10k70		
Nucleotide diversity ~	Spearman's rho	*p*-value
Temperature	-0.475	0.197
Salinity	0.452	0.222
Connection	0.433	0.250
Area	-0.067	0.880

u3k70		
Expected Heterozygosity ~	Spearman's rho	*p*-value
Temperature	-0.203	0.600
Salinity	0.243	0.529
Connection	0.533	0.148
Area	-0.067	0.880

u10k70		
Expected Heterozygosity ~	Spearman's rho	*p*-value
Temperature	-0.203	0.600
Salinity	0.159	0.683
Connection	0.533	0.148
Area	-0.217	0.581

u3k90		
Nucleotide diversity ~	Spearman's rho	*p*-value
Temperature	-0.610	0.081
Salinity	0.368	0.330
Connection	0.433	0.250
Area	0.033	0.948

u10k90		
Nucleotide diversity ~	Spearman's rho	*p*-value
Temperature	-0.356	0.347
Salinity	0.393	0.295
Connection	0.683	0.050
Area	-0.133	0.744

u3k90		
Expected Heterozygosity ~	Spearman's rho	*p*-value
Temperature	-0.276	0.472
Salinity	0.545	0.129
Connection	0.678	0.045
Area	0.458	0.215

u10k90		
Expected Heterozygosity ~	Spearman's rho	*p*-value
Temperature	-0.196	0.614
Salinity	0.197	0.611
Connection	0.577	0.104
Area	-0.251	0.515

u3k95		
Nucleotide diversity ~	Spearman's rho	*p*-value
Temperature	-0.576	0.104
Salinity	0.427	0.252
Connection	0.417	0.270
Area	0.017	0.982

u10k95		
Nucleotide diversity ~	Spearman's rho	*p*-value
Temperature	-0.509	0.162
Salinity	0.544	0.130
Connection	0.683	0.050
Area	-0.083	0.843

u3k95		
Expected Heterozygosity ~	Spearman's rho	*p*-value
Temperature	-0.213	0.583
Salinity	0.660	0.053
Connection	0.845	0.004
Area	-0.460	0.213

u10k95		
Expected Heterozygosity ~	Spearman's rho	*p*-value
Temperature	-0.102	0.794
Salinity	0.294	0.442
Connection	0.611	0.081
Area	-0.469	0.203

u3k99		
Nucleotide diversity ~	Spearman's rho	*p*-value
Temperature	-0.373	0.323
Salinity	0.444	0.232
Connection	0.567	0.121
Area	-0.083	0.843

u10k99		
Nucleotide diversity ~	Spearman's rho	*p*-value
Temperature	-0.102	0.795
Salinity	0.561	0.116
Connection	0.700	0.043
Area	-0.500	0.178

u3k99		
Expected Heterozygosity ~	Spearman's rho	*p*-value
Temperature	-0.545	0.129
Salinity	0.786	0.012
Connection	0.745	0.021
Area	-0.084	0.831

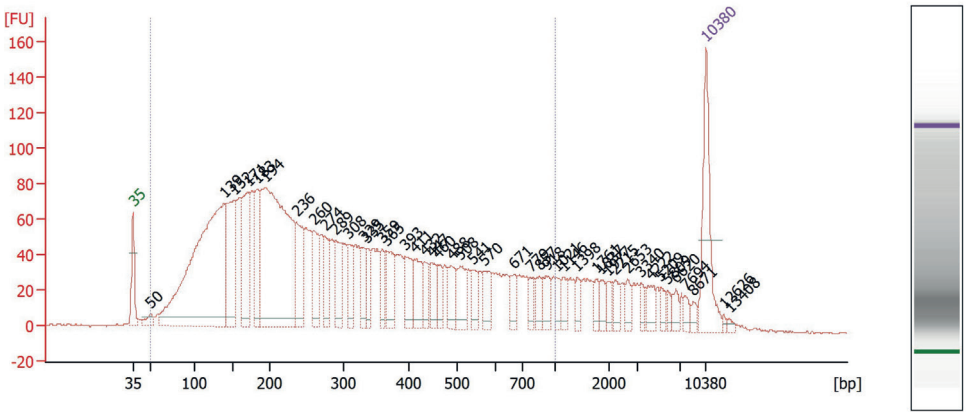
u10k99		
Expected Heterozygosity ~	Spearman's rho	*p*-value
Temperature	-0.128	0.742
Salinity	0.460	0.213
Connection	0.544	0.130
Area	-0.538	0.135

Supplemental Information A: Results from BioAnalyzer for restriction enzymes

Assay Class: High Sensitivity DNA Assay
Data Path: C:\...gh Sensitivity DNA Assay_DE13804097_2014-10-17_16-31-56.xad

Created: 10/17/2014 4:31:55 PM
Modified: 10/17/2014 6:17:12 PM

Electropherogram Summary Continued ...



Overall Results for sample 2 :

Number of peaks found: 45
Noise: 0.5
Corr. Area 1: 3,462.3

Peak table for sample 2 : S2

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations	Area	Aligned Migration Time [s]	Peak Height	Peak Width	% of Total	Time corrected area
1	35	125.00	5,411.3	Lower Marker	29.0	43.00	63.5	1.0	0.0	66.5
2	50	15.85	480.5		6.9	45.22	6.3	1.4	0.5	15.1
3	139	616.53	6,735.5		339.7	54.13	69.6	8.1	21.1	609.8
4	152	152.85	1,525.5		87.6	55.33	71.1	1.2	5.3	153.7
5	171	139.06	1,232.1		85.3	57.08	75.7	1.1	5.0	144.9
6	183	78.18	648.7		49.9	58.15	77.6	0.6	2.9	83.2
7	194	490.88	3,839.2		325.5	59.16	78.6	4.3	18.4	532.3
8	236	88.04	564.4		65.7	62.99	59.2	1.1	3.5	100.6
9	260	63.94	373.3		50.6	65.07	54.8	0.9	2.6	74.9
10	274	45.90	253.4		37.7	66.41	51.3	0.7	1.9	54.7
11	289	53.83	281.8		45.9	67.75	49.2	0.9	2.3	65.1
12	308	41.86	206.2		37.3	69.32	47.2	0.7	1.8	51.7
13	329	31.05	142.8		29.2	71.07	45.4	0.6	1.4	39.4
14	335	30.22	136.7		28.8	71.54	44.6	0.6	1.3	38.6
15	359	28.12	118.6		28.4	73.47	44.5	0.6	1.3	37.0
16	365	49.87	206.8		51.1	73.98	43.3	1.1	2.3	66.1
17	393	43.60	167.9		47.6	76.25	40.7	1.1	2.1	59.7
18	411	40.98	151.2		45.8	77.40	39.1	1.1	2.0	56.6
19	432	27.72	97.2		31.5	78.65	37.9	0.8	1.3	38.3
20	447	30.07	102.0		34.6	79.48	37.1	0.9	1.4	41.6
21	460	23.02	75.9		26.8	80.22	36.2	0.7	1.1	31.9
22	488	33.64	104.4		40.0	81.88	35.5	1.1	1.6	46.6
23	508	38.11	113.6		46.3	82.94	35.5	1.2	1.8	53.3
24	541	24.50	68.7		31.3	84.46	33.5	0.9	1.2	35.3

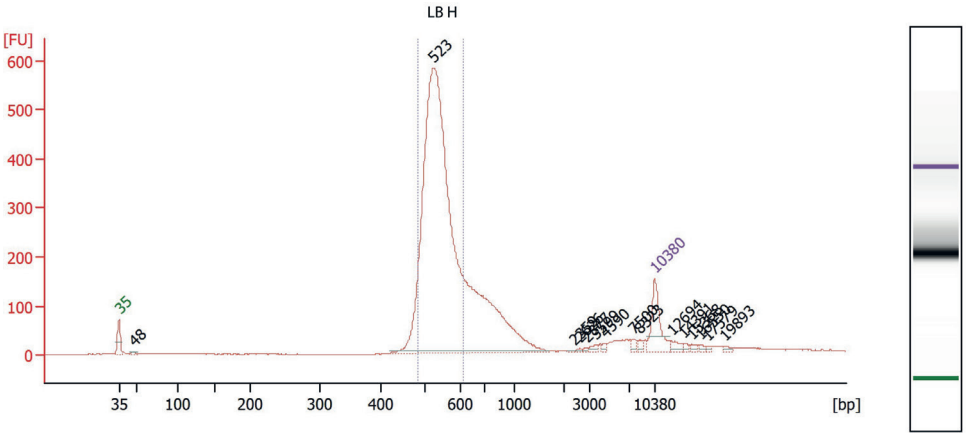
Supplemental Information B: BioAnalyzer PippinPrep results

2100 expert_High Sensitivity DNA Assay_DE13804097_2015-06-16_16-22-57.xad

Assay Class: High Sensitivity DNA Assay
Data Path: C:\...gh Sensitivity DNA Assay_DE13804097_2015-06-16_16-22-57.xad

Created: 6/16/2015 4:22:57 PM
Modified: 6/16/2015 6:23:58 PM

Electropherogram Summary Continued ...



Overall Results for sample 4 :

Number of peaks found: 15
Noise: 0.6
Corr. Area 1: 2,625.9

Peak table for sample 4 :

Peak	Size [bp]	Conc. [pg/µl]	Molarity [pmol/l]	Observations	Area	Aligned Migration Time [s]	Peak Height	Peak Width	% of Total	Time corrected area
1	35	125.00	5,411.3	Lower Marker	32.5	43.00	72.4	1.1	0.0	77.6
2	48	10.72	337.2		5.1	44.98	5.7	1.2	0.3	11.7
3	523	2,274.50	6,586.8		3,110.1	84.03	582.5	20.9	97.7	3,726.9
4	2,359	2.16	1.4		4.7	102.42	5.0	1.1	0.1	4.7
5	2,626	2.87	1.7		6.4	103.31	8.1	0.9	0.2	6.2
6	2,877	3.30	1.7		7.4	104.15	10.6	0.8	0.2	7.1
7	3,599	6.98	2.9		15.7	105.34	15.7	1.2	0.4	15.0
8	4,590	6.02	2.0		13.6	106.62	18.9	0.7	0.3	12.8
9	7,500	7.05	1.4		16.4	110.23	26.4	0.6	0.4	14.9
10	8,323	7.63	1.4		18.1	111.02	24.6	0.7	0.4	16.4
11	10,380	75.00	10.9	Upper Marker	189.0	113.00	148.8	3.1	0.0	167.6
12	12,694	0.00	0.0		35.0	115.22	23.1	1.8	0.0	30.4
13	14,391	0.00	0.0		11.8	116.86	16.5	0.7	0.0	10.1
14	15,368	0.00	0.0		13.9	117.80	14.9	1.0	0.0	11.8
15	16,550	0.00	0.0		11.9	118.93	13.4	0.9	0.0	10.0
16	17,579	0.00	0.0		9.6	119.92	11.8	0.8	0.0	8.0
17	19,893	0.00	0.0		11.4	122.15	10.6	1.2	0.0	9.3

Region table for sample 4 :

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/µl]	Molarity [pmol/l]	Color
486	612	2,625.9	67	537	5.7	1,581.37	4,469.8	

Chapter 5

Rapid divergence of mussel populations despite incomplete barriers to dispersal

Diede L. Maas, Stefan Prost, Ke Bi, Lydia Smith, Ellie E. Armstrong, Ludi P. Aji, Abdul H.A. Toha, Rosemary G. Gillespie, Leontine E. Becking

Molecular Ecology ***2018****, 27, 1556-1571*

Abstract

Striking genetic structure among marine populations at small spatial scales is becoming evident with extensive molecular studies. Such observations suggest isolation at small scales may play an important role in forming patterns of genetic diversity within species. Isolation-by-distance, isolation-by-environment, and historical priority effects are umbrella terms for a suite of processes that underlie genetic structure, but their relative importance at different spatial and temporal scales remains elusive. Here, we use marine lakes in Indonesia to assess genetic structure and test relative roles of the processes in shaping genetic differentiation in populations of a bivalve mussel (*Brachidontes* sp.). Marine lakes are landlocked water bodies of similar age (6,000 – 12,000 years), but with heterogeneous environments and varying degrees of connection to the sea. Using a population genomic approach (double-digest Restriction-site Associated DNA sequencing), we show strong genetic structuring across populations (range F_{ST}: 0.07 – 0.24) and find limited gene flow through admixture plots. At large spatial scales (>1400km), a clear isolation-by-distance pattern was detected. At smaller spatial scales (<200km), this pattern is maintained, but accompanied by an association of genetic divergence with degree of connection. No signatures of isolation-by-environment were found. We hypothesize that (incomplete) dispersal barriers can cause initial isolation, allowing priority effects to give the numerical advantage necessary to initiate strong genetic structure. Priority effects may be strengthened by local adaptation, which our data potentially corroborates by showing a high correlation between mussel genotypes and temperature. Our study indicates an often-neglected role of evolution-mediated priority effects in shaping population divergence.

Keywords: population genomics, ddRADseq, isolation-by-distance, isolation-by-environment, priority effects, marine biodiversity

Introduction

In the marine realm, barriers to dispersal have been assumed to be few, and speciation rates to be slow compared to terrestrial ecosystems (Palumbi et al., 1994; Cowen et al., 2000; Carr et al., 2003). Recently, these long-held assumptions have been overturned by studies in marine taxa showing high population genetic structuring at small spatial scales (Barber et al., 2006; Marshall et al., 2010; Carpenter et al., 2011; Neves et al., 2016; Gonzalez et al., 2017). Isolating processes at small spatial scales may therefore play an important role in forming genetic diversity within species. The relative importance of processes shaping genetic structure on different spatial and temporal scales, however, remains elusive (Vellend, 2005; Bowen et al., 2013). Anthropogenic impacts on ecosystems via climate change and fragmentation of habitats are evident (Hoegh-Guldberg and Bruno, 2010; Haddad et al., 2015). Insights into drivers of biodiversity, and how populations may respond and disperse to colonize new habitats will be valuable knowledge for adequate management of for example marine protected areas.

Isolation acts to decrease the rate of dispersal, and thus decrease of gene flow, which can enhance speciation by allowing populations to adapt, diverge, and ultimately form new species (Hendry et al., 2009). Within the context of population genetics, generally two modes of isolation are considered: (a) isolation-by-distance due to geographic distance or physical barriers, irrespective of environment (Wright, 1943), and (b) isolation-by-environment due to environmental dissimilarity, irrespective of geographic distance (Wang and Summers, 2010; Wang and Bradburd, 2014). Recently, a third mode has come into more consideration in the context of population genetics: (c) historical priority effects (Orsini et al., 2013; Fukami, 2015; De Meester et al., 2016). The three modes are not mutually exclusive, and may have different relative importance at different spatial and temporal scales in the formation of population structure (Orsini et al., 2013). The three patterns provide insights into the underlying types of processes that shape diversity, such as larval mobility, selection against migrants, and competition.

The three modes differ in the predictions they make on the relationship between genetic distance and geographic and environmental distance. Isolation-by-distance patterns arise when there is a spatial reduction in movement of individuals through the process of dispersal limitation, and is opposed to a panmictic population with freely dispersing propagules (Wright, 1943). It predicts that genetic differentiation increases with geographic distance between populations. With increasingly larger distances, propagules become more diluted, which lowers the chance of gene flow or colonization (Johannesson, 1988). With restricted gene flow, populations will locally accumulate genetic differences via stochastic genetic drift. Isolation-by-environment predicts gene flow between environmentally divergent habitats is limited due to biased dispersal, or selection against migrants via reduced fitness of immigrants in the new environment or reduced hybrid fitness (Rundle and Nosil, 2005; Schluter, 2009; Wang and Bradburd, 2014). Environmentally similar habitats are in contrast expected to have ongoing gene flow, which may benefit local adaptation through

incorporation of pre-adapted alleles (Wang and Bradburd, 2014). Finally, priority effects emphasize the importance of first colonizers in shaping subsequent population genetic structure, which is also termed historical contingency (Orsini et al., 2013; Fukami, 2015). Priority effects may have an ecological and an evolutionary component. The ecological advantage of being the first colonizer through competitor or predator release may result in a numerical dominance of certain genotypes that can be difficult to overcome by subsequent immigrating genotypes (De Meester et al., 2016). The advantage may be enhanced by an evolutionary component of priority effects via rapid local adaptation of the first genotypes. This gives a head start to early colonizers towards locally adapting and therefore becoming stronger competitors to late-arrivers, effectively fixing the initial stochastic numerical advantage in the long term (de Meester et al., 2002; De Meester et al., 2016). Priority effects predict that even on small spatial scales, and with incomplete barriers to dispersal, gene flow can still be limited between populations due to density-dependent and evolution-mediated dominance of early genotypes, irrespective of geographic and environmental distance. Priority effects are expected to be stronger in more isolated habitat patches, as the time lag between immigration events will be sufficient to attain a numerical advantage for early colonists, potentially aided by a head start for local adaptation (Fukami, 2015).

A key issue is that modes of isolation can be difficult to distinguish, as they are frequently confounded (Legendre, 1993; Lee and Mitchell-Olds, 2011). Environmental data is often spatially structured at multiple scales, and may change over time, which clouds the ultimate cause of divergence. Island systems have classically been used as model systems since they alleviate confounding factors as islands provide a clearly defined spatial, temporal and environmental context (MacArthur and Wilson, 1967; Warren et al., 2015). Therefore, island-like systems are ideal systems to test relative importance of modes of isolation underlying genetic structure. Here, we focus on marine lakes which, like terrestrial islands, are 'natural laboratories' in that they harbor discrete marine populations, replicated over space and time, and present different environmental regimes (Becking et al., 2011). Marine lakes are land-locked bodies of seawater with varying degrees of connection to the surrounding ocean via subterranean fissures and pores (Holthuis, 1973; Tomascik and Mah, 1994; Hamner and Hamner, 1998; Dawson et al., 2009). The lakes originated after the Last Glacial Maximum (20,000 years ago) when natural depressions in karstic landscapes filled with rising sea water level (Sathiamurthy and Voris, 2006). Hence, they are relatively young systems estimated to be about 6,000 – 12,000 years old (Dawson et al., 2009). Large numbers of marine lakes are found in the Coral Triangle in Indonesia (Becking et al., 2011, 2015). This region is characterized by extremely high marine biodiversity (Hoeksema, 2007; Mangubhai et al., 2012). As all marine lakes maintain a connection to the surrounding sea, there is a continued vector for propagules to move in and out the lakes, namely via tidal fluctuations.

Most marine lakes in Indonesia harbor the diploid bivalve mussel *Brachidontes* spp. (Swainson 1840) (Mollusca; Bivalvia; Mytilidae). *Brachidontes* spp. shells have been found in the deepest layers of sediment cores from marine lakes, and thus have been some of the

first colonists of the marine lakes. Species of the genus *Brachidontes* can form large beds by attaching themselves to substrates in and below intertidal areas (Terranova et al., 2007). They are broadcast spawners, and have a dispersive planktonic larval stage for a duration of up to four weeks (Monteiro-Ribas et al., 2006). Previous work on mussels has shown that *Brachidontes* spp. from marine lakes and mangroves in the Indo-Pacific fall into six genetically distinct lineages that likely represent six separate species (Goto, Tamate & Hanzawa, 2011; Becking et al., 2016). Recently, studies using a mitochondrial DNA marker (COI) found that for 22 marine lakes studied each lake contained only one lineage, which may be diverging *in situ* (Becking et al. 2016; de Leeuw, et al., *under review*). We currently expand on these studies that were based on single markers, by using multiple markers generated from high-throughput double-digest Restriction-site Associated DNA sequencing (ddRAD) to assess genome-wide signals.

By comparing marine lakes in Indonesia on multiple spatial scales, with similar ages and sizes, but with varying degrees of connection to the sea and differing environmental regimes, we tested the relative contribution of isolation-by-distance, isolation-by-environment and historical priority effects in the formation of population genetic structure of a bivalve mussel *(Brachidontes* sp.). On large spatial scales (>1400km), we expect to find a pattern of isolation-by-distance, as differentiation in genotypes most likely pre-dated the origin of marine lakes. On scales where propagule dispersal is still expected (<200km), we expect low levels of genetic differentiation which will only show moderate isolation-by-distance. Contrastingly, if isolation-by-environment plays a role we expect marine lakes that are environmentally similar will be genetically similar as well, regardless of geographic scale. Finally, if priority effects are at play, we would see high levels of differentiation even on small spatial scales (<40km), where more connected lakes would be more similar to each other as they would have less time lag to new genotypes coming into the lake.

Materials and Methods

Sample locations

The study encompasses two regions: Berau in East-Kalimantan (Fig. 1A) and Raja Ampat (including Gam and Misool islands) in West Papua (Fig 1B, C), and three spatial scales (1400km, 200km, and 40km). The islands of Berau are part of the Berau marine protected area (Becking et al., 2013a). Berau has a tropical rainforest climate with no clear difference in rainy and dry seasons, except for an increase in winds between December and March (Tomascik and Mah, 1994; Becking et al., 2013a). We included a lake from the Berau region as a distant population when comparing its patterns to those found in West Papua. Raja Ampat is part of the Coral Triangle, which is famous for its high marine biodiversity (Hoeksema, 2007). Raja Ampat lies on the equator and has a tropical climate with yearly precipitation ranging from 2500-4500 mm, with monsoons being the main characteristics of seasonal change. North-western monsoons from November-March and south-eastern monsoons from May-October are characterized by persistent winds. Within the Raja Ampat Regency there is a multitude of islands of karstic rock, which causes the coastline to be

irregular (Becking et al., 2011). Currents among complex coastlines create local turbulence, which is expected to benefit larval connectivity among reefs (Starger et al., 2015). Reefs of Raja Ampat are mostly shallow fringing, lagoon, and atoll reefs (Mangubhai et al., 2012). Average sea surface temperature is 29.0°. At least 45 marine lakes have been identified in Raja Ampat, with the highest density found in the Misool area (Becking et al., 2011, Fig. 1C). Most marine lakes in this area are in a pristine state with no apparent anthropogenic influence.

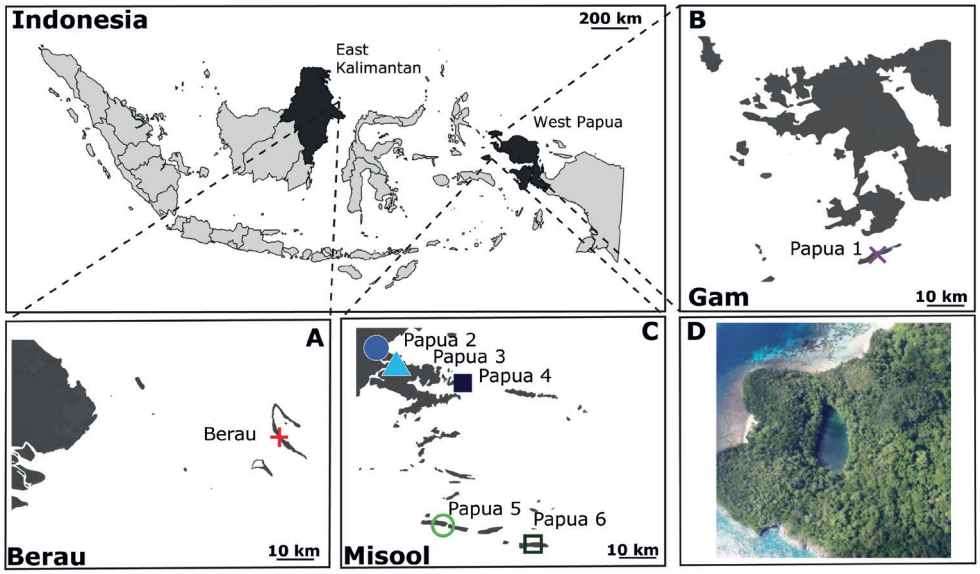


Figure 1: Overview of Brachidontes *sp. sampling sites in seven marine lakes in Indonesia. Land is indicated in dark grey, and sea is indicated in white. Marine lakes are located inland on different landmasses and are indicated with a symbol. (a) Berau, East Kalimantan, (b) Papua 1, Gam, Raja Ampat. (c) Papua 2-6, Misool, Raja Ampat. (d) Marine lake photographed by L.E. Becking. There is approximately 1,400km distance between East Kalimantan and West Papua.*

Sample collection and environmental profiling

We collected specimens of the bivalve mussel *Brachidontes* sp. from seven marine lakes in 2011. From sediment cores we know these mussels were present at the early stages of formation of the lake (*Unpublished data*). All *Brachidontes* sp. samples were confirmed to be of the same genetic lineage (Lineage A, as defined by Goto et al. (2011) and Becking et al. (2016). One lake was sampled in East-Kalimantan (Fig. 1A), and six in Raja Ampat (Fig. 1B, C). Codes, locations, number of specimens per lake, and lake characteristics are recorded in Table 1. In total, 125 *Brachidontes* sp. samples were collected. The study encompassed three geographical scales: 1) >1400km, including Berau, Gam and Misool areas, 2) <200 km, including Gam and Misool areas, and 3) <40km, including only the Misool area. We logged lake coordinates using a Global Positioning System (Garmin Gpsmap 64S). Adductor muscles from *Brachidontes* sp. samples were excised and preserved in 99%

ethanol at 0-4°C while in the field (4-8 weeks). In the laboratory, tissue samples were stored in a -80°C freezer until further use.

Lake coordinates were used to calculate minimum pairwise geographic distances between lakes via the R function *distm* implemented in the geosphere package. Environmental characteristics defined as temperature (°C), salinity (ppt) and pH were measured using a YSI Professional Plus Multimeter from an inflatable boat (Table 1). Measurements were taken at 1-meter intervals from 1m to 5m depth in each lake, as the bulk of *Brachidontes* sp. bivalves are found within this range. In each lake, we measured at least 10 different sites. Pairwise environmental distance between lakes was defined via calculating Euclidean distances between lakes in a Principal Component Analysis ran with the three environmental variables. The tidal amplitude was measured simultaneously inside of the lakes as well as in the adjacent sea using Hobo water level loggers. Physical connection of the lake to the sea was defined as the ratio of maximum tidal amplitude in meters of the lakes as compared to the sea (this was termed 'Connection', Equation 1).

$$Connection = \frac{Max.tidal\ amplitude\ lake\ (m)}{Max.tidal\ amplutide\ sea\ (m)} \quad (1)$$

To get a measure of 'Isolation' instead of 'Connection', Connection was subtracted from 1. (Equation 2). Now, lakes with a high connection ratio (e.g., 0.8) had a low value for Isolation (e.g., 0.2). Conversely, low connected lakes (e.g., connection ratio of 0.1) had high values for Isolation (e.g., 0.9).

$$Isolation = 1 - Connection \quad (2)$$

Finally, pairwise distances between lakes were obtained by multiplying the Isolation values of each pairwise combination of lakes (i and j, Equation 3), making sure the diagonal (lake i with itself) was put at zero. Low connection distance values indicate two lakes both with high connection (e.g., Isolation value 0.2 * Isolation value 0.2 = 0.04), and high connection distance values indicate two lakes with low connection (e.g., Isolation value 0.9 * Isolation value 0.9 = 0.81). Intermediate connection distance values indicate lakes with varying degrees of connection (e.g., 0.9 * 0.2 = 0.18).

$$Connection\ distance_{i,j} = Isolation_i * Isolation_j \quad (3)$$

Library preparation and sequencing

DNA from the excised adductor muscles was extracted using Qiagen DNeasy kit (Qiagen, Germantown, MD, USA). We built ddRAD-seq libraries for the 125 individuals following an adapted protocol of Peterson *et al* (2012) See Supporting Information for protocol). Briefly, we started the ddRAD-seq protocol with 600 nanograms of DNA. The DNA was digested using SphI-HF (rare-cutting) and MluCI (frequent-cutting) restriction enzymes. The fragmentation was investigated with Bioanalyzer High Sensitivity chip (Agilent). Following

the spreadsheet of Peterson *et al* (2012), "Locus count from Bioanalyzer % in region", the combination of enzymes was predicted to yield 64,770 sequencable fragments when excising a region of 400-500 bp, assuming a genome size of 1 Gb. We pooled 17-18 individuals per ddRAD library and came to a total of 7 libraries for 125 specimens. We used a Sage Science Pippen Prep to size select 476-576 bp (including internal adapters) fragments. We confirmed the sizes by Bioanalyzer High Sensitivity chip (Agilent). Then, ten indexing polymerase chain reaction cycles (PCRs) were run on each library to enrich for double-digested fragments and to incorporate a unique external index for each library pool. The 7 libraries were sequenced as 100bp single reads (SR) sequences on two Illumina HiSeq 2500 lanes at the Vincent J. Coates Genomic Sequencing Facility at UC Berkeley (SR100).

Assembly reference and filtering

We used pipelines implemented in a custom perl script invoking a variety of external programs to process ddRAD-seq data (RADToolKit v0.13.10; https://github.com/CGRL-QB3-UCBerkeley/RAD). Raw fastq reads were first de-multiplexed based on the sequences of internal barcodes with a maximum tolerance of one mismatch. De-multiplexed reads were removed if expected cutting sites were not found at the beginning of the 5'-end of the sequences. The resulting reads were then filtered using cutadapt (Martin, 2011), and Trimmomatic (Bolger et al., 2014) to trim off Illumina adapter contaminations and low quality reads. We also removed low complexity and potential bacterial, viral and human contamination sequences, using the genomes from GenBank represented in Supplemental Table 1. Exact duplicates, either derived from PCR or from sequencing, could not be determined due to the nature of ddRAD data, so these were not eliminated.

The resulting cleaned reads of each individual were clustered using cd-hit (Li and Godzik, 2006; Fu et al., 2012), and only clusters with at least three reads supported were kept. For

Table 1: Overview of measured variables in sampled locations. Codes and sizes, physiographic, environmental and genetic parameters are displayed. Temperature, salinity and pH were measured at 1-m intervals from 1m to 5m depth in at least 10 locations per lake. Genetic measures nucleotide diversity (θ_π) and Tajima's D neutrality values (D) are displayed. Means are displayed with standard deviations. Codes correspond with Figure 1.

Code	Sample size (n)	Surface area (m^2)	Max. depth (m)	Fraction tidal amplitude (Lake/ Sea)	Temp. (°C)	Salin. (ppt)	pH	θ_π (%)	Tajima's D
Berau	15	$10*10^4$	-	0.38	29.5 (0.71)	26.0	-	0.90 (0.82)	-0.24 (0.85)
Papua 1	18	$8.6*10^4$	19	0.07	32.3 (0.13)	24.0 (0.71)	7.6 (0.05)	0.81 (0.76)	-0.30 (0.85)
Papua 2	19	$1.5*10^4$	7	0.31	33.6 (0.58)	25.6 (0.52)	7.8 (0.08)	0.87 (0.78)	-0.33 (0.90)
Papua 3	15	$1.0*10^4$	15	0.51	32.6 (0.62)	30.7 (0.84)	8.0 (0.06)	0.93 (0.76)	-0.62 (0.77)
Papua 4	16	$1.3*10^4$	20	0.80	31.7 (0.35)	25.9 (0.84)	8.1 (0.06)	0.94 (0.78)	-0.52 (0.84)
Papua 5	19	$0.6*10^4$	5	0.26	31.5 (0.28)	28.9 (0.26)	8.1 (0.05)	0.88 (0.76)	-0.52 (0.84)
Papua 6	23	$0.3*10^4$	12	0.78	31.9 (0.31)	28.3 (0.26)	7.9 (0.03)	0.89 (0.75)	-0.36 (0.88)

each cluster, the representative sequence determined by cd-hit was retained. The markers were then masked for putative repetitive elements, low complexities and short repeats with Ns using RepeatMasker (Smit et al., 2014), with "Mytilidae" as a database. After masking, we eliminated markers if more than 60% of the nucleotides were Ns. The resulting RAD markers from each individual were then combined and clustered for all individuals to search for those shared by at least 70% of all the individuals, which then served as the reference.

Cleaned sequence reads from all individuals were aligned to the reference using Novoalign (http://www.novocraft.com), and only reads that mapped uniquely to the reference were kept. We used Picard (http://www.picard.sourceforge.net) to add read groups and GATK (McKenna et al., 2010) to perform realignment on alignment files in BAM format, generated by SAMtools (Li et al., 2009). We then used SAMtools/BCFtools to generate data quality control information in VCF format. We filtered out any markers where more than two alleles were called on any site. We masked SNPs/sites within 10 bp upstream and downstream around an indel. Individual sites were eliminated if their depth fell outside 1st and 99th percentile of the overall coverage (among all samples). These data were further filtered using a custom sites filtering program, SNPcleaner (Bi et al., 2013), which was modified and implemented in our pipelines.

Genotype likelihoods and allele frequency estimation

In order to assess potential artefacts of coverage of SNPs and percentage included individuals we tested four filtering options in total (Supplemental Table 2), varying these two parameters. We tested coverage of 3X, 5X and 10X, and percentage included individuals of 40% and 70%. We found that downstream patterns remained highly similar (but see Supplemental Figs 1 and 2, and Supplemental Table 3). This indicated that there is no large influence of minimum coverage or percentage of included individuals on biological patterns. All further results are based on SNPs having at least 3X coverage and a minimum of 70% included individuals.

SNP and genotype calls based on allele counts might show high uncertainty and could cause potential bias or introduce noise in downstream analyses (Johnson and Slatkin, 2008; Lynch, 2008). To account for uncertainty in our data, we used genotype likelihoods instead of genotype calls whenever possible. Genotype likelihoods were calculated in an empirical Bayesian framework, implemented in ANGSD (http://www.popgen.dk/angsd/index.php/ANGSD) (Korneliussen et al., 2014). This software is specialized in analysing low to medium coverage next generation sequencing data. Most of the downstream analyses implemented in ANGSD were performed based on likelihood of site allele frequencies, genotype likelihood or genotype posterior probabilities. For some analyses done by external programs that rely on called genotypes, we used genotype posterior probability of 0.95 as a cut-off to output a list of high confidence variants.

Within-population diversity and demography

Overall genetic variation within marine lakes was estimated via nucleotide diversity: through the calculation of average number of pairwise differences between sequences, theta pi (θ_π; Nei 1987), and through Watterson's Theta as the total number of segregating sites (θ_W; Watterson 1975). We ran Pearson correlations of nucleotide diversity (θ_π) versus environmental variables and connection fractions to see how these variables affect within-population diversity. From the thetas, we computed Tajima's D as a neutrality test to examine genomic evidence for population expansion or decline (Tajima, 1989). We computed per-individual inbreeding coefficients (F), calculated from degree of deviation from Hardy-Weinberg equilibrium via ngsF (Vieira et al., 2013). Furthermore, we computed Stairway plots (Liu and Fu, 2015) to estimate changes in effective population size (N_e) over time, using a generation time of 1 year (Morton, 1988) and a range of mutation rates from $1.0 * 10^{-8}$ to $3.5 * 10^{-8}$ per site per generation.

Population genetic structure

We summarized genotypic differentiation among marine lake populations using different strategies. First, we performed a Principal Components Analysis (PCA) of the covariance matrix of posterior genotype probabilities as implemented in ngsTools (http://github.com/mfumagalli/ngsTools) (Fumagalli et al., 2014). PCAs are commonly used in analysing SNP data, since it is an unsupervised clustering method, which may discern population structure in an unbiased manner. The first six Principal Components were included based on their eigenvalues. We further explored the data by performing a neighbor-joining network (NeighborNet) analysis based on uncorrelated p-distances in Splitstree (Huson, 1998; Huson and Bryant, 2006). This shows how well the data would fit a phylogenetic tree, without forcing a tree-like structure onto the data. Furthermore, a genetic distance matrix was computed from genotype probabilities via the program ngsDist (Vieira et al., 2016). A bootstrapped Neighbor-Joining tree matrix was computed from 1000 possible trees via RAxML (Stamatakis, 2014), converted to a phylogenetic tree via FastME (Lefort et al., 2015), and visualized in FigTree v.1.4.2 (Rambaut, 2009).

Genetic differentiation among populations as summarized by the fixation index (F_{ST}) was calculated in ANGSD (Analysis of Next-Generation Sequencing Data) using the shared site frequency spectrum of each pairwise combination of lakes (2dSFS) (Korneliussen et al., 2014). Next, we explored genetic structure across populations via admixture analysis implemented in ngsAdmix (Skotte et al., 2013). By calculating admixture proportions per individual, the ancestry of populations could be defined. Finally, a connectivity network was computed using the Nei's G_{ST} calculation implemented in the diveRsity package of R (Sundqvist et al., 2016).

Inferences on modes of isolation

We used F_{ST} values to calculate pairwise genetic differentiation ($F_{ST}/(1-F_{ST})$) (Slatkin, 1995). We used Mantel's tests (Mantel, 1967; Slatkin, 1993) to test significance among distance matrices to elucidate the importance the different processes causing isolation. We ran

Mantel's tests via the function *mantel* from the R package *vegan* (Oksanen et al., 2016), with 1,000 permutations. A correlation of r > 0.6 was considered strong, and significance was assigned when the *P*-value was smaller than 0.05. We confirmed the absence of a correlation between geographic and environmental distance (Mantel's test, r = -0.25, *P*-value = 0.78), which allowed us to distinguish between scenarios of dispersal limitation and adaptation to the environment. We did the same test for geographic and connection distance, and environmental and connection distance, and similarly no correlations were found (Mantel's test, r = 0.34, *P*-value = 0.21, and r = 0.25, *P*-value = 0.24, respectively).

Finally, we further explored the effect of environmental variables influencing genetic structure by correlating the first principal component from a principal component analysis run for the lakes in Raja Ampat (explaining 10.73% of total genetic variation, Supplemental Fig. 13) to environmental variables.

Results

Filtering

We obtained 103 million reads after demultiplexing from the Illumina sequencer. The total number of reads per individual ranged from to 43536 to 2966453. From these reads, we obtained 761,014 ddRAD loci and 116,416 anonymous SNPs usable in downstream analyses. Principal Component Analyses, Splitstree and F_{ST} estimates remained highly similar with different filtering options (coverage: 3X, 5X, or 10X, included individuals 40% or 70%) (Supplemental Fig. 1 and 2, Supplemental Table 3). Therefore, all subsequent results are based on SNPs having at least 3X coverage and 70% included individuals. On average, samples had a coverage of 12X (range: 3 – 42X) across all loci.

Within-population diversity and demography

First, we investigated genetic diversity within *Brachidontes* sp. populations from seven marine lakes. Per population nucleotide diversity (θ_π) was similar among all lakes (Table 1, θ_π). The highest nucleotide diversity was observed for Papua 4 (0.0094), and lowest for Papua 1 (0.008). Although the standard deviation of both θ_π and θ_W was high, they showed consistent patterns per lake (Table 1, Supplemental Table 4, Supplemental Fig. 4). We observed a trend towards higher nucleotide diversity with increasing connection to the sea (Pearson's correlation, r = 0.78, p = 0.04) (Supplemental Fig. 5). Inbreeding coefficients (F) were generally found to be low: range (0.003 – 0.041) (Supplemental Table 4).

All lakes were found to have slightly negative Tajima's D values (average D-value: -0.41, standard deviation: 0.14) (Table 1, Supplemental Fig. 6). The lowest value of Tajima's D was observed for Papua 3 (-0.62) and the highest for Berau (-0.24). Negative Tajima's D values are generally associated with populations showing recent expansion after a bottleneck. Our Stairway Plots consistently showed a population decline at 4,000 to 6,000 years before present for each population, potentially indicating the colonization of the lakes by *Brachidontes* sp. (Supplementary Fig. S7).

Population differentiation and structure

Next, we investigated population structure. Clear structuring was observed in the first four principal components of the Principal Component Analysis, together explaining 28.81% of total genetic variation (Fig. 2B). The first principal component, explaining 10.87% of the total variation, separated East-Kalimantan (Berau) from the rest. The second principal component (explaining 8.50%) separated Papua 1 and 2 from the remaining lakes in Misool. Principal components 3 and 4 (together explaining 9.44% of the variation) clearly separated the lakes in southern Misool (Papua 5 and 6) from the rest. None of the principal components separated Papua 3 and 4.

The pattern observed in the principal components analyses was supported by the neighbor-joining network constructed using Splitstree (Fig. 2C). The network was based on 390 splits and had a fit of 99.3. The degree of reticulation was assessed with a delta-score of 0.224

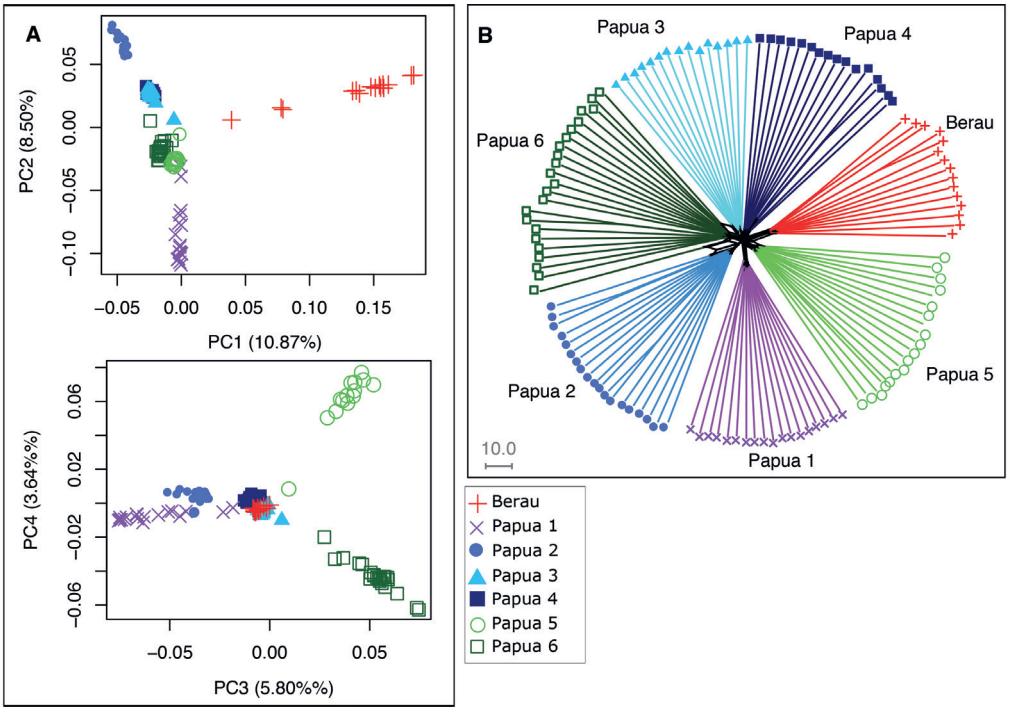


Figure 2: Genetic structure of seven populations of Brachidontes *sp. mussels in marine lakes in Indonesia. (a) Principal Component Analysis (PCA) based on pairwise genetic covariance among 125 individuals with 116,415 SNPs, from all seven locations. First four axes represent 28.8% of total genetic variation. Each dot is one individual. (b) Neighbor-Joining network with equal angles computed in Splitstree from pairwise genetic distances. Splitstree based on 390 splits and a delta score of 0.224. Scale bar indicates number of substitutions per site. Colors and codes correspond to Fig. 1 and Table 1.*

and a Q-residual score of 9.28 $*10^{-5}$. These relatively low scores show that the data followed a tree-like pattern (Holland et al., 2002). All lakes showed distinct separation, with Berau and

Papua 1 being most distant from the rest. Papua 6 showed additional splits with Papua 2 and 3, indicating that some individuals might share genetic variation (Supplemental Fig. 8). The Neighbor-Joining tree showed congruent patterns (Supplemental Fig. 9). Our connectivity network confirmed patterns from F_{ST} comparisons and the Splitstree network (Supplemental Fig. 10). Lowest bi-directional connectivity was found between Berau to all other lakes (range 0.18 – 0.28). Highest bi-directional connectivity was found between Papua 3 and Papua 4 (0.88 – 1).

Pairwise population fixation indices (F_{ST}), showed moderate to high genetic structuring between all pairwise lake comparisons (Table 2, Supplemental Fig. 11). Pairwise F_{ST} values varied between 0.069 (Papua 3 and 4), and 0.235 (Berau and Papua 1). On average, marine lakes had a global fixation index of 0.145 (standard deviation = 0.05). Admixture analysis showed clear distinction between lakes (Fig. 3). Conversion of likelihood values of 5 replicate runs showed the lowest variance for 7 ancestral populations (Supplemental Fig. 12). When putative ancestral populations were set to K=7, all seven marine lakes were separated. Increasing the number of ancestral populations from K=3 to K=7 showed a high similarity of Papua 3 and 4, as these lakes only became separate at K=7. Some individual admixture between Papua 2, 3 and 6 could be seen, consistent with the additional splits in the Neighbor-Joining Network (Fig. 2B).

Table 2: Genetic differentiation among seven Brachidontes *sp. populations from marine lakes in Indonesia. Genetic fixation indices (F_{ST}) are displayed, showing strong genetic differentiation. Codes correspond with Fig. 1 and Table 1.*

	Berau	Papua 1	Papua 2	Papua 3	Papua 4	Papua 5
Papua 1	0.235	-				
Papua 2	0.213	0.179	-			
Papua 3	0.184	0.148	0.082	-		
Papua 4	0.191	0.155	0.104	0.069	-	
Papua 5	0.198	0.153	0.138	0.106	0.116	-
Papua 6	0.198	0.151	0.126	0.089	0.108	0.102

Relative importance of modes of isolation

We found clear evidence for isolation-by-distance on large spatial scales (> 1400km) with a strong and significant correlation between geographic distance and genetic distance (Mantel's test, $r = 0.92$, $p < 0.001$). Whereas environmental and connection distance both were not significant (Mantel's test, $r = 0.44$, $p = 0.10$, and $r = 0.46$, $p = 0.11$, respectively) On the spatial scale of 200km, we found associations with geographic distance (Mantel's test, $r = 0.82$, $P = 0.01$) and connection distance (Mantel's test, $r = 0.63$, $P = 0.03$), but not for environmental distance (Mantel's test, $r = 0.21$, $P = 0.29$) (Fig. 4.). Remarkable is the wide range of genetic differentiation even within relatively well-connected lakes (Fig. 4C). On the smallest spatial scale (40km), all associations with geographic, environmental and connection distance become less pronounced and insignificant (Mantel's tests, geography: $r = 0.50$, $p = 0.05$, environment: $r = 0.16$, $p = 0.35$, connection: $r = 0.40$, $p = 0.19$). Furthermore, we observed a significant correlation of water temperature to the first principal

component (Fig. 1) on the two smaller spatial scales (200km and 40km) (Spearman's correlation, rho = 0.49 and 0.78, respectively, $p < 0.001$) (Fig. 5).

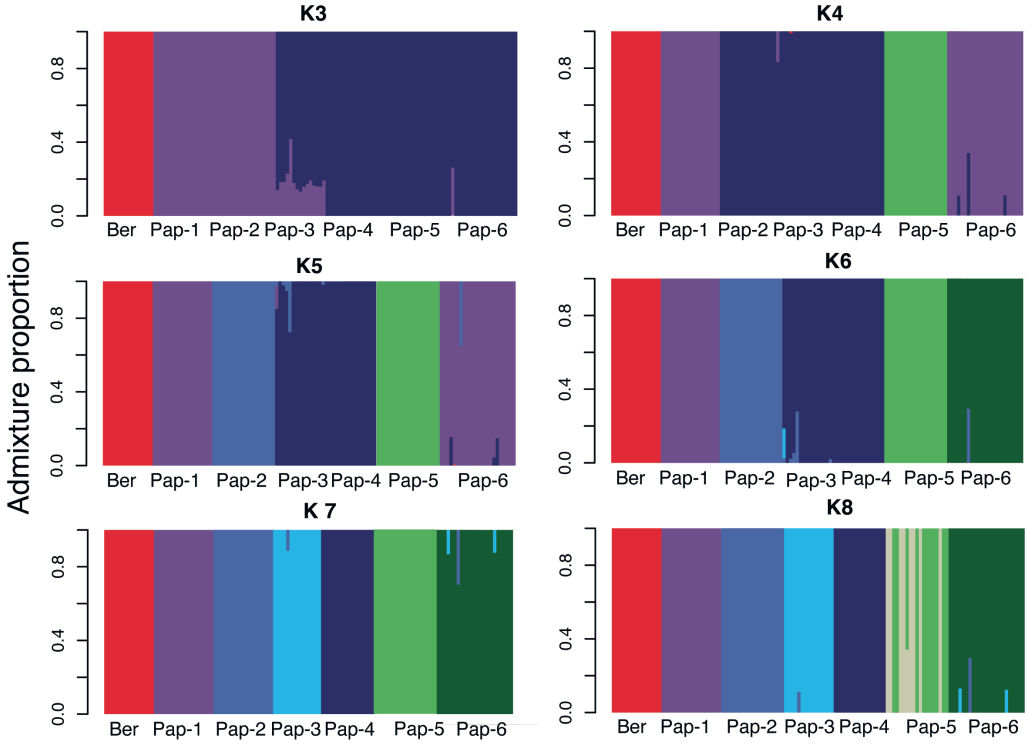


Figure 3: Individual admixture plots of Brachidontes *sp. populations from seven marine lakes in Indonesia. Bayesian admixture analysis for a range of putative ancestral populations (K) based on genotype likelihoods via ngsAdmix. Highest likelihood was obtained for K = 7 (Supplemental Fig. 12). Each bar represents one individual. Colors within bars represent admixture proportions and correspond to Fig. 1 and Table 1.*

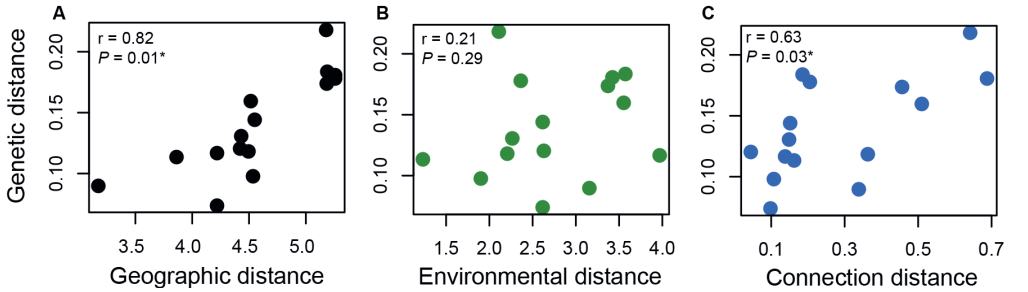


Figure 4: Relative importance of three modes of isolation tested for Brachidontes sp. populations from six marine lakes in West Papua, Indonesia (scale <200km). Mantel test results of genetic distance (normalized F_{ST}) versus (a) geographic distance (log-transformed m), (b) environmental distance (Euclidean distances based on PCA of temperature, salinity and pH), (c) connection distance. Mantel observation values (r) and p-values are displayed. Each dot represents a pairwise comparison between two populations. Asterisks represent significance at the alpha = 0.05 level.

Discussion

Understanding patterns underlying population genetic structure in the marine realm is critical for predicting ecosystem responses to natural and anthropogenic change. By comparing marine lakes in Indonesia with similar ages and sizes, but varying degrees of connection to the sea and differing environmental regimes, we tested the relative contribution of isolation-by-distance, isolation-by-environment, and historical priority effects in the formation of population genetic patterns of a bivalve (*Brachidontes* sp.) in marine lakes. Our results indicate strong genetic structure despite incomplete dispersal barriers, and provides important insights into the role that (evolution-mediated) priority effects may play in influencing rapid population divergence in peripheral environments.

As marine lakes are landlocked water bodies with subterranean connection to the sea, they can be regarded as the marine equivalent of terrestrial islands. Marine lakes provided novel habitat to be colonized, and we assume that the floodwaters that filled the lakes brought in the propagules of populations from the source pool and that these were the progenitors of current marine lake populations (Dawson and Hamner, 2005). Preliminary studies of sediment cores from the marine lakes indicate that *Brachidontes* was present at the onset of the lakes in Raja Ampat (*Unpublished data*). We assume that the degree of connection between lakes and the sea has remained similar since the formation of the lake, and the potential remained for novel migrants to come into the lakes.

We show that short term (<6,000 years) reduction in connection between populations can lead to strong population divergence. The role of priority effects is often neglected in population genetic and phylogeography studies, but may be relevant in the context of ongoing environmental change and habitat fragmentation, which all influence landscape connectivity. We first discuss within-lake structure and demography, then elaborate on the observed high population differentiation, and finally disentangle the importance of the three patterns underlying the structure: isolation-by-distance, isolation-by-environment and (evolution-mediated) priority effects.

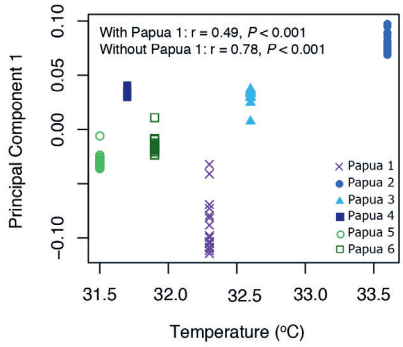


Figure 5: Correlation of genetic variation of Brachidontes *sp. populations with water temperature in marine lakes from West Papua, Indonesia (scale <200km). Correlations between first principal component of Supplemental Fig. 13 (explaining 10.73% of genetic variation among populations Papua 1-6) and temperature. Spearman's correlation values (rho) and p-values are displayed for correlations including Papua 1 (spatial scale 200km) and excluding Papua 1 (spatial scale 40km). Colors and codes correspond to Fig. 1 and Table 1.*

Marine lake colonization and population structure

We investigated early colonization of *Brachidontes* sp. mussels of the marine lakes based on analyses for bottlenecks (Table 1, Supplementary Fig. 7). Our findings are consistent with mitochondrial DNA data (COI) based on the same mussel lineage analysed in this study (Becking et al., 2016; de Leeuw et al., *under review*). Their mismatch distributions followed a unimodal distribution which is typical for bottlenecked populations and subsequent expansions (Becking et al., 2016). In the current study, the estimated timing of bottlenecks events found for the majority of lakes is approximately 4,000 to 6,000 years before present, which corresponds to the presumed dates of filling of the marine lakes (Sathiamurthy and Voris, 2006). We do note that the estimations may be crude, with large confidence intervals, and are dependent on a variety of assumptions (mutation rate and generation time). Bottlenecks with subsequent expansions have also been found in studies of other peripheral environments (Dawson and Hamner, 2005; Goto et al., 2011; Gotoh et al., 2011; Hanzawa et al., 2012), which show rapid accumulation of mutations in populations after colonization of the habitat in a typical pattern of founder effects.

We found nucleotide diversities of marine lake populations to be relatively high (ranging between 0.008 and 0.010, Table 1) (Pazmiño et al., 2017). High nucleotide diversity is generally not expected in bottleneck scenarios due to a loss of rare alleles and increased inbreeding (Fauvelot et al., 2003; Hohenlohe et al., 2010; Gotoh et al., 2011). Indeed, inbreeding coefficients for marine lake populations were found to be low (F mean 0.018 ± 0.024), where severe inbreeding is indicated by F values larger than 0.1 (Vieira et al., 2013; Hazzouri et al., 2015; Wang et al., 2016). We also observed a trend towards higher nucleotide diversity with increasing connection of the lake to the surrounding sea. With higher connection to the sea, the influx of genetically different propagules in highly connected lakes will be more frequent. Higher nucleotide diversity in more connected lakes may therefore be the result of a higher rate of (successful) immigration.

Rapid population differentiation

We observed striking genetic structure among the seven marine lake populations, indicating limitations to gene flow, even on small spatial (<200km) and temporal (~6,000 years) scales. Compared to other studies using RADseq in marine organisms, mussels from marine lakes show moderate to high pairwise F_{ST} values, and clear structure in admixure analyses (Lal et al., 2016; Tariel et al., 2016; Van Wyngaarden et al., 2016). There is not a complete barrier to the sea, as the vector of water that can carry propagules in and out of the lakes is ongoing. There may be other barriers to dispersal or successful establishment of immigrating genotypes, however. Such barriers would effectively block individuals from outside the lakes to interbreed with the standing population. In the subheading 'Isolation-by-distance, isolation-by-environment and priority effects' we discuss which of these modes of isolation may underlie the high genetic structure.

Since colonization of the lakes occurred a relatively short time ago (~6,000 years), any *in situ* differentiation can be considered rapid. Assuming that all marine lakes were colonized

from the same ancestral source pool in the sea, at least on the scale of <200km, genetic fixation indices of >0.07 are relatively high. Multiple studies of threespine stickleback, African rift lake cichlids, and organisms in anchialine ponds have found genetic and morphological diversification at even shorter time scales, as little as 150 years (Genner et al., 2007; Weese et al., 2012; Lucek et al., 2013; Marques et al., 2016; Weber et al., 2016; Gonzalez et al., 2017). Even though post-glacial derived mutations would imply rapid evolutionary rates, such rates are not uncommon (e.g. Genner et al., 2007; Ho et al., 2011). Furthermore, post-glacial divergence may also come from divergent selection on standing genetic mutation (Wang and Bradburd, 2014).

Emerging patterns for multiple marine taxa in population genetic studies suggest spatial genetic structure and limited gene flow at small spatial scales, despite a lack of clear physical barriers and high dispersal potential (Hoeksema, 2007; Barber, 2009; Carpenter et al., 2011; Waters et al., 2013; Starger et al., 2015). For example, genetic structure was found for stomatopods and caridean shrimp (range 200-300 km) (Barber et al., 2006; Haig et al., 2010), damselfish (range 25-150 km) (Timm and Kochzius, 2008), giant boring clams (range 25-50) (DeBoer et al., 2008), and starfish (range 10-15 km) (Crandall et al., 2008). A number of processes have been suggested underlying the unexplained marine biodiversity in the Coral Triangle, such as glacial cycles, heterogeneous environments, complex bathymetries and ocean current systems forming eddies to effectively trap larvae (Hoeksema, 2007). Glacial cycles and complex bathymetries may enhance patterns of isolation-by-distance or priority effects. Since sea levels were approximately 120m lower than modern sea levels at the Last Glacial Maximum, the reef setting of the Sahul shelf, including West Papua, would have consisted of many semi-isolated basins with reduced connection among each other (Sathiamurthy and Voris, 2006; Hoeksema, 2007). Hence, peripheral environments, such as modern-day marine lakes, have likely been present throughout history in the Coral Triangle area, potentially contributing to its high biodiversity. Additionally, heterogeneous environments may facilitate isolation-by-environment as populations may become locally adapted.

Elucidating the mechanisms of divergence is challenging as means to speciation are often obscure (Palumbi et al., 1994; Coyne and Orr, 2004), and typically a reliable historical component is lacking (Sanderson, 1997; Ho et al., 2011). The existence of multiple independently derived populations in landlocked marine lakes with varied environments provides an opportunity for fundamental research into the role of short term and incomplete isolation in population divergence.

Isolation-by-distance, isolation-by-environment and priority effects

Finally, we aimed to disentangle relative importance of three modes of isolation (Fig. 6). We assume that small populations colonized the marine lakes after their origin following the Last Glacial Maximum. The populations will likely have experienced founder effects (Mayr, 1954; Barton and Charlesworth, 1984). Genetic drift may be particularly strong in small, founding

populations, as alleles rare in the source population may stochastically be driven to fixation due to a subset of individuals colonizing new habitats. Therefore, founding populations can be expected to be genetically distinct from the ancestral populations and from conspecifics colonizing different marine lakes within several thousand generations. There has, however, been a continued connection between lakes and the sea and a potential for new migrants. Genetic signatures resulting from founder effects can therefore be expected to be overwhelmed by ongoing dispersal from the source population or from other marine lakes, or by forces of natural selection (Mayr, 1963; Waters et al., 2013).

On large spatial scales (>1400km) we find support for isolation-by-distance, formed likely by dispersal limitation. Due to the geographic distance, currents and obstructing landmasses between the lake in Berau and lakes in Papua, larvae are unlikely to be able to maintain genetic connectivity (Barber et al., 2011). With decreased dispersal between populations, there is subsequently decreased gene flow, and genetic drift may cause populations to become genetically distinct (Wright, 1943; Slatkin, 1985). Theoretical models have showed that larvae of marine organisms are able to traverse large distances, via passive dispersal

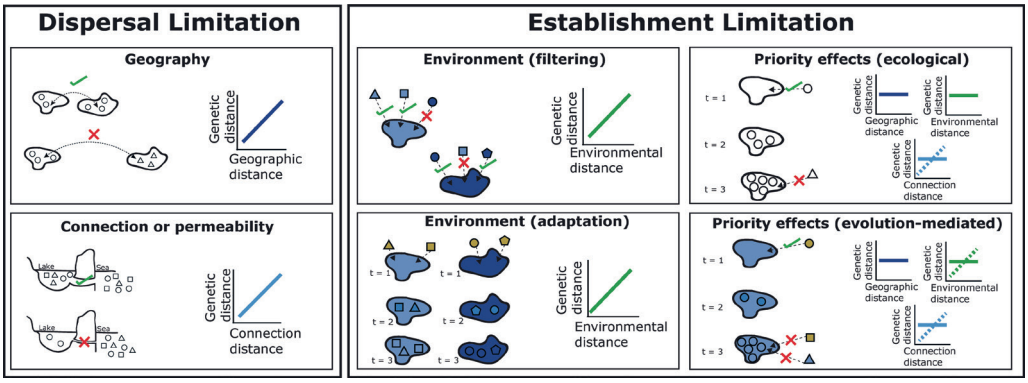


Figure 6: Modes of isolation and predictions for population genetic differentiation (F_{ST}). Dispersal limitation is caused by a reduction in homogenizing gene flow due to reduced dispersal success. Geographic distance can play a role when genetic distance between populations will increase with increasing geographic distance in a pattern of isolation-by-distance. Alternatively, permeability of habitats or landscape barriers can play a role. Connection distance represents the permeability of habitats to entrance of propagules, where higher connection distance means less permeable habitats. This results in a pattern of high genetic structure among populations inhabiting less permeable habitats, compared to populations from more permeable habitats. Alternatively, establishment limitation is caused by a reduction in homogenizing gene flow due to reduced establishment success of propagules within a habitat. It arises when there are differential environments, which either act as filters or which cause local adaptation of populations. Both result in an isolation-by-environment pattern, which entails populations in environmentally similar habitats will have ongoing gene flow, while populations in environmentally dissimilar habitats will not. Establishment limitation can also be caused by historical priority effects, which results in colonizing populations being able to outcompete any new immigrants via a numerical advantage, potentially aided by evolutionary local adaptation. Dotted lines in predictions for connection distance indicate the time lag necessary for priority effects to occur and be maintained is more likely in habitats that have low connection. Dotted lines in environmental distance indicate local adaptation may result in a pattern of isolation-by-environment.

via oceanic currents (Armsworth, 2002). Species of the genus *Brachidontes* are known to have long dispersive stages of up to four weeks (Monteiro-Ribas et al., 2006). However, these predictions are often overestimations of actual dispersal (Marshall et al., 2010), as shown particularly for the Coral Triangle (Treml et al., 2015). Furthermore, there is an issue of dilution: a high density of larvae is needed to maintain gene flow between habitats at large distances (Johannesson, 1988).

On smaller spatial scales (<200km), we find that geographic distance also plays a significant role, although it becomes less pronounced and not significant at the scale of 40km (Fig. 4). No signatures of isolation-by-environment were found on any spatial scale, thus we do not find evidence for isolation-by-environment based on our environmental data (temperature, salinity and pH) (Orsini et al., 2013; Wang and Bradburd, 2014). It is possible we have not accounted for certain environmental variables that may be important in population adaptation. Including parameters such as nutrient level or biotic interactions may increase resolution in finding signatures of local adaptation. On the scale of <200km we do find a positive relationship between genetic variation and degree of connection to the sea, indicating that the extent of connection to a source of propagules is important in influencing genetic structure among populations, which can also be thought of as an isolation-by-resistance pattern (McRae, 2006). We expect more isolated lakes will have a more pronounced delay in the arrival of new genotypes, which is corroborated by our finding of higher nucleotide diversity within populations inhabiting marine lakes with higher connections to the sea. Our observation of more isolated lakes being more genetically distinct may be consistent with the hypothesis that priority effects are stronger in more isolated environments, (Orsini et al., 2013; Fukami, 2015). Density-dependent ecological priority effects may further be mediated by evolution via adaptation of first colonizers to local conditions (De Meester et al., 2016). By having a head start in locally adapting to the new environment, early colonizers may become even more strong competitors to any future immigrants. If this were the case, we would expect that genotypes show a correlation to important environmental parameters. Our data corroborates this by showing a strong correlation to water temperature, particularly on the scale of 40km, which potentially indicates local adaptation may have occurred. Furthermore, priority effects predict extensive small-scale genetic differentiation (De Meester et al., 2016), which we observe even at the scale of 40km. Incomplete dispersal barriers facilitated by physical barriers may provide the ideal setting for priority effects to arise, and subsequently be fixed in the long-term via (evolution-mediated) adaptation.

The influence of priority effects on genetic structure we find is supported by the broadcast spawning life strategy of mussels, which allows them to reproduce rapidly, which is beneficial to the colonization of novel habitats. Mussels show a range of tolerance towards temperature and salinity (Spidle, 1995; Sarà et al., 2008; Dowd and Somero, 2012). Particularly species of the genus *Brachidontes*, such as *Brachidontes pharaonis* appear to be highly tolerant to low salinities and high temperatures (Sarà et al., 2008). The high plasticity would benefit

colonization of diverse environments. Since we find potential influences of temperature in determining genetic variation, we hypothesize that environment has a role in short term population differentiation. Initial ecological priority effects may be supported by subsequent local adaptation (De Meester et al., 2016). To further explore the extent of evolution-mediated priority effects in marine lake systems, we need to distinguish neutral and adaptive loci and look for signatures of local adaptation.

High differentiation in peripheral populations which are assumed to have the potential for ongoing gene flow is found in multiple marine taxa, and could support a role for priority effects (Dawson and Hamner, 2005; Goto et al., 2011; Gotoh et al., 2011; Swift et al., 2016; Pinheiro et al., 2017). These studies showed evidence for genetic isolation on similar time scales as our study for jellyfish ($\phi_{ST} > 0.74$) (Dawson and Hamner, 2005), fish (ϕ_{ST} ranging from 0.040 – 0.728) (Gotoh et al., 2011), and within the *Brachidontes* genus (p-distance = 0.146) (Goto et al., 2011). However, the three underlying modes of isolation tested were not explicitly tested as in the current study. Isolation-by-distance, isolation-by-environment and priority effects are not mutually exclusive, can be confounded, and may have different relative importance on different spatial and temporal scales. They form umbrella patterns for specific types of processes driving genetic structure, such as larval mobility (isolation-by-distance), selection against immigrants (isolation-by-environment), and competition (priority effects). Though we do not yet know the processes, we do find signatures of the overarching patterns. We hypothesize that density-dependent priority effects, potentially strengthened by local adaptation, may be more prevalent than previously assumed. In coastal reef systems, ongoing habitat fragmentation and climate change may accelerate processes of priority effects (Legrand et al., 2017), which may have important implications for conservational efforts. Our study supports an often-neglected role of eco-evolutionary dynamics in early stages of habitat colonization and subsequent population divergence.

Low-cost method for non-model organisms

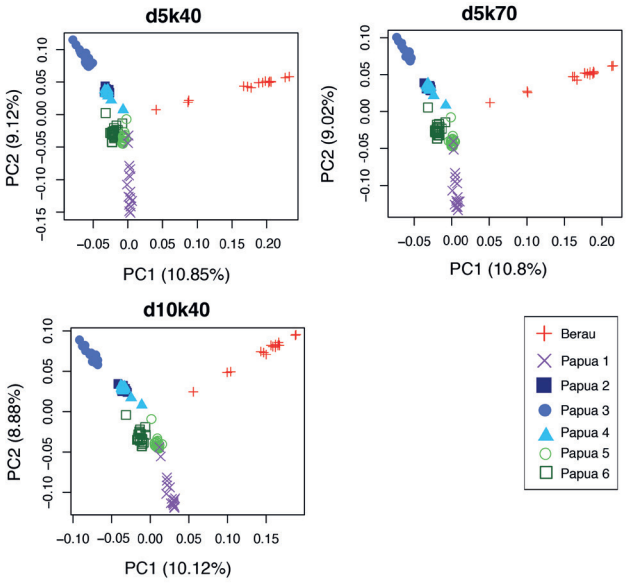
On a final note, we would like to stress the importance of developing low-cost Next-Generation Sequencing methods for non-model organisms. We used a low-cost method of population genomic library preparation, and an extensive step-by-step protocol is provided in the Supplementation Information to enhance the use in low-budget projects. We found major structures and conclusions to remain the same when trying different filtering options, showing that low coverage data can still provide accurate information. Hence, high numbers of individuals can be sequenced on one lane, reducing costs. This is especially relevant for laboratories with limited funding and working on non-model organisms, for example in the Coral Triangle. We hope that low-cost protocols such as the one used in the current study promotes and facilitates more extensive studies into marine biodiversity, which is still largely unknown and unchartered.

Acknowledgements

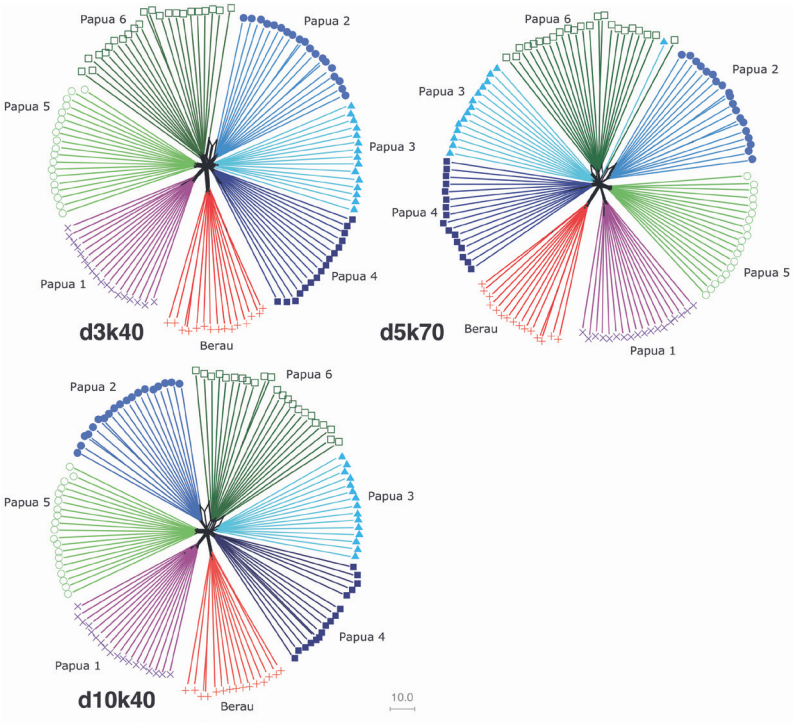
We would like to thank two anonymous reviewers and the subject-editor for valuable comments on the original manuscript. We would like to thank The Nature Conservancy and Conservation International for their logistical support during our sampling period in Indonesia. Furthermore, we would like to thank the following people for assistance with the library preparation in the laboratory and consultation about analyses: Cerise Chen, Emily Farrer, Maria Meijer, Milan Malinksy, George Roderick and Norah Saarman. The following people provided help with logistics for field- and/or labwork: RCO-LIPI, RISTEK, Bahruddin, Bert Hoeksema, Purwanto, A.S. Sidik, Suharsono, Ricardo Tapilatu, Yosephine Tuti, and the staff of Nabucco Island Dive Resort, of Derawan Dive Resort, and of Misool Eco Resort. For that, they have our sincere thanks.

This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation grants S10RR029668 and S10RR027303. The Netherlands Organisation for Scientific Research provided funding to Leontine E. Becking through the grants RUBICON #825.12.007, and VENI #863.14.020. Fieldwork in Indonesia was made possible through additional financial support of National Geographic Society Waitt Grant, De Beukelaar- van der Hucht Stichting, World Wildlife Fund, Netherlands-INNO Fund, the Schure-Beijerinck-Popping Fund of the Royal Dutch Academy of Science (KNAW), Conservation International Ecosystem Based Management program (funded by the David and Lucile Packard Foundation), the Treub-Maatschappij Fund, the Leiden University Fund (LUF)/Slingelands, Singapore Airlines, the A.M. Buitendijk Fund, and the J.J. ter Pelkwijk Fund (Naturalis Biodiveristy Center). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Permits were obtained from RISTEK (0094/FRP/SM/V/2009; 098/SIP/FRP/SM/V/2011).

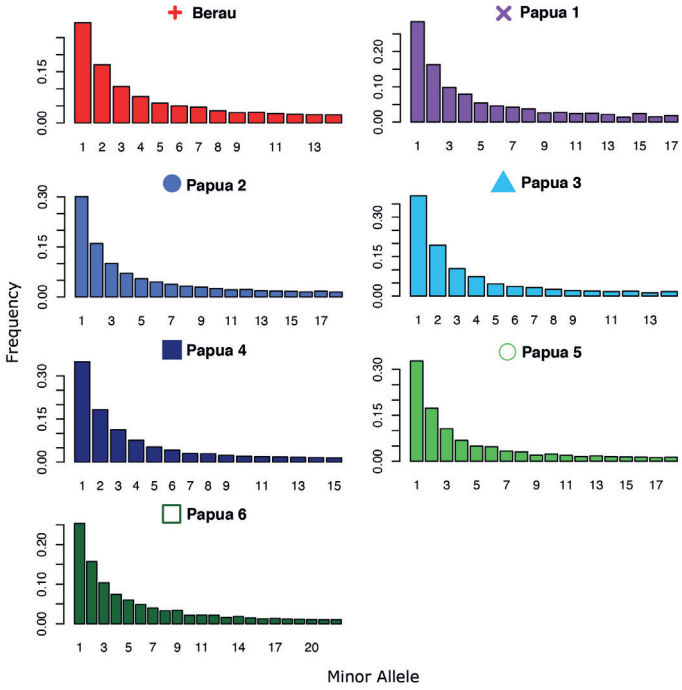
Supplemental Figures



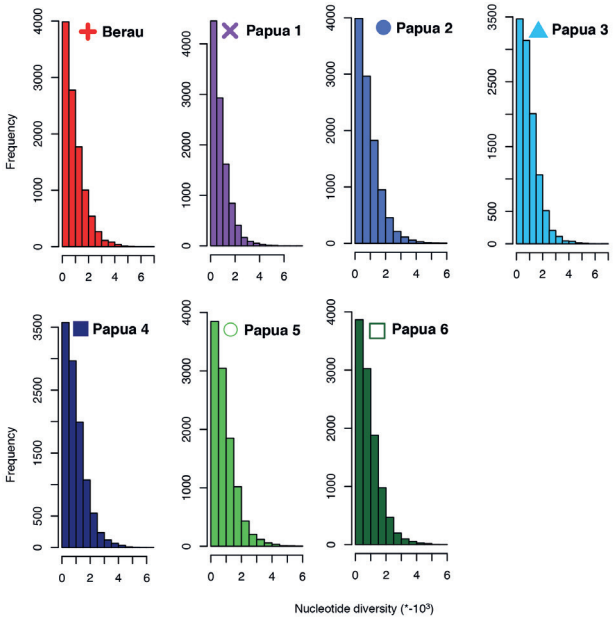
Supplemental Figure 1: Principal Component Analyses (PCA) of three alternative filtering options based on covariance matrices. Filtering options correspond to Table S1. Colours and codes correspond to Fig.1 and Table 1.



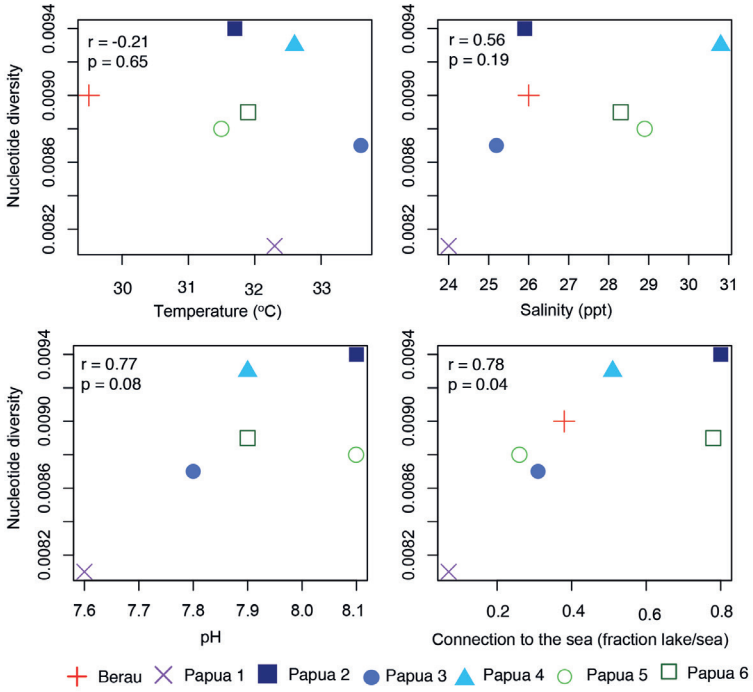
Supplemental Figure 2: Splitstree analysis for three alternative filtering options. Filtering options correspond with table S1. Colours and codes correspond to Fig. 1 and Table 1.



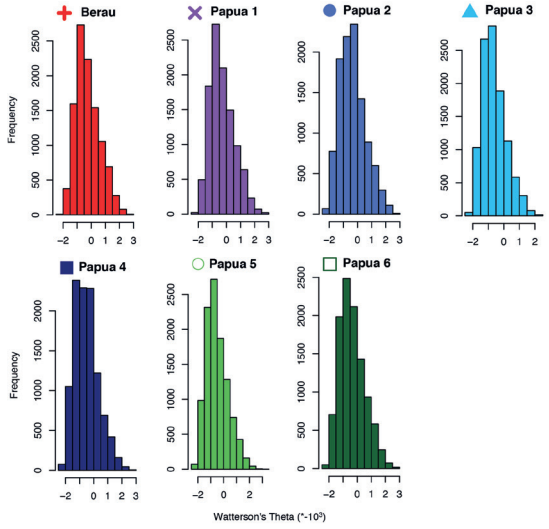
Supplemental Figure 3: Site Frequency Spectra (SFS). Frequency of shared minor alleles for individuals are shown. Colours and codes correspond to Fig. 1 and Table 1.



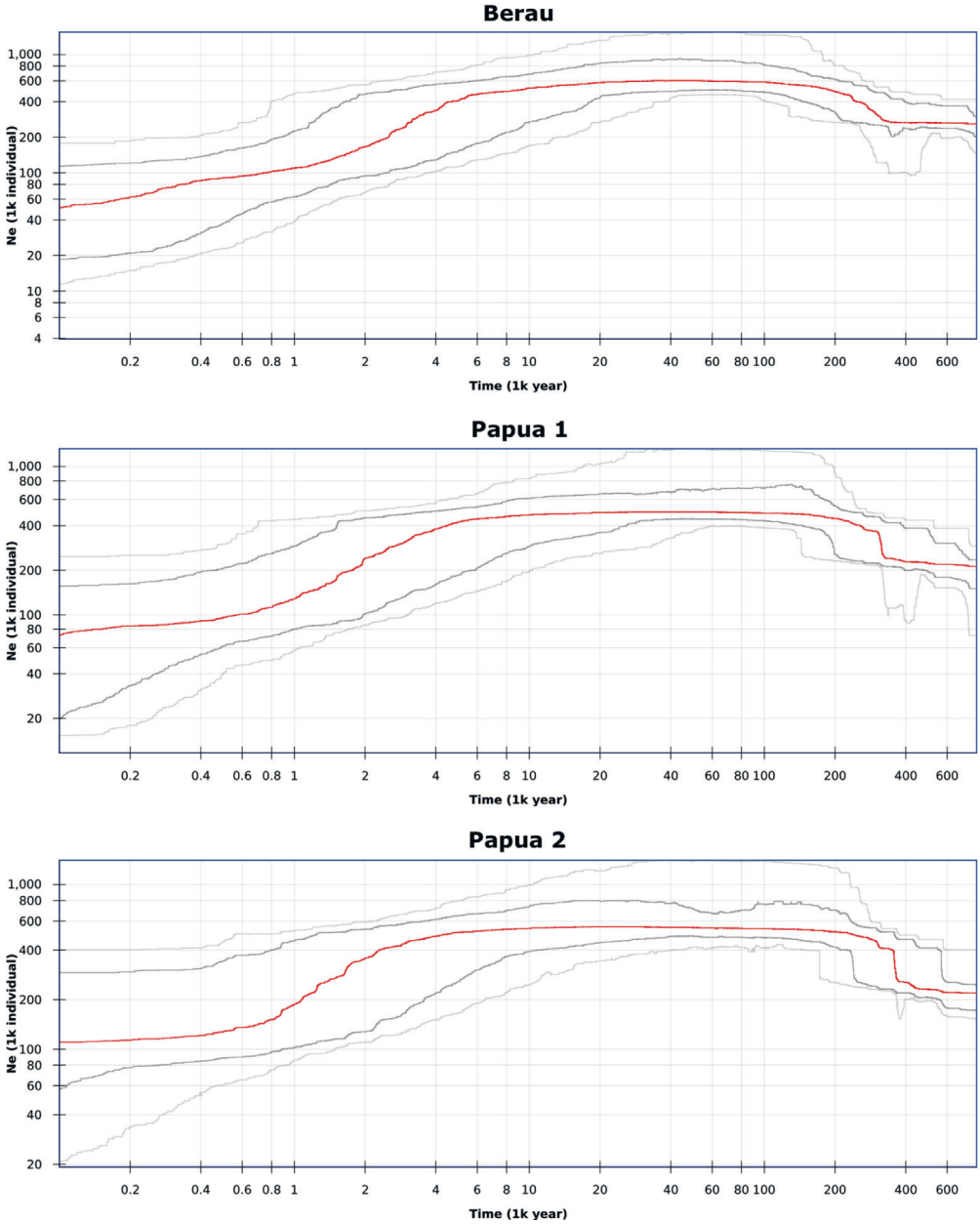
Supplemental Figure 4: Graphical representations nucleotide diversity (θ_π). Colours and codes correspond to Fig. 1 and Table 1.



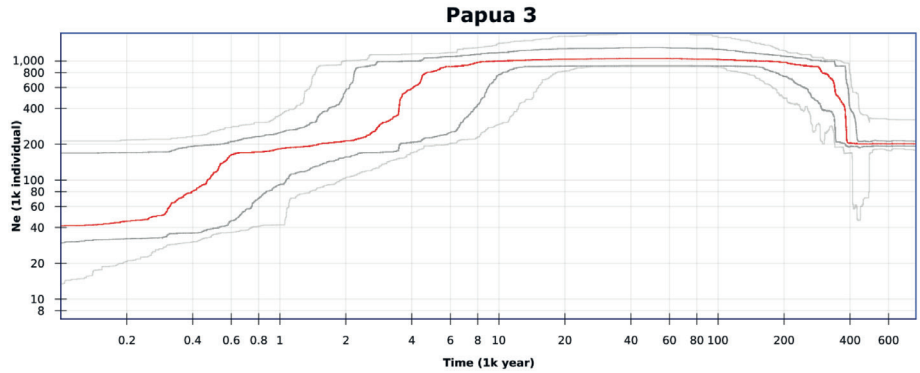
Supplemental Figure 5: Nucleotide diversity (θ_π) versus environmental variables (temperature, salinity and pH), and versus connectivity (fraction lake/sea). Pearson's correlation (r) and P-values are displayed. Colours and codes correspond to Fig. 1. and Table 1.

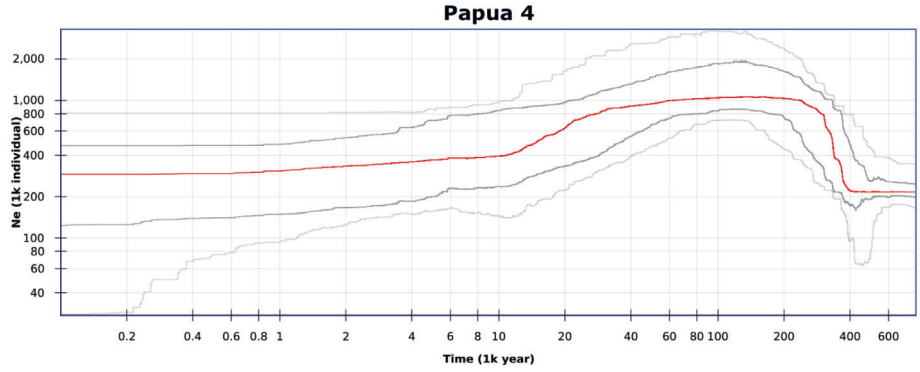


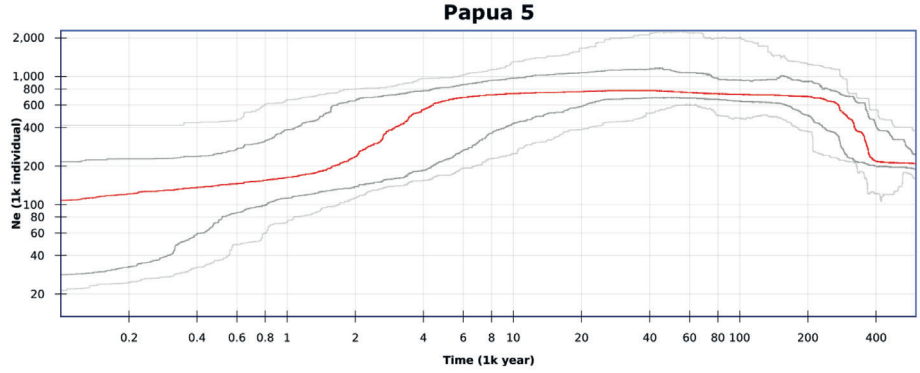
Supplemental Figure 6: Graphical representations of Tajima's D coefficients. Colours and codes correspond to Fig. 1 and Table 1.

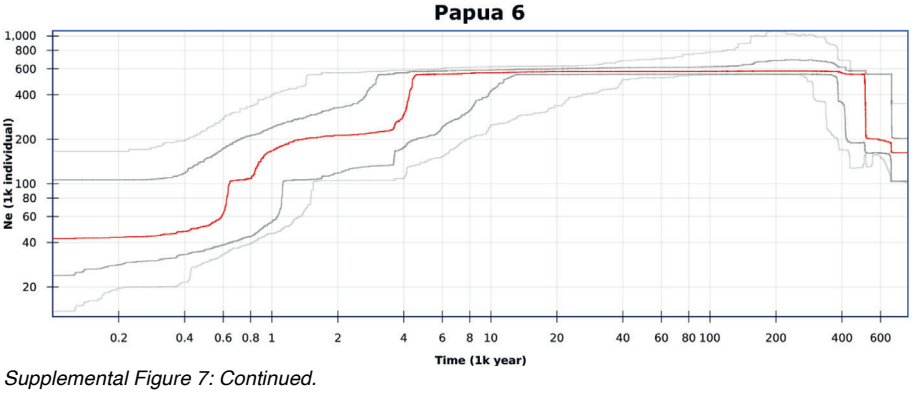


Supplemental Figure 7: Stairway plots modelling changes in effective population size (Ne) over time. Continued below.

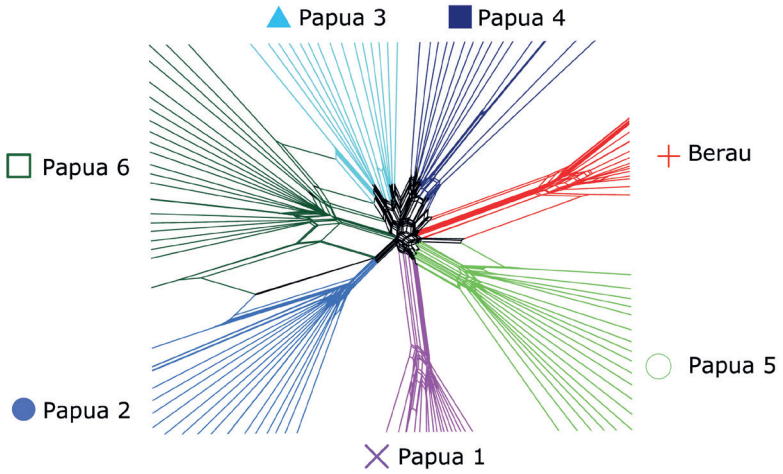




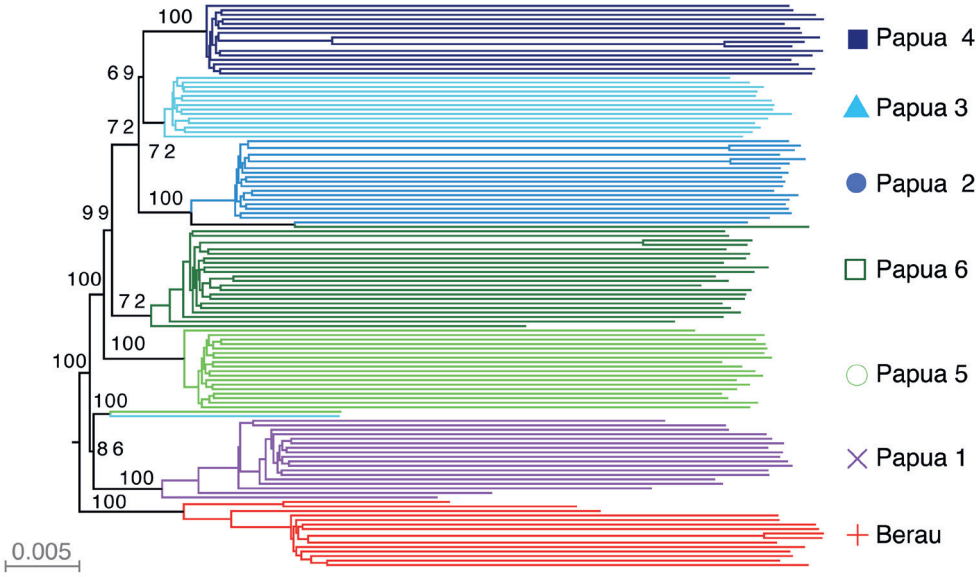




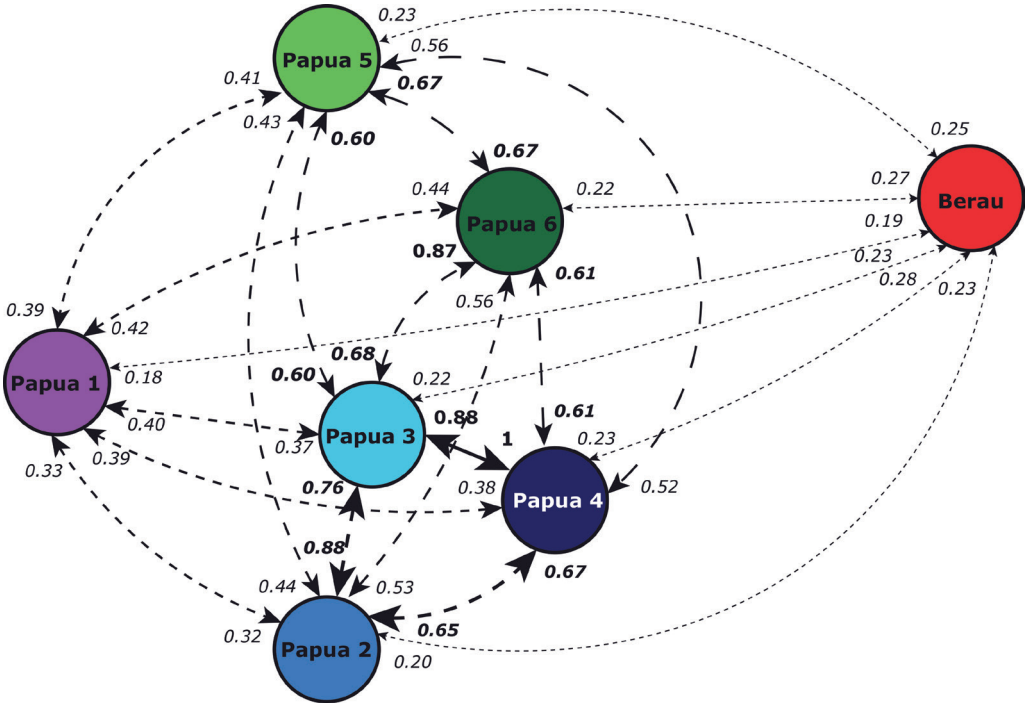
Supplemental Figure 7: Continued.



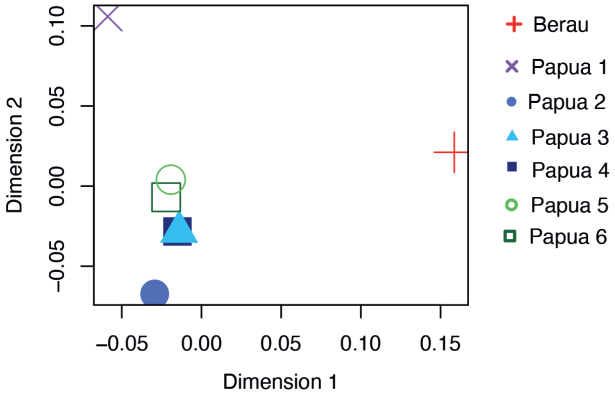
Supplemental Figure 8: Zoom of the Splitstree network in Fig. 2.



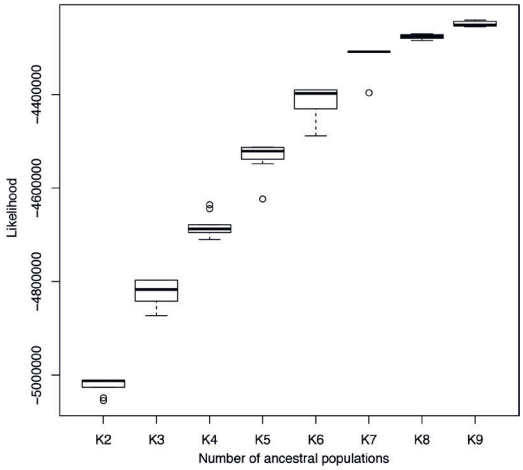
Supplemental Figure 9: Neighbour-Joining tree from on pairwise genetic distances. Bootstrap support values are displayed. Each branch represents one individual. Colours and codes correspond to Fig. 1 and Table 1.



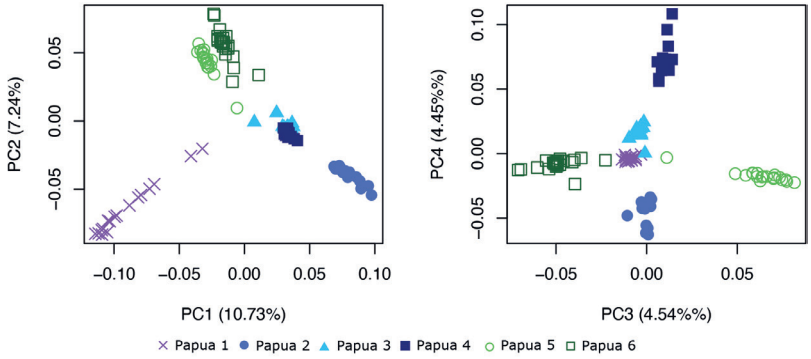
Supplemental Figure 10: Connectivity network. Fractions of relative connectivity are displayed. Colours and codes correspond to Fig. 1 and Table 1.



Supplemental Figure 11: Multi-Dimensional Scaling (MDS) plot of FST values from Table 3. Colours and codes correspond to Fig. 1 and Table 1.



Supplemental Figure 12: Boxplot of likelihood values obtained from admixture analysis. Values are generated by running each K-value for 5 independent times from ngsAdmix analyses.



Supplemental Figure 13: Principal Component Analysis (PCA) based on pairwise genetic covariance among 110 individuals with 116,415 SNPs, from six locations in Raja Ampat, Indonesia. First four axes represent 26.96% of total genetic variation. Each dot is one individual.

Supplemental Tables

Supplemental Table 1: Sequences blasted to check for contamination. All sequences were obtained from Genbank.

Organism	Strain	Accession number	Method
Neptuniibacter caesariensis **Gammaproteobacterium**	strain MED92 1099521380378	gil89083557lgblAAOW01000001.1	Whole genome shotgun sequence
Vibrio splendidus **Gammaproteobacterium**	strain 12B01 1099451319047	gil84379704lgblAAMR01000001.1	Whole genome shotgun sequence
Escherichia coli **Gammaproteobacterium**	strain K-12 substr. MG1655	gil556503834lreflNC_000913.3	Complete genome
Claviceps purpurea **Sordariomycetes**	20.1 WGS project CAGA00000000 data	gil399172385lemblCAGA01000001.1 contig scaffold00001,	Whole genome shot gun sequence
Rhizobiales bacterium **Alphaproteobacterium**	JGI 001012-O08 C600DRAFT_contig_0.1_C	gil554773868lgblAUSA02000001.1	Whole genome shotgun sequence
Homo sapiens	Chromosome 2	gil568336022lgblCM000664.2l	GRCh38 reference primary assembly

Supplemental Table 2: Total number of loci and number of selected SNPs for four different filtering options. Here, 'd' indicates the minimum read depth, 'k' indicates minimum percentage of individuals included that have at least a read depth of 'd'. Option in bold is used in all downstream analyses.

Filtering	Total loci	SNPs
Options: d=3, k=70	**761,014**	**116,415**
Options: d=5, k=40	127,903	120,320
Options: d=5, k=70	84,650	80,086
Options: d=10, k=40	68506	64,957

Supplemental Table 3: Pairwise F_{ST} indices for three filtering options.

u5k40	Berau	Papua 1	Papua 2	Papua 3	Papua 4	Papua 5
Papua 1	0.250	-				
Papua 2	0.227	0.189	-			
Papua 3	0.195	0.156	0.088	-		
Papua 4	0.202	0.163	0.110	0.072	-	
Papua 5	0.213	0.162	0.145	0.111	0.121	-
Papua 6	0.212	0.159	0.134	0.094	0.112	0.108

u5k70	Berau	Papua 1	Papua 2	Papua 3	Papua 4	Papua 5
Papua 1	0.212	-				
Papua 2	0.193	0.166	-			
Papua 3	0.164	0.138	0.078	-		
Papua 4	0.173	0.144	0.099	0.066	-	
Papua 5	0.179	0.141	0.132	0.101	0.111	-
Papua 6	0.178	0.139	0.119	0.084	0.102	0.097

u10k40	Berau	Papua 1	Papua 2	Papua 3	Papua 4	Papua 5
Papua 1	0.223	-				
Papua 2	0.203	0.173	-			
Papua 3	0.172	0.142	0.082	-		
Papua 4	0.181	0.150	0.104	0.069	-	
Papua 5	0.188	0.146	0.137	0.104	0.114	-
Papua 6	0.188	0.146	0.125	0.089	0.106	0.102

Supplemental Table 4: Watterson's theta (θ_W), and inbreeding coefficient (F). Between brackets are standard deviations.

Location	θ_W	F
Berau	0.0100 (0.0085)	0.009 (0.020)
Papua 1	0.0091 (0.0078)	0.008 (0.013)
Papua 2	0.0103 (0.0086)	0.017 (0.022)
Papua 3	0.0124 (0.0092)	0.027 (0.019)
Papua 4	0.0121 (0.0092)	0.041 (0.032)
Papua 5	0.0113 (0.0088)	0.028 (0.024)
Papua 6	0.0104 (0.0082)	0.003 (0.009)

Supplemental methods, results and information

Can be downloaded from https://onlinelibrary.wiley.com/doi/full/10.1111/mec.14556.

Chapter 6

Genetic divergence does not follow environmental acclimatization in marine lake mussels

Diede L. Maas, Stefan Prost, Emiliano Trucchi, Awaludinnoer Ahmad, Ludi P. Aji, Martine Bèrubè, Rosemary G. Gillespie, Christiaan A. de Leeuw, Purwanto, Detmer Sipkema, Ricardo F. Tapilatu, Hamid H.A. Toha, Leontine E. Becking

Submitted

Abstract

Relative importance of mechanisms of speciation on small spatial scales in the marine realm is not fully understood. For highly dispersive marine organisms, mechanisms that cause differentiation despite ongoing gene flow such as ecological speciation are expected to play an important role in generating current genetic diversity. Here, we used the replicated system of marine lakes, island-like bodies of seawater surrounded by land, to study the relative importance of ecology-driven versus neutral parapatric population differentiation in the sea. We sampled *Brachidontes* sp. mussels in Raja Ampat, Indonesia from 9 coastal locations and 18 marine lakes along a gradient of connection to the surrounding sea and with varying local environments. We used restriction site-associated DNA sequencing to quantify population genetic variation and complimented this by measuring phenotypic variation as a proxy for direct responses to the local environment (shell morphology, associated microbial communities and trophic niche space). Compelling evidence for ecological speciation would consist of concordance between genetic and phenotypic divergence. Regardless of using 10,294 neutral Single Nucleotide Polymorphisms (SNPs) or including 681 SNPs under putative selection, we saw clear clustering per lake. We observed strong genetic differentiation between marine lakes (average F_{ST} = 0.11), yet panmixia among coastal ocean locations. For the lakes, geographic location and degree of connection to the surrounding sea were found to be significant predictors of genetic variation. Most outlier SNPs were found in contrasts with isolated lakes, potentially indicating local adaptation. Yet we found no association between genetic and phenotypic divergence. Therefore, we found no evidence for ecological speciation driving population differentiation. Instead, neutral dispersal limitations such as geographic distance and (incomplete) physical dispersal barriers appear to play an important role in structuring parapatric populations. However, while population divergence may have ultimately been formed by stochastic processes, eco-evolutionary dynamics such as priority effects may have contributed to the current diversity found.

Keywords: population genomics, ddRADseq, marine biodiversity, genetic divergence, phenotypic divergence, ecological speciation

Introduction

The relative importance of different mechanisms of speciation in marine systems particular in species rich areas is not well understood (Puebla, 2009). The role of ecological speciation in promoting the formation of reproductive isolation has been receiving renewed interest as an alternative mechanism to chance events such as genetic drift and founder events (Coyne and Orr, 2004; Rundle and Nosil, 2005; Faria et al., 2014). However, the study of ecological speciation in the marine realm has been limited so far. This is surprising since due to the presumed incomplete barriers to dispersal in the marine realm (Palumbi, 1994; Ward, 1994; Bowen et al., 2013), processes that mediate speciation in the presence of gene flow are expected to be more important (Puebla, 2009). Ecological speciation, or the arising of reproductive barriers due to divergent selection between ecologically different habitats (Rundle and Nosil, 2005; Schluter, 2009), is therefore thought to play an important role in causing microevolutionary processes of marine population differentiation. It predicts that populations adapting in parallel to similar environments should genetically also be similar, in a pattern of isolation-by-environment (Schluter, 2009; Wang and Bradburd, 2014). Ecological speciation is particularly likely to cause patterns of differentiation if it is accompanied by (replicated) phenotypic diversification. Examples of rapid divergence presumably due to ecological speciation have been reported in European flounders in the Baltic Sea (Momigliano et al., 2017), and in repeated colonization by marine stickleback of freshwater lakes (Jones et al., 2012). On the other hand, neutral processes of divergence can include processes of founder events or historical contingency (Mayr, 1942; Fukami, 2015). In founder events, a subset of the source population splits off and by chance consists of distinct allele frequencies. This effect can be exacerbated through genetic drift in the generally smaller effective size of the population that has split off (Frankham et al., 2002). Effects of historical contingency, or arrival order, can be effectively fixed through numerical priority effects (where new genetic variants or species are outcompeted by resident populations), potentially aided by rapid local adaptation (De Meester et al., 2016). The mechanisms of speciation can be difficult to distinguish because they can generate similar patterns, are not mutually exclusive, or differ on different spatial and/or temporal scales (Rundle and Nosil, 2005; Schluter, 2009). Yet recent advances in genomic tools are allowing further exploration of adaptive and neutral marine microevolutionary processes.

Genomic tools that allow for the sequencing of high resolution genetic marker panels are becoming more affordable (Catchen et al., 2017; Luikart et al., 2018). The increased resolution allows marine scientists to dive into fine-scale spatial and environmental factors underlying population differentiation, and to distinguish between neutral markers and markers under putative selection (Riginos et al., 2016; Selkoe et al., 2016; Liggins et al., 2020). While generally high dispersal potential of marine organisms is being confirmed by neutral markers, signatures of population divergence are being detected in a subset of outlier SNPs (Single Nucleotide Polymorphisms) (Bierne et al., 2016; Van Wyngaarden et al., 2016; Bernatchez et al., 2018; Xuereb et al., 2018). An isolation-by-distance pattern was found only for the subset of outlier SNPs in the sea scallop despite high dispersal potential,

indicating restricted connectivity potentially owing to local adaptation (Van Wyngaarden et al., 2016). For the giant California sea cucumber, the subset of outlier SNPs fitted better to an isolation-by-resistance pattern mediated by asymmetric currents (Xuereb et al., 2018). Genetic structure of putatively adaptive genetic variation of the eastern oyster was found to be best explained by environmental variables or temperature and salinity (Bernatchez et al., 2018). The studies indicate marine population structure on smaller spatial scales than assumed previously assumed based on larval duration periods, with possible mechanisms of barriers formed by currents or adaptation to local environments. However, an issue generally complicating population genomic work in the marine realm is that it is often difficult to establish the boundaries of populations *a priori*. Furthermore, environmental factors are often spatially and/or temporally variable, clouding ultimate causes of divergence (Legendre, 1993; Lee and Mitchell-Olds, 2011).

Islands, or island-like systems, have classically provided clearly defined model systems to test evolutionary and ecological theories (Warren et al., 2015; Patiño et al., 2017; Whittaker et al., 2017), and marine lakes are proposed to be marine equivalents (Dawson et al., 2009; Itescu, 2018). Marine lakes are defined as bodies of seawater entirely surrounded by land, but which are connected to the surrounding sea via caves or pores in karstic rock (Holthuis, 1973; Dawson et al., 2009). Natural basins in karstic were filled up when seawater rose after the Last Glacial Maximum (~20,000 BCE), giving rise to the approximately 200 marine lakes known to science today (Tomascik and Mah, 1994; Dawson et al., 2009). The extent of connection of a marine lake to the surrounding sea can be observed through the presence of tidal fluctuations (Dawson et al., 2009; Dawson, 2016). Beyond determining the degree of isolation of a marine lake, the extent of in- and outflux of seawater also determines the amount of mixing which consequently may influence local environment of the lakes. The Indo-Pacific has several hotspots of marine lakes, notably in Palau (Dawson et al., 2001; Dawson and Hamner, 2005) and Indonesia (Becking et al., 2011, 2015). Raja Ampat in Indonesia in particular has a high number of marine lakes, which can be seen as independent replicates of evolution. Raja Ampat lies in the center of the Coral Triangle, an area known for its peaks in marine species richness (Hoeksema, 2007; Mangubhai et al., 2012). The high biodiversity of the area and the replicated nature of marine lakes gives an ideal setting to study spatial and environmental processes that may underlie population differentiation.

The independent replicates of marine lakes provide a unique opportunity to estimate genetic and phenotypic population differentiation in marine lake mussels in West-Papua (~300km radius). We use this setting to test for relative importance of ecological speciation and neutral mechanisms of speciation such as founder events. Mussels of the genus *Brachidontes* sp. (Swainson 1840) (Mollusca; Bivalvia; Mytilidae) are ubiquitously present in marine lakes of Raja Ampat where they can form large mussel beds. We determined genetic divergence via a double digest restriction-site associated sequencing approach (ddRADseq), which allows for the identification of thousands of anonymous SNPs (Peterson et al., 2012; Franchini et

al., 2017). To quantify phenotypic divergence, which represents acclimatization to local environments, we measured mussel shell morphology, sequenced mussel-associated microbes, and determined stable isotopes. For shell morphology, we quantified shell shape as an indicator of local biotic (competition and predation (Reimer and Tedegren, 1996; Lauzon-Guay et al., 2005)) and abiotic (water quality) factors, and shell thickness as calcification is expected to be affected by water temperature and pH (Gazeau et al., 2007; Zhao et al., 2020). The microbiome of mussels was determined via 16S ribosomal RNA gene sequencing and identification of Operational Taxonomic Units (OTUs) to add another dimension to reflect the local environments and selection by the mussel. And finally, stable isotopes represent the trophic niche space of the mussels, which might shift in different local environments as food sources change (Katzenberg, 2000).

We quantified genetic and phenotypic divergence of mussel populations in 18 marine lakes and included 9 coastal locations as outgroups. The marine lakes represent a range in connection to the sea and in local environments, from isolated lakes with distinct environments to highly connected lagoon-like lakes. We examined population structure to verify if marine lakes can be considered independent replicates or if they showed signatures of ongoing gene flow. If ecological speciation underlies population differentiation, we expect to find signatures of panmixia in neutral markers, but putative outlier markers to follow an isolation-by-environment pattern. The highest number of outlier SNPs should then be present between contrasts of ocean populations and isolated lakes with the most distinct environments. If non-ecological mechanisms such as founder events are more important drivers of structure, we expect lower genetic diversity in the insular systems compared to open ocean populations along with a population bottleneck. We then also expect population differentiation to show no difference between neutral and putatively selective markers, and to only show a correlation to connection to the sea, as bottlenecks of population size are expected to be more stringent in more isolated lakes. To further distinguish ecological and neutral mechanisms, we tested for associations between genetic and phenotypic datasets. If genetic and phenotypic divergence match, it would be compelling evidence of local adaptation. In contrast, if we find no match between the two, other mechanisms of speciation might play a more important role in differentiation, and phenotypic differentiation reflect plasticity. With our setting, we expect we will be able to make predictions on the effects of degree of isolation and local environments on population differentiation and connectivity. Accurate knowledge on population diversity connectivity among marine populations on small spatial scales and the factors acting thereupon is required to design effective Marine Protected Areas for population resilience and persistence in the face of global change (Allendorf et al., 2010; Carr et al., 2017).

Material and Methods

Sample collection and marine lake setting

The *Brachidontes* sp. samples were collected from 27 locations in West-Papua Indonesia (Fig. 1A), in the islands of Wayag, Gam and Misool in Raja Ampat (Fig. 1B) as well as along

the island of Biak (Fig. 1C). Sampling included 9 coastal locations and 18 marine lakes (Fig. 1D). *Brachidontes* sp. mussels (Fig. 1E) were collected for sequencing by hand-picking and snorkeling from 2016 to 2018 (Table 1). Within each lake, 30 mussels were sampled around the perimeter of the lake at about 1m below the lowest tide point. Mussels were taken at at least 5m distance from each other. Where possible, large individuals (>2cm) were collected to ensure maturity. In the field, the adductor muscle was immediately extracted and preserved in custom made RNAlater or 95% ethanol. The adductor muscle was taken to avoid microbial contamination as much as possible. Remaining tissue and shells were preserved in 95% ethanol. Upon returning to the lab, extracted muscles were preserved in -80°C until DNA extraction, and remaining tissue and shells in -20°C or 5°C.

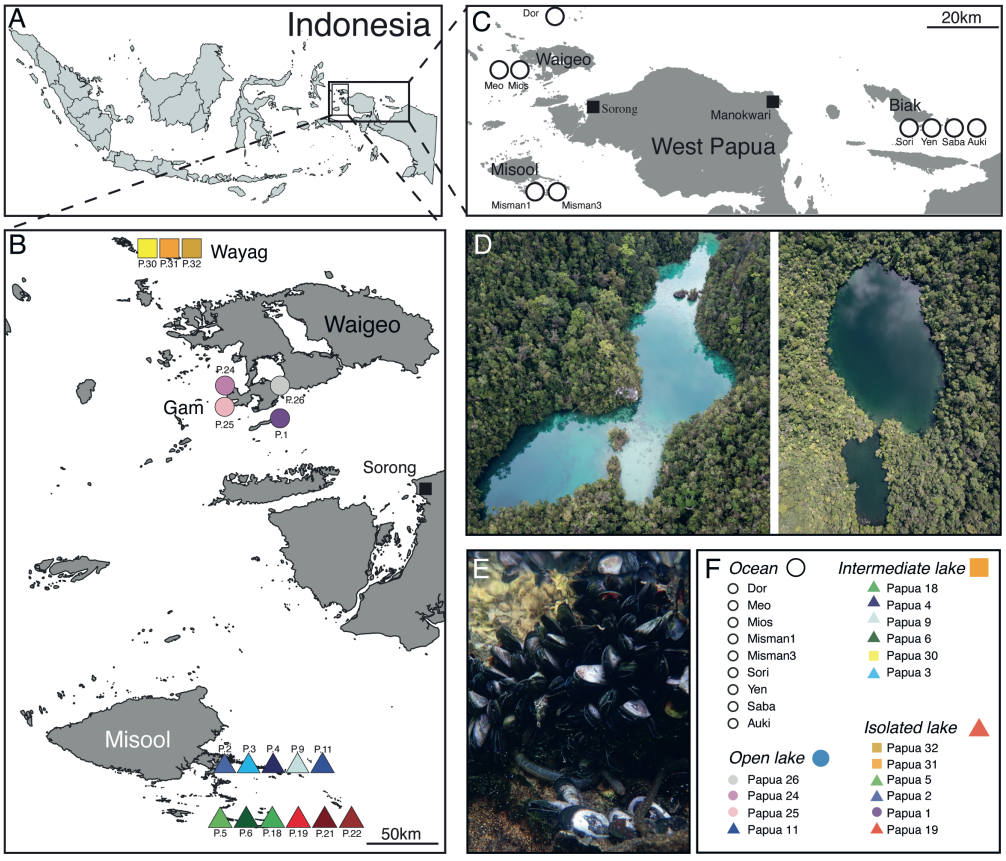


Figure 1: Map of sampling locations of Brachidontes sp. mussels in West-Papua, Indonesia. Coastal sea locations are indicated by a white circle, marine lakes with colored shapes depending on amount of connection with the sea. (A) Overview of Indonesia. (B) Raja Ampat, West-Papua, including 18 marine lakes. (C) Broader Bird's Head Seascape, West-Papua, including 5 coastal locations in Raja Ampat and 4 coastal locations around the island of Biak. (D) Two examples of marine lakes (Left: open lake Papua 11, right: isolated lake Papua 2). (E) Brachidontes sp. mussels. (F) Categorization of sampling localities into four groups: Coastal ocean locations, open lakes, intermediate lakes and isolated lakes. Location codes and fraction measurements correspond with Table 1.

Table 1: Samples genotyped total (all lineages), samples genotyped subset (only lineage A and without outlying group), genetic lineage (Goto et al., 2011, Appendix Table 1), physical lake characteristics (area, depth and connection calculated as a fraction of maximum tidal amplitude of the lake over the sea), water quality characteristics (temperature, salinity and pH), category of locations based on connection, nucleotide diversity (π) and expected heterozygosity (H_e) of Brachidontes sp. samples from 27 locations in West-Papua, Indonesia (18 marine lakes, 9 coastal sea locations). Locations ordered by connection to the surrounding sea.

Code	#Samples total	#Samples subset	Genetic lineage	Lake area (m^2)	Lake depth (m)	Con. (Lake/ Sea)	Temp. (°C)	Salinity (ppt)	pH	Category	Nucleotide diversity (π)	Expected heterozygosity (H_e)
Dor	10	10	A			1	30.1	30.8	8.10	Ocean	0.064	0.060
Meo	10	10	A			1	30.1	30.8	8.10	Ocean	0.064	0.060
Mios	10	10	A			1	30.1	30.8	8.10	Ocean	0.065	0.061
Misman1	3	0	B			1	30.4	30.1	8.30	Ocean		
Misman3	5	0	B			1	30.4	30.1	8.30	Ocean		
Sori	5	5	A			1	30.1	30.8	8.10	Ocean	0.063	0.055
Yen	6	6	A			1	30.1	30.8	8.10	Ocean	0.065	0.058
Saba	5	5	A			1	30.1	30.8	8.10	Ocean	0.063	0.055
Auki	6	6	A			1	30.1	30.8	8.10	Ocean	0.064	0.058
Papua 26	14	0	B	16700	4.5	0.86	30.4	29.9	8.00	Open lake		
Papua 24	8	8	A	4200	6	0.85	30.3	30.7	7.87	Open lake	0.061	0.056
Papua 25	18	16	A	21500	8.3	0.85	31.7	29.6	7.79	Open lake	0.061	0.059
Papua 11	17	15	A	27300	8.9	0.83	30.7	28.3	8.06	Open lake	0.060	0.057
Papua 4	21	19	A	13750	20.4	0.8	31.7	25.9	8.10	Intermediate lake	0.058	0.056
Papua 9	12	12	A	33600		0.8	31.7	25.5	7.80	Intermediate lake	0.056	0.052
Papua 6	20	20	A	2950	12.4	0.78	31.9	28.3	7.91	Intermediate lake	0.055	0.054
Papua 18	21	9	A	7000	4.6	0.78	31.5	28.4	7.76	Intermediate lake	0.056	0.052
Papua 30	22	22	A	13000	4.1	0.75	32.4	28.9	7.64	Intermediate lake	0.061	0.059
Papua 3	21	20	A	20800	7.5	0.51	32.6	30.8	7.79	Intermediate lake	0.057	0.055
Papua 32	20	18	A	6100	5.5	0.45	31.2	30.7	7.33	Isolated lake	0.052	0.050
Papua 31	9	7	A	6200	7.1	0.3	30.1	29.2	7.17	Isolated lake	0.050	0.045
Papua 5	17	8	A	3700	4.8	0.26	31.5	28.9	8.11	Isolated lake	0.054	0.049
Papua 2	21	19	A	12200	7.3	0.23	33.6	25.2	7.81	Isolated lake	0.053	0.051
Papua 21	21	0	A	18950	13	0.11	35.9	23.9	7.91	Isolated lake		
Papua 1	19	18	A	88530	19	0.07	32.3	24.0	7.58	Isolated lake	0.05	0.049
Papua 22	22	0	A	23100	12.3	0.06	35.6	16.3	8.08	Isolated lake		
Papua 19	19	8	A	61050	31.5	0.05	35.4	19.6	8.21	Isolated lake	0.055	0.051

Environmental and physical profiling of the sampling locations was performed consistent with Maas et al. (2018). In brief, location coordinates logged via Gpsmap 64S (Garmin) were used to calculate pairwise straight-line distances using the R package *geosphere*. Water quality parameters temperature (°C), salinity (ppt) and pH measurements were taken in at least 10 sites per lake from the lake surface to 5m depth by using a YSI Professional Plus Multimeter (Table 1). Lake depth was measured via a handheld sonar system (H22PX) at approximately 100 points per lake. Google Earth was used to calculate approximate lake area (in m^2). Degree of connection was calculated from tidal amplitudes from the sea and the directly adjacent sea. Using Hobo water-level loggers, the maximum tidal amplitude (in m) was obtained, and the degree of connection was defined as the fraction of the maximum tidal amplitude of the lake compared to the sea. The combination of degree of connection of the sea as well as the local water quality parameters determined whether lakes were defined as 'open', 'intermediate' or 'isolated' (Fig. 1F). Open lakes had high connection to the surrounding sea (0.85±0.01), low temperature (30.8°C±0.6) and high salinity (29.6 ppt±1.0). Isolated lakes had low connection to the surrounding sea (0.19±0.1), generally high temperature (33.2°C±2.2) and low salinity (24.7 ppt±4.9). Intermediate lakes showed intermediate values for connection (0.74±0.1), temperature (32.0°C±0.4) and salinity (28.0 ppt±2.0).

DNA extraction, library preparation and sequencing

Genomic DNA was extracted using the DNeasy Blood & Tissue Kit from Qiagen (Quiagen, Germantown, MD, USA). Quantity and quality were checked on a QuBit DNA assay from Thermo Fisher and using 1% agarose gel electrophoresis. First, mitochondrial DNA was extracted via Polymerase Chain Reaction (PCR) amplifying the COI fragment using jgLCO/HCO primers designed by (Geller et al., 2013). The protocol as described by (de Leeuw et al., 2020) was performed to amplify COI fragments. A maximum likelihood tree was built, to ensure all samples were of Lineage A (Goto et al., 2011; de Leeuw et al., 2020) (Appendix Fig. 1A, B). Mussels from ocean locations Misman1 and Misman3 and marine lake Papua 26 were found to consist of Lineage B. Lake Papua 26 was included in downstream analyses as an outgroup where appropriate.

Next, a modified RADseq protocol based on Peterson et al. (2012) and Franchini et al. (2017) was followed to prepare genomic libraries (Supporting Information A). In brief, we employed two restriction enzymes (*SphI* and *MluCI*) to double-digest genomic fragments. To allow for the detection of PCR duplicates, four to six random bases were added to the adapters. The distribution of resulting fragments was investigated with a BioAnalyzer High Sensitivity chip (Agilent). We selected fragments of 300-500bp in size using a Pippin Prep (Sage Science). Genomic libraries were pooled and paired-end 2x150bp sequenced on the Illumina HiSeq2500, HiSeq4000 and NovaSeq platforms at GATC Biotech and LUMC (Leiden University Medical Center).

Bioinformatics and genotype calling

After sequencing, quality of reads was evaluated using FastQC v.0.11.9 (Andrews, 2010). We then used STACKS v.2.2 and v.2.4 to detect Single Nucleotide Polymorphisms (SNPs) using the *de novo* pipeline as no reference genome was available (Catchen et al., 2013; Rochette and Catchen, 2017). PCR duplicates were filtered out using the *filter_clones* module. Next, the module *process_radtags* was used to demultiplex samples, remove inline barcodes, and apply quality filters. Resulting reads were truncated to a common length of 130bp via Trimmomatic v.0.39 (Bolger et al., 2014). Then, loci were assembled in stacks via *ustacks* with a minimum depth of coverage set at 3 (m) and a maximum of 6 nucleotide mismatches (M). A catalog of all loci was assembled via *cstacks* where a distance of 4 was allowed between loci (n). The *populations* module was used to retain loci present in at least 70% of individuals and retaining only one random SNP per locus. Individuals with more than 30% missing data were excluded from the analyses.

Identifying neutral and putatively outlier loci

We used BayeScan (Foll and Gaggiotti, 2008) and PCAdapt (Luu et al., 2017) to assemble a set of candidate loci under selection. Since different methods may detect different putatively outlying loci we took two different approaches (Narum and Hess, 2011). BayeScan v2.1 is a Bayesian method and was run with default settings of 5,000 iterations, a burn-in 50,000 steps of and a fall discovery rate (FDR) value of 0.05. PCAdapt is a Principal Component Analysis method that uses correlations of genetic variation to Principal Components to detect outliers. We used a screeplot to determine the number of Principal Components (K). We used three methods to detect outlying SNPs with an alpha value of 0.01 as recommended by the authors, but results from the q-value, Benjamin-Hochberg and Bonferroni procedures were similar. Population genomic patterns may vary between using a neutral marker panel versus an outlier SNP panel (Luikart et al., 2003), therefore we established three datasets: (i) Including both neutral and putative outlier SNPs, (ii) including only neutral SNPs, so filtering out all indicated SNPs under selection, and (iii) including only putatively outlier SNPs that were identified by both methods.

Within population genetic diversity and demography

Within-population genetic diversity indices nucleotide diversity (π) and observed and expected heterozygosity (H_o and H_e, respectively) were obtained via the *populations* module in STACKS (Catchen et al., 2013), and the latter two verified in GenoDive (Meirmans and Van Tienderen, 2004). We compared categories (Ocean, Open lake, Intermediate lake and Isolated lake) with a Kruskal-Wallis Chi-squared test in R (v.3.6.3 (R Core Team, 2018)) as data were not normally distributed or showed homogeneity of variance. Pairwise Wilcoxon tests were run to detect which comparisons showed significant differences from each other. Finally, we used Spearman correlations to determine correlations between genetic diversity (nucleotide diversity and expected heterozygosity) and environmental parameters (temperature, salinity, pH, connection, lake area and depth). Since multiple tests were run, we adjusted the alpha-value to 0.008 to indicate significance.

To detect changes in effective population size (N_e) over time Stairway plots were built (Liu and Fu, 2015). We set the generation time at 1 year (Morton, 1988), and calculated an expected mutation rate of 1.3×10^{-9} per site per generation using the regression of Lynch, (2010). Other mutation rates were tested but did not alter the observed pattern.

Population structure and genetic differentiation

Population structure and differentiation was explored in several different ways. We explored the phylogenetic relationships of 382 *Brachidontes* sp. specimens from all locations sampled via a genetic distance matrix calculated from genotype probabilities in ngsDist (Vieira et al., 2016). A neighbor-joining tree was computed via RAxML (Stamatakis, 2014), and converted to a phylogenetic tree via FastME (Lefort et al., 2015), visualized in FigTree v.1.4.2 (Rambaut, 2009). Then, we performed Principal Component Anayses (PCA) based on scores provided by GenoDive (Meirmans and Van Tienderen, 2004). When including samples from Papua 21 and Papua 22, and some individuals from Papua 19, Papua 5 and Papua 6, we noticed a distinct clustering of these samples away from the rest (Supplemental Fig. 2). As this may indicate another distinct genetic lineage, we took these samples out and continued all downstream analyses with 271 individuals. Furthermore, we built a co-ancestry matrix using haplotype data (i.e., including all SNPs present in each locus) via FineRADstructure with standard settings (Malinsky et al., 2018). We generated a dataset with the full haplotype per RADtag using the RADPainter flag in the populations module of STACKS. After checking the distribution of SNPs and alleles per locus running the Stacks2fineRAd.py script in the FineRADstructure package, we subsampled the dataset to include loci with a maximum of 70 SNPs. The extremely high polymorphism has been found to be characteristic for marine bivalves (Hedgecock et al., 2005; Zhang et al., 2012; Harrang et al., 2013). To verify whether observed patterns were not driven by the extreme variability or missing data we ran the analysis again to test the robustness of the inferred co-ancestry matrix. We excluded individuals with more than 50% missing data and with more than 30 SNPs, but overall patterns remained the same. Next, we calculated pairwise genetic differentiation (F_{ST}) among populations in GenoDive with 1,000 permutations. Grouping F_{ST} per category, we tested whether F_{ST} was significantly different between comparisons of categories via Kruskal-Wallis chi-squared test and Wilcoxon post-hocs in R. Significance was assigned when the p-value was below 0.005. Finally, we ran an AMOVA (hierarchical analysis of molecular variance) (Excoffier et al., 1992) in GenoDive with 1,000 permutations to test for structuring within populations and within categories.

Next, we investigated links of genetic variation to environmental parameters. We ran a redundancy analysis (RDA) on only marine lake populations, taking out coastal ocean locations. We used genetic variation as the first four Principal Components from the PCA as response variables. We computed a distance-based Moran's eigenvector maps (dbMEM) (Dray et al., 2006) via decomposing geographic (Euclidean) distances in meters between sites. The first axis explained 95.8% of the variation in geographic distance and was chosen as an explanatory variable. Temperature (°C) and connection to the sea were also chosen

as explanatory variables to explore the effect of local environment and degree of isolation in structuring genetic variation, respectively. We ran an Anova to determine whether the model was significant and used the function *envfit* as implemented in vegan to assess explained variation of explanatory variables to the genetic structure with 1,000 permutations (Oksanen et al., 2016). To compare these results, we also ran Mantel tests on normalized genetic distance ($F_{ST}/(1-F_{ST})$) and geographic, environmental and connection distance matrices (Mantel, 1967; Slatkin, 1993). Environmental distance (using temperature, salinity and pH) and connection distance were calculated as outlined in Maas et al., (2018).

Potentially outlier loci were visually assessed using pairwise genetic differentiation (F_{ST}) per locus between all population comparisons. We used Manhattan plots to visualize outlier loci. Grouping population comparisons per category, we also calculated number of unique and shared outlying loci ($F_{ST}>0.8$) among the groups.

Phenotypic acclimatization

We quantified phenotypic acclimatization in terms of shell morphology, associated microbe communities and trophic niche space. Morphological diversity was defined as shell shape and shell thickness variation. Relative warp analysis on based on semi-landmarks was performed to assess shell shape (Bookstein, 1989; Rohlf, 1995). The approach of semi-landmarks was chosen because in mussel shell outlines there are no true landmarks except for the beak (umbo) (Bookstein, 1996). Shell variation from 428 mussels originating from 22 locations was analyzed. Using a Nikon D100 SLR camera, shells were photographed in a standardized way including a scale. Images were processed using the TPS suites TpsUtil (Rohlf, 2010a), TpsDig (Rohlf, 2010b), and TpsRelw (Rohlf, 2010c). Along each shell outline, 69 semi-landmarks were placed with equal spacing. The Tps softwares calculated centroid size and relative warps (latter are comparable with Principal Components). We used an adonis test implemented in vegan in R to test for significant differences between categories and used Redundancy Analysis to test for explained variation by geographic location, connection and temperature.

For a subset of 75 mussels, we derived shell thickness from cross-section images made by a SkyScan 1072 micro-CT scanner. Approximately 1250 cross-sectional images were produced per specimen, from which shell area was calculated by a custom MatLab script. The area of each shell was corrected for shell size. We used boxplots to display variation in shell area and tested for differences among categories using an Anova. Spearman correlations were run between shell thickness and explanatory variables.

Microbial communities were determined via DNA extractions from gill tissue of 96 mussel specimens from 20 locations. From 6 locations microbial communities were also extracted from water filters to get an indication of ambient microbial communities. We amplified the 16S rRNA gene using barcoded primers 515F and 806R (Walters et al., 2015). In brief, the FastDNA Spin Kit for soil (MP Biomedicals) was used to extract DNA, eluting in a final

volume of 80-100μL. Then, PCR reactions were performed in 36.5μL nuclease free water, 10μL 5x HF-buffer, 1μL dNTPs, 0.5μL Phusion Hot Start II DNA polymerase, 1.5μL DMSO and 1%, 1μL primers and 1μL DNA template. PCR cycles were set at 30s at 98°C for activation, 30 cycles of denaturation for 10s at 98°C, annealing for 10s at 50°C, elongation for 10s at 72°C and finally extension of 7min at 72°C. Final libraries were sequenced on an Illumina HiSeq2000 at Eurofins Genomics. In total, 3,376 Operational Taxonomic Units (OTUs) were identified. We used a non-metric Multidimensional Scaling plot to visualize variation among locations and categories. We used an *adonis* test to test for significant differences between categories and used *envfit* to determine the influence of explanatory variables.

Then, trophic level via stable isotope analysis (Katzenberg, 2000) was determined for 90 mussel specimens from 17 locations. By making use of stable isotopes $\delta^{13}C$ (carbon) and $\delta^{15}N$ (nitrogen), the technique allows for the tracking of carbon sources and trophic positions in the food web (Fry, 2006). Mussel tissue was dried at 60°C and grinded into a homogenized powder using liquid nitrogen. Weight was standardized and stable isotopes were determined using a mass spectrometer (Thermo Scientific Flash2000 Organic Elemental Analyser) at NIOZ (Royal Netherlands Institute for Sea Research). A biplot was made displaying the $\delta^{13}C$ and δ^{15} signatures per individual. We made separate boxplots for carbon and nitrogen per location to test significant differences between categories (Anova) and ran Spearman correlations to test effects of explanatory variables.

Lastly, we contrasted genetic divergence with phenotypic divergence. We compared the first two Principal Components of the genetic variation as described above to morphological, microbial community and trophic level diversity. Comparisons were made visually and tested by Mantel tests to find significant correlations between distance matrices. We also performed a hierarchical clustering analysis using *hclust* in R using the UPGMA (unweighted pair group method with arithmetic mean) method using genetic, morphological, microbial community and isotope distance matrices. Next, we tested the alignment of each comparison to the genetic dendrogram and measured the entanglement score via the package dendextend in R (Galili, 2015). The entanglement score ranges from 0 (perfect alignment) to 1 (total misalignment).

Results

Read statistics and SNP dataset subsetting

We obtained 2,816,573,029 raw reads from Illumina and NovaSeq platforms after correcting for PCR duplicates. The median number of reads per sample was 8,656,830, and we discarded individuals with less than 2 million reads. After filtering and selection for one SNP (Single Nucleotide Polymorphism) per locus we retained different numbers of SNPs depending on the genetic lineages included. When including populations from both genetic Lineage A and B (Supplemental Fig. 1, Goto et al., 2011), we retained 8,573 SNPs (382 mussels total). When including only Lineage A (360 individuals), we retained 10,464 SNPs,

but an outlying cluster was observed in a neighbor-joining tree, potentially representing a different lineage (Supplemental Fig. 2 and 3). Excluding the individuals from this cluster (Papua 21 and 22 entirely, and some individuals from Papua 5, 18 and 19) we performed subsequent population genetic analyses in the subset of 271 individuals and 10,836 SNPs. From this last dataset, BayeScan identified 139 SNPs to be under putative selection, and PCAdapt identified 542 SNPs. Of the two methods, 43 SNPs overlapped. So finally, for the neutral dataset we filtered out SNPs under putative selection to retain 10,294 SNPs for population genetic analyses.

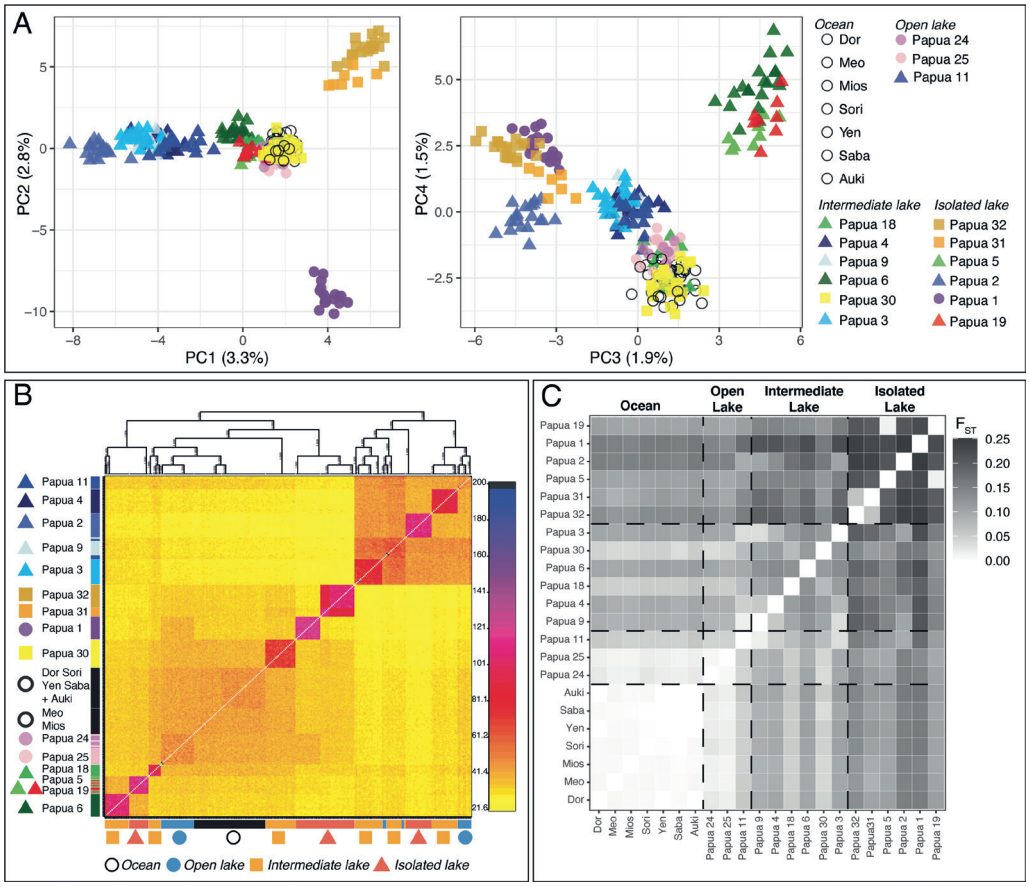


Figure 2: Population genetic structure of 271 Brachidontes sp. specimens based on 10,294 neutral SNPs. (A) Principal Component Analysis (PCA). First four axes represent 9.5% of the total genetic variation. Principal Component 1 vs 2 and 3 vs 4 are displayed. Each symbol represents an individual and codes correspond to Table 1. Locations are colored individually, lake shapes represent major geographic areas (Wayag = square, Gam = circle, Misool = triangle). Refer to Appendix for PCAs based on all individuals sampled, both neutral and outlier SNPs, only outlier SNPs and only representing ocean locations (Appendix Figure 2). (B) Co-ancestry matrix showing numbers of SNPs shared among pairwise comparisons of individuals. Individual populations are indicated as vertical axis, categories are indicated on horizontal axis. (C) Heatmap of pairwise genetic differentiation between populations (F_{ST}). Dashed lines separate categories.

Population genetic diversity and demography

Genetic diversity measured by nucleotide diversity (π) and expected heterozygosity (H_e) showed consistent patterns (Table 1). Averaging at 0.058, nucleotide diversity was highest in coastal ocean locations (0.063-0.065), and lowest in isolated lakes Papua 1 and Papua 31 (0.050). Expected heterozygosity on average was 0.055, and also was highest in coastal ocean locations (0.055-0.060), and lowest in isolated lake Papua 31 (0.045). When categorizing populations in Ocean, Open lake, Intermediate lake and Isolated lake groups, significant differences in genetic diversity between groups could be observed (nucleotide diversity: Kruskal-Wallis chi-squared test (1,3) = 19.08, p = 0.0002, expected heterozygosity: Kruskal-Wallis chi-squared test (1,3) = 15,4, p = 0.001) (Supplemental Fig. 4A). Lowest diversity was seen in the group of isolated lakes. When testing individual populations for correlations to explanatory variables, only a significant correlation was observed for both nucleotide diversity and expected heterozygosity with degree of connection (Spearman correlations, rho = 0.79, p = 0.0003 and rho = 0.77, p = 0.0004, respectively) (Supplemental Table 1, Supplemental Fig. 4B).

Stairway plots representing changes in effective population size (N_e) did not show different patterns among the four categories (Supplemental Fig. 5), nor for any individual population. Consistently, a pattern of increasing population sizes up until approximately 10,000-20,000 years before present could be seen, after which effective population size dropped for all categories. Effective population sizes for intermediate and isolated lake populations were estimated to be lower than for coastal ocean and open lake locations.

Population structure and differentiation

The neighbor-joining tree of 383 *Brachidontes* sp. specimens showed consistent results to what was expected from the single marker COI maximum likelihood and the RADtag neighbor-joining trees (Supplemental Fig. 1 and 2). Populations sampled from coastal locations Misman1 and Misman3, as well as marine lake Papua 26 clustered together as an outgroup from Lineage B. Remarkably, lakes previously presumed to contain Lineage A (Papua 21 and 22) were found to be genetically highly distinct from the other lakes (Supplemental Fig. 2, 3). Along with some individuals from Papua 5, 6, 18 and 19 these specimens may represent yet another genetic lineage splitting off from Lineage A. Therefore, in subsequent analyses they were taken out.

The Principal Component Analysis (PCA) based on the subset of individuals from Lineage A showed clear clustering per lake (Fig. 2A). The first four axes of the PCA explained 9.5% of total genetic variation. Principal Component 1 and 2 together separated Papua 1 and Papua 31/32 from the other populations. Coastal ocean locations centered in the middle of the graph and overlapped with Papua 24 and 25 from Gam island and Papua 30 from Wayag. Principal Component 3 and 4 further separated Papua 2, and a cluster of lakes in South Misool: Papua 5, 6 and 19. The patterns seen in multiple axes of the PCA could also be retrieved from the co-ancestry matrix (Fig. 2B). All individuals from marine lakes clustered

in distinct groups according to the lake of origin, with the exception of the open lake Papua 11 (with some individuals clustering together with Papua 9) the open lakes Papua 24 and 25, and isolated lakes Papua 5 and 19. All individuals from coastal ocean popualtions cluster in two separate groups, one including the localities from Biak together with Dor, and the other including only populations from Waigeo. Individuals from coastal ocean populations and open lakes show lower level of co-ancestry among each other as compared to individuals from intermediate and, in particular, isolated lakes. At higher hierarchical level, lakes are grouped into three clear geographical clusters: North Misool, South Misool (with notable exception of Papua 18), and Gam/Wayag. In the Gam/Wayag group, open lakes Papua 24 and 25 are the most similar to the ocean populations. Papua 1 does retain higher non-treelike co-ancestry with geographically close Papua 24 and 25 populations. Papua 18 clusters together with the ocean and Wayag/Gam group instead of the geographically closer South Misool lakes.

All pairwise comparisons of genetic differentiation (F_{ST}) except for some comparisons among coastal ocean populations were significant (Fig. 2C, Supplemental Table 2). Significant estimates of F_{ST} ranged from 0.007 (ocean populations Meo vs. Auki) to 0.251 (isolated lakes Papua 1 vs. Papua 2). Comparisons among categories showed significant differences per group (Kruskal-Wallis: Chi-squared test (1,9) = 173, $p < 0.0001$), with comparisons with isolated lakes having significantly higher genetic differentiation values compared to other groups (Supplemental Fig. 6). Finally, the analysis of molecular variance (AMOVA) showed that 10.2% of genetic variation could significantly be explained by the assigned categories ($p = 0.001$), while 32.5% could be explained by variation between populations ($p = 0.001$) (Table 2).

Linking genetic variation to explanatory variables

The redundancy analysis (RDA) showed that genetic variation was significantly influenced by explanatory variables (Anova: F(3,11) = 2.47, $p = 0.009$) (Fig. 3A). Here, geographic location ($R^2 = 0.77$, $p = 0.001$) and connection to the sea ($R^2 = 0.60$, $p = 0.006$) were significant predictors, but not temperature ($R^2 = 0.31$, $p = 0.010$). The results from the RDA were related to the more generally used Mantel tests. In order of strength of correlations, we found significant isolation-by-distance (Mantel $r = 0.30$, $p = 0.01$), isolation-by-environment (Mantel $r = 0.43$, $p = 0.01$) and isolation-by-resistance, as determined by the connection to the sea, effects (Mantel $r = 0.73$, $p < 0.001$). Environmental distance and connection distance matrices were highly correlated (Mantel $r = 0.54$, $p = 0.004$), and environmental and geographic distance matrices as well (Mantel $r = 0.28$, $p = 0.02$), making absolute effects difficult to disentangle.

Table 2: Analysis of Molecular Variance (AMOVA) among Brachidontes sp. populations.

Source of variation	Percentage variation	F-value	Standard deviation	p-value
Between categories	10.2%	0.104	0.002	0.001
Between populations	32.5%	0.364	0.004	0.001
Among individuals	56.7%	0.433	0.004	-

Finding effects of degree of isolation on genetic variation and differentiation, we assessed pairwise F_{ST} values per locus for pairwise comparisons of categories using both neutral SNPs and those under putative selection (Fig. 3B). Contrasting open lakes with each other showed a signature of relatively low overall F_{ST} without outliers above 0.6. Instead, comparisons with isolated lakes, be it with ocean locations, open lake locations or other isolated lakes, showed high numbers of outlier SNPs. This observation was also reflected when we counted SNPs above an F_{ST} value of 0.8 when running pairwise comparisons between and among categories (Table 3). Here, most outlier SNPs were observed between ocean vs. isolated lake (83 SNPs) and isolated lake vs. isolated lake (77 SNPs) comparisons, with about half the SNPs for open lake vs. isolated lake comparisons (41 SNPs).

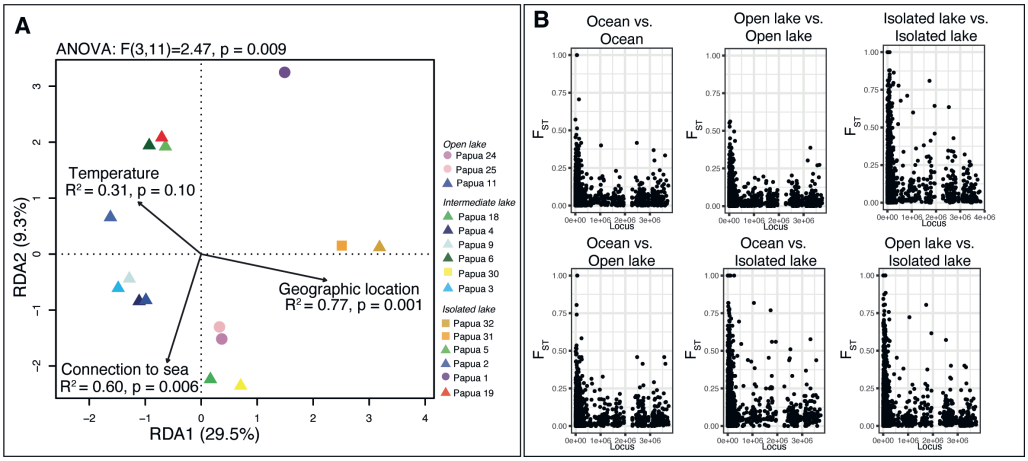


Figure 3: Relationships of genetic variation to explanatory variables. (A) Redundancy Analysis (RDA) based on first four Principal Component Analyses axes from Figure 2, with geographic location, temperature and connection to the sea as explanatory variables. ANOVA and environmental fitting results are displayed. (B) Manhattan plots of category comparisons based on F_{ST} value per locus.

Table 3: Number of outlier Single Nucleotide Polymorphisms (F_{ST}>0.8) as seen from Figure 3B calculated per pairwise comparison of populations in categories Ocean, Open lake, and Isolated lake. Unique indicates SNPs unique to a category comparison and shared indicates SNPs that were shared between category comparisons. For example, in Ocean versus Ocean locations, 2 SNPs were unique, 6 were shared with Ocean vs. Open lake and Ocean vs. Isolated lake, 2 were shared with Ocean vs. Isolated lake only, and 1 was shared with Ocean vs. Open lake only (9 shared total).

Category	Unique	Shared	Total
Ocean vs. Ocean	2	9	11
Ocean vs. Open lake	4	8	12
Ocean vs. Isolated lake	25	58	83
Open lake vs. Open lake	0	0	0
Open lake vs. Isolated lake	3	38	41
Isolated lake vs. Isolated lake	22	55	77

Genetic versus phenotypic divergence

Finally, we tested whether genetic divergence showed associations to phenotypic acclimatization (morphological, microbial communities and trophic level) (Fig. 4 and 5). Morphological variation in terms of shell outlines showed significant differences between lake categories (adonis: $F(3,368) = 28.84$, $p < 0.001$), with all categories being significantly different from each other ($p < 0.001$) (Fig. 4B, see Supplemental Fig. 7 for individual populations). However, we did not observe an association between genetic diversity and shell outlines (Mantel $r = 0.07$, $p = 0.25$). Shell thickness (measured as total area of shell) did not vary significantly among categories (Anova: $F(2,11) = 0.73$, $p = 0.51$), nor was there a correlation to genetic variation (Mantel $r = 0.10$, $p = 0.26$). Microbial communities also showed significant differences between categories (adonis: $F(3,83) = 6.28$, $p < 0.001$), with all groups being significantly different from each other (Fig. 4C, see Supplemental Fig. 8 for individual populations). We again found no correlation between genetic diversity and microbial communities (Mantel $r = 0.0009$, $p = 0.489$). Finally, we observed a significant difference among categories for $\delta^{13}C$ (Anova: $F(3,15) = 16.72$, $p < 0.001$), but not for $\delta^{15}N$ (Anova: $F(3,15) = 1.208$, $p = 0.34$) (Fig. 4D, see Supplement Fig. 9 for individual populations). Samples from coastal ocean locations showed higher $\delta^{13}C$ than the lake locations. The trophic levels of mussels from different populations were not significantly related to genetic variation (Mantel $r = 0.12$, $p = 0.23$). Links between genetic variation and other diversity measurements was also visualized in comparisons of dendrograms, where no clear relationships could be observed (Fig. 5). Based on the entanglement score the comparison between genetic and trophic level diversity measures showed the best fit (0.207).

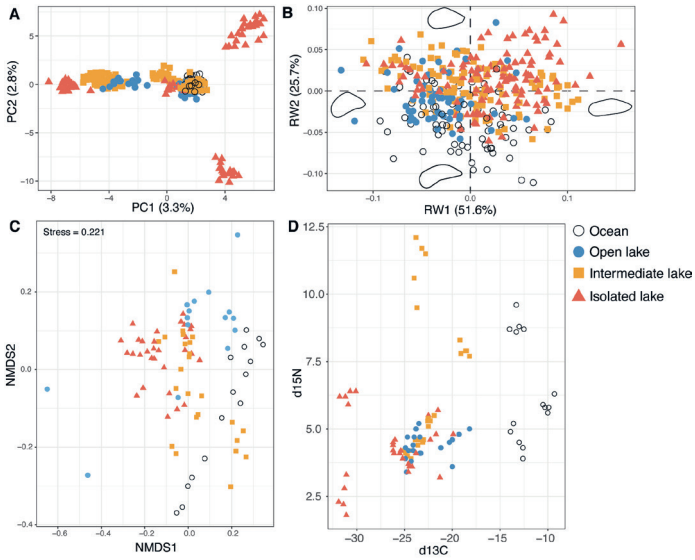


Figure 4: Comparison of structure of different measures of diversity. (A) Genetic diversity represented by first two Principal Component Axes. (B) Morphological diversity represented by geometric morphometrics on shell outlines. First two relative warp axes are displayed. Shell outlines of axis extremes are displayed. (C) Microbial community diversity as depicted in a non-metric multidimensional scaling plot. (D) Trophic level diversity as displayed as a biplot of $\delta^{13}C$ on the x-axis and $\delta^{15}N$ on the y-axis. Each dot represents an individual, colors and shapes correspond categories as presented in Figure 1.

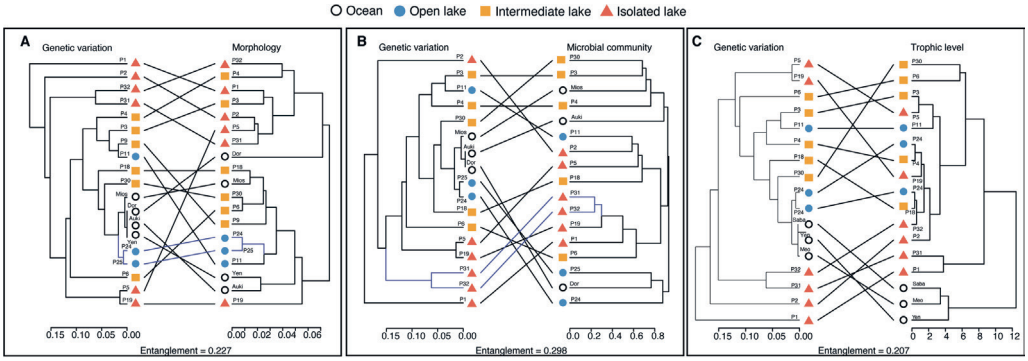


Figure 5: Comparisons of dendrograms of genetic distance versus (A) morphological, (B) microbial community, and (C) isotope distance measures. Entanglement scores are indicated.

Relationships between the other diversity measures and explanatory variables consisting of geographic location, water quality and connection to the sea were also tested. An RDA on shell outlines followed up by envfit showed temperature ($R^2 = 0.45$, $p = 0.04$) and connection to the sea ($R^2 = 0.44$, $p = 0.03$) to be significant predictors, but not geographic location ($R^2 = 0.05$, $p = 0.75$). We observed no correlations between shell thickness or connection to the sea (Spearman: rho = 0.34, p = 0.23) (Supplementary Fig. 10), nor to temperature (rho = -0.41, p = 0.15) or pH (rho = -0.49, p = 0.08). For microbial communities the envfit analysis on the nMDS showed no significant influence of temperature, connection or geographic location (all p-values > 0.18). Lastly, for trophic level diversity only $\delta^{13}C$ showed a significant relationship to connection to the sea (Spearman: rho = 0.41, $p < 0.001$) (Supplemental Fig. 11).

Discussion

Understanding the factors underlying population genetic diversity and connectivity among marine organisms will allow us to better ensure population persistence in times of global change. In this study, we aimed to elucidate relative importance of ecological and neutral processes of mussel population differentiation by using the replicated island-like setting of marine lakes in Raja Ampat, Indonesia. Regardless of including neutral or putatively adaptive genetic markers, we observed clear clustering per lake, contrasted by panmixia found in oceanic populations. Geographic location and connection to the surrounding sea were found to be an important predictors of population structure. Comparing patterns of genetic versus phenotypic divergence, we saw no concordance. Hence, we did not find clear evidence of ecological speciation underlying population differentiation in marine lake populations. Other driving mechanisms of speciation could have a higher importance, such as founder events potentially fixated by eco-evolutionary dynamics of priority effects and monopolization. Below we discuss population genomics patterns in insular marine systems, compare genetic and phenotypic divergence, and discuss consequences for marine population connectivity and conservation.

Population genomics in marine lakes

We first tested whether marine lake populations showed distinct signatures from coastal ocean populations, and then asked whether the structure could be attributed to ecological or neutral processes. We found that mussel populations from coastal locations showed higher genetic diversity than populations sampled from marine lakes. That oceanic populations harbor higher genetic diversity than lake populations has also been observed in other marine lake mussel studies in Indonesia (Becking et al., 2016; de Leeuw et al., 2020), and in the marine lakes in Palau for jellyfish (Dawson and Hamner, 2005) and fish (Gotoh et al., 2009, 2011). Founder events predict that genetic diversity of populations within colonized habitats is lower than that of the source population (Mayr, 1942). Only a subset of the source population splits off to colonize a new habitat and as a result may stochastically display a distinct subset of the genetic variation. Furthermore, since the colonizing population is likely smaller, processes of genetic drift have larger effects and can cause declines in genetic diversity. In the current study, the severity of diversity loss increased with increasing degree of isolation (Table 1, Supplemental Fig. 4). Perhaps in more connected lakes propagules from the source population can more continuously enter the lake, providing a 'rescue effect' (Brown and Kodric-Brown, 1977). Founder events are generally associated with bottlenecks with subsequent population growth. While a clear bottleneck signature was found in marine lake mussels using single markers (de Leeuw et al., 2020), surprisingly all mussel populations analyzed in the current study showed signatures of population size decline without recovery, even the coastal ocean populations (Supplemental Fig. 5). The demographic similarity of all populations may be an artefact of the high level of diversity in RADtags which has been found for marine bivalves before (Hedgecock et al., 2005; Zhang et al., 2012; Harrang et al., 2013). The high diversity may overwhelm subtle recent patterns of population size changes. Still, we observed clear structuring per marine lake where we did not observe this in ocean locations, indicating a process causing differentiation is at play. The initially stochastic effect of founder events may be strengthened if environments act as a filter to effective colonization and/or establishment (Rundle and Nosil, 2005; Orsini et al., 2013; Wang and Bradburd, 2014).

If local environmental conditions pose selective pressures for the propagules that successfully establish within a new habitat, as has been shown for *Mytilus* mussels (Bierne et al., 2003), we would expect that similar environments show similar directions of population differentiation. We did observe increasing pairwise population differentiation with increasing isolation (and subsequent more distinct local environments) (Fig. 2C, 3, Supplemental Fig. 6), although there also was a clear geographic component to structuring the lakes in the PCA and co-ancestry analyses (i.e., North Misool, South Misool and Gam/Wayag). The differentiation was in clear contrast to the panmixia observed among oceanic populations, although there did seem to be a divergence between locations from Biak and two locations in Gam. However, in contrast to expectations from ecological speciation, we did not observe a isolation-by-environment pattern (Wang and Bradburd, 2014), and isolated populations did not cluster together. Hence, the genetic variation seemed to consist of a random sampling

of the source population, the severity of which may be dependent on the degree of isolation of the marine lake. This could explain why distant, isolated populations Papua 1, 31 and 32 from Gam and Wayag were found to cluster together, as they may have experienced similar levels of genetic drift associated with stringent founder events, and consequent differentiation from other more connected lakes. Yet, while genetic divergence may ultimately be formed by stochastic non-ecological processes, the accumulation of differentiation may be sped up by ecologically based divergent selection.

Another signature indicating an adaptive component of genetic differentiation is the indication of the presence of SNPs under putative selection. In total, 681 of the 10,836 SNPs retained in the current study were indicated as being potential loci under directional selection. The percentage of outlier loci detected in our study (6.3%) is relatively high compared to other marine molluscs: 3.2% in red abalone (*Haliotis rufescens*) (De Wit and Palumbi, 2013), 1.6% in sea scallop (*Placopecten magellanicus*) (Van Wyngaarden et al., 2016), 0.2% in eastern oyster (*Crassostre virginica*) (Bernatchez et al., 2018). Still, it falls within the expected 5-10% of the genome showing signatures of selection (Nosil et al., 2009). When investigating pairwise F_{ST} per contrast between categories (ocean, open lake, intermediate lake and isolated lake), we found that contrasts with isolated lakes showed the highest number of SNPs above F_{ST} 0.8. While this could indicate reduced gene flow due to local adaptation as a first stage in ecological speciation (Rundle and Nosil, 2005; Nosil and Schluter, 2011; Wang and Bradburd, 2014), instead it may also be caused by the reduced genetic diversity characteristic of especially isolated lakes (Cruickshank and Hahn, 2014). Here, genomic islands of differentiation could arise in the low diversity/recombination regions by chance. If ecological speciation underlies the patterns of differentiation, we expect similar phenotypic divergence to be related to genetic divergence.

Genetic adaptation versus phenotypic acclimatization

Further disproving a strong contribution of ecological speciation towards the differentiation of marine lake populations is the lack of correlations between genetic and phenotypic diversity (Fig. 4 and 5). This is in contrast to predictions of ecological speciation, where a phenotypic component is often found to be associated to genetic divergence (Rundle and Nosil, 2005; Schluter, 2009). Although for all phenotypic measurements we saw an influence of connection to the surrounding sea, the pattern and/or direction of this influence was dissimilar to those observed for the genetic diversity. For shell shape and thickness, this may imply the potential of phenotypic plasticity instead of genetic adaptation (Miner et al., 2005; Fierst, 2011), as has also been found for pelagic marine molluscs (Mariani et al., 2012; Mekkes et al., 2021). Members of the *Brachidontes* genus are known to be robust to variable environmental conditions (Sarà et al., 2008, 2013), and may morphologically or physiological adjust to diverse local environments rapidly without showing concordant genetic adaptation. Alternatively, shell shape may be more related to predation, the specific settlement location, and mussel density rather than local environments. We did observe significant differences

between categories of degree of connection to the surrounding sea. Perhaps indirectly connection to the sea determines likelihood of predator/competitor presence.

The absence of a correlation between genetic diversity and microbial communities associated to mussel gill tissue firstly indicates that we did not retrieve high microbial contamination in the RADtags. Secondly, we did see a distinct signature for microbial communities from water filters as compared to the signatures retrieved from gill tissue, indicating mussels may exert some selective pressure in associated microbial communities. We observed a significant effect of degree of connection to the surrounding sea in structuring microbial communities, yet no associations to host genetic diversity was found. Perhaps the local environment plays a larger role in determining the presence of certain microbes than host population structure (Webster and Thomas, 2016; Xu et al., 2017).

Trophic level of mussels is influenced by the food sources they take in (Katzenberg, 2000; Fry, 2006). As local environments change, potentially the trophic level may also shift (Nagelkerken et al., 2020). Of all diversity measurements, the trophic positioning of mussels showed highest concordance with genetic diversity (albeit still insignificant). The ocean populations showed distinctly higher $\delta^{13}C$ signatures, indicating a higher influence of marine resources compared to the marine lakes, which likely have more terrestrial input (Fry, 2006). As the lakes are relatively small and are completely surrounded by dense tropical forest, this is a likely scenario. We also observed clustering towards lower carbon levels for isolated lakes Papua 1 and Papua 31, which could be explained by Papua 1 being one of the most isolated lakes with the least influence of the surrounding sea, and Papua 31 being surrounded by mangrove trees (personal observation). Papua 6 showed higher $\delta^{15}N$ ratios, which is associated with a higher trophic level (Fry, 2006). Perhaps this lake lacks certain taxonomic groups which has allowed *Brachidontes* sp. to shift its level.

At similar time scales as marine lake populations, Momigliano et al. (2017) did find rapid ecological speciation for European flounders (*Platichthys flesus*) in the Baltic sea. They found differences in reproductive behavior (demersal versus pelagic spawning) which resulted in strong reproductive isolation. Repeated colonization of marine threespine stickleback (*Gasterosteus aculeatus*) into freshwater lakes has also occurred on similar time scales, and is associated with changes in morphology, behavior and physiology (Jones et al., 2012). That we did not find concordance between genetic and phenotypic divergence for marine lake mussels may relate back to the stochastic colonization and persistence of mussels in marine lakes due to their high robustness to a wide range in environmental conditions, as found for other members of the *Brachidontes* genus (Sarà et al., 2008, 2013). It is also possible we have not measured the morphological or physiological traits that are under divergent selection between lake categories. Using the unique replicated setting of marine lakes allows for further investigation of relative importance of mechanisms of microevolutionary population differentiation. Common garden and/or reciprocal transplant

experiments may further elucidate local adaptation and phenotypic plasticity in marine lake mussels (Nuismer and Gandon, 2008; De Villemereuil et al., 2016).

Marine population connectivity and conservation

In general, we found strong population structure even when including only putatively neutral genomic loci. This is in stark contrast to other RADseq studies in marine organisms, which generally find population panmixia when studying neutral loci, and differentiation only when studying loci under putative selection (Hemmer-Hansen et al., 2014; Milano et al., 2014; Van Wyngaarden et al., 2016; Bernatchez et al., 2018; Xuereb et al., 2018). Unlike expectations for marine populations (Puebla, 2009), we did not obtain compelling evidence for ecological speciation. Instead, we did observe an influence of geographic location and an isolation-by-distance pattern (Wright, 1943), and we found an influence of connection to the sea. The combination of strong patterns of neutral genetic structure, along with an influence of degree of isolation may indicate a potential for ongoing priority effects (Fukami, 2015; De Meester et al., 2016), also suggested by previous marine lake mussel studies (Maas et al., 2018; de Leeuw et al., 2020). In priority effects, first colonizers have a numerical advantage potentially strengthened by local adaptation, which results in their dominance (Waters et al., 2013; De Meester et al., 2016). Priority effects can persist for thousands of years, as observed in lake zooplankton (Ventura et al., 2014), even in scenarios with high dispersal. Hence, early eco-evolutionary dynamics in marine populations may have long-lasting effects on population connectivity. Our results show that differentiation in peripheral habitats which have a reduction in gene flow such as marine lakes may have contributed to the high marine biodiversity in the area.

The establishment of effective Marine Protected Areas (MPAs) is greatly aided by accurate knowledge on population connectivity (Allendorf et al., 2010; Carr et al., 2017). The replicated nature of marine lakes, along with the potential to elucidate small scale population structure by using high resolution marker panels gives us a unique opportunity to elucidate the influence of spatial and environmental factors in shaping population structure and connectivity. The patterns observed in the current study indicate an important influence of founder effects and priority effects that may persist for generations, despite the potential of ongoing gene flow. Habitat fragmentation due to anthropogenic influence and climate change may strengthen priority effects (Legrand et al., 2017). Designation of MPAs in affected or newly restored areas should take into account that first colonizers may impact the resulting community for generations. In any case, each marine lake, particularly the most isolated ones, should be considered unique evolutionary units, and should be included in MPAs of Indonesia (Maas et al., 2020).

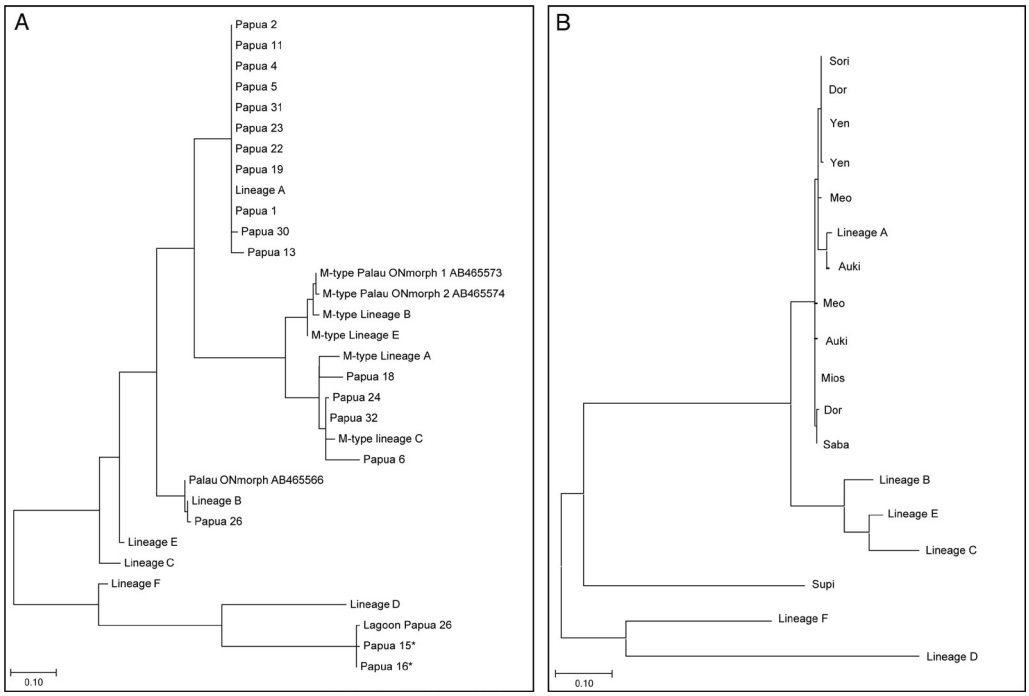
Certain marine lakes already show local conditions comparable to predictions for the oceans in the year 2100 by the Intergovernmental Panel on Climate Change (IPCC, 2014), namely elevated temperatures, and lower salinities and pH. The populations residing in marine lakes therefore may have the opportunity to pre-adapt to climate change. Future studies into

marine lakes may shed light into how communities and populations have assembled and potentially allow for a look into the future for coastal reef ecosystems.

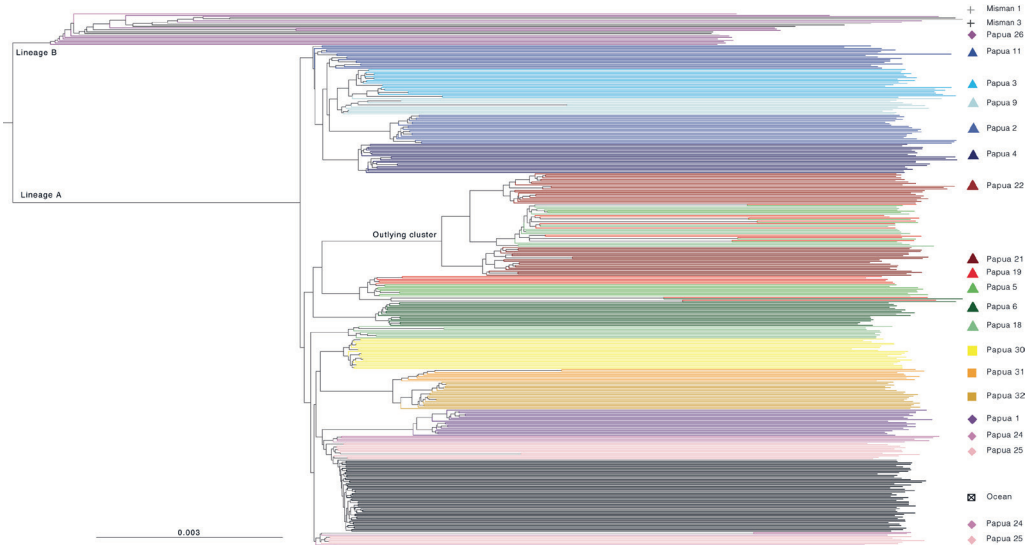
Acknowledgements

We would like to thank the following people who provided help with logistics and/or fieldwork: Max Ammer, Andy Miners, Prof. Tinka Murk, Misool Eco Resort, Papua Diving, and The Nature Conservancy. The following people provided invaluable contributions to genetic, morphological, microbial community and isotope analyses: Remco Pieters, Marina Martin, Merethe Zalm, Yinci Yan, Joris Geurts, Stephanie Martinez, Emma Pratt, Saskia Oosterbroek, Jennifer Koelmans. This study was funded by the Netherlands Organisation for Scientific Research provided to LEB through grant VENI# 863.14.020. Field work in Indonesia was made possible through additional financial support of the National Geographic grants (W162-11 and EC-181C-18), National Geographic Young Explorers grant (#Asia 62-16), WWF-INNO (16-032), Rufford Foundation, the Schure-Beijerinck-Popping fund of the Royal Dutch Academy of Science, and the Treub-Maatschappij fund. We are grateful to the Indonesian Institute of Sciences (LIPI) and the Indonesian State Ministry of Research and Technology and Higher Education (KEMENRISTEK-DIKTI) for providing research permits (098/SIP/FRP/SM/V/2011; #3B/TKPIPA/E5/Dit.KI/III/2016; 30/VI-P2SP2/ UNIPA/2017; 56/SIP/FRP/E5/Dit.KI/III/2018).

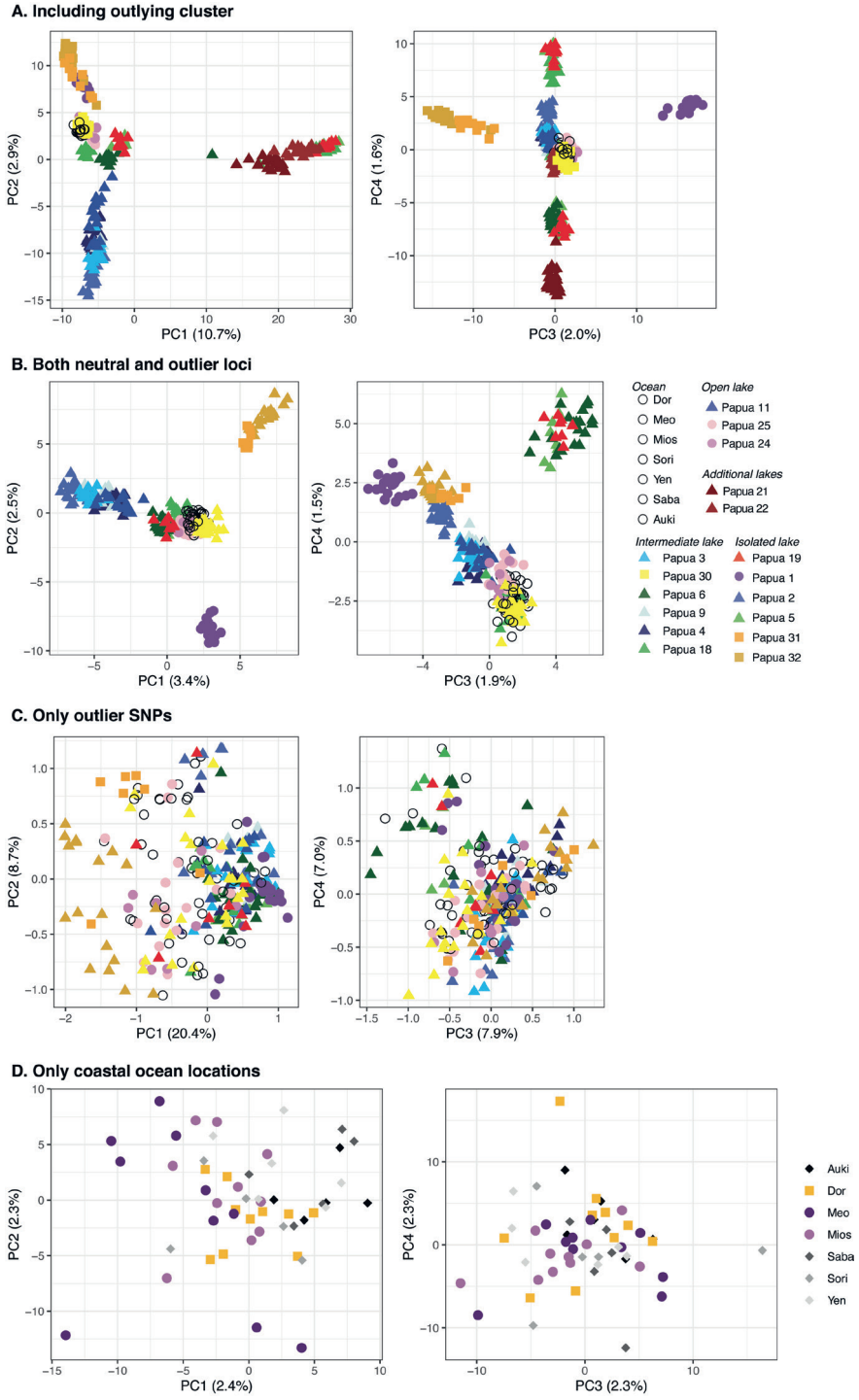
Supplemental Figures



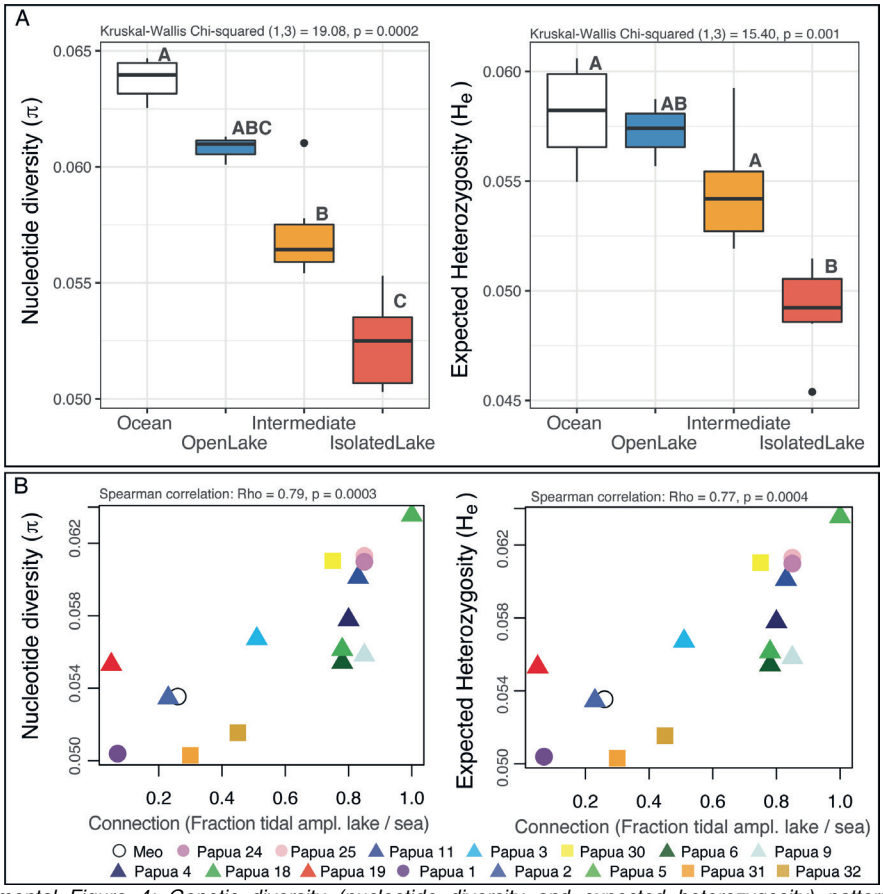
Supplemental Figure 1: Major genetic lineages of Brachidontes sp. based on COI (Goto et al., 2011) of (A) Marine lakes, and (B) Coastal ocean locations. M-type indicates male-type COI. Populations with asterisks were not genotyped.



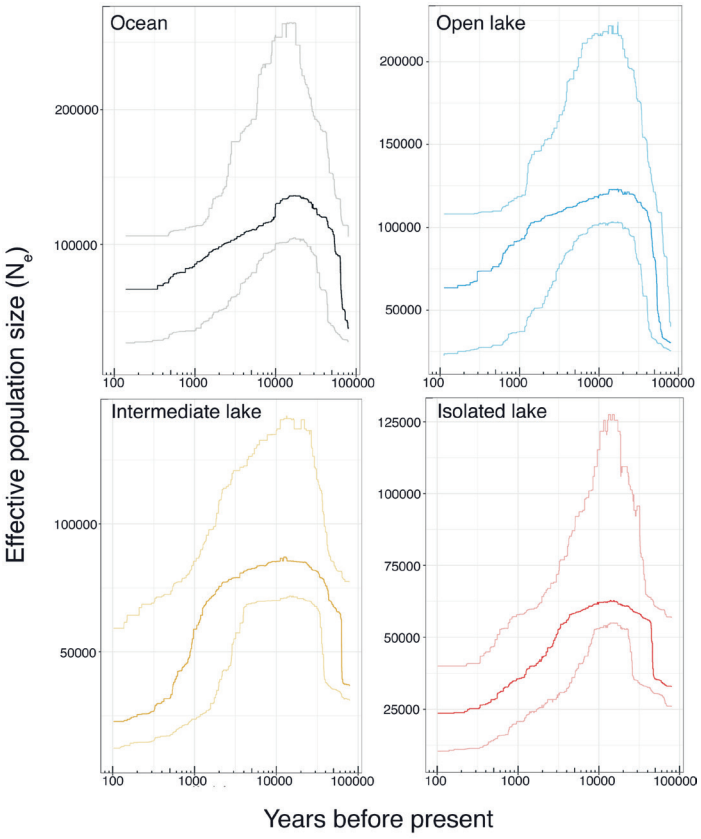
Supplemental Figure 2: Neighbor-Joining tree from on pairwise genetic distances. Preliminary tree based on 1 bootstrap run. Each branch represents one individual.



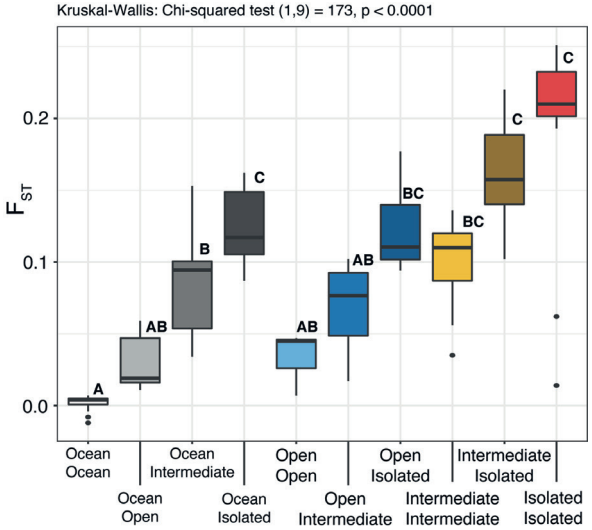
Supplemental Figure 3: Principal Component Analyses (PCA) on populations including (A) outlying cluster with 10,464 SNPs, (B) including both neutral and outlier loci with 10,836 SNPs, (C) only including outlier loci with 43 SNPs, (D) and including only coastal ocean locations with 15,055 SNPs.



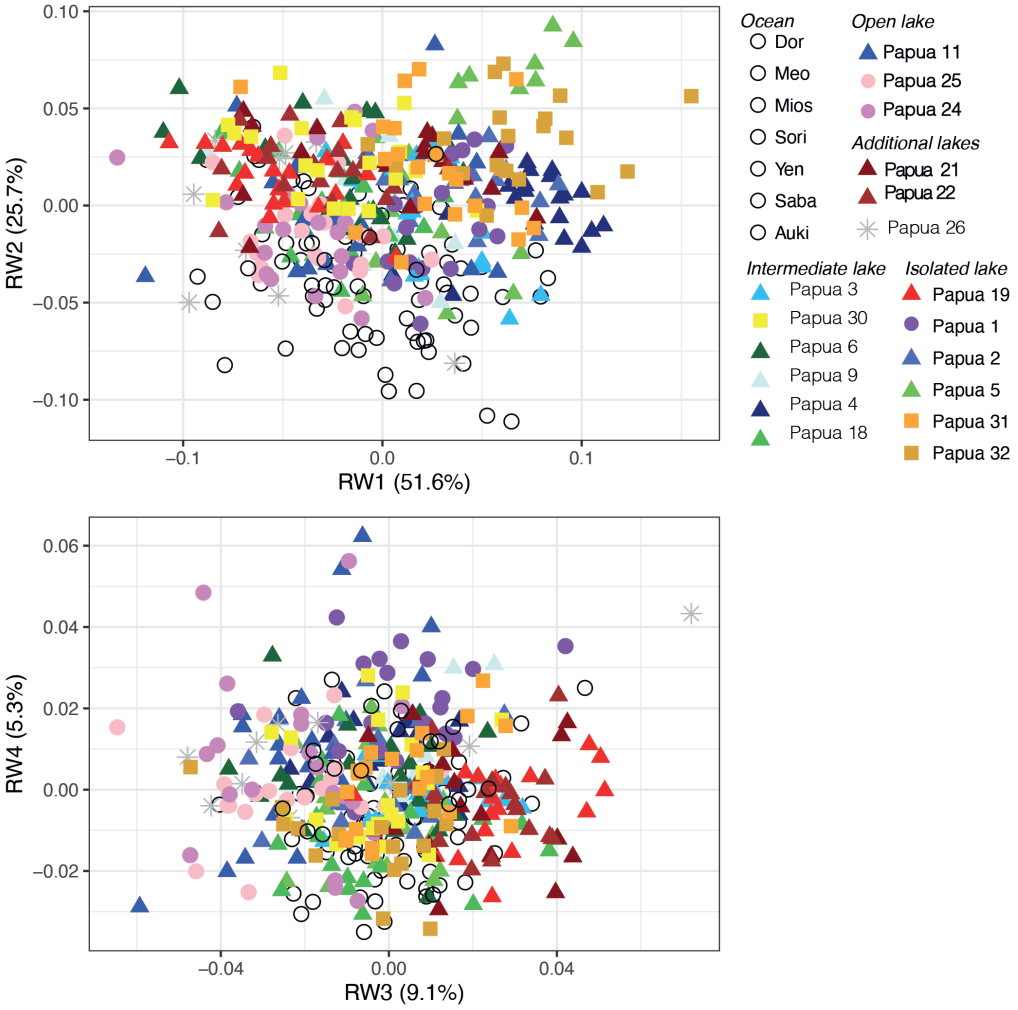
Supplemental Figure 4: Genetic diversity (nucleotide diversity and expected heterozygosity) patterns over categories and individual coastal and marine lake locations. (A) Boxplots representing genetic diversity within categories, tested for significance by a Kruskal-Wallis chi-squared test. Wilcoxon posthoc results are displayed in graph by letters, where if groups share the same letter they are not significantly different. (B) Spearman correlations of genetic diversity with degree of connection to the sea.



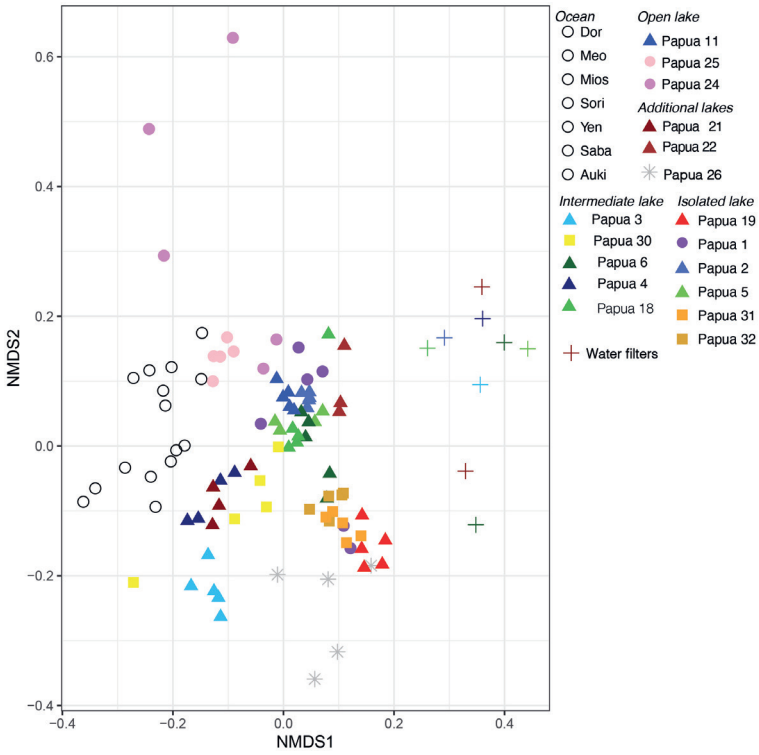
Supplemental Figure 5: Stairway plots representing changes in effective population size (N_e) of Brachidontes sp. populations. Four populations are represented: a coastal ocean population, a population from an open lake, a population from an intermediate lake, and a population from an isolated lake.



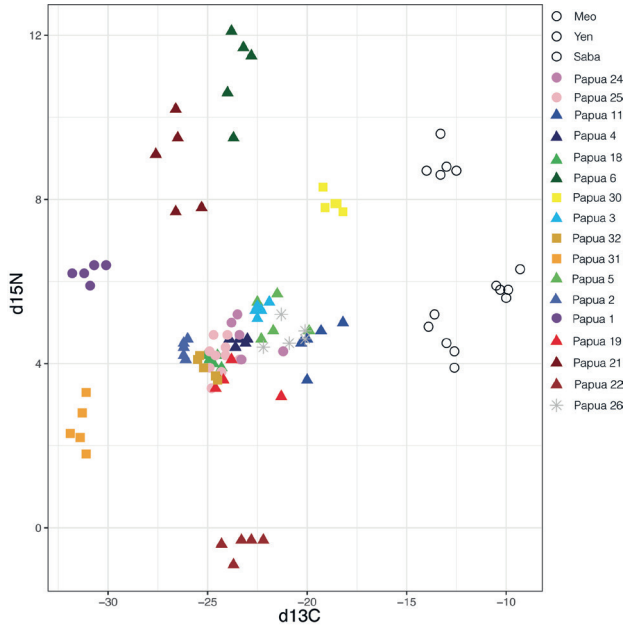
Supplemental Figure 6: Pairwise genetic differentiation (F_{ST}) between and among categories (ocean, open lake, intermediate lake and isolated lake). Kruskal-Wallis chi-squared test indicates differences among groups, and significant differences between comparisons are indicated with different letters.



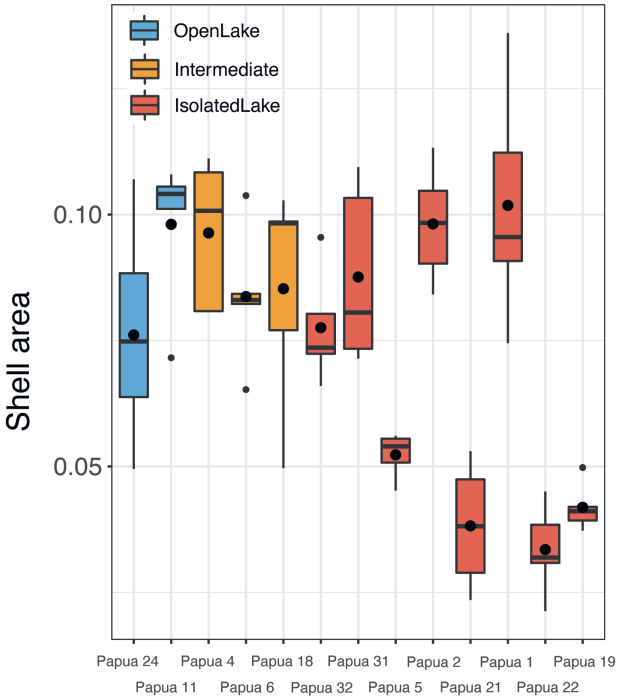
Supplemental Figure 7: Morphological variation displayed per lake through variation in shell outlines. Relative warps 1-4 of the outlines are displayed. Including Lineage B (Papua 26) and outlying lakes Papua 21 and Papua 22.



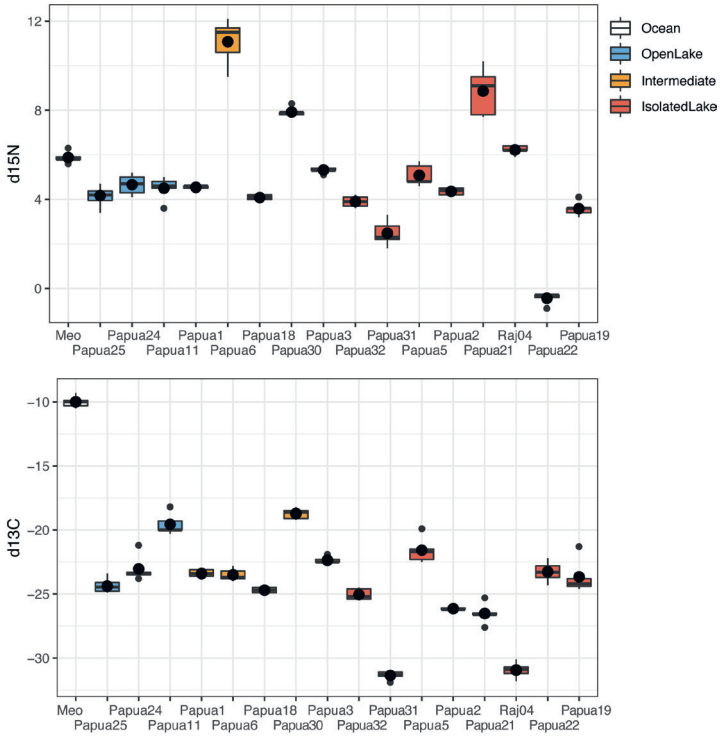
Supplemental Figure 8: Diversity of microbial communities (mussel tissue and water filters) as resulting from a non-metric Multidimensional Scaling analysis. Including Lineage B (Papua 26) and outlying lakes Papua 21 and Papua 22.



Supplemental Figure 9: Trophic level of mussels measured in $\delta^{13}C$ and $\delta^{15}N$. Including Lineage B (Papua 26) and outlying lakes Papua 21 and Papua 22.



Supplemental Figure 10: Shell thickness (measured as total area of shell as via microCT scanning). Ordered by connection (high to low), colored by category.



Supplemental Figure 11: Individual levels of $\delta^{13}C$ and $\delta^{15}N$. Ordered by connection, colored by category.

Supplemental Tables

Supplemental Table 1: Spearman correlations between nucleotide diversity and expected heterozygosity and temperature, salinity, pH, connection, lake area and lake depth. As multiple tests were run, the alpha was adjusted to 0.008 in order to be considered significant.

Nucleotide diversity	Rho	p-value
Temperature	-0.16	0.57
Salinity	0.29	0.28
pH	0.32	0.23
Connection	0.79	0.0003
Area	0.12	0.68
Depth	-0.06	0.83

Expected heterozygosity	Rho	p-value
Temperature	-0.10	0.72
Salinity	0.37	0.16
pH	0.27	0.31
Connection	0.77	0.0004
Area	0.10	0.71
Depth	-0.05	0.88

Supplemental Table 2: Pairwise population differentiation (F_{ST}) for 271 Brachidontes sp. individuals based on 10,294 neutral SNPs. Above diagonal significance values based on 1,000 bootstraps. Coloration ranging from high F_{ST} (red) to low F_{ST} (green). Locations colored by category: ocean (white), open lake (blue), intermediate lake (orange), isolated lake (red).

	Dor	Meo	Mios	Sori	Yen	Saba	Auki	Papua 24	Papua 25	Papua 11	Papua 9	Papua 4	Papua 18	Papua 6	Papua 30	Papua 3	Papua 32	Papua 31	Papua 5	Papua 2	Papua 1	Papua 19
Dor	--	0.039	0.041	0.151	0.482	0.233	0.807	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002
Meo	0.006	--	0.061	0.102	0.179	0.339	0.035	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Mios	0.005	0.004	--	0.165	0.068	0.359	0.243	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Sori	0.005	0.006	0.005	--	0.413	1	0.168	0.002	0.005	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.001	0.002
Yen	0	0.004	0.005	0.001	--	0.989	0.542	0.001	0.003	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Saba	0.004	0.002	0.001	-0.015	-0.012	--	0.915	0.002	0.025	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.005	0.001	0.001	0.001	0.001
Auki	-0.004	0.007	0.003	0.006	0	-0.008	--	0.001	0.002	0.001	0.001	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.001	0.002
Papua 24	0.024	0.013	0.014	0.025	0.019	0.016	0.021	--	0.037	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002
Papua 25	0.016	0.011	0.011	0.018	0.016	0.012	0.016	0.007	--	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Papua 11	0.056	0.052	0.046	0.05	0.05	0.047	0.059	0.045	0.047	--	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Papua 9	0.093	0.085	0.085	0.093	0.089	0.086	0.096	0.083	0.082	0.017	--	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Papua 4	0.099	0.092	0.088	0.096	0.096	0.089	0.1	0.091	0.09	0.048	0.056	--	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Papua 18	0.053	0.049	0.047	0.054	0.052	0.051	0.063	0.056	0.051	0.068	0.099	0.11	--	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Papua 6	0.1	0.097	0.097	0.103	0.108	0.103	0.112	0.102	0.093	0.096	0.125	0.136	0.125	--	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Papua 30	0.047	0.04	0.04	0.047	0.048	0.034	0.051	0.037	0.042	0.071	0.102	0.11	0.074	0.114	--	0.001	0.001	0.001	0.001	0.001	0.001	0.001
Papua 3	0.102	0.1	0.097	0.107	0.105	0.098	0.11	0.1	0.095	0.035	0.035	0.075	0.113	0.134	0.115	--	0.001	0.001	0.001	0.001	0.001	0.001
Papua 32	0.132	0.119	0.119	0.135	0.137	0.132	0.142	0.137	0.121	0.156	0.194	0.19	0.163	0.188	0.138	0.196	--	0.001	0.001	0.001	0.001	0.001
Papua 31	0.103	0.087	0.094	0.107	0.115	0.106	0.115	0.112	0.098	0.137	0.178	0.172	0.141	0.177	0.11	0.184	0.062	--	0.001	0.001	0.001	0.001
Papua 5	0.108	0.101	0.099	0.106	0.108	0.105	0.11	0.103	0.095	0.109	0.143	0.145	0.142	0.145	0.117	0.152	0.21	0.193	--	0.001	0.001	0.008
Papua 2	0.153	0.148	0.149	0.153	0.151	0.145	0.16	0.152	0.148	0.094	0.102	0.13	0.172	0.191	0.161	0.103	0.242	0.236	0.207	--	0.001	0.001
Papua 1	0.156	0.151	0.146	0.159	0.159	0.153	0.162	0.147	0.133	0.177	0.209	0.21	0.192	0.22	0.161	0.22	0.242	0.229	0.223	0.251	--	0.001
Papua 19	0.109	0.103	0.1	0.103	0.105	0.102	0.109	0.104	0.095	0.109	0.147	0.145	0.137	0.13	0.117	0.154	0.207	0.197	0.014	0.206	0.223	--

Supplemental Information A: QuadRADseq protocol

Available upon request: diede.maas@wur.nl

Chapter 7

General Discussion

The world's oceans are vast, and seemingly without boundaries. Together with the pelagic larval stage of most marine organisms, this leads to expectations of high dispersal abilities, large effective population sizes, and subsequent low population differentiation (Crawford and Oleksiak, 2016). However, with the advance of high-throughput sequencing, marine populations are shown to be more structured than previously thought (Oleksiak and Rajora, 2020b). The underlying processes that shape marine diversity patterns are not fully understood, although an important role of processes that allow diversification despite ongoing gene flow such as ecological speciation are expected (Rundle and Nosil, 2005; Puebla, 2009).

In this thesis, I used marine lakes in Indonesia as an opportunity to study microevolutionary patterns of marine invertebrate populations and the underlying processes most likely to have shaped them. My objectives were to set environmental and biological baselines of marine lakes, assess how benthic groups and populations of three invertebrates conformed to expectations of Island Biogeography Theory, test whether population genomics revealed patterns of panmixia or genetic structure, and to quantify the relative importance of neutral (geographic and physical barriers to dispersal) and selective (environmental) processes on population differentiation. I conclude that historical processes during colonization may have long-lasting effects on population structure in marine lakes, regardless of the species investigated. The effects could be exacerbated by decreasing connection to the surrounding sea and potentially associated distinct local environments. Below, major findings per chapter are summarized (Figure 1).

In **Chapter 2**, I identified 32 marine lakes in Raja Ampat, of which 17 were new to science. For a subset of 20 lakes, environmental and biological surveys were done to set baselines. The lakes had variable connections to the surrounding sea (relative tidal amplitude 5-89% compared to the surrounding sea) and span a temperature gradient of 6.8°C (30.0-36.8°C). Three categories could be distinguished: lagoon-like lakes with rich communities, highly isolated lakes with distinct local environmental conditions and depauperate communities, and intermediate lakes. Connection to the surrounding sea and water temperature were found to be important predictors of coverage and richness of major benthic groups consisting of broad taxonomic groups (corals, sponges, molluscs, cyanobacterial mats, macroalgae and other marine invertebrates). Indications of species-isolation, but not of species-area relationships were found.

In **Chapter 3**, I investigated genetic, ecological and social context of *Mastigias papua* (Lesson, 1830) to provide scientific basis to protect marine lakes as individual management units. In Raja Ampat, these lakes are increasingly becoming a tourist attraction, yet they are not included in conservation management plans. Strong genetic differentiation was observed among lakes using the genetic marker *COI* (ϕ_{ST}: 0.30-0.86), potentially including unique subspecies. The differentiation was also reflected in significant morphological differences. Major fluctuations in jellyfish abundance were observed, with no temporally consistent

pattern across lakes. Tenure disputes were indicated, and perceived threats included conversion of lakes to aquaculture ponds, introduction of invasive species and unregulated tourism. A 30-fold increase in tourism to Raja Ampat was found since 2007, which likely poses an increasing pressure on the local ecosystem and warrants restrictions on jellyfish lake visitation. Based on unique genetic and morphological diversity, there is scientific basis to delimit the lakes as individual management units.

In **Chapter 4**, I assessed population structure and effects of underlying processes of the sponge *Suberites diversicolor* (Becking and Lim, 2009), and compared results from high-resolution genetic markers with previous low-resolution genetic marker work. Both low- and high-resolution marker sets resolved two broad genetic lineages, yet the high-resolution marker panel was able to describe previously undetected population structure. Genetic diversity was observed to be higher in ocean populations compared to lake populations. Lake populations also showed signatures of having gone through population bottlenecks. Strong genetic differentiation was observed between populations. I found no evidence for an influence of geographic distance, connection to the sea or local environments on the observed strong population structure, suggesting other drivers of differentiation may be at play.

In **Chapter 5**, I assessed the genetic structure and the relative roles of spatial and environmental factors for the mussel *Brachidontes* sp. (Swainson, 1840). Genetic diversity within populations showed a trend of increasing with increasing connection to the surrounding sea. *Brachidontes* sp. show high population differentiation between lakes (F_{ST}: 0.07-0.24), that was influenced by geographic distance on broad spatial scales (>1,400km), accompanied by an influence of connection to the sea on a smaller spatial scale (<200km). No influence of isolation-by-environment was observed. It is suggested that even currently incomplete barriers to dispersal could have caused the initial isolation necessary for population differentiation.

In **Chapter 6**, I explicitly tested for relative importance of ecology-driven versus neutral parapatric population differentiation in the mussel by combining genomic and phenotypic data to evaluate whether ecological speciation could underlie population structure. I distinguished neutral SNP markers from potential SNPs under selection. Neutral markers allow the investigation of processes such as gene flow and changes in population size, whereas markers under putative selection allow for the investigation of the adaptive potential of a population. Including 18 marine lakes and 9 coastal ocean populations, I saw significant population differentiation among lake and lake-ocean comparisons, contrasted by panmixia among ocean locations. Strong clustering per marine lake was observed, regardless of using only neutral SNPs or when SNPs under putative selection were also included. Geographic location and the degree of connection to the sea were observed to have a significant influence on population genetic structure. No associations between genetic and phenotypic divergence were found, suggesting that ecological speciation did not play a

dominant role in generating the observed patterns. Instead, neutral dispersal limitations such as geographic distance, (incomplete) physical dispersal barriers and priority effects appear to play a more important role.

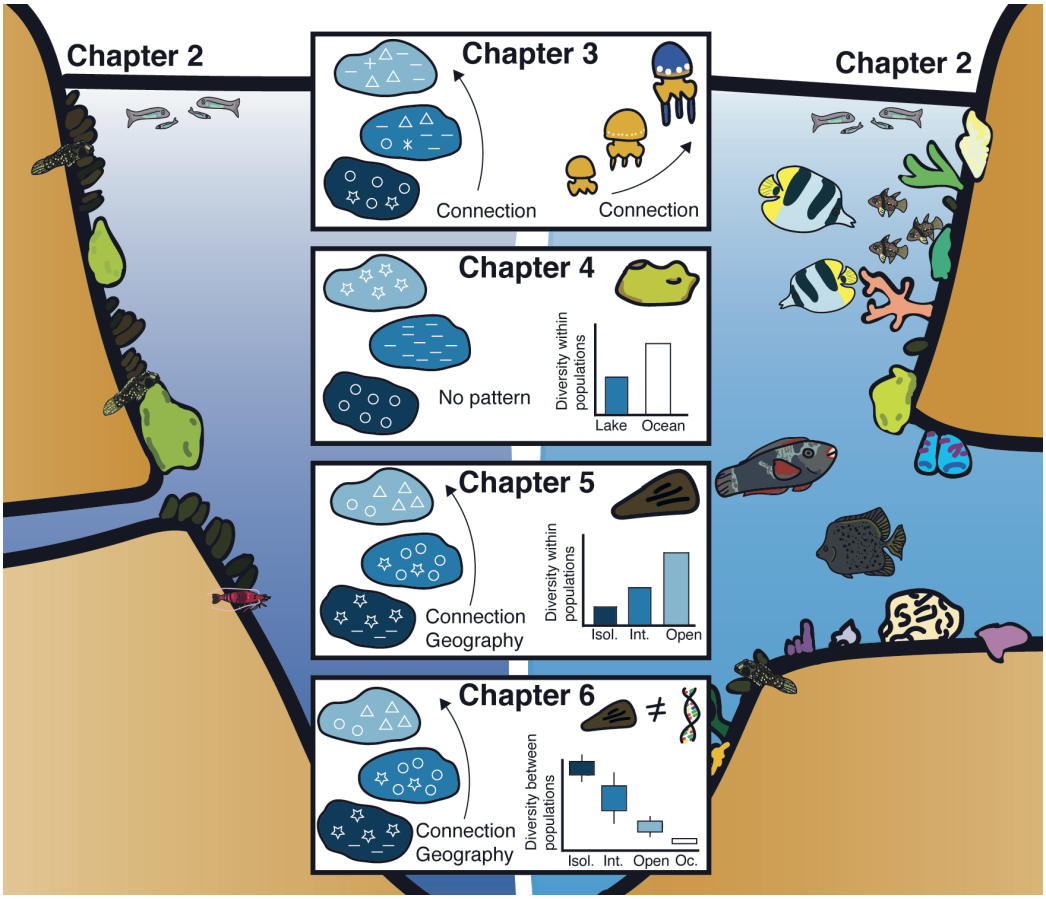


Figure 1: Schematic overview of major findings per chapter. Within the lakes, the shade of blue indicates the connection to the surrounding sea (dark blue = low connection, light blue = high connection). The symbols within the lakes in Chapter 3-6 indicate genetic diversity (different shapes = different genotypes). Simplified results for within and between population genetic diversity are displayed. In Chapter 6, the discordance between phenotypic and genetic divergence is indicated. Abbreviations: Isol. = Isolated, Int. = Intermediate, Oc. = Ocean.

Marine lakes as island-like systems

Islands have been at the basis of important ecological and evolutionary theories, being 'natural laboratories' (Darwin, 1872; MacArthur and Wilson, 1967; Warren et al., 2015). Geographically, islands are defined as land masses completely surrounded by water, yet biologically, the definition of what makes an island is more flexible (Dawson, 2016; Itescu, 2018). Insular systems can therefore be defined to include both true islands (portion of land surrounded by water), or island-like systems (any isolated habitat) (Itescu, 2018). Characteristics shared by all insular systems are limited size, spatial fragmentation, and low to no connection between systems. How patterns of biodiversity compare across different island-like systems has been defined as one of the remaining questions in island biology (Patiño et al., 2017), and this thesis aimed to add a marine perspective on insular biodiversity patterns and dynamics using marine lakes.

Marine lakes share the characteristics defined for insular systems: they are of limited size, are spatially fragmented by being completely surrounded by land, and have limited connection to the surrounding sea and thus each other. As such, marine lakes physically and geographically conform to characteristics defined for insular systems. Marine lakes formed *de novo* as new features in the landscape when natural basins in karstic islands were filled with seawater after the Last Glacial Maximum (Tomascik and Mah, 1994). Communities arising via successional colonists within marine lakes thus are of known maximum age of around 6,000-12,000 years (Sathiamurthy and Voris, 2006; Dawson, 2016). Isolation of marine lakes is assumed to be both physical, by a limited connection to the surrounding sea, and physiological, as the local conditions of marine lakes usually show higher water temperatures, lower salinity, and sometimes stratification (Hamner et al., 1982, **Chapter 2**). Below, I will discuss how patterns observed from marine lake communities and populations relate to the biogeographical predictions on species-area and species-isolation relationships (MacArthur and Wilson, 1967), and predictions of the microevolutionary processes of genetic differentiation.

Following Island Biogeography Theory, the species-area relationship predicts that the number of species is larger in larger islands (MacArthur and Wilson, 1967). Contrasting to this expectation, I found that neither richness (number of different major benthic groups) nor Shannon diversity (calculated from coverage of major benthic groups, also including an evenness component) were correlated to lake area (**Chapter 2**). It has been suggested that species diversity and genetic diversity are structured by the same factors and thus should show similar patterns (Vellend, 2003; Vellend and Geber, 2005). If I assume this to be true, I expect to find positive relationships between genetic diversity and lake area. However, for the sponge (*S. diversicolor*) and the mussel (*Brachidontes* sp.) I found no relationship between genetic diversity and total lake area (**Chapter 4, Chapter 6**). A study into species-area relationship in freshwater lakes found that until an area of 1,000km^2 they could not detect a correlation between cichlid species richness and lake area (Wagner et al., 2014). An explanation could then be that the marine lakes investigated in this thesis are too small

(average area 16,500m^2, **Chapter 2**) for lake size to play a role in facilitating diversification. This is in contrast to a study using a single genetic marker to investigate genetic patterns of *Brachidontes* sp. in the same marine lake setting, where they did find a significant relationship between genetic diversity and lake area (de Leeuw et al., 2020). For the mitochondrial marker cytochrome oxidase subunit 1 (*COI*) they found that when lake area decreased, haplotype diversity did as well. A relationship between genetic diversity using COI and lake area was also found for jellyfish in marine lakes in Palau (Dawson and Hamner, 2005). Mitochondrial DNA is considered to have lower effective population size than nuclear DNA due to its maternal inheritance and haploidy (Lynch, 1997). Hence, decreasing habitat size may impact mitochondrial DNA diversity more severely than nuclear DNA diversity, as with lower effective population size, diversity is more easily lost. Another factor contributing to high nuclear diversity is the generally high nuclear genome-wide polymorphism found for marine bivalves, which may further explain the discrepancy (Rosa et al., 2015; Halanych and Kocot, 2017).

Another prediction arising from Island Biogeography Theory is that with increasing isolation of an island, the number of species will be lower (MacArthur and Wilson, 1967). For true islands, the degree of isolation is defined as the distance from the mainland, the source pool for potential colonizing species. For marine lakes, it is difficult to define true isolation, as it may be affected by multiple factors. The 'mainland' for marine lakes is the surrounding sea (Dawson et al., 2009). Calculating the 'distance to the sea' may be one way of defining marine lake isolation. In **Chapter 2**, I did not find a correlation between distance to sea and number of major benthic groups or Shannon diversity. In addition, and for the other chapters, I decided to use a different proxy for connection to the sea. I measured maximum tidal amplitude in the lake and the directly adjacent sea and defined the connection as their ratio (Hamner and Hamner, 1998). Using this proxy, I observed strong relationships between connection and benthic group richness (although not for Shannon diversity) (**Chapter 2**), for some bioinformatic filtering options between connection and genetic diversity in *S. diversicolor* (**Chapter 4**), and between connection and *Brachidontes* sp. genetic diversity (**Chapter 5** and **6**). Increasing diversity with increasing connection to the sea could be explained by either the initial colonizing population showing higher effective size and thus retained higher ancestral polymorphisms, or by highly connected lakes having ongoing gene flow with the source population from the surrounding sea, providing a rescue effect (Brown and Kodric-Brown, 1977). Using mitochondrial DNA, no correlation between genetic diversity and degree of connection to the sea was observed for *Brachidontes* sp. (de Leeuw et al., 2020). Perhaps other factors such as lake size more severely impact mitochondrial diversity. De Leeuw et al. (2020) suggest eco-evolutionary dynamics of priority effects to underlie the absence of a correlation between connection and genetic diversity. If early colonizers gain a numerical and/or evolutionary advantage, they may effectively outcompete novel migrants. Hence, the genetic diversity present in the colonizing population would effectively be fixed over generation. If this is the case, we would expect strong population differentiation between lakes, despite degree of connection to the sea. To further investigate such eco-

evolutionary effects, the next step is to look at microevolutionary processes shaping genetic diversity and differentiation.

Islands have been the basis of studying for macro- and microevolutionary processes. The Island Biogeography Theory originally did not contain an evolutionary component (Whittaker et al., 2017). However, if an island is isolated enough, it will allow for the arising of new species through cladogenesis (macroevolutionary process), of which population differentiation is a first stage (microevolutionary process). For jellyfish in the marine lakes of Palau, molecular differences between lakes and lagoon specimens were found to be so extensive that they were split into five new subspecies (Dawson, 2005b). Following this framework, I also observed potential subspecies of *Mastigias papua* in the marine lakes of Raja Ampat (**Chapter 3**). Finding divergent genetic lineages that potentially represent cryptic species is common for anchialine systems (see references in **Chapter 1**, **Chapter 3**). Moving beyond the exact definition of (sub)species, we observed marine lakes to have unique population genetic variation, as exemplified by clear genetic clustering per marine lake and significant differentiation between lakes (**Chapter 3-6**), and between pairwise comparisons of ocean locations versus lakes (**Chapter 6**). In contrast, I observed signatures of panmixia among ocean populations (**Chapter 6**). Other population genetic studies from marine lakes confirm the observations differences between lake and ocean population differentiation (Dawson and Hamner, 2005; Gotoh et al., 2009, 2011; Becking et al., 2013b, 2016; Swift et al., 2016; de Leeuw et al., 2020). What is also consistently found, and what may also relate to the findings of high genetic differentiation, is that populations from oceanic locations show higher genetic diversity than populations from marine lakes (**Chapter 4** and **6**) (Dawson and Hamner, 2005; Gotoh et al., 2009, 2011). Even the lakes with a high connection to the surrounding sea showed trends towards having lower diversity, although this pattern was most apparent for the most isolated lakes (**Chapter 6**). Therefore, even a slight reduction in connection to the surrounding sea already decreases the genetic diversity within the lakes.

Both the lower genetic diversity within lake populations and the higher genetic differentiation between populations could represent classic consequences of founder events (Mayr, 1942; Frankham et al., 2002). Here, a small subset of the main population splits off to colonize a new habitat, resulting in a loss of diversity through genetic drift associated with smaller population sizes. Indeed, for the sponge we observed severe signatures of former bottlenecks for the marine lake populations (**Chapter 4**), potentially also owing to their asexual reproductive strategies. Contrastingly, for the mussel we did not see a difference in the demographic histories between lakes and ocean locations, as they all showed a gradual decrease in effective population size (**Chapter 5** and **6**). This is interesting, because we did see clear clustering per lake, a signature of founder events. Furthermore, for the mitochondrial marker, de Leeuw et al., (2020) did observe signatures of bottlenecks marine lake mussel populations. As mentioned before, the lower effective population size of mitochondrial DNA as compared to nuclear DNA may have caused why bottlenecks are more easily detectable for mitochondrial DNA than for nuclear DNA (Lynch, 1997), which is

known to have high heterozygosity in marine bivalves (Rosa et al., 2015; Halanych and Kocot, 2017). Alternatively, it has been questioned whether mitochondrial DNA is the neutral marker it is generally assumed to be (Dowling et al., 2008). The demographic patterns of bottlenecks with associated population differentiation in mitochondrial DNA may therefore also have a selective component.

Although the marine lakes of Indonesia do not conform to all predictions on species/genetic diversity versus island size and isolation established for true islands (i.e., no relationship with lake area was found), we reaffirm their usefulness as insular systems. The populations residing within marine lakes seem to be on distinct eco-evolutionary paths, and harbor unique biodiversity. Marine lakes have clearly defined populations, are replicated over space, and can be selected to reflect a gradient in local environmental conditions (connection, water quality and physical lake characteristics). As such, they can be used as model systems to test the influence of processes of diversification, and to unravel mechanisms difficult to investigate in open ocean systems.

Processes underlying marine population genomic structure

If signatures of population genetic structure are found, the null hypothesis of panmixia can be refuted. The next step is to understand what underlying processes are causing the observed structuring of populations (Orsini et al., 2013). Several theories predict how populations may become isolated to such an extent that they show differentiation, such as theories of isolation-by-distance (Wright, 1943), isolation-by-environment (Wang and Bradburd, 2014), isolation-by-resistance (McRae, 2006), and historical contingency patterns such as priority effects (Fukami, 2015; De Meester et al., 2016). Marine lakes provide an opportunity to test the relative contribution of these different processes towards population differentiation.

Depending on the spatial scale, for the mussel I observed strong effects of geographic distance on genetic structure (spatial scales ~1,400km and ~200km, **Chapter 5** and **6**), and effects of connection to the surrounding sea (spatial scale ~200km, **Chapter 5** and **6**). In **Chapter 6** also a significant correlation to local environments was found, yet throughout the thesis autocorrelation between connection to the sea and local environments was observed (explicitly reported in **Chapter 2** and **6**), making those two difficult to disentangle completely. Interestingly, even though the study investigating the sponge *S. diversicolor* was performed on a large spatial scale (~1,400km) and we found strong population genetic structure, we did not find any correlations of genetic variation to geographic location, connection to the sea or local environmental conditions (**Chapter 4**). Other studies using in marine lakes using low-resolution markers also found contradicting patterns depending on the species investigated. For *Brachidontes* sp. on similar spatial scales, de Leeuw et al. (2020) found no effects of either geographic location or connection to the sea using mitochondrial DNA data. However, a pattern of isolation-by-distance was observed for the jellyfish *Mastigias papua* in Palau using mitochondrial DNA (Dawson and Hamner, 2005). Processes shaping genetic

diversity are complex, not mutually exclusive, and may vary over different spatial and temporal scales. While this complicates their elucidation, marine population genomic studies are working towards uncovering genetic-environment associations.

Seascape genomics (and genetics) is performed to reveal how space and environment shape microevolutionary processes (Selkoe et al., 2016; Liggins et al., 2020). The null hypothesis here is an association between genetic and geographic distance (isolation-by-distance). This association has been observed in RADseq studies on marine invertebrates for example in black-lip pearl oyster (Lal et al., 2016), sea scallop (van Wyngaarden et al., 2017), and the eastern oyster (Bernatchez et al., 2018). Contrastingly, for the California sea cucumber isolation-by-resistance through ocean currents was a better predictor of population structure than isolation-by-distance (Xuereb et al., 2018). A major observation in oceanic population genomic studies is that while neutral markers tend to show panmixia, a subset of SNPs under putative selection do often show population structure (Crawford and Oleksiak, 2016). These outlying SNPs could indicate adaptive divergence in a process of ecological speciation. For example, for red abalone the majority of SNPs attained via transcriptome analyses showed no population differentiation, but a subset of outlying SNPs showed significant structure (De Wit and Palumbi, 2013). The authors found that these SNPs coded for genes involved in biomineralization, tolerance to heat, disease or hypoxia and energy metabolism, and thus provide candidate loci for local adaptation. For the American lobster, neutral variation was found to be structured by ocean currents, while thermal tolerance was thought to underlie adaptive variation (Benestan et al., 2016). With increasing genomic resources, it is becoming more and more apparent that adaptation to local environments can arise even in the absence of physical boundaries (Oleksiak, 2018).

Contrasting to seascape genomic studies, in this thesis I consistently find strong population structure even for neutral genomic marker datasets (**Chapter 3-6**). In contrast to ocean marine populations where only few adaptive SNPs show signatures of ecological speciation in the face of ongoing gene flow (Crawford and Oleksiak, 2016; Oleksiak, 2018), marine lakes potentially reflect true boundaries to effective gene flow via their limited connection to the surrounding sea, potentially aided by distinct local environments. With the rising of seawater levels after the Last Glacial Maximum, propagules may have inoculated the marine lakes after which they became isolated (Tomascik and Mah, 1994; Dawson et al., 2009; Dawson, 2016). Founder events may have caused allele frequencies in the subset of a population to be dissimilar from the source population by chance (Mayr, 1942; Frankham et al., 2002). Furthermore, processes of genetic drift, where random alleles are lost and fixed, are expected to have a larger effect in small populations, exacerbating the initial founder events (Frankham et al., 2002). However, the potential of ongoing gene flow in and out of the marine lakes is still there because the lakes show some degree of tidal fluctuation (**Chapter 2**). Especially in the lakes with extensive tidal fluctuations and connection to the surrounding sea via caves we can expect propagules to continuously enter the lake. Assuming ongoing gene flow, the pattern of finding distinct populations on small spatial

scales could point towards eco-evolutionary dynamics of priority effects playing an important role (Fukami, 2015; De Meester et al., 2016).

Priority effects arise from a numerical advantage of first colonizers which effectively blocks subsequent species or genotypes from successful establishment (Fukami, 2015; De Meester et al., 2016). For freshwater lake populations, such priority effects have been shown to structure populations over thousands of generations (Ventura et al., 2014). For marine lakes, if priority effects are important, such persistence is also observed for populations that have been isolated for approximately 12,000 years (**Chapter 3-6**). For priority effects to be able to arise there needs to be a sufficient time lag between first and subsequent colonizers (Fukami, 2015). The decrease in connection to the surrounding sea observed in marine lakes may provide this opportunity. This would also explain why we find correlations between genetic variation and connection to the surrounding sea, as this effect can be expected to be more important with increasing degrees of isolation. Recently, an adaptive component has been added to the theory of priority effects, termed monopolization (De Meester et al., 2016). It suggests that besides first colonizers having a numerical advantage, they also have the opportunity to locally adapt, giving them a 'head start' and making them even stronger competitors towards new colonizers. While it is difficult to explicitly test for this prediction, it would explain why we find persisting patterns of unique variation between lakes, even lakes which are within 10km to each other. In any case, the patterns we observed are dissimilar to expectations of ecological speciation in marine systems, in that we see persistent patterns in neutral genomic markers. Coalescent and model-based analyses of divergence timing and migration may shed more light on how the marine lake populations are connected (Excoffier and Foll, 2011).

On the road to speciation

Our findings of high genetic structure on small spatial and temporal scales have parallels in other aquatic ecosystems. For European flounders in the Baltic Sea, rapid ecological speciation was found (Momigliano et al., 2017). Strong reproductive isolation over only a few thousand generations was observed and related to divergent reproductive behavior of this species. Highest genetic differentiation was found in comparisons between pelagic and demersal spawning strategies, regardless of geographic distance, suggesting ecological speciation to underlie differentiation in parapatric flounder populations. On a similar timescale (after the Last Glacial Maximum), the marine lake invertebrates of this thesis also show strong clustering per lake and appear to be at the beginning of the speciation continuum (**Chapter 3-6**), regardless of their differences in life histories. For the mussel, I observed neutral processes to play a role (isolation-by-distance and by-resistance) (**Chapter 5-6**) and found no relationships to phenotype (**Chapter 6**). I observed high genetic differentiation between contrasting habitats (low versus highly connected lakes) but highest differentiation was observed within habitat types, namely among isolated lakes (**Chapter 6**). The Baltic sea populations represent only one instance of eco-evolutionary dynamics and comparing marine lakes to other replicated systems could provide further insights.

Multiple replicated systems have been defined for threespine stickleback fish. The fish show interesting patterns of parallel evolution between marine and freshwater (Jones et al., 2012), and between lake and stream habitats (Hendry et al., 2009; Roesti et al., 2012, 2015). Having colonized the lakes and streams from the marine realm after the Last Glacial Maximum, and showing repeated independent population pairs, the stickleback system shows clear overlap with the marine lake system (**Chapter 2-6**). Marine and freshwater stickleback populations differ phenotypically in body shape, plate armor, how they deal with osmotic pressures and mating preferences, among others (Jones et al., 2012). Lake and stream populations most prominently differ in foraging strategies, pelagic versus benthic foraging, respectively (Roesti et al., 2012, 2015). Jones et al. (2012) sequenced the genomes of stickleback populations across the two habitats and found a set of loci associated with the phenotypic divergence. In sliding window analyses, for neutrally evolving markers they found patterns of panmixia, while for certain coding and/or regularly loci thought to underlie the phenotypic differences they observed parallel divergence and high structure. For marine lake invertebrates, I observed strong clustering per lake regardless of including neutral or putatively selective loci (**Chapter 3-6**). I did, however, observe more unique outlier SNPs in pairwise comparisons made among the isolated lakes than for comparisons among connected lakes (**Chapter 6**). Although this may consist of stochastic differences between genomic variants, potentially these SNPs indicate functional responses to local environments.

For lake and stream populations of stickleback a high number of genes across the genome were found to be involved in divergence between the two habitats, accompanied by strong differentiation even in neutral markers on small spatial scales and in the absence of dispersal barriers (Berner et al., 2009; Roesti et al., 2012). This provides a clearer parallel to the findings from marine lake invertebrates. However, divergence patterns among replicates of lake and stream stickleback population pairs were inconsistent, making generalizations difficult (Roesti et al., 2012). Furthermore, in some cases the hypothesis of parallel evolution was rejected, and divergence might instead have been established via ecological vicariance (Roesti et al., 2015). For marine lake invertebrates, the exact processes underlying differentiation are not fully understood, but the ecological component is further explored in the section 'Adaptation and phenotypic plasticity'.

Threespine stickleback fish have been used to define different stages in the speciation continuum (Hendry et al., 2009), including 1) panmixia, 2) incomplete patterns of variation with minor reproductive isolation, 3) strong patterns of variation yet still reversible, and 4) complete, irreversible reproductive isolation. For sticklebacks, transitions between state 1 and 2 involved divergent selection coupled with at least some geographic isolation, be it through allopatry (different lakes) or parapatry (lake and stream). Reaching state 4 is thought to involve environment-independent genetic restructuring such as chromosomal rearrangements. All marine lake invertebrates investigated in this study appear to be at the

transition between stage 1 and 2 (**Chapter 3-6**), where a limitation in dispersal through the connection to the surrounding sea appeared to be enough to generate strong population structure.

The explosive adaptive radiation of cichlid fish in African rift lakes likely represent full speciation on short time scales. (Genner et al., 2007; Berner and Salzburger, 2015; Brawand et al., 2015). For Lake Malawi, it was shown that cycles of habitat isolation and reconnection established the current fish diversity (Genner et al., 2007). Diversifying in a satellite lake, a species of cichlid fish developed different assortative mating strategies associated with male nuptial color patterns. After reconnecting with Lake Malawi, it contributed to the overall fish diversity. Hence, peripheral habitats were shown to contribute to biodiversity patterns. Marine lake populations appear to follow a similar trend, by harboring unique genetic diversity (**Chapter 3-6**), with perhaps even subspecies (**Chapter 3**). If the diversity within lakes spills over to the open ocean, the marine lakes could represent an argument towards the Coral Triangle being a Center of Origin (Tornabene et al., 2015). Marine lakes could then add the component of diversification in peripheral habitats to the debate on what is causing the high marine biodiversity in the area (Hoeksema, 2007).

Adaptation and phenotypic plasticity

For the jellyfish (**Chapter 3**) and the mussel (**Chapter 6**) I studied the variation in selected morphological and ecological characters, reflecting phenotypic variation and possible acclimatization to the environment. In both species, significant differences in phenotypes were observed among the marine lakes. Strikingly, the phenotypic divergence showed no concordance to genetic divergence. Below I will elaborate on the signatures observed for the two invertebrates.

Consistent with previous studies in jellyfish from Palau (Dawson, 2005a; Dawson and Hamner, 2005; Swift et al., 2016), I saw morphological variation in *Mastigias papua* that was to a certain extent related to the local environment of the lakes. In more isolated lakes, I observed an ecotype distinct from the morphology of jellyfish found in lagoons (**Chapter 3**), with distinct coloration and loss of terminal clubs. One lake (Papua 4) held a dense population of jellyfish displaying the 'lake' ecotype up until May 2017 (**Chapter 3**). Then, there was a crash of the population, going from an estimated half million in May 2017 to zero to a few individuals in November 2017. Not half a year later (May 2018) the lake was populated again with jellyfish, which showed a significant shift in morphology more towards the 'lagoon' ecotype with blue coloration and long terminal clubs. Remarkably, we saw no change in genetic markers, implying that the morphological shift might have been purely phenotypic plasticity. Recently, the same phenomenon was observed for a jellyfish population in Palau (Swift and Dawson, 2020). I suggest that current differences in jellyfish morphology are the consequence of phenotypic plasticity instead of genetic adaptation. However, over time phenotypic plasticity may facilitate local adaptation reflected in genetic signatures (Miner et al., 2005; Fierst, 2011; Valladares et al., 2014).

In the mussel, I studied shell morphology, microbial community structure and trophic niche space (**Chapter 6**). Morphological variation in shell shape was not related to genetic divergence. Furthermore, I observed high within-population variation of shell outline morphologies, indicating that even within the same lake shell shape can be different between individuals. Discrepancy between genetic and morphological divergence and high within-lake variation indicates a strong influence of phenotypic plasticity in shell shape. Still, a significant influence of the degree of connection to the surrounding sea on shell shape was found. Ecologically, variation in shell shape is likely to be related to mussel bed density and predation (Reimer and Tedegren, 1996; Lauzon-Guay et al., 2005), factors that indirectly may be affected by the degree of connection the sea. For instance, a higher diversity of predators can be expected in more connected lakes, potentially resulting in differently shaped shells.

Similar to shell shape, also no concordance between genetic variation and shell thickness was found (**Chapter 6**). Interestingly, I found no correlations of shell thickness to pH, a pattern commonly found for calcifying organisms in experimental and field studies, where a lower pH results in the dissolution of shell material or a lessened ability to incorporate calcium carbonate (Melatunan et al., 2013; Fitzer et al., 2015; Manno et al., 2017; Mekkes et al., 2021). It is unlikely that the range in pH found for marine lakes in this thesis (**Chapter 2**) is caused by altered carbon chemistry of the lakes as is the case for ocean acidification through CO_2. Instead, it could be due to bacterial decomposition of (mangrove) leaf litter resulting from a higher terrestrial input (Middelburg et al., 1996; Alongi et al., 1998). As carbon chemistry and alkalinity may not have been affected, perhaps marine lake mussels could still incorporate calcium carbonate into their shells. The high terrestrial input of isolated may represent abundant food sources for mussels (corroborated by a shift in the trophic niche towards terrestrial sources in isolated lakes), which may provide the energy needed to compensate for low pH environments (Osinga et al., 2011). Mussel shells were observed to be the thinnest in highly isolated lakes (Papua 19, 21 and 22) (**Chapter 6**). These lakes show local water quality regimes most different from the surrounding sea, namely extremely high water temperatures (35.4-35.9°C) and low salinities (16.4-23.9 ppt). Although *Brachidontes* mussels are known for their robustness towards large ranges in temperature and salinity (Sarà et al., 2008, 2013), perhaps the conditions in these lakes affect energy budgeting (Grant and Bacher, 1998), leading to mussels having to spend more energy on keeping homeostasis compared to shell growth. Energy reallocation towards growth instead of calcification has also been observed for other marine molluscs (pteropods) (Mekkes et al., 2021).

Even though I quantified multiple mussel characteristics to represent adaptation to the local environment (morphology, microbial communities and trophic niche space), I did not find concordance between phenotypic and genomic divergence (**Chapter 6**). Therefore, no compelling evidence was found for ecological speciation to underlie population differentiation, and instead phenotypic plasticity and acclimatization may underlie the current

morphological and ecological variation in marine lakes (Miner et al., 2005; Fierst, 2011). Instead, neutral limitations to dispersal in combination with founder events appear to play a more important role in generating current population differentiation in marine lake invertebrates, potentially aided by eco-evolutionary dynamics of priority effects and monopolization. Historical eco-evolutionary dynamics are often neglected in population genomic studies yet may have consequences for population connectivity and subsequently may have to be considered when applying conservation management.

Marine connectivity and conservation

Discontinuity in the marine realm may be more prevalent than previously thought (Oleksiak, 2018). Be it via gradients in temperature or salinity, via oceanographic currents, or via physical, although permeable, boundaries such as those presented in marine lakes. Even for marine zooplankton in the open ocean, patterns of genetic structure and abilities of rapid adaptive responses have been suggested (Peijnenburg and Goetze, 2013). Genomic DNA sequencing approaches provide the resolution necessary to elucidate small scale population structure, allow for unprecedented insights into migration and demography, and are unveiling signatures of local adaptation (Oleksiak and Rajora, 2020b). Conservation genetics investigates patterns of connectivity and demographic changes such as bottlenecks, and aims to preserve genetic diversity and consequently the evolutionary potential of populations (Frankham et al., 2002). Information on population connectivity and whether populations are sources or sinks of genetic diversity can aid in the designation of Marine Protected Areas (MPAs) (Carr et al., 2017; Xuereb et al., 2019). Furthermore, with global change, it also becomes more and more important to preserve the adaptive potential of marine species in order to ensure persistence.

Seascape genomic work is revealing population structure and gene flow even on small spatial scales (see references within Xuereb et al., 2019), also including adaptive differentiation to preserve evolutionary potential. However, most studies focus on effects gene flow by gradients in environmental factors and/or ocean currents. What is lacking in these studies, and what is brought forward by my thesis is a historical component of founder events and early colonizer advantages that may affect gene flow (**Chapter 4-6**). In addition, even a slight decrease in the permeability of barriers to dispersal (colonization) may have effects that persist over thousands of generations. Furthermore, anthropogenic habitat fragmentation and global change may exacerbate these historical contingency effects (Legrand et al., 2017; Oleksiak, 2018). Such historical eco-evolutionary dynamics are often neglected in studies of population genomics, yet these dynamics may especially have persistent effects in peripheral habitats.

Within the context of peripheral marine habitats, this thesis shows strong population structure for three invertebrate species in marine lakes, and each lake harbors unique genetic diversity (**Chapter 3-6**). From a conservation perspective, **Chapter 3** argued that the lakes containing jellyfish should each be treated as a unique management unit. My

results from the other chapters support the notion that this recommendation can be extended to all marine lakes, especially to the most isolated lakes (**Chapter 4-6**). Even though we see reductions in genetic diversity in marine lake populations compared to populations in the surrounding sea, we also see unique SNPs in isolated lakes (**Chapter 6**). As the most isolated lakes both show the highest population differentiation (**Chapter 4-6**) and the most distinct local environments compared to the surrounding ocean (**Chapter 2**), this indicates that adaptation to local abiotic and biotic conditions can play a role in these lakes. Temperature and salinity regimes in the isolated lakes represent worst-case scenarios predicted for the year 2100 by the IPCC for the world's oceans (IPCC, 2014). The potential local adaptations currently arising in the marine lakes may be thus of importance in the persistence of marine populations in the future via genetic rescue.

Conclusions and future perspectives

This thesis aimed to better understand small scale population genomic patterns of different invertebrates found in marine lakes in the highly diverse area in the center of the Coral Triangle (Mangubhai et al., 2012; Starger et al., 2015). The major finding of this thesis is that historical colonization of peripheral habitats such as marine lakes may have long-lasting effects on population differentiation, particularly with decreasing connection to the surrounding ocean and potentially associated distinct local environments. As seawater levels have fluctuated over history, partially isolated habitats such as marine lakes likely have been present throughout these changes (even if not in their current positions) (Tomascik and Mah, 1994; Dawson et al., 2009). Since we observed distinct genetic patterns in marine lake populations of a maximum age of around 12,000 years (Sathiamurthy and Voris, 2006; Dawson, 2016), the short term isolation in these peripheral habitats may have contributed to the high marine biodiversity found in the area (Hoeksema, 2007).

The delineation of Marine Protected Areas is informed by population connectivity and adaptive potential (Carr et al., 2017; Xuereb et al., 2019). Advances in genomic sequencing strategies have allowed marine population genomic studies to investigate patterns on small spatial scales and assess effects of SNPs under putative selection (Oleksiak and Rajora, 2020b). My findings encourage the incorporation of historical eco-evolutionary dynamics in defining marine connectivity, particularly in a time of global change and habitat fragmentation (Palsbøll et al., 2007; Allendorf et al., 2010; Legrand et al., 2017). Yet, there are still many avenues for future research to further elucidate the impacts of eco-evolutionary dynamics on population genomic structure in peripheral habitats. Below, I highlight a few for marine lakes in particular.

I would recommend using the opportunities provided by marine lakes to tackle remaining questions regarding the functional genetic and phenotypic basis of local adaptation. By representing replicated and independent models of ecology and evolution, pairwise comparisons can be made between low-connected and high-connected lakes to further the search for loci underlying selection. Taking inspiration from a similar study system, the

threespine stickleback (Jones et al., 2012; Roesti et al., 2014), whole genome sequencing should be applied to obtain more in-depth insights into the divergence and adaptation of invertebrates inhabiting marine lakes (Ekblom and Wolf, 2014). The current thesis showed clear per-lake clustering of populations, justifying the selection of a few individuals per lake to allow for the sequencing of the entire genome. Assembling and annotating a high-quality draft genome will provide the resource needed to obtain detailed maps of genome architecture and gene content. In turn, this will allow a much more detailed investigation into the functionality and structural genetic variants found in populations in marine lakes. In combination with genomic resources (such as whole genomes and/or transcriptomes), including a (phenotypic) fitness component can further improve research into local adaptation in the marine realm. Physiological and/or morphological adaptations can be assessed by common garden and/or reciprocal transplant experiments. Such experiments allow for further disentanglement of phenotypic plasticity versus genetic adaptation. I would suggest extending research to fish, for example the goby species that are found in most marine lakes (**Chapter 2**). Fish provide more opportunities to link genetic variation to phenotypic variation, as they show more behavioral variation than the marine lake invertebrates studied in this thesis. For fish, changes in morphology and behavior may also be exacerbated via sexual selection, further linking environment to genetics.

Next, more explicitly defining two fundamental aspects of marine lakes will aid the investigation of eco-evolutionary dynamics. Firstly, setting up long-term regular or continuous water quality monitoring in at least one, but preferably several marine lakes will give insight into the stability of the lakes. Marine lakes have classically been assumed to be stable in water quality conditions over time (Hamner et al., 1982), but regular monitoring in a marine lake has only been done in the Jellyfish Lake in Palau so far (Cimino et al., 2018). Taking advantage of the tourism interest in the lake in Raja Ampat also known to harbor multitudes of *Mastigias papua* (Papua 4 in this study), perhaps this is a good lake to start. Monitoring water quality here could then be related to the dramatic shifts in jellyfish abundance to better understand their boom-bust dynamics. Secondly, the pressure of incoming propagules should be more exactly defined. In this thesis I used a proxy for connection to the surrounding sea to estimate how likely it is for new colonizers to enter the lake. However, it would be more informative if for example through plankton nets at subterranean entrances to the lake the influx of propagules could be quantified. Of course, first the actual entrances of the lakes need to be specifically defined. Bathymetric measurements using sonar could be used to map the bottom of the lake to find potential caves. Hamner and Hamner (1998) also used sodium flueorescein dye to search for where water emerges in the lake, a relatively simple method that could be repeated for marine lakes in Indonesia. More accurately defining propagule influx will give a more accurate indication of how isolated the marine lakes actually are.

Finally, from a conservation genomic angle, marine lakes also provide opportunities to further investigate the effects of fragmentation of habitats in the sea. Fragmentation affects

all levels of diversity, from genetic to ecosystem diversity (Legrand et al., 2017). The effects of fragmentation mainly come from the increased isolation and the reduction in habitat size. Marine lakes represent fragmented habitats that arise *de novo* instead of the splitting into fragments of existing habitat, and thus aspects of colonization are more important than survival of resident species after the split. Populations and communities in the marine lakes provide opportunities to assess how ecological interactions and evolutionary forces (together defined as eco-evolutionary dynamics) are modified by the fragmentation. A short-term equivalent to *de novo* fragmented habitats in the marine realm are artificial reefs (Baine, 2001). The placement of artificial reefs is getting more and more attention as a tool to assist coral reef persistence. For artificial reefs, it is known that they provide opportunities for invasive species to colonize the new structures rapidly, leading to unexpected and unintended outcomes of the placement of these structures (Glasby et al., 2007; Sheehy and Vik, 2010; Dafforn et al., 2012). Marine lakes could provide a long-term natural perspective to artificial reefs to investigate lasting effects and successional stages of colonizing species. Beyond invasive species, artificial reefs and marine lakes can provide insights into the connectivity between fragments, with important implications for population persistence. Together, the short- and long-term studies of marine fragmentation could inform marine conservation strategies.

This thesis takes a first step in elucidating population genomic patterns of marine lake invertebrates. It paves the way for future studies to assess how insular systems can inform fundamental questions underlying generation and maintenance of biodiversity, and how they can function as models to assess conservation genetics and conservation.

References

Agostini, V.N., Grantham, H.S., et al. (2012). Achieving Fisheries and Conservation Objectives within Marine Protected Areas: Zoning the Raja Ampat Network.
Allendorf, F.W., Hohenlohe, P.A., et al. (2010). Genomics and the future of conservation genetics. Nat. Rev. Genet. *11*, 697–709.
Almén, M.S., Lamichhaney, S., et al. (2016). Adaptive radiation of Darwin's finches revisited using whole genome sequencing. BioEssays *38*, 14–20.
Alongi, D.M., Sasekumar, A., et al. (1998). The influence of stand age on benthic decomposition and recycling of organic matter in managed mangrove forests of Malaysia. J. Exp. Mar. Bio. Ecol. *225*, 197–218.
Altschul, S.F., Gish, W., et al. (1990). Basic Local Alignment Search Tool. J. Mol. Biol. *215*, 403–410.
Altschul, S.F., Madden, T.L., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.
Andrews, K.R., Good, J.M., et al. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. Nat. Rev. Genet. *17*, 81–92.
Arai, Y., Gotoh, R.O., et al. (2017). Phylogenetic relationships and morphological variations of upside-down jellyfishes, Cassiopea spp. inhabiting Palau Islands. Biogeography *19*, 133–141.
Armsworth, P.R. (2002). Recruitment limitation, population regulation, and larval connectivity in reef fish metapopulations. Ecology *83*, 1092–1104.
Arnott, S. (2009). Lakes, as islands. In Encyclopedia of Islands, (University of California Press), pp. 526–531.
Asaad, I., Lundquist, C.J., et al. (2018). Delineating priority areas for marine biodiversity conservation in the Coral Triangle. Biol. Conserv. *222*, 198–211.
Avise, J.C. (2000). Phylogeography: the history and formation of species. (Harvard University Press).
Avise, J.C. (2004). Molecular Markers, Natural History, and Evolution (Sunderland, MA.: Sinauer Associates, Inc.).
Avise, J.C. (2009). Phylogeography: Retrospect and prospect. J. Biogeogr. *36*, 3–15.
Ayling, A.L. (1980). Patterns of sexuality, asexual reproduction and recruitment in some subtital marine Demospongiae. Biol. Bull. *158*, 271–282.
Azzini, F., Calcinai, B., et al. (2007). Sponges of the marine karst lakes and of the coast of the islands of Ha Long Bay (North Vietnam). In Porifera Research: Biodiversity, Innovation and Sustainability, pp. 157–164.
Baine, M. (2001). Artificial reefs: a review of their design, application, management and performance. Ocean Coast. Manag. *44*, 19.
Baird, N.A., Etter, P.D., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One *3*, 1–7.
De Bakker, D.M., Meesters, E.H.W.G., et al. (2016). Population genetic structure, abundance, and health status of two dominant benthic species in the saba bank national park, caribbean netherlands:Montastraea cavernosa and xestospongia muta. PLoS One *11*, 1–20.
Barber, P.H. (2009). The challenge of understanding the Coral Triangle biodiversity hotspot. J. Biogeogr. *36*, 1845–1846.
Barber, P.H., Erdmann, M.V., et al. (2006). Comparative Phylogeography of Three Codistributed Stomatopods : Origins and Timing of Regional Lineage Diversification in the Coral Triangle. Evolution (N. Y). *60*, 1825–1839.
Barber, P.H., Cheng, S.H., et al. (2011). Evolution and conservation of marine biodiversity in the Coral Triangle: insights from stomatopod Crustacea. Crustac. Issues *19*, 129–156.
Barton, N.H., and Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. Annu. Rev. Ecol. Syst. *15*, 133–164.
Bauzà-Ribot, M.M., Jaume, D., et al. (2011). Islands beneath islands: Phylogeography of a groundwater amphipod crustacean in the Balearic archipelago. BMC Evol. Biol. *11*.
Bayha, K.M., and Graham, W.M. (2011). First confirmed reports of the rhizostome jellyfish Mastigias (Cnidaria: Rhizostomeae) in the Atlantic basin. Aquat. Invasions *6*, 361–366.
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., and Lim, S. (2009). A new Suberites (Demospongiae: Hadromerida: Suberitidae) from the tropical Indo-West Pacific. Zool. Meded. *83*, 853–862.
Becking, L.E., Renema, W., et al. (2011). Recently discovered landlocked basins in Indonesia reveal high habitat diversity in anchialine systems. Hydrobiologia *677*, 89–105.
Becking, L.E., Cleary, D.F.R.R., et al. (2013a). Sponge species composition, abundance, and cover in marine lakes and coastal mangroves in Berau, Indonesia. Mar. Ecol. Prog. Ser. *481*, 105–120.
Becking, L.E., Erpenbeck, D., et al. (2013b). Phylogeography of the Sponge Suberites diversicolor in Indonesia: Insights into the Evolution of Marine Lake Populations. PLoS One *8*.
Becking, L.E., de Leeuw, C.A., et al. (2015). Newly discovered "jellyfish lakes" in Misool, Raja Ampat, Papua, Indonesia. Mar. Biodivers. *45*, 597–598.
Becking, L.E., de Leeuw, C.A., et al. (2016). Highly divergent mussel lineages in isolated Indonesian marine lakes. PeerJ *4*, e2496.

Bell, J.J. (2008). The functional roles of marine sponges. Estuar. Coast. Shelf Sci. *79*, 341–353.
Bell, J.J., Davy, S.K., et al. (2013). Could some coral reefs become sponge reefs as our climate changes? Glob. Chang. Biol. *19*, 2613–2624.
Bell, L.J., Ucharm, G., et al. (2017). Sunscreen pollution analysis in Jellyfish Lake coral reef research foundation Palau - Coral Reef Research Foundation.
Belyea, L.R., and Lancaster, J. (1999). Assembly rules within a contingent ecology. Oikos *86*, 402–416.
Benestan, L., Quinn, B.K., et al. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (Homarus americanus). Mol. Ecol. *25*, 5073–5092.
Bentlage, B., and Wörheide, G. (2007). Low genetic structuring among Pericharax heteroraphis (Porifera: Calcarea) populations from the Great Barrier Reef (Australia), revealed by analysis of nrDNA and nuclear intron sequences. Coral Reefs *26*, 807–816.
Bernatchez, S., Xuereb, A., et al. (2018). Seascape genomics of eastern oyster (Crassostrea virginica) along the Atlantic coast of Canada. Evol. Appl. *12*, 587–609.
Berner, D., and Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. Trends Genet. *31*, 491–499.
Berner, D., Grandchamp, A.C., et al. (2009). Variable progress toward ecological speciation in parapatry: Stickleback across eight lake-stream transitions. Evolution (N. Y). *63*, 1740–1753.
Bi, K., Linderoth, T., et al. (2013). Unlocking the vault: Next-generation museum population genomics. Mol. Ecol. *22*, 6018–6032.
Bi, K., Linderoth, T., et al. (2019). Temporal genomic contrasts reveal heterogeneous evolutionary responses within and among montane chipmunk species during recent climate change. PLoS Genet. *15*, e1008119.
Bierne, N., Bonhomme, F., et al. (2003). Habitat preference and the marine-speciation paradox. Proc. R. Soc. London B Biol. Sci. *270*, 1399–1406.
Bierne, N., Bonhomme, F., et al. (2016). Dedicated population genomics for the silent world: The specific questions of marine population genetics. Curr. Zool. *62*, 545–550.
Blanquer, A., and Uriz, M.J. (2010). Population genetics at three spatial scales of a rare sponge living in fragmented habitats. BMC Evol. Biol. *10*.
Bolger, A.M., Lohse, M., et al. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
Bookstein, F.L. (1989). Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. IEEE Trans. Pattern Anal. Mach. Intell. *11*, 567–585.
Bookstein, F.L. (1996). Landmark methods for forms without landmarks: localizing group differences in outline shape. Proc. Work. Math. Methods Biomed. Image Anal. *1*, 279–289.
Botello, A., and Alvarez, F. (2010). Genetic variation in the stygobitic shrimp Creaseria morleyi (Decapoda: Palaemonidae), evidence of bottlenecks and re-invasions in the Yucatan Peninsula. Biol. J. Linn. Soc. *99*, 315–325.
Bowen, B.W., Rocha, L.A., et al. (2013). The origins of tropical marine biodiversity. Trends Ecol. Evol. *28*, 359–366.
Bradbury, I.R., Hamilton, L.C., et al. (2015). Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. Mol. Ecol. *24*, 5130–5144.
Brawand, D., Wagner, C.E., et al. (2015). The genomic substrate for adaptive radiation in African cichlid fish. Nature *513*, 375–381.
Brown, J.H., and Kodric-Brown, A. (1977). Turnover rates in insular biogeography: effect of immigration on extinction. Ecology *58*, 445–449.
Brown, R.R., Davis, C.S., et al. (2014). SNP discovery in a reef-forming glass sponge, Aphrocallistes vastus, using the Ion Torrent next generation sequencing platform. Conserv. Genet. Resour. *6*, 49–51.
Brown, R.R., Davis, C.S., et al. (2017). Clones or clans: the genetic structure of a deep-sea sponge, Aphrocallistes vastus, in unique sponge reefs of British Columbia, Canada. Mol. Ecol. *26*, 1045–1059.
Cabezas, P., Alda, F., et al. (2012). Genetic characterization of the endangered and endemic anchialine squat lobster Munidopsis polymorpha from Lanzarote (Canary Islands): management implications. ICES J. Mar. Sci. *69*, 1030–1037.
Cadotte, M.W., and Tucker, C.M. (2017). Should Environmental Filtering be Abandoned? Trends Ecol. Evol. *32*, 429–437.
Calderón-Gutiérrez, F., Sánchez-Ortiz, C.A., et al. (2018). Ecological patterns in anchialine caves. PLoS One *13*, 1–19.
Cammen, K.M., Andrews, K.R., et al. (2016). Genomic methods take the plunge: Recent advances in high-throughput sequencing of marine mammals. J. Hered. *107*, 481–495.
Carpenter, K.E., Barber, P.H., et al. (2011). Comparative phylogeography of the coral triangle and implications for marine management. J. Mar. Biol. *2011*.
Carr, M.H., Neigel, J.E., et al. (2003). Comparing Marine and Terrestrial Ecosystems : Implications for the Design of Coastal Marine Reserves. Ecol. Appl. 90–130.
Carr, M.H., Robinson, S.P., et al. (2017). The central importance of ecological spatial connectivity to effective

coastal marine protected areas and to meeting the challenges of climate change in the marine environment. Aquat. Conserv. Mar. Freshw. Ecosyst. *27*, 6–29.
Catchen, J.M., Hohenlohe, P.A., et al. (2013). Stacks: An analysis tool set for population genomics. Mol. Ecol. *22*, 3124–3140.
Catchen, J.M., Hohenlohe, P.A., et al. (2017). Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. Mol. Ecol. Resour. *17*, 362–365.
Cerrano, C., Azzini, F., et al. (2006). Marine lakes of karst islands in Ha Long Bay (Vietnam). Chem. Ecol. *22*, 489–500.
Charlesworth, B. (2010). Molecular population genomics: A short history. Genet. Res. (Camb). *92*, 397–411.
Chaves-Fonnegra, A., Feldheim, K.A., et al. (2015). Population structure and dispersal of the coral-excavating sponge Cliona delitrix. Mol. Ecol. *24*, 1447–1466.
Chi, X., Mueller-Navarra, D.C., et al. (2019). Food quality matters: Interplay among food quality, food quantity and temperature affecting life history traits of Aurelia aurita (Cnidaria: Scyphozoa) polyps. Sci. Total Environ. *656*, 1280–1288.
Cimino, M.A., Patris, S., et al. (2018). Jellyfish distribution and abundance in relation to the physical habitat of Jellyfish Lake, Palau. J. Trop. Ecol. *34*, 17–31.
Cleary, D.F.R., Becking, L.E., et al. (2013). Habitat- and host-related variation in sponge bacterial symbiont communities in Indonesian waters. FEMS Microbiol. Ecol. *85*, 465–482.
Collins, A.G. (2002). Phylogeny of the Medusozoa and the evolution of cnidarian life cycles. J. Evol. Biol. *15*, 418–432.
Cornwell, B.H., Fisher, J.L., et al. (2016). Chaotic genetic patchiness without sweepstakes reproduction in the shore crab Hemigrapsus oregonensis. Mar. Ecol. Prog. Ser. *548*, 139–152.
Costello, M.J., and Chaudhary, C. (2017). Marine Biodiversity, Biogeography, Deep-Sea Gradients, and Conservation. Curr. Biol. *27*, R511–R527.
Cowen, R.K., Lwiza, K.M.M., et al. (2000). Connectivity Populations: Open or Closed? Science (80-.). *287*, 857–859.
Coyne, J.A., and Orr, H.A. (2004). Speciation (Sunderland, MA.: Sinauer Associates, Inc.).
Craft, J.D., Russ, A.D., et al. (2008). Islands under islands: The phylogeography and evolution of Halocaridina rubra Holthuis, 1963 (Crustacean: Decapoda: Atyidae) in the Hawaiian archipelago. Limnol. Oceanogr. *53*, 675–689.
Crandall, E.D., Jones, M.E., et al. (2008). Comparative phylogeography of two seastars and their ectosymbionts within the Coral Triangle. Mol. Ecol. *17*, 5276–5290.
Crawford, D.L., and Oleksiak, M.F. (2016). Ecological population genomics in the marine environment. Brief. Funct. Genomics *15*, 342–351.
Cruickshank, T.E., and Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. *23*, 3133–3157.
D'Aloia, C.C., Andrés, J.A., et al. (2020). Unraveling hierarchical genetic structure in a marine metapopulation: A comparison of three high-throughput genotyping approaches. Mol. Ecol. *29*, 2189–2203.
Dafforn, K.A., Glasby, T.M., et al. (2012). Comparing the invasibility of experimental "reefs" with field observations of natural reefs and artificial structures. PLoS One *7*, 1–16.
Darriba, D., Taboada, G.L., et al. (2012). jModelTest 2: more models, new heuristics and high-performance computing Europe PMC Funders Group. Nat. Methods *9*, 772.
Darwin, C. (1872). The origin of species.
Dawson, M.N. (2005a). Morphological variation and systematics in the Scyphozoa: Mastigias (Rhizostomeae, Mastigiidae) - A golden unstandard? Hydrobiologia *537*, 185–206.
Dawson, M.N. (2005b). Five new subspecies of Mastigias (Scyphozoa: Rhizostomeae: Mastigiidae) from marine lakes, Palau, Micronesia. J. Mar. Biol. Assoc. UK *85*, 679–694.
Dawson, M.N. (2016). Island and island-like marine environments. Glob. Ecol. Biogeogr. *25*, 831–846.
Dawson, M.N., and Hamner, W.M. (2005). Rapid evolutionary radiation of marine zooplankton in peripheral environments. Proc. Natl. Acad. Sci. U. S. A. *102*, 9235–9240.
Dawson, M.N., Martin, L.E., et al. (2001). Jellyfish swarms, tourists, and the Christ-child. Hydrobiologia *451*, 131–144.
Dawson, M.N., Martin, L.E., et al. (2009). Marine Lakes. In Encyclopedia of Islands, R. Gillespie, and D.A. Clague, eds. (University of California Press), pp. 603–607.
Dawson, M.N., Cieciel, K., et al. (2015). Population-level perspectives on global change: genetic and demographic analyses indicate various scales, timing, and causes of scyphozoan jellyfish blooms. Biol. Invasions *17*, 851–867.
DeBiasse, M.B., Nelson, B.J., et al. (2014). Evaluating summary statistics used to test for incomplete lineage sorting: Mito-nuclear discordance in the reef sponge Callyspongia vaginalis. Mol. Ecol. *23*, 225–238.
DeBoer, T.S., Subia, M.D., et al. (2008). Phylogeography and limited genetic connectivity in the endangered boring giant clam across the coral triangle. Conserv. Biol. *22*, 1255–1266.
DKP Dinas Kelautan dan Perikanan Provinsi Papua Barat. (2018). Rencana Pengelolaan dan Zonasi Kawasan Konservasi Taman Pesisir Teluk Berau dan Taman Pesisir Teluk Nusalasi-Van Den Bosch. (Manokwari).
Dowd, W.W., and Somero, G.N. (2012). Behavior and survival of Mytilus congeners following episodes of

elevated body temperature in air and seawater. J. Exp. Biol. 502–514.
Dowling, D.K., Friberg, U., et al. (2008). Evolutionary implications of non-neutral mitochondrial genetic variation. Trends Ecol. Evol. *23*, 546–554.
Dray, S., Legendre, P., et al. (2006). Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Modell. *196*, 483–493.
Duran, S., Pascual, M., et al. (2004a). Low levels of genetic variation in mtDNA sequences over the western Mediterranean and Atlantic range of the sponge Crambe crambe (Poecilosclerida). Mar. Biol. *144*, 31–35.
Duran, S., Pascual, M., et al. (2004b). Strong population structure in the marine sponge Crambe crambe (Poecilosclerida) as revealed by microsatellite markers. Mol. Ecol. *13*, 511–522.
Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.
Ekblom, R., and Wolf, J.B.W. (2014). A field guide to whole-genome sequencing, assembly and annotation. Evol. Appl. *7*, 1026–1042.
Ekins, M., Erpenbeck, D., et al. (2016). Staying well connected-Lithistid sponges on seamounts. J. Mar. Biol. Assoc. United Kingdom *96*, 437–451.
Evanno, G., Regnaut, S., et al. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. *14*, 2611–2620.
Excoffier, L., and Foll, M. (2011). fastsimcoal: A continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics *27*, 1332–1334.
Excoffier, L., and Lischer, H.E.L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. *10*, 564–567.
Excoffier, L., Smouse, P.E., et al. (1992). Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. Genetics *131*, 479–491.
Faria, R., Renaut, S., et al. (2014). Advances in Ecological Speciation: An integrative approach. Mol. Ecol. *23*, 513–521.
Fattorini, S., Mantoni, C., et al. (2018). Island biogeography of insect conservation in urban green spaces. Environ. Conserv. *45*, 1–10.
Fauvelot, C., Bernardi, G., et al. (2003). Reductions in the mitochondrial DNA diversity of coral reef fish provide evidence of population bottlenecks resulting from Holocene sea-level change. Evolution *57*, 1571–1583.
Feder, J.L., Egan, S.P., et al. (2012). The genomics of speciation-with-gene-flow. Trends Genet. *28*, 342–350.
Fernandez-Palacios, J.M., Kueffer, C., et al. (2015). A new golden era in island biogeography. Front. Biogeogr. *7*.
Fierst, J.L. (2011). A history of phenotypic plasticity accelerates adaptation to a new environment. J. Evol. Biol. *24*, 1992–2001.
Fitzer, S.C., Vittert, L., et al. (2015). Ocean acidification and temperature increase impact mussel shell shape and thickness: Problematic for protection? Ecol. Evol. *5*, 4875–4884.
Foll, M., and Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics *180*, 977–993.
Franchini, P., Monné Parera, D., et al. (2017). quaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. Mol. Ecol. *26*, 2783–2795.
Frankham, R., Briscoe, D.A., et al. (2002). Introduction to conservation genetics (Cambridge University Press).
Fry, B. (2006). Stable isotope ecology (New York: Springer).
Fu, L., Niu, B., et al. (2012). CD-HIT: Accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.
Fukami, T. (2015). Historical Contingency in Community Assembly: Integrating Niches, Species Pools, and Priority Effects. Annu. Rev. Ecol. Evol. Syst. *46*, 1–23.
Fumagalli, M., Vieira, F.G., et al. (2014). ngsTools : methods for population genetics analyses from Next-Generation Sequencing data © The Authors 2014 . Published by Oxford University Press on behalf of the Japanese. Bioinformatics 1–2.
Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics *31*, 3718–3720.
Gaman, A.G., Nebore, A., et al. (2012). Rencana pengelolaan taman pulau-pulau kecil daerah Raja Ampat.
Gazeau, F., Quiblier, C., et al. (2007). Impact of elevated CO2 on shellfish calcification. Geophys. Res. Lett. *34*, 1–5.
Geller, J., Meyer, C., et al. (2013). Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. Mol. Ecol. Resour. *13*, 851–861.
Genner, M.J., Nichols, P., et al. (2007). Evolution of a cichlid fish in a Lake Malawi satellite lake. Proc. Biol. Sci. *274*, 2249–2257.
Gerovasileiou, V., Martínez, A., et al. (2016). World Register of marine Cave Species (WoRCS): a new Thematic Species Database for marine and anchialine cave biodiversity. Res. Ideas Outcomes *2*, e10451.
Gibson, L.A., Cowan, M.A., et al. (2017). Island refuges: Conservation significance of the biodiversity patterns resulting from 'natural' fragmentation. Biol. Conserv. *212*, 349–356.
Giles, E.C., Saenz-Agudelo, P., et al. (2015). Exploring seascape genetics and kinship in the reef sponge Stylissa carteri in the Red Sea. Ecol. Evol. *5*, 2487–2502.

Gillespie, R. (2004). Community assembly through adaptive radiation in Hawaiian spiders. Science (80-.). *303*, 356–359.
Gillespie, R.G. (2016). Island time and the interplay between ecology and evolution in species diversification. Evol. Appl. *9*, 53–73.
Gillespie, R.G., and Clague, D.A. (2009). Encyclopedia of islands (University of California Press).
Glasby, T.M., Connell, S.D., et al. (2007). Nonindigenous biota on artificial structures: Could habitat creation facilitate biological invasions? Mar. Biol. *151*, 887–895.
De Goeij, J.M., Van Oevelen, D., et al. (2013). Surviving in a marine desert: The sponge loop retains resources within coral reefs. Science (80-.). *342*, 108–110.
Gonzalez-Zapata, F.L., Bongaerts, P., et al. (2018). Holobiont diversity in a reef-building coral over its entire depth range in the mesophotic zone. Front. Mar. Sci. *5*, 1–13.
Gonzalez, B.C., Martínez, A., et al. (2017). Genetic spatial structure of an anchialine cave annelid indicates connectivity within - but not between - islands of the Great Bahama Bank. Mol. Phylogenet. Evol. *109*, 259–270.
Goto, T. V, Tamate, H.B., et al. (2011). Phylogenetic characterization of three morphs of mussels (Bivalvia, Mytilidae) inhabiting isolated marine environments in Palau Islands. Zoolog. Sci. *28*, 568–579.
Gotoh, R.O., Sekimoto, H., et al. (2009). Peripatric differentiation among adjacent marine lake and lagoon populations of a coastal fish, Sphaeramia orbicularis (Apogonidae, Perciformes, Teleostei). Genes Genet. Syst. *84*, 287–295.
Gotoh, R.O., Chiba, S.N., et al. (2011). Population genetic structure of the striped silverside, Atherinomorus endrachtensis (Atherinidae, Atheriniformes, Teleostei), inhabiting marine lakes and adjacent lagoons in Palau: marine lakes are “Islands” for marine species. Genes Genet. Syst. *86*, 325–337.
Graham, N.R., Gruner, D.S., et al. (2017). Island ecology and evolution: Challenges in the Anthropocene. Environ. Conserv. *44*, 323–335.
Grant, J., and Bacher, C. (1998). Comparative models of mussel bioenergetics and their validation at field culture sites. J. Exp. Mar. Bio. Ecol. *219*, 21–44.
Grantham, H.S., Agostini, V.N., et al. (2013). A comparison of zoning analyses to inform the planning of a marine protected area network in Raja Ampat , Indonesia. Mar. Policy *38*, 184–194.
Guardiola, M., Frotscher, J., et al. (2016). High genetic diversity, phenotypic plasticity, and invasive potential of a recently introduced calcareous sponge, fast spreading across the Atlanto-Mediterranean basin. Mar. Biol. *163*, 1–16.
Gutiérrez, J.L., Jones, C.G., et al. (2003). Mollusks as ecosystem engineers: The role of shell production in aquatic habitats. Oikos *101*, 79–90.
Hachich, N.F., Bonsall, M.B., et al. (2015). Island biogeography: Patterns of marine shallow-water organisms in the Atlantic Ocean. J. Biogeogr. *42*, 1871–1882.
Haddad, N.M., Brudvig, L.A., et al. (2015). Habitat fragmentation and its lasting impact on Earth’s ecosystems. Sci. Adv. *1*, e1500052–e1500052.
Haig, J.A., Connolly, R.M., et al. (2010). Little shrimp left on the shelf: The roles that sea-level change, ocean currents and continental shelf width play in the genetic connectivity of a seagrass-associated species. J. Biogeogr. *37*, 1570–1583.
Halanych, K.M., and Kocot, K.M. (2017). Shellfish genes. Nat. Ecol. Evol. *1*, 1–2.
Hamner, W.M. (1982). Strange world of Palau’s salt lakes. Natl. Geogr. Mag. *161*, 264–282.
Hamner, W., and Hamner, P. (1998). Stratified Marine Lakes of Palau (Western Caroline Islands). Phys. Geogr. *19*, 175–220.
Hamner, W.M., Gilmer, R.W., et al. (1982). The physical, chemical, and biological characteristics of a stratified, saline, sulfide lake in Palau. Limnol. Oceanogr. *27*, 896–909.
Hanzawa, N., Sekimoto, H., et al. (2012). Genetic Diversity and Evolution of Marine Animals Isolated in Marine Lakes. INTECH Open Access Publ.
Harrang, E., Lapègue, S., et al. (2013). A high load of non-neutral amino-acid polymorphisms explains high protein diversity despite moderate effective population size in a marine bivalve with sweepstakes reproduction. G3 Genes, Genomes, Genet. *3*, 333–341.
Hasegawa, M., Kishino, H., et al. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. *22*, 160–174.
Hauser, L., and Carvalho, G.R. (2008). Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. Fish Fish. *9*, 333–362.
Hazzouri, K.M., Flowers, J.M., et al. (2015). Whole genome re-sequencing of date palms yields insights into diversification of a fruit tree crop. Nat. Commun. *6*, 8824.
Hedgecock, D. (1986). Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? Bull. Mar. Sci. *39*, 550–564.
Hedgecock, D., Gaffney, P.M., et al. (2005). The case for sequencing the pacific oyster genome. J. Shellfish Res. *24*, 429–441.
Helmus, M.R., Mahler, D.L., et al. (2014). Island biogeography of the Anthropocene. Nature *513*, 543–546.
Hemmer-Hansen, J., Therkildsen, N.O., et al. (2014). Conserving marine biodiversity: Insights from life-history trait candidate genes in Atlantic cod (Gadus morhua). Conserv. Genet. *15*, 213–228.

Hendry, A.P., Bolnick, D.I., et al. (2009). Along the speciation continuum in sticklebacks. J. Fish Biol. *75*, 2000–2036.
Hernawan, U.E., van Dijk, K.J., et al. (2017). Historical processes and contemporary ocean currents drive genetic structure in the seagrass Thalassia hemprichii in the Indo-Australian Archipelago. Mol. Ecol. *26*, 1008–1021.
Ho, S.Y.W., Lanfear, R., et al. (2011). Time-dependent rates of molecular evolution. Mol. Ecol. *20*, 3087–3101.
Hoegh-Guldberg, O., and Bruno, J.F. (2010). The impact of climate change on the world's marine ecosystems. Science (80-.). *328*, 1523–1528.
Hoeksema, B.W. (2007). Delineation of the Indo-Malayan centre of maximum marine biodiversity: the Coral Triangle. In Biogeography, Time, and Place: Distributions, Barriers, and Islands, (Springer, the Netherlands), pp. 118–178.
Hohenlohe, P.A., Bassham, S., et al. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. *6*.
Holland, B.R., Huber, K.T., et al. (2002). Delta plots: a tool for analyzing phylogenetic distance data. Mol. Biol. Evol. *19*, 2051–2059.
Holst, S., Sötje, I., et al. (2007). Life cycle of the rhizostome jellyfish Rhizostoma octopus (L.) (Scyphozoa, Rhizostomeae), with studies on cnidocysts and statoliths. Mar. Biol. *151*, 1695–1710.
Holthuis, L.B. (1973). Caridean shrimps found in land-locked salt-water pools at four Indo-West Pacific localities (Sinai Peninsula, Funafuti Atoll, Maui and Hawaii Islands), with the description of one new genus and four new species. Zool. Verh. *128*, 1–48.
Horn, H., Slaby, B.M., et al. (2016). An Enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. Front. Microbiol. *7*, 1–15.
Howard, B.C. (2016). The famous Jellyfish Lake is running out of jellyfish. Natl. Geogr. Mag.
Huang, D., Meier, R., et al. (2008). Slow mitochondrial COI sequence evolution at the base of the metazoan tree and its implications for DNA barcoding. J. Mol. Evol. *66*, 167–174.
Hubbell, S.P. (2001). The Unified Neutral Theory of Biodiversity and Biogeography (Princeton University Press, Princeton, NJ).
Hughes, T.P., Barnes, M.L., et al. (2017). Coral reefs in the Anthropocene. Nature *546*, 82–90.
Huson, D.H. (1998). SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics *14*, 68–73.
Huson, D.H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. *23*, 254–267.
IPCC (2014). Climate Change 2014: Synthesis Report. Contributions of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Geneva, Switzerland).
Itescu, Y. (2018). Are island-like systems biologically similar to islands? A review of the evidence. Ecography (Cop.). 1298–1314.
Jeffery, N.W., Jardine, C.B., et al. (2013). A first exploration of genome size diversity in sponges. Genome *56*, 451–456.
Johannesson, K. (1988). The paradox of Rockall: why is a brooding gastropod (Littorina saxatilis) more widespread than one having a planktonic larval dispersal stage (L. littorea)? Mar. Biol. *99*, 507–513.
Johnson, P.L.F., and Slatkin, M. (2008). Accounting for bias from sequencing error in population genetic estimates. Mol. Biol. Evol. *25*, 199–206.
Jones, C.G., Lawton, J.H., et al. (1994). Organisms as ecosystem engineers. In Ecosystem Management, (Springer, New York, NY), pp. 130–147.
Jones, F.C., Grabherr, M.G., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. Nature *484*, 55–61.
Kano, Y., and Kase, T. (2004). Genetic exchange between anchialine cave populations by means of larval dispersal: The case of a new gastropod species Neritilia cavernicola. Zool. Scr. *33*, 423–437.
Kartasapoetra, G., Kartasapoetra, R.G., et al. (1985). Hukum Tanah jaminan UUPA bagi keberhasilan pendayagunaan tanah. PT. Bina Aksara Jakarta.
Kasmidi, M., and Gunadharma, A. (2017). Diving in Manta Sandy? Here's what you need to know. Bird's Head Seascape.
Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
Katzenberg, M.A. (2000). Stable isotope analysis: a tool for studying past diet, demography, and life history. In Biological Anthropology of the Human Skeleton, M.A. Katzenberg, and S.R. Saunders, eds. (Wiley-Liss, New York), pp. 305–327.
Keenan, K., Mcginnity, P., et al. (2013). DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods Ecol. Evol. *4*, 782–788.
Kelley, J.L., Brown, A.P., et al. (2016). The life aquatic: Advances in marine vertebrate genomics. Nat. Rev. Genet. *17*, 523–534.
Klautau, M., Russo, C.A.M., et al. (1999). Does cosmopolitanism result from overconservative systematics? A case study using the marine sponge Chondrilla nucula. Evolution (N. Y). *53*, 1414–1422.
Kohler, K.E., and Gill, S.M. (2006). Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. Comput.

Geosci. *32*, 1259–1269.
Korneliussen, T.S., Albrechtsen, A., et al. (2014). ANGSD: Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*, 356.
Kraft, N.J.B., Adler, P.B., et al. (2015). Community assembly, coexistence and the environmental filtering metaphor. Funct. Ecol. *29*, 592–599.
Kumar, S., Stecher, G., et al. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. *33*, 1870–1874.
Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome assemblies. F1000Research *6*, 1287.
Lal, M.M., Southgate, P.C., et al. (2016). Fishing for divergence in a sea of connectivity : The utility of ddRADseq genotyping in a marine invertebrate , the black-lip pearl oyster. Mar. Genomics *25*, 57–68.
Lal, M.M., Southgate, P.C., et al. (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.
Lauzon-Guay, J.S., Dionne, M., et al. (2005). Effects of seed size and density on growth, tissue-to-shell ratio and survival of cultivated mussels (Mytilus edulis) in Prince Edward Island, Canada. Aquaculture *250*, 652–665.
Lee, C.-R., and Mitchell-Olds, T. (2011). Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. Mol. Ecol. *20*, 4631–4642.
de Leeuw, C.A., Peijnenburg, K.T.C.A., et al. (2020). First come, first served: Possible role for priority effects in marine populations under different degrees of dispersal potential. J. Biogeogr. 1649–1662.
Lefort, V., Desper, R., et al. (2015). FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. Mol. Biol. Evol. *32*, 2798–2800.
Legendre, P. (1993). Spatial Autocorrelation : Trouble or New Paradigm ? *74*, 1659–1673.
Legendre, J.V., and Legendre, L.F. (2012). Numerical Ecology (Elsevier, Amsterdam, Netherlands).
Legrand, D., Cote, J., et al. (2017). Eco-evolutionary dynamics in fragmented landscapes. Ecography (Cop.). *40*, 9–25.
Leiva, C., Taboada, S., et al. (2019). Population substructure and signals of divergent adaptive selection despite admixture in the sponge Dendrilla antarctica from shallow waters surrounding the Antarctic Peninsula. Mol. Ecol. 3151–3170.
Lemopoulos, A., Prokkola, J.M., et al. (2019). Comparing RADseq and microsatellites for estimating genetic diversity and relatedness — Implications for brown trout conservation. Ecol. Evol. *9*, 2106–2120.
Li, W., and Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.
Li, H., Handsaker, B., et al. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
Liggins, L., Treml, E.A., et al. (2013). Taking the Plunge: An Introduction to Undertaking Seascape Genetic Studies and using Biophysical Models. Geogr. Compass *7*, 173–196.
Liggins, L., Treml, E.A., et al. (2020). Seascape genomics: contextualizing adaptive and neutral genomic variation in the ocean environment. In Population Genomics: Marine Organisms, (Springer, Cham), pp. 171–218.
Liu, X., and Fu, Y.-X. (2015). Exploring population size changes using SNP frequency spectra. Nat. Genet. *47*, 555–559.
Longhurst, R. (2003). Semi-structured interviews and focus groups. In Key Methods in Geography, pp. 177–132.
Lopez-Maldonado, Y., and Berkes, F. (2017). Restoring the environment, revitalizing the culture: Cenote conservation in Yucatan, Mexico. Ecol. Soc. *22*.
Losos, J.B., and Ricklefs, R.E. (2010). The Theory of Island Biogeography Revisited (Princeton University Press).
Lowry, D.B., and Gould, B.A. (2016). Speciation Continuum. Encycl. Evol. Biol. *105*, 159–165.
Lucek, K., Sivasundar, A., et al. (2013). Repeated and predictable patterns of ecotypic differentiation during a biological invasion: Lake-stream divergence in parapatric Swiss stickleback. J. Evol. Biol. *26*, 2691–2709.
Luikart, G., England, P.R., et al. (2003). The power and promise of population genomics: From genotyping to genome typing. Nat. Rev. Genet. *4*, 981–994.
Luikart, G., Kardos, M., et al. (2018). Population Genomics: Advancing Understanding of Nature. 3–79.
Luu, K., Bazin, E., et al. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol. Ecol. Resour. *17*, 67–77.
Lynch, M. (1997). Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. Mol. Biol. Evol. *14*, 914–925.
Lynch, M. (2008). Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects. Mol. Biol. Evol. *25*, 2409–2419.
Lynch, M. (2010). Evolution of the mutation rate. Trends Genet. *26*, 345–352.
Maas, D.L., Prost, S., et al. (2018). Rapid divergence of mussel populations despite incomplete barriers to dispersal. Mol. Ecol. *27*, 1556–1571.
Maas, D.L., Capriati, A., et al. (2020). Recognizing peripheral ecosystems in marine protected areas: A case study of golden jellyfish lakes in Raja Ampat, Indonesia. Mar. Pollut. Bull. *151*, 110700.
MacArthur, R.H., and Wilson, E.O. (1963). An Equilibrium Theory of Insular Zoogeography. Evolution (N. Y). *17*, 373.

MacArthur, R.H., and Wilson, E.O. (1967). The theory of island biogeography (Princeton University Press).
Maldonado, M. (2006). The ecology of the sponge larva. Can. J. Zool. *84*, 175–194.
Malinsky, M., Trucchi, E., et al. (2018). RADpainter and fineRADstructure : Population Inference from RADseq Data. Mol. Biol. Evol. *35*, 1284–1290.
Mangubhai, S., Erdmann, M. V., et al. (2012). Papuan Bird's Head Seascape: Emerging threats and challenges in the global center of marine biodiversity. Mar. Pollut. Bull. *64*, 2279–2295.
Manno, C., Bednaršek, N., et al. (2017). Shelled pteropods in peril: Assessing vulnerability in a high CO2 ocean. Earth-Science Rev. *169*, 132–145.
Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. Cancer Res. *27*, 209–220.
Mariani, S., Peijnenburg, K.T.C.A., et al. (2012). Independence of neutral and adaptive divergence in a low dispersal marine mollusc. Mar. Ecol. Prog. Ser. *446*, 173–187.
Marques, D.A., Lucek, K., et al. (2016). Genomics of Rapid Incipient Speciation in Sympatric Threespine Stickleback. PLoS Genet. *12*, 1–34.
Marshall, D.J., Monro, K., et al. (2010). Phenotype-environment mismatches reduce connectivity in the sea. Ecol. Lett. *13*, 128–140.
Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.
Martin, L.E., Dawson, M.N., et al. (2006). Marine lake ecosystem dynamics illustrate ENSO variation in the tropical western Pacific. Biol. Lett. *2*, 144–147.
Martínez, A., García-Herrero, G., et al. (2019). Lanzarote and Chinijo Islands: An Anchialine UNESCO Global Geopark. In Lanzarote and Chinijo Islands Geopark: From Earth to Space, (Springer, Cham), pp. 109–121.
Masonjones, H., Rose, E., et al. (2019). High density, early maturing, and morphometrically unique Hippocampus erectus population makes a Bahamian pond a priority site for conservation. Endanger. Species Res. *39*, 35–49.
Mayr, E. (1942). Systematics and the Origin of Species (New York, USA: Columbia University Press).
Mayr, E. (1954). Change of genetic environment and evolution (London: Allen & Unwin).
Mayr, E. (1963). Animal species and evolution (Cambridge, Massachusetts: Harvard Univeristy Press).
McKenna, A., Hanna, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303.
McLeod, E., Szuster, B., et al. (2009). Sasi and marine conservation in Raja Ampat, Indonesia. Coast. Manag. *37*, 656–676.
McRae, B.H. (2006). Isolation by resistance. Evolution (N. Y). *60*, 1551–1561.
de Meester, L., Gómez, A., et al. (2002). The Monopolization Hypothesis and the dispersal – gene flow paradox in aquatic organisms. Acta Oecologica *23*, 121–135.
De Meester, L., Vanoverbeke, J., et al. (2016). Evolving Perspectives on Monopolization and Priority Effects. Trends Ecol. Evol. *31*, 136–146.
Meirmans, P.G., and Hedrick, P.W. (2011). Assessing population structure: FST and related measures. Mol. Ecol. Resour. *11*, 5–18.
Meirmans, P.G., and Van Tienderen, P.H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes *4*, 792–794.
Mekkes, L., Renema, W., et al. (2021). Pteropods make thinner shells in the upwelling region of the California Current Ecosystem. Sci. Rep. *11*, 1–11.
Melatunan, S., Calosi, P., et al. (2013). Effects of ocean acidification and elevated temperature on shell plasticity and its energetic basis in an intertidal gastropod. Mar. Ecol. Prog. Ser. *472*, 155–168.
Mercado-Salas, N.F., Morales-Vela, B., et al. (2013). Conservation status of the inland aquatic crustaceans in the Yucatan Peninsula, Mexico: Shortcomings of a protection strategy. Aquat. Conserv. Mar. Freshw. Ecosyst. *23*, 939–951.
Metzker, M.L. (2010). Sequencing technologies the next generation. Nat. Rev. Genet. *11*, 31–46.
Meyer, C.P. (2003). Molecular systematics of cowries (Gastropoda: Cypraeidae) and diversification patterns in the tropics. Biol. J. Linn. Soc. *79*, 401–459.
Middelburg, J.J., Nieuwenhuize, J., et al. (1996). Sediment biogeochemistry in an east African mangrove forest (Gazi Bay, Kenya). Biogeochemistry *34*, 133–155.
Milano, I., Babbucci, M., et al. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (Merluccius merluccius). Mol. Ecol. *23*, 118–135.
Miller, K.J., Baird, H.P., et al. (2018). Complex genetic structure revealed in the circum-Antarctic broadcast spawning sea urchin Sterechinus neumayeri. Mar. Ecol. Prog. Ser. *601*, 153–166.
Miner, B.G., Sultan, S.E., et al. (2005). Ecological consequences of phenotypic plasticity. Trends Ecol. Evol. *20*, 685–692.
Momigliano, P., Jokinen, H., et al. (2017). Extraordinarily rapid speciation in a marine fish. Proc. Natl. Acad. Sci. U. S. A. *114*, 6074–6079.
Monteiro-Ribas, W., Rocha-Miranda, F., et al. (2006). Larval development of Brachidontes solisianus (Bivalvia, Mytilidae), with notes on differences between its hinge system and that of the mollusk Perna perna. Braz. J. Biol. *66*, 109–116.

Moritz, C. (1994). Defining ' Evolutionarily Significant Units ' for conservation. Trends Ecol. Evol. 373–375.
Morrow, C., and Cárdenas, P. (2015). Proposal for a revised classification of the Demospongiae (Porifera). Front. Zool. *12*, 1–27.
Morton, B. (1988). The population dynamics and reproductive cycle of Brachidontes variabilis (Bivalvia: Mytilidae) in a Hong Kong mangrove. Malacol. Rev. *21*, 109–117.
Muscatine, L., Wilkerson, F., et al. (1986). Regulation of population density of symbiotic algae in a tropical marine jellyfish (Mastigias sp.) . Mar. Ecol. Prog. Ser. *32*, 279–290.
Mustaghfirin, Urbinas, M.P., et al. (2012). Buku 1 & 2 Rencana pengelolaan taman pulaupulau kecil dan daerah (TPPKD) Raja Ampat. Unit pelaksana teknis daerah TPPKD Raja Ampat. (Waisai).
Nagelkerken, I., Goldenber, S.U., et al. (2020). Trophic pyramids reorganize when food web architecture fails to adjust to ocean change. Science (80-.). *369*, 829–832.
Narum, S.R., and Hess, J.E. (2011). Comparison of FST outlier tests for SNP loci under selection. Mol. Ecol. Resour. *11*, 184–194.
Nei, M. (1987). Molecular evolutionary genetics (New York, USA: Columbia University Press).
Neves, J.M.M., Lima, S.M.Q., et al. (2016). Population Structure of the rockpool blenny entomacrodus vomerinus shows source-sink dynamics among ecoregions in the tropical Southwestern Atlantic. PLoS One *11*, 1–15.
Ngan-Tillard, D., and Huisman, H. (2017). Micro-CT scanning. In Archaeological Soil and Sediment Micromorphology, C. Nicosia, and G. Stoops, eds. (John Wiley & Sons, Ltd, Chichester), pp. 441–449.
Nguyen, L.T., Schmidt, H.A., et al. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
Nosil, P., and Schluter, D. (2011). The genes underlying the process of speciation. Trends Ecol. Evol. *26*, 160–167.
Nosil, P., Funk, D.J., et al. (2009). Divergent selection and heterogeneous genomic divergence. Mol. Ecol. *18*, 375–402.
Noyer, C., and Becerro, M.A. (2012). Relationship between genetic, chemical, and bacterial diversity in the Atlanto-Mediterranean bath sponge Spongia lamella. Hydrobiologia *687*, 85–99.
Nuismer, S.L., and Gandon, S. (2008). Moving beyond Common-Garden and Transplant Designs: Insight into the Causes of Local Adaptation in Species Interactions. Am. Nat. *171*, 658–668.
Oksanen, J., Blanchet, F.G., et al. (2016). Vegan: community ecology package. R Package 2.3-3.
Oleksiak, M.F. (2018). Adaptation Without Boundaries: Population Genomics in Marine Systems. In Population Genomics, (Springer, Cham), pp. 587–612.
Oleksiak, M.F., and Rajora, O.P. (2020a). Marine population genomics: challenges and opportunities. In Population Genomics: Marine Organisms, (Springer, Cham), pp. 3–35.
Oleksiak, M.F., and Rajora, O.P. (2020b). Population Genomics: Marine Organisms.
van Oppen, M.J.H., Wörheide, G., et al. (2002). Nuclear markers in evolutionary and population genetic studies of scleractinian corals and sponges. Proc. 9th Int. Coral Reef Symp. *1*.
Orsini, L., Vanoverbeke, J., et al. (2013). Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. Mol. Ecol. *22*, 983–5999.
Osinga, R., Schutter, M., et al. (2011). The Biology and Economics of Coral Growth. Mar. Biotechnol. *13*, 658–671.
Palsbøll, P.J., Bérubé, M., et al. (2007). Identification of management units using population genetic data. Trends Ecol. Evol. *22*, 11–16.
Palumbi, S.R. (1994). Genetic Divergence, Reproductive Isolation, and Marine Speciation. Annu. Rev. Ecol. Syst. *25*, 547–572.
Palumbi, S.R., Reviews, A., et al. (1994). Genetic Divergence , Reproductive Isolation , and Marine Speciation Author (s): Stephen R . Palumbi Source : Annual Review of Ecology and Systematics , Vol . 25 (1994), pp . 547-572 Published by : Annual Reviews Stable URL : http://www.jstor.org/stabl. *25*, 547–572.
Paradis, E. (2010). Pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics *26*, 419–420.
Patiño, J., Whittaker, R.J., et al. (2017). A roadmap for island biology: 50 fundamental questions after 50 years of The Theory of Island Biogeography. J. Biogeogr. *44*, 963–983.
Patris, S., Martin, L.E., et al. (2019). Expansion of an introduced sea anemone population, and its associations with native species in a tropical marine lake (Jellyfish Lake, Palau). Front. Biogeogr. *11*, 0–12.
Pazmiño, D.A., Maes, G.E., et al. (2017). Genome-wide SNPs reveal low effective population size within confined management units of the highly vagile Galapagos shark (Carcharhinus galapagensis). Conserv. Genet. 1–13.
Peijnenburg, K.T.C.A., and Goetze, E. (2013). High evolutionary potential of marine zooplankton. Ecol. Evol. *3*, 2765–2783.
Pérez-Portela, R., and Riesgo, A. (2018). Population Genomics of Early-Splitting Lineages of Metazoans. In Population Genomics, P. Rajora, ed. (Cham, Switzerland: Springer), pp. 103–137.
Pérez-Portela, R., Noyer, C., et al. (2015). Genetic structure and diversity of the endangered bath sponge Spongia lamella. Aquat. Conserv. Mar. Freshw. Ecosyst. *25*, 365–379.
Peterson, B.K., Weber, J.N., et al. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One *7*.

Pinheiro, H.T., Bernardi, G., et al. (2017). Island biogeography of marine organisms. Nature.
Pitt, K.A., and Lucas, C.H. (2014). Jellyfish blooms. Jellyfish Bloom. *9789400770*, 1–304.
Pretty, J., and Smith, D. (2004). Social Capital in Biodiversity Conservation and Management. Conserv. Biol. *18*, 631–638.
Prieto, L., Astorga, D., et al. (2010). Environmental control of phase transition and polyp survival of a massive-outbreaker jellyfish. PLoS One *5*.
Puebla, O. (2009). Ecological speciation in marine v. freshwater fishes. J. Fish Biol. *75*, 960–996.
Purba, G.Y., Haryono, E., et al. (2018). Jellyfish Lakes at Misool Islands, Raja Ampat, West Papua, Indonesia. Jellyfish Lakes at Misool Islands Biodiversitas *19*, 2085–4722.
Puritz, J.B., Matz, M. V., et al. (2014). Demystifying the RAD fad. Mol. Ecol. *23*, 5937–5942.
R Core Team (2018). R: A language and environment for statistical computing.
Rambaut, A. (2009). FigTree.
Reimer, O., and Tedegren, M. (1996). Phenotypical Improvement of Morphological Defences in the Mussel Mytilus edulis Induced by Exposure to the Predator Asterias rubens. Oikos *75*, 383–390.
Richards, V.P., Bernard, A.M., et al. (2016). Patterns of population structure and dispersal in the long-lived "redwood" of the coral reef, the giant barrel sponge (Xestospongia muta). Coral Reefs *35*, 1097–1107.
Riesgo, A., Taboada, S., et al. (2019). Genetic diversity, connectivity and gene flow along the distribution of the emblematic Atlanto-Mediterranean sponge Petrosia ficiformis (Haplosclerida, Demospongiae). BMC Evol. Biol. *19*, 1–18.
Riginos, C., Crandall, E.D., et al. (2016). Navigating the currents of seascape genomics: How spatial analyses can augment population genomic studies. Curr. Zool. *62*, 581–601.
Roberts, C.M., Mcclean, C.J., et al. (2002). Marine Biodiversity Hotspots for Priorities and Conservation Tropical Reefs. Science (80-.). *295*, 1280–1284.
Rochette, N.C., and Catchen, J.M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. Nat. Protoc. *12*, 2640–2659.
Roesti, M., Hendry, A.P., et al. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. Mol. Ecol. *21*, 2852–2862.
Roesti, M., Gavrilets, S., et al. (2014). The genomic signature of parallel adaptation from shared genetic variation. Mol. Ecol. *23*, 3944–3956.
Roesti, M., Kueng, B., et al. (2015). The genomics of ecological vicariance in threespine stickleback fish. Nat. Commun. *6*, 1–14.
Rohlf, F.J. (1995). Statistical analysis of shape using partial-warp scores. Proc. Curr. Issues Stat. Shape Anal. 154–158.
Rohlf, F.J. (2010a). Tpsutil, file utility prgram.
Rohlf, F.J. (2010b). Tpsdig, digitize landmarks and outlines.
Rohlf, F.J. (2010c). Tpsrelw, relative warps analysis.
Rosa, R.D., Alonso, P., et al. (2015). High polymorphism in big defensin gene expression reveals presence-absence gene variability (PAV) in the oyster Crassostrea gigas. Dev. Comp. Immunol. *49*, 231–238.
Rundle, H.D., and Nosil, P. (2005). Ecological speciation. Ecol. Lett. *8*, 336–352.
Russ, A., Santos, S.R., et al. (2010). Genetic population structure of an anchialine shrimp, Metabetaeus lohena (Crustacea: Alpheidae), in the Hawaiian Islands. Rev. Biol. Trop. *58*, 159–170.
Salzburger, W., Ewing, G.B., et al. (2011). The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. Mol. Ecol. *20*, 1952–1963.
Sanderson, M.J. (1997). A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. *14*, 1218–1231.
Santos, S.R. (2006). Patterns of genetic connectivity among anchialine habitats: A case study of the endemic Hawaiian shrimp Halocaridina rubra on the island of Hawaii. Mol. Ecol. *15*, 2699–2718.
Santos, S.R., and Weese, D.A. (2011). Rocks and clocks: Linking geologic history and rates of genetic differentiation in anchialine organisms. Hydrobiologia *677*, 53–64.
Santos, A.M.C., Field, R., et al. (2016). New directions in island biogeography. Glob. Ecol. Biogeogr. *25*, 751–768.
Sarà, G., Romano, C., et al. (2008). Effect of salinity and temperature on feeding physiology and scope for growth of an invasive species (Brachidontes pharaonis - MOLLUSCA: BIVALVIA) within the Mediterranean sea. J. Exp. Mar. Bio. Ecol. *363*, 130–136.
Sarà, G., Palmeri, V., et al. (2013). Predicting biological invasions in marine habitats through eco-physiological mechanistic models: A case study with the bivalve Brachidontes pharaonis. Divers. Distrib. *19*, 1235–1247.
Sathiamurthy, E., and Voris, K.H. (2006). Maps of Holocene Sea Level Transgression and Submerged Lakes on the Sunda Shelf.
Scheffer, M. (1997). Ecology of shallow lakes (Springer Science & Business Media).
Schluter, D. (2009). Evidence for ecological speciation and its alternative. Science *323*, 737–741.
Selkoe, K.A., Henzler, C.M., et al. (2008). Seascape genetics and the spatial ecology of marine populations. Fish Fish. *9*, 363–377.
Selkoe, K.A., Gaggiotti, O.E., et al. (2014). Emergent patterns of population genetic structure for a coral reef community. Mol. Ecol. *23*, 3064–3079.

Selkoe, K.A., D'Aloia, C.C., et al. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. Mar. Ecol. Prog. Ser. *554*, 1–19.
Sheehy, D.J., and Vik, S.F. (2010). The role of constructed reefs in non-indigenous species introductions and range expansions. Ecol. Eng. *36*, 1–11.
Skotte, L., Korneliussen, T.S., et al. (2013). Estimating individual admixture proportions from next generation sequencing data. Genetics *195*, 693–702.
Slatkin, M. (1985). Gene flow in natural populations. Annu. Rev. Ecol. Syst. *16*, 393–430.
Slatkin, M. (1993). Isolation by Distance in Equilibrium and Non-Equilibrium Populations. Evolution (N. Y). *47*, 264.
Slatkin, M. (1995). A Measure of Population Subdivision Based on Microsatellite Allele Frequencies. Genetics *139*, 457–462.
Slice, D. (2007). Geometric morphometrics. Annu. Rev. Anthropol. *36*, 261–281.
Smit, A.F.A., Hubley, R., et al. (2014). RepeatMasker Open-4.0. 2013-2015.
Sonnenfeld, P., and Hudec, P.P. (1980). Heliothermal lakes. Dev. Sedimentol. *28*, 93–100.
Souza, M.R.D.E., Dawson, M.N., et al. (2018). Redescription of Mastigias papua (Scyphozoa, Rhizostomeae) with designation of a neotype and recognition of two additional species. Zootaxa *4457*, 520–536.
Spalding, M.D., Fox, H.E., et al. (2007). Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. Bioscience *57*, 573.
Spidle, a P.E. Al (1995). Limits to tolerance of temperature and salinity in the quagga mussel and the zebra mussel. Can. J. Fish Aquat. Sci. *52*, 2108–2119.
Srivastava, M., Simakov, O., et al. (2010). The Amphimedon queenslandica genome and the evolution of animal complexity. Nature *466*, 720–726.
Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
Starger, C.J., Van Nydeck Erdmann, M., et al. (2015). Strong genetic structure among coral populations within a conservation priority region, the Bird's Head Seascape (Papua and West Papua, Indonesia). Front. Biogeogr. *7*.
Suchan, T., Pitteloud, C., et al. (2016). Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. PLoS One *11*, 1–22.
Sugiura, Y. (1965). On the Life-History of Rhizostome Medusae. III. On the Effects of Temperature on the Strobilation of Mastigias papua. Biol. Bull. *128*, 493–496.
Sunde, J., Yıldırım, Y., et al. (2020). Comparing the Performance of Microsatellites and RADseq in Population Genetic Studies: Analysis of Data for Pike (Esox lucius) and a Synthesis of Previous Studies. Front. Genet. *11*, 1–17.
Sundqvist, L., Keenan, K., et al. (2016). Directional genetic differentiation and relative migration. Ecol. Evol. *6*, 3461–3475.
Swierts, T., Peijnenburg, K.T.C.A., et al. (2013). Lock, Stock and Two Different Barrels: Comparing the Genetic Composition of Morphotypes of the Indo-Pacific Sponge Xestospongia testudinaria. PLoS One *8*, 1–12.
Swift, H.F., and Dawson, M.N. (2020). Demographic, environmental, and phenotypic change but genetic consistency in the jellyfish Mastigias papua. Biol. Bull. *239*, 80–94.
Swift, H.F., Gómez Daglio, L., et al. (2016). Three routes to crypsis: Stasis, convergence, and parallelism in the Mastigias species complex (Scyphozoa, Rhizostomeae). Mol. Phylogenet. Evol. *99*, 103–115.
Taboada, S., Riesgo, A., et al. (2018). Implications of population connectivity studies for the design of marine protected areas in the deep sea: An example of a demosponge from the Clarion-Clipperton Zone. Mol. Ecol. *27*, 4657–4679.
Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. Genetics *105*, 437–460.
Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.
Takao, M., Okawachi, H., et al. (2014). Natural predators of polyps of Aurelia aurita s.l. (Cnidaria: Scyphozoa: Semaeostomeae) and their predation rates. Plankt. Benthos Res. *9*, 105–113.
Tariel, J., Longo, G.C., et al. (2016). Tempo and mode of speciation in Holocanthus angelfishes based on RADseq markers. Mol. Phylogenet. Evol. *98*, 84–88.
Terranova, M.S., Lo Brutto, S., et al. (2007). A mitochondrial phylogeography of Brachidontes variabilis (Bivalvia: Mytilidae) reveals three cryptic species. J. Zool. Syst. Evol. Res. *45*, 289–298.
Timm, L.E. (2020). A fair fight between molecular marker types in a seascape genetics setting. Mol. Ecol. *29*, 2133–2136.
Timm, J., and Kochzius, M. (2008). Geological history and oceanography of the Indo-Malay Archipelago shape the genetic population structure in the false clown anemonefish (Amphiprion ocellaris). Mol. Ecol. *17*, 3999–4014.
Tomascik, T., and Mah, A.J. (1994). The ecology of Halimeda Lagoon: An anchialine Lagoon of a raised atoll, Kakaban Island, Eas kalimantan, Indonesia. Trop. Biodiversoty *2*, 385–399.
Tornabene, L., Valdez, S., et al. (2015). Support for a "Center of Origin" in the Coral Triangle: Cryptic diversity, recent speciation, and local endemism in a diverse lineage of reef fishes (Gobiidae: Eviota). Mol. Phylogenet. Evol. *82*, 200–210.

Trave, C., Brunnschweiler, J., et al. (2017). Are we killing them with kindness? Evaluation of sustainable marine wildlife tourism. Biol. Conserv. *209*, 211–222.
Treml, E.A., Roberts, J., et al. (2015). The emergent geography of biophysical dispersal barriers across the Indo-West Pacific. Divers. Distrib. *21*, 465–476.
Uriz, M.J., and Turon, X. (2012). Sponge Ecology in the Molecular Era (Elsevier Ltd.).
Valladares, F., Matesanz, S., et al. (2014). The effects of phenotypic plasticity and local adaptation on forecasts of species range shifts under climate change. Ecol. Lett. *17*, 1351–1364.
Vellend, M. (2003). Island Biogeography of Genes and Species. Am. Nat. *162*, 358–365.
Vellend, M. (2005). Species diversity and genetic diversity: parallel processes and correlated patterns. Am. Nat. *166*, 199–215.
Vellend, M., and Geber, M.A. (2005). Connections between species diversity and genetic diversity. Ecol. Lett. *8*, 767–781.
Ventura, M., Petrusek, A., et al. (2014). Local and regional founder effects in lake zooplankton persist after thousands of years despite high dispersal potential. Mol. Ecol. *23*, 1014–1027.
Vieira, F.G., Fumagalli, M., et al. (2013). Estimating inbreeding coefficients from NGS data : impact on genotype calling and allele frequency estimation. Genome Res. *23*, 1852–1861.
Vieira, F.G., Lassalle, F., et al. (2016). Improving the estimation of genetic distances from Next-Generation Sequencing data. Biol. J. Linn. Soc. *117*, 139–149.
De Villemereuil, P., Gaggiotti, O.E., et al. (2016). Common garden experiments in the genomic era: New perspectives and opportunities. Heredity (Edinb). *116*, 249–254.
Vu, N.T.T., Zenger, K.R., et al. (2020). Fine-scale population structure and evidence for local adaptation in Australian giant black tiger shrimp (Penaeus monodon) using SNP analysis. BMC Genomics *21*, 1–18.
Waddington, C.H. (1942). Canalization of development and the inheritance of acquired characters. Nature *150*, 563–565.
Wagner, C.E., Harmon, L.J., et al. (2014). Cichlid species-area relationships are shaped by adaptive radiations that scale with area. Ecol. Lett. *17*, 583–592.
Walters, W., Hyde, E.R., et al. (2015). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. MSystems *1*, e0009-15.
Wang, I.J., and Bradburd, G.S. (2014). Isolation by environment. Mol. Ecol. *23*, 5649–5662.
Wang, I.J., and Summers, K. (2010). Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. Mol. Ecol. *19*, 447–458.
Wang, H., Xu, X., et al. (2016). The Power of Inbreeding: NGS-Based GWAS of Rice Reveals Convergent Evolution during Rice Domestication. Mol. Plant *9*, 975–985.
Ward, R. (1994). A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. J. Fish Biol. *44*, 213–232.
Warren, B.H., Simberloff, D., et al. (2015). Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. Ecol. Lett. *18*, 200–217.
Waters, J.M., Fraser, C.I., et al. (2013). Founder takes all: Density-dependent processes structure biodiversity. Trends Ecol. Evol. *28*, 78–85.
Watterson, G.A. (1975). On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. *7*, 256–276.
Weber, J.N., Bradburd, G.S., et al. (2016). Partitioning the effects of isolation by distance, environment, and physical barriers on genomic divergence between parapatric threespine stickleback. Evolution (N. Y). 1–15.
Webster, N.S., and Thomas, T. (2016). The sponge hologenome. MBio *7*, 1–14.
Weese, D.A., Fujita, Y., et al. (2012). The Long and Short of It: Genetic Variation and Population Structure of the Anchialine Atyid Shrimp Caridina rubella on Miyako-Jima, Japan. J. Crustac. Biol. *32*, 109–117.
Weese, D.A., Fujita, Y., et al. (2013). Multiple colonizations lead to cryptic biodiversity in an island ecosystem: Comparative phylogeography of anchialine shrimp species in the Ryukyu Archipelago, Japan. Biol. Bull. *225*, 24–41.
Whalan, S., de Nys, R., et al. (2008). Low genetic variability within and among populations of the brooding sponge rhopaloeides odorabile on the central great barrier reef. Aquat. Biol. *3*, 111–119.
Whittaker, R.J., Fernández-Palacios, J.M., et al. (2017). Island biogeography: Taking the long view of natures laboratories. Science (80-.). *357*.
Wickham, H. (2016). ggplot2: elegant graphics for data analysis (Springer).
Wiens, J.J., Camacho, A., et al. (2019). Climate change, extinction, and Sky Island biogeography in a montane lizard. Mol. Ecol. *28*, 2610–2624.
De Wit, P., and Palumbi, S.R. (2013). Transcriptome-wide polymorphisms of red abalone (Haliotis rufescens) reveal patterns of gene flow and local adaptation. Mol. Ecol. *22*, 2884–2897.
Wood, D.E., and Salzberg, S.L. (2014). Kraken: Ultrafast metagenomic sequence classification using exact alignments. Genome Biol. *15*.
Wörheide, G., Hooper, J.N.A., et al. (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). Mol. Ecol. *11*, 1753–1768.

Wörheide, G., Solé-Cava, A.M., et al. (2005). Biodiversity, molecular ecology and phylogeography of marine sponges: Patterns, implications and outlooks. Integr. Comp. Biol. *45*, 377–385.
Wörheide, G., Epp, L.S., et al. (2008). Deep genetic divergences among Indo-Pacific populations of the coral reef sponge Leucetta chagosensis (Leucettidae): Founder effects, vicariance, or both? BMC Evol. Biol. *8*, 1–18.
Wright, S. (1943). Isolation by distance. Genetics *28*, 114–138.
van Wyngaarden, M., Snelgrove, P.V.R., et al. (2017). Identifying patterns of dispersal, connectivity and selection in the sea scallop, Placopecten magellanicus, using RADseq-derived SNPs. Evol. Appl. *10*, 102–117.
Van Wyngaarden, M., Snelgrove, P.V.R., et al. (2016). Identifying patterns of dispersal, connectivity, and selection in the sea scallop, *Placopecten magellanicus,* using RAD-seq derived SNPs. Evol. Appl. 1–16.
Xu, T., Sun, J., et al. (2017). Genome-wide discovery of single nucleotide polymorphisms (SNPs) and single nucleotide variants (SNVs) in deep-sea mussels: Potential use in population genomics and cross-species application. Deep. Res. Part II Top. Stud. Oceanogr. *137*, 318–326.
Xuereb, A., Benestan, L., et al. (2018). Asymmetric oceanographic processes mediate connectivity and population genetic structure, as revealed by RADseq, in a highly dispersive marine invertebrate (Parastichopus californicus). Mol. Ecol. *27*, 2347–2364.
Xuereb, A., D'Aloia, C.C., et al. (2019). Marine conservation and marine protected areas. In Population Genomics: Marine Organisms, (Springer, Cham), pp. 423–446.
Zhang, G., Fang, X., et al. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. Nature *490*, 49–54.
Zhao, X., Han, Y., et al. (2020). CO2-driven ocean acidification weakens mussel shell defense capacity and induces global molecular compensatory responses. Chemosphere *243*, 125415.

Summary
Samenvatting
Ringkasan

SUMMARY

The marine realm is seemingly boundless, and the absence of physical barriers to dispersal has led to assumptions of high dispersal abilities and subsequent slow speciation rates for marine organisms. Recently, these assumptions are being overturned by studies showing high genetic structuring for marine populations even at small spatial scales. Isolating processes at small scales may therefore play an important role in shaping genetic diversity within species. Isolation decreases effective gene flow between populations, which allows populations to diverge. Isolation may consist of processes affecting dispersal limitation or establishment limitation. In this thesis, four modes of isolation are considered, including neutral and selective processes. Neutral processes of dispersal limitation include geographic distance between populations (isolation-by-distance) and permeability of the matrix between populations (isolation-by-resistance). A selective process that affects establishment is ecological speciation, or restrictions in gene flow between environmentally divergent habitats, resulting in isolation-by-environment. Lastly, establishment success can also be affected by historical contingency. Here, eco-evolutionary dynamics of priority effects with potential adaptive components may result in first colonizers shaping the subsequent structure of populations in such a way that later immigrants cannot successfully establish. The relative importance of processes shaping genetic structure on different spatial and temporal scales in the marine realm remains unclear.

A key issue that impedes the unraveling of relative importance of isolating processes is that they are often spatially and/or temporally confounded. Islands are ideal systems to test for the modes of isolation, as they represent clearly defined spatial and temporal contexts. Islands have classically been used in ecological and evolutionary studies and have been at the basis of theories such as the Island Biogeography Theory. Anchialine systems such as marine lakes provide a marine analogue to oceanic islands by being bodies of seawater completely surrounded by land but maintaining a subterranean connection to the surrounding sea. Hence, they represent independent replicates of ecology and evolution. For this thesis, I selected marine lakes in Indonesia with a range of connection to the surrounding sea and variable local environments, but in the same spatial and temporal context. I was especially interested in the relative contribution of neutral (geographic and physical barriers to dispersal) versus selective (local environments) factors on the generation of genetic variation and differentiation. I used two types of genetic marker panels: a single marker (cytochrome oxidase subunit I) and a reduced representation genomic strategy (double digest restriction site-associated DNA sequencing, or ddRAD). Using reduced representation genomics allows for the generation of high-resolution neutral loci (giving information on demographic processes) and increases the chance of detecting potential loci under selection (giving information on adaptive potential). I complemented the strong genetic focus of this thesis with phenotypic measurements to assess associations between genetic adaptation and phenotypic acclimatization. For this thesis, I sampled three marine invertebrate species with different life histories: a jellyfish (*Mastigias papua*), a sponge

(*Suberites diversicolor*) and a mussel (*Brachidontes* sp.). My objectives were to 1) set environmental and biological baselines in the marine lakes of Indonesia, 2) assess how benthic communities and three invertebrates conform to expectations of Island Biogeography Theory, 3) test whether invertebrate populations show panmixia or if they show patterns of genetic structuring, and 4) quantify the relative importance of neutral and selective factors shaping population differentiation.

Chapter 2 presents a comprehensive overview of marine lakes in Raja Ampat, West Papua, Indonesia. I aimed to set ecological and biological baselines, test which factors best explained variation in benthic communities, and test hypotheses of species-area and species-isolation relationships following from Island Biogeography theory. I measured water quality parameters (temperature, salinity, pH and dissolved oxygen), and determined physical lake characteristics (i.e., depth, surface area, distance to sea, and connection). Benthic transects were performed to estimate coverage of broad benthic taxa (benthic cyanobacterial mats, bivalves, (hard) corals, crustose coralline algae, echinoderms, gastropods, macroalgae, polychaetes, sponges, and tunicates). I made qualitative diversity assessments for sponge, mollusc and fish communities. I identified 32 marine lakes, of which 17 were new to science. The lakes showed large variability in local environments (e.g., temperature ranging from 30.0° to 36.8°C) and the connection to the surrounding sea (relative tidal amplitude ranged from 5% to 89% compared to the sea). Three categories of marine lakes were apparent: lagoon-like lakes with high connection to the surrounding sea, highly isolated lakes with distinct local environmental conditions and depauperate communities, and intermediate lakes. I found a significant influence of temperature of the distribution of benthic group coverage. Connection to the surrounding sea and temperature were important predictors of qualitative diversity distributions. Richness of benthic groups increased with increasing connection to the surrounding sea and decreasing water temperatures. Lake area did not influence benthic group coverage or diversity. Hence, I found indications of species-isolation, but not for species-area relationships. I conclude that marine lakes can be seen as marine analogues of island-like systems and can be used in ecological and evolutionary studies.

In **chapter 3** I selected seven marine lakes in Raja Ampat, West Papua, containing dense populations of the golden jellyfish (*Mastigias papua*) as a case study to include peripheral ecosystems in Marine Protected Areas (MPAs). While these lakes receive increasing tourism attention, they are not yet included in conservation management plans. My aim was to provide a scientific basis to treat these 'jellyfish lakes' as individual management units. Between 2016-2018, I used the single genetic marker *COI* to determine phylogenetic position of jellyfish and quantify genetic variation between lakes and I performed morphometric analyses to assess morphological variation. I used qualitative and quantitative observations on jellyfish abundance over 11 years to log changes in population size. Finally, semi-structured stakeholder interviews were performed to assess lake use, customary tenure and perceived threats. I found strong genetic differentiation between lakes, including

one subspecies defined by previous research and potential unique subspecies. I also observed significant morphological differences between lakes. Major fluctuations in jellyfish abundance were observed, with no temporally consistent pattern across lakes. Interestingly, a morphological shift was observed for a jellyfish population in one lake after a crash in abundance, with no concordance in genetics, indicating potential phenotypic plasticity. Stakeholder interviews indicated that this lake was visited most by tourists. Identified threats included tenure disputes, conversion to aquaculture, introduced species and unregulated tourism. A 30-fold increase of tourism to Raja Ampat was observed since 2007. The increase in tourism and identified threats warrant the incorporation of jellyfish lakes into conservation management plans. Chapter 3 provides the scientific basis in the form of unique genetic and morphological variation to distinguish the lakes as individual management units.

In **Chapter 4** I take a population genomic approach to assess population structure and associated drivers of the sponge *Suberites diversicolor* and compare the results of genome-wide sequencing (ddRAD) to previously published results using single markers. I selected two coastal locations and six marine lakes on a broad geographic scale (<10 to >1,400km, East Kalimantan, West Papua and Australia) to test the relative influence of geography, physical dispersal barriers and local environments on genetic differentiation. Both low- and high-resolution markers were able to detect two major genetic lineages which might represent distinct species. However, with the high-resolution marker panel, I was able to provide new evidence of strong population structure within one of the lineages, even at spatial scales of <10km. Interestingly, I found no indications of geographic distance, permeability of dispersal barriers or local environments to play a role in shaping population genomic structure. Other processes are likely more important in determining structure of this poorly dispersing sponge. The discrepancy in genetic structure assumed for high- versus low-resolution markers perhaps calls for the reassessment of other benthic marine organisms where previous panmixia was reported based on low-resolution markers.

Selecting seven marine lakes in East Kalimantan and West Papua, I sampled the mussel *Brachidontes* sp. for ddRAD sequencing from seven marine lakes to assess whether isolation-by-distance, isolation-by-environment or historical contingency contribute most to population structure in **chapter 5**. Again, I found clear patterns of strong genetic differentiation among marine lake populations and limited gene flow. However, in contrast to the sponge, I found a clear isolation-by-distance pattern on both large (>1,400km) and small (<200km) spatial scales. Within West Papua (spatial scale <200km), this pattern was accompanied by an association of genetic divergence to the degree of connection to the surrounding sea. I conclude that the isolation provided by even incomplete barriers to dispersal could be enough to establish long-lasting population divergence, perhaps through eco-evolutionary dynamics such as priority effects.

In **chapter 6**, I explicitly tested the effects of ecological speciation versus neutral processes for *Brachidontes* sp. mussels by combining genomic and phenotypic data. Sampling nine

coastal ocean locations and 18 marine lakes in Raja Ampat, West-Papua, I aimed to test whether predictions of ecological speciation are met. I provide a unique dataset consisting of genomic data (ddRAD) and phenotypic data representing environmental acclimatization including morphological shell characteristics, mussel-associated microbial communities, and trophic niche space. Compelling evidence for ecological speciation would consist of concordance between genetic and phenotypic divergence. In contrast to most population genomic studies in the marine realm, I found strong clustering per marine lake regardless of only using neutral loci or also including loci under putative selection. I observed significant genetic differentiation among marine lakes, and panmixia for coastal ocean locations. Geographic location and connection to the surrounding sea were found to be important predictors of genomic variation. Outlier loci were detected mostly in comparisons to the most isolated lakes, which could point towards local adaptation. Yet no associations between genomic and phenotypic differentiation were found.

In the last chapter (**chapter 7**), I discuss the utility of marine lakes as insular systems and compare patterns observed to other population genomic studies in the marine realm. The population genomic data in this thesis consistently showed strong genetic structuring regardless of the marine invertebrate studied, and regardless of using neutral loci or loci under putative selection. I found no clear evidence of ecological speciation to underlie the observed structure, despite it being a mechanism of speciation thought to be particularly important in the marine realm. For peripheral systems such as marine lakes, apparently other processes have a stronger influence such as for example founder events. Even though initial diverging mechanisms may have been stochastic, it is remarkable that the patterns have persisted over thousands of generations, particularly as connection to the sea has remained. Hence, eco-evolutionary dynamics such as priority and monopolization effects may have effectively been fixed, underlining the often-neglected importance of historical contingency in shaping population genomic patterns. This thesis takes a first step in elucidating population genomic patterns of marine lake invertebrates. It paves the way for future studies to assess how insular systems can inform fundamental questions underlying generation and maintenance of biodiversity, and how they can function as models to assess conservation genetics and conservation.

SAMENVATTING

De oceanen van de wereld zijn uitgestrekt en schijnbaar grenzeloos. De afwezigheid van fysieke barrières heeft geleid tot de veronderstellingen dat mariene organismen een hoog dispersievermogen hebben en trage soortvorming vertonen. Deze veronderstellingen worden echter steeds meer weerlegd door studies die een hoge genetische structurering in mariene populaties aantonen, zelfs op kleine ruimtelijke schaal. Isolerende processen op kleine schaal kunnen daardoor een belangrijke rol spelen bij de vorming van genetische diversiteit binnen soorten. Isolatie vermindert effectieve uitwisseling van genetische variatie tussen populaties (gene flow genoemd), waardoor populaties kunnen divergeren. Isolatie kan bestaan uit processen die bijvoorbeeld dispersiebeperking of vestigingsbeperking veroorzaken. In dit proefschrift worden vier vormen van isolatie beschouwd, die zowel genetisch neutrale als selectieve processen omvatten. Neutrale processen van dispersiebeperking zijn bijvoorbeeld de geografische afstand tussen populaties (isolatie-door-afstand) en het gemak waarmee organismen zich door de matrix tussen populaties bewegen (isolatie-door-resistentie). Een selectief proces dat de vestiging van organismen beïnvloedt, is bijvoorbeeld ecologische soortvorming, of de beperking van gene flow tussen ecologisch verschillende habitats (isolatie-door-omgeving). Ten slotte kan succesvolle vestiging worden beïnvloed door historische eventualiteiten. Eco-evolutionaire dynamieken zoals prioriteitseffecten kunnen ertoe leiden dat de eerste kolonisatoren van een habitat de populatie structuur zodanig vormgeven dat latere immigranten zich er niet kunnen vestigen. Het relatieve belang van al deze processen voor de vorming van genetische structuur op verschillende ruimtelijke en temporele schalen is niet geheel duidelijk in het mariene rijk.

Een belangrijk probleem dat de ontrafeling van het relatieve belang van isolerende processen belemmert, is dat zij vaak ruimtelijk en/of temporeel moeilijk te scheiden zijn. Eilanden zijn ideale systemen om deze vormen van isolatie te testen, omdat zij een duidelijk afgebakende ruimtelijke en temporele context vertegenwoordigen. Eilanden worden al tijden gebruikt in ecologische en evolutionaire studies en vormen de basis van theorieën zoals de eilandbiogeografie. Anchialiene systemen zoals mariene meren kunnen een mariene analogie van oceanische eilanden zijn, omdat zij natuurlijke bassins gevuld met zeewater zijn die volledig door land omgeven zijn, maar een ondergrondse verbinding met de omringende zee behouden. Zij vertegenwoordigen dus onafhankelijke replica's van ecologie en evolutie. Voor dit proefschrift heb ik mariene meren in Indonesië geselecteerd die in verschillende mate in verbinding staan met de omringende zee. Ze hebben verschillende lokale omgevingen, maar bevinden zich in dezelfde ruimtelijke en temporele context. Ik was vooral geïnteresseerd in de relatieve bijdrage van neutrale (geografische en fysieke barrières voor dispersie) versus selectieve (lokale omgevingen) factoren aan het ontstaan van genetische variatie en differentiatie. Ik gebruikte twee types genetische merkerpanels: de enkele merker (cytochroomoxidase-subeenheid I, *COI*) en een strategie waarbij een subset van het hele genoom wordt getarget (double digest restriction site-associated DNA sequencing, of ddRAD). De genomische strategie maakt het mogelijk neutrale loci op hoge

resolutie te genereren (wat informatie geeft over demografische processen) en verhoogt de kans om potentiële loci onderhevig aan selectie te detecteren (wat informatie geeft over adaptief potentieel). Ik heb de sterke genetische focus van dit proefschrift aangevuld met fenotypische metingen om verbanden tussen genetische aanpassing en fenotypische acclimatisatie te testen. Voor dit proefschrift bemonsterde ik drie soorten mariene ongewervelden met verschillende soortspecifieke eigenschappen: een kwallensoort (*Mastigias papua*), een sponssoort (*Suberites diversicolor*) en een mosselsoort (*Brachidontes* sp.). Mijn doelstellingen waren: 1) het leggen van ecologische en biologische nulmetingen in de mariene meren van Indonesië, 2) nagaan in hoeverre benthische gemeenschappen en drie ongewervelden voldoen aan de verwachtingen uit de eilandbiogeografietheorie, 3) testen of ongewervelde populaties panmixie of patronen van genetische structuur vertonen en 4) het relatieve belang kwantificeren van neutrale en selectieve factoren die de populatiedifferentiatie bepalen.

Hoofdstuk 2 geeft een uitgebreid overzicht van de mariene meren in Raja Ampat, West-Papoea, Indonesië. Mijn doel was om ecologische en biologische uitgangspunten vast te stellen, te testen welke factoren de variatie in benthische gemeenschappen het beste verklaren en hypotheses te testen over de toe- of afname van de hoeveelheid soorten in relatie tot de grootte en de mate van isolatie van de mariene meren, volgend uit verwachtingen van de eilandbiogeografietheorie. Ik heb waterkwaliteitsparameters gemeten (temperatuur, zoutgehalte, pH en opgeloste zuurstof) en de fysische kenmerken van het meer bepaald (diepte, oppervlakte, afstand tot de zee, mate van verbinding). Benthische transecten zijn uitgevoerd om de bedekking van benthische taxa te schatten (benthische cyanobacteriële matten, echinodermata, gastropoden, koraalalgen, (harde) koralen, macroalgen, polychaeten, sponzen, tunicaten, en tweekleppigen). Ik maakte kwalitatieve diversiteitsbeoordelingen voor spons-, weekdier- en visgemeenschappen. Ik identificeerde 32 mariene meren, waarvan er 17 nieuw waren voor de wetenschap. De meren vertoonden een grote variabiliteit in de lokale omgeving (bijvoorbeeld een temperatuur variërend van 30,0 tot 36,8°C) en in de verbinding met de omringende zee (relatieve getijdenamplitude variërend van 5% tot 89% ten opzichte van de zee). Drie categorieën van mariene meren zijn onderscheiden: laguneachtige meren met een sterke verbinding met de omringende zee, geïsoleerde meren met diverse lokale milieuomstandigheden en verarmde biologische gemeenschappen en de meren die daartussen liggen. Ik heb een significante invloed van temperatuur op de distributie van de bedekking van benthische groepen waargenomen. De verbinding met de omringende zee en de temperatuur waren belangrijke voorspellers van kwalitatieve diversiteitmetingen. Het aantal benthische groepen nam toe naarmate de verbinding met de omringende zee toenam en de watertemperaturen afnamen. De oppervlakte van het meer had geen invloed op de bedekking van de bentische groepen of op de diversiteit. Ik vond dus aanwijzingen voor een effect van de mate van isolatie, maar niet voor een effect van de grootte van de meren. Ik concludeer dat mariene meren gezien kunnen worden als mariene analogieën van eilandsystemen en gebruikt kunnen worden in ecologische en evolutionaire studies.

Voor **hoofdstuk 3** heb ik zeven mariene meren in Raja Ampat, West-Papoea, met talrijke populaties van de gouden kwal (*Mastigias papua*) geselecteerd als casestudy om te kijken of perifere ecosystemen moeten worden opgenomen in zeereservaten. Hoewel deze meren steeds meer toeristische aandacht krijgen, zijn ze nog niet opgenomen in beheerplannen voor natuurbehoud. Mijn doel was om een wetenschappelijke basis te verschaffen voor de behandeling van deze 'kwallenmeren' als individuele beheereenheden. Tussen 2016 en 2018 gebruikte ik de genetische merker *COI* om de fylogenetische positie van kwallen te bepalen en de genetische variatie tussen meren te kwantificeren en voerde ik morfometrische analyses uit om morfologische variatie te bepalen. Ik gebruikte kwalitatieve en kwantitatieve waarnemingen van kwallenabundantie gedurende 11 jaar om veranderingen in populatiegrootte te loggen. Ten slotte zijn semigestructureerde interviews met belanghebbenden uitgevoerd om het gebruik van het meer, het eigendomsrecht en de waargenomen bedreigingen vast te stellen. Ik vond sterke genetische differentiatie tussen de meren, waaronder één ondersoort gedefinieerd door eerder onderzoek en potentieel andere unieke ondersoorten. Er waren ook significante morfologische verschillen tussen de meren. Grote fluctuaties in populatiegrootte zijn waargenomen, maar er was geen sprake van een consistent patroon tussen de meren. In één meer is een morfologische verandering in de kwallenpopulatie waargenomen voor en na een crash in populatiegrootte, terwijl de genetische structuur gelijk bleef, wat mogelijk wijst op fenotypische plasticiteit. Uit interviews met belanghebbenden bleek dat dit meer het meest werd bezocht door toeristen van de zeven kwallenmeren. Als bedreigingen werden genoemd: geschillen over eigendomsrecht, conversie naar aquacultuur, geïntroduceerde soorten en ongereguleerd toerisme. Sinds 2007 is het toerisme naar Raja Ampat met een factor 30 toegenomen. De toename van het toerisme en de bedreigingen hiervan rechtvaardigen de opname van individuele kwallenmeren in natuurbeschermingsplannen. In hoofdstuk 3 wordt op basis van de unieke genetische en morfologische variatie de stelling onderbouwd dat de meren als beheereenheden moeten worden onderscheiden.

Voor **hoofdstuk 4** heb ik de populatiestructuur van de spons *Suberites diversicolor* onderzocht aan de hand van een genomische benadering. De resultaten van genoombrede sequencing (ddRAD) worden er vergeleken met eerder gepubliceerde resultaten op basis van enkelvoudige genetische merkers. Ik selecteerde twee kustlocaties en zes mariene meren op een brede geografische schaal (<10 tot >1.400 km, Oost-Kalimantan, West-Papoea en Australië) om de relatieve invloed van geografie, fysieke dispersiebarrières en lokale milieus op genetische differentiatie te testen. Zowel lage- als hogeresolutiemarkers waren in staat om twee belangrijke genetische lijnen te detecteren die mogelijk verschillende soorten vertegenwoordigen. Met het hogeresolutiemerkerpanel kon ik echter nieuwe bewijzen leveren voor een sterke populatiestructuur in één van de lijnen, zelfs op ruimtelijke schalen van minder dan 10 kilometer. Het is opvallend dat ik geen aanwijzingen vond dat geografische afstand, doorlaatbaarheid van dispersiebarrières of lokale milieus een rol spelen bij de vorming van de genetische structuur van de populatie. Andere processen zijn

waarschijnlijk belangrijker om de structuur van deze spons met lage dispersiekwaliteiten te bepalen. De discrepantie tussen de uitkomsten van de hoge- en lageresolutiemarkers voor de veronderstelde genetische structuren vraagt wellicht om een herbeoordeling van andere benthische mariene organismen waar eerder panmixie werd gerapporteerd op basis van lageresolutiemarkers.

Om na te gaan of de populatiestructuur het meest wordt beïnvloed door isolatie-door-afstand, isolatie-door-omgeving of historische eventualiteiten bemonsterde ik de mossel *Brachidontes* sp. in zeven mariene meren in Oost-Kalimantan en West-Papoea voor ddRAD sequencing. De resultaten hiervan zijn opgenomen in **hoofdstuk 5**. Ook hier vond ik duidelijke patronen van sterke genetische differentiatie tussen populaties in mariene meren en patronen van beperkte gene flow. Echter, anders dan bij de spons, vond ik een duidelijk isolatie-per-afstandpatroon op zowel grote (>1.400km) als kleine (<200km) ruimtelijke schaal. Binnen West-Papoea (ruimtelijke schaal <200km) ging dit patroon gepaard met een associatie tussen genetische divergentie en de mate van verbinding met de omringende zee. Ik concludeer dat de isolatie door zelfs onvolledige barrières voor dispersie voldoende kan zijn om langdurige populatiedivergentie tot stand te brengen, misschien door eco-evolutionaire dynamieken zoals prioriteitseffecten.

Voor **hoofdstuk 6** testte ik expliciet de effecten van ecologische soortvorming versus neutrale processen op mosselen van de soort *Brachidontes* sp. door genomische en fenotypische gegevens te combineren. Door 9 kustlocaties en 18 mariene meren in Raja Ampat, West-Papoea, te bemonsteren, wilde ik testen of voorspellingen van ecologische soortvorming bevestigd konden worden. Ik heb een unieke dataset samengesteld, bestaande uit genomische informatie (ddRAD) en fenotypische informatie die omgevingsacclimatisatie kunnen weergeven, inclusief morfologische schelpkenmerken, mosselgeassocieerde microbiële gemeenschappen, en trofische nicheruimte. Een overtuigend bewijs voor ecologische soortvorming zou bestaan uit concordantie tussen genetische en fenotypische divergentie. In tegenstelling tot de meeste mariene populatiegenomische studies, vond ik sterke clustering per marien meer. Het maakte daarbij niet uit of uitsluitend neutrale loci werden gebruikt of ook loci potentieel onder selectie. Ik nam significante genetische differentiatie tussen de mariene meren waar, evenals panmixie voor de kustlocaties. De geografische locatie en de verbinding met de omringende zee bleken belangrijke voorspellers te zijn voor genomische variatie. Uitbijterloci werden vooral gedetecteerd in vergelijkingen tussen de meest geïsoleerde meren, wat zou kunnen wijzen op lokale adaptatie. Het is opvallend dat geen associaties tussen genomische en fenotypische differentiatie werden gevonden.

In het laatste hoofdstuk (**hoofdstuk 7**) bespreek ik het functioneren van mariene meren als insulaire systemen en vergelijk ik de waargenomen patronen met andere mariene populatie-genomische. De populatie-genomische data in dit proefschrift vertoonden consistent een sterke genetische structurering, ongeacht de bestudeerde mariene ongewervelde en

ongeacht het gebruik van neutrale loci of loci potentieel onder selectie. Ik vond geen duidelijk bewijs dat ecologische soortvorming ten grondslag ligt aan de waargenomen structuur, ondanks het feit dat dit een mechanisme van soortvorming is waarvan gedacht wordt dat het van belang is in mariene systemen. Andere processen hebben blijkbaar een sterkere invloed op perifere systemen als mariene meren, zoals stichtereffecten. Hoewel de initiële divergentiemechanismen stochastisch kunnen zijn geweest, is het opmerkelijk dat de patronen duizenden generaties lang zijn blijven bestaan, vooral omdat er waarschijnlijk een voortdurende de verbinding met de zee is geweest. Eco-evolutionaire dynamieken zoals prioriteit- en monopolisatie-effecten kunnen dus effectief gefixeerd zijn, wat het vaak verwaarloosde belang onderstreept van historische eventualiteiten gedurende de vorming van populatie-genomische patronen. Dit proefschrift zet een eerste stap in het belichten van populatie-genomische patronen van ongewervelden in mariene meren. Het effent de weg voor toekomstige studies om na te gaan hoe insulaire systemen fundamentele vragen kunnen beantwoorden over het ontstaan en behoud van biodiversiteit en hoe ze kunnen functioneren als modellen van conservatiegenetica en conservatiebeheer.

RINGKASAN

Alam laut terlihat tidak terbatas, dan tidak adanya hambatan fisik untuk penyebaran telah menimbulkan asumsi akan tingginya kemampuan penyebaran dan tingkat spesiasi organisme laut yang lambat. Baru-baru ini, asumsi ini digulingkan oleh penelitian yang menunjukkan struktur genetik populasi laut yang tinggi bahkan pada skala spasial yang kecil. Oleh karena itu, proses isolasi pada skala kecil dapat memiliki peran penting dalam membentuk keragaman genetik pada spesies. Isolasi mengurangi aliran gen efektif antar populasi, yang memungkinkan penyimpangan populasi. Isolasi dapat terdiri dari proses yang mempengaruhi batasan penyebaran atau batasan pembentukan. Dalam tesis ini, akan dibahas empat mode isolasi, termasuk proses netral dan selektif. Proses netral dari batasan penyebaran termasuk jarak geografis antar populasi (isolasi-oleh-jarak) dan permeabilitas matriks antar populasi (isolasi-oleh-resistensi). Proses selektif yang mempengaruhi pembentukan adalah spesiasi ekologis, atau pembatasan pada aliran gen antara habitat yang berbeda di lingkungan, yang mengakibatkan isolasi oleh lingkungan. Pada akhirnya, keberhasilan pembentukan juga dapat dipengaruhi oleh kemungkinan dalam sejarah. Di sini, dinamika eko-evolusioner dari efek prioritas dengan komponen adaptif yang potensial dapat mengakibatkan para penjajah pertama dalam suatu lingkungan membentuk struktur populasi berikutnya sedemikian rupa sehingga para imigran kemudian tidak dapat berhasil membentuknya. Kepentingan relatif dari proses yang membentuk struktur genetik pada skala spasial dan temporal yang berbeda di alam laut masih belum dapat dijelaskan.

Masalah utama yang menghambat penjelasan pentingnya proses isolasi yang relatif adalah bahwa proses tersebut sering kali membingungkan secara spasial dan/atau temporal. Pulau adalah sistem yang ideal untuk menguji mode isolasi, karena pulau dapat mewakili konteks spasial dan temporal yang terdefinisi dengan jelas. Pulau secara klasik telah digunakan dalam studi ekologi dan evolusi, serta telah menjadi dasar teori seperti Teori Biogeografi Pulau. Sistem Anchialine seperti danau laut memberikan analogi kelautan dengan pulau-pulau di samudra dengan menjadi badan air laut yang sepenuhnya dikelilingi oleh daratan tetapi mempertahankan koneksi bawah tanah dengan laut sekitarnya. Oleh karena itu, pulau tersebut mewakili replikasi independen ekologi dan evolusi. Untuk tesis ini, saya memilih danau laut di Indonesia dengan jangkauan koneksi ke laut sekitarnya dan lingkungan lokal yang bervariasi, tetapi dalam konteks spasial dan temporal yang sama. Saya secara khusus tertarik pada kontribusi relatif faktor netral (hambatan geografis dan fisik terhadap penyebaran) dibandingkan dengan faktor selektif (lingkungan lokal) pada generasi variasi dan diferensiasi genetik. Saya menggunakan dua jenis panel penanda genetik: penanda tunggal (subunit I sitokrom oksidase) dan strategi genomik representasi tereduksi (pengurutan double digest restriction site-associated DNA, atau ddRAD). Menggunakan genomik representasi yang direduksi memungkinkan pembentukan lokus netral dengan resolusi tinggi (memberikan informasi tentang proses demografis) dan meningkatkan kemungkinan dalam mendeteksi lokus yang potensial yang sedang diseleksi (memberikan informasi tentang potensi adaptif). Saya melengkapi fokus genetik yang kuat untuk tesis ini

dengan pengukuran fenotipik untuk menilai hubungan antara adaptasi genetik dan aklimatisasi fenotipik. Untuk tesis ini, saya mengambil sampel tiga spesies invertebrata laut dengan riwayat hidup yang berbeda: ubur-ubur (Mastigias papua), bunga karang (Suberites diversicolor), dan remis (Brachidontes sp.). Tujuan saya adalah untuk 1) menetapkan dasar lingkungan dan biologis di danau laut Indonesia, 2) menilai bagaimana komunitas bentik dan tiga invertebrata sesuai dengan harapan dalam Teori Biogeografi Pulau, 3) menguji apakah populasi invertebrata menunjukkan panmixia atau apakah mereka menunjukkan pola penataan genetik, dan 4) mengukur kepentingan relatif dari faktor netral dan selektif yang membentuk diferensiasi populasi.

Bab 2 menyajikan gambaran menyeluruh tentang danau laut di Raja Ampat, Papua Barat, Indonesia. Saya bertujuan untuk menetapkan dasar ekologi dan biologi, menguji faktor mana yang paling baik dalam menjelaskan variasi dalam komunitas bentik, dan menguji hipotesis hubungan antara spesies-area dan spesies-isolasi mengikuti teori Biogeografi Pulau. Saya mengukur parameter kualitas air (suhu, salinitas, pH dan oksigen terlarut), dan menentukan karakteristik fisik danau (yaitu kedalaman, luas permukaan, jarak ke laut, dan koneksi). Transek bentik dilakukan untuk memperkirakan cakupan taksa bentik yang luas (benthic cyanobacterial mats, bivalvia, karang keras, crustose coralline algae, echinodermata, gastropoda, makroalga, polychaetes, bunga karang, dan tunicate). Saya membuat penilaian keragaman kualitatif untuk komunitas bunga karang, moluska, dan ikan. Saya mengidentifikasi 32 danau laut, 17 di antaranya merupakan danau laut yang baru diketahui dalam ilmu pengetahuan. Danau menunjukkan variabilitas besar di lingkungan lokal (misalnya, suhu berkisar dari 30,0^{o} from sampai 36,8^{o}C) dan hubungan ke laut sekitarnya (amplitudo pasang surut relatif berkisar antara 5% sampai 89% dibandingkan dengan laut). Terdapat tiga kategori danau laut yang jelas: danau mirip laguna dengan koneksi tinggi ke laut sekitarnya, danau yang sangat terisolasi dengan kondisi lingkungan lokal yang berbeda dan komunitas yang rusak, dan danau intermediate. Saya menemukan pengaruh suhu yang signifikan dari penyebaran cakupan kelompok bentik. Koneksi dengan laut sekitar dan suhu merupakan prediktor penting dari penyebaran keragaman kualitatif. Kekayaan kelompok bentik meningkat dengan meningkatnya koneksi atau hubungan dengan laut sekitarnya dan penurunan suhu air. Kawasan danau tidak mempengaruhi cakupan atau keanekaragaman kelompok bentik. Oleh karena itu, saya menemukan indikasi isolasi spesies, tetapi tidak untuk hubungan spesies-area. Saya menyimpulkan bahwa danau laut dapat dilihat sebagai analog laut dari sistem mirip pulau dan dapat digunakan dalam studi ekologi dan evolusi..

Pada Bab 3, Saya memilih tujuh danau laut di Raja Ampat, Papua Barat, yang memiliki populasi ubur-ubur emas (Mastigias papua) yang padat sebagai studi kasus untuk memasukkan ekosistem perifer di Kawasan Konservasi Laut (KKL). Meskipun danau-danau ini mendapat perhatian pariwisata yang semakin meningkat, danau-danau tersebut belum dimasukkan dalam rencana pengelolaan konservasi. Tujuan saya adalah memberikan dasar ilmiah untuk memperlakukan 'danau ubur-ubur' ini sebagai unit manajemen individu. Antara

tahun 2016-2018, saya menggunakan penanda genetik tunggal COI untuk menentukan posisi filogenetik ubur-ubur dan mengukur variasi genetik antar danau dan saya melakukan analisis morfometri untuk menilai variasi morfologi. Saya menggunakan pengamatan kualitatif dan kuantitatif pada kelimpahan ubur-ubur selama 11 tahun untuk mencatat perubahan ukuran populasi. Terakhir, wawancara dengan pemangku kepentingan semi terstruktur dilakukan untuk menilai pemanfaatan danau, kedudukan adat dan persepsi ancaman. Saya menemukan perbedaan genetik yang kuat antara danau, termasuk satu subspesies yang ditentukan oleh penelitian sebelumnya dan subspesies unik yang potensial. Saya juga mengamati perbedaan morfologi yang signifikan antar danau. Fluktuasi yang besar dalam kelimpahan ubur-ubur diamati, dengan tidak ada pola yang konsisten di seluruh danau. Menariknya, pergeseran morfologi pada populasi ubur-ubur diamati di satu danau setelah rusaknya kelimpahan, tanpa adanya kesesuaian dalam genetika, yang menunjukkan potensi plastisitas fenotipik. Wawancara dengan pemangku kepentingan menunjukkan bahwa danau ini paling banyak dikunjungi oleh wisatawan. Ancaman yang teridentifikasi termasuk sengketa tenurial, konversi ke budidaya, spesies pendatang, dan pariwisata yang tidak diatur. Peningkatan pariwisata sebesar 30 kali lipat ke Raja Ampat telah diamati sejak tahun 2007. Peningkatan pariwisata dan ancaman yang teridentifikasi memerlukan masuknya danau ubur-ubur ke dalam rencana pengelolaan konservasi. Bab 3 memberikan landasan ilmiah berupa keragaman genetik dan morfologi yang unik untuk membedakan danau sebagai unit pengelolaan tersendiri.

Pada Bab 4, Saya mengambil pendekatan genom populasi untuk menilai struktur populasi dan penggerak terkait bunga karang Suberites diversicolor dan membandingkan hasil pengurutan seluruh genom (ddRAD) dengan hasil yang dipublikasikan sebelumnya menggunakan penanda tunggal. Saya memilih dua lokasi pesisir dan enam danau laut pada skala geografis yang luas (<10 hingga >1.400 km, Kalimantan Timur, Papua Barat dan Australia) untuk menguji pengaruh relatif dari geografi, penghambat penyebaran fisik dan lingkungan lokal pada diferensiasi genetik. Penanda dengan resolusi rendah dan resolusi tinggi mampu mendeteksi dua garis keturunan genetik utama yang mungkin mewakili spesies yang berbeda. Namun demikian, dengan panel penanda dengan resolusi tinggi, saya dapat memberikan bukti baru tentang struktur populasi yang kuat dalam salah satu garis keturunan, bahkan pada skala spasial <10 km. Menariknya, saya tidak menemukan indikasi jarak geografis, permeabilitas hambatan penyebaran atau lingkungan lokal yang berperan dalam membentuk struktur genom populasi. Proses lain mungkin lebih penting dalam menentukan struktur bunga karang yang menyebar dengan buruk ini. Perbedaan dalam struktur genetik yang diasumsikan untuk penanda dengan resolusi tinggi versus rendah mungkin memerlukan penilaian ulang pada organisme bentik laut lainnya di mana panmixia sebelumnya dilaporkan berdasarkan penanda dengan resolusi rendah.

Memilih tujuh danau laut di Kalimantan Timur dan Papua Barat, saya mencoba remis Brachidontes sp. untuk pengurutan ddRAD dari tujuh danau laut untuk menilai apakah isolasi oleh jarak, isolasi oleh lingkungan, atau kemungkinan sejarah berkontribusi paling

besar pada struktur populasi di bab 5. Sekali lagi, saya menemukan pola yang jelas dari diferensiasi genetik yang kuat di antara populasi danau laut dan aliran gen yang terbatas. Namun, berbeda dengan bunga karang, saya menemukan pola isolasi oleh jarak yang jelas pada skala spasial besar (> 1.400 km) dan kecil (<200 km). Di Papua Barat (skala spasial <200km), pola ini disertai dengan asosiasi divergensi genetik dengan derajat pada koneksi dengan laut di sekitarnya. Saya menyimpulkan bahwa isolasi yang diberikan bahkan oleh penghambat penyebaran yang tidak lengkap sudah cukup untuk membangun divergensi populasi yang bertahan lama, mungkin melalui dinamika eko-evolusioner seperti efek prioritas.

Pada Bab 6, Saya secara eksplisit menguji efek spesiasi ekologi dibandingkan dengan proses netral pada kerang Brachidontes sp. dengan menggabungkan data genomik dan fenotipik. Dengan mengambil sampel dari sembilan lokasi pesisir laut dan 18 danau laut di Raja Ampat, Papua Barat, saya bertujuan untuk menguji apakah prediksi spesiasi ekologi dapat terpenuhi. Saya menyediakan set data unik yang terdiri dari data genomik (ddRAD) dan data fenotipe yang mewakili aklimatisasi lingkungan termasuk karakteristik morfologi cangkang, komunitas mikroba terkait kerang, dan ruang relung trofik. Bukti kuat untuk spesiasi ekologi terdiri dari kesesuaian antara divergensi genetik dan fenotipik. Berbeda dengan kebanyakan studi genom populasi laut, saya menemukan pengelompokan yang kuat untuk danau laut terlepas dari hanya menggunakan lokus netral atau juga termasuk lokus di bawah seleksi yang diharapkan. Saya mengamati perbedaan genetik yang signifikan antara danau laut, dan panmixia untuk lokasi pesisir laut. Lokasi geografis dan koneksi ke laut sekitarnya ditemukan menjadi prediktor penting dari variasi genom. Lokus outlier sebagian besar terdeteksi dalam perbandingan dengan danau yang paling terisolasi, yang dapat mengarah ke adaptasi lokal. Namun tidak ditemukan adanya hubungan antara diferensiasi genomik dan fenotipik.

Dalam bab terakhir (Bab 7), saya membahas kegunaan danau laut sebagai sistem pulau dan membandingkan pola yang diamati dengan studi lain tentang genom populasi alam laut. Data genom populasi dalam tesis ini secara konsisten menunjukkan struktur genetik yang kuat terlepas dari invertebrata laut yang diteliti, dan terlepas dari penggunaan lokus atau lokus netral dalam seleksi yang diharapkan. Saya tidak menemukan bukti yang jelas tentang spesiasi ekologi yang mendasari struktur yang diamati, meskipun itu merupakan mekanisme spesiasi yang dianggap sangat penting di alam laut. Untuk sistem periferal seperti danau laut, ternyata proses lain memiliki pengaruh yang lebih kuat seperti misalnya peristiwa pembentukan. Meskipun mekanisme awal divergen mungkin bersifat tidak pasti, sungguh luar biasa bahwa pola tersebut telah bertahan selama ribuan generasi, terutama karena hubungan dengan laut tetap ada. Oleh karena itu, dinamika eko-evolusioner seperti prioritas dan efek monopolisasi mungkin telah diperbaiki secara efektif, menggarisbawahi pentingnya kemungkinan sejarah yang sering diabaikan dalam membentuk pola genom populasi. Tesis ini mengambil langkah awal dalam menjelaskan pola genom populasi invertebrata danau laut. Ini membuka jalan bagi studi di masa depan untuk menilai bagaimana sistem pulau

dapat menginformasikan pertanyaan mendasar yang mendasari generasi dan pemeliharaan keanekaragaman hayati, dan bagaimana mereka dapat berfungsi sebagai model untuk menilai genetika konservasi dan konservasi itu sendiri.

Appendices

About the author

Diede Maas was born on October 3rd, 1990 in Breda, the Netherlands. She went to Wageningen to pursue both a BSc and MSc in Biology, with specific focus on marine biology. She carried out major research projects during her Master. A major thesis together with Dr. Ronald Osinga of WUR and Jan Jansen of Ouwehands Zoo on the interactions of Acoela flatworms and a soft coral where she also looked at the effectiveness of biological pest controls. She went to Oviedo, Spain, for her minor thesis on mirror dimporphism and morphological diversity in the Portuguese Man-o-War, where she was supervised by Dr. Leo Nagelkerke from WUR and Dr. José Acuña from the University of Oviedo. Finally, for her internship Diede went to Naturalis to work on morphological population structure of marine lake mussels, under the supervision of Dr. Katja Peijnenburg from Naturalis, Dr. Rudi Roijackers from WUR, and Dr. Lisa Becking from then the University of Berkeley.

Diede's fascination with 'weird animals' (as Leo Nagelkerke kept referring to them) and interesting ecosystems continued during her PhD project. Dr. Lisa Becking had an opportunity in 2015 to continue the work on marine lakes but then with a more genetic focus. Lisa tried luring in Diede by promising work on the golden jellyfish, but it was not necessary. The marine lakes were interesting enough by itself, regardless of the organisms inhabiting them (although the jellyfish did become part of Diede's thesis!). The PhD focused on genetic population structure of different invertebrate species where several state-of-the-art genetic strategies were employed. Of course, Diede could not resist adding a phenotypic component to these studies. With this work, Diede aimed to better understand marine population genetics and connectivity to come closer to fundamental questions on how biodiversity arises and is maintained.

For the future, Diede will continue as a lecturer at the Marine Animal Ecology group. Here, her focus will be on education innovation and data analysis.

Publications

Maas, D.L., Capriati, A., Ahmad, A., Erdmann, M.V., Lamers, M., de Leeuw, C.A., Prins. L., Purwanto, Putri, 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. (**Chapter 3 this thesis**).

de Leeuw, C.A., Peijnenburg, K.T.C.A., 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.

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 despite incomplete barriers to dispersal. *Molecular Ecology* 27(7): 1556-1571. (**Chapter 5 this thesis**).

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

Training and education

With the training and education activities listed below, Diede Louise Maas has complied with the requirements set by the Graduate School of the Wageningen Institute of Animal Sciences (WIAS) which comprises of a total of 34 ECTs (1 ECT equals a study load of 28 hours).

TRAINING AND EDUCATION	Year	ECTs*
A. The Basic Package		2
WIAS Introduction Day, WUR	2015	
Research Integrity & Animal Ethics, WUR	2016	
B. Disciplinary Competences		13
WIAS Research Proposal, WUR	2015	
Population Genomics and Speciation, Czech Republic	2016	
Genomics course, WUR	2016	
Design of Experiments, WUR	2018	
C. Professional Competences		9
Language course Bahasa Indonesia, Yogyakarta	2015	
Presenting with Impact, WUR	2016	
Teaching and Supervising MSc students, WUR	2017	
Efficient Writing Strategies, WUR	2017	
Volunteering at Zoology conference + Chairing a session	2017	
Career Orientation	2018	
Organizing PhD writing ring		
D. Presentation Skills		4
ESEB, the Netherlands (poster)	2017	
NERN - NAEM, the Netherlands (poster)	2017	
World Sponge Conference, Ireland (poster)	2017	
Acroporanet, the Netherlands (oral presentation)	2017	
World Conference Marine Biodiversity, Canada (oral presentation)	2018	
E. Teaching Competences		6
Supervising BSc student (1)	2016	
Supervising MSc students (7)	2015-2020	
Designing and supervising Marine Animal Ecology course practical & tutorials	2017-2020	
Organizing and chairing BSc and MSc thesis rings	2017-2020	
Total (minimum of 30 ECTs)		**34**

*ECTs: European Credit Transfer and Accumulation System

Acknowledgements

Dear reader, you are now at the end of the thesis (or perhaps you started here?). This section is where I admit that in all actuality, I could not have made it through my PhD without the help of quite a bunch of people. What a journey it has been, both literally and metaphorically.

First and foremost, I would like to thank my co-promotors and promotor.

Lisa thank you for introducing me to the wonderful world of marine lakes already during my internship at Naturalis during my Master's, albeit then only remotely via the morphological analysis of a whole bunch of mussel shells. Thank you for thinking of me when you had an exciting PhD opportunity working on these fascinating systems - even despite my lack in genetic knowledge. I remember the very first trip we took to Indonesia in 2015 to lay the foundation for this project. I will never forget the seemingly endless karaoke songs at night or how incredibly nice that one hotel in Yogya was after our time in Raja Ampat (it smelled so much of cinnamon!). I admire you for your dedicated ambition and how you organize your network to be able to do this very interesting research. That is definitely something I can learn from you. Thank you for letting me take care of your cat Tasch now and then! Animal cuddles are the best when working on a PhD. Lastly, thank you for all the effort you put in this PhD project, even until the last moment.

Stefan, thank you for all the effort you put into trying to teach me how genomics works and how to assemble a genome (even though we did not use it in the end). It was fun to learn all the different analyses from you and together with you, and you always responded so quickly to my endless command line questions. Thank you for that. I am also really glad we got to meet in person, even if it was only once. I enjoyed listening to all your cool rescue worker stories. How you manage to combine high-end scientific work with all these extracurricular activities AND still help out PhD students is beyond me.

Tinka, thank you for allowing me to be the first 'real' MAE PhD student. What an interesting brief period of time that was, with just you, Lisa and me. How we have grown as a group, from fighting for our chairs to now a fully-fledged chair group with more than 30 people. I am very happy to be part of this group and I feel very lucky that you gave me the opportunity to be part of our team beyond my PhD. Although you were not highly involved in this project, I still would like to thank you for your input, creative thoughts and advice.

I also have a whole list of coauthors and collaborators to thank, because without you, this thesis literally would not exist. Some of you I will mention at other places in these acknowledgements, but for now I would like to extend my gratitude to the following people in alphabetical order.

Abdul, thank you for facilitating this research. I hope we get to meet some time.

Detmer, thank you for your contributions to the less visible part of mussels and sponges: their microbial communities.

Ellie, thank you for sharing your analysis strategies with me, I hope you will do well in your future career.

Emiliano, while you only joined the project at a relatively late stage, this is not reflected in the immense amount of support you gave me. Truly thank you for your clear explanations, your ideas on analyses and your effort. I can honestly say that also thanks to you I know a little bit better what I am doing. I hope we get to collaborate more in the future.

Ke, thank you for your *endless* patience in trying to explain command line, Python and Perl to me. Thank you for always responding quickly and thoroughly to my questions, even when you moved on to a different job.

Lydia, I unfortunately never got to meet you in person, but I indirectly thank you for providing Lisa such good lab support in order for me to work on great data!

Machiel, thank you for your much needed input from the social science angle, and **Amanda** thank you for doing the actual work interviewing stakeholders!

Mark, thank you for your valuable feedback on the jellyfish conservation manuscript. I'm glad I briefly got to meet you during my field work in 2016 where you were studying some very cool sharks. I hope you will continue to get to name a bunch of fish species in the future!

Martine, thank you for teaching the 'tri-RAD' method to me and hosting me at the University of Groningen. Every time I went to Groningen you made me feel very welcome. Please extend my warmest regards to the rest of the group (I guess most of the PhDs are finishing up as well!).

Purwanto, thank you for everything, and I am glad we will stay in touch via Lisa's work. I also greatly enjoyed the presentations you gave here in Wageningen.

Remco, thank you for helping me sneak morphological work in my thesis. Your patience in explaining the microCT scanner and camera settings both to me and my students is very much appreciated. I hope we stay in touch in the future.

Ricardo, thank you for all the effort you put into this thesis. I'm glad Christiaan and I got to travel to Manokwari to give some lectures. I admire you for your perseverance and I am sure we will stay in contact in the future.

Rosemary, thank you for your inspiring work on islands and island-like systems. Thank you for your comprehensive feedback and for making me aware how to come across best when submitting manuscripts to scientific journals. I'm very glad we got to meet.

Wawan, thank you for your help in the field and sunny attitude. You take the most beautiful pictures! I hope to see you again in the future.

Then, a big thanks to everyone in the MAE team! I can honestly say each and every one of you is amazing. Thank you for making me feel right at home in Zodiac-Radix(-Zodiac again?). In order of how I myself put you on our People page:

Maya, thank you for your support and organizing the fun MAE activities. You are an important, warm and sunny presence in our group.

Ronald, thank you for your humor and your calm attitude. I feel like we are great 'short-response' and 'witty remarks' buddies in meetings. Thank you as well for supervising me as an MSc student. I am so glad to have you in our group! Also, thank you for explaining to me what butter is.

Reindert, thank you for always allowing me my silly remarks and my unending enthusiasm for game-related stuff (Gathertown, HELL yeah). Thank you for involving me in your amazing AME course! Thank you for your endless excitement for the MinION machine (how many flow cells fit into a PromethION again?). I hope we get to work together more in the nearby future.

Rosa, our newest addition! You got to (digitally) meet me at the best possible time: during the final stretch of my PhD. I'm sorry for that. Thank you for your support and wise words (and illustrations) on this most fun time of a PhD life. Beyond that, thank you so much you applied for a job at MAE because I feel we will be a fantastic team in further innovating our education. MOOCs here we come!

Tim, thank you for your enthusiasm. It is always great to listen to you going on about your coral work with the latest technologies (and **Robbert-Jan** should also feel addressed here). Tim, I will never forget our lunches where you brought an assortment of vegetables every time (mostly cucumber, I believe) which you had to eat, much to your dismay. If you ever need someone to give your veggies to, you know where to find me.

Saskia, I am super stoked that you are part of the team after being one of our excellent students. I admire your calm attitude in life and strive towards adopting it myself more. I cannot wait to come over again, eat good food, talk about coding, play video games and pet Dude!

Edwin, thank you for your unrelenting positive outlook. In a different life I would have loved to be your PhD student, but instead I'll hope we get to collaborate more in the future. After all, how big of a step is it from marine lakes to marine mesocosms?

Dolfi, thank you for basically working on everything and for always asking such interesting questions during meetings. I'm excited to see what projects you will do next.

Martine, thank you for bringing the polar perspective in our group. It is much needed with all these tropical and North Sea subjects! Thank you also for always being cheerful in meetings.

Pauline, thank you for your efforts of restoring oyster reefs in the North Sea! I will never get used to your picture on the People page being a little crooked, though (I know, I know, it's to make the horizon horizontal, but still!). Thank you for being part of MAE.

Rene, although we don't see you often, thank you for being part of the team. I hope I get to send a lot of students your way to join your exciting projects!

Then. The PhD Band. The greatest and best band ever. So great, in fact, that we have not yet even *needed* to perform on stage. That's how great we are. Thank you for all the helpful comments, the dinners, the laughs, the activities (KARAOKE), and especially the Coca Cola's. Please accept my thanks in reversed alphabetical order (I must keep things fresh).

Sahri, thank you for being such an inspiration for the rest of us. You insane paper-writing machine! I am so happy for you that you managed to finish everything in time, and I hope you're having a blast back in Indonesia. Thank you so much for asking me to be your paranymph, I loved every moment of it. Thank you as well for all the Indonesian food you brought to thesis rings and provided us with in general (although I guess I should thank your wife for that).

Park, my Chinese brother. Please thank your wife for the lovely dinners she made us and thank you for being such a happy addition to our group. I always have fun when we talk, and I hope that will continue in the future. I feel like Sahri and you have always been such a calm, stable presence in our group - which is much needed!

Mischa. Gosh, where do I start. I am immensely grateful to have you in my life. Your contagiously positive attitude will be missed now you have moved on to AEW, but I'm sure our singing and whistling and screaming will not be missed by our roommates. I'm so glad we got to start together in this group. It wouldn't have been the same without you. Thank you for being there for me whenever I need. Thank you for all the walks with the Kafluny. Thank you for everything. I have learned a lot from you, and I am sure this will continue in the future. The world has not seen the last of Dikke Mik.

Mert, wow, I am amazed by how you can combine doing a PhD and being a great documentary maker at the same time. I admire your passion a lot, and I always love to listen to your stories. You will do your defense soon after me, so here's to us! I hope we get to hang out again in the future.

Niels, I always enjoy our talks, even if they are quite few and far between. I am curious to see how your PhD project will turn out!

Ludi, wow we do have some history together, don't we! From meeting during the 2016 field work of Christiaan and mine to now doing your own PhD! I am sure you will do great, and I am so glad to have you be part of MAE. Your unending positivity is inspiring. I hope we get to supervise a lot of students together in the future!

Karlijn, just you saying "heyyyyy" in meetings never neglects to make me smile. I absolutely love having you in our group. You're such a fun and passionate person, you always bring the right energy. I cannot wait to see what the rest of your PhD project will bring (please do remember to say no to things every now and then).

Joshua, our social science guy! I am in awe of how you always appear to be right at home. Thank you for providing new perspectives in our group. You always are happy, I like that!

Gibbs, skipper Gibbs! I'm glad you have also joined MAE and not only ABG. Your story-telling abilities are much appreciated. I hope you get to fully start your project soon! And that we can hang out again soon (with some game, mayhaps?).

Ewout, you're not in the Netherlands a lot (and I am so sorry on behalf of the Netherlands for the times that you *were* forced to be here), but every time you were I enjoyed your presence a lot. I hope you're having a lot of fun there in Kenya and I hope we get to collaborate in the future. If you do find yourself in the Netherlands at some point again, please do come over for some more Abzu!

Erik, mr. Fresh Eyes. Thank you for all the pizza nights, for eating WAY TOO MUCH asparagus (I promise to stop you in time next time), for all game knights (BROFORCE) and for all concerts we went to together. I cannot wait for this corona to be over so we can go to see some bands again and head bang (YOU MESSED WITH THE WRONG GENERATION). Let me apologize in black and white once again for all the weird songs and noises that I bother you with all the time (*Chickiiiiiin*). I know you hate being above water, and I am sure I do not make it better. Sorry.

And then we have **Christiaan**. Jij zieke vrucht. Een van de favoriete collega's. The very first time I met you was during my internship with Lisa, where you had to explain haplotype networks and other very abstract genetic stuff to me (admittedly, genetic stuff hasn't become that much less abstract during the PhD). I tried to be funny, you weren't having it. I thought that was the last time I would see you. But I am so incredibly glad you also joined the marine lake team and we got to go on this PhD adventure together. And boy, adventures we did have. From the rats that were *definitely not* present on Yellit, to the time I thought the python got you, but it turned out to be a wasp, to following courses abroad. I have learned a lot from you, and I greatly admire your critical way of thinking. You always manage to ask the right, difficult questions. I think we make a great team. I am very glad that you found a new exciting job that seems to fit you really well, but I refuse to let you out of my life, at the very least not until we get 100% platinum in Overcooked.

Celia, I have missed having you in the office. You always provided a great combination of working hard (making the rest of us work hard as well) and yelling OH MY GOD in several experiment-induced stress moments. I am so glad it all worked out in the end and having you be the first MAE student to do your defense! Thank you so much for asking me to be your paranymph, it was an honor. I can't wait to have dinner together again and talk about inappropriate subjects.

Annemiek, thank you for your enthusiasm! I cannot wait to start collaborating and work on those elasmobranchs! Let's see if we can quantify their behavior as we have envisioned it in our experimental design talks. I'm very glad to have you in our group.

Alwin, we did not get to meet often, sadly. Hopefully that changes in the future! In the meantime, I am very happy passing along students to your awesome projects!

Agustin. I could have placed you in so many categories in these acknowledgements. You have been a great student, a great co-author, and now I am sure you will deliver a great PhD thesis. I admire your work ethic, and always had fun having meetings with you. I heard you can come to the Netherlands soon, and I can't wait until we get to hang out! Thank you for all your work on the marine lakes project, my thesis would have been a lot less complete without you.

Finally, an additional shout-out to **Mischa** and **Erik**! Thank you for being my paranymphs and keeping me happy during this last stretch. You guys are the best.

I have alluded to it already in my thanks to Agustin, but my thesis would be a lot different without all the effort put into it by the best students ever! Let me thank you in the order I got to supervise you:

Maria, you were the first student I ever supervised, and boy did you set the bar high for the rest! You were such a sunny addition to the team and your work ethic was impeccable. I am so glad we got to learn so much together. From planning and performing fieldwork to bioinformatics. I enjoyed your company a lot. Thank you for the times you hosted me at your house. If I could, I would erase all butterflies and moths from existence for you (Robert Paauwe would also thank me for that), but I think they do have their place in the ecosystem.

Mark, thank you for your calm presence in the generally stressful world that is science. I was impressed with how you picked up all the new analyses. You helped me learn about them more as well.

Luca, thank you for all of your efforts and your incredible passion for jellyfish! Your motivation never wavered. Without you the paper on golden jellyfish lakes would not have been the same. I'm glad I could be your supervisor and am sure you will do well in your future career!

Olaf, the only Bachelor student I ever supervised! I hope you enjoyed the project and thank you for being part of it.

Joris, this is going to sound horrible, but I am sorry for having you spent so much time in the Zodiac basement (which also goes for Jennifer and Yinci). Joris, thank you for developing such a great basis for all my future morphological work. Thank you for our talks and our laughs. Thank you for actually taking the lead in the morphological work and explaining it so well to all future students. I wish you well and hope to see you again!

Jennifer, thank you for your unending enthusiasm about... I mean, actually everything, right? I always feel super energized after talking to you. You are one of the few people that gets so excited over mussels, and I'm glad I could provide you with a project on them (although they were a little too dead for your liking, maybe). Thank you for your support, and let's hold that game night soon, no?

Yinci, thank you for your incredible work ethic! I learned a lot from you. You were really tenacious in making sure all the analyses were just right and worked well. I am inspired by you. I hope you're doing well and also hope we get to meet again at some point.

Emma, it was a joy to work with you. Man, I could not have figured out those fastsimcoal analyses by myself, and you really held on until it worked. Well, it sort of worked, at least. We had good times together and I wish you all the best on Bonaire! Perhaps I'll get to visit you some time.

Merethe (and also a shout-out to **Stephanie**), thank you for your unrelenting work on mussel morphology and for all the memes. Thank you both as well for your hard work in the field! It was a great pleasure to work with you both!

Additionally, a big thank you to all the students whom I had to pleasure to have in my thesis ring. You all are great, and you have made me laugh consistently every Friday. Thank you for that!

Next, a massive thank you to my friends, obviously. You have been the best distraction from work one could ask for. You've reminded me that there's more to life than a PhD, while also giving me food for thought to incorporate in this thesis. In Excel-based randomly generated order (gosh these orders are getting more and more difficult), here are my personal thanks.

Sanne, thank you for being such a lovely constant in my life. From writing Harvest Moon manuals at primary school, to going to Ireland together (with our mums), to doing arts and crafts together as recently as this year. I am also thankful for all the lovely food we have eaten over the years. I admire your ambition and cannot wait until I get to hold your book in my hands and until I get to take a front seat at the gala premiere of your first musical. Thank

you for always being there for me and for all your wise words. I am extremely glad we stayed in touch over all these years, and I hope we will continue this streak far into the future.

Prinsesjes, also known as Hanneke, Myrthe and Sheryl. I am very glad the misery of high school gym class brought us together. It definitely formed a very strong bond. Thank you for all the holidays, dinners, games and just chilling in Hanneke's garden together. We held out pretty good in this online year. We had some nice online meetings drawing beavers (otters?) and watched some great bad movies. **Hanneke**, thank you for sharing a love for oldschool emo bands. Thank you for all the Wednesday dinners (we should pick those up again!) and thank you for all the good talks and laughs. At some point we should breed the hamster the size of an ottoman (apparently the English word for 'poef', simultaneously also a great name for the hamster: Lieutenant Ottoman!). I am so glad that we continue to live exactly 1 minute walking distance to each other. I may not see you in person every day, but I always look in your direction and give a little mental wave. **Myrthe**, firstly thank you for improving my Dutch in the summary, I did not realize how bad it had gotten. But most of all thank you for being such a good and thoughtful friend. Your dry sense of humor is impeccable, and I am so glad to have you in my life. Like with Sanne, I also cannot wait to have your book in my hands. If I could, I would delete all trees in springtime, apples, nuts and soybeans from the world for you. Thank you as well for never being able to distinguish Ryan Reynolds and Ryan Gosling, I love it. **Sheryl**, thank you for you! You always seem to just take life as it is, which is a great inspiration to me. Thank you for getting me through high school maths, science and physics. I will never forget the moment our maths teacher had to desperately search for a mistake to not give you a 10 (crazy). Thank you for setting up the game Pandemic so we could play it together remotely, one of the high points during corona for sure. Thank you for hosting such a nice Christmas dinner last year. Please also extend my thanks to Juno and Strip. Do let me know when I can come over to work on your house again (and also reconsider placing a jacuzzi). Prinsesjes, I hope we get to play mini golf together very soon again!

Wouter, thank you for being an awesome ACT teammate and staying in touch after. I hope the cinemas open at some point soon again so we can watch some action movie! Also thank you for generously hosting the guinea pigs. I miss them, and I will come give them pea flakes soon.

Speaking of GUINAS!! **Maaike**, thank you for sharing this VERY IMPORTANT AND RELEVANT passion! I am SO glad we got to work on the stadstuin together during ACT. As I've already said in that tv program in Brabant (what was its name again), for me it was love at first sight. Thank you for your wisdom, for all your help, for everything. I can always call you when I need, and I hope you know that you can always call me as well. Thank you for inspiring me to be a better person. I have no words to express how much I appreciate you, so I'll just divert the attention to silly things again (as you know I like to do). I hope we get to make a lot of Molcars together in the future, and I hope I never have to stop to ask you what kind of bird this or that is. Thank you as well for all the cracking sounds your back makes (you old lady). I cannot wait to see what the future will bring you (IEBELTJE PERHAPS?)

Eline, your outlook in life and how you move through it is truly inspiring. Always when I spend time with you, I am in total awe. I am also very proud to have some of your artwork hanging in my home. I cannot wait to see what other pieces you will make next. Thank you for everything, and also especially for keeping the Nakama group together. I enjoyed our times together! The Harry Potter city walk, the VR escape room (although you did not enjoy that too much...), getting jealous when you showed your Japan pictures, to name but a few. I hope we get to see each other soon again!

De Zwarte Hoek including Arnout, Jeroen, Sara, Niels, Maikel, Sherdina, Harm and Marieke. I am so glad to be part of this group! We had some really nice dinners together and weekends in Ameland. I am impressed as well on how we kept things going, even when we

were done with Uni. **Arnout**, thank you for all our silly talks about stuff I will not mention here. I hope we get to go to see that comedian together with Sara. **Jeroen**, or just Chef, of course. Thank you for all the game nights with Skittles and Halo. **Sara**, thank you for your beer knowledge (I recently discovered I like Kasteel Rouge!) and thank you for your spontaneity. Let's start the knutselclub soon! **Niels**, thank you for all the music you shared! I always enjoyed listening to Harm and you in particular speaking of politics. **Maikel**, thank you for always being there to lift me up. You never neglect to make me feel better about myself. I hope I do the same for you. I cannot wait to come visit Julia soon! **Sherdina**, thank you for always lighting up the entire room wherever you are. Thank you for your ability to make everyone (including me) feel right at ease when talking to you. You inspire me. **Harm** thank you for your kindness and calmness. Thank you for hosting us at some Dutch island or at your parents'! **Marieke** thank you for all our talks and walks together. Thank you for loving that same emo bands Hanneke and I love. Thank you for the Gilmore Girls nights!

Tom. Tomu. Desuka. Thank you for inventing a new language with me. Thank you for having conversations only consisting of gifs. Thank you for your passion when you talk about games, and immediately my apologies for not remembering everything. Please do not walk on thin ice anymore and video calling me in the middle of it. You never fail to make me laugh, and you also never fail to give me new insights on humanity. I loved playing Undertale together (that flipping Flowey). Let's find a new game to play together soon! AAAAA!

Robin, thank you for giving me the parakeet photo of your petting zoo! I look at it every day and it makes me happy. Thank you as well for the Mexican Games in the summer. You're so creative with them, and although I never win, I always enjoy. You are such a kind person, and I am very glad I got to meet you.

Romain, thank you for being such a consistent pen pal! I don't think a lot of people still receive letters on a regular basis, and I feel very lucky for our friendship (it does remind me that it's my turn writing a letter, oops). Apologies for delaying our letter correspondence so much during this hectic PhD. I will be better. I am very glad I got to meet you and your family last summer! It was very special, and you made me feel right at home. I cannot wait to see how Alice and Sunny will grow up (REINE DES NEIGES DEUUUUX). Please also give my regards to Gaëlle! You inspire me to work on my drawing more, since you're an absolute beast with how productive you are. Cannot wait to see what you will do next!

Then, the **GameBois!** Robert, Mariëlle, Nick, Buur, Ellard, Allison, Jessica, thank you for being the most awesome but also the most disturbing group I am happy to be part of. Thank you for all the events we have together, from holidays in spooky mansions to LAN parties, to just seeing how much food we can consume. **Robert**, thank you for your enthusiasm in helping me create the cover! I could not have done it without you. Your art inspires me, as well as your general thought processes (langpootmugpatronus?). **Mariëlle**, thank you firstly (because I just ate one) for introducing me to Humble & Crumble. My blood cholesterol levels are rising as we speak. Thank you for your cheerfulness and your advice on how to survive a PhD. We are finally both done! **Nick**, thank you for having so many boardgames we can play together! I cannot wait until we get to play some of them together soon again. Thank you as well for your kindness. Give my regards to Pietje! **Buur**, dankjewel voor je blije hoofd! Je bent altijd zo vrolijk, het is ongelofelijk. Dankjewel voor het altijd meteen fixen als er iets is in mijn huis (de koelkast bijvoorbeeld). Dankjewel ook dat ik je in paniek kon bellen toen ik mijn verwarming aan het ontluchten was maar niet realiseerde dat ik dan ook de kraan weer dicht moest draaien. Dankjewel voor al het lekkere eten wat je altijd voor ons maakt. Dank voor al je gezelligheid tijdens vakanties, concerten en spelletjesdagen/avonden. Ik hoop dat we elkaar snel weer zien voor bijvoorbeeld een barbecue in de tuin van je nieuwe huis! **Ellard**, oof. We had quite a run, haven't we? You've definitely seen me both at my best and at my worst during the PhD (and in life, to be honest). Thank you so much for sticking with me and always making me feel better. Thank you for

your passion, thank you for your humor. I am so glad to still have you in my life. Our regular walks together are always inspiring and fun, thank you for all your advice and help. Please say hi to Junior to me (at the day of my defense it's finally niet meiors maar juniors!). **Allison**, what NOT to thank you for? Thank you for always believing in me. Thank you for being my friend for such a long time already. From our first kip Tandoori together, to going to all kinds of concerts, to whizzing through practicals at University like it's nothing, to participating in several food studies together. Especially that last one was good for our bond (spoonful of butter, anyone?). Thank you for your yellow-faced budgies (and also those without a yellow face). I'm glad I got to budgie-sit them several times. I am also SUPER GLAD you and Robert are coming back to the Netherlands. I've missed you. You are an inspiration. **Jessica**, thank you for your creativity and for making me aware that De Beren exists (those burgers are yummy!). I hope I get to commission you to make a jellyfish painting again (I do not have enough in my home). **Bois**, my life would be a lot less fun (but also a lot less poop-talk-filled) without you. Thank you for everything.

Hier wil ik graag nog even wat ruimte innemen om de beste ouders en de beste broer ter wereld te bedanken. Jullie geloven altijd in mij en ik hou van jullie. Op volgorde van leeftijd (de laatste volgorde die ik kon bedenken):

Toon, dankjewel voor je immer wijze woorden. Dankjewel dat ik altijd naar Breda mag komen om gekke liedjes tegen je te zingen en op de grond van je appartement te liggen tijdens het gamen. Het voelt altijd als een soort mini-vakantie als ik er ben. Dankjewel ook voor al het lekkere eten! Bedankt voor al onze gesprekken over zeer uiteenlopende onderwerpen. Dankjewel ook dat je zo gek was om naar Wageningen te komen fietsen een keer! Sowieso vind ik het altijd erg gezellig als je langskomt en blijft slapen. Ik waardeer onze zondagwandelingen langs de brede A enorm (of was het toch zaterdag?). Snel weer doen? Misschien een keer gaan fietsen en dan met een pondje terug?).

Marijn, dankjewel voor je spontaniteit en creativiteit. Dat is echt iets wat ik van je geleerd heb en nog steeds aan het leren ben. Je geeft me het goede voorbeeld om meer mezelf te zijn. Ook jij dankjewel dat ik altijd langs mag komen, ook bij jou voelt het als een vakantie. Wat een verwennerij altijd met pokebowls, taartjes, chipjes. Ongelofelijk! Dankjewel ook dat je hebt besloten een hondje te nemen! Yuki is het zoetste hondje wat er bestaat en ik kan niet wachten om weer met jullie te gaan wandelen in het Oosterhouts bos (of in Wageningen, of ergens anders). Dankjewel ook **Kian** voor het delen van enthousiasme voor actiefilms en LEGO. Bedankt ook voor alle praktische tips over klussen en voor het ophangen van de laatste schilderijtjes in mijn huis.

Ludo, dankjewel voor al je verhalen! Wauw man, jij kan echt alles! Leonardo da Vinci? More like, LeoLuti da Maasi (hm werkt niet helemaal, ik ben niet zo goed met woorden als jij). Ik ben benieuwd wat je allemaal nog gaat doen in de toekomst (toch nog die PhD misschien?). Dankjewel voor alle grapjes en de oneindigende herinneren aan de enige echte Ap Snap the Greatest. Dank voor je aandachtsspan van ongeveer 20 seconden tijdens het spelen van games (al ging het laatst best goed!). Laten we binnenkort weer eens een poging doen om Comix Zone uitgespeeld te krijgen. Dankjewel voor het er altijd voor me zijn.

Ook **Appi** en **Kiki** zijn vast trots, terwijl ze aan het genieten zijn in de eeuwige bleekselderij- en trosgierstvelden. Ok, misschien zijn ze trouwens veel te druk met ehte.

For my last thanks, I would like to thank Nintendo, music, and puzzles for existing.

Let me close with a word of apology to all the mussels, sponges and jellyfish that were killed or harmed during this research. I'm sorry.

The research described in this thesis was financially supported by the Netherlands Organization for Scientific Research (NWO-VENI) (grant number #863.14.020), the Graduate School of the Wageningen Institute of Animal Sciences (WIAS) and Marine Animal Ecology (MAE).

Lay out and cover design by Diede Louise Maas
Chapter inset photographs by Christiaan A.t de Leeuw
Printed by ProefschriftMaken ll www.proefschriftmaken.nl