

## **Genomic diversity and connectivity of small giant clam (*Tridacna maxima*) populations across the Cook Islands**

Commissioned by the Ministry of Marine Resources, Government of the Cook Islands

May 2021

Liggins, L and Carvajal, J. I. (2021). Genomic diversity and connectivity of small giant clam (*Tridacna maxima*) populations across the Cook Islands. Report for the Ministry of Marine Resources, Government of the Cook Islands. 18 p.

### **Contact**

Dr. Libby Liggins

Senior Lecturer in Marine Ecology, Massey University Auckland

Research Associate, Auckland Museum, Tāmaki Paenga Hira

Diversity of the Indo-Pacific Network Team Member, DIPnet, <http://diversityindopacific.net/>

Director, Ira Moana – Genes of the Sea – Network, <https://sites.massey.ac.nz/iramoana/>

Massey University Auckland

School of Natural & Computational Sciences

Room 5.06, Building 5, Oteha Rohe Campus

Albany, Auckland 0745

New Zealand

Email: [L.Liggins@massey.ac.nz](mailto:L.Liggins@massey.ac.nz)

Telephone: +64 21 082 823 50

## EXECUTIVE SUMMARY

**Liggins, L and Carvajal, JI (2021). Genomic diversity and connectivity of small giant clam (*Tridacna maxima*) populations across the Cook Islands. Report for the Ministry of Marine Resources, Government of the Cook Islands. 18 p.**

This report contributes preliminary results for a study of the genomic diversity and population connectivity of the small giant clam (*Tridacna maxima*) in the Cook Islands. Single Nucleotide Polymorphisms (SNPs) were generated using a Genotype-By-Sequencing (GBS) approach to recover genome-wide, multilocus genotypes for *T. maxima* individuals sampled from ten island populations. Several individuals morphologically identified as *Tridacna noae* (Noah's clam) were also included, to confirm whether they were from a taxon distinct from *T. maxima*. The SNP data was analysed to describe the genomic diversity and patterns of genomic differentiation (or connectivity) among island populations of *T. maxima*, with a focus on the genomic relationship between the populations of Aitutaki and Manuae.

Our preliminary results suggest that the individuals identified as *T. noae* are a distinct taxon. Verification of the identity of this taxon as *T. noae* will be possible based on the mitochondrial DNA sequences being generated by our project collaborator. Population genomic analysis of *T. maxima* revealed that the island population of Manihiki was most differentiated from the other sampled islands, and the island population of Palmerston Atoll likely provides a stepping-stone population between Manihiki and the southern island populations. The southern group of island populations of *T. maxima* were not greatly genetically differentiated and likely have regular gene flow among them. In particular, the island populations of Aitutaki, Manuae, Atiu, and Takutea are most genetically similar. All the island populations of *T. maxima* had a heterozygote (i.e. genetic diversity) deficit, and high estimated values of inbreeding, although this was less the case for Manihiki and Palmerston Atoll. Such patterns of genetic diversity, and inbreeding values, may be indicative of small or reduced population sizes, owing to harvesting pressure.

Further analyses focused on the genomic diversity and connectivity among *T. maxima* populations of the Cook Islands are planned. These analyses will investigate the levels of genetic diversity and inbreeding, and whether they correlate with local clam density, harvest history, or reef type, as well as the meta-population structure (i.e. gene flow pattern, source-sink dynamics) among the southern group of islands to inform fisheries management.

## SCOPE OF REPORT

In collaboration with the Ministry of Marine Resources, several research questions relevant to the conservation and fisheries management of *Tridacna* clams in the Cook Islands were proposed (see Research Questions below). All research to address these questions has been coordinated among the collaborating research groups, so as to maximise the research outputs and understanding of *Tridacna* clams in the Cook Islands. This report addresses Q2, but also provides preliminary information regarding Q1, and genotypes that can be included in analyses to inform Q3. Herein we describe the methodological approach taken to address Q2 (primarily), our results to date, and the insights from these preliminary analyses. Furthermore, we outline further planned analyses and anticipated insights (see Future Directions of the Study below).

## RESEARCH QUESTIONS

### **Q1. What *Tridacna* clam species are in the Cook Islands, and what genetic diversity is present within these species of the Cook Islands relative to neighbouring regions of the Pacific?**

**Rationale:** Despite *Tridacna* clams being widespread throughout the tropical regions of the Indo-Pacific Ocean, their importance to several cultures as a food source and treasure, and regulation in their harvest and trade according to national and international legislation (e.g. Wells et al. 1983, Wells 2006) - several cryptic species have been described in over recent decades. In some cases, these ‘new cryptic species’ are widespread and co-exist with other *Tridacna* species over small spatial scales. For instance, since the first description of *T. noae* in the Solomon Islands (Huelsen et al. 2013) the species has been recorded in Western Australia, Indonesia, and New Caledonia. In all cases, these records of new *Tridacna* species for a region have relied upon the use of genetic markers to verify morphological identification. As part of our collaboration, we are using genetic markers to verify the morphological identifications of *Tridacna* species resident in the Cook Islands.

Despite most *Tridacna* species being widespread, genetic studies based on mitochondrial DNA (mtDNA) have described distinct clades (genetic lineages) within species, highlighting regions of highly restricted gene flow (e.g. De Boer et al. 2008, 2014, Huelsen et al. 2013). The geographic extent of these clades within species, helps us to understand the historical population (or stock) structure of these clams, identifying which populations are interdependent. Furthermore, the relative levels of genetic diversity, and unique genetic diversity within these populations, provide a measure of ‘value’ for each population in the conservation or management of a particular clade, or species. Recent initiatives have consolidated all existing genetic data and metadata for reef associated marine organisms in the Indo-Pacific, including *Tridacna* clams (i.e. Diversity of the Indo-Pacific Network, DIPnet, <http://diversityindopacific.net/>). These efforts now provide the data necessary to quickly contextualise any genetic data created for these species in new, previously understudied regions of the Indo-Pacific. To contextualise the clades and genetic diversity of *Tridacna* clams of the Cook Islands, and their relationships to other neighbouring regions of the Pacific, we will analyse the mtDNA data generated by project collaborators alongside existing mtDNA data from throughout the Indo-Pacific.

### **Q2. What is the genomic diversity and connectivity of small giant clam (*Tridacna maxima*) populations across the Cook Islands? In particular, what level of gene flow is likely between Aitutaki and Manuae?**

**Rationale:** Over smaller spatial scales and within national jurisdictions, genetic diversity is an important consideration for conservation prioritisation and fisheries management. As a unit of

biodiversity analogous to species diversity, genetic diversity has been evidenced to increase the productivity and resilience of populations (Aguirre et al. 2012) as well as that of communities and ecosystems (Reusch et al. 2005, Bernhardt and Leslie 2012). Genetic diversity also serves as a measurable proxy for other desirable demographic attributes and evolutionary processes (Laikre et al. 2010, Hoban et al. 2020). For example, genetic diversity typically correlates with population size (Hare et al. 2011) and in some systems, signals high connectivity, both of which contribute to the resilience of a population or metapopulation. Furthermore, genetic diversity can increase the resilience of a population to future challenges, providing the raw genetic material for increased adaptive potential within a population (Barret and Schluter 2008, Sgro and Hoffmann 2011).

The conservation or fisheries value of a spatial location based on genetic data is usually qualified by genetic diversity. Genetic diversity will increase in a population over time through the processes of mutation accumulation (higher in larger populations), and immigration from other populations. Genetically diverse locations are often assumed to be demographic sources and sources of genetic diversity, however immigration into sinks from more than one population can also elevate genetic diversity (Liggins et al. 2015). In contrast, genetic diversity will decrease due to drift (i.e. the stochastic loss, or change in frequency, of genetic variants within a population; stronger in small populations), population bottlenecks (such as through fishing), and/or selection in favour of a particular genetic variant rather than diversity per se. In this way, low genetic diversity populations may include rare, or unique genetic variants that may be locally adapted. For these reasons, measures reflecting the relationship among populations (i.e. genetic differentiation, covariance and gene flow) should also be considered alongside measures of genetic diversity.

Genetic diversity and differentiation patterns are influenced by a combination of interacting factors such as: differences and changes in (effective) population size, demographic and colonisation history, and natural selection. For these reasons, the direct interpretation of spatial genetic patterns in terms of gene flow among spatial locations should be cautious (Whitlock and McCauley 1999, Hart and Marko 2010, Lowe and Allendorf 2010, Marko and Hart 2011, Karl et al. 2012). In the present report, we have estimated genetic distinctiveness using measures of genetic differentiation, clustering methods, and multivariate discrimination analyses. In general, a differentiated or distinct population/group of populations will likely have low genetic connectivity (low gene flow), but we intend to complement these inferences with further planned analyses (see Future Directions of the Study).

Within the Cook Islands there is a particular interest in source-sink relationships among island populations of *Tridacna* (particularly the most common species, *T. maxima*) that can inform local conservation and fisheries management of stocks. Of most urgency, is understanding the levels of gene flow between Aitutaki and Manuae, and to a lesser extent, between Atiu and Takutea. In both cases, the former island is inhabited by humans and has a history of clam harvest, whereas the latter island is not inhabited by humans, but clam populations are under increasing clam harvest pressure. In both cases, to inform management, there is interest in whether the clam populations are interdependent and which are important sources of clam recruits to other islands. Determining patterns of genetic diversity and connectivity over these small spatial scales, requires higher-resolution genetic markers from throughout a species genome. Recent advances in laboratory techniques, and sequencing technologies, now enable such genome-wide, multilocus genotypes to be derived *de novo*, without requiring existing genomic information for the species (Elshire et al. 2011, Willette et al. 2014). In the present study we use one of these techniques, Genotype-By-Sequencing, to create a population genomic dataset appropriate to address Q2.

**Q3. What are the zooxanthellae symbiont communities of *Tridacna* in the Cook Islands? Are the zooxanthellae symbiont communities differentiated according to *Tridacna* host species, or other geographic, or environmental determinants across the Cook Islands?**

**Rationale:** *Tridacna* clams host an algal symbiont (called zooxanthellae, *Symbiodinium*) that help to fulfil their nutritional requirements through photosynthesis. These zooxanthellae communities (of clams and corals) have been described to differ depending on the host species, depth, and environmental conditions (such as heat stress), for example. The characterisation of the zooxanthellae communities of *Tridacna* clams of the Cook Islands is being coordinated by our project collaborator, but the genotypes of the *T. maxima* derived herein, will also form the basis of an analysis exploring whether a clam's genotype may have a role in determining the symbiont zooxanthellae community it supports (for more explanation see Future Directions of the Study).

## METHODS

### Sampling

The Ministry of Marine Resources provided clam mantle samples preserved in ethanol from ten islands within the Cook Islands including: Manihiki - the only island from the northern group; Palmerston - intermediary to the northern and southern group; and Aitutaki, Manuae, Takutea, Mitiaro, Atiu, Mauke, Mangaia and Rarotonga - all from the southern group of islands. Included in these samples were *Tridacna squamosa* (the fluted clam) and *Tridacna noae* (Noah's clam), but the majority were *T. maxima* (the small giant clam). The focus of the present study was *T. maxima*, but based on the difficulty in morphologically identifying giant clam species, six individuals of *T. noae* were included in the genomic analysis in order to confirm their morphological identification as a discrete taxon from those individuals identified as *T. maxima*. From each island population of *T. maxima*, between 25 and 50 individuals were chosen for genetic laboratory work.

### Genetic laboratory methods

Genomic DNA was extracted from 369 *T. maxima* and 6 *T. noae* samples using the Qiagen DNeasy Blood and Tissue Kit following the manufacturers protocols for animal tissue with the following modifications: the lysis step was conducted overnight at 56°C with 20µl Proteinase K added; after digestion, 3ul of RNase A (Monarch) was added and samples were incubated for 10 minutes at 37°C; and to maximize DNA yield, the elution step using TE buffer was repeated (each 50 ul in volume) providing a DNA extraction volume of 100 ul. Extractions were initially assessed for quality by running 2µl of DNA on a 1% agarose gel and DNA concentrations were measured using a Qubit 2.0 Fluorometer (Fisher Scientific). Samples with DNA concentrations of less than 10 ng/µl were concentrated where possible or re-extracted from remaining tissue.

### Genotyping

Single Nucleotide Polymorphisms (SNPs) were generated using a Genotyping-by-Sequencing (GBS) approach. Enzyme and adapter optimisation, size selection, library preparation, and NGS sequencing was conducted by the Elshire Group (Manawatu, New Zealand). Three restriction enzymes were tested to optimize the GBS result by digesting 7.2 ng of Illumina adapter and 500 or 200ng of DNA from one *T. maxima* extraction. Either, ApeKI, PstI, or EcoT22i were used to digest the DNA and the digest was run on a fragment analyzer. Based on these results, EcoT22i was selected as the best restriction enzyme for the GBS experiment. Library preparation was conducted for 375 DNA extractions using 100ng of genomic DNA and 1.44 ng of adapter. Genotype-by-sequencing followed the Elshire et al. (2011) method with the following specifications: genomic DNA was fragmented using EcoT22i restriction enzyme and the library was amplified with 18 PCR cycles, yielding 26-34 ng/µl of pooled libraries. Samples were sequenced on 4 lanes of 2 x 150bp Illumina HiSeq.

The GBS data was demultiplexed with AXE and trimmed for adapter and barcode sequences using the `batch_trim.pl` script from <https://github.com/Lanilen/GBS-PreProcess>. All reads were trimmed to a fixed 64 bp length using Skewer (Jiang et al. 2014) with no quality filtering thresholds and discarding reads shorter than 64bp as recommended by the Stacks pipeline. Next, the trimmed reads were put through the standard Stacks 2.5 pipeline (Rochette et al. 2019) to build loci using the `denovo_map` function. The 'populations' part of the Stacks workflow was run with a minor allele frequency filter of 5% and the resulting SNP file was exported as a VCF and Fasta file. To ensure that we were including loci from the target organism and not the endosymbiont, we removed any loci that mapped to *Symbiodinium*. Fasta sequences from the Stacks output were mapped to the *Symbiodinium* genome (Clade A and C; Aranda et al. 2016 and Liu et al. 2018) using the NGM mapping tool (Sedlazeck et al. 2013). Mapped loci were removed from the dataset using VCFtools (Danecek et al. 2011).

The SNP file was further filtered using the Radiator R package and VCFtools. Individuals with more than 80% missing data were removed. Loci with more than 80% missing data, average read depths <14 or >60 and minor allele counts of <20 were all removed. From this dataset of 17,446 loci, one SNP per locus was retained to ensure that the SNPs were not linked. Another version of this dataset was created to further explore relationships among the southern islands by excluding *T. noae* samples and all individuals from Palmerston and Manihiki. This version of the dataset had the same number of loci but only 303 individuals.

### **Population genomic analysis**

We calculated several genetic summary statistics for the sampled island populations (and taxa), including measures of genetic diversity and pairwise genetic differentiation ( $D_{est}$ ,  $G''_{ST}$  and  $\phi'_{ST}$ ; Weir and Cockerham 1984) using Stacks (Rochette et al. 2019) and the R package mmod (Winter 2012).

Patterns of genetic differentiation among *T. maxima* individuals were examined using two complementary approaches. First, to establish whether there was any significant genetic differentiation according to island populations, we undertook Discriminant Analysis of Principal Components (DAPC) in the R package Adegenet (Jombart 2008). DAPC is a multivariate method for clustering genetically related individuals. It performs a Principal Components Analysis (PCA) and then a Discriminatory Analysis (DA) to discriminate individuals into prior groups (i.e. island populations) by maximising the between group variance and minimising the within group variance. In our DAPC analysis, we used island population (and identification as *T. noae*) as our prior. We determined 120 to be the best number of PCs to retain in the analysis using a cross validation with 30 iterations.

Second, to identify any patterns of genetic differentiation that were not related to island population, we undertook a STRUCTURE analysis (Pritchard 2000). STRUCTURE is a model-based clustering method to infer population structure and assign individuals to these clusters. These types of Bayesian inferences can be very slow on large datasets, so a subset of the data is normally used. We ran STRUCTURE twice on a random subset of 4,000 SNPs from our complete dataset and again on our reduced dataset of just *T. maxima* from the southern islands. Both analyses were run for 15 replicates (8,500 MCMC chains with 1,500 discarded as burn-in) that tested which number (from 1 to 10) of clusters (k) best fitted the data.

## RESULTS

### Genotyping

A total of 977 million paired end reads were sequenced, with an average of 2.4 million reads per sample (including no failed samples and 6 samples with less than 10% of the average number of reads). Following quality control and filtering, all 376 individuals from across the sampling design were retained for analysis (370 *T. maxima* and 6 *T. noae*). The dataset had an average of 38% missing data across the 17,466 loci.

### Population genomic analysis

A DAPC clearly distinguished the individuals morphologically identified as *T. noae* as genetically distinct from all individuals identified as *T. maxima* (Figure 1). The pairwise population genetic differentiation statistics (Table 1) confirmed a much higher level of differentiation of these *T. noae* individuals from all *T. maxima* island populations (e.g.  $\phi'_{ST} = 0.26 - 0.30$ ), than was identified among the *T. maxima* island populations ( $\phi'_{ST} = 0.01 - 0.08$ ). Based on the population summary statistics (Table 2), *T. noae* also had much lower estimates of genetic diversity (e.g. % polymorphic loci, *T. noae* = 11.71) than found in any of the populations of *T. maxima* (% polymorphic loci, *T. maxima* island populations = 86.16 - 97.57). As the GBS method was optimised for the detection of genetic diversity within *T. maxima*, these results are likely a further indication that these individuals are a distinct species. Accordingly, the *T. noae* individuals were excluded from all subsequent population genomic analysis.

Focusing on the *T. maxima* individuals, the DAPC identified three distinct genetic clusters (i.e. groupings or stocks, Figure 1). The most distinct genetic group comprised individuals from Manihiki, the only population from the northern islands of the Cook Islands. A second distinct genetic grouping consisted of individuals sampled from Palmerston Atoll - an island that sits intermediary to the northern and southern island group of the Cook Islands, albeit much closer to the south. The third genetic group comprised individuals from all of the southern island populations. The STRUCTURE analyses (Figure 2) indicated that the *T. maxima* individuals of Palmerston, were intermediary in their genetic composition to the southern island populations and Manihiki, suggesting a stepping-stone role of the Palmerston population. Accordingly, the pairwise genetic differentiation statistics between Palmerston and all other island populations were typically intermediary values (i.e. for  $D_{est}$  and  $G''_{ST}$ , coloured mid-green) relative to values between Manihiki and the southern island populations (coloured darker green, according to values, Table 1).

A subsequent STRUCTURE analysis of just the southern island populations nominated that two distinct genetic groups could be detected, but these did not align with island geography (Figure 3). Initial Principal Components Analyses (PCA, not shown) also did not reveal any association of these two distinct genetic groups with any procedural biases (i.e. lab or genotyping methods/mistakes), habitat (i.e. depth, water temperature, macroalgal cover, live coral cover) or age (using clam size as a proxy). Similarly, although the DAPC supported the distinction of Manihiki and Palmerston from all other island populations, none of the southern island populations could be discriminated (Figure 1). These results suggest the island populations of *T. maxima* in the southern Cook Islands are a large meta-population that have recently, or regularly exchanged gene flow. Future planned analyses will further investigate the meta-population structure of these islands (see Future Directions of the Study).

Based on genetic differentiation statistics (Table 1), the island population of Takutea is least differentiated from all other island populations. In particular, *T. maxima* clams of Takutea are most genetically similar to the island populations of Aitutaki, Manuae, and Atiu (e.g.  $D_{est} = 0.02$ ). These four geographically proximal island populations are most similar in their genetic composition ( $D_{est}$

between 0.02 and 0.03). In increasing order of genetic differentiation from this cluster of four islands, is the island population of Rarotonga, Mitiaro, Mangaia and Mauke (followed by Palmerston and Manihiki). These results suggest that genetic differentiation increases over increasing geographic distance (i.e. geneflow is limited over larger geographic distances, Wright 1943). However, not all values of genetic differentiation among the island populations reflect their geographic proximity – for instance, although the island population of Mangaia is most similar to the nearby island of Rarotonga, it is equally similar to the *T. maxima* population of Takutea, a relatively distant island ( $D_{\text{est}} = 0.02$  in both cases). These results suggest there may be other important determinants of gene flow not apparent from the island geography alone (e.g. ocean currents, source-sink dynamics; Liggins et al. 2014, 2016, Riginos et al. 2016, Liggins et al. 2019).

Across all island populations of *T. maxima*, there was a heterozygote deficit (i.e. the difference between the expected and observed heterozygosity, Table 2). Heterozygosity (also often called gene diversity) indicates how much genetic variation there is in the population, and how this variation is distributed across the alleles in the loci of individuals. Despite high values being expected across all island populations, observed values were low. This deficiency of heterozygotes could be a consequence of inbreeding. The inbreeding coefficient,  $F_{\text{IS}}$ , assesses genetic variation in individuals, relative to the variation in their population. In particular, the inbreeding coefficient is highest in the island populations of Aitutaki, Manuae, Rarotonga, Atiu, and Takutea, and relatively low in Palmerston and Manihiki. These high  $F_{\text{IS}}$  values could be as a consequence of reduced population sizes due to harvesting pressure on these wild populations.

## FUTURE DIRECTIONS OF THE STUDY

**We will confirm the identity of the genetically distinct taxon, morphologically identified as *Tridacna noae*.** Although our analyses confirm that the individuals morphologically identified as *T. noae* are a distinct taxon, verification of the species identity will be possible from the mtDNA sequencing being conducted by project collaborators (in relation to Q1).

**For *T. maxima*, patterns of adaptive diversity will be characterised with further analyses.** Within our genome-wide SNP dataset, some loci may be under selection, or linked to regions of the genome that are under selection. Distinguishing the patterns of diversity and differentiation due to selective processes, such as genotype-environment associations with reef type or environmental variables, will additionally help to refine estimates of gene flow. We will use two methods of outlier loci detection to identify putatively adaptive loci and directly look for association of particular loci with habitat and environmental data (based on sampling metadata and remote sensed data) using Redundancy Analyses (RDA; see Liggins et al. 2019 for more information).

**Directionality of gene flow, source-sink, and stepping-stone relationships among island populations will be characterised for *T. maxima*.** Analyses to estimate asymmetrical migration among some of the island populations will be conducted based on research priorities (i.e. among Aitutaki and Manuae, Q2) and informed by prior analyses suggesting genetic connectivity. Estimates of migration are computationally intensive when using a large number of SNPs, but our prior analysis (in this report) and further planned analyses of genetic covariance among populations will inform a minimal set of island populations to include in order to gain a view of genetic connectivity across the island populations. Specifically, the genetic covariance among populations, and the minimum set of codependent island populations will be depicted as a population graph. Population graphs are a network created using graph theory with nodes which represent locations connected by edges which represent the genetic covariance between populations at those locations (Dyer and Nason 2004). In this way, the relationships among island populations can be examined and visualised simultaneously rather than through a number of pairwise relationships (as in this report). In the case of *T. maxima* in the Cook Islands, we anticipate that only populations that have had recent gene flow, and/or represent source-sink relationships would retain edges between them. Furthermore, several graph metrics (e.g. betweenness and centrality) can be used to indicate the relative roles of island populations as sources, and stepping-stones for other populations (Urban et al. 2009), helping to highlight their relative importance for conservation or fisheries management.

**The genetic connectivity of *T. maxima* of the Cook Islands with neighbouring regions of the Pacific will be derived.** The mtDNA sequences generated by our project collaborator will be combined with existing mtDNA datasets (e.g. as provided by DIPnet; and Liggins and Arranz, 2018), and available in the publicly accessible Genomic Observatory Metadatabase (GEOME, [www.geome-db.org](http://www.geome-db.org); Deck et al. 2017, Riginos et al. 2020). The compiled data will be analysed to describe the genetic diversity, genetic uniqueness, and genetic distinctiveness of the Cook Islands *Tridacna* clams and populations relative Pacific locations.

**The genotype-community relationship between *T. maxima* and zooxanthellae communities will be examined.** Where possible, the same individuals that were genotyped using SNPs were included in the characterisation of the zooxanthellae communities being led by our project collaborator. Using the multilocus SNP genotype of each *T. maxima* and their zooxanthellae community profiles, we can explore whether the *T. maxima*'s genotype, or any particular loci, determine the community of zooxanthellae its supports in its tissues.

**Table 1.** Genetic differentiation statistics for studied *Tridacna maxima* populations in the Cook Islands, and the included *Tridacna noae* individuals (*T. noae*). Island populations include: AIT Aitutaki, ATU Atiu, MAN Manuae, MHX Manihiki, MIT Mitiaro, MKE Mauke, MNG Mangaia, PLM Palmerston, TAK Takutea, RAR Rarotonga. **a.**  $\phi'_{ST}$ , **b.**  $D_{est}$ , **c.**  $G''_{ST}$ . Higher levels of differentiation indicated by deeper shades of green.

**a.**

	AIT	ATU	MAN	MHX	MIT	MKE	MNG	<i>T. noae</i>	PLM	RAR	TAK
AIT	0	0.02	0.02	0.07	0.02	0.03	0.02	0.26	0.03	0.02	0.01
ATU	0.02	0	0.02	0.07	0.02	0.03	0.02	0.26	0.03	0.02	0.02
MAN	0.02	0.02	0	0.07	0.02	0.03	0.02	0.26	0.03	0.02	0.01
MHX	0.07	0.07	0.07	0	0.07	0.08	0.08	0.30	0.06	0.06	0.06
MIT	0.02	0.02	0.02	0.07	0	0.03	0.03	0.27	0.04	0.02	0.02
MKE	0.03	0.03	0.03	0.08	0.03	0	0.04	0.29	0.04	0.03	0.03
MNG	0.02	0.02	0.02	0.08	0.03	0.04	0	0.28	0.04	0.02	0.02
<i>T. noae</i>	0.26	0.26	0.26	0.30	0.27	0.29	0.28	0	0.27	0.26	0.26
PLM	0.03	0.03	0.03	0.06	0.04	0.04	0.04	0.27	0	0.03	0.03
RAR	0.02	0.02	0.02	0.06	0.02	0.03	0.02	0.26	0.03	0	0.01
TAK	0.01	0.02	0.01	0.06	0.02	0.03	0.02	0.26	0.03	0.01	0

**b.**

	AIT	ATU	MAN	MHX	MIT	MKE	MNG	<i>T. noae</i>	PLM	RAR	TAK
AIT	0	0.03	0.02	0.09	0.03	0.04	0.03	0.26	0.05	0.03	0.02
ATU	0.03	0	0.03	0.09	0.03	0.04	0.04	0.25	0.05	0.03	0.02
MAN	0.02	0.03	0	0.09	0.03	0.04	0.03	0.25	0.05	0.02	0.02
MHX	0.09	0.09	0.09	0	0.09	0.11	0.10	0.26	0.08	0.09	0.08
MIT	0.03	0.03	0.03	0.09	0	0.05	0.05	0.26	0.05	0.03	0.03
MKE	0.04	0.04	0.04	0.11	0.05	0	0.05	0.27	0.06	0.04	0.04
MNG	0.03	0.04	0.03	0.10	0.05	0.05	0	0.26	0.06	0.04	0.03
<i>T. noae</i>	0.26	0.25	0.25	0.26	0.26	0.27	0.26	0	0.25	0.25	0.25
PLM	0.05	0.05	0.05	0.08	0.05	0.06	0.06	0.25	0	0.05	0.04
RAR	0.03	0.03	0.02	0.09	0.03	0.04	0.04	0.25	0.05	0	0.02
TAK	0.02	0.02	0.02	0.08	0.03	0.04	0.03	0.25	0.04	0.02	0

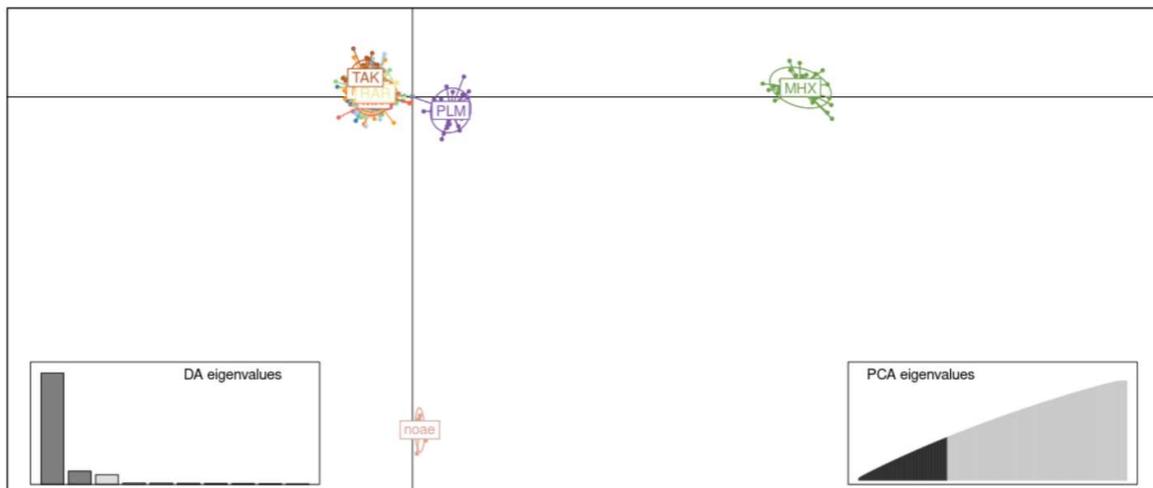
**c.**

	AIT	ATU	MAN	MHX	MIT	MKE	MNG	<i>T. noae</i>	PLM	RAR	TAK
AIT	0	0.06	0.06	0.20	0.07	0.10	0.08	0.57	0.10	0.06	0.05
ATU	0.06	0	0.06	0.21	0.08	0.09	0.08	0.57	0.11	0.06	0.05
MAN	0.06	0.06	0	0.20	0.08	0.09	0.08	0.56	0.11	0.06	0.05
MHX	0.20	0.21	0.20	0	0.21	0.24	0.23	0.60	0.19	0.20	0.19
MIT	0.07	0.08	0.08	0.21	0	0.11	0.10	0.58	0.13	0.08	0.07
MKE	0.10	0.09	0.09	0.24	0.11	0	0.12	0.60	0.14	0.09	0.09
MNG	0.08	0.08	0.08	0.23	0.10	0.12	0	0.59	0.13	0.08	0.07
<i>T. noae</i>	0.57	0.57	0.56	0.60	0.58	0.60	0.59	0	0.57	0.56	0.56
PLM	0.10	0.11	0.11	0.19	0.13	0.14	0.13	0.57	0	0.11	0.10
RAR	0.06	0.06	0.06	0.20	0.08	0.09	0.08	0.56	0.11	0	0.05
TAK	0.05	0.05	0.05	0.19	0.07	0.09	0.07	0.56	0.10	0.05	0

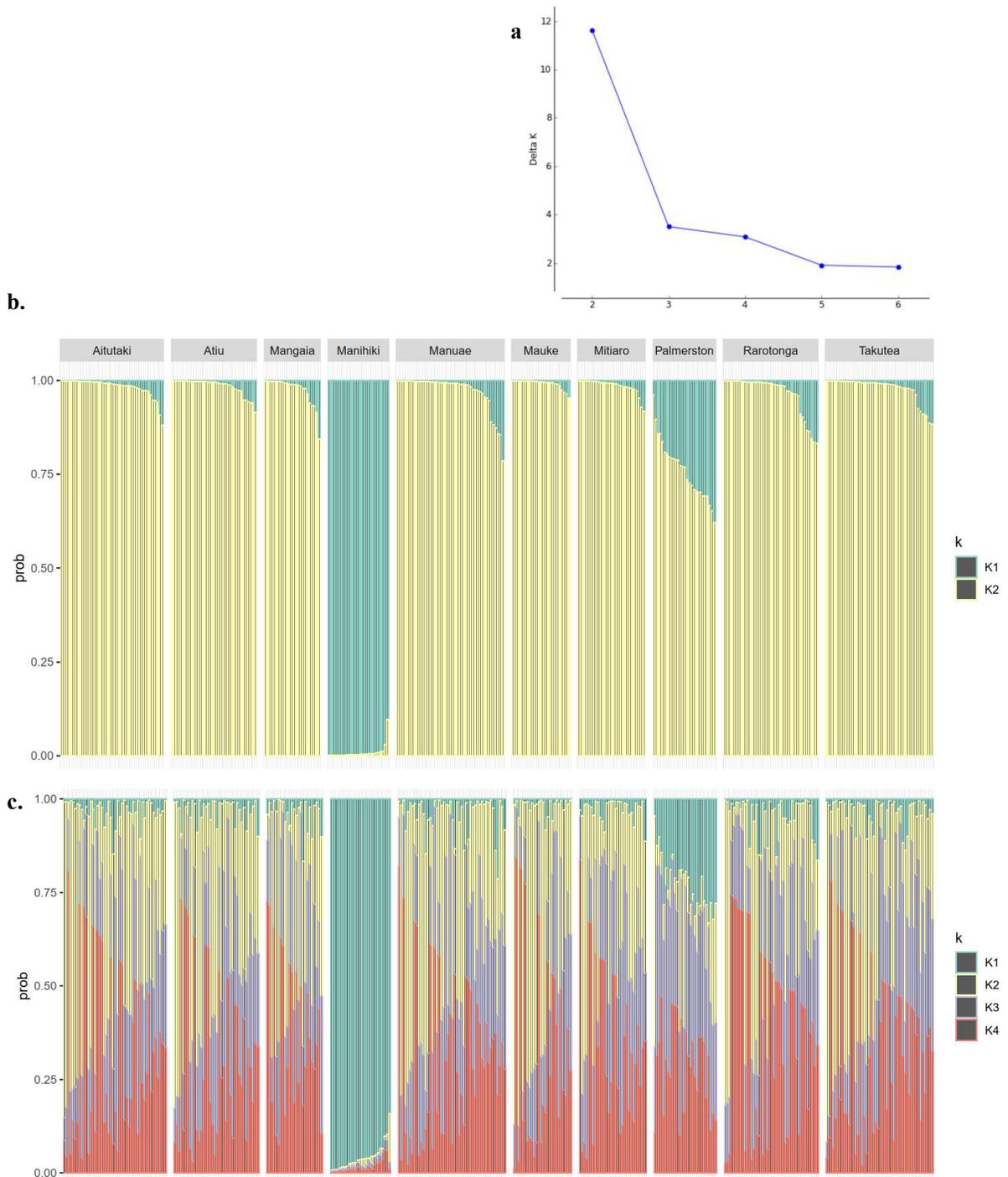
**Table 2.** Genetic summary statistics for studied *Tridacna maxima* populations in the Cook Islands, and the included *Tridacna noae* individuals (*T. noae*). Island populations include: AIT Aitutaki, ATU Atiu, MAN Manuae, MHX Manihiki, MIT Mitiaro, MKE Mauke, MNG Mangaia, PLM Palmerston, TAK Takutea, RAR Rarotonga. Summary statistics include: the mean number of individuals represented per locus; number of polymorphic sites; the % polymorphic loci; mean frequency of the major allele at each locus, P; mean observed heterozygosity; mean observed homozygosity; mean expected heterozygosity; mean expected homozygosity; mean value of  $\pi$ ; and the inbreeding coefficient,  $F_{is}$  (including standard errors).

Summary statistic	Island population										<i>T. noae</i>
	AIT	ATU	MAN	MHX	MIT	MKE	MNG	PLM	TAK	RAR	
Number of Individuals	23.30	18.84	24.37	15.50	14.66	12.75	12.79	14.47	25.18	21.29	3.46
Std. Error	0.09	0.08	0.10	0.06	0.06	0.05	0.05	0.06	0.10	0.08	0.03
Polymorphic Sites	21507	21314	21480	18735	20597	19089	20253	20240	21740	21490	574
% Polymorphic Loci	97.57	96.66	97.42	86.16	93.62	89.28	92.20	92.34	98.52	97.44	11.71
P	0.69	0.69	0.69	0.74	0.70	0.71	0.70	0.71	0.68	0.69	0.97
Std. Error	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Obs. Heterozygosity	0.18	0.18	0.18	0.17	0.18	0.17	0.18	0.18	0.19	0.18	0.04
Std. Error	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Obs. Homozygosity	0.82	0.82	0.82	0.83	0.82	0.83	0.82	0.82	0.81	0.82	0.96
Std. Error	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Exp. Heterozygosity	0.40	0.40	0.40	0.33	0.38	0.37	0.38	0.37	0.41	0.40	0.04
Std. Error	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Exp. Homozygosity	0.60	0.60	0.60	0.67	0.62	0.63	0.62	0.63	0.59	0.60	0.96
Std. Error	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$\pi$	0.42	0.42	0.42	0.35	0.41	0.40	0.40	0.40	0.42	0.42	0.06
Std. Error	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$F_{is}$	0.56	0.54	0.54	0.45	0.51	0.49	0.50	0.49	0.55	0.55	0.02
Std. Error	0.09	0.08	0.10	0.06	0.06	0.05	0.05	0.06	0.10	0.08	0.03

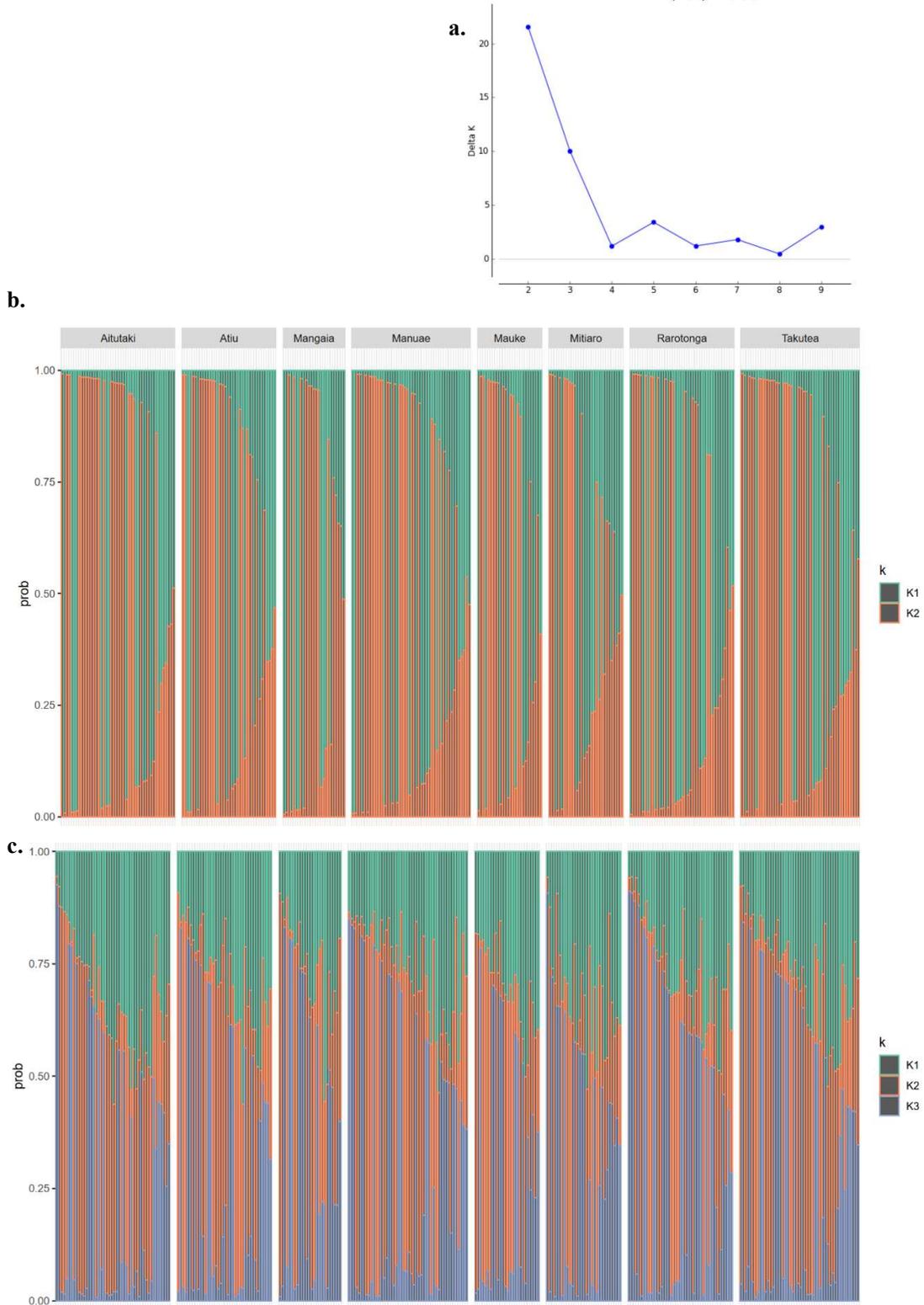
**Figure 1.** Discriminant Analysis of Principal Components distinguishing the island populations (and taxa) based on the multilocus SNP genotypes of individuals (points) morphologically identified as *T. maxima* and *T. noae*. Individuals identified as *T. noae* (noae) were genetically distinct from all samples identified as *T. maxima*. The analysis also reveals the island population of Manihiki (MHX) was distinct from the island population of Palmerston Atoll (PLM, also genetically distinct), and all southern island populations (including AIT Aitutaki, ATU Atiu, MAN Manuae, MIT Mitiaro, MKE Mauke, MNG Mangaia, , TAK Takutea, RAR Rarotonga). These southern island populations were not genetically distinct from each other based on this analysis.



**Figure 2.** STRUCTURE analyses distinguishing the major genetic groups within *T. maxima* based on the multilocus SNP genotypes of individuals. **a.** This analysis identified two distinct genetic groupings within *T. maxima*. **b.** Each vertical bar represents an individual clam (sorted according to island population of origin, top) and its probability of being assigned to each genetic grouping (colour). All individuals of Manihiki had a high probability of being assigned to the first genetic grouping (K1), and all individuals of Palmerston Atoll also had some probability of being assigned to this genetic grouping. Overall, all individuals of other island populations, including Palmerston Atoll, had a higher probability of being assigned to the second genetic grouping (K2). **c.** The probability of belonging to a different (third or fourth) genetic grouping did not improve for any population, when the number of genetic groups was increased.



**Figure 3.** STRUCTURE analyses distinguishing the major genetic groups within the southern island populations of *T. maxima* based on the multilocus SNP genotypes of individuals. **a.** This analysis identified two distinct genetic groupings within these island populations that did not correspond to island geography. **b.** The number of individuals assigned to each genetic grouping was similar across island populations. **c.** The probability of belonging to a different (third) genetic grouping did not improve for any population, when the number of genetic groups was increased.



## **DATA ACCESSIBILITY**

Upon completion of the study, sample metadata will be uploaded to the Genomics Observatories Metadatabase (GEOME, [www.geome-db.org](http://www.geome-db.org)) and linked to the genetic and genomic data in appropriate repositories. At the discretion of the Ministry of Marine Resources we can apply Traditional Knowledge Notices and Biocultural Notices to these genetic resources and derived data (see: <https://localcontexts.org/notices/biocultural-notices/>; Liggins et al. 2021). Curated tissue samples, metadata and analysed datasets are available from Dr. Libby Liggins by request.

## **PERMIT INFORMATION**

The Ministry of Marine Resources of the Cook Islands were responsible for all required permissions for the collections. Import of CITES-listed species (Appendix II, *Tridacna* spp.) into New Zealand was allowed by the New Zealand Department of Conservation in accordance with the New Zealand Import Health Standard and Certificates from the Government of the Cook Islands for the export from a non-party to a party (Certificate Numbers: CK/2020 16146, CK/2020 16147, CK/2020 16148).

## LITERATURE CITED

- Aguirre JD, Marshall DJ. 2012. Genetic diversity increases population productivity in a sessile marine invertebrate. *Ecology* 93: 1134–42.
- Aranda M, Li Y, Liew YJ, Baumgarten S, Simakov O, Wilson MC, Piel J, Ashoor H, Bougouffa S, Bajic VB, Ryu T. 2016. Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Scientific Reports* 6:1-5.
- Barrett RD, Schluter D. 2008. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* 23:38–44
- Bernhardt JR, Leslie HM. 2013. Resilience to climate change in coastal marine ecosystems. *Annual Review of Marine Science* 5:371–392
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G. 2011. The variant call format and VCFtools. *Bioinformatics* 27:2156-8.
- DeBoer TS, Ambariyanto, Baker, Erdmann MV, Barber PH. 2014. Concordant phylogenetic patterns inferred from mitochondrial and microsatellite DNA in the giant clam *Tridacna crocea*. *Bulletin of Marine Science* 90:301–329.
- DeBoer TS, Subia MD, Ambariyanto, Erdmann MV, Kovitvongsa K, Barber PH. 2008. Phylogeography and limited genetic connectivity in the endangered boring giant clam across the Coral Triangle. *Conservation Biology* 22: 1255-1266.
- Deck J, Gaither MR, Ewing R, Bird CE, Davies N, Meyer C, Riginos C, Toonen RJ, Crandall ED. 2017. The Genomic Observatories Metadatabase (GeOME): A new repository for field and sampling event metadata associated with genetic samples. *PLoS Biology* 8: e2002925.
- Diversity of the Indo-Pacific Network, DIPnet. <http://diversityindopacific.net/>
- Dyer RJ, Nason JD. 2004. Population graphs: the graph theoretic shape of genetic structure. *Molecular ecology* 13:1713-27.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379.
- Genomic Observatories Metadatabase, GEOME. [www.geome-db.org](http://www.geome-db.org)
- Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples RS, Palstra F. 2011. Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology* 25:438–449.
- Hart MW, Marko PB. 2010. It's about time: divergence, demography, and the evolution of developmental modes in marine invertebrates. *Integrative and Comparative Biology* 50: 643–661.
- Hoban S, Bruford M, Jackson JD, Lopes-Fernandes M, Heuertz M, Hohenlohe PA, Paz-Vinas I, Sjögren-Gulve P, Segelbacher G, Vernesi C, Aitken S. 2020. Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved. *Biological Conservation* 248:108654.
- Huelsken T, Keyse J, Liggins L, Penny S, Trembl EA, Riginos C. 2013. A novel widespread cryptic species and phylogeographic patterns within several giant clam species (Cardiidae: *Tridacna*) from the Indo-Pacific Ocean. *PLoS One* 8:e80858.
- Ira Moana – Genes of the Sea – Project, <https://sites.massey.ac.nz/iramoana/>
- Jiang H, Lei R, Ding SW, Zhu S. 2014. Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15:1-2.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24:1403-5.

- Karl SA, Toonen RJ, Grant WS, Bowen BW. 2012. Common misconceptions in molecular ecology: echoes of the modern synthesis. *Molecular Ecology* 21: 4171–4189.
- Laikre L. 2010. Genetic diversity is overlooked in international conservation policy implementation. *Conservation Genetics* 11:349–354
- Liggins L, Arranz V. 2018. Genetic diversity and connectivity of Beveridge Reef's marine biodiversity in an Indo-Pacific wide context. Report commissioned by The Department of Agriculture, Forestry and Fisheries of Niue. <https://doi.org/10.6084/m9.figshare.12287069.v1>
- Liggins L, Booth DJ, Figueira WF, Treml EA, Tonk L, Ridgway T, Harris DA, Riginos C. 2015. Latitude-wide genetic patterns reveal historical effects and contrasting patterns of turnover and nestedness at the range peripheries of a tropical marine fish. *Ecography* 38: 1212–24.
- Liggins L, Gleeson L, Riginos C. Evaluating edge-of-range genetic patterns for tropical echinoderms, *Acanthaster planci* and *Tripneustes gratilla*, of the Kermadec Islands, southwest Pacific. 2014. *Bulletin of Marine Science* 90: 379–97.
- Liggins L, Treml EA, Possingham HP, Riginos C. 2016. Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. *Journal of Biogeography* 2: 256–67.
- Liggins L, Treml EA, Riginos C. 2019. Seascape genomics: contextualizing adaptive and neutral genomic variation in the ocean environment. In: *Population genomics: Marine organisms* (pp. 171–218). Springer, Cham.
- Liggins L, Hudson M, Anderson J. Creating space for Indigenous perspectives on access and benefit-sharing: Encouraging researcher use of the Local Contexts Notices. *Molecular Ecology*: <https://onlinelibrary.wiley.com/doi/10.1111/mec.15918>
- Liu H, Stephens TG, González-Pech RA, Beltran VH, Lapeyre B, Bongaerts P, Cooke I, Aranda M, Bourne DG, Forêt S, Miller DJ. 2018. Symbiodinium genomes reveal adaptive evolution of functions related to coral-dinoflagellate symbiosis. *Communications Biology* 1:1–11.
- Lowe WH, Allendorf FW. 2010. What can genetics tell us about population connectivity? *Molecular Ecology* 19: 3038–3051.
- Marko PB, Hart MW. 2011. The complex analytical landscape of gene flow inference. *Trends in Ecology & Evolution* 26: 448–456.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–59.
- Reusch TB, Ehlers A, Hämmerli A, Worm B. 2005. Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences of the United States of America* 8: 2826–31.
- Riginos C, Crandall ED, Liggins L, Bongaerts P, Treml EA. 2016. Navigating the currents of seascape genomics: how spatial analyses can augment population genomic studies. *Current Zoology* 62:581–601.
- Riginos C, Crandall ED, Liggins L, Gaither MR, Ewing RB, Meyer C, Andrews KR, Euclide PT, Titus BM, Therkildsen NO, Salces-Castellano A. 2020. Building a global genomics observatory: Using GEOME (the Genomic Observatories Metadatabase) to expedite and improve deposition and retrieval of genetic data and metadata for biodiversity research. *Molecular Ecology Resources* 20:1458–69.
- Rochette NC, Rivera-Colón AG, Catchen JM. 2019. Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology* 28:4737–54.
- Sedlazeck FJ, Rescheneder P, Von Haeseler A. 2013. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics* 29:2790–1.
- Sgro CM, Lowe AJ, Hoffmann AA. 2011. Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications* 4: 326–337.

Urban DL, Minor ES, Treml EA, Schick RS. 2009. Graph models of habitat mosaics. *Ecology Letters* 12:260-73.

Weir BS and Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358-1370.

Wells S. 1996. *Tridacna maxima*. The IUCN Red List of Threatened Species 1996: e.T22138A9362499. <http://dx.doi.org/10.2305/IUCN.UK.1996.RLTS.T22138A9362499.en>.

Wells SM, Pyle RM, and Collins M. 1983. The IUCN invertebrate red data book. The IUCN invertebrate red data book.

Whitlock MC, McCauley DE. 1999. Indirect measures of gene flow and migration:  $F_{ST}$  not equal  $1/(4Nm+1)$ . *Heredity* 82: 117–125

Willette DA, Allendorf FW, Barber PH, Barshis DJ, Carpenter KE, Crandall ED, Cresko WA, Fernandez-Silva I, Matz MV, Meyer E, Santos MD. 2014. So, you want to use next-generation sequencing in marine systems? Insight from the Pan-Pacific Advanced Studies Institute. *Bulletin of Marine Science* 90:79-122.

Winter DJ. 2012. MMOD: an R library for the calculation of population differentiation statistics. *Molecular ecology resources* 12:1158-60.

Wright S. 1943. Isolation by distance. *Genetics* 28: 114-38.