A peer-reviewed version of this preprint was published in PeerJ on 13 May 2019.

View the peer-reviewed version (peerj.com/articles/6898), which is the preferred citable publication unless you specifically need to cite this preprint.

Pochon X, Wecker P, Stat M, Berteaux-Lecellier V, Lecellier G. 2019. Towards an in-depth characterization of Symbiodiniaceae in tropical giant clams via metabarcoding of pooled multi-gene amplicons. PeerJ 7:e6898 https://doi.org/10.7717/peerj.6898

Towards an in-depth characterization of Symbiodiniaceae in tropical giant clams via metabarcoding of pooled multi-gene amplicons

Xavier Pochon[1,2*], Patricia Wecker[3], Michael Stat[4], Véronique Berteaux-Lecellier[5], Gaël Lecellier[5,6]

[1] Coastal and Freshwater Group, Cawthron Institute, Nelson, New Zealand

[2] Institute of Marine Science, University of Auckland, Auckland, New Zealand

[3] Consultant in marine Microbiology, 7 Av. des garennes, 62170 Montreuil sur mer, France

[4] Department of Biological Sciences, Macquarie University, North Ryde, NSW 2109, Australia

[5] UMR250/9220 ENTROPIE, IRD-CNRS-UR, LabEx CORAIL, 101, promenade Roger-Laroque, BP A5 98848 Nouméa Cedex, New-Caledonia.

[6] Université Paris -Saclay, UVSQ, 55 Avenue de Paris 78035 Versailles Cedex, France

Competing Interests: The authors have declared that no competing interests exist. Xavier Pochon is an Academic Editor for PeerJ.

* E-mail: xavier.pochon@cawthron.org.nz

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

ABSTRACT

High-throughput sequencing is revolutionizing our ability to comprehensively characterize free-living and symbiotic Symbiodiniaceae, a diverse dinoflagellate group that plays a critical role in coral reef ecosystems. Most studies however, focus on a single marker for metabarcoding Symbiodiniaceae, potentially missing important ecological traits that a combination of markers may capture. In this proof-of-concept study, we used a small set of symbiotic giant clam (*Tridacna maxima*) samples obtained from nine French Polynesian locations and tested a dual-index sequence library preparation method that pools and simultaneously sequences multiple Symbiodiniaceae gene amplicons per sample for in-depth biodiversity assessments. The rational for this approach was to allow the metabarcoding of multiple genes without extra costs associated with additional single amplicon dual indexing and library preparations. Our results showed that the technique effectively recovered very similar proportions of sequence reads and dominant Symbiodiniaceae clades among the three pooled gene amplicons investigated per sample, and captured varying levels of phylogenetic resolution enabling a more comprehensive assessment of the diversity present. The pooled Symbiodiniaceae multi-gene metabarcoding approach decribed here is readily scalable, offering considerable analytical cost savings while providing sufficient phylogenetic information and sequence coverage.

Keywords (6-10 words): Biodiversity, Marine Ecology, Multi-gene Metabarcoding, High-Throughput Sequencing, South Pacific Ocean, symbiosis, *Tridacna*.

INTRODUCTION

Giant clams (Family Tridacnidae) play important roles in reef systems, acting as shelter for a number of organisms (Mercier & Hamel 1996), contributing to primary production through their symbiosis with dinoflagellates (Neo et al. 2015), and as effective filter feeders (Klumpp & Griffiths 1994). Due to their large size, relative abundance and longevity, giant clams can be considered as centennial barometers of reef health (Knop 1996). Unfortunately, as a highly prized resource throughout much of their Indo-Pacific range, giant clams also contain some of the most endangered species due to habitat degradation and overfishing, i.e. wild stock depletion and local extinctions (IUCN Red List (List 2018).

Giant clams on shallow reefs allow for the establishment of a diverse *in-situ* reservoir of interacting fungal, bacterial, and micro-algal communities (Baker 2003; Neo et al. 2015). Importantly, they form obligatory symbioses with, and release living cells of, Symbiodiniaceae *sensu* LaJeunesse et al. (2018), a group of dinoflagellates that are critical for the survival of a myriad of tropical invertebrates, including corals. Despite these dynamic interactions, very little is known about the extent of symbiont diversity within giant clams and the potential exchange with other reef invertebrates engaged in similar symbiotic associations (i.e. nudibranchs and corals; Wecker et al. 2015). Unlike traditional molecular techniques (e.g. PCR-based fingerprinting methods and Sanger sequencing) that have been extensively used to shed light on Symbiodiniaceae diversity in reef organisms (reviewed in Coffroth & Santos 2005; Stat et al. 2006), recent advances in High-Throughput Sequencing (HTS) technologies now enable unprecedented sequencing depth for global biodiversity assessments of symbiotic and free-living communities of Symbiodiniaceae (Boulotte et al. 2016; Cunning et al. 2015; Edmunds et al. 2014; Hume

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

et al. 2018; Shinzato et al. 2018; Thomas et al. 2014). Nevertheless, such studies usually focus on metabarcoding analyses of single molecular markers in isolation, in particular the *ITS2* marker (but see Smith et al. 2017; Thomas et al. 2014), potentially overlooking intrinsic phylogenetic differences known to occur between distinct Symbiodiniaceae genes (Pochon et al. 2012; Pochon et al. 2014).

A variety of HTS library preparation methods exist for metabarcoding biological samples using Illumina™ sequencing platforms, including the use of fusion tag primers (Stat et al. 2017), the ligation of Illumina™ adapters using TruSeq™ PCR-free kits (Rhodes et al. 2014), and the addition of Illumina™ adapters via dual-index sequencing (Kozich et al. 2013). The latter technique requires two distinct rounds of PCR analyses. The first round uses gene-specific primers modified to include Illumina™ adapter tails. Following purification of the PCR products, a second short round of PCR is applied using Nextera™ library construction kits that involve individual primer sets containing the Illumina™ adapter and sequencing primer sequence. This second PCR step is usually performed on individual PCR amplicon products before the pooling and sequencing of multiple samples so that demultiplexing of sequence data results in appropriate identification of input samples. For laboratories that use the services of external genomic facilities for the preparation of their dual-index libraries, an increased sample set usually correlates positively with the analytical cost due, in part, to the use of additional Nextera™ indexed primers. Therefore, one solution for reducing costs when performing multi-gene analyses of individual samples, is to pool the PCR amplicon products prior to the second PCR step, followed by the sequencing and gene-specific demultiplexing per sample.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

Here we conducted a preliminary assessment of a dual-index multi-gene metabarcoding approach via the pooling and side-by-side HTS analysis of PCR amplicons from three commonly employed nuclear and chloroplastic Symbiodiniaceae markers. The ability to combine multiple gene amplicon targets per sample offers considerable analytical cost savings while providing sufficient phylogenetic information and sequence coverage. This study describes a multi-marker metabarcoding approach using giant clam *Tridacna maxima* as a model and discusses future applications for improving analyses of coral reef holobionts.

MATERIAL AND METHODS

Sample collection and DNA extraction

For this study, twelve DNA extracts from *Tridacna maxima* biopsies, previously collected between February 1st 2011 and November 2nd 2013 from nine islands in the French Polynesian Archipelagos (Figure 1, Table S1) were used (Dubousquet et al. 2018).

Preparation of Multiplexed Amplicons High-Throughput Sequencing Libraries

Three sets of Symbiodiniaceae-specific primers with Illumina™ adapter tails (Table S2) were used to amplify each sample (S141-S152; Table 1) in separate Polymerase Chain Reactions (PCR). Three markers were amplified: (*i*) the Internal Transcribed Spacer 2 (*ITS2*) of the nuclear ribosomal RNA array using primers ITSD_illu and ITS2rev2_illu , (*ii*) the D1-D2 region of the 28S large subunit (*LSU*) nuclear ribosomal RNA gene using the newly designed primers LSU1F_illu and LSU1R_illu, and (*iii*) the hyper-variable region of the chloroplast 23S (*23S*) ribosomal RNA gene using primers 23SHyperUP_illu and 23SHyperDN_illu (Manning & Gates 2008; Pochon et al. 2010).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

PCR was performed for each sample and for each gene separately in 50 µL volumes, with the reaction mixture containing 45 µL of Platinum PCR SuperMix High Fidelity (Life Technologies), 10 uM of each primer, and 10-20 ng of template DNA. In order to maximize specificity to Symbiodiniaceae, a touchdown PCR protocol was used for each reaction as follows: (*i*) 95 °C for 10 min; (*ii*) 25 cycles of 94 °C for 30 s, 65 °C for 30 s (decreasing the annealing temperature 0.5 °C for every cycle after cycle 1), and 72 °C for 1 min; (*iii*) 14 cycles of 94 °C for 30 s, 52 °C for 30 s and 72 °C for 1 min; and (*iv*) a final extension of 72 °C for 10 min. Amplicons of the correct size were purified using Agencourt AMPure XP PCR Purification beads following the manufacturers' instructions. In order to sequence the three gene amplicons per sample in multiplex using HTS, individual purified products for each marker originating from the same giant clam were pooled together to enable the attachment of the same Nextera™ index (i.e. 12 samples). This was achieved by quantifying the amplicons using a Qubit Fluorometer 2.0 (Life Technologies), diluting to 1 ng/µL using Milli-Q water and mixing 5 µL of each of gene amplicon from the same giant clam together. To assess the levels of cross-contamination between samples potentially arising during the library indexing step, nine unmixed amplicon products (i.e. *ITS2*, *LSU* and *23S* amplicons from three haphazardly selected giant clams; samples S141-S143; Table 1), each with their own unique index to be added, were also prepared.

The resulting 21 samples were sent to New Zealand Genomics Ltd. (University of Auckland, New Zealand) for HTS library preparation which involved a second round of PCR to attach the Nextera™ indexes on to the amplicons for MiSeq Illumina™ sequencing. PCR products were combined in equimolar concentrations and the final library paired-end

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

147 sequenced on an Illumina™ MiSeq using a 500 cycle (2 x 250) MiSeq® v2 Reagent Kit
148 and standard flow cell.
149
150
151 ***Bioinformatics***
152 Illumina™ sequence datasets were prepared using the read preparation and dereplication
153 pipeline of USEARCH (Edgar 2010). Firstly, paired reads were merged (fastq_mergepairs
154 command) and filtered (fastq_filter command) with an expected number of error of 0.25.
155 More than 90% of the base pairs had a Q score > 40. Next, samples were demultiplexed in
156 three groups, primers were trimmed and a global trimming was operated according to the
157 recommendations for *ITS* amplicon reads (Edgar 2013). The sequence data were
158 dereplicated and unique singletons found across the complete dataset were discarded.
159
160 For phylogenetic assignments of Symbiodiniaceae, three distinct annotated reference
161 databases (*ITS2*, *LSU* and *23S*) were generated in fasta format, including sequence
162 representatives from each of nine Symbiodiniaceae clades (A to I), with (*i*) 409
163 representative *ITS2* phylotypes from GeoSymbio (Franklin et al. 2012), (*ii*) 37
164 representative *LSU* sequences from Pochon et al. (2012), and (*iii*) 104 sequences of *23S*
165 from Takabayashi et al. (2011). Symbiodiniaceae assignments were performed using the
166 software 'Kallisto' (Bray et al. 2016) which provides unprecedented speed and accuracy
167 for optimal analysis of large-scale datasets (e.g. large RNA-Seq data) without the need for
168 time-consuming alignment steps.
169
170 Because the main goal of the present pilot study was to investigate the sequencing depth

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

and potential inter-marker biases of the multi-marker metabarcoding approach using giant clam samples as a proof-of-concept, as opposed to describing potentially novel Symbiodiniaceae diversity in these samples, we modified the Kallisto pipeline to only retain HTS reads yielding exact matches (i.e. without ambiguity amongst k-mers) to individual referenced genotypes in each gene. This approach transforms each sequence from reference databases into pseudo-alignments of k base-pairs (bp) k-mers which slide along the sequence of reference one bp at a time. Individual sequences generated via HTS were then blasted against all pseudo-alignments and exact matches against the entire population of k-mers are recorded. To reduce mis-assignments, all merged reads with ambiguities between k-mers of different reference genotypes were determined as chimeric and removed from the dataset. The sequences that did not result in exact matches could correspond to non-Symbiodiniacea sequences or to sequences not comprised in our custom database. Therefore, a second comparison using BLASTn against the National Center for Biotechnology Information (NCBI) nucleotide databases was performed and the accession numbers yielding exact matches were retained for downstream analyses. The number of unique sequences matching genotypes in the reference databases and GenBank was recorded (Table S3). Raw sequence data were submitted to the BioProject Archive under accession PRJNA471926 (SRR7181922-SRR7181942).

Sequence Diversity Analyses

Unique sequence genotypes found at or above a 0.05% threshold from the total sequence abundance per sample were scored (Table S3) and the specific genotypes of reference (i.e. from in-house reference databases and GenBank) were retained for sequence diversity and phylogenetic analyses. Global sequence diversity from each of the three

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

datasets (*23S*, *ITS2*, and *LSU*) were visualized using the plug-in DataBurst implemented in Excel (Microsoft Office version 2013 or later).

One sequence alignment was generated for each of the three investigated gene datasets using the sequence alignment software BioEdit v7.2.5 (Hall 1999). Owing to the difficulty in aligning sequences from *Symbiodinium* (clade A) and *Cladocopium* (clade C) genera when using the *23S* and *ITS2* genes, and between Symbiodiniaceae and non-Symbiodiniaceae (i.e. clams, fungi, and plants) sequences, phylogenetic reconstructions only aimed at depicting pair-wise relationships between retained sequence genotypes. Therefore, unrooted phylogenetic inferences were generated using the neighbor-joining method implemented in the program MEGA v. 7.0 (Kumar et al. 2016), with the p-distance model and gaps treated as pairwise deletions. Internal nodes support was tested using the bootstrap method (Felsenstein 1985) and 500 replicates.

RESULTS

A total of 1,590,047 sequences were obtained from the 21 samples (75,716 +/- 41,576 sequences per sample), which included 12 amplicon samples (S141-S152) each containing three pooled gene products (*23S*, *ITS2*, and *LSU*) and nine amplicon samples from three selected giant clam isolates (S141, S142, and S143) which only contained a single gene amplicon as internal controls (Table 1; Table S3). One sample (internal control S143 for ITS2) failed the sequencing step with only 130 raw reads produced. After read cleaning, the total number of high-quality sequences was 1,104,687 (52,604 +/- 29,250 sequences per sample). The proportion of total reads (Table 1) between the three investigated genes was well-balanced with 398,442 reads (*23S*), 339,780 reads (*ITS2*), and 359,768 reads

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

(*LSU*). In contrast, unique reads varied between 23,779 sequences for the *23S* gene and 71,776 sequences for the *LSU* gene (Table S3). The inclusion of nine positive controls, representing three amplicon products per gene sequenced in isolation, revealed the presence of low levels of sequence cross-contamination between samples (mean of 4.5 sequences ± 4.6 SD) (Table 1). This low-level of background contamination (1 to 23 sequences per sample) represented <0.003% of the total reads per sample (Table S3). Therefore, as a conservative measure, we chose to remove sequences that represented < 0.05% of the total sequence abundance per sample.

Our bioinformatics pipeline identified 43 Symbiodiniaceae *23S* chloroplast genotypes, including 16 that matched the *23S* reference database and another 27 that matched sequences in GenBank. After exclusion of genotypes represented by less than 0.05% of the sequence abundance in each sample (Table S3), the number of unique *23S* Symbiodiniaceae sequences retained for phylogenetic analysis was eleven (Figure S1). Similarly, blasting *ITS2* and *LSU* datasets against both types of databases led to the identification of 117 and 93 unique sequences when using the original datasets, and to 46 and 51 unique sequences following the 0.05% filtering threshold, respectively.

Diversity diagrams were generated to visualize the sequence abundance of Symbiodiniaceae generic and sub-generic sequences recovered from the twelve giant clam samples and among the three investigated genes (Figure 2). The pooled multi-gene approach yielded similar proportions of dominant genera, but with some notable differences. The genus *Symbiodinium* (previously Clade A) dominated in all three markers, particularly in *23S* (91.8%; dominant sub-generic sequence chvA2), with lower but similar

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

proportions between *ITS2* (81.7%; dominant sub-generic sequences A3/A6) and *LSU* (83.9%; dominant sub-generic sequences A3/A13). The genus *Cladocopium* (previously Clade C) represented 7.9% (dominant sub-generic sequence chvC1), 18.2% (dominant sequence C1), and 15.0% (dominant sequence C1) of reads for the *23S*, *ITS2*, and *LSU* markers, respectively. *Geracladium* (previously clade G) was only detected using the chloroplast *23S* gene (0.2% of reads), whereas the nuclear *LSU* gene displayed reduced specificity for Symbiodiniaceae as indicated by ~1% of sequence reads matching other organisms such as streptophytes (*Mitchella repens* and *Asclepias verticillata*), and the host giant clam *T. maxima*. Overall, the proportion of dominant Symbiodiniaceae generic and sub-generic sequences recovered between the multiplexed samples and the positive (single gene) controls were very similar (Table S4).

DISCUSSION

Multi-gene metabarcoding: more for less

The concept of pooled multi-gene amplicons for dual-indexed metabarcoding, i.e. the tagging and pooling of distinct gene amplicons before Illumina™ adapter indexing and simultaneous sequencing of samples, has been used extensively in other research fields (e.g., Keeley et al. 2018; Marcelino & Verbruggen 2016; von Ammon et al. 2018; Zhang et al. 2018), but has never been applied to Symbiodiniaceae dinoflagellates. In this proof-of-concept study, we show that the technique effectively recovered similar proportions of sequence reads and Symbiodiniaceae genera among the three multiplexed genes investigated per sample, providing more confidence that single gene primer biases did not occur during Nextera™ indexing. Another advantage is the ability to simultaneously visualize varying levels of phylogenetic resolution, enabling a more comprehensive

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

assessment of the diversity present. For example, while the traditional 'species-level' *ITS2* marker (LaJeunesse 2001) enabled characterization of 46 Symbiodiniaceae sub-generic sequences, the *LSU* marker, interestingly, offered both a similarly high resolution for Symbiodiniaceae (46 sub-generic sequences) and a reduced specificity by also enabling identification of other host-associated organisms such as streptophytes, as well as the host *Tridacna*. The hyper-variable region of the chloroplast *23S* marker used here is more conserved, but has been successfully used for specifically targeting low abundance free-living Symbiodiniaceae cells from environmental samples (Decelle et al. 2018; Manning & Gates 2008; Pochon et al. 2010; Takabayashi et al. 2011). The unique detection of *Gerakladium* (clade G) using the 23S marker highlights the added value of the multi-gene approach for broader Symbiodiniaceae screening efficiency. Analytical cost is an important consideration for any research group aiming to monitor coral reef ecosystems, and the budget needed to include HTS for biodiversity assessments is highly variable. The cost ranges between AU $40-$100 per sample (Stat et al. 2018) and depends on the number of gene regions investigated, method of library preparation, sequencing depth, and whether amplicon multiplexing is employed as shown here. In this context, our approach is readily scalable and has the potential to offer substantial savings in terms of both time and cost, for example, by enabling coral reefs researchers to generate multi-gene Symbiodiniaceae data in a 96-well format for the price of a single dual-indexed Illumina™ MiSeq run. Additional studies are required to investigate whether 'true multiplexing', i.e. the mixing of multiple primer sets in the original PCR to produce multi-gene amplicons (De Barba et al. 2014; Fiore-Donno et al. 2018) would result in similar proportions of Symbiodiniaceae genotypes between markers such as shown in the present study. Such approach, if validated, would allow significant additional cost savings.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

Paving the way for comprehensive biodiversity assessment of giant clams

Giant clams on shallow reefs allow for the establishment of a diverse *in-situ* reservoir of interacting fungal, bacterial, and micro-algal communities (Baker 2003; Neo et al. 2015). For example, they commonly harbor Symbiodiniaceae from at least three distinct genera (*Symbiodinium* [clade A], *Cladocopium* [clade C], and/or *Durusdinium* [clade D]) simultaneously or in isolation within one host, with *Symbiodinium* being the dominant symbiont genus in most clams (Baillie et al. 2000; DeBoer et al. 2012; Ikeda et al. 2017; Ikeda et al. 2016; Pappas et al. 2017; Trench et al. 1981). Similar to coral symbiosis, it is assumed that the genotypic composition of Symbiodiniaceae in giant clams is influenced by environmental or physical parameters (e.g. temperature, irradiance), or by life stages and taxonomic affiliation (Ikeda et al. 2017; Pappas et al. 2017). Giant clam larvae (veliger) acquire free-living Symbiodiniaceae cells 'horizontally' from their surrounding environment (Fitt & Trench 1981). When mature, giant clams (e.g. *Tridacna derasa*) expel high numbers of intact symbionts in their faeces at rates of 4.9×10^5 cells d^{-1} (Buck et al. 2002; Maruyama & Heslinga 1997). Despite the dynamic interaction of symbionts between Tridacnidae and the environment, very little is known about the extent of symbiont diversity within giant clams and the potential exchange with other reef invertebrates engaged in similar symbiotic associations

In this preliminary study, we found that genera *Symbiodinium* and *Cladocopium* dominated in adult giant clams in French Polynesia (Figure S2). Symbiodinium was the major genus in our samples and in particular the closely related sub-generic ITS2 genotypes A3 and A6, previously described as *Symbiodinium tridacniadorum*, and therefore associated

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

315 with *Tridacna* clams (Lee et al. 2015). A3 is the most dominant genotype in *T. Maxima*
316 around the world and both A3/A6 are more likely to be sampled in giant clams from
317 shallow reefs (Weber 2009).
318
319 Furthermore, for *Cladocopium* we found that the generalist ITS2 genotype C1 (LaJeunesse
320 et al. 2003) co-dominated in our samples, which is consistent with a previous study
321 showing C1 as a common genotype in *T. Maxima* from around the world (Weber 2009).
322 Noteworthy, we also found a smaller percentage of C3z, Cspd and C50 ITS2 genotypes,
323 which to our knowledge have not yet been found in *T. maxima* before, but are usually
324 restricted to corals (LaJeunesse et al. 2004; LaJeunesse et al. 2010; Macdonald et al. 2008;
325 Shinzato et al. 2018). Finally, we did not detect any symbiont from the genus *Durusdinium*
326 (Clade D) despite the in-depth sequencing afforded by our multiplexed amplicons method.
327 However, *Durusdinium* has never been detected in *T. maxima* from French Polynesia
328 compared to other regions such as the Indian Ocean (DeBoer et al. 2012; Weber 2009). As
329 we only worked with adult clams from shallow water, it would be interesting to confirm
330 the hypotheses of Ikeda et al. (2017) and Weber (2009) who argued that *Durusdinium*
331 symbionts might be restricted to 'young' *T. squamosa* clams (less than 11 cm) or that giant
332 clams harbored this dinoflagellate genus only when sampled from deeper reefs,
333 respectively. Nevertheless, the small dataset used in the present study precludes us from
334 making any relevant assumptions about potentially novel symbiont diversity in giant clams.
335 In particular, the use of the Kallisto bioinformatics pipeline which restricted the analysis to
336 100% sequence similarity hits is likely not suitable for the many studies where a high
337 degree of sequence novelty is found. It is our hope, however, that our multi-gene approach
338 will be investigated further using a more comprehensive giant clam dataset along with the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

development of an alternative bioinformatics method guiding users on the assignment of genus to species-level taxon ID to novel multi-gene sequences for deposition to GenBank.

CONCLUSIONS

This pilot project explored the use of multiplexed amplicons metabarcoding for rapid, cost-effective and in-depth characterization of Symbiodiniaceae dinoflagellates using the giant clam *T. maxima* as a model. Our results showed that the technique effectively recovered similar proportions of sequence reads and Symbiodiniaceae diversity among the three multiplexed gene amplicons investigated per sample enabling a more comprehensive assessment of the diversity present, while also offering appreciable analytical cost savings. We also found that *Symbiodinium* (clades A) and *Cladocopium* (clade C) were the dominant genera in adult giant clams in French Polynesia, with similar sub-generic genotypes (*ITS2* A3, A6, and C1) previously described as commonly associated with giant clams from around the world. Our approach paves the way for more comprehensive surveys of this important yet endangered group of reef invertebrates and its potential role as an important *Symbiodinium* reservoir for declining coral reefs. Future investigations may also expand on this method to clarify species-level differentiation among *Symbiodinium* taxa using other markers (e.g. nuclear Actin, chloroplast *psbA*, mitochondrial *COI* and *16S*), or simultaneously characterize all organisms (viruses, bacteria, fungi, and other eukaryotes) associated with a more diverse host range. Such holistic diversity assessments will improve our knowledge on the ecology and evolution of tropical holobionts and better predict the adaptation of coral reefs in a rapidly changing environment.

ACKNOWLEDGMENTS

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

We thank Jonathan Drew for his support in the laboratory, and Charley Waters, Sam Murray and Chris Cornelisen for valuable discussions during manuscript preparation.

Funding Statement

This work was supported by Cawthron Institute Internal Investment Fund (IIF) #BST16931 and by the French National Research Center (CNRS). The funders have no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Table 1 Number of DNA sequences recovered from each sample (S141-S152), before and after quality filtration, and after demultiplexing into each gene. Samples S141 to S143 were used as control samples, each targeting only one of three PCR amplicons. Columns highlighted in grey show a low background contamination.

Sample ID	Source reads	Filtered reads	23S reads	ITS2 reads	LSU reads
Multiplexed					
S141	75731	53654	22072	17813	13435
S142	89975	65312	26504	24395	14040
S143	78009	48881	21061	10256	17321
S144	172319	126860	48941	39131	38128
S145	147293	104743	31048	34457	38662
S146	72548	51886	23268	16817	11537
S147	118815	79339	29870	32449	16332
S148	50176	34810	12577	11695	10264

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

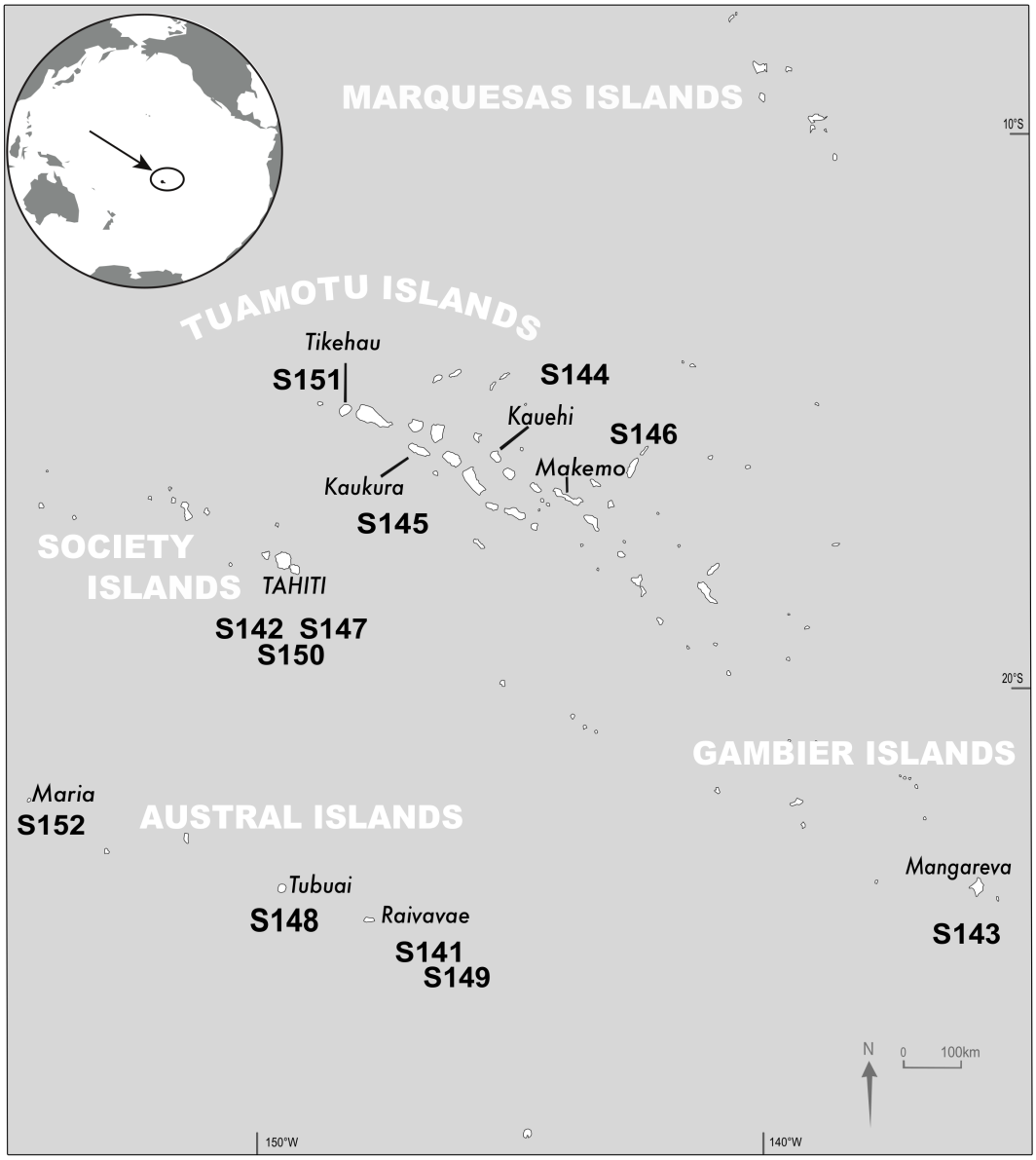
S149	4728	3381	2400	366	599
S150	88926	59387	20788	22068	16216
S151	53016	38314	15964	12882	9298
S152	60107	42239	17075	13108	11707
Controls					
ITS2 only					
S141	85824	52588	8	52335	1
S142	81924	52270	10	51988	6
S143*	130	13	5	6	2
LSU only					
S141	56565	31134	8	7	30758
S142	92110	62629	23	0	62129
S143	114431	69823	9	0	69318
23S only					
S141	77522	66763	66399	3	3
S142	42004	36422	36263	3	9
S143	27894	24239	24149	1	3
Total reads	**1590047**	**1104687**	**398442**	**339780**	**359768**

376 *One control sample (S143 ITS2) failed at sequencing, resulting in only 130 raw reads.

377

378

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019



379

380 **Figure 1** Location and sample identification for the twelve *Tridacna maxima* samples investigated

381 in this study (credit to R. Canavesio).

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

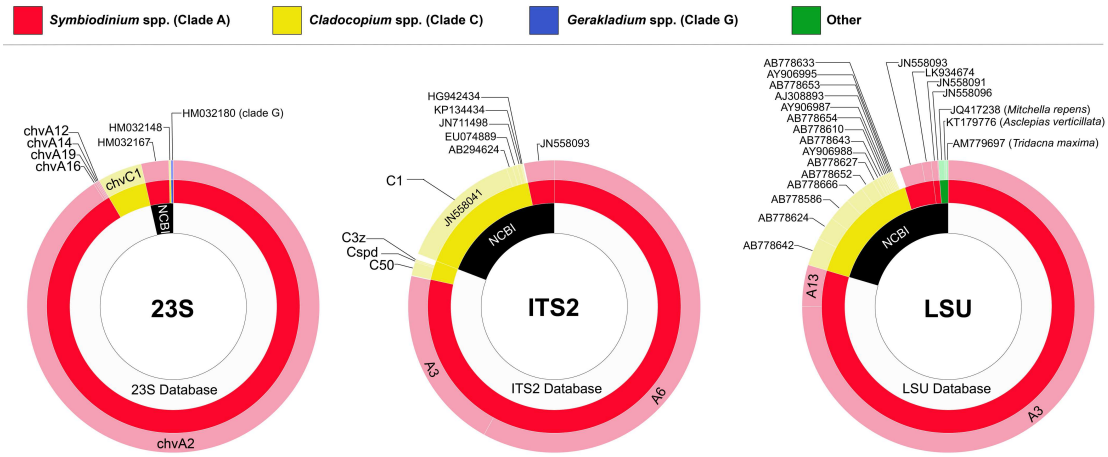


Figure 2 Global Symbiodiniaceae diversity charts obtained from each of the three datasets (left to right: *23S*, *ITS2*, and *LSU*). The proportion of sequences matching one of the three in-house reference databases or NCBI (inner circles) and their corresponding phylogenetic affiliation at genus (i.e clade; middle circles) and sub-generic (i.e. subclade; outer circles) levels. Sequence reads representing <0.1% of total read abundance are not included.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

Supplementary Information

Table S1 Identification numbers, collection localities and date collected for the twelve samples of *Tridacna maxima* investigated in this study.

Table S2 List of primers used for generating PCR amplicons. Illumina adaptors are shown in bold.

Table S3 Sequence counts and blast annotations for the 21 PCR amplicon samples analysed in multiplex and individually (controls) over three distinct genes (*23S*, *ITS2*, *LSU*). Sheet 1 includes the merged counts and dereplicated data; Sheet 2 includes exact *23S* sequence matches against the Takabayashi et al. (2012) database and NCBI; Sheet 3 includes retained *23S* genotypes following the 0.05% abundance threshold; Sheet 4 includes exact *ITS2* sequence matches against the GeoSymbio database and NCBI; Sheet 5 includes retained *ITS2* genotypes following the 0.05% abundance threshold; Sheet 6 includes exact *LSU* sequence matches against the Pochon et al. (2012) database and NCBI; and Sheet 7 includes retained *LSU* genotypes following the 0.05% abundance threshold.

Table S4 Percentage comparison of each Symbiodiniaceae sub-generic genotype recovered using the three amplicon markers in 'Multiplex' versus single 'Control' markers (see Table 1). The proportion of each sub-generic type between 'Multiplex' and 'Control' is almost identical for the *23S* marker, but shows some minor differences for the *ITS2* and *LSU* markers. For example, four *ITS2* types were detected in the 'Multiplex' but not in the 'Control' samples, and there were five instances where *LSU* types were detected in the 'Control' but not in the 'Multiplex' samples. These minor differences are likely attributable to PCR or sequencing biases.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

Figure S1 Unrooted circled trees of Symbiodiniaceae genotypes inferred using the Neighbor-Joining method, with (**A**) 11 *23S* sequences, (**B**) 46 *ITS2* sequences, and (C) 51 *LSU* sequences.

Figure S2 Distribution of Symbiodiniaceae genera (i.e. clades) in *Tridacna maxima* obtained from each of the three datasets (left to right: *23S*, *ITS2*, and *LSU*) per sample identification (S141-152).

REFERENCES

Baillie BC, Belda-Baillie CA, Silvestre V, Sison M, Gomez AV, Gomez ED, and Monje V. 2000. Genetic variation in *Symbiodinium* isolates from giant clams based on random-amplified-polymorphic DNA (RAPD) patterns. *Marine Biology* **136**:829-836.

Baker AC. 2003. Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of *Symbiodinium*. *Annual Review of Ecology, Evolution, and Systematics* **34**:661-689. DOI 10.1146/annurev.ecolsys.34.011802.132417

Boulotte NM, Dalton SJ, Carroll AG, Harrison PL, Putnam HM, Peplow LM, and van Oppen MJ. 2016. Exploring the *Symbiodinium* rare biosphere provides evidence for symbiont switching in reef-building corals. *ISME J* **10**:1-9. DOI 10:2693-2701. 10.1038/ismej.2016.54

Bray NL, Pimentel H, Melsted P, and Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**:525-527. DOI 10.1038/nbt.3519

Buck BH, Rosenthal H, and Saint-Paul U. 2002. Effect of increased irradiance and thermal stress on the symbiosis of *Symbiodinium microadriaticum* and *Tridacna gigas*. *Aquatic Living Resources* **15**:107-117.

Coffroth MA, and Santos SR. 2005. Genetic diversity of symbiotic dinoflagellates in the genus *Symbiodinium*. *Protist* **156**:19-34. DOI 10.1016/j.protis.2005.02.004

Cunning R, Yost DM, Guarinello ML, Putnam HM, and Gates RD. 2015. Variability of *Symbiodinium* communities in waters, sediments, and corals of thermally distinct reef pools in American Samoa. *PLoS One* **10**:e0145099. DOI 10.1371/journal.pone.0145099

De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, Coissac E, and Taberlet P. 2014. DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources* **14**:306-323. DOI 10.1111/1755-0998.12188

DeBoer TS, Baker AC, Erdmann MV, Ambariyanto, Jones PR, and Barber PH. 2012. Patterns of *Symbiodinium* distribution in three giant clam species across the biodiverse Bird's Head region of Indonesia. *Marine Ecology Progress Series* **444**:117-132. DOI 10.3354/meps09413

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

469 Decelle J, Carradec Q, Pochon X, Henry N, Romac S, Mahe F, Dunthorn M, Kourlaiev
470 A, Voolstra CR, Wincker P, and de Vargas C. 2018. Worldwide occurrence and
471 activity of the reef-building coral symbiont *Symbiodinium* in the open ocean.
472 *Current Biology* **28**:3625-3633. DOI 10.1016/j.cub.2018.09.024
473 Dubousquet V, Berteaux-Lecellier V, Lopes V, Raharivelomanana P, and Lecellier G.
474 2018. Existence of two novel and non-sectorized clades of the giant clam
475 *Tridacna maxima* in French Polynesia: implications for connectivity and origin.
476 *HAL Archives Ouvertes*:hal-01713884.
477 Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST.
478 *Bioinformatics* **26**:2460-2461. DOI 10.1093/bioinformatics/btq461
479 Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon
480 reads. *Nature Methods* **10**:996-998. DOI 10.1038/nmeth.2604
481 Edmunds PJ, Pochon X, Levitan DR, Yost DM, Belcaid M, Putnam HM, and Gates RD.
482 2014. Long-term changes in *Symbiodinium* communities in *Orbicella annularis* in
483 St. John, US Virgin Islands. *Marine Ecology Progress Series* **506**:129-144. DOI
484 10.3354/meps10808
485 Felsenstein J. 1985. Phylogenies and the Comparative Method. *American Naturalist*
486 **125**:1-15.
487 Fiore-Donno AM, Rixen C, Rippin M, Glaser K, Samolov E, Karsten U, Becker B, and
488 Bonkowski M. 2018. New barcoded primers for efficient retrieval of cercozoan
489 sequences in high-throughput environmental diversity surveys, with emphasis on
490 worldwide biological soil crusts. *Molecular Ecology Resources* **18**:229-239. DOI
491 10.1111/1755-0998.12729
492 Fitt WK, and Trench RK. 1981. Spawning, development, and acquisition of
493 zooxanthellae by *Tridacna squamosa* (mollusca, bivalvia). *Biological Bulletin*
494 **161**:213-235.
495 Franklin EC, Stat M, Pochon X, Putnam HM, and Gates RD. 2012. GeoSymbio: a hybrid,
496 cloud-based web application of global geospatial bioinformatics and
497 ecoinformatics for *Symbiodinium*-host symbioses. *Molecular Ecology Resources*
498 **12**:369-373. DOI 10.1111/j.1755-0998.2011.03081.x
499 Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis
500 program for Windows 95/98/NT. *Nucleic Acids Research* **41**:95-98.
501 Hume BCC, Ziegler M, Poulain J, Pochon X, Romac S, Boissin E, de Vargas C, Planes S,
502 Wincker P, and Voolstra CR. 2018. An improved primer set and amplification
503 protocol with increased specificity and sensitivity targeting the *Symbiodinium*
504 ITS2 region. *PeerJ* **6**:e4816. DOI 10.7717/peerj.4816
505 Ikeda S, Yamashita H, Kondo SN, Inoue K, Morishima SY, and Koike K. 2017.
506 Zooxanthellal genetic varieties in giant clams are partially determined by species-
507 intrinsic and growth-related characteristics. *PLoS One* **12**:e0172285. DOI
508 10.1371/journal.pone.0172285
509 Ikeda S, Yamashita H, Liao LM, and Koike K. 2016. A simple and rapid determination
510 method for zooxanthellal genetic diversity in giant clams using multiplex PCR.
511 *Fisheries Science* **82**:747-753. DOI 10.1007/s12562-016-1004-x
512 Keeley N, Wood SA, and Pochon X. 2018. Development and preliminary validation of a
513 multi-trophic metabarcoding biotic index for monitoring benthic organic
514 enrichment. *Ecological Indicators* **85**:1044-1057.
515 DOI 10.1016/j.ecolind.2017.11.014

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

Klumpp D, and Griffiths C. 1994. Contributions of phototrophic and heterotrophic nutrition to the metabolic and growth requirements of four species of giant clam (Tridacnidae). *Marine Ecology Progress Series* **115**:103-115.

Knop D. 1996. Giant clams - a comprehensive guide to the identification and care of tridacnid clams. *Dähne Verlag, Ettlingen*:p. 255.

Kozich JJ, Westcott SL, Baxter NT, Highlander SK, and Schloss PD. 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied Environmental Microbiology* **79**:5112-5120. DOI 10.1128/AEM.01043-13

Kumar S, Stecher G, and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**:1870-1874. DOI 10.1093/molbev/msw054

LaJeunesse T, Thornhill D, Cox E, Stanton F, Fitt W, and Schmidt G. 2004. High diversity and host specificity observed among symbiotic dinoflagellates in reef coral communities from Hawaii. *Coral Reefs* **23**:596-603. DOI 10.1007/s00338-004-0428-4

LaJeunesse TC. 2001. Investigating the biodiversity, ecology, and phylogeny of endosymbiotic dinoflagellates in the genus *symbiodinium* using the its region: in search of a "species" level marker. *Journal of Phycology* **37**:866–880.

LaJeunesse TC, Loh WKW, van Woesik R, Hoegh-Guldberg O, Schmidt GW, and Fitt WK. 2003. Low symbiont diversity in southern Great Barrier Reef corals, relative to those of the Caribbean. *Limnology and Oceanography* **48**:2046-2054.

LaJeunesse TC, Parkinson JE, Gabrielson PW, Jeong HJ, Reimer JD, Voolstra CR, and Santos SR. 2018. Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. *Current Biology* **28**:2570-2580. DOI 10.1016/j.cub.2018.07.008

LaJeunesse TC, Smith R, Walther M, Pinzon J, Pettay DT, McGinley M, Aschaffenburg M, Medina-Rosas P, Cupul-Magana AL, Perez AL, Reyes-Bonilla H, and Warner ME. 2010. Host-symbiont recombination versus natural selection in the response of coral-dinoflagellate symbioses to environmental disturbance. *Proceedings of the Royal Society B* **277**:2925-2934. DOI 10.1098/rspb.2010.0385

Lee SY, Jeong HJ, Kang NS, Jang TY, Jang SH, and Lajeunesse TC. 2015. *Symbiodinium tridacnidorum* sp. nov., a dinoflagellate common to Indo-Pacific giant clams, and a revised morphological description of *Symbiodinium microadriaticum* Freudenthal, emended Trench & Blank. *European Journal of Phycology* **50**:155-172. DOI 10.1080/09670262.2015.1018336

List IR. 2018. The IUCN Red List of Threatened Species [WWW Document].

Macdonald AHH, Sampayo EM, Ridgway T, and Schleyer MH. 2008. Latitudinal symbiont zonation in *Stylophora pistillata* from southeast Africa. *Marine Biology* **154**:209-217. DOI 10.1007/s00227-008-0915-1

Manning MM, and Gates RD. 2008. Diversity in populations of free-living *Symbiodinium* from a Caribbean and Pacific reef. *Limnology and Oceanography* **53**:1853–1861.

Marcelino VR, and Verbruggen H. 2016. Multi-marker metabarcoding of coral skeletons reveals a rich microbiome and diverse evolutionary origins of endolithic algae. *Scientific Reports* **6**:31508. DOI 10.1038/srep31508

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

562 Maruyama T, and Heslinga GA. 1997. Fecal discharge of zooxanthellae in the giant clam
563 *Tridacna derasa*, with reference to their in situ growth rate. *Marine Biology*
564 **127**:473-477.
565 Mercier A, and Hamel JF. 1996. The secret of the giant clam. *Freshwater and Marine*
566 *Aquaculture* **19**:112-113.
567 Neo ML, Eckman W, Vicentuan K, Teo SLM, and Todd PA. 2015. The ecological
568 significance of giant clams in coral reef ecosystems. *Biological Conservation*
569 **181**:111-123. DOI 10.1016/j.biocon.2014.11.004
570 Pappas MK, He S, Hardenstine RS, Kanee H, and Berumen ML. 2017. Genetic diversity
571 of giant clams (*Tridacna* spp.) and their associated *Symbiodinium* in the central
572 Red Sea. *Marine Biodiversity* **47**:1209-1222. DOI 10.1007/s12526-017-0715-2
573 Pochon X, Putnam HM, Burki F, and Gates RD. 2012. Identifying and characterizing
574 alternative molecular markers for the symbiotic and free-living dinoflagellate
575 genus *Symbiodinium*. *PLoS One* **7**:e29816.
576 DOI 10.1371/journal.pone.0029816
577 Pochon X, Putnam HM, and Gates RD. 2014. Multi-gene analysis of *Symbiodinium*
578 dinoflagellates: a perspective on rarity, symbiosis, and evolution. *PeerJ* **2**:e394.
579 DOI 10.7717/peerj.394
580 Pochon X, Stat M, Takabayashi M, Chasqui L, Chauka LJ, Logan DDK, and Gates RD.
581 2010. Comparison of endosymbiotic and free-living *Symbiodinium* (Dinophyceae)
582 diversity in a Hawaiian reef environment1. *Journal of Phycology* **46**:53-65. DOI
583 10.1111/j.1529-8817.2009.00797.x
584 Rhodes J, Beale MA, and Fisher MC. 2014. Illuminating choices for library prep: a
585 comparison of library preparation methods for whole genome sequencing of
586 *Cryptococcus neoformans* using Illumina HiSeq. *PLoS One* **9**:e113501.
587 DOI 10.1371/journal.pone.0113501
588 Shinzato C, Zayasu Y, Kanda M, Kawamitsu M, Satoh N, Yamashita H, and Suzuki G.
589 2018. Using seawater to document coral-zoothanthella diversity: a new approach
590 to coral reef monitoring using environmental DNA. *Frontiers in Marine Science*
591 **5**: 28. DOI 10.3389/fmars.2018.00028
592 Smith EG, Ketchum RN, and Burt JA. 2017. Host specificity of *Symbiodinium* variants
593 revealed by an ITS2 metahaplotype approach. *ISME J* **11**:1500-1503. DOI
594 10.1038/ismej.2016.206
595 Stat M, Carter D, and Hoegh-Guldberg O. 2006. The evolutionary history of
596 *Symbiodinium* and scleractinian hosts—Symbiosis, diversity, and the effect of
597 climate change. *Perspectives in Plant Ecology, Evolution and Systematics* **8**:23-
598 43. DOI 10.1016/j.ppees.2006.04.001
599 Stat M, Huggett MJ, Bernasconi R, DiBattista JD, Berry TE, Newman SJ, Harvey ES,
600 and Bunce M. 2017. Ecosystem biomonitoring with eDNA: metabarcoding across
601 the tree of life in a tropical marine environment. *Scientific Reports* **7**:12240. DOI
602 10.1038/s41598-017-12501-5
603 Stat M, John J, DiBattista JD, Newman SJ, Bunce M, and Harvey ES. 2018. Combined
604 use of eDNA metabarcoding and video surveillance for the assessment of fish
605 biodiversity. *Conservervation Biology*, *in press*.
606 DOI 10.1111/cobi.13183
607 Takabayashi M, Adams LM, Pochon X, and Gates RD. 2011. Genetic diversity of free-
608 living *Symbiodinium* in surface water and sediment of Hawai‘i and Florida. *Coral*
609 *Reefs* **31**:157-167. DOI 10.1007/s00338-011-0832-5

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019

610 Thomas L, Kendrick GA, Kennington WJ, Richards ZT, and Stat M. 2014. Exploring
611 *Symbiodinium* diversity and host specificity in *Acropora* corals from geographical
612 extremes of Western Australia with 454 amplicon pyrosequencing. *Molecular*
613 *Ecology* **23**:3113-3126. DOI 10.1111/mec.12801
614 Trench R, Wethey D, and Porter J. 1981. Observations on the symbiosis with
615 zooxanthellae among the tridacnidae (Mollusca, Bivalvia). *Biological Bulletin*
616 **161**:180-198.
617 von Ammon U, Wood SA, Laroche O, Zaiko A, Tait L, Lavery S, Inglis G, and Pochon
618 X. 2018. The impact of artificial surfaces on marine bacterial and eukaryotic
619 biofouling assemblages: A high-throughput sequencing analysis. *Marine*
620 *Environmental Research* **133**:57-66.
621 DOI 10.1016/j.marenvres.2017.12.003
622 Weber M. 2009. The Biogeography and Evolution of *Symbiodinium* in Giant Clams
623 (Tridacnidae). *PhD Dissertation, UC Berkeley*:103 pages + appendices.
624 Wecker P, Fournier A, Bosserelle P, Debitus C, Lecellier G, and Berteaux-Lecellier V.
625 2015. Dinoflagellate diversity among nudibranchs and sponges from French
626 Polynesia: insights into associations and transfer. *Comptes Rendus Biologies*
627 **338**:278-283. DOI 10.1016/j.crvi.2015.01.005
628 Zhang GK, Chain FJJ, Abbott CL, and Cristescu ME. 2018. Metabarcoding using
629 multiplexed markers increases species detection in complex zooplankton
630 communities. *Evolutionary Applications* **11**:1901-1914.
631 DOI 10.1111/eva.12694
632

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27313v2 | CC BY 4.0 Open Access | rec: 9 Jan 2019, publ: 9 Jan 2019