EXTENDED MATERIAL AND METHODS

Preparation of Multiplex 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 and Gates 2008; Pochon et al. 2010; this study).

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 genes 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 Illumina 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 MiSeq Illumina™ indexes on to the amplicons. PCR products were combined in equimolar

concentrations and the final library paired-end sequenced on an Illumina MiSeq using a 500 cycle (2 x 250) MiSeq® v2 Reagent Kit and standard flow cell.

Bioinformatics

Illumina™ sequence datasets were prepared using the read preparation and dereplication pipeline of USEARCH (Edgar, 2010). Firstly, paired reads were merged (fastq_mergepairs command) and filtered (fastq_filter command) with an expected number of error of 0.25. More than 90% of the base pairs had a Q score > 40. Next, samples were demultiplexed in three groups, primers were trimmed and a global trimming was operated according to the recommendations for *ITS* amplicon reads (Edgar, 2013). The sequence data were dereplicated and unique singletons found across the complete dataset were discarded.

For phylogenetic assignments of Symbiodiniaceae, three distinct reference databases (*ITS2*, *LSU* and *23S*) were generated in fasta format, including sequence representatives from each of nine Symbiodiniaceae clades (A to I), with (*i*) 409 unique sequences of *ITS2* types from GeoSymbio (Franklin et al., 2012), (*ii*) 37 representative sequences of *LSU* from Pochon et al. (2012), and (*iii*) 104 sequences of *23S* from Takabayashi et al. (2012). Symbiodiniaceae assignments were performed using a novel algorithm approach called 'Kallisto' (Bray et al. 2016) which provides unprecedented speed and accuracy for optimal analysis of large-scale datasets (e.g. large RNA-Seq data) without the need for time-consuming alignment steps.

Because the main goal of the present study was to investigate the sequencing depth and potential inter-marker biases of the multiplex HTS 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 miss-assignments, all merged reads with ambiguities between k-mers of different reference genotypes were determined as chimeric and removed from the dataset. For sequences that did not result in exact matches to reference databases, 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 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.

EXTENDED RESULTS & DISCUSSION

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 multiplexed 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 filtering, 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 (*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 Symbiodiniaceae genotypes 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 genotypes when using the original datasets, and to 46 and 51 unique genotypes following the 0.05% filtering threshold, respectively.

High-throughput sequencing of rDNA barcodes is widely used to study microbial communities. Typically, sequences are clustered into Operational Taxonomic Units (OTUs) assigned by using a pre-defined reference database of known sequences (reference-based) or constructed *de novo*. Many recent studies have demonstrated the difficulties in determining robust clustering and assignments (i.e. reusability, reproducibility, comprehensiveness), and proposed new algorithms and methods for error-correcting (Callahan et al. 2016, 2017; Chen et al. 2016; Edgar 2016, 2017; Edgar and Valencia 2018). Moreover, as for bacteria, the family Symbiodiniaceae represents a large microbial assemblage, consisting of nine clades (seven of which corresponding to newly described genera; Lajeunesse et al. 2018), and subdivided into numerous sub-generic genotypes likely comprising hundreds of species and an estimated 100,000 types of *ITS2* sequences (Arif et al. 2014; Pochon et al. 2014; Thornhill et al. 2017; Ziegler et al. 2017). Because the aim of this work is focused on the comparison between single and multiplex approaches using three genetic markers, we performed assignments without clustering by applying exact matches between the merged reads and k-mers from reference databases. While this method provides a more accurate description of

the diversity present because it is not limited to identifying reference OTU sequences alone, it is not suitable for the discovery of novel diversity.

Diversity diagrams were generated to visualize the sequence abundance of Symbiodiniaceae clade and subclade genotypes recovered from the twelve giant clam samples and among the three investigated genes (Figure 2). The multiplexing approach yielded similar proportions of dominant clades, but with some notable differences. The genus *Symbiodinium* (Clade A) dominated in all three markers, particularly in *23S* (91.8%; dominant subclade type chvA2), with lower but similar proportions between *ITS2* (81.7%; dominant types A3/A6) and *LSU* (83.9%; dominant types A3/A13). The genus *Cladocopium* (Clade C) represented 7.9% (dominant type chvC1), 18.2% (dominant type C1), and 15.0% (dominant type C1) of sequence reads for the *23S*, *ITS2*, and *LSU* markers, respectively. *Geracladium* (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 clades and subclades recovered between the multiplexed samples and the positive (single gene) controls were very similar (Table S4).

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 in most clams (Baillie et al. 2000; DeBoer et al. 2012; Ikeda et al. 2017, 2016; Pappas et al. 2017; Trench et al. 1981). Similar to coral symbiosis, it is assumed that the clade composition 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 and 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} (Maruyama and Heslinga 1997; Buck 2002). 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 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 with *Tridacna* clams (Lee et al. 2015). A3 is the most dominant type in *T. Maxima* around the world and both A3/A6 are more likely to be sampled in giant clams from shallow reefs (Weber, 2009).

Furthermore, for *Cladocopium* we found that the generalist ITS2 type C1 (LaJeunesse et al. 2003) co-dominated in our samples, which is consistent with a previous study showing C1 as a common type in *T. Maxima* from around the world (Weber 2009). Noteworthy, we also found a smaller percentage of C3z, Cspd and C50 ITS2 types, which to our knowledge have not yet been found in *T. maxima* before, but are usually restricted to corals (LaJeunesse et al. 2004, 2010; Macdonald et al. 2008; Shinzato et al. 2018). Finally, we did not detect any symbiont from the genus *Durusdinium* (Clade D) despite the in-depth sequencing afforded by our multiplex method. However, *Durusdinium* has never been detected in *T. maxima* from French Polynesia compared to other regions such as the Indian Ocean (DeBoer et al. 2012; Weber 2009). As we only worked with adult clams from shallow water, it would be interesting to confirm the hypotheses of Ikeda et al. (2017) and Weber (2009) who argued that *Durusdinium* symbionts might be restricted to 'young' *T. squamosa* clams (less than 11 cm) or that giant clams harbored this dinoflagellate genus only when sampled from deeper reefs, respectively.

REFERENCES

Arif C, Daniels C, Bayer T, Banguera-Hinestroza E, Barbrook A, Howe CJ, LaJeunesse TC, Voolstra CR (2014) Assessing *Symbiodinium* diversity in scleractinian corals via next-generation sequencing-based genotyping of the ITS2 rDNA region. Mol Ecol 23:4418-4433. https://doi.org/10.1111/mec.12869

Baillie B, Belda-Baillie C, Silvestre V, Sison M, Gomez A, Gomez ED, Monje VD (2000) Genetic variation in *Symbiodinium* isolates from giant clams based on random-amplified-polymorphic DNA (RAPD) patterns. Mar Biol 136:829-836.

Baker A (2003) Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of *Symbiodinium*. Annu Rev Ecol Evol Syst 34:661-689.

Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525.

Buck B (2002) Effect of increased irradiance and thermal stress of the symbiosis of *Symbiodinium microadriaticum* and *Tridacna gigas*. Aquat Living Resour 15:107-117.

Callahan BJ, McMurdie PJ, Holmes SP (2017) Exact sequence variants should replace opertational taxonimic units in marker-gene data analysis. ISME J 11:2639-2643. https://doi.org/10.1038/ismej.2017.119

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581-583. https://doi.org/10.1038/nmeth.3869

Chen Y, Deng F, Jia X, Liu YP, Lai SJ (2016) bioOTU: An improved method for simultaneous taxonomic assignments and operationals taxonomic units clustering of 16sRNA gene sequences. J Comput Biol 23:229-238. https://doi.org/10.1089/cmb.2015.0214

DeBoer T, Baker A, Erdmann M, Ambariyanto, Jones P, Barber P (2012) Patterns of *Symbiodinium* distribution in three giant clam species across the biodiverse Bird's Head region of Indonesia. Mar Ecol Prog Ser 444:117-132. https://doi.org/10.3354/meps09413

Dubousquet V, Gros E, Berteaux-Lecellier V, Viguier B, Raharivelomanana P, Bertrand C, Lecellier GJ (2016) Changes in fatty acid composition in the giant clam *Tridacna maxima* in response to thermal stress. Biol Open 5:1400-1407. https://doi.org/10.1242/bio.017921

Edgar RC, Valencia A (2018) Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics 34:2371-2375. https://doi.org/10.1093/bioinformatics/bty113

Edgar RC (2017) SEARCH_16S: A new algorithm for identifying 16S ribosomal RNA genes in contigs and chromosomes. bioRxiv 124131. https://doi.org/10.1101/124131

Edgar RC (2016) UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. bioRxiv 081257. https://doi.org/10.1101/081257

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods 10:996.

Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460-2461.

Felsenstein J (1985) Phylogenies and the Comparative Method. Am Nat 125:1-15.

Fitt W, Trench R (1981) Spawning, development, and acquisition of zooxanthellae by *Tridacna squamosa* (Mollusca, Bivalvia). Biol Bull 161:213-235.

Franklin EC, Stat M, Pochon X, Putnam HM, Gates RD (2012) GeoSymbio: a hybrid, cloud-based web application of global geospatial bioinformatics and ecoinformatics for *Symbiodinium*-host symbioses. Mol Ecol Resour 12:369-373. https://doi.org/10.1111/j.1755-0998.2011.03081.x

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 45:95-98.

Ikeda S, Yamashita H, Kondo S, Inoue K, Morishima S, Koike K (2017) Zooxanthellal genetic varieties in giant clams are partially determined by species-intrinsic and growth-related characteristics. PLOS ONE 12:e0172285. https://doi.org/10.1371/journal.pone.0172285

Ikeda S, Yamashita H, Koike K, Liao LM (2016) simple and rapid determination method for zooxanthellal genetic diversity in giant clams using multiplex PCR. Fish Sci 82:747-753.

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870-1874. https://doi.org/10.1093/molbev/msw054

LaJeunesse TC, Parkinson JE, Gabrielson PW, Jeong HJ, Reimer JD, Voolstra CR, Santos SR (2018). Systematic revision of Symbiodiniaceae highlights the antiquity and diversity of coral endosymbionts. Curr Biol 28:1-11

LaJeunesse TC, Smith R, Walther M, Pinzón J, Pettay DT, McGinley M, Aschaffenburg M, Medina-Rosas P, Cupul-Magaña AL, Pérez AL, Reyes-Bonilla H, Warner ME (2010) Host-symbiont recombination versus natural selection in the response of coral-dinoflagellate symbioses to environmental disturbance. Proc Biol Sci 277:2925-2934. https://doi.org/10.1098/rspb.2010.0385

LaJeunesse TC, Thornhill DJ, Cox EF, Stanton FG, Fitt WK, Schmidt GW (2004) High diversity and host specificity observed among symbiotic dinoflagellates in reef coral communities from Hawaii. Coral Reefs 23:596-603. https://doi.org/10.1007/s00338-004-0428-4

LaJeunesse TC, Loh WKW, Woesik R van, Hoegh-Guldberg O, Schmidt GW, Fitt WK (2003) Low symbiont diversity in southern Great Barrier Reef corals, relative to those of the Caribbean. Limnol Oceanogr 48:2046-2054. https://doi.org/10.4319/lo.2003.48.5.2046

Lee SY, Jeong HJ, Kang NS, Jang TY, Jang SH, 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. Eur J Phycol 50:155-172. https://doi.org/10.1080/09670262.2015.1018336

Macdonald AH, Sampayo EM, Ridway T, Schleyer MH (2008) Latitudinal symbiont zonation in *Stylophora pistillata* from southeast Africa. Mar Biol 154:209-217.

Manning MM, Gates RD (2008) Diversity in populations of free-living *Symbiodinium* from a Caribbean and Pacific reef. Limnol Oceanogr 53:1853-1861. https://doi.org/10.4319/lo.2008.53.5.1853

Maruyama T, Heslinga GA (1997) Fecal discharge of zooxanthellae in the giant clam *Tridacna derasa*, with reference to their in situ growth rate. Mar Biol 127:473-477. https://doi.org/10.1007/s002270050035

Neo ML, Eckman W, Vicentuan K, Teo SLM, Todd PA (2015) The ecological significance of giant clams in coral reef ecosystems. Biol Conserv 181:111-123. https://doi.org/10.1016/j.biocon.2014.11.004

Pappas MK, He S, Hardenstine RS, Kanee H, Berumen ML (2017) Genetic diversity of giant clams *Tridacna* spp. and their associated *Symbiodinium* in the central Red Sea. Mar Biodivers 47:1209-1222. https://doi.org/10.1007/s12526-017-0715-2

Pochon X, Stat M, Takabayashi M, Chasqui L, Chauka LJ, Logan DDK, Gates RD (2010) Comparison of endosymbiotic and free-living *Symbiodinium* (dinophyceae) diversity in a hawaiian reef environment. J Phycol 46:53-65. doi:10.1111/j.1529-8817.2009.00797.x

Pochon X, Putnam HM, Burki F, Gates RD (2012) Identifying and characterizing alternative molecular markers for the symbiotic and free-living dinoflagellate genus *Symbiodinium*. PloS ONE 7:e29816. https://doi.org/10.1371/journal.pone.0029816

Pochon X, Putnam HM, Gates RD (2014) Multi-gene analysis of *Symbiodinium*

dinoflagellates: a perspective on rarity, symbiosis, and evolution. PeerJ 2:e394. https://doi.org/10.7717/peerj.394

Shinzato C, Zayasu Y, Kanda M, Kawamitsu M, Satoh N, Yamashita H, Suzuki G (2018) Using seawater to document Coral-Zoothanthella diversity: A new approach to coral reef monitoring using environmental DNA. Front Mar Sci 5:28. https://doi.org/10.3389/fmars.2018.00028

Takabayashi M, Adams LM, Pochon X, Gates RD (2012) Genetic diversity of free-living *Symbiodinium* in surface water and sediment of Hawai`i and Florida. Coral Reefs 31:157-167. https://doi.org/10.1007/s00338-011-0832-5

Thornhill DJ, Howells EJ, Wham DC, Steury TD, Santos SR (2017) Population genetics of reef coral endosymbionts (*Symbiodinium, Dinophyceae*). Mol Ecol 26:2640–2659. https://doi.org/10.1111/mec.14055

Trench R, Wethey D, Porter J (1981) Observations on the symbiosis with zooxanthellae among the *tridacnidae* (Mollusca, Bivalvia) 161:180-198.

Weber M (2009) The Biogeography and Evolution of *Symbiodinium* in Giant Clams (Tridacnidae). UC Berkeley (103 pages + appendices).

Ziegler M, Eguíluz VM, Duarte CM, Voolstra CR (2018) Rare symbionts may contribute to the resilience of coral–algal assemblages. ISME J 12:161-172. https://doi.org/10.1038/ismej.2017.151