An improved primer set and amplification protocol with increased specificity and sensitivity targeting the *Symbiodinium* ITS2 region

Benjamin CC Hume[1]†, Maren Ziegler[1], Julie Poulain[2,3,4], Xavier Pochon[3,5,6], Sarah Romac[4,7], Emilie Boissin[5,8], Colomban de Vargas[4,7], Serge Planes[5,8], Patrick Wincker[2,3,4], Christian R Voolstra[1]†

[1]Red Sea Research Center, Division of Biological Sciences and Engineering, KAUSTKing Abdullah University of Science and Technology, Thuwal, 23955-6900, Saudi Arabia

[2]2CEA - Institut de Biologie François Jacob, Genoscope, 2 rue Gaston Crémieux, 91057 Evry, France. Institut de Génomique, Génoscope, Evry, France

[3] CNRS, UMR 8030, CP5706, Evry, France.

[4] Université d'Evry, UMR 8030, CP5706, Evry, France.

[3] 5 Coastal and Freshwater Group, Cawthron Institute, Private Bag 2, Nelson, 7042, New Zealand.

6 Institute of Marine Science, University of Auckland, Private Bag 349, Warkworth 0941, New Zealand.

[4,7] CNRS, UMR 7144, EPEP & Sorbonne Universités, UPMC Université Paris 06; Station Biologique de Roscoff, 29680 Roscoff, France.

[5,8] PSL Research University: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France

†Corresponding authors:

Christian R Voolstra

Bldg 2, Rm 2226

22Red Sea Research Center, KAUST

23Phone: +966-12-808 2377

24Email: christian.voolstra@kaust.edu.sa

25

26Benjamin C C Hume

27Bldg 2, Rm 2227 WS01

28Red Sea Research Center, KAUST

29Phone: +966-544-701-291

30Email: benjamin.hume@kaust.edu.sa

31

32Running title: Assessment of *Symbiodinium* ITS2 primers ~~pairssets~~

33~~Keywords: Coral Reef, Metabarcoding, Metagenomics, Next-generation sequencing, Microbial~~

34~~Biology, *Symbiodinium*~~

Abstract

The Internal Transcribed Spacer 2 (ITS2) rRNA gene is a commonly targeted genetic marker to assess diversity of *Symbiodinium*, a dinoflagellate genus of algal endosymbionts that is pervasively associated with marine invertebrates, and notably reef-building corals. Here we tested three commonly used ITS2 primer pairs (SYM_VAR_5.8S2/SYM_VAR_REV, ITSintfor2/ITSReverse, and ITS-DINO/ITS2Rev2) with regard to amplification specificity and sensitivity towards *Symbiodinium*, as well as sub-genera taxonomic bias. We tested these primers over a range of sample types including three coral species, coral surrounding water, reef surface water, and open ocean water to assess their suitability for use in large-scale next generation sequencing projects and to develop a standardized PCR protocol. We found the SYM_VAR_5.8S2/SYM_VAR_REV primers to perform superior to the other tested ITS2 primers. We therefore used this primer pair to develop a standardised PCR protocol. To do this, we tested the effect of PCR-to-PCR variation, annealing temperature, cycle number, and different polymerase systems on the PCR efficacy. The *Symbiodinium* ITS2 PCR protocol developed here delivers improved specificity and sensitivity towards *Symbiodinium* with apparent minimal sub-genera taxonomic bias across all sample types. In particular, the protocol's ability to amplify *Symbiodinium* from a range of environmental sources will facilitate the study of *Symbiodinium* populations across biomes.

Introduction

Coral reefs sustain some of the highest levels of biodiversity on Earth and provide a range of services to communities totalling millions of people (Moberg & Folke 1999; Plaisance et al. 2011). However, these ecosystems are being lost at an alarming rate (Hughes et al. 2017; Norstrom et al. 2016). This loss is primarily due to anthropogenic stressors degrading the scleractinian~~Scleractinian~~ corals that build and support these reefs. Interestingly, susceptibility to these stressors amongst reef-building corals is often not homogenous (Kemp et al. 2014; Pandolfi et al. 2011; Rowan et al. 1997).

Coral resilience is determined not only by the animal genotype but also by the diversity of microbes that associate with the animal host (Bourne et al. 2016; Levin et al. 2017; Peixoto et al. 2017; Pogoreutz et al. 2017; Radecker et al. 2015; Torda et al. 2017; Ziegler et al. 2017c). Of the microbial components that make up the coral holobiont - the consideration of the animal host and its associating microbes as a functional ecological unit – the algal symbionts of the genus *Symbiodinium* have received the most attention for their role in affording resilience to the coral host (Hume et al. 2016; Thornhill et al. 2017). The coral-*Symbiodinium* association is generally obligate with the algal symbiont providing towards the nutritional needs of the animal host, in exchange for a stable, beneficial environment (Muscatine 1990). The efficacy and character of this association is~~,~~ determined by the genotype of alga (Kemp et al. 2014; Rowan et al. 1997; Silverstein et al. 2015; Thornhill et al. 2017). Accordingly, the ability to resolve taxa within the genus *Symbiodinium* is essential to better understanding resilience of the coral holobiont and reef ecosystems as a whole.

Taxonomic resolution within *Symbiodinium* is primarily achieved genetically (Wham et al. 2017). To this end, a range of genetic markers and associated analytical approaches exist (Pochon et al. 2014; Thornhill et al. 2017). Initially, genetic characterisations resolved *Symbiodinium* into broad taxonomic groupings referred to as clades A-I (Pochon & Gates 2010). However, due to the significant genetic and phenotypic diversity found within these clades, contemporary genetic resolutions are conducted at a sub-cladal level. For such analyses, the most commonly used marker is the ITS2 region of the nrDNA (LaJeunesse 2001). This marker is multi-copy in nature meaning that a single *Symbiodinium* cell may contain 100s-1000s of copies of the ITS2 region (Arif et al. 2014; LaJeunesse 2002). As such, 100s of distinct ITS2 sequences may be generated from a single genotype, referred to as intragenomic diversity. Given that corals may associate with multiple *Symbiodinium* genotypes, intergenomic diversity may also exist in samples. Disentangling these two diversities can complicate analyses, but ultimately, the intragenomic diversity of the ITS2 region has proven to be a rich source of information that has been effectively leveraged to improve resolutions within *Symbiodinium* **(Smith et al. 2017; Wham et al. 2017)**.

Symbiodinium ITS2 intragenomic diversity was first employed taxonomically using denaturing gradient gel electrophoresis (DGGE) methodologies, enabling the separation of PCR amplicon sequence fragments according to melting temperature, a proxy for sequence (LaJeunesse 2002). Using such techniques, *Symbiodinium* taxa that shared the same most abundant ITS2 sequence in common, yet displayed significantly different functional phenotypes, were able to be resolved specifically according to differences in rarer sequences in their intragenomic diversity. Importantly, *Symbiodinium* taxa began to be resolved in part by~~partby~~ specific sets of these

intragenomic ITS2 sequences referred to as ITS2 types or profiles. For example, *S. trenchii* and *S. glynnii* have the D1 ITS2 sequence in common but can be differentiated according to different presence-absence combinations of the D4, and D6 sequences (Wham et al. 2017). Similarly, *S. thermophilum*, a symbiont prevalent in corals living in the world's warmest reefs, can be differentiated from other C3 dominated symbionts, that have a global distribution, through the identification of a specific intragenomic sequence, *S. thermo*.-indel (Hume et al. 2015).

More recently, studies have been taking advantage of the sequencing benefits afforded by next generation sequencing (NGS) of PCR amplicons to elucidate *Symbiodinium* diversity using the ITS2 region. Multiple approaches to dealing with the sequence diversity have been taken, including, but not limited to, OTU clustering (Arif et al. 2014; Cunning et al. 2017; Ziegler et al. 2017a; Ziegler et al. 2017b), searching for specific taxa-defining ITS2 sequences (Boulotte et al. 2016; Hume et al. 2016; Ziegler et al. 2017a) or looking for ITS2 profiles found in common between samples (Smith et al. 2017).

Given the ongoing surge in *Symbiodinium* diversity studies employing NGS approaches using the ITS2 marker, and the large-scale projects that rely on them, such as The Global Coral Microbiome Project (http://vegathurberlab.oregonstate.edu/global-coral-microbiome-project) and the *Tara*~~*TARA*~~ Pacific expedition (http://oceans.taraexpeditions.org/en/m/environment/ocean-climate/new-expedition-tara-pacific/), we sought to establish, and future proof, a standardised and improved PCR protocol. Particularly, we assessed the specificity (preferential amplification of *Symbiodinium* DNA), sensitivity (ability to amplify *Symbiodinium* from *Symbiodinium*-rare environments), and relative sub-genera taxonomic bias (within *Symbiodinium* bias) of existing ITS2 primer pairs, as well as the effect of different polymerase systems. Notably, we tested

amplification on a range of sample types including coral, coral surrounding water, reef surface water, and open ocean water due to increasing efforts to elucidate environmental reservoirs and assess free-living *Symbiodinium* (Mordret et al. 2016; Nitschke et al. 2016; Pochon et al. 2010; Thornhill et al. 2017).

Materials & Methods

1. Sample collection, DNA isolation, ITS2 primer pairs, and experimental setup

To establish a robust ITS2 NGS-based amplification protocol for *Symbiodinium*, we assessed the effectiveness of three commonly used primer pairs (Table 1) on samples from different environments (Table 2). Further, we tested the robustness of the most effective primer pair, SYM_VAR, by documenting the effects of PCR-to-PCR variation, annealing temperature, cycle number, and polymerase system used with regards to: specificity, sensitivity, relative sub-genera amplification, and which sequences were returned. Chimeric sequence formation was additionally assessed as a function of PCR cycles.

Sample collection: In total, 27 samples were used in our analyses from 4 different environments: coral tissue (CO), coral surrounding water (CSW), coral reef surface water (SRF-CO), and open ocean (SRF-OO~~OA~~). Samples were collected either as part of the *Tara~~TARA~~* Oceans or *Tara~~TARA~~* Pacific expeditions. Field Permits were obtained as part of the *Tara~~TARA~~* Oceans and *Tara~~TARA~~* Pacific expeditions and the approving bodies were the respective governments and/or ministries for the environment of the countries where the samples were collected. The authors of this work are part of the *Tara~~TARA~~* Oceans and *Tara~~TARA~~* Pacific expeditions, and thus, participated in sample collection. ***Tara*** Expeditions are global scientific voyages aimed at probing morphological and molecular diversity, evolution and ecology of marine plankton (viruses, bacteria, archaea, protists and planktonic metazoans) in the photic layer of the world oceans to research how they are impacted by Earth's climate changes. ***Tara*** meteorological, oceanographic, biochemical, and plankton morphological data are archived in the comprehensive

Sample Registry of the PANGAEA database (https://www.pangaea.de).

~~(https://www.pangaea.de).~~ The Sample Registry links physical samples to metadata about sampling and analysis methodology performed on each sample. Whilst the *Tara* Pacific registry is still under construction at the time of writing of this manuscript, the *Tara* Oceans registry can be accessed at https://doi.pangaea.de/10.1594/PANGAEA.842237.

Six CO samples from three genera, 2 *Porites lobata* (samples: CO-0000150 and CO-0000151),~~,~~ 2 *Pocillopora meandrina* (samples: CO-0000208 and CO-0000209),~~,~~ and 2 *Millepora platyphylla* (samples: CO-0000302 and CO-0000303),~~,~~ were collected on SCUBA off the coast of Panama in July 2016. Corals were stored in DNA/RNA Shield (Zymo Research, USA).

Two CSW and one SRF-CO samples were also collected off the coast of Panama in July 2016. CSW samples were collected (15-17 L) from water surrounding 2 *Pocillopora* corals using a boat-mounted vacuum pump. SRF-CO samples were collected (18 L) directly from the surface of the reef. Water was pre-filtered through a 20 µm metallic sieve, and then vacuum filtered on a 3 µm polycarbonate filter of 142 mm diameter for 15 mins. After filtration, filters were immediately flash frozen in liquid nitrogen.

In addition, six SRF-OO~~OA~~ samples were collected off the coast of Panama and were processed as described above for the CSW and SRF-CO samples. Twelve additional samples were collected as part of the *Tara* Oceans expeditions~~TARA Ocean~~ from a range of ocean basins including the Pacific, North Atlantic, and Indian Ocean. These samples were filtered to a size fraction of 5-20 µm as described in section 6a of Pesant *et al.* 2015.

DNA isolation: DNA was extracted using the ZR-Duet DNA/RNA MiniPrep Plus (Zymo Research, USA) kit for all coral samples. Further, DNA was extracted from all water samples as

described in Alberti~~Alerti~~ et al.~~,~~ (2017)~~)(Alberti et al. 2017)~~ for the 5-20 µm and 3-20 µm fractions. In brief, the protocol was based on simultaneous extraction of DNA and RNA by cryogenic grinding of cryopreserved membrane filters, followed by nucleic acid extraction with NucleoSpin RNA kits (Macherey-Nagel, Germany) combined with DNA Elution buffer kit (Macherey-Nagel, Germany).

ITS2 primer pairs: Three pairs~~sets~~ of published ITS2 primer pairs specifically designed to amplify *Symbiodinium* DNA were tested. These were: SYM_VAR_5.8S2/SYM_VAR_REV(Hume et al. 2013; Hume et al. 2015), ITSintfor2/ITS-reverse (Coleman et al. 1994; LaJeunesse 2002), and ITS-DINO/ITS2Rev2 (Pochon et al. 2001; Stat et al. 2009). Primer details and amplification protocols are given in Table 1. For ease of comprehension, each primer pair will be referred to as the SYM_VAR, ITSintfor2, or ITS-DINO primer pair.

Experimental setup: Sequencing for this study was conducted in two consecutive efforts. In the first part, aimed at directly comparing the three primer pairs, each primer pair was used to amplify each of the 27 samples. In the second part, a subset of samples was amplified using the SYM_VAR primer pair with different PCR protocols. We tested the effect of number of cycles, annealing temperature, and polymerase system on selectivity, sensitivity, and stability to optimise the reaction. Details of the samples as well as the associated number of cycles and annealing temperatures used can be found in Table 2 and Supplementary Table 1. All PCRs were carried out in 25 µl reactions with the following conditions: 12.5 µl Phusion High-Fidelity PCR Master Mix 2X (ThermoFisher Scientific, USA) or equivalent amount of Advantage 2 PCR system (Takara Bio, Japan), 1 µl of forward and reverse primer at a concentration of 10 µM, 0.75 µl of

DMSO, 1 µl of genomic DNA at a concentration of between 5-10 ng and 8.75 µl of ddH_20 with a GeneAmp PCR system 9700 (Perkin Elmer, USA).~~.~~

2. Library preparation, sequencing, and sequence quality control

For both sequencing efforts, all PCRs were conducted in triplicate before being pooled for sequencing. Pooled PCR products were purified using 0.8 x AMPure XP beads (Beckmann Coulter Genomics, USA), then aliquots of purified amplicons were run on an Agilent Bioanalyzer (Agilent technologies, USA) using the DNA High Sensitivity LabChip kit to verify lengths and were quantified using a Qubit Fluorometer (ThermoFisher Scientific, USA).

One hundred ng aliquots of the amplicons were directly end-repaired, A-tailed, and ligated to Illumina adapters (Illumina, USA) on a Biomek FX Laboratory Automation Workstation (Beckman Coulter, USA). Then, library amplification was performed using the KAPA Hifi HotStart NGS kit (KAPA Biosystems, USA). After library profile **analysis by** LabChip GX (Perkin Elmer, USA) and qPCR quantification using PicoGreen (ThermoFisher Scientific, USA) ~~on~~in 96-well plates, each library was sequenced using 250 bp read length chemistry in a paired-end flow cell on a MiSeq instrument (Illumina, USA).

During sequencing, Illumina's Real Time Analysis (RTA) was run with default settings to remove clusters with the least reliable data. Output BCL files were converted to demultiplexed fastq files using Illumina's *bcl2fastq* package and in-house filtering was applied to reads that passed the Illumina quality filters (Alberti et al. 2017). Read pairs that mapped onto run quality control sequences (Enterobacteria phage PhiX174 genome) were removed using the *bowtie* package.

For each of the sequenced samples analysed in this study the same sequence quality control (QC) was performed. Briefly, mothur~~mother~~ _1.39.5 (Schloss et al. 2009) was used to create contigs from paired forward and reverse demultiplexed fastq.gz files using the *make.contigs* command. The *screen.seqs* command was then used (*maxambig=0, maxhomop=5*) to discard sequences putatively generated from sequencing errors. The *unique.seqs* command was used to create a non-redundant collection of sequences represented by a .name and .fasta file. The remainder of the QC was performed using both the .name and .fasta files produced. The *pcr.seqs* command (*pdiffs=2, rdiffs=2*) was used to trim the primer sequence regions from the returned sequences. In addition, this command was used to discard sequences in which the specific primer pairs could not be found – indicative of poor sequencing quality – allowing a deviation of no more than 2 nucleotides differences in either of the forward or reverse primer sequences. Next, *split.abund* was run (*cutoff=2*) to discard sequences that were not found more than 2 times in each sequenced sample – again, to reduce incorporating sequences with sequencing errors. Sequences were once again made non-redundant using the *unique.seqs* command. Finally, the *summary.seqs* command was used to calculate the size distribution of the remaining sequences. Sequences below the 2.5% and above the 97.5% size percentiles were removed using the *screen.seqs minlength* and *maxlength* parameters. A summary of the number of sequences retained at three stages of the QC process (pre-QC, post-QC, and *Symbiodinium* only sequences) are given in Supplementary Table 1 and Supplementary Fig. 1.

3. Sequence data analysis – primer pair comparisons

To assess which of the three primer pairs performed the best, each primer pair was assessed for specificity (i.e., preferential amplification of *Symbiodinium* DNA) and sensitivity (i.e., ability to

amplify *Symbiodinium* from *Symbiodinium*-rare environments). Additionally, potential within-*Symbiodinium* taxonomic biases were assessed by comparing the proportions of sequences for each primer pair that belonged to each of the 9 major cladal groups of *Symbiodinium*, clades A-I (Pochon & Gates 2010). Our assumption was that if no or only small ~~cladal~~clade biases exist, the proportion of sequences from each clade should be similar between primers. Consequently, a large deviation in ~~cladal~~clade proportions by any one of the primer pairs~~'~~ was regarded as indicative of taxonomic bias.

The specific PCR conditions used for each primer pair~~set~~ are given in Table 1 with PCR reagents as detailed above. The sequences returned from each PCR were annotated using *blastn* and the NCBI 'nt' database according to their closest match to one of the following categories: *Symbiodinium*, dinoflagellate, stony coral, Hydrozoan, uncultured eukaryote, other, or 'no match'. To do this, the .fasta file produced from the initial QC from each sample was run against NCBI's 'nt' database with the *max_target_seqs* argument set to 1 and an output format string of *'6 qseqid sseqid staxids sscinames sblastnames evalue*'. For sequences to be categorised as *Symbiodinium*, an e-value > 50 was required (representing approximately ~~a~~ 100% coverage match at 80-85% nt identity) in addition to the closest match being of *Symbiodinium* origin. Additionally, closest match subject sequences were screened for two sequences in particular (JN406302 and JN406301), which are mis-annotated as *Symbiodinium* (highly divergent from any other *Symbiodinium* sequences) in the 'nt' database. Thus, query sequences matching these sequences were annotated as 'Unclutured eukaryote'. Notably, before controlling for this, 58% of all '*Symbiodinium*' sequences amplified by the ITSintfor2 and ITS-DINO primer pairs (0% for

the SymVar primer pair) in the SRF-OO~~OA~~ samples matched these sequences as their closest match (Supplementary Fig. 2~~1~~).

blastn was also used to associate *Symbiodinium* sequences to one of the 9 ~~cladal groups~~clades. To do this, a .fasta file was created for each sample that contained all sequences that had previously been categorised as *Symbiodinium*. This file was used as an input for *blastn* with the *max_target_seqs* argument set to 1 and an output format string of *'6 qseqid sseqid evalue pident gcovs*'. A custom BLAST database was used that contained a single ~~cladal~~ representative sequence for each of the 9 clades (Supplementary Data 1).

4. Sequence data analysis – optimizing the SYM_VAR PCR protocol

Having determined SYM_VAR as the most effective primer pair for the amplification of the *Symbiodinium* ITS2 region, we undertook a second sequencing effort to optimize the PCR protocol. Specifically, we investigated: 1) how consistent the PCR was (i.e. PCR-to-PCR effect) by conducting PCRs with the exact same reaction conditions but sequenced ~~on~~in different sequencing runs, 2)~~),~~ whether increasing the annealing temperature could increase sensitivity in non-coral sample environments, and 3) whether a decrease~~an increase~~ in the number of PCR cycles would still allow for a satisfactory ~~actually aids in the~~ amplification of *Symbiodinium* whilst decreasing~~from samples representing *Symbiodinium*-rare environments. Cladal variation~~ ~~and~~ quantity of chimeric sequence~~s return~~. ~~formation were additionally considered, when investigating the effects of increased annealing temperature and number of PCR cycles.~~

PCR conditions were the same as for the SYM_VAR_5.8S2/SYM_VAR_REV primer pair (Table 1), but with a 30 sec denaturing step, following the manufacturer's recommendation (Phusion High Fidelity PCR; ThermoFisher Scientific, USA).~~.~~ Taxonomic identities and ~~the~~

~~cladal~~clade breakdown (*Symbiodinium* sequences only) of sequences returned from each sample (annealing temperature and cycle number comparisons only) were determined in the same way as for the initial primer pair~~set~~ comparisons described above.

As a final optimisation ~~to~~of the SYM_VAR amplification protocol, we tested whether the polymerase enzyme system used could affect the outcome of the PCR amplification. For 11 samples (Table 2), we contrasted amplification with Phusion High-Fidelity PCR Master Mix (ThermoFisher Scientific, USA) and the Advantage 2 PCR Enzyme System (Takara Bio, Japan). PCR amplification with the Advantage 2 PCR Enzyme System were conducted as shown in Table 1, but with 30 cycles.

To further assess the potential effect of PCR protocol modifications, we additionally assessed whether PCR-to-PCR variation, or any of the protocol optimisation modifications would affect which sequences were amplified. To assess this, the top 20 most abundant sequences (by relative abundance) were identified and plotted by decreasing abundance for pairings of sequenced samples for which PCR conditions were identical, or only differed in one of the investigated parameters (i.e., cycle number,~~r~~ annealing temperature, or polymerase system). This analysis was conducted for two samples, AW-0000035 and IW-0000015 and only for *Symbiodinium* sequences contained in each. It should finally be noted that although we did not test for it here, variability between different PCR cyclers (e.g. in ramping rate and accuracy of temperature binning) also has the potential to cause variation in a given PCR.

An online version of the final optimized PCR protocol is available through the doi: dx.doi.org/10.17504/protocols.io.n8edhte.

5. Design of the SYM_VAR primer pair

The SYM_VAR primer pair were designed to minimise ~~cladal~~clade bias and coral-host complementarity. Originally, the reverse primer of the SYM_VAR primer pair that binds in the 28S region of the rDNA, SYM_VAR_REV, was designed to be complimentary to the forward primer SYM_VAR_FWD that binds in the 18S (Hume et al. 2013). This primer pair was designed to amplify both ITS regions as well as the 5.8S region of the rDNA. These primers were designed with the incorporation of degenerate bases based upon a multiple sequence alignment containing representatives of all *Symbiodinium* clades so that, where possible, the primer sequences had an equal number of nucleotide mis-matches to each of the clades. Later, to be able to amplify only the ITS2 region of the rDNA, the SYM_VAR_5.8SII primer was designed (Hume et al. 2015). Compatibility to the SYM_VAR_REV primer and a minimisation of ~~cladal~~clade bias were maintained whilst complementarity to coral host ITS2 rDNA sequences was also taken into account by including representative sequences from *P. lobata* and *P. lutea* (ITS2 sequences from which, had been amplified using alternative primers) to the multiple sequence alignment (Hume 2013). It should be noted that this primer pair was successfully used to recover *Symbiodinium* DNA from *Acropora downingi*, *Cyphastrea microphthalma*, *Favia pallida*, *Platygyra daedalea*, *Porites harrisoni*, and *P. lutea* in Hume et al (2015).

~~6.~~ Sampling permits

The following permits are were required for the collection of the samples analysed as part of this study: Panama scientific sampling: SE/AP-18-16; Panama CITES: SEX/A-72-16, no. 05567; ~~Portulgal~~Portugal scientific sampling: 080522/2009/Proc.º E.17.b; Maldives scientific sampling: FA-D2/33/2010/02; Chile scientific sampling: 13270/24/74.

Results

1. Primer pair comparisons for amplification of *Symbiodinium* ITS2

In the CO samples, the SYM_VAR primer pair exclusively amplified *Symbiodinium* DNA. For the ITSintfor2 and ITS-DINO primer pairs *Symbiodinium* sequences represented 84% (SD 11%) and 79% (SD 12%) of total reads, respectively. Remaining sequences associated with Scleractinia, Hydrozoa, or resulted in no BLAST match (Fig 1).

In the CSW water samples, the SYM_VAR primer pair maintained a high specificity for *Symbiodinium* (98%; SD 3%). In contrast, the alternative pairs amplified 51% (SD 23%) and 42% (SD 24%) of *Symbiodinium* reads, respectively, with the remaining sequences characterised as dinoflagellate or falling into the uncultured eukaryote category (Fig 1).

In the SRF-CO sample, *Symbiodinium* amplifications were considerably lower compared to the CO and CSW samples. SYM_VAR amplified 43% of *Symbiodinium* reads compared to 1% in both alternative primer pairs. The remaining sequences in the SYM_VAR pair were primarily of dinoflagellate origin. In the other two primer pairs, although a large proportion were also dinoflagellate in origin, sequences primarily fell into the uncultured eukaryote category (Fig 1).

In the SRF-OO~~OA~~ samples, the ITSintfor2 and ITS-DINO primer pairs were more similar in their amplifications compared to the SYM_VAR pair. In contrast to the other primer pairs, SYM_VAR amplified less ‘uncultured eukaryotic’, ‘other’, and ‘no match’ sequences, and more dinoflagellate and *Symbiodinium* sequences. Specifically, amplification of *Symbiodinium* was achieved in 12, 7 and 8 of the SRF-OO~~OA~~ samples using the SYM_VAR, ITSintfor2 and ITS-DINO primer pairs, respectively. All of the 5 libraries that achieved a *Symbiodinium*

amplification above 5% of total reads were amplified with the SYM_VAR primer pair. In total, the SYM_VAR primer pair achieved a *Symbiodinium* amplification of 7% (SD 15%) compared to <0.5% (SDs < 0.5%) for the alternative primer pairs (Fig 1).

Across all samples in which all three primer pairs amplified > 5% of *Symbiodinium* reads, the proportions of sub-genus taxa amplified (*Symbiodinium* clades) were comparable (Fig 1).

2. Establishment of an optimal PCR protocol using the SYM_VAR primer pair

First, we assessed whether the same *Symbiodinium* ITS2 sequences were returned from PCRs performed on the same samples, with identical reaction conditions, but sequenced ~~on~~in different sequencing runs. For the two samples tested, the 20 most abundant sequences were present in both PCRs ~~pairings~~ with relative abundances of each sequence being comparable (Figure 2).

Next, we increased the annealing temperature from 56°C to 59°C to see whether a gain in *Symbiodinium* selectivity, and therefore sensitivity, could be achieved without yielding sub-genus taxonomic bias. In the CSW and SRF-CO samples, increasing annealing temperature increased the proportion of returned *Symbiodinium* sequences from 95 to 98% and from 28 to 47%, respectively. However, in both samples a small increase in the proportion of clade D versus clade C was observed, 45 to 47% and from 57 to 66%, respectively. In the two SRF-OO~~OA~~ samples tested, no *Symbiodinium* amplification occurred at either 56°C or 59°C (Fig 3). The 20 most abundant sequences were found in both libraries for the CSW samples tested. However, 2 of the 20 sequences were missing from one of the libraries in the SRF-CO sample tested (the 10th and 16th most abundant sequences; Figure 2).

Additionally, we tested whether *Symbiodinium* could be sufficiently amplified using a decreased ~~an increase in the~~ number of PCR cycles ~~actually improves the amplification of *Symbiodinium* from samples representing *Symbiodinium*-rare environments without leading~~ to reduce~~an increase in~~ chimeric sequence formation or taxonomic bias. To accomplish this, we compared ~~by comparing~~ the above-used protocol to one with only~~a reduced number of~~ 25 cycles. In the CSW and SRF-CO samples there was a greater~~an increase in the~~ proportion of *Symbiodinium* amplified at 35 compared to 25 cycles (~5% greater~~increase~~ in the CSW samples and 15% in the SRF-CO samples). However, in both of the SRF-OO~~OA~~ samples no *Symbiodinium* amplification was seen at~~in~~ either ~~of the~~ cycle conditions. Only non-*Symbiodinium*, dinoflagellate taxa sequences were amplified. Of the 9 samples considered in the chimeric analysis (the 3~~three~~ CSW and SRF-CO samples considered above, as well as 6~~six~~ CO samples) greater number of~~an increase in predicted~~ chimeric sequences were putatively identified~~was observed~~ at 35 cycles versus 25 cycles in 3 of them with an increase of 1% in each case (Table 3). The proportions of sub-genus taxa amplified (*Symbiodinium* clades) were comparable (Fig 3) and all but one (the 20th) of the most abundant sequences were amplified in common (Fig 2).

Last, we also tested to see whether there would be any effect of using different polymerase systems. In the CO and SRF-OO~~OA~~ samples tested, no difference in proportion of returned *Symbiodinium* sequences was observed between polymerase systems used (100% and 0% amplification for all libraries, respectively). However, in the CSW and SRF-CO samples we observed a relative decrease of *Symbiodinium* reads in the libraries that used the Advantage 2 system (Fig. 4). Despite, this relative decrease, the polymerase system used had minimal effect on which *Symbiodinium* sequences were returned (top 20 most abundant; Fig 2).

Discussion

NGS approaches are quickly superseding their traditional gel-based counterparts due to advances in sequencing power and ease of handling (Arif et al. 2014; Thomas et al. 2014). In particular, increases in sequencing depths by several orders of magnitude ~~have~~has enabled the examination of low-abundance ITS2 sequences for the first time. The illumination of these rare-sequences holds the potential to characterising the rare *Symbiodinium* biosphere *in hospite* and in free-living contexts (e.g. Boulotte et al. 2016; Mordret et al. 2016; Ziegler et al. 2017b). Also, the incorporation of rarer ITS2 sequences into phylogenetic analyses has the potential to vastly improve the taxonomic resolving power of the ITS2 marker (Smith et al. 2017).

As rarer ITS2 sequence variants become more integral to the biological inferences made (Boulotte et al. 2016; Ziegler et al. 2017b), the effect of differences in ITS2 amplification strategies will become more significant. For example, small changes in amplification strategy may have a large effect on the rarer sequence variants returned. Despite this potential sensitivity to amplification variation, NGS studies published to date have used a range of different primer pairs developed for use in less-sensitive analytical contexts, without first assessing the effect of primer choice. For example, ineffective amplifications have the potential to compromise our ability to infer taxonomic identities, and inter-primer pair variation could hinder cross-study comparisons. In particular, since several large-scale projects including the Global Coral Microbiome Project and *Tara*~~*TARA*~~ Pacific, in which thousands of samples are being analysed for *Symbiodinium* ITS2 diversity, many of which represent *Symbiodinium*-rare environments, it is important to standardise and test proof the primer pair and PCR protocol employed. Here, we

compared three commonly used *Symbiodinium* ITS2 primer pairs, and show that the SYM_VAR primer pair outperforms the two alternative pairs, in specificity and sensitivity whilst maintaining an apparent minimal taxonomic bias.

Our data show that primer pair had a considerable effect on amplification efficacy of the *Symbiodinium* ITS2 gene, especially in *Symbiodinium*-rare environments. In the SRF-OO~~O~~A and SRF-CO samples, the SYM_VAR primer pair significantly outperformed the alternative primers amplifying more *Symbiodinium* reads in every sample tested (Fig 1). Only SYM_VAR were able to amplify *Symbiodinium* in the majority of SRF-OO~~O~~A samples with proportions as high as 46%. In contrast, the maximum amplification of the alternative primers was 2% with both pairs returning averages of <1% from successful amplifications in less than half the samples. In more *Symbiodinium*-abundant samples, such as the CO or CSW samples, the use of the SYM_VAR primers enabled a significant gain in the proportion of *Symbiodinium* sequences returned, once again amplifying more *Symbiodinium* in every sample tested (Fig 1). In practical terms, this translates to needing a shallower sequencing effort per sample to return the same amount of *Symbiodinium* information, thus enabling more samples to be sequenced per unit cost.

As part of optimising the SYM_VAR amplification protocol we tested whether increasing annealing temperature could increase the primers' taxonomic specificity for *Symbiodinium*. An increase in annealing temperature from 56°C to 59˚C resulted in a higher proportion of *Symbiodinium* being amplified (Fig. 3). However, this was accompanied by a minor increase in the proportion of clade D *Symbiodinium* amplified. Given that any change in ~~cladal~~clade proportion resulting from an increased annealing temperature is most likely due to preferential amplification of certain clades over others, this increase may be symptomatic of a weak

~~cladal~~clade bias. Considered alongside the fact that sufficient *Symbiodinium* sequences were amplified to enable a robust analysis in the CO, CSW, and SRF-CO samples at 56ºC, and that increasing annealing temperature did not enable the amplification of *Symbiodinium* DNA in the SRF-OO~~O~~A samples, we recommend the 56ºC annealing temperature to ~~preference~~ ~~minimising~~minimise ~~cladal~~clade bias at the expense of additional *Symbiodinium* amplification. Notably, although every effort has been made to design the SYM_VAR primer pair in a manner that minimises *Symbiodinium* ~~cladal~~clade bias, further testing for ~~cladal~~clade bias in particular when working with *Symbiodinium* clades not recovered in this study may be needed to unequivocally confirm minimal bias. To this end, amplification, sequencing and analysis of pre-determined, artificially mixed communities of *Symbiodinium* may offer an effective way to test for ~~cladal~~clade-bias.

We also tested whether a reduced number of PCR cycles could still effectively amplify *Symbiodinium* DNA and whether the cycle number would affect chimeric sequence production. Although no *Symbiodinium* amplification was achieved in the 2 SRF-OO~~O~~A samples tested, a ~~increased~~greater amplification of *Symbiodinium* at 35 rather than 25 cycles in the CSW and SRF-CO samples (Fig. 3), accompanied by only minimal ~~increase~~reduction in chimeric sequence production (Fig. 2) at the lower cycles~~2),~~ advocates the use of the higher cycles in the standardised protocol. Notably, any cycles required for the preparation of NGS libraries should be included in the number of total cycles. For example, if adapter and index attachment require a total of 7 cycles, then 28 cycles should be used for the initial PCR. Otherwise 35 cycles should be used for non-NGS applications.

Finally, we investigated whether the polymerase system used could modify the specificity of the SYM_VAR pair. Importantly, our results showed that the use of the Advantage 2 rather than the Phusion system led to a decreased specificity for *Symbiodinium* DNA when used in conjunction with the SYM_VAR pair (Fig. 4). Whilst it is not our objective to recommend the use of one polymerase system over any other, and it is likely unrealistic to expect research groups to diverge from polymerase systems established in their labs, our results do highlight that ~~the~~a polymerase system used can have an effect on specificity and should therefore be taken into consideration during PCR optimisations. Reassuringly, the profile of ITS2 sequences returned from each of the polymerase systems were comparable (Fig 2). This would suggest that although polymerase system may affect the efficiency of amplification of *Symbiodinium*, it will have no effect on inferences made from data.

Alongside considerations of *Symbiodinium* amplification efficiency (i.e., whether *Symbiodinium* can be amplified), it is also important to assess amplifications in terms of what taxonomic inferences they may offer. Due to its multicopy nature and the circumstance that more than one *Symbiodinium* taxa can be associated with any given host, the level of taxonomic resolution offered by the ITS2 genetic marker within *Symbiodinium* is dependent on the depth of sequencing. As the returned number of *Symbiodinium* sequences decreases, so does the power to resolve. When 100s to 1000s of sequences (per clade) are returned, ITS2 type profiles (sets of intragenomic sequences diagnostic of *Symbiodinium* taxa) can be assessed in detail (Smith et al. 2017). Analyses that make use of intragenomic variance in such a way are very recent or still in development. Thus, species descriptions have yet to make use of NGS-derived ITS2 type profiles, but their identification has the ability to resolve at, and beyond, the species level within

Symbiodinium **(Hume et al. 2015; LaJeunesse et al. 2014; Smith et al. 2017)**. Whilst all of the primer pairs were able to return such sequencing depth in the CO and CSW samples, for the SRF-CO and SRF-OO~~OA~~ samples this depth was only achieved by the SYM_VAR primer pair (and not consistently across all samples). With this resolution, correlations between *in hospite* and free-living *Symbiodinium* populations can be addressed with a high degree of certainty. Notably, the level of sequencing depth becomes rather an issue of sequencing effort. Hence, differences in sequencing efforts for different environments will need to be taken into account if comparisons are planned to be made between different environments at the same resolution.

Conclusions

Given the ongoing popularity and increasing interest in phylotyping *Symbiodinium* in host as well as non-host environments employing the ITS2 region, we set out to develop a standardised *Symbiodinium* ITS2 amplification protocol, in particular for the use of NGS approaches. We show that the SYM_VAR_5.8S2/SYM_VAR_REV primer pair represent a superior primer pair able to amplify a greater proportion of *Symbiodinium* DNA with minimal cladal bias in comparison to other commonly used primers. In particular, the SYM_VAR primer pair's ability to amplify from *Symbiodinium*-rare environments whilst maintaining maximum taxonomic resolving power holds potential for better elucidating the role of free-living *Symbiodinium* in reef ecology. Our standardised protocol should therefore help ~~to~~ maximize return of information from a diverse range of sample types whilst maintaining comparability to future and existing projects.

Acknowledgements

~~This project has been funded through the *Tara* consortium, France Genomique grant number ANR-10-INBS-09, and KAUST baseline and KAUST BESE division funds to CRV.~~

Tara~~*TARA*~~ Pacific consortium acknowledgement: We are keen to thank the commitment of the people and the following institutions and sponsors who made this singular expedition possible: CNRS, CSM, PSL, KAUST, Genoscope/CEA, ANR-CORALGENE, France Genomique (specifically grant number ANR-10-INBS-09), agnès b., the Veolia Environment Foundation, Region Bretagne, Serge Ferrari, Billerudkorsnas, AmerisourceBergen Company, Lorient Agglomeration, Oceans by Disney, the Prince Albert II de Monaco Foundation, L'Oréal, Biotherm, France Collectivités, Kankyo Station, Fonds Français pour l'Environnement Mondial (FFEM), Etienne Bourgois, UNESCO-IOC, the ***Tara*** Foundation teams and crew. ***Tara*** Pacific would not exist without the continuous support of the participating institutes. This work is contribution number 4 of ***Tara*** Pacific.

References

Alberti A, Poulain J, Engelen S, Labadie K, Romac S, Ferrera I, Albini G, Aury JM, Belser C, Bertrand A, Cruaud C, Da Silva C, Dossat C, Gavory F, Gas S, Guy J, Haquelle M, Jacoby E, Jaillon O, Lemainque A, Pelletier E, Samson G, Wessner M, Acinas SG, Royo-Llonch M, Cornejo-Castillo FM, Logares R, Fernandez-Gomez B, Bowler C, Cochrane G, Amid C, Ten Hoopen P, De Vargas C, Grimsley N, Desgranges E, Kandels-Lewis S, Ogata H, Poulton N, Sieracki ME, Stepanauskas R, Sullivan MB, Brum JR, Duhaime MB, Poulos BT, Hurwitz BL, Pesant S, Karsenti E, Wincker P, Team GT, and Consortium TO. 2017. Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans expedition. *Scientific Data* 4. ARTN 170093 10.1038/sdata.2017.93

Arif C, Daniels C, Bayer T, Banguera-Hinestroza E, Barbrook A, Howe CJ, Lajeunesse TC, and Voolstra CR. 2014. Assessing *Symbiodinium* diversity in scleractinian corals via next-generation sequencing-based genotyping of the ITS2 rDNA region. *Molecular Ecology* 23:4418-4433. 10.1111/mec.12869

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

Bourne DG, Morrow KM, and Webster NS. 2016. Insights into the Coral Microbiome: Underpinning the Health and Resilience of Reef Ecosystems. *Annual Review of Microbiology, Vol 70* 70:317-340. 10.1146/annurev-micro-102215-095440

Coleman AW, Suarez A, and Goff LJ. 1994. Molecular Delineation of Species and Syngens in Volvocacean Green-Algae (Chlorophyta). *Journal of Phycology* 30:80-90. DOI 10.1111/j.0022-3646.1994.00080.x

Cunning R, Gates RD, and Edmunds PJ. 2017. Using high-throughput sequencing of ITS2 to describe Symbiodinium metacommunities in St. John, US Virgin Islands. *Peerj* 5. ARTN e347210.7717/peerj.3472

Hughes TP, Barnes ML, Bellwood DR, Cinner JE, Cumming GS, Jackson JBC, Kleypas J, van de Leemput IA, Lough JM, Morrison TH, Palumbi SR, van Nes EH, and Scheffer M. 2017. Coral reefs in the Anthropocene. *Nature* 546:82-90. 10.1038/nature22901

Hume B, D'Angelo C, Burt J, Baker AC, Riegl B, and Wiedenmann J. 2013. Corals from the Persian/Arabian Gulf as models for thermotolerant reef-builders: Prevalence of clade C3 *Symbiodinium*, host fluorescence and *ex situ* temperature tolerance. *Marine Pollution Bulletin* 72:313-322. 10.1016/j.marpolbul.2012.11.032

Hume BCC. 2013. Are specific coral-*Symbiodinium* partnerships associated with survival in extreme temperature environments of the Persian/Arabian Gulf? PhD thesis. University of Southampton.

Hume BCC, D'Angelo C, Smith EG, Stevens JR, Burt J, and Wiedenmann J. 2015. *Symbiodinium thermophilum* sp nov., a thermotolerant symbiotic alga prevalent in corals

of the world's hottest sea, the Persian/Arabian Gulf. *Scientific Reports* 5. 10.1038/srep08562

Hume BCC, Voolstra CR, Arif C, D'Angelo C, Burt JA, Eyal G, Loya Y, and Wiedenmann J. 2016. Ancestral genetic diversity associated with the rapid spread of stress-tolerant coral symbionts in response to Holocene climate change. *Proceedings of the National Academy of Sciences of the United States of America* 113:4416-4421. 10.1073/pnas.1601910113

Kemp DW, Hernandez-Pech X, Iglesias-Prieto R, Fitt WK, and Schmidt GW. 2014. Community dynamics and physiology of Symbiodinium spp. before, during, and after a coral bleaching event. *Limnology and Oceanography* 59:788-797. 10.4319/lo.2014.59.3.0788

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. DOI 10.1046/j.1529-8817.2001.01031.x

LaJeunesse TC. 2002. Diversity and community structure of symbiotic dinoflagellates from Caribbean coral reefs. *Marine Biology* 141:387-400. 10.1007/s00227-002-0829-2

LaJeunesse TC, Wham DC, Pettay DT, Parkinson JE, Keshavmurthy S, and Chen CA. 2014. Ecologically differentiated stress-tolerant endosymbionts in the dinoflagellate genus *Symbiodinium* (Dinophyceae) Clade D are different species. *Phycologia* 53:305-319. 10.2216/13-186.1

Levin RA, Voolstra CR, Weynberg KD, and van Oppen MJH. 2017. Evidence for a role of viruses in the thermal sensitivity of coral photosymbionts. *Isme Journal* 11:808-812. 10.1038/ismej.2016.154

Moberg F, and Folke C. 1999. Ecological goods and services of coral reef ecosystems. *Ecological Economics* 29:215-233. Doi 10.1016/S0921-8009(99)00009-9

Mordret S, Romac S, Henry N, Colin S, Carmichael M, Berney C, Audic S, Richter DJ, Pochon X, de Vargas C, and Decelle J. 2016. The symbiotic life of Symbiodinium in the open ocean within a new species of calcifying ciliate (Tiarina sp.). *Isme Journal* 10:1424-1436. 10.1038/ismej.2015.211

Muscatine L. 1990. The role of symbiotic algae in carbon and energy flux in reef corals. In: Dubinsky Z, ed. *Ecosystems of the world 25: Coral Reefs*. Amsterdam: Elsevier.

Nitschke MR, Davy SK, and Ward S. 2016. Horizontal transmission of Symbiodinium cells between adult and juvenile corals is aided by benthic sediment. *Coral Reefs* 35:335-344. 10.1007/s00338-015-1349-0

Norstrom AV, Nystrom M, Jouffray JB, Folke C, Graham NAJ, Moberg F, Olsson P, and Williams GJ. 2016. Guiding coral reef futures in the Anthropocene. *Frontiers in Ecology and the Environment* 14:490-498. 10.1002/fee.1427

Pandolfi JM, Connolly SR, Marshall DJ, and Cohen AL. 2011. Projecting Coral Reef Futures Under Global Warming and Ocean Acidification. *Science* 333:418-422. 10.1126/science.1204794

576Peixoto RS, Rosado PM, Leite DCD, Rosado AS, and Bourne DG. 2017. Beneficial
577 Microorganisms for Corals (BMC): Proposed Mechanisms for Coral Health and
578 Resilience. *Frontiers in Microbiology* 8. ARTN 34110.3389/fmicb.2017.00341

579Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, Iudicone D, Karsenti E,
580 Speich S, Trouble R, Dimier C, Searson S, and Consortium TO. 2015. Open science
581 resources for the discovery and analysis of Tara Oceans data. *Scientific Data* 2. ARTN
582 15002310.1038/sdata.2015.23

583Plaisance L, Caley MJ, Brainard RE, and Knowlton N. 2011. The Diversity of Coral Reefs: What
584 Are We Missing? *Plos One* 6. 10.1371/journal.pone.0025026

585Pochon X, and Gates RD. 2010. A new Symbiodinium clade (Dinophyceae) from soritid
586 foraminifera in Hawai'i. *Molecular Phylogenetics and Evolution* 56:492-497.
587 10.1016/j.ympev.2010.03.040

588Pochon X, Pawlowski J, Zaninetti L, and Rowan R. 2001. High genetic diversity and relative
589 specificity among Symbiodinium-like endosymbiotic dinoflagellates in soritid
590 foraminiferans. *Marine Biology* 139:1069-1078.

591Pochon X, Putnam HM, and Gates RD. 2014. Multi-gene analysis of Symbiodinium
592 dinoflagellates: a perspective on rarity, symbiosis, and evolution. *Peerj* 2. ARTN
593 e39410.7717/peerj.394

594Pochon X, Stat M, Takabayashi M, Chasqui L, Chauka LJ, Logan DDK, and Gates RD. 2010.
595 Comparison of Endosymbiotic and Free-Living Symbiodinium (Dinophyceae) Diversity
596 in a Hawaiian Reef Environment. *Journal of Phycology* 46:53-65. 10.1111/j.1529-
597 8817.2009.00797.x

598Pogoreutz C, Radecker N, Cardenas A, Gardes A, Voolstra CR, and Wild C. 2017. Sugar
599 enrichment provides evidence for a role of nitrogen fixation in coral bleaching. *Global*
600 *Change Biology* 23:3838-3848. 10.1111/gcb.13695

601Radecker N, Pogoreutz C, Voolstra CR, Wiedenmann J, and Wild C. 2015. Nitrogen cycling in
602 corals: the key to understanding holobiont functioning? *Trends in Microbiology* 23:490-
603 497. 10.1016/j.tim.2015.03.008

604Rowan R, Knowlton N, Baker A, and Jara J. 1997. Landscape ecology of algal symbionts creates
605 variation in episodes of coral bleaching. *Nature* 388:265-269. Doi 10.1038/40843

606Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley
607 BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, and Weber
608 CF. 2009. Introducing mothur: Open-source, platform-independent, community-supported
609 software for describing and comparing microbial communities. *Applied and*
610 *Environmental Microbiology* 75:7537-7541. 10.1128/Aem.01541-09

611Silverstein RN, Cunning R, and Baker AC. 2015. Change in algal symbiont communities after
612 bleaching, not prior heat exposure, increases heat tolerance of reef corals. *Global Change*
613 *Biology* 21:236-249. 10.1111/gcb.12706

614Smith EG, Ketchum RN, and Burt JA. 2017. Host specificity of Symbiodinium variants revealed
615 by an ITS2 metahaplotype approach. *ISME J* 11:1500-1503. 10.1038/ismej.2016.206

616Stat M, Pochon X, Cowie ROM, and Gates RD. 2009. Specificity in communities of
617 Symbiodinium in corals from Johnston Atoll. *Marine Ecology Progress Series* 386:83-96.
618 10.3354/meps08080

619Thomas L, Kendrick GA, Kennington WJ, Richards ZT, and Stat M. 2014. Exploring
620 Symbiodinium diversity and host specificity in Acropora corals from geographical
621 extremes of Western Australia with 454 amplicon pyrosequencing. *Molecular Ecology*
622 23:3113-3126. 10.1111/mec.12801

623Thornhill DJ, Howells EJ, Wham DC, Steury TD, and Santos SR. 2017. Population genetics of
624 reef coral endosymbionts (Symbiodinium, Dinophyceae). *Molecular Ecology* 26:2640-
625 2659. 10.1111/mec.14055

626Torda G, Donelson JM, Aranda M, Barshis DJ, Bay L, Berumen ML, Bourne DG, Cantin N,
627 Foret S, Matz M, Miller DJ, Moya A, Putnam HM, Ravasi T, van Oppen MJH, Thurber
628 RV, Vidal-Dupiol J, Voolstra CR, Watson SA, Whitelaw E, Willis BL, and Munday PL.
629 2017. Rapid adaptive responses to climate change in corals. *Nature Climate Change*
630 7:627-636. 10.1038/Nclimate3374

631Wham DC, Ning G, and LaJeunesse TC. 2017. Symbiodinium glynnii sp nov., a species of
632 stress-tolerant symbiotic dinoflagellates from pocilloporid and montiporid corals in the
633 Pacific Ocean. *Phycologia* 56:396-409. 10.2216/16-86.1

634Ziegler M, Arif C, Burt JA, Dobretsov S, Roder C, LaJeunesse TC, and Voolstra CR. 2017a.
635 Biogeography and molecular diversity of coral symbionts in the genus Symbiodinium
636 around the Arabian Peninsula. *Journal of Biogeography* 44:674-686. 10.1111/jbi.12913

637Ziegler M, Eguíluz VM, Duarte CM, and Voolstra CR. 2017b. Rare symbionts may contribute to
638 the resilience of coral–algal assemblages. *The Isme Journal*. 10.1038/ismej.2017.151

639Ziegler M, Seneca FO, Yum LK, Palumbi SR, and Voolstra CR. 2017c. Bacterial community
640 dynamics are linked to patterns of coral heat tolerance. *Nature Communications* 8. ARTN
641 1421310.1038/ncomms14213

642

~~Conflict of interest~~

~~The authors declare no conflict of interest.~~

Data accessibility

Sequence data can be accessed through ~~B~~bioproject ~~A~~accession PRJNA430028 at NCBI (~~https://www.ncbi.nlm.nih.gov/~~https://www.ncbi.nlm.nih.gov/bioproject/PRJNA430028).

~~Author Contributions~~

~~Conceived and designed the experiments: BCCH, MZ, JP, PW, CRV; generated data, provided tools/reagents/samples: BCCH, MZ, JP, XP, SR, EB, CdV, SP, PW, CRV; analysed data: BCCH, MZ, JP, CRV; wrote manuscript: BCCH, MZ, CRV; revised manuscript: BCCH, MZ, JP, XP, PW, CRV~~

Sampling permi~~ts~~

Panama scientific sampling: SE/AP-18-16; Panama CITES: SEX/A-72-16, no. 05567; Portulgal scientific sampling: 080522/2009/Proc.º E.17.b; Maldives scientific sampling: FA-D2/33/2010/02; Chile scientific sampling: 13270/24/74

Table headers

Table 1 Primer pairs tested.

Table 2 Sample overview and PCR conditions. Samples analysed for primer pair comparisons and optimisations are denoted.

Table 3 Effect of number of PCR cycles on chimeric amplicon formation.

Figure legends

Figure 1 Comparison and taxonomic composition of sequences returned using three *Symbiodinium* ITS2 primer pairs. Each plot represents one sample and contains three pairs of stacked columns, one for each primer pair in the order of: SYM_VAR, ITSintfor, ITS-DINO (from left to right). For each pair of stacked columns, the left coloured bar denotes the taxonomic breakdown of all returned sequences into 7 categories, and the right greyscale bar denotes the sub-genus (cladal) distribution for the subset of sequences classified as *Symbiodinium* (see colour legend). Plots are annotated with the number of sequences returned after QC and the proportion of those sequences that were annotated as *Symbiodinium* underneath the coloured and grayscale bars, respectively. For each environment type (indicated to the left of the plot groupings) a summary plot is given at the right side, showing the average taxonomic breakdowns, total post-QC sequences returned, and average proportion of *Symbiodinium* sequences.

Figure 2 Robustness of PCR protocol using the SYM_VAR ITS2 primer pair. Each subplot represents a pairwise comparison of PCRs from the same sample. The sample, environment type, and PCR conditions are noted above each subplot. Depending on which pair of PCRs are being compared, these subplots graphically represent effects due to (A~~1~~) PCR-to-PCR variation, (B~~2~~) annealing temperature, (C~~3~~) number of PCR cycles, or (D~~4~~) polymerase system used. For each subplot, the relative presence of the 20 most abundant amplicons (calculated across both PCRs) is plotted for each PCR, denoted as black and grey bars. For example, ~~the top left subplot of the figure~~Aii represents a pairwise comparison of PCRs run using the same annealing temp and number of cycles on sample AW-0000035. It is therefore testing the PCR-to-PCR effect.

Figure 3 Effect of annealing temperature and number of PCR cycles on sequences returned using the SYM_VAR ITS2 primer pair. Two annealing temperatures (a; 56 and 59°C; columns) and two cycle numbers (b; 25 and 35 cycles; columns) were tested on samples from three different environment types (rows; indicated to the left of the plot groupings). Each plot represents a different pooled PCR amplicon with sample name, sample type, number of cycles, and annealing temperature noted to the left. For each plot, coloured bars on the left denote the taxonomic breakdown of all returned sequences into 7 categories. The right greyscale bars denote the sub-genus (cladal) distribution for the subset of sequences classified as *Symbiodinium*. Plots are annotated with the number of sequences returned after QC and the proportion of those sequences that were annotated as *Symbiodinium* underneath the coloured and grayscale bars, respectively. *Symbiodinium*-derived sequences were further subcategorized according to their cladal identity (greyscale stacked bars).

Figure 4 Effect of polymerase type on sequences returned using the SYM_VAR ITS2 primer pair. Each plot represents one sample and contains two pairs of stacked columns, one for each polymerase with Phusion (ThermoFisher Scientific, USA) on the left, and Advantage 2 (Takara bio, Japan) on the right. Coloured bars denote the taxonomic breakdown of all returned sequences into 7 categories. The greyscale bars denote the sub-genus (cladal) distribution for the subset of sequences classified as *Symbiodinium*. Plots are annotated with the number of sequences returned after QC and the proportion of those sequences that were annotated as *Symbiodinium* underneath the coloured and grayscale bars, respectively. Plots are ordered by

environment (rows; CO - coral, CSW - coral surrounding water, SRF-CO - surface water, SRF-OOOA - open ocean). A summary plot is provided for each environment type at the end of the row.

Supplemental Information

Supplementary Table 1 Sample details and PCR conditions. Samples analysed for primer pair comparisons and optimisations are denoted.

Supplementary Figure 1. **Mean sequences returned per sample and proportion retained during the quality control pipeline (QC) in relation to primer pair used.** The number of sequences retained at three stages of the QC pipeline are given: 'raw_contigs' (number of sequences returned from mothur's *make.contigs* command), 'post-qc' (the number of sequences remaining after all quality control and taxonomic identification), and '*Symbiodinium*' (the number of sequences identified as *Symbiodinium*). Samples are grouped according to their environment and amplification primers. Abbreviations as follows: CO – coral tissue, CSW – coral surrounding water, SRF-CO – coral reef surface water, SRF-OO – open ocean surface water.

Supplementary Figure 2. Comparison and taxonomic composition of sequences returned using three *Symbiodinium* ITS2 primer pairs before removal of sequences JN406302 and JN406301. Each plot represents one sample and contains three pairs of stacked columns, one for each primer pair in the order of: SYM_VAR, ITSintfor, ITS-DINO (from left to right). For each pair of stacked columns, the left coloured bar denotes the taxonomic breakdown of all returned sequences into 7 categories, and the right greyscale bar denote the sub-genus (cladal) distribution for the subset of sequences classified as *Symbiodinium*. Plots are annotated with the number of sequences returned after QC and the proportion of those sequences that were annotated as Symbiodinium underneath the coloured and grayscale bars, respectively. For each environment type (indicated to the left of the plot groupings) a summary plot is given at the right side, showing the average taxonomic breakdowns, total post-QC sequences returned, and average proportion of *Symbiodinium* sequences.

Supplementary Data 1 – The fasta sequences used to generate the *Symbiodinium* cladal database.