MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 6: 89–99
DOI 10.3897/mbmg.6.78762

Data Paper

18S rDNA amplicon sequence data (V1–V3) of the Palmyra Atoll National Wildlife Refuge, Central Pacific

Brendan N. Reid[1*], **Jennifer A. Servis**[2,3*], **Molly Timmers**[4,5], **Forest Rohwer**[6], **Eugenia Naro-Maciel**[7]

1 *Department of Ecology, Evolution, and Natural Resources, Rutgers University, 14 College Farm Road, New Brunswick, New Jersey 08901, USA*
2 *Biology Dept., College of Staten Island, City University of New York, 2800 Victory Boulevard, Staten Island, New York 10314, USA*
3 *US Fish and Wildlife Service, 5275 Leesburg Pike, Falls Church, Virginia, 22041-3803, USA*
4 *Hawai'i Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Honolulu, Hawaii, 96744, USA*
5 *Pristine Seas, National Geographic Society, Washington, DC 20036, USA*
6 *Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego, California, 92182, USA*
7 *Liberal Studies, New York University, 726 Broadway, New York, New York, 10003, USA*

Corresponding author: Eugenia Naro-Maciel (enmaciel@nyu.edu)

Academic editor: Anastasija Zaiko | **Received** 2 December 2021 | **Accepted** 23 March 2022 | **Published** 12 April 2022

Abstract

To address the global biodiversity crisis, standardized data that are rapidly obtainable through minimally invasive means are needed for documenting change and informing conservation within threatened and diverse systems, such as coral reefs. In this data paper, we describe 18S rRNA gene amplicon data (V1–V3 region) generated from samples collected to begin characterizing coral reef eukaryotic community composition at the Palmyra Atoll National Wildlife Refuge in the Central Pacific Ocean. Sixteen samples were obtained across four sample types: sediments from two sieved fractions (100–500 µm, n = 3; 500 µm-2 mm, n = 3) and sessile material scrapings (n = 3) from Autonomous Reef Monitoring Structures (ARMS) sampled in 2015, as well as seawater from 2012 (n = 7). After filtering and contaminant removal, 3,861 Amplicon Sequence Variants (ASVs) were produced from 1,062,238 reads. The rarefaction curves demonstrated adequate sampling depth, and communities grouped by sample type. The dominant orders across samples were polychaete worms (Eunicida), demosponges (Poecilosclerida), and bryozoans (Cheilostomatida). The ten most common orders in terms of relative abundance comprised ~60% of all sequences and 23% of ASVs, and included reef-building crustose coralline algae (CCA; Corallinophycidae) and stony corals (Scleractinia), two taxa associated with healthy reefs. Highlighting the need for further study, ~21% of the ASVs were identified as uncultured, *incertae sedis*, or not assigned to phylum or order. This data paper presents the first 18S rDNA survey at Palmyra Atoll and serves as a baseline for biodiversity assessment, monitoring, and conservation of this remote and pristine ecosystem.

Key Words

ARMS, eDNA, next-generation sequencing, PANWR, QIIME2, seawater, sediment, sessile

Introduction

Tropical coral reefs are among the most biologically diverse, complex, and productive of ocean ecosystems (Knowlton et al. 2010). However, they are also highly endangered habitats, with one-third of reef-building corals under extinction threat (Carpenter et al. 2008; Knowlton et al. 2010). Reef systems are declining due to pollution, overexploitation, global climate change, bleaching, ocean acidification, changing storms, and other factors

* These authors contributed equally to this work.

(Carpenter et al. 2008; Knowlton and Jackson 2008; Sandin et al. 2008). Yet, basic understanding of taxa being affected is limited due to a combination of collection biases and challenges, lack of taxonomic expertise, and labor-intensive morphological evaluations, leading to uncertain species identifications and biodiversity estimates (Plaisance et al. 2011; Leray and Knowlton 2016).

To address this gap, baseline biodiversity assessment using standardized methods to enable comparison among sites represents a fundamental first step (Sandin et al. 2008; Leray and Knowlton 2016; Lafferty et al. 2021). While surface-dwelling taxa such as fish and corals can be visually and morphologically assessed, microscopic organisms and the small and cryptic groups living within the reef framework are generally more readily surveyed using next-generation metabarcoding technologies (Knowlton et al. 2010; Plaisance et al. 2011; Leray and Knowlton 2016; Servis et al. 2020; Deiner et al. 2021). High-throughput community-wide metagenomic approaches provide orders of magnitude more data than morphological methods or traditional DNA barcoding of individuals, identifying multiple organisms simultaneously from living biomass or environmental DNA (eDNA) in water or sediment (Taberlet et al. 2018; Deiner et al. 2021). Although there are limitations due to incomplete reference databases and other factors, this analysis is minimally destructive, less reliant on taxonomic expertise, increasingly cost-effective, and capable of identifying elusive, rare, endangered, and invasive taxa. Furthermore, it is especially useful for sampling remote areas that are difficult or costly to access (Kelly et al. 2017), like the Palmyra Atoll National Wildlife Refuge (PANWR) (Fig. 1).

Targeted metabarcoding research at Palmyra has established baselines and patterns for fish (Lafferty et al. 2021), prokaryotes, and viruses (Dinsdale et al. 2008; Smriga et al. 2010; McDole et al. 2012). DNA barcoding studies there have uncovered crustacean and brachyuran richness from *Pocillopora* coral heads (Plaisance et al. 2009; Knowlton et al. 2010) and sampling units known as ARMS, or Autonomous Reef Monitoring Structures (Leray and Knowlton 2015; Ransome et al. 2017; Pearman et al. 2018; Nichols et al. 2021). ARMS are composed of PVC plates stacked in a tier of alternating open and semi-closed layers and affixed to the reef benthos, where they attract organisms by mimicking structural complexity (Ransome et al. 2017). ARMS capture poorly understood cryptofauna (e.g., bryozoans, polychaetes, sponges, tunicates) and sediments containing microscopic taxa, secretions, excretions, shed cells, and other sources of eDNA. As such, ARMS constitute a standardized, replicable method for assessing both the metazoan diversity of coral reefs and their broader associated eukaryotic communities, and are increasingly deployed globally (Leray and Knowlton 2015; Ransome et al. 2017; Pearman et al. 2018; Nichols et al. 2021).

The data set presented here constitutes the first broad eukaryotic biodiversity survey of coral reef systems at the Palmyra Atoll National Wildlife Refuge. Both seawater and ARMS samples were examined using 18S rDNA, a promising genetic marker offering advantages of broad amplification across eukaryotic kingdoms, a rapidly growing reference database, and wide use (Taberlet et al. 2018). In this work the V1–V3 segment, with demonstrated utility for uncovering a broad range of aquatic taxa (Ingala et al. 2021), was sequenced. Our analysis of preliminary data showed that V1–V3 captured groups of interest, including corals and algae, while also contributing information about a less frequently sequenced 18S gene region. The versatility of the 18S PCR primers represents a convenient tool to screen the entire eukaryotic domain, providing a valuable baseline assessment of eukaryotic life in this relatively undisturbed coral reef ecosystem, and a significant contribution to the global effort to characterize reef biodiversity.

Methods

Study site

Palmyra Atoll is located in the Central Pacific Ocean approximately 1,700 km southwest of Hawai‘i, lying on the northwest end of the Northern Line Islands (Fig. 1; 5°53'N; 162°5'W). The PANWR, currently uninhabited except for research and management staff, includes about 2.5 km^2 of land area and 155 km^2 of reefs, and comprises small islands and islets surrounding western, central, and eastern lagoons (Collen et al. 2009). While U.S. military activities in the 1940s increased the area and volume of the islands, resulting in sedimentation and decreased water level and flow in the lagoons, the surrounding shallow reef flats and submerged reef terraces to the east and west remained relatively undisturbed (Collen et al. 2009; Sterling et al. 2013).

ARMS collection and DNA extraction

Three sites around the atoll were selected for ARMS deployment (Fig. 1). In May 2012 NOAA researchers placed one ARMS unit on the forereef at each of these sites, at depths of 10.7–18.3 m. The monitoring structures were secured to the hard bottom reef habitat using stainless steel stakes and heavy-duty zip ties, where they remained for three years. Upon recovery in 2015, each ARMS was encapsulated with a mesh-lined crate, detached from the reef, brought to the surface, transferred to a disassembly tub, and transported to the NOAA ship where it was taken apart plate by plate.

Once disassembled, the seawater within the tub was sieved through adjoining 2 mm and 500 µm pans and an attachable 100 µm mesh net, creating two separate sediment size fractions (100–500 µm and 500 µm-2 mm). These fractions were stored in 95% ethanol within a -20 °C freezer. Back at a NOAA lab onshore, each sediment fraction underwent a decantation process to isolate organic material from sediment particles following Leray and Knowlton (2015). Resulting organic material was stored in 95% ethanol until being sent out for DNA extraction.

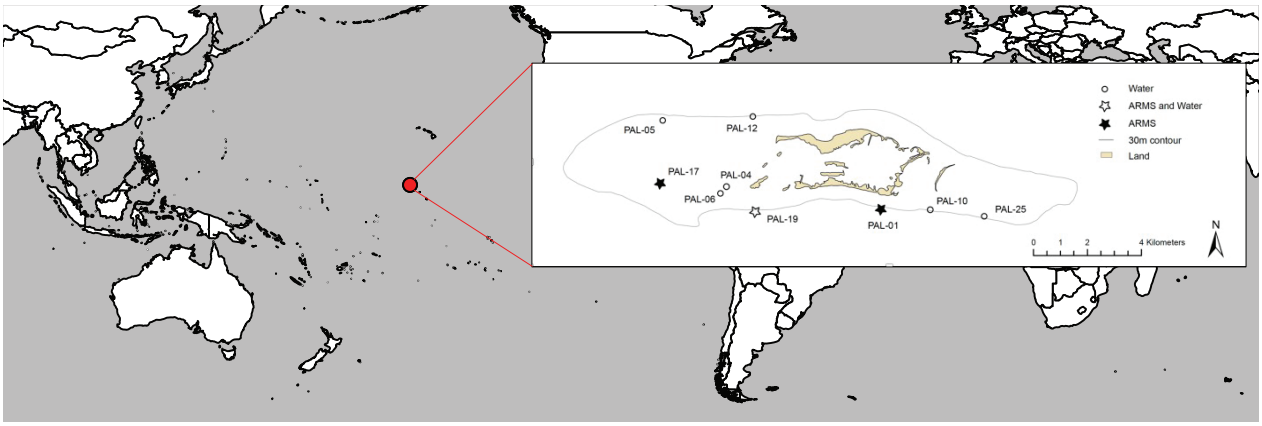

Figure 1. The Palmyra Atoll National Wildlife Refuge in the Central Pacific. The inset shows sampling sites at the PANWR. Autonomous Reef Monitoring Structure (ARMS) and water samples were collected from nine coral reef sites at the Palmyra Atoll National Wildlife Refuge, Central Pacific. ARMS samples were obtained at sites marked by stars, water at sites marked by circles, and at site PAL-19 both ARMS and water samples were collected.

Plates were scoured using a paint scraper to remove the sessile material. Scrapings were transferred into a blender and homogenized for one minute. The blended homogenate was poured into a 40 µm hand net and drained with the aid of filtered seawater. Approximately 10 g subsamples were placed into 50 ml falcon tubes containing 95% ethanol and stored at -20 °C until being shipped for DNA extraction.

At the Sackler Institute for Comparative Genomics, four replicate extractions (~ 0.25 g each) were performed using the MOBIO PowerSoil DNA Isolation Kit on each of the ARMS sediment and sessile samples. Although extraction blanks and positive controls were not used, standard decontamination, disinfection, and sterilization practices stringently in place at this dedicated molecular laboratory, including use of PCR-free areas, were assiduously followed, and *in silico* contaminant removal was carried out as described below. The DNA eluates from the four replicates were pooled (combined and mixed into one tube) and stored at -20 °C prior to being sent out for sequencing.

Water collection and DNA extraction

Water samples (n = 7) were collected in May 2012 from seven forereef sites < 1 m above the benthos at depths ranging from 3.7–13.7 m (Fig. 1). Following established collection and processing protocols, at each site, 5 L of reef water were obtained and passed through a 20 µm Nitex, followed by a 0.22 µm Sterivex filter (2 filters, ~ 2.5 L each) (Haas et al. 2014). Once filtered, samples were stored at -20 °C and shipped to a dedicated eDNA lab at San Diego State University for DNA isolation following Haas et al. (2014).

DNA sequencing

Sixteen samples, 7 from water and 9 from the three ARMS (two sediment fractions and one sessile material scraping sample per unit) (Fig. 1, Table 1), were quantified using a Qubit Fluorometer according to manufacturer instructions (Thermo Scientific, USA) and then sent to MRDNA (www.mrdnalab.com, Molecular Research LP, Shallowater, TX, USA) for purification, PCR amplification, sequencing, and analysis using standard and previously described procedures (Dowd et al. 2008; Ingala et al. 2021). First, preliminary runs were conducted and examined, and once deemed successful for satisfactory amplification of organisms of interest (e.g. plankton, corals, and algae), all of the remaining samples were submitted for processing. The DNA underwent PCR amplification targeting a 272-bp segment from the V1–V3 region of the 18S rRNA gene using the primers Euk7F (Medlin et al. 1988) tagged with a unique 8-bp identifier barcode, and Euk570R (Weekers et al. 1994). Each of the ARMS samples had a replicate sequenced, for a total of 25 PCR reactions. These reactions were carried out in 20 µl total volume containing 10 µl of the QIAGEN HotStarTaq Plus Master Mix Kit (Qiagen, USA), 9 µl of water, and 1 µl of template (Qubit 1/200 concentration range: 5.1 – 76.8 ng/mL). The PCR program consisted of an initial denaturation of 3 min at 94 °C followed by 28 cycles of 30 s at 94 °C, 40 s at 53 °C, and 60 s at 72 °C, and a final extension of 5 min at 72 °C. PCR product quality was assessed by gel electrophoresis in 2% agarose gels, and amplification success and proportions for sample pooling were informed by the band length and intensity. The uniquely tagged PCR products were pooled together in equal proportions based on DNA concentrations and electrophoresis band characteristics, and purified using Ampure XP beads (Agencourt Bioscience, USA) with a ratio of beads to PCR products of 0.75× following Illumina guidelines. They were then ligated to Illumina adapters using the Illumina TruSeq protocol, and sequenced on an Illumina MiSeq platform using a paired-end 2 × 300bp procedure (Dowd et al. 2008; Naro-Maciel et al. 2020; Ingala et al. 2021).

Demultiplexing, sequence filtering, and quality control

Forward and reverse reads extracted from MRDNA's Fastq Processor (MRDNA 2021) were imported into

Table 1. Amplicon Sequence Variant (ASV) diversity measurements for each sample type. Included are the total number of samples, sequences retained after filtering, and the total and mean numbers of ASVs detected per sample. Number of (unique) ASVs, Shannon-Weaver and Faith's phylogenetic diversity for each sample type, before and after rarefaction to the lowest combined number of reads per sample type (152,590), are given to the left and to the right of slashes respectively.

	Sediment (100- 500 µm)	Sediment (500 µm-2 mm)	Sessile material scrapings	Water	Total
Number of samples	3	3	3	7	16
Sample sites	ARMS: PAL 01, 17, and 19	ARMS: PAL 01, 17, and 19	ARMS: PAL 01, 17, and 19	PAL 04, 05, 06, 10, 12, 19, and 25	
Sampling year	2015	2015	2015	2012	
Number of replicates	6 (2 per site)	6 (2 per site)	6 (2 per site)	7	25
Number of sequences after filtering	164,799	232,724	152,590	512,125	1,062,328
Total no. ASVs	815 / 808	420 / 416	631 / 625	2,646 / 2,474	
Unique ASVs	445 / 442	143 / 140	338 / 336	2,487 / 2,322	
H (Shannon-Weaver) before / after rarefaction	3.98 / 3.97	2.63 / 2.63	3.79 / 3.79	5.47 / 5.46	
Faith's phylogenetic diversity before / after rarefaction	84.98 / 84.01	53.31 / 52.68	59.82 / 58.84	238.10 / 215.42	

the QIIME2 v. 2019.1 pipeline (Bolyen et al. 2019), demultiplexed as paired-end sequences, and examined for overall read quality. While the quality of forward reads was generally high, reverse reads varied across runs, and as a result, only forward reads were used. The DADA2 plugin in QIIME2 (Callahan et al. 2016) was employed in single-read mode to denoise and remove chimeric sequences (including discarding singleton sequences) and to identify amplicon sequence variants (ASVs), using the default value (2) for the maximum number of expected errors, a 230-bp read length, and trimming the first 30 bases. Default values were also used for removing chimeras and sequences of poor quality and for all other parameters (QIIME2 2021). All scripts used in this paper are publicly available at https://github.com/nerdbrained/palmyra_edna.

Taxonomy assignment

Amplicon-specific reference databases were used to obtain more robust taxonomic classifications (Bokulich et al. 2018). A primary database was created with the RESCRIPt QIIME plugin (Robeson et al. 2021) using sequences from the curated SILVA 138 rRNA reference database (Quast et al. 2013). Reference sequences were removed if 5 or more ambiguous bases or homopolymer runs greater than 8-bp in length were detected. Remaining sequences were then filtered to include only eukaryotes. From this pool, amplicon-specific reads were extracted from the filtered SILVA 138 database using the forward and reverse sequencing primers, and any replicate segments present after filtering were removed. Finally, a naïve Bayes classifier was fit to the reference sequence database. Sequences were aligned and a phylogenetic tree was created using the denoised forward sequences with the align-to-tree-mafft-fasttree command. The Bayes classifier was used to assign a taxonomic identification to each ASV in the tree using the classify-sklearn command in QIIME2.

To improve taxonomic assignments, two alternate assignment methods were also used. For the first method, the steps above were repeated employing an alternate curated 18S database with a focus on planktonic sequences (pr2 v.4.14.0) (Guillou et al. 2013, del Campo et al. 2018). For the second method, a sequence search was conducted against the NCBI database (downloaded 1/27/22) using the blastn algorithm with default parameters in BLAST+ v.2.11.0. (Camacho et al. 2009). BLAST hits were then used to assign sequences to taxa using the LCA-assignment algorithm in MEGAN Community Edition v.6.2.17 (Huson et al. 2016). Assignments from these methods were compiled for all sequences that did not receive a designation at the order level using the SILVA database. If either alternate method provided an assignment at the order level, that assignment was substituted in for downstream analyses. If the pr2 and BLAST identifications conflicted, the assignment from the pr2 database was preferentially retained as this database is actively curated by taxonomic experts.

Contaminant identification and removal

The QIIME2 outputs were imported into R v.4.0.0 (R Core Team 2020) using the R package PHYLOSEQ (McMurdie and Holmes 2013). All R code employed in this work can be found at https://github.com/nerdbrained/palmyra_edna. Potential contaminants were identified and subsequently removed from the dataset *in silico* using the R package DECONTAM (Davis et al. 2018). This option was selected because lab procedures were conducted prior to the widespread use of extraction blanks and sequencing of PCR negative controls in metabarcoding studies. The program's frequency method flagged taxa whose proportions were inversely correlated with sample DNA concentration across all samples. A conservative threshold value of $p < 0.10$ was employed to identify potential contaminants, and any taxa meeting this criterion were removed from the dataset. After contaminant removal, the

ARMS duplicate samples were merged, as these were not true biological replicates (Table 1).

Community composition

To assess if the number of sequences was appropriate for accurately estimating community composition and taxonomic diversity, rarefaction curves for each sample were produced using the R package VEGAN v. 2.5.5 (McMurdie and Holmes 2013; Oksanen 2019). Variation in community composition among sample types and collection sites was analyzed using a Jaccard distance matrix based on presence/absence within a permutational analysis of variance (PERMANOVA) with 999 permutations. Presence/absence rather than abundance-based analyses are presented here, as sequence and naturally occurring organismal frequencies do not necessarily match. However, abundance-based analyses using Bray-Curtis distance gave similar results (data not shown). Before calculating Jaccard distance, sequence abundance data were converted to presence/absence information using the function vegan_stand in the R package QSRUTILS (Zhang et al. 2017). Homogeneity of dispersion among groups was tested using the betadisper function in VEGAN. Community composition was primarily visualized using principal coordinate analysis (PCoA). The number and percent of taxa shared and unique to each sample type were visualized using quasi-proportional Venn diagrams of ASVs in the R package NVENNR (Quesada 2020).

Taxonomic and phylogenetic diversity

Sequences were grouped by sample type, the total number of ASVs was calculated, and the number of taxa was estimated at the second- and fourth-highest taxonomic levels present in the SILVA138 database. These levels are roughly consistent with conventional "phylum" and "order" classifications, respectively, with the caveat that taxonomic ranks in SILVA are assigned to preserve roughly the same level of evolutionary divergence at a given rank across the tree (Yilmaz et al. 2014). As such, some groupings reported here are clades encompassing several related orders. The proportion of sequences from each sample type belonging to each identified order was calculated, and these proportions were summed across sample types to identify the ten most common orders in the dataset.

The Shannon-Weaver diversity index and ACE richness estimate were calculated using VEGAN for individual samples, pooled sample types, and the collective dataset. Faith's phylogenetic diversity (defined as the sum of all branch lengths in the phylogenetic tree for all samples in a given group) was calculated at the sample type levels and for the entire dataset with the R package PICANTE (Kembel et al. 2010). A Kruskal-Wallis test was used to determine whether differences in diversity among sample types were larger than expected under a null hypothesis of no difference. To explore the effect of varying read number on diversity metrics, all statistics were then re-calculated after rarefying with replacement to the lowest read depth for the site- and sample type-level analyses. To determine how inclusion of unicellular eukaryotes influenced each metric, diversity metrics were also calculated for each sample without rarefaction using only ASVs assigned to metazoan phyla. Visual representations of the abundance of the top 10 orders and the top 25 orders were created using a bar plot and a heat map, respectively, generated in PHYLOSEQ.

Results and discussion

This project generated promising baseline data for characterizing overall eukaryotic diversity of the remote and relatively pristine PANWR reef community, and for contributing to the ongoing global assessment of reef biodiversity. A total of 1,610,301 raw sequences were generated, with on average 100,000 sequences per sample (range: 54,090–190,997). After denoising and sequence filtration, 1,113,657 (70.5%) sequences remained, resulting in 3,936 ASVs. As noted above, 75 ASVs (51,419 sequences) were removed as putative contaminants, leaving 1,062,238 sequences and 3,861 ASVs for analysis (Suppl. material 3: Table S1).

Rarefaction at the sample level (n = 16) indicated that sequencing depth was sufficient for characterizing biodiversity, as ASV accumulation curves reached an asymptote in each case (Fig. 2). In all instances, rarefied diversity statistics were highly similar to measures calculated without rarefaction, indicating that variation in read number among samples and sample types did not account for substantive differences in diversity (Table 1; Suppl. material 4: Table S2). Dispersion did not differ significantly among sample types ($p > 0.05$). Community composition varied among sample types (PERMANOVA; $p = 0.001$) but not among sites (PERMANOVA; $p > 0.1$). PCoA ordination showed clustering of ARMS sample types, with the two sediment fractions showing the greatest similarity to one another (Fig. 3). Only 0.1% of the ASVs were found within all sample types, and 88.4% were unique to a single sample type, with roughly 72.9% of the unique ASVs found within the water samples (Suppl. material 1: Fig. S1).

Taxonomic composition consisted of 73 different phyla and 261 orders (Suppl. material 5, 6: Tables S3, S4). The pr2 and BLAST identification methods provided order-level identifications for over 400 ASVs that were initially unidentified using the SILVA database (Suppl. material 6: Table S4). However, even after using multiple assignment methods, nearly 14% of the ASVs remained identified as either uncultured, *incertae sedis*, or were not assigned a phylum, and an additional 7.5% were not assigned to order. The unidentified ASVs represented a relatively low number of sequences (67,932 sequences, or 6.4% of the total sequences retained after filtering), and over 80%

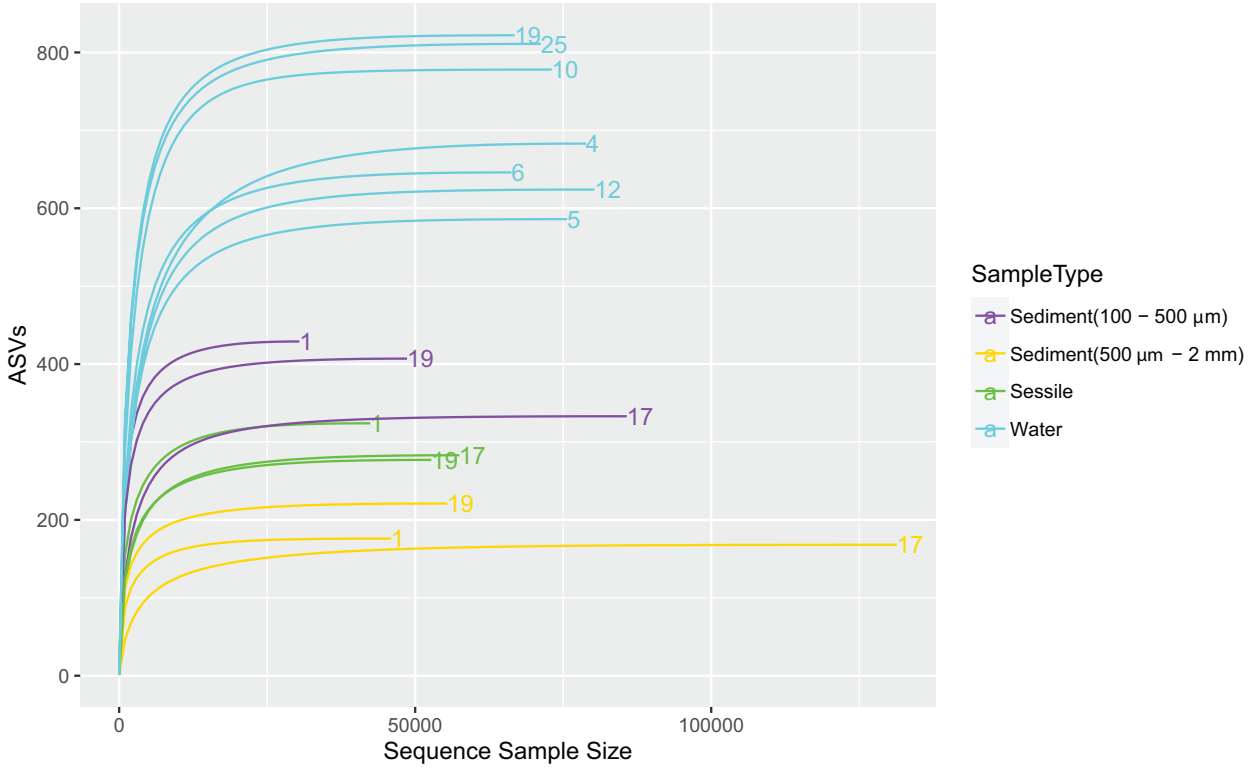

Figure 2. Sample-based rarefaction curves of ASV diversity detected in 18S rRNA (V1–V3) gene amplicon analysis of Palmyra Atoll. Sample types include 100-500 µm sediment, 500 µm -2 mm sediment, sessile material scrapings, and reef water. Given the uneven sampling depths, for some comparative analyses samples were later rarefied to even sequencing depth. Numbers indicate sampling site as shown in Fig. 1.

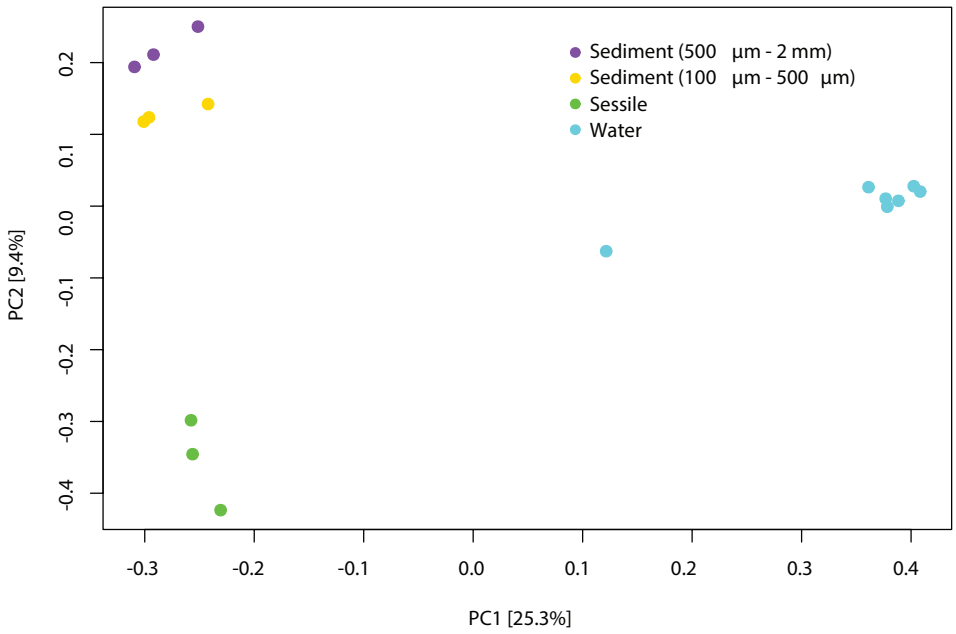

Figure 3. Principal coordinate analysis (PCoA) depicting distinct clusters of communities at the Palmyra Atoll National Wildlife Refuge. In addition to sea water obtained in 2012, three types of samples were collected in 2015 from the Autonomous Reef Monitoring Structures (ARMS): two separate sediment size fractions (100-500 µm and 500 µm-2 mm), and sessile material scraped from the plates. The PCoA is based on ASV presence or absence (Jaccard distance) in each of the four sample types

of these sequences were detected in water samples. This finding underscores the needs both for further research and more comprehensive reference databases. Among ASVs identified to order, few orders (mostly metazoan) had high relative abundance (proportion summed across sample types > 0.1), while many had lower relative abundance (with proportion summed across sample types < 0.1; Fig. 4; Suppl. material 3, 5, 6: Tables S1, S3, S4). For example, the top 10 orders by relative abundance were represented by 534,776 sequences and 784 ASVs, comprising ~50% of all sequences and ~20% of ASVs. However, given documented PCR amplification biases, estimates of relative abundance should be interpreted with care (Fonseca 2018).

The most common groups across samples were polychaete worms (order Eunicida), followed by demosponges (order Poecilosclerida) and bryozoans (order Cheilostomatida). Order Podocopida (ostracod crustaceans) was the fourth most frequent, the fifth was the dinoflagellate order Syndiniales, and the sixth was the Peracarida crustaceans (a group containing amphipods and isopods, ranked as an order in the SILVA 138 taxonomy but classified as a superorder in other taxonomies). The Corallinophycidae red algae, which contains the order Corallinales (crustose coralline algae, CCA), was the seventh most common. Order Scleractinia, or stony corals, was the eight most frequent, followed by another demosponge order (Dendroceratida) and Calanoida (a marine copepod group). Notably, sequences from two orders critical to healthy reefs (Corallinophycidae and Scleractinia) were detected in both water and sessile ARMS samples. Corallinophycid genera identified by eDNA included *Amphiroa*, *Hydrolithon*, *Jania*, *Lithothamnion*, *Mesophyllum*, *Neogoniolithon*, *Porolithon*, *Sporolithon*, *Titanoderma*, and the scleractinian coral genera *Acropora*, *Favites*, and *Montipora* were also detected (Suppl. Material 3, 5: Tables S1, S3). Coral symbiotic zooxanthellae (genus *Symbiodinium*), which are ejected when corals bleach, were also found, mainly in water but also in the sessile and coarse sediment fractions of the ARMS samples.

Sample types differed significantly for the number of ASVs (Kruskal-Wallis $p < 0.004$), Shannon-Weaver diversity (Kruskal-Wallis $p < 0.05$), and phylogenetic diversity (Kruskal-Wallis $p < 0.005$). Water samples contained many ASVs assigned to orders that were not represented in the list of top ten most prevalent orders, most notably non-metazoan eukaryotes / protists, and green algae (Fig. 4, Suppl. material 2, 3, 5, 6: Fig. S2, Tables S1, S3, S4). A large proportion of the variation identified in seawater was planktonic, with Calanoida (copepods) and Syndiniales (dinoflagellates) representing the two most common taxa. When non-metazoan taxa were excluded, however, water samples exhibited the lowest richness and phylogenetic diversity of the sample types (Suppl. material 4: Table S2). Although relatively few studies have examined

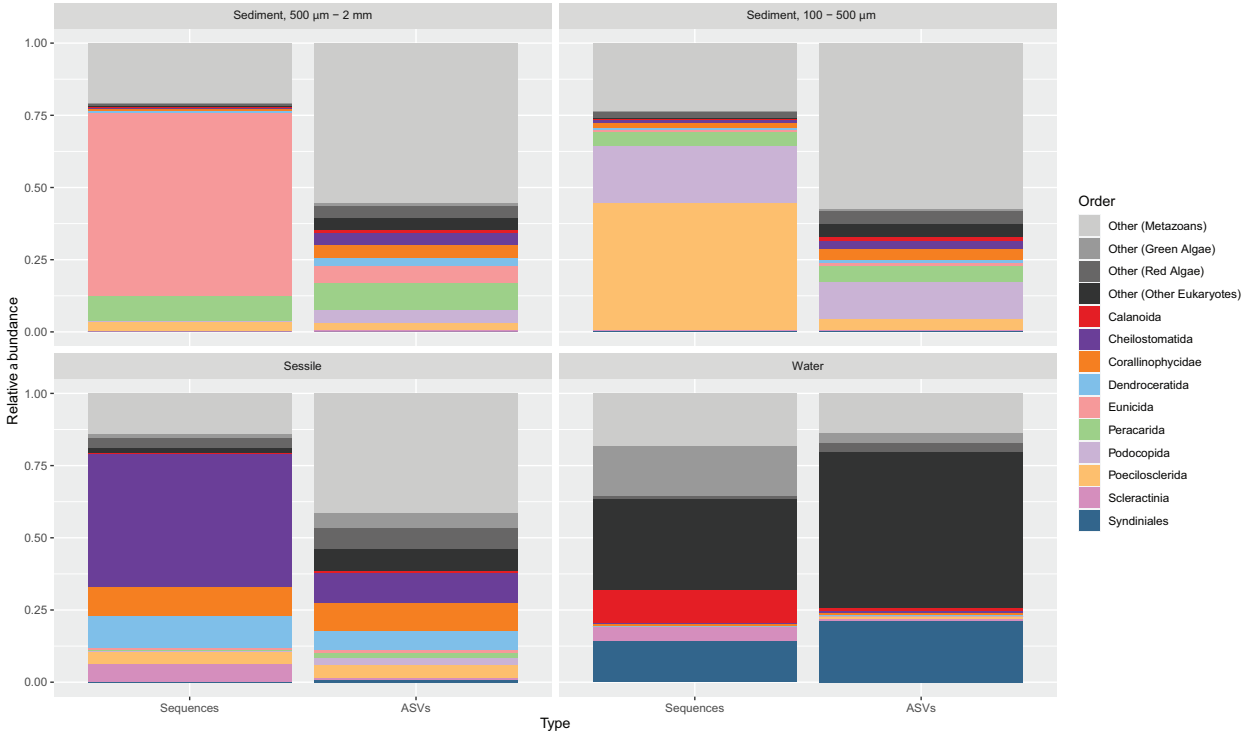

Figure 4. Relative ASV abundance of the ten most common orders sequenced for 18S rDNA (V1–V3). Samples were obtained from four sample types (500 µm-2 mm sediment fraction, 100-500 µm sediment fraction, sessile material scrapings; and water) collected from nine sites at Palmyra Atoll (Fig. 1). The top ten orders were Calanoida (copepods), Cheilostomatida (bryozoans), Corallinophycidae (red algae), Dendroceratida (sponges), Eunicida (polychaete worms), Peracarida (malacostracan crustaceans), Podocopida (ostracod crustaceans), Poecilosclerida (demosponges), Scleractinia (stony corals), and Syndiniales (dinoflagellates). Bar heights show the relative abundance of each taxon in terms of both sequences and number of ASVs. Relative abundances of ASVs not in the top ten orders and classified as metazoans, green algae, red algae, and other eukaryotic groups are shown in gray.

Table 2. Summary of sample data. The sample name (see Fig. 1), barcode sequence, and number of raw of reads per sample are depicted along with site and sample type information and the NCBI Sequence Read Archive (SRA) accession number within BioProject PRJNA804389. The linker primer sequence for all samples was AACCTGGTTGATCCTGCCAGT.

Sample Name	Barcode Sequence	Site	Sample type	Sample Year	Raw # of reads	SRA Accession
01A115	CTCTGACT	1	Sediment (100-500µm)	2015	17036	SRS11916995
01A115b	CTCTGAGA	1	Sediment (100-500µm)	2015	17606	SRS11916994
01A515	CTCTGTCA	1	Sediment (500µm-2mm)	2015	23522	SRS11917005
01A515b	CTCTGTGT	1	Sediment (500µm-2mm)	2015	22814	SRS11917012
01A15	CTCTTCAG	1	Sessile	2015	32664	SRS11917011
01A15b	CTCTTCTC	1	Sessile	2015	29761	SRS11917013
17A115	CTGACTCT	17	Sediment (100-500µm)	2015	29888	SRS11916996
17A115b	CTCTGTCA	17	Sediment (100-500µm)	2015	58338	SRS11916997
17A515	CTGACTGA	17	Sediment (500µm-2mm)	2015	49053	SRS11916998
17A515b	CTCTGTGT	17	Sediment (500µm-2mm)	2015	82789	SRS11916999
17A15	CTCTTGAC	17	Sessile	2015	32198	SRS11917000
17A15b	CTCTTGTG	17	Sessile	2015	29883	SRS11917001
19A115	CTGAGACT	19	Sediment (100-500µm)	2015	21791	SRS11917003
19A115b	CTCTTGAC	19	Sediment (100-500µm)	2015	33386	SRS11917004
19A515	CTGAGAGA	19	Sediment (500µm-2mm)	2015	20800	SRS11917002
19A515b	CTCTTGTG	19	Sediment (500µm-2mm)	2015	35722	SRS11917006
19A15	CTGAACAC	19	Sessile	2015	32291	SRS11917007
PAL.04	CTCTGACT	4	Water	2012	78908	SRS11917014
PAL.05	CTCTGAGA	5	Water	2012	76076	SRS11917016
PAL.06	CTCTGTCA	6	Water	2012	66775	SRS11917015
PAL.10	CTCTGTGT	10	Water	2012	73617	SRS11917017
PAL.12	CTCTTCAG	12	Water	2012	80428	SRS11916993
PAL.19	CTCTTCTC	19	Water	2012	67378	SRS11917009
PAL.25	CTCTTGAC	25	Water	2012	71342	SRS11917010

the different communities recovered from metabarcoding seawater and ARMS from the same location, metazoan diversity between multiple sample types, including combined ARMS biomass and seawater, has been compared previously and confirmed higher metazoan diversity in the combined ARMS biomass (Nichols et al. 2021). In our work, the higher overall diversity and uniqueness of seawater compared to the combined ARMS biomass is dependent on inclusion of unicellular eukaryotes, as dinoflagellates and other unicellular organisms made up the majority of the seawater sequences and ASVs. However, caution is warranted as the water samples were collected in 2012 and the ARMS samples in 2015.

Among the ARMS samples, the sessile material scraping and 100–500 µm sediment fractions tended to have higher diversity than the 500 µm-2 mm ones (Table 1). Sequences from the sessile fraction represented mostly colonial metazoans that make up the physical structure of the reef, with scleractinian corals, demosponges, bryozoans, and CCA all present. The coarser 500 µm-2 mm sample displayed the lowest overall diversity and the fewest unique ASVs. However, this sample type did contain many polychaete sequences (Eunicida) that were not found in the other ARMS fractions. As reported in other ARMS studies, the 100–500 um fraction showed higher diversity than the 500 um-2 mm sample (Leray and Knowlton 2015; Ransome et al. 2017; Pearman et al. 2018). The two sediment fractions both recovered small malacostracan crustaceans (Peracarida), and the fine sample also contained many sequences from demosponges and ostracod crustaceans not found in the coarser fraction.

Notably, all four of the sample types contained unique ASVs, suggesting that complete characterization of reef biodiversity requires metabarcoding of a range of different sample types over time. This provides a complementary picture of reef biodiversity that recovers many sessile components of reef structure and small metazoans that would otherwise be missed in water column sampling. In conclusion, this data set paves the way for a better understanding of eukaryotic biodiversity in this largely pristine reef system, as well as guidance for future studies that should pay careful attention to updated protocols. This includes careful use of replicates at each site, sequencing of extraction blanks and negative / positive PCR controls, and meticulous attention to potential errors in relative abundance estimates (Fonseca 2018; Taberlet et al. 2018; Minamoto et al. 2021). This work demonstrated the breadth of biodiversity accessible through 18S rDNA metabarcoding of different types of marine environmental and biomass samples, and also provided a valuable baseline from a relatively intact and pristine reef system to which other reefs may be compared.

Data availability

The 18S rDNA amplicon gene sequences from this work are posted on the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA804389 (Table 2). Scripts and R code used for processing and analyzing data are available at https://github.com/nerdbrained/palmyra_edna. All DNA extracts are stored in Forest Rohwer's lab or at NOAA.

Conflicts of interest

The authors declare no competing interests.

Acknowledgments

Funding for this project was provided by the Professional Staff Congress of the City University of New York (to ENM) and the Lerner Gray Fund for Marine Research of the American Museum of Natural History (to JAS). All work was carried out under authorized permits to Forest Rohwer and Rusty Brainard (NOAA). We would like to thank Kevin Green and Ben Knowles at San Diego State University for the logistical processing of water samples, as well as Kerry Reardon at NOAA for ARMS collections, and Seth Wollney and Vasiliki Stergioula at CSI for assistance with bioinformatics and lab work. Finally, we extend thanks to Eleanor Sterling and George Amato for their initial advice and direction on the project. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. We are grateful to our Editor, Anastasija Zaiko, as well as Reviewers David Stankovic, Florian Leese, and anonymous for helpful comments.

References

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Caporaso JG (2018) Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6(1): e90. https://doi.org/10.1186/s40168-018-0470-z

Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope EK, Da Silva R, Diener C, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibbons SM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley GA, Janssen S, Jarmusch AK, Jiang L, Kaehler BD, Kang K Bin, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton JT, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso JG (2019) QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. Nature Biotechnology 37: 852–857. https://doi.org/10.7287/peerj.preprints.27295v2

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13(7): 581–583. https://doi.org/10.1038/nmeth.3869

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST +: Architecture and applications. BMC Bioinformatics 10(1): e421. https://doi.org/10.1186/1471-2105-10-421

Carpenter KE, Abrar M, Aeby G, Aronson RB, Banks S, Bruckner A, Chiriboga A, Cortés J, Delbeek JC, DeVantier L, Edgar GJ, Edwards AJ, Fenner D, Guzmán HM, Hoeksema BW, Hodgson G, Johan O, Licuanan WY, Livingstone SR, Lovell ER, Moore JA, Obura DO, Ochavillo D, Polidoro BA, Precht WF, Quibilan MC, Reboton C, Richards ZT, Rogers AD, Sanciangco J, Sheppard A, Sheppard C, Smith J, Stuart S, Turak E, Veron JEN, Wallace C, Weil E, Wood E (2008) One-third of reef-building corals face elevated extinction risk from climate change and local impacts. Science 321(5888): 560–563. https://doi.org/10.1126/science.1159196

Collen JD, Garton DW, Gardner JPA (2009) Shoreline changes and sediment redistribution at Palmyra Atoll (Equatorial Pacific Ocean): 1874–present. Journal of Coastal Research 253: 711–722. https://doi.org/10.2112/08-1007.1

Davis NM, Proctor Di M, Holmes SP, Relman DA, Callahan BJ (2018) Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6(1): 1–14. https://doi.org/10.1186/s40168-018-0605-2

Deiner K, Yamanaka H, Bernatchez L (2021) The future of biodiversity monitoring and conservation utilizing environmental DNA. Environmental DNA 3(1): 3–7. https://doi.org/10.1002/edn3.178

del Campo X, Kolisko M, Boscaro V, Santoferrara LF, Nenarokov S, Massana R, Guillou L, Simpson A, Berney C, Vargas C De, Brown MW, Keeling PJ, Parfrey LW (2018) EukRef: Phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. PLoS Biology 16: e2005849. https://doi.org/10.1371/journal.pbio.2005849

Dinsdale EA, Pantos O, Smriga S, Edwards RA, Angly F, Wegley L, Hatay M, Hall D, Brown E, Haynes M, Krause L, Sala E, Sandin SA, Thurber RV, Willis BL, Azam F, Knowlton N, Rohwer F (2008) Microbial Ecology of Four Coral Atolls in the Northern Line Islands. PLoS ONE 3: e1584. https://doi.org/10.1371/journal.pone.0001584

Dowd SE, Sun Y, Wolcott RD, Domingo A, Carroll JA (2008) Bacterial Tag-Encoded FLX Amplicon Pyrosequencing (bTEFAP) for Microbiome Studies: Bacterial Diversity in the Ileum of Newly Weaned Salmonella-Infected Pigs. Foodborne Pathogens and Disease 5(4): 459–472. https://doi.org/10.1089/fpd.2008.0107

Fonseca VG (2018) Pitfalls in relative abundance estimation using eDNA metabarcoding. Molecular Ecology Resources 18(5): 923–926. https://doi.org/10.1111/1755-0998.12902

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, Del Campo J, Dolan J, Dunthorn M, Edvardsen B, Holzmann M, Kooistra W, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R (2013) The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41(D1): D597–D604. https://doi.org/10.1093/nar/gks1160

Haas AF, Knowles B, Lim YW, McDole Somera T, Kelly LW, Hatay M, Rohwer F (2014) Unraveling the Unseen Players in the Ocean - A Field Guide to Water Chemistry and Marine Microbiology. Journal of Visualized Experiments 93: e52131. https://doi.org/10.3791/52131

Huson DH, Beier S, Flade I, Górska A, El-hadidi M (2016) MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. Computational Biology 12: e1004957. https://doi.org/110.1371/journal.pcbi.1004957

Ingala MR, Werner IE, Fitzgerald AM, Naro-Maciel E (2021) 18S rRNA amplicon sequence data (V1-V3) of the Bronx River estuary, New York. Metabarcoding and Metagenomics 5: 153–162. https://doi.org/10.3897/mbmg.5.69691

Kelly RP, Closek CJ, O'Donnell JL, Kralj JE, Shelton AO, Samhouri JF (2017) Genetic and manual survey methods yield different and complementary views of an ecosystem. Frontiers in Marine Science 3: 1–11. https://doi.org/10.3389/fmars.2016.00283

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, Blomberg SP, Webb CO (2010) Picante: R tools for integrating phylogenies and ecology. Bioinformatics (Oxford, England) 26(11): 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Knowlton N, Jackson JBC (2008) Shifting Baselines, Local Impacts, and Global Change on Coral Reefs. PLoS Biology 6(2): e54. https://doi.org/10.1371/journal.pbio.0060054

Knowlton N, Brainard RE, Fisher R, Moews M, Plaisance L, Caley MJ (2010) Coral Reef Biodiversity. In: Life in the World's Oceans. Wiley-Blackwell, Oxford, UK, 65–78. https://doi.org/10.1002/9781444325508.ch4

Lafferty KD, Garcia-Vedrenne AE, McLaughlin JP, Childress JN, Morse MF, Jerde CL (2021) At Palmyra Atoll, the fish-community environmental DNA signal changes across habitats but not with tides. Journal of Fish Biology 98(2): 415–425. https://doi.org/10.1111/jfb.14403

Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proceedings of the National Academy of Sciences of the United States of America 112(7): 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray M, Knowlton N (2016) Censusing marine eukaryotic diversity in the twenty-first century. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 371(1702): e20150331. https://doi.org/10.1098/rstb.2015.0331

McDole T, Nulton J, Barott KL, Felts B, Hand C, Hatay M, Lee H, Nadon MO, Nosrat B, Salamon P, Bailey B, Sandin SA, Vargas-Angel B, Youle M, Zgliczynski BJ, Brainard RE, Rohwer F (2012) Assessing Coral Reefs on a Pacific-Wide Scale Using the Microbialization Score. PLoS ONE 7(9): e43233. https://doi.org/10.1371/journal.pone.0043233

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE 8(4): e61217. https://doi.org/10.1371/journal.pone.0061217

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71(2): 491–499. https://doi.org/10.1016/0378-1119(88)90066-2

Minamoto T, Miya M, Sado T, Seino S, Doi H, Kondoh M, Nakamura K, Takahara T, Yamamoto S, Yamanaka H, Araki H, Iwasaki W, Kasai A, Masuda R, Uchii K (2021) An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols. Environmental DNA 3(1): 8–13. https://doi.org/10.1002/edn3.121

MRDNA (2021) FASTQ Processor. www.mrdnafreesoftware.com

Naro-Maciel E, Ingala MR, Werner IE, Fitzgerald AM (2020) 16S rRNA Amplicon Sequencing of Urban Prokaryotic Communities in the South Bronx River Estuary. Microbiology Resource Announcements 9(22): e00182-20. https://doi.org/10.1128/MRA.00182-20

Nichols PK, Timmers M, Marko PB (2021) Hide 'n seq: Direct versus indirect metabarcoding of coral reef cryptic communities. Environmental DNA 4(1): 93–107. https://doi.org/10.1002/edn3.203

Oksanen J (2019) Vegan: ecological diversity. https://cran.r-project.org/package=vegan

Pearman JK, Leray M, Villalobos R, Machida RJ, Berumen ML, Knowlton N, Carvalho S (2018) Cross-shelf investigation of coral reef cryptic benthic organisms reveals diversity patterns of the hidden majority. Scientific Reports 8(1): e8090. https://doi.org/10.1038/s41598-018-26332-5

Plaisance L, Knowlton N, Paulay G, Meyer C (2009) Reef-associated crustacean fauna: Biodiversity estimates using semi-quantitative sampling and DNA barcoding. Coral Reefs 28(4): 977–986. https://doi.org/10.1007/s00338-009-0543-3

Plaisance L, Caley MJ, Brainard RE, Knowlton N (2011) The Diversity of Coral Reefs: What Are We Missing? PLoS ONE 6(10): e25026. https://doi.org/10.1371/journal.pone.0025026

QIIME2 (2021) denoise-single: Denoise and dereplicate single-end sequences. https://docs.qiime2.org/2021.4/plugins/available/dada2/denoise-single/ [July 19, 2021]

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Research 41(D1): D590–D596. https://doi.org/10.1093/nar/gks1219

Quesada V (2020) nVennR: Create n-Dimensional, Quasi-Proportional Venn Diagrams.

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/

Ransome E, Geller JB, Timmers M, Leray M, Mahardini A, Sembiring A, Collins AG, Meyer CP (2017) The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. PLoS ONE 12(4): 1–19. https://doi.org/10.1371/journal.pone.0175066

Robeson MS, O'Rourke DR, Kaehler BD, Ziemski M, Dillon MR, Foster JT, Bokulich NA (2021) RESCRIPt: Reproducible sequence taxonomy reference database management. PLoS Computational Biology 17(11): e1009581. https://doi.org/10.1371/journal.pcbi.1009581

Sandin SA, Smith JE, DeMartini EE, Dinsdale EA, Donner SD, Friedlander AM, Konotchick T, Malay M, Maragos JE, Obura D, Pantos O, Paulay G, Richie M, Rohwer F, Schroeder RE, Walsh S, Jackson JBC, Knowlton N, Sala E (2008) Baselines and degradation of coral reefs in the Northern Line Islands. PLoS ONE 3(2): e1548. https://doi.org/10.1371/journal.pone.0001548

Servis JA, Reid BN, Timmers MA, Stergioula V, Naro-Maciel E (2020) Characterizing coral reef biodiversity: Genetic species delimitation in brachyuran crabs of Palmyra Atoll, Central Pacific. Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis 31(5): 178–189. https://doi.org/10.1080/24701394.2020.1769087

Smriga S, Sandin SA, Azam F (2010) Abundance, diversity, and activity of microbial assemblages associated with coral reef fish guts and feces. FEMS Microbiology Ecology 73: 31–42. https://doi.org/10.1111/j.1574-6941.2010.00879.x

Sterling EJ, McFadden KW, Holmes KE, Vintinner EC, Arengo F, Naro-Maciel E (2013) Ecology and conservation of marine turtles in a Central Pacific foraging ground. Chelonian Conservation and Biology 12(1): 2–16. https://doi.org/10.2744/CCB-1014.1

Taberlet P, Bonin A, Zinger L, Coissac E (2018) Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press, London, 253 pp. https://doi.org/10.1093/oso/9780198767220.001.0001

Weekers P, Gast R, Fuerst P, Byers T (1994) Sequence variations in small-subunit ribosomal RNAs of *Hartmannella vermiformis* and their phylogenetic implications. Molecular Biology and Evolution 11: 684–690. https://doi.org/10.1093/oxfordjournals.molbev.a040147

Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO (2014) The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research 42(D1): D643–D648. https://doi.org/10.1093/nar/gkt1209

Zhang B, Penton CR, Xue C, Quensen JF, Roley SS, Guo J, Garoutte A, Zheng T, Tiedje JM (2017) Soil depth and crop determinants of bacterial communities under ten biofuel cropping systems. Soil Biology & Biochemistry 112: 140–152. https://doi.org/10.1016/j.soilbio.2017.04.019

Supplementary material 1
Figure S1
Author: Brendan Reid
Data type: jpg. file
Explanation note: Venn diagram indicating the number of shared and unique ASVs (in bold) in the full dataset. Numbers 1-4 in parentheses correspond to the following sample types: 100-500 µm sediment fraction (1); 500 µm-2 mm sediment fraction (2); sessile material scrapings (3); and water (4).

Link: https://doi.org/10.3897/mbmg.6.78762.suppl1

Supplementary material 2
Figure S2
Author: Brendan Reid
Data type: jpg.file
Explanation note: Heatmap of sequence abundances for the top 25 order-level classifications. RAD_A represents a Retarian group, and Subclade_B is a group of chlorophyte algae.

Link: https://doi.org/10.3897/mbmg.6.78762.suppl2

Supplementary material 3
Table S1
Author: Brendan Reid
Data type: xslx.file
Explanation note: The number of sequences for each ASV detected by sample type. When possible, assignments at the lowest taxonomic level in the SILVA database were associated with higher levels (genus and family) consistent with the NCBI's currently accepted taxonomy. ASVs identified as putative contaminants are included at the bottom of the table (available at https://github.com/nerdbrained/palmyra_edna).

Link: https://doi.org/10.3897/mbmg.6.78762.suppl3

Supplementary material 4
Table S2
Author: Brendan Reid
Data type: pdf. file
Explanation note: Site-level diversity statistics for either all eukaryotes, or metazoans only, in each sample type. Included are Amplicon Sequence Variants before (ASVs) and after (ASVsrare) rarefaction to the lowest sample size (30,443 reads). H = Shannon-Weaver diversity. PD = Faith's phylogenetic diversity.

Link: https://doi.org/10.3897/mbmg.6.78762.suppl4

Supplementary material 5
Table S3
Author: Brendan Reid
Data type: xslx. file
Explanation note: Ranked phyla, with class and orders, by relative abundance summed across sample type types.

Link: https://doi.org/10.3897/mbmg.6.78762.suppl5

Supplementary material 6
Table S4
Author: Brendan Reid
Data type: xslx. file
Explanation note: Additional order-level identifications added after pr2 and BLAST analyses.

Link: https://doi.org/10.3897/mbmg.6.78762.suppl6