Biogeography of reef water microbes from within reef to global scales

Lei Ma[1,2], Cynthia Becker[1,2], Laura Weber[1], Chris Sullivan[3], Brian Zgliczynski[3], Stuart Sandin[3], Marilyn Brandt[4], Tyler B. Smith[4], Amy Apprill [1*]

1 Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA

2 MIT-WHOI Joint Program in Oceanography/Applied Ocean Science & Engineering, Cambridge and Woods Hole, MA, USA

3 Scripps Institution of Oceanography, UC San Diego, La Jolla, CA 92037, USA

4 University of Virgin Islands, St. Thomas 00802, USVI

* Corresponding author: aapprill@whoi.edu

Keywords: 16S rDNA, Biogeography, Coral reef communities, Free-living bacteria, Marine microbial ecology, meso-scale spatial variations, secondary analysis, Microbial diversity

Abstract

Seawater microorganisms play an important role in coral reef ecosystem functioning and can be influenced by biological, chemical, and physical features of reefs. As coral reefs continue to respond to environmental changes, the reef seawater microbiome has been proposed as a conservation tool for monitoring perturbations. However, the spatial variability of reef seawater microbial communities is not well studied, limiting our ability to make generalizable inferences across reefs. In order to better understand how

19 microorganisms are distributed at multiple spatial scales, we examined seawater microbial
20 communities in Florida Reef Tract and U.S. Virgin Islands reef systems using a nested
21 sampling design. On three reefs per reef system, we sampled seawater at regular spatial
22 intervals close to the benthos. We assessed the microbial community composition of these
23 waters using ribosomal RNA gene amplicon sequencing. Our analysis revealed that reef
24 water microbial communities varied as a function of reef system and individual reefs, but
25 communities did not differ within reefs and were not significantly influenced by benthic
26 composition. For the reef system and inter-reef differences, abundant microbial taxa were
27 found to be potentially useful indicators of environmental difference due to their high
28 prevalence and variance. We further examined reef water microbial biogeography on a
29 global scale using a secondary analysis of five studies, which revealed that microbial
30 communities were more distinct with increasing geographic distance. These results suggest
31 that biogeography is a distinguishing feature for reef water microbiomes, and that
32 development of monitoring criteria may necessitate regionally-specific sampling and
33 analyses.

34

1 Introduction

Coral reefs are currently experiencing significant challenges due to global and local factors (Hughes et al. 2017). Among them, climate change and ocean acidification affect corals worldwide while stressors such as human impacts and disease outbreaks are more localized. These crises are driving the development of new management and conservation strategies to preserve and monitor reef biodiversity. Awareness of the coral as a holobiont - an assemblage of a host and all of its associated symbiotic microorganisms (Knowlton & Rohwer 2003, Rosenberg et al. 2007) - has spurred research into establishing microbial solutions to reef stress, such as coral probiotics and microbiome-based monitoring (Peixoto et al. 2017, Glasl et al. 2017). In particular, a holistic characterization of microbes in coral reefs will aid in predicting reef resilience and environmental threats (Kelly et al. 2018).

Reefs harbor many distinct niches for bacterial and archaeal communities, including corals, sponges, sediments, and the water-column itself (Tout et al. 2014, McDevitt-Irwin et al. 2017). Free living water column microbes, residing above the reef substrate, are influenced by hydrological conditions (Sweet et al. 2010, Becker et al. 2020), general benthic community composition (Haas et al. 2011, Kelly et al. 2014), local nutrient regimes (van Duyl & Gast 2001, Nelson et al. 2011), and temporal dynamics (Weber & Apprill 2020, Becker et al. 2020). When combined, these influences cause reef-associated seawater microbiomes to be readily distinguishable between reefs, as well as between zones within a reef (Jeffries et al. 2015, Salerno et al. 2016, Frade et al. 2020). Microbial communities can be powerful indicators of reef health and environmental conditions (Glasl et al. 2017).

Indeed, as reefs transition from coral to algae dominated, the exudates released from the benthos also likely shift, causing increased heterotrophy and decreased oligotrophy in the seawater microbiome (Haas et al. 2011, Nelson et al. 2013). In a proccess called microbialization, the heterotroph-dominated microbial community further depresses growth of coral and encourages the growth of algae (Haas et al. 2016, Kelly et al. 2018). This microbial phase shift may be an important process to monitor in at-risk reefs. Additionally, reef microorganisms respond rapidly to nutrient and temperature fluctuations, potentially providing a sensitive and non-invasive diagnostic or predictive tool for perturbations that may provide knowledge prior to visible reef changes (Glasl et al. 2019, Becker et al. 2020).

Implementation of large scale reef water monitoring efforts for reef microorganisms is partially limited by our understanding of reef seawater microbial diversity across spatial scales (Bourne et al. 2016, Glasl et al. 2017). Biogeographic patterns of coral reef microbial assemblages have been found at a variety of spatial scales. Small scale patterns such as within a coral skeleton (Marcelino et al. 2018), in the boundary layer overlying the coral mucus (Weber et al. 2019), and in micro-habitats generated by coral structures (Schöttner et al. 2012) highlight potential mechanisms affecting reef microbial composition, but may not represent the state of an entire reef. On the other hand, studies and models of marine microbial distribution at the scale of oceans (Amend et al. 2012, Hellweger et al. 2014) provide insight into the global drivers of microbial abundance, but are not specific to the unique environments of reefs. Therefore, a better understanding of the biogeography of coral reef seawater microbes across distinct spatial scales is warranted.

The goal of this study is to understand the variability of coral reef seawater microbial communities across different spatial scales. We examined this question in two parts. For the first part, we examined reef water microbial communities within and between two reef systems to understand the influence of both reef and reef benthic composition on microbial diversity (Figure 1A). Secondly, in order to quantify the impact of larger geographic distances on reef water microbial communities, we conducted a secondary analysis of aggregated 16S rRNA gene sequences from five studies that used similar sampling methodology (Figure 1B, Table S1). We predicted that microbial community structure would differ primarily on the scale of individual reefs but and secondarily on underlying benthic structure. Additionally, we expected the secondary analysis to recapitulate the individual reef and reef system-based biogeographic patterns seen in the Fl and VI systems on a more global scale.

2 Materials and Methods

2.1 FL and VI transects: Sampling

The first part of this study took place in two reef systems, the Florida Reef Tract (Fl) in June 2019 and off the southern coast of St. Thomas in the U.S. Virgin Islands (VI) in February 2020. A total of three reefs were sampled in the Florida Keys; the northernmost reef was Biscayne, located within the boundaries of Biscayne National Park, the reef Grecian was located at the Grecian Rocks reef off the coast of Key Largo, and the reef Dry Tortugas was located within Dry Tortugas National Park. All reefs in the Florida Reef Tract (Fl) were forereefs within the barrier reef. Similarly, three reefs were sampled in St. Thomas (VI);

Brewers Bay and Black Point were forereef zones on fringing reefs a few hundred meters from the coast while Flat Cay was a fringing reef located near an uninhabited island named Flat Cay about two kilometers off the coast (Table 1). Average sampling depth was between 5.0 and 7.1 meters, with the exception of the Dry Tortugas, which was deeper with an average depth of 18.0 meters. Due to the difference in season of sampling, the average temperature in the Florida reefs was slightly higher than in the Virgin Islands reefs (28.7 °C and 26.9 °C respectively).

At each reef, three 10 meter transects were taken by laying down a 10 m weighted line that was marked every meter. Water samples were taken by a diver using a 60 or 100 mL syringe positioned approximately 5 cm above the benthos at each meter line. The transects were laid haphazardly, but did not intersect with each other. Because of inclement conditions, only 1 transect was collected at the Biscayne reef. At the Fl reefs, benthic composition - represented by percent cover of coral skeleton, crustose coralline algae, cyanobacteria, hard coral, macroalgae, non biological, other invertebrates, soft coral, sponge, and turf algae - was determined using large-area imagery collected from 10 by 10 m area plots. All transects were placed within these 100 m^2 plots. Stratified random points (2500) were dropped across the reef area and classified to generate reef-wide cover estimates - see full methods in Fox et al. (2019). At the VI reefs, benthic composition was recorded at the precise location of each syringe sample using a video survey of the transect line as well as noted in writing by a diver during sampling. Video and written record were cross-referenced and each sample was then classified into a single category from algae, dead coral, live coral, rock, sand, sponge, and undetermined.

To capture the seawater microbial community, 60 mL of the seawater was filtered through a 0.22 µm Supor filter (25 mm; Pall Corporation). The water volume of 60 mL has previously been found to be comparable to larger volumes (1-2L) for characterizing seawater microbial communities using amplicon sequencing (Weber et al. 2019). Filters were placed in 2 mL cryovials, flash frozen in a liquid nitrogen dry shipper, and processed upon returning to Woods Hole, MA.

2.2 FL and VI transects: DNA extraction, PCR amplification, and sequencing

DNA was extracted from the filters using the DNeasy PowerBiofilm Kit (Qiagen) according to manufacturer protocols. Seven DNA extraction controls, consisting of unused 0.22 µm filters, were processed alongside samples. Extracted DNA was quantified using the Qubit 2.0 fluorometer HS dsDNA assay (ThermoFisher Scientific). Primers 515FY (Parada et al. 2016) and 806RB (Apprill et al. 2015) containing Illumina overhang adapter sequences were used to amplify the V4 region of the small subunit rRNA gene in bacteria and archaea. PCR reactions contained 14.75 µL molecular grade water, 5 µL GoTaq Flexi 5X buffer (Promega Corporation), 2.5 µL of 25 mM $MgCl_2$, 1 µL of 10 mM dNTPs, 1 µL of 10 mM forward and reverse primers, 0.5 µL GoTaq DNA polymerase (Promega) and 1 µL of DNA template. Three PCR controls consisting of 1 µL of PCR-grade water as template were also included, as well as microbial genomic DNA from a Human Microbiome Project mock community (BEI Resources, NIAID, NIH as part of the Human Microbiome Project: Genomic DNA from Microbial Mock Community B (Even, Low Concentration), v5.1L, for 16S rRNA Gene Sequencing, HM-782D). The first stage PCR conditions were: 28 cycles (95°C 20s, 55°C 20s, 72°C 5 min) with a 2 min 95°C hot start and 10 min 72°C final elongation. PCR

products were screened for quality using gel electrophoresis and purified using the MinElute PCR purification kit (Qiagen). PCR products were then barcoded using the Nextera XT Index Kit v2 set A primers (Illumina) using the following conditions: 8 cycles (95°C 30s, 55°C 30s, 72°C 30s) with 3 min 95° hot start and 5 min 72°C final elongation. Barcoded products were purified as above and concentrations of the purified products were assessed using the HS dsDNA assay on the Qubit 2.0 fluorometer (ThermoFisher Scientific). Products were diluted with Tris HCl to 5 nM before being pooled randomly into two libraries. The libraries were diluted to a final loading concentration of 50 pM with a 5% spike-in of 50 pM PhiX. The libraries were then sequenced on the iSeq 100 System (Illumina) using paired-end 150 bp reads. Data are accessible in the NCBI Sequence Read Archive under bioproject PRJNA733652.

2.3 FL and VI transects: Data analysis

All code used to generate the figures and analyses in this paper is publicly available on GitHub (https://github.com/microlei/AME_biogeography_2021). Sequences were processed using the DADA2 package (v. 1.12.1) in R (v. 3.6.2) (Callahan et al. 2016). Due to the short length of iSeq reads (150bp), it was not possible to merge the reads and therefore only forward reads were used in analysis. Forward reads were filtered using the default parameters of the function filterAndTrim in DADA2 except: trimLeft=20 (to remove the primer), truncLen=125, truncQ=2, maxEE=1. The parameter of truncLen was determined after observing quality dropping during the last 5 bp of the fastq reads. Chimera removal and amplicon sequence variant (ASV) generation was also done by DADA2. Taxonomy was assigned, without percent identity clustering, using the naive Bayesian classifier method of

Wang et al. (2007) trained on the Silva SSU rRNA database (Version 138) (Pruesse et al. 2007). Putative contaminant reads were identified using the prevalence method in the R package decontam (v. 1.4.0) (Davis et al. 2018) by using the negative controls to identify contaminants; contaminant reads were subsequently removed. Reads matching the Kingdom Eukaryota or Order Chloroplast were also removed.

Data analysis was completed in RStudio (v 1.2.5.001) (RStudio Team 2019) using, primarily, the packages phyloseq (v 1.28.0) and vegan (v 2.5-7) (McMurdie & Holmes 2013, Oksanen et al. 2020). Graphics were generated using ggplot2 (v 3.3.3) (Wickham 2016). Alpha diversity metrics were estimated using the *estimate_richness* function in vegan with unrarefied read counts. Differences in alpha diversity metrics were assessed using pairwise t tests corrected for multiple comparisons using the holm method (Holm 1979). To understand the variability of microbial community composition across all samples, ASV counts were transformed to relative abundances and Bray-Curtis dissimilarity was calculated between each sample pair. Dissimilarity values were plotted using non-metric multidimensional scaling ordination (NMDS). Taxa that were differentially abundant were identified using the package corncob (v 0.2.0)(Bryan D Martin et al. 2021), with a false discovery rate cutoff of 0.05 using Benjamini-Hochberg correction.

In order to account for differences in read counts arising from variances in sequencing depth and read quality as well as to improve the quality of the distance-decay analyses, the Aitchison distance was used instead of Bray-Curtis when comparing community similarity across studies in the secondary analysis and, for consistency, between samples in the transect-based study (Gloor et al. 2017, Clark et al. 2021). The package zCompositions (v

1.3.4) (Palarea-Albaladejo & Martín-Fernández 2015) was used to impute zeroes before performing a centered-log-ratio (CLR) transformation on the count data. Taking the euclidean distances of the CLR transformed data generated the Aitchison distances. Geographic distances were calculated using the *gdist* function in the package Imap (v 1.32) (John R Wallace 2012), which uses the Vincenty inverse formula for ellipsoids. The *adonis2* function in the vegan package was used to perform PERMANOVA analysis (999 iterations) on the dissimilarity indices at the scales of transect (within reef), reef, and reef system. Differences in dispersion at various spatial scales were calculated using the vegan function *betadisper* and tested using the vegan function *permutest*, which performs a permutation test (999 permutations) of multivariate homogeneity of groups dispersions. The *mantel* function in the vegan package was used (999 permutation) to test for correlation between the geographic distance matrix and the community similarity for both the secondary analysis and transect-based study.

2.4 Secondary analysis: Sample information

The methods for sample collection, DNA extraction, PCR amplification, and sequencing used by the five studies in the secondary analysis are highly similar with small variations. The studies collected seawater from reefs at a variety of depths, ranging from surface (0.3 m) to benthic (13 m). Seawater sampling was done by filtering replicate 2 liter volumes of seawater through 0.22 µm pore size, 25 mm Supor® filters using a peristaltic pump. For DNA extraction, all studies used bead beating followed by spin column purification, although the DNA extraction reagents differed. Neave et al. (2017) used the PowerPlant Pro DNA isolation kit (Qiagen) while Weber et al. (2020) used a sucrose-lysis with bead beating

method followed by column purification with the Qiagen DNeasy Blood and Tissue Kit (Santoro et al. 2010) as well as a phenol chloroform protocol (Urakawa et al. 2010) and pooled the extracts. Unpublished data from Becker et al. used the DNeasy PowerBiofilm kit (Qiagen). Full methods for these unpublished data are included in the Supplemental Text S1. The remaining two studies used the extraction method described in Santoro et al. (2010).

All studies amplified the V4 hypervariable region of the 16S rRNA gene using the reverse primer described in Apprill et al. (2015), but two studies (Neave et al. 2017, and Apprill et al. 2021) used the forward primer not optimized for Thaumarchaeota (Caporaso et al. 2011) while the others used the forward primer described in Parada et al. (2016). All studies used the 250 bp paired-end Illumina MiSeq platform, although Weber et al. (2020) used the Fluidgm® platform (Fluidgm Corporation) for library preparation while others followed the methods described in Kozich et al. (2013). Primer choice, sequencing technology, and DNA extraction method are known to influence downstream 16S rRNA gene sequence analysis, such as in marine biofilms and seawater (Urakawa et al. 2010, Corcoll et al. 2017). A comparison of different DNA extraction techniques on aquatic samples concluded that rare taxa are more affected by differing extraction technique, driving small but significant differences in Bray-Curtis distances (Liu et al. 2019). However, the secondary analysis is based on pairwise Aitchison distances, which are less influenced by presence/absence of individual taxa (Gloor et al. 2017) and does not seek to compare groups of samples based on distances.

2.5 Secondary Analysis: Data acquisition and processing

Raw sequence data from the five studies in the secondary analysis were collected from the NCBI Sequence Read Archive (SRA) and for the unpublished study, with consent from the authors. Because the transect comparisons collected in Fl and VI were sequenced with shorter reads, these samples were excluded from the secondary analysis. Using metadata from the studies, sequence files were filtered such that only samples taken from reef associated seawater (and not controls) were included. Samples were classified based on the reef that was sampled as well as the overall reef system (Table S1). Primer sequences were removed using cutadapt (Martin 2011). Sequences were processed in DADA2 as above with the parameters trimLeft=(20,20), truncLen=(205,205), truncQ=(2,2), maxEE=(1,1), and error estimation was performed by pooling all sequences into one error model. Paired forward and reverse reads were assembled into one contig and trimmed to 230 bp. Chimera removal, ASV generation, and taxonomy assignment were performed as in section 2.3. Four samples with fewer than 10,000 reads were removed. Because negative controls are specific to each study, contaminant reads were not identified or removed, but reads matching the Kingdom Eukaryota and Order Chloroplast were removed. Data analysis was performed as described in section 2.3. Briefly, the package zCompositions (Palarea-Albaladejo & Martín-Fernández 2015) was used to impute zeroes before using the center-log-ratio transform to normalize the read counts. The Aitchison distance was then plotted against the geographic distance between samples to examine the distance-decay relationship between samples.

3 Results

3.1 Fl and VI transects: Site characteristics

In the Florida reef system, Dry Tortugas and Grecian were dominated by macroalgae (56-67% of cover), while Biscayne was dominated by turf algae (45% of cover). Hard coral was more abundant at Dry Tortugas (21%) and Biscayne (20 %), but only comprised 3% of Grecian (Figure S1A). In the Virgin Islands system, live coral predmoninated at Brewer's Bay and Flat Cay (40-43%), but algae was slightly more prevalent at Black Point (40% of cover) (Figure S1B).

At the time of sampling, all reefs with the exception of Dry Tortugas had been experiencing outbreaks of Stony Coral Tissue Loss Disease (SCTLD) to varying degrees of severity and duration (Precht et al. 2016, Brandt et al. 2021).

3.2 Fl and VI transects: Sequence output

After quality control of the Fl and VI 16S rRNA gene amplicons from reef water transects, a total of 13,382,051 reads were retained and the number of reads per sample ranged from 19,686 to 186,683 with a median of 78,976. A total of 20,488 amplicon sequence variants (ASVs) were identified over 156 samples. Per sample unique ASVs averaged 461. The abundance matrices of the ASV counts per sample were very sparse, comprising of 97.7% zeros, indicating that a small number of taxa comprised the majority of the dataset. Specifically, only 1,228 ASVs make up the top 90% of observations across all samples.

3.3 Fl and VI transects: Alpha and beta diversity metrics

Alpha diversity metrics of the reef water microbiomes, including observed ASV richness, Shannon index (a measure of evenness), and Simpson's index (a measure of dominance) measured at the transect (within reef), individual reef, and reef system level in Fl and VI showed comparable values with some notable differences. At the transect level (within reefs), observed ASVs were most variable (highest and lowest values) at the Dry Tortugas reefs, with some outliers at both Fl and VI reefs. Simpson's index was most variable at the Dry Tortugas and Brewer's Bay reef (Figure S2, pairwise t-test: $p<0.0001$ for all significant comparisons involving Dry Tortugas and Brewer's Bay). At the individual reef level, significantly lower Simpson's index values were detected at Dry Tortugas (mean of 0.93) and Brewer's Bay (mean of 0.94) compared to the other reefs (mean of 0.97). Both reefs also displayed significantly lower Shannon index values than three other reefs, except Biscayne (Figure S2, pairwise t-test: $p<0.002$ for significant differences, $p>0.1$ for nonsignificant differences). At the reef system level (Fl compared to VI), each of the diversity metrics were significantly different (Figure S2).

An NMDS ordination of Bray-Curtis dissimilarities between reef water microbiomes showed that the VI reefs clustered together along with the Dry Tortugas reef water microbial communities, and Fl reefs Grecian and Biscayne were more separated (Figure 2A). Bray-Curtis dissimilarities for reef system (Fl and VI) and individual reefs differed significantly, but transects within a reef were not significantly different from each other (PERMANOVA)(Table 2).

A comparison of microbial community beta dispersion, calculated as the distance from the centroid of each reef's reef microbial communities in principal coordinate space, showed significant differences at the level of reef system (permutest; F=30.12, p=0.001), individual reef (F=4.87, p=0.001), and within-reef transects (permutest; F=1.90, p=0.026). However, a post hoc test (Tukey's Honest Significant Differences) on the reef-based and transect-based beta dispersions found that the difference was driven solely by the comparisons between the Dry Tortugas reef versus Grecian and Black Point (Figure 2B). The Dry Tortugas reef had low variance among its samples while Grecian and Black Point had a larger variance among samples. When Dry Tortugas was removed from the analysis, there was no longer a significant effect of individual reef or transect on beta dispersion.

While only reef-wide benthic composition was recorded for the Fl reefs, we recorded the underlying benthic composition for each sample in the VI reefs to enable comparison between substrate type and the overlying seawater microbiome. When all samples from VI were considered together, Bray-Curtis dissimilarities weakly correlated with benthic substrate type (PERMANOVA; R^2=0.093, Pseudo-F=1.64, p=0.042). However, a comparison of live coral compared to other categories (grouped together) did not show a correlation (PERMANOVA; R^2=0.0091, Pseudo-F= 0.77, p=0.62), and when samples were nested by their respective reef, the correlation with benthic substrate was no longer significant (PERMANOVA; Pseudo-F=1.64, p= 0.23). Group dispersions were not different between benthic substrate classes or between live coral and other substrates and no taxa were found to be differentially abundant in either of these contrasts (permutest; F=2.32, p=0.063).

3.4 Fl and VI transects: Differentially abundant taxa

To investigate which taxa may be driving differences in community composition with respect to individual reefs, a differential abundance (DA) analysis was performed on the dataset. DA analysis revealed 138 taxa that were significantly differentially abundant ($p<0.05$, false discovery rate corrected using the Benjamini-Hochberg method) across the six reefs sampled. These significant taxa included common oligotrophic marine groups, such as the SAR11 and SAR86 clade, *Cyanobiaceae*, SAR116, and the Archaean Marine Group II. Opportunistic copiotrophs, such as *Flavobacteriaceae*, *Rhodobacteraceae*, and *Vibrionaceae* were also well represented. All significant taxa were among the most abundant and most variable (displayed the highest variance in their relative abundances across samples) in the dataset (Figure S3). A list of the significant taxa along with their sequences is provided in Supplemental Table 2.

3.5 Secondary analysis: Sequence output and methods analysis

After assembly and quality control of the raw sequence reads from the five studies comprising the secondary analysis, a total of 8,761,462 reads were retained. The number of reads per sample ranged from 10,441 to 159,388 with a median of 35,967. A total of 15,005 ASVs were identified across the samples of all studies. Per sample unique ASVs averaged 272. Although the Fl and VI transect samples had on average greater sequencing depth and ASV count per sample than the studies in the secondary analysis, the relationship between sequencing depth and observed ASVs does not appear to have been saturated in either case (Figure S4).

Because two studies in the analysis used the 806R primer while the other three used the 806RB primer, we evaluated the impact of this difference on the study. The group dispersions between the two primer sets are not significantly different (permutest; F=2.11, p=0.14), indicating that primer choice did not significantly contribute to community variability. While DNA extraction methods were generally similar across studies, this was not similarly tested as a factor because only two studies shared the same method.

3.6 Distance-Decay relationship

In order to investigate the impact of geographic distance on the microbial community, a geographic distance matrix was generated using the samples' physical location and compared to the Aitchison distances calculated between microbial communities (a measure of dissimilarity). The comparisons of samples within transects at an individual reef spanned <10 m, while individual reefs were 1-3 km apart for VI reefs and 34-279 km for Fl reefs. Mantel tests of these two matrices revealed that physical distance was significantly correlated with the Aitchison distance between samples within each (Fl and VI) reef system, with a stronger relationship in Florida (r=0.45, p=0.001 for Fl and r=0.10, p=0.004 for VI) (Table 3). There was also a relationship between geographic distances and microbial communities for both Fl and VI reef systems combined, with comparisons spanning 0-1,978 km (r=0.34, p=0.001).

Incorporating the additional reef water microbiomes from the secondary analysis provided us with the opportunity to extend this study to thousands of kilometers. Distances between reefs within a reef system ranged from 1-2775 km while reef systems were separated by

421-16,874 km. The Aitchison distance between these secondary analysis samples showed a significant relationship to geographic distance (r=0.28, p=0.001)(Table 3).

In the Fl and VI study, distance-decay plots showed no relationship between geographic distance and microbial community similarity within reefs (at the transect level) at the scale of meters (p>0.05) (Figure 3A). However, at the scale of kilometers, there was a negative relationship between geographic distance and microbial community similarity for the Fl and VI study (Figure 3B) as well as for the more expansive secondary analysis (Figure 3C), with microbial communities becoming less similar with increasing distance (R^2 = 0.11, p<0.001; R^2 = 0.08, p<0.001, respectively). The slope of the negative distance-decay relationship increased in magnitude as the geographic extent of the samples increased (Table 3, Figure 3C).

3.7 Drivers of distance-decay relationship

In the secondary analysis, the effects of collection depth, temperature, and reef type were examined as potential drivers of community similarity. A PERMANOVA assessing the marginal impacts of these abiotic factors as well the effect of study found that study accounted for the most variation (R^2=0.14, Pseudo-F=9.32, p=0.001), distantly followed by reef type (R^2=0.03, Pseudo-F=2.06, p=0.001), collection depth (R^2=0.017, Pseudo-F=4.64, p=0.001), and finally temperature (R^2=0.017, Pseudo-F=4.49, p=0.001). Despite explaining the least amount of variation, difference in temperature was significantly correlated with Aitchison distance (Mantel: r=0.18, p=0.001), meaning communities that were more different (distant) in temperature were also more dissimilar. A similar correlation for depth was not found (Mantel: r=-0.00035, p=0.49).

4 Discussion

In this study we used a nested distance sampling design to examine how reef seawater microbiomes vary at multiple spatial scales, including within reefs, between individual reefs in a reef system as well as across northern Caribbean (Fl and VI) forereef systems. Overall, we found that individual reef and reef system-related features had the largest influence on microbial community diversity and composition. No differences in microbial community diversity or composition were detected within different locations on individual reefs, and there was a weak correlation with the benthic substrate underlying the sample. Despite the large number of observed microbial taxa in the transect-based study, just over one hundred of the most abundant were identified as differentially abundant between reefs, suggesting that these abundant taxa may be useful indicators of reef change. We also used data from five previous studies in a secondary analysis to understand the biogeography of more distant reef seawater microbiomes, and this revealed that microbial communities are more distinct with increasing geographic distance.

4.1 Microbial communities differentiate by individual reef (>1 km) and reef system (>100 km), but not within each reef (<10 m)

Counter to expectations, the benthic substrate did not have a strong influence on the composition of the seawater microbiome. Although there is evidence that substrate type influences the surrounding seawater microbial community (Schöttner et al. 2012, Tout et al. 2014), we did not find a correlation in these data, nor did we find differentially abundant microbial taxa between substrate types. It is likely that differential hydrodynamic

399 conditions, which have not yet been measured in the context of coral reef benthic-pelagic
400 microbial interactions, may play a role in these differential results. As such, additional
401 research concerning benthic-pelagic exchange on coral reefs is needed to understand the
402 impact of substrate seawater microorganisms close to the reef surface.

403 All six reefs were distinguishable in terms of the composition of their reef water microbial
404 communities, despite the three VI reefs being separated by only 1 to 3 kilometers. These
405 results align with our expectation that microbes in the water column above reefs would
406 display reef-specific signatures because marine microbial communities are reflective of
407 their physical and chemical environment (Azam & Malfatti 2007, Kelly et al. 2018). While
408 benthic substrate had a weak influence on the seawater microbes, there were other
409 influences that varied between the reefs, including season, time of sampling, depth, and
410 reef type, but these generally were consistent within VI or Fl and therefore difficult to
411 statistically examine. Indeed, in our secondary analysis, which included a larger number of
412 reef sites and more geographic locations, temperature, depth, and reef type were small but
413 significant contributors to the community variation. It must be noted, however, that all
414 variables examined were highly confounded by the specific sampling scheme of each study.
415 For example, Becker et al. (2020) only sampled reef seawater from four forereefs in the
416 Virgin Islands at 0.3 meter depth, and Weber et al. (2020) contains a disproportionate
417 number of samples at cooler temperatures (67 out of 82 samples below average
418 temperature of the secondary analysis), all from Cuban reefs. In addition to our study,
419 other studies of the Indian Ocean and Northwestern Hawaiian islands have shown strong
420 microbial biogeographic signatures (Jeffries et al. 2015, Salerno et al. 2016).
421 Hydrodynamics likely plays a major role in explaining some of the biogeographical

422 portioning between reef water microbial communities because it impacts distribution and
423 transport of nutrients and facilitates dispersal of pelagic microorganisms. Previous studies
424 have suggested links between water masses and microbial community composition (Varela
425 et al. 2008, Galand et al. 2010, Jeffries et al. 2015). Comparison of reef water microbial
426 communities within and between hydrographic regimes and current systems could help us
427 better understand this influence.

428 Surprisingly, we did not identify consistent differences in the microbial communities
429 within each reef (at the transect level). Samples collected within meters of each other were
430 indistinguishable in the VI and Fl, but trends in beta dispersion did suggest some within-
431 reef variability, indicating differences in beta diversity among one or more groups
432 (Anderson et al. 2006). Dispersion was greatest at Grecian and lowest at Dry Tortugas, both
433 Florida reefs. Grecian reef is located in the Upper Keys, which on average has elevated
434 nutrients, organic carbon, and turbidity compared to the Lower Keys (Lirman & Fong
435 2007). In contrast, the Dry Tortugas reef is located within a marine protected zone (US
436 National Park), is more distant from the shore, was the deepest reef sampled (60 feet), and
437 was the only reef not experiencing active outbreaks of Stony Coral Tissue Loss Disease at
438 the time of sampling. These factors could contribute to the relative homogeneity of the
439 samples collected at Dry Tortugas. Reef depth and coastal influence may be among the
440 regional geographic conditions that influence the variability of microbial communities in
441 reef associated seawater (Weber et al. 2020, Frade et al. 2020).

4.2 Community similarity decays with distance beginning at the kilometer scale

The distance-decay of community similarity - a widely studied relationship in ecology (Soininen et al. 2007) - quantifies the decrease in community similarity with increased geographic distance. Typically, communities that are closer geographically are also more similar to each other compositionally (Soininen et al. 2007). One mechanism that drives this relationship is spatial structuring, where locations closer together have more similar environments, thus leading to selection of more similar communities. In the absence of selection (e.g., in a homogeneous environment), neutral drift interacts with dispersal limitation to differentiate communities over space (Soininen et al. 2007, Hanson et al. 2012). These mechanisms represent two hypotheses for what drives species distributions: environmental selection and historical contingency (Martiny et al. 2006). A number of studies have examined the distance-decay relationship in both soil and marine environments and found that microorganisms tend to display a weaker (i.e., less negative) relationship compared to macroorganisms on the same scale, a phenomenon attributed to the small size and large populations of microorganisms leading to greater dispersal (Martiny et al. 2006, Green & Bohannan 2006, Meyer et al. 2018).

The sampling pattern in this study allowed us to assess this biogeographic pattern in the context of coral reef associated seawater. Within transects in a reef (<10 m scale), no distance-decay relationship was found, likely due to high mixing rates on the reef. However, there was a significant correlation between community similarity and geographic distance beginning at the reef level (1 km scale), and the steepness of the relationship increased with an increase in geographic extent (10,000 km scale) (Table 3). Differences in

464 the steepness and strength of correlation of the distance-decay relationship may reflect
465 different mechanisms driving the decay at multiple spatial scales (Martiny et al. 2011). The
466 larger correlation and slope observed in the Fl reefs compared to VI reefs may reflect the
467 orientation of Fl reefs in a north-south line along the Florida current, with the most distant
468 reef upstream of the two closer reefs, while the VI reefs were closer together and not
469 oriented in relation to the surrounding Caribbean current. The autocorrelation of distance
470 and environmental similarity (Lirman & Fong 2007) along the Florida Reef Tract likely
471 drives the stronger correlation compared to the VI reefs. The steep slope and weaker
472 correlation found in the secondary analysis likely reflects historical factors such as
473 dispersal limitation and drift as distant reefs recruit from different metapopulations
474 (Hellweger et al. 2014, Clark et al. 2021).

475 ## 4.3 Abundant taxa are most variable and more likely to differentiate individual
476 ## reefs

477 Although we recovered a total of over twenty thousand microbial ASVs from the Fl and VI
478 transect sampling, the vast majority were rare and samples were dominated by just over
479 one thousand highly abundant taxa. This is a common occurance in microbiome
480 sequencing, especially with the advent of high throughput deep sequencing, and there is
481 debate over the importance of these rare taxa (McMurdie & Holmes 2014, Cao et al. 2021).
482 Abundant taxa tend to be the most prevalent, and in this study, also displayed the highest
483 variance in their relative abundance values between samples. The ASVs that were
484 identified as differentially abundant between reefs were also among the most abundant
485 taxa (Figure S3). Glasl et al. (2019) found that the relative abundances of indicator taxa in

coral reef seawater that best predict environmental conditions range from 0.5-20%, and those taxa were also prevalent throughout that study's sampling period. Overlap in taxonomic assignment between these indicator taxa and differentially abundant taxa found within this study include *Synechococcus*, *Prochlorococcus*, Rhodobacteraceae, unclassified Alphaproteobacteria, and others. Because extremely rare taxa can be difficult to reliably detect, more frequent, shallower sequencing may be more important for capturing the salient variability of a reef.

4.4 Caveats

Although the nested design surveyed seawater microbial communities across multiple spatial scales, the temporal scale of seawater variability was not considered in this study. The Florida samples were taken in June, while the St. Thomas samples were taken in February. Each reef was only sampled once and not throughout the day. Seasonal as well as diurnal/tidal cycles in reef seawater microbial communities are well documented (Weber & Apprill 2020, Glasl et al. 2020, Frade et al. 2020, Becker et al. 2020). The differences between the Florida and VI microbial communities may be in part due to the different seasons in which the samples were taken. Temporal differences in sampling can make direct comparisons between distant reefs challenging, even within the same study (Weber & Apprill 2020). While microbial communities are sensitive to environmental conditions, coral reef seawater remains distinct from other seawater habitats (Becker et al. 2020) and variation of the microbial community is better explained by reef-level environmental parameters rather than seasonal differences (Glasl et al. 2019). Within a reef, repeated sampling throughout the calendar year may be needed to establish the baseline variability.

In contrast to many other seawater microbiome studies, including those in the secondary analysis, this study filtered a small volume of seawater (60 mL) for each sample rather than the more typical 1-2 L. Weber et al. (2019) directly examined the effect of sampling volume and found that while species richness modestly increased with larger volumes due to sampling rare taxa, beta diversity and overall community composition was not influenced by sampling volume.

4.5 Conclusions

In conclusion, we suggest that due to reef and reef system-level influences, the development of reef water microbiome monitoring criteria may need to be regionally tailored. We found that the community composition of reef seawater microbiomes are distinguishable even when reefs are a few kilometers apart and that there can be large differences in the beta dispersion within a reef. Detecting a shift in the community composition as a whole will necessitate an understanding of each reef region's variability. Individual reefs within a reef system may also be experiencing different regional stressors, such as varying degrees of anthropogenic influence. Such differences may be reflected in both the baseline microbial community composition and variability as reef conditions change. Additionally, microbial taxa common between reef regions are vastly outnumbered by taxa that are unique, making it difficult to develop a generalized database of indicator microbial taxa for reef environental conditions. Overall, we found that the seawater microbial communities of reefs closer together are more similar, and that the local oceanographic conditions which differentiate these communities are important to investigate.

Acknowledgements

We thank Carolyn Miller, Joseph Townsend, Kathryn Cobleigh, Sonora Meiling, Kelsey Beavers, Erinn Muller, Abigail Clark and the captain and crew of the M/V Alucia for field assistance and Illumina Inc. for technical support. Samples were collected under Florida Keys National Marine Sanctuary permit #FKNMS-2019-046, US National Park Service permit DRTO-2019-SCI-0010 and and USVI territory permit DFW19057U. This project was supported by an OceanX sponsored cruise on the M/V Alucia, with additional funding provided to A.A. by the Rockefeller Philanthropy Advisors, Dalio Foundation, and other generous donors of the Oceans 5 project and NSF OCE-1938147 and NSF OCE-1928761 awards.

References

Amend AS, Oliver TA, Amaral-Zettler LA, Boetius A, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM, Martiny AC, Ramette A, Zinger L, Sogin ML, Martiny JBH (2012) Macroecological Patterns of Marine Bacteria on a Global Scale. Journal of Biogeography 40:800–811.

Anderson MJ, Ellingsen KE, McArdle BH (2006) Multivariate Dispersion as a Measure of Beta Diversity. Ecology Letters 9:683–693.

Apprill A, Holm H, Santoro AE, Becker C, Neave M, Hughen K, Donà AR, Aeby G, Work T, Weber L, McNally S (2021) Microbial Ecology of Coral-Dominated Reefs in the Federated States of Micronesia. Aquatic Microbial Ecology 86:115–136.

Apprill A, McNally S, Parsons R, Weber L (2015) Minor Revision to V4 Region SSU rRNA 806R Gene Primer Greatly Increases Detection of Sar11 Bacterioplankton. Aquatic Microbial Ecology 75:129–137.

Azam F, Malfatti F (2007) Microbial Structuring of Marine Ecosystems. Nature Reviews Microbiology 5:782–791.

Becker CC, Weber L, Suca JJ, Llopiz JK, Mooney TA, Apprill A (2020) Microbial and Nutrient Dynamics in Mangrove, Reef, and Seagrass Waters over Tidal and Diurnal Time Scales. Aquatic Microbial Ecology 85:101–119.

560 Bourne DG, Morrow KM, Webster NS (2016) Insights into the Coral Microbiome:
561 Underpinning the Health and Resilience of Reef Ecosystems. Annual Review of
562 Microbiology 70:317–340.

563 Brandt ME, Ennis RS, Meiling SS, Townsend J, Cobleigh K, Glahn A, Quetel J, Brandtneris V,
564 Henderson LM, Smith TB (2021) The Emergence and Initial Impact of Stony Coral Tissue
565 Loss Disease (SCTLD) in the United States Virgin Islands. Frontiers in Marine Science
566 8:1105.

567 Bryan D Martin, Daniela Witten, Amy D Willis (2021) Corncob: Count Regression for
568 Correlated Observations with the Beta-Binomial.

569 Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) Dada2: High-
570 resolution Sample Inference from Illumina Amplicon Data. Nature Methods 13:581–583.

571 Cao Q, Sun X, Rajesh K, Chalasani N, Gelow K, Katz B, Shah VH, Sanyal AJ, Smirnova E (2021)
572 Effects of Rare Microbiome Taxa Filtering on Statistical Analysis. Front Microbiol 11.

573 Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N,
574 Knight R (2011) Global Patterns of 16S rRNA Diversity at a Depth of Millions of Sequences
575 Per Sample. PNAS 108:4516–4522.

576 Clark DR, Underwood GJC, McGenity TJ, Dumbrell AJ (2021) What Drives Study-Dependent
577 Differences in Distance–Decay Relationships of Microbial Communities? Global Ecology and
578 Biogeography 30:811–825.

579 Corcoll N, Österlund T, Sinclair L, Eiler A, Kristiansson E, Backhaus T, Eriksson KM (2017)
580 Comparison of Four DNA Extraction Methods for Comprehensive Assessment of 16S rRNA

581 Bacterial Diversity in Marine Biofilms Using High-Throughput Sequencing. FEMS
582 Microbiology Letters 364.

583 Davis NM, Proctor DM, Holmes SP, Relman DA, Callahan BJ (2018) Simple Statistical
584 Identification and Removal of Contaminant Sequences in Marker-Gene and Metagenomics
585 Data. Microbiome 6:226.

586 Fox MD, Carter AL, Edwards CB, Takeshita Y, Johnson MD, Petrovic V, Amir CG, Sala E,
587 Sandin SA, Smith JE (2019) Limited Coral Mortality Following Acute Thermal Stress and
588 Widespread Bleaching on Palmyra Atoll, Central Pacific. Coral Reefs 38:701–712.

589 Frade PR, Glasl B, Matthews SA, Mellin C, Serrão EA, Wolfe K, Mumby PJ, Webster NS,
590 Bourne DG (2020) Spatial Patterns of Microbial Communities Across Surface Waters of the
591 Great Barrier Reef. Communications biology 3:442.

592 Galand PE, Potvin M, Casamayor EO, Lovejoy C (2010) Hydrography Shapes Bacterial
593 Biogeography of the Deep Arctic Ocean. ISME J 4:564–576.

594 Glasl B, Bourne DG, Frade PR, Thomas T, Schaffelke B, Webster NS (2019) Microbial
595 Indicators of Environmental Perturbations in Coral Reef Ecosystems. 1–13.

596 Glasl B, Robbins S, Frade PR, Marangon E, Laffy PW, Bourne DG, Webster NS (2020)
597 Comparative Genome-Centric Analysis Reveals Seasonal Variation in the Function of Coral
598 Reef Microbiomes. The ISME Journal 14:1435–1450.

599 Glasl B, Webster NS, Bourne DG (2017) Microbial Indicators as a Diagnostic Tool for
600 Assessing Water Quality and Climate Stress in Coral Reef Ecosystems. Marine Biology
601 164:1–18.

602 Gloor GB, Macklaim JM, Pawlowsky-Glahn V, Egozcue JJ (2017) Microbiome Datasets Are
603 Compositional: And This Is Not Optional. Front Microbiol 8.

604 Green J, Bohannan BJM (2006) Spatial Scaling of Microbial Biodiversity. Trends in Ecology
605 & Evolution 21:501–507.

606 Haas AF, Fairoz MFM, Kelly LW, Nelson CE, Dinsdale EA, Edwards RA, Giles S, Hatay M,
607 Hisakawa N, Knowles B, Lim YW, Maughan H, Pantos O, Roach TNF, Sanchez SE, Silveira CB,
608 Sandin S, Smith JE, Rohwer F (2016) Global Microbialization of Coral Reefs. Nature
609 Microbiology 1:1–7.

610 Haas AF, Nelson CE, Kelly LW, Carlson CA, Rohwer F, Leichter JJ, Wyatt A, Smith JE (2011)
611 Effects of Coral Reef Benthic Primary Producers on Dissolved Organic Carbon and
612 Microbial Activity. PLOS ONE 6:e27973.

613 Hanson CA, Fuhrman JA, Horner-Devine MC, Martiny JBH (2012) Beyond Biogeographic
614 Patterns: Processes Shaping the Microbial Landscape. Nature Reviews Microbiology
615 10:497–506.

616 Hellweger FL, Sebille E van, Fredrick ND (2014) Biogeographic Patterns in Ocean Microbes
617 Emerge in a Neutral Agent-Based Model. Science 345:1346–1349.

618 Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian
619 Journal of Statistics 6:65–70.

620 Hughes TP, Barnes ML, Bellwood DR, Cinner JE, Cumming GS, Jackson JBC, Kleypas J, van de
621 Leemput IA, Lough JM, Morrison TH, Palumbi SR, van Nes EH, Scheffer M (2017) Coral
622 Reefs in the Anthropocene. Nature 546:82–90.

623 Jeffries TC, Ostrowski M, Williams RB, Xie C, Jensen RM, Grzymski JJ, Senstius SJ, Givskov M,
624 Hoeke R, Philip GK, Neches RY, Drautz-Moses DI, Chénard C, Paulsen IT, Lauro FM (2015)
625 Spatially Extensive Microbial Biogeography of the Indian Ocean Provides Insights into the
626 Unique Community Structure of a Pristine Coral Atoll. Scientific Reports 5:15383.

627 John R Wallace (2012) Imap.

628 Kelly LW, Haas AF, Nelson CE (2018) Ecosystem Microbiology of Coral Reefs: Linking
629 Genomic, Metabolomic, and Biogeochemical Dynamics from Animal Symbioses to
630 Reefscape Processes. mSystems 3.

631 Kelly LW, Williams GJ, Barott KL, Carlson CA, Dinsdale EA, Edwards RA, Haas AF, Haynes M,
632 Lim YW, McDole T, Nelson CE, Sala E, Sandin SA, Smith JE, Vermeij MJA, Youle M, Rohwer F
633 (2014) Local Genomic Adaptation of Coral Reef-Associated Microbiomes to Gradients of
634 Natural Variability and Anthropogenic Stressors. PNAS 111:10227–10232.

635 Knowlton N, Rohwer F (2003) Multispecies Microbial Mutualisms on Coral Reefs: The Host
636 as a Habitat. The American naturalist 162:S51–62.

637 Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a Dual-
638 Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on
639 the MiSeq Illumina Sequencing Platform. Appl Environ Microbiol 79:5112–5120.

640 Lirman D, Fong P (2007) Is Proximity to Land-Based Sources of Coral Stressors an
641 Appropriate Measure of Risk to Coral Reefs? An Example from the Florida Reef Tract.
642 Marine Pollution Bulletin 54:779–791.

643 Liu M, Xue Y, Yang J (2019) Rare Plankton Subcommunities Are Far More Affected by DNA
644 Extraction Kits Than Abundant Plankton. Frontiers in Microbiology 10:454.

645 Marcelino VR, van Oppen MJ, Verbruggen H (2018) Highly Structured Prokaryote
646 Communities Exist Within the Skeleton of Coral Colonies. The ISME Journal 12:300–303.

647 Martin M (2011) Cutadapt Removes Adapter Sequences from High-Throughput Sequencing
648 Reads. EMBnetjournal 17:10–12.

649 Martiny JBH, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine
650 MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Øvre\aas L, Reysenbach A-L, Smith
651 VH, Staley JT (2006) Microbial Biogeography: Putting Microorganisms on the Map. Nature
652 Reviews Microbiology 4:102–112.

653 Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC (2011) Drivers of Bacterial β-
654 Diversity Depend on Spatial Scale. PNAS 108:7850–7854.

655 McDevitt-Irwin JM, Baum JK, Garren M, Vega Thurber RL (2017) Responses of Coral-
656 Associated Bacterial Communities to Local and Global Stressors. Frontiers in Marine
657 Science 4:289–16.

658 McMurdie PJ, Holmes S (2013) Phyloseq: An R Package for Reproducible Interactive
659 Analysis and Graphics of Microbiome Census Data. PLoS ONE 8:e61217.

660 McMurdie PJ, Holmes S (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is
661 Inadmissible. PLoS Computational Biology 10:e1003531–12.

662 Meyer KM, Memiaghe H, Korte L, Kenfack D, Alonso A, Bohannan BJM (2018) Why Do
663 Microbes Exhibit Weak Biogeographic Patterns? ISME J 12:1404–1413.

664 Neave MJ, Rachmawati R, Xun L, Michell CT, Bourne DG, Apprill A, Voolstra CR (2017)
665 Differential Specificity Between Closely Related Corals and Abundant Endozoicomonas
666 Endosymbionts Across Global Scales. The ISME Journal 11:186–200.

667 Nelson CE, Alldredge AL, McCliment EA, Amaral-Zettler LA, Carlson CA (2011) Depleted
668 Dissolved Organic Carbon and Distinct Bacterial Communities in the Water Column of a
669 Rapid-Flushing Coral Reef Ecosystem. The ISME Journal 5:1374–1387.

670 Nelson CE, Goldberg SJ, Wegley Kelly L, Haas AF, Smith JE, Rohwer F, Carlson CA (2013)
671 Coral and Macroalgal Exudates Vary in Neutral Sugar Composition and Differentially Enrich
672 Reef Bacterioplankton Lineages. ISME J 7:962–979.

673 Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB,
674 Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2020) Vegan: Community
675 Ecology Package.

676 Palarea-Albaladejo J, Martín-Fernández JA (2015) zCompositions — R Package for
677 Multivariate Imputation of Left-Censored Data Under a Compositional Approach.
678 Chemometrics and Intelligent Laboratory Systems 143:85–96.

679 Parada AE, Needham DM, Fuhrman JA (2016) Every Base Matters: Assessing Small Subunit
680 rRNA Primers for Marine Microbiomes with Mock Communities, Time Series and Global
681 Field Samples. Environmental Microbiology 18:1403–1414.

682 Peixoto RS, Rosado PM, Leite DC de A, Rosado AS, Bourne DG (2017) Beneficial
683 Microorganisms for Corals (BMC): Proposed Mechanisms for Coral Health and Resilience.
684 Frontiers in Microbiology 8:100–16.

685 Precht WF, Gintert BE, Robbart ML, Fura R, van Woesik R (2016) Unprecedented Disease-
686 Related Coral Mortality in Southeastern Florida. Scientific Reports 6:31374.

687 Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO (2007) SILVA: A
688 Comprehensive Online Resource for Quality Checked and Aligned Ribosomal RNA Sequence
689 Data Compatible with ARB. Nucleic Acids Research 35:7188–7196.

690 Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I (2007) The Role of
691 Microorganisms in Coral Health, Disease and Evolution. Nat Rev Microbiol 5:355–362.

692 RStudio Team (2019) RStudio: Integrated Development Environment for R. RStudio, Inc,
693 Boston, MA.

694 Salerno JL, Bowen BW, Rappé MS (2016) Biogeography of Planktonic and Coral-Associated
695 Microorganisms Across the Hawaiian Archipelago. FEMS Microbiology Ecology 92:fiw109.

696 Santoro AE, Casciotti KL, Francis CA (2010) Activity, Abundance and Diversity of Nitrifying
697 Archaea and Bacteria in the Central California Current. Environmental Microbiology
698 12:1989–2006.

699 Schöttner S, Wild C, Hoffmann F, Boetius A, Ramette A (2012) Spatial Scales of Bacterial
700 Diversity in Cold-Water Coral Reef Ecosystems. PLoS ONE 7:e32093.

701 Soininen J, McDonald R, Hillebrand H (2007) The Distance Decay of Similarity in Ecological
702 Communities. Ecography 30:3–12.

703 Sweet MJ, Croquer A, Bythell JC (2010) Temporal and Spatial Patterns in Waterborne
704 Bacterial Communities of an Island Reef System. Aquatic Microbial Ecology 61:1–11.

705 Tout J, Jeffries TC, Webster NS, Stocker R, Ralph PJ, Seymour JR (2014) Variability in
706 Microbial Community Composition and Function Between Different Niches Within a Coral
707 Reef. Microbial Ecology 67:540–552.

708 Urakawa H, Martens-Habbena W, Stahl DA (2010) High Abundance of Ammonia-Oxidizing
709 Archaea in Coastal Waters, Determined Using a Modified DNA Extraction Method. Appl
710 Environ Microbiol 76:2129–2135.

711 van Duyl FC, Gast GJ (2001) Linkage of Small-Scale Spatial Variations in DOC, Inorganic
712 Nutrients and Bacterioplankton Growth with Different Coral Reef Water Types. Aquatic
713 Microbial Ecology 24:17–26.

714 Varela MM, Van Aken HM, Sintes E, Herndl GJ (2008) Latitudinal Trends of Crenarchaeota
715 and Bacteria in the Meso- and Bathypelagic Water Masses of the Eastern North Atlantic.
716 Environmental Microbiology 10:110–124.

717 Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian Classifier for Rapid
718 Assignment of rRNA Sequences into the New Bacterial Taxonomy. Appl Environ Microbiol
719 73:5261–5267.

720 Weber L, Apprill A (2020) Diel, Daily, and Spatial Variation of Coral Reef Seawater
721 Microbial Communities. PLOS ONE 15:e0229442.

722 Weber L, Gonzalez-Díaz P, Armenteros M, Apprill A (2019) The Coral Ecosphere: A Unique
723 Coral Reef Habitat That Fosters Coral–Microbial Interactions. Limnology and Oceanography
724 64:2373–2388.

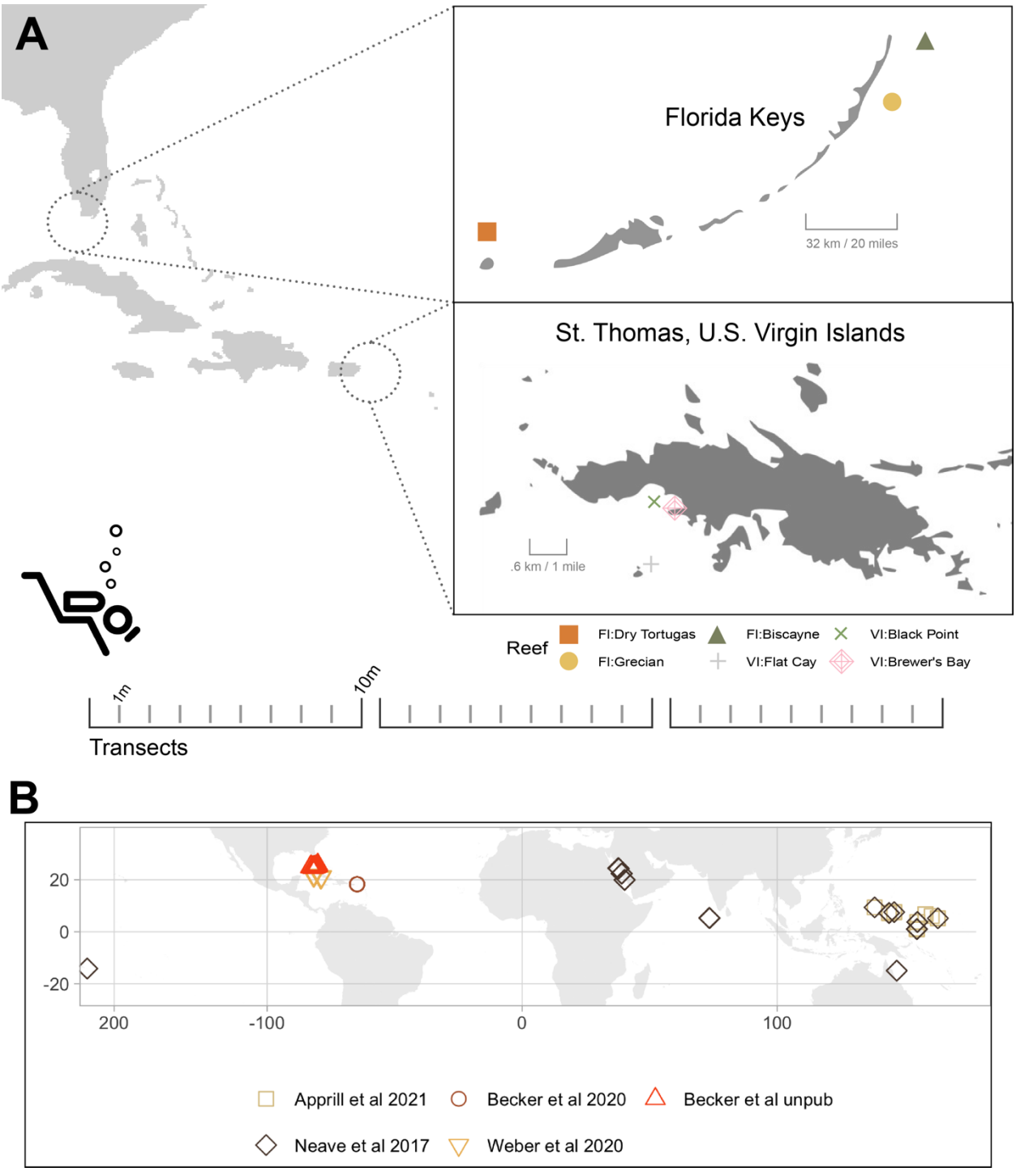
725 Weber L, Gonzalez-Díaz P, Armenteros M, Ferrer VM, Bretos F, Bartels E, Santoro AE,
726 Apprill A (2020) Microbial Signatures of Protected and Impacted Northern Caribbean
727 Reefs: Changes from Cuba to the Florida Keys. Environmental Microbiology 22:499–519.

728 Wickham H (2016) Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

729

730 ## Figures and tables

731



732

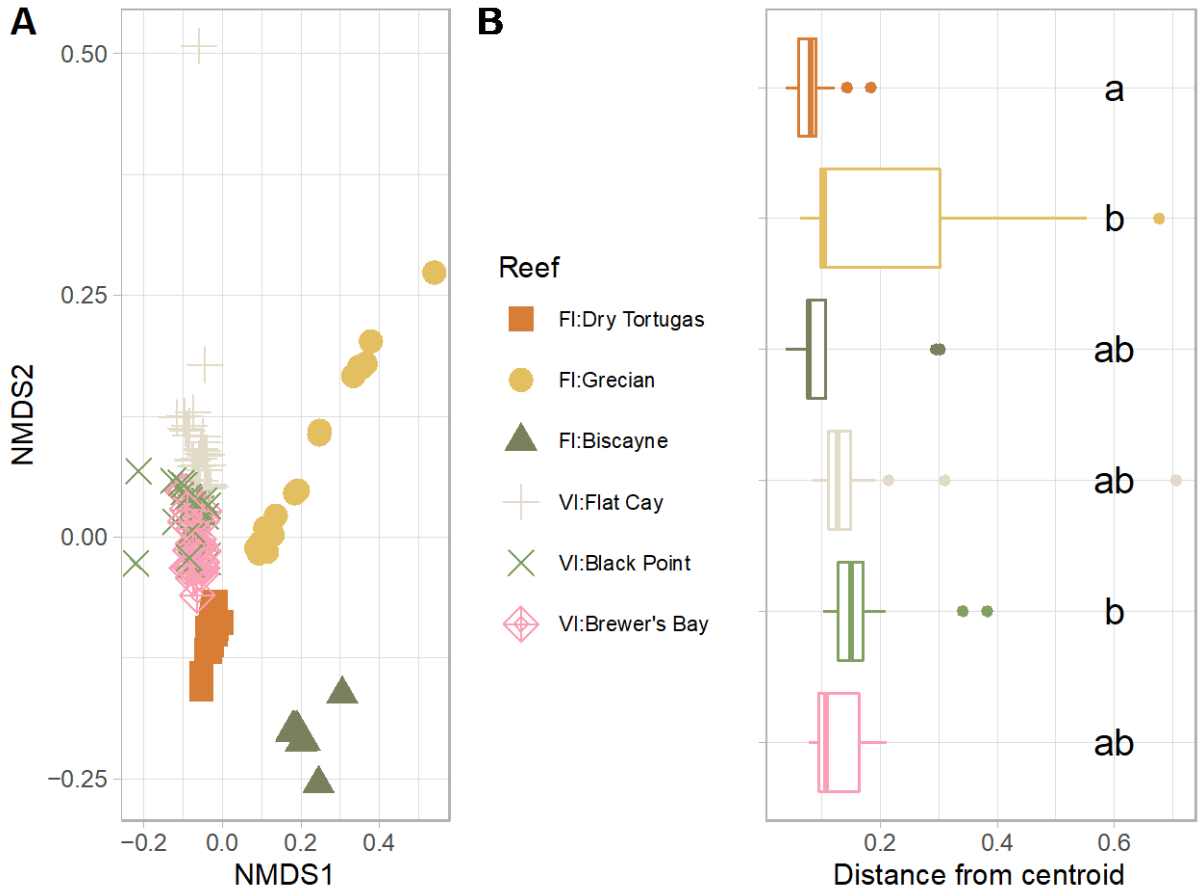
733 **Figure 1. Locations of reefs in the present study and the secondary analysis** A.

734 Sampling locations and method for the transect-based study. Six reefs were surveyed

735 across two reef systems. At each reef, three 10 meter transects were laid and divers used a

736 syringe to sample seawater just above the benthos at 1 meter intervals. B. The locations of
737 the five studies in the secondary analysis span many major reef systems across the globe.
738 Collections for all studies were performed by the same lab group using nearly identical
739 techniques in the field and in the lab.

740



741

742 **Figure 2. Reefs have distinct microbial communities and variable group dispersions**

743 A. An NMDS ordination using Bray-Curtis dissimilarities. Tries=20, 2D stress=0.1006136. B.

744 The group dispersions calculated as distance to centroid using the Bray-Curtis dissimilarity

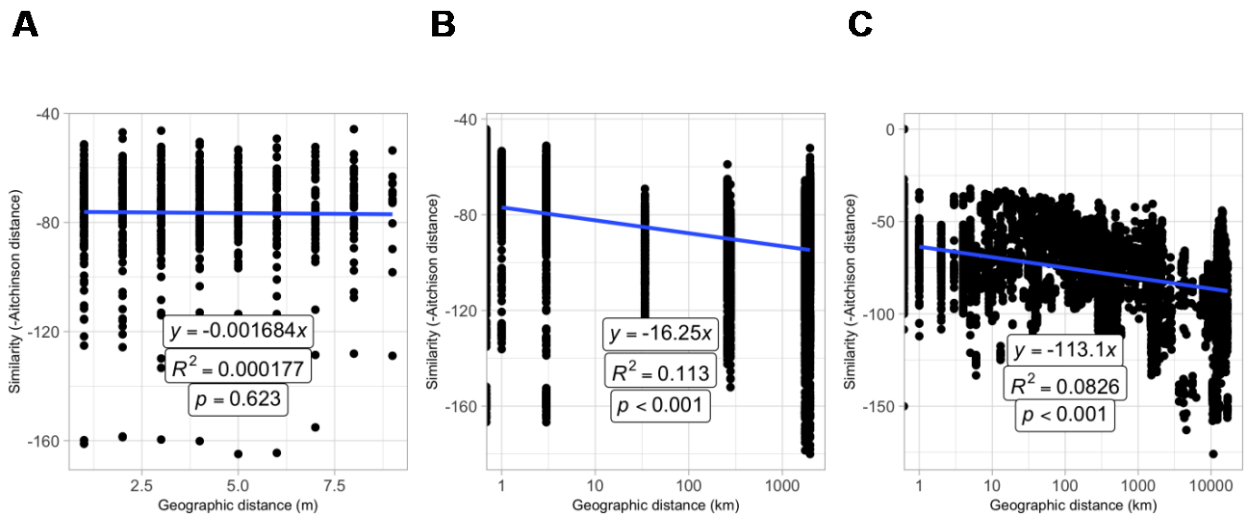
745 metric. The vertical line in each boxplot indicates the median, the box hinges represent the

746 first and third quartiles while the whiskers extend to 1.5*IQR (interquartile range). Colors

747 correspond to the individual reef and letters indicate significance groups (pairwise t test;

748 p-adjusted<0.05).

749



750

751 **Figure 3. Reef seawater communities exibit a distance-decay relationship at the scale**
752 **of 100 km but not at <10 m scale** A. The pairwise geographic distances between samples
753 within each transect in the Fl/VI-based study is plotted on the x-axis and the corresponding
754 Aitchison distances are plotted on the y axis. Between transect distances are not known
755 and therefore not included. B. Distance-decay plot of all pairwise distances (i.e., not just
756 within a transect) between samples collected in the Fl/VI-based study. C. Distance-decay
757 plot using the samples from the five studies included in the secondary analysis.

758

759

Table 1: Florida reef tract (Fl) and St. Thomas (VI) Sampling locations

Reef System	Reef	Transects	Samples	Avg. Depth (m)	GPS	Date
Florida Keys	Dry Tortugas	3	30	17.98	24.722 N 82.828 W	7-Jun-19
	Grecian	3	30	6.71	25.110 N 80.303 W	14-Jun-19
	Biscayne	1	9	5.18	25.386 N 80.162 W	17-Jun-19
St. Thomas, USVI	Flat Cay	3	30	7.12	18.316 N 64.987 W	12-Feb-20
	Black Point	3	30	7.00	18.344 N 64.986 W	12-Feb-20
	Brewer's Bay	3	27	5.92	18.343 N 64.980 W	12-Feb-20

Table 2: Summary of PERMANOVA results based on Bray-Curtis dissimilarties of microbial community abundances grouped at the spatial scales of within reefs, between reefs, and between reef regions

Scope	Df	Sum Sq	R2	Pseudo-F	P
Transect (within reef)	15	8.366559	0.6524673	17.52265	0.412
Residual	140	4.456397	0.3475327		
Total	155	12.822956	1.0000000		
Site (between reefs)	5	8.050737	0.6278378	50.61002	**0.001**
Residual	150	4.772219	0.3721622		
Total	155	12.822956	1.0000000		
Reef system (between reef regions)	1	3.284216	0.2561200	53.02264	**0.001**
Residual	154	9.538740	0.7438800		
Total	155	12.822956	1.0000000		

Note: Df - degrees of freedom; R2 - R-squared value; Pseudo-F values derived from 999 permutations. Values in bold are significant at P<0.05.

Table 3: Distance decay relationships at multiple spatial scales

Scope	Spatial scale	Correlation (Mantel r)	Mantel p value	Slope of linear fit
Within transects	0-9 m	NA*	NA*	-0.002
(Reef system) USVI	0-3 km	0.1009	0.004	-0.008†
(Reef system) Florida Keys	0-279 km	0.4492	0.001	-3.208†
All samples in Fl/VI-based study	0-1,978 km	0.3358	0.001	-16.246†
All samples in meta-analysis	0-16,874 km	0.2840	0.001	-113.128†

* Mantel test not performed because distances exist for samples within transects but not between, therefore matrix had >50% missing values

† Linear fit significant at $p < 0.01$