Supporting Information

The coral ecosphere: a unique coral reef microbial habitat that fosters coral-microbial interactions

Laura Weber[1, 2], Patricia Gonzalez[3], Maickel Armenteros[3], and Amy Apprill[1*]

[1]Marine Chemistry and Geochemistry Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543
[2]Massachusetts Institute of Technology-Woods Hole Oceanographic Institution Joint Program in Oceanography/Applied Ocean Science and Engineering, Biological Oceanography, Cambridge, MA 02139
[3]Centro de Investigaciones Marinas, Universidad de la Habana; 16 #114, Playa, CP11300; La Habana, Cuba; CP 11300
[*]For correspondence: aapprill@whoi.edu

This document includes:
1. Supporting Methods
2. Supporting Figures (S1-S6)
3. Supporting Tables (S1, S2)
4. Additional Supporting Methods: Seawater volume experiment
5. Supporting Results: Seawater volume experiment
6. Supporting Figures: Seawater volume experiment (S7-S12)
7. References

Supporting methods

Cell enumeration using flow cytometry

Flow cytometry samples were analyzed at the University of Hawaii using an Altra flow cytometer (Beckman Coulter Life Sciences, Inc, Indianapolis, IN, USA) with a laser excitation wavelength of 488 nm. Unstained and stained (SybrGreen I, Invitrogen™, Waltham, MA, USA) subsamples of each sample were run on the instrument to estimate the concentration of fluorescent picocyanobacteria (*Prochlorococcus* and *Synechococcus*) and picoeukaryotes (unstained) as well as the concentration of unpigmented (stained) cells, respectively. The abundance of unpigmented cells generally serves as a proxy for heterotrophic bacteria within the sample (Marie et al. 1997). Fluorescence spectra were binned, analyzed, and transformed into count data using FlowJo (v. 6.4.7) (FlowJo, LLC) software. The number of cells per ml of seawater was estimated using the original sample volume (1 ml).

To determine overall trends in picoplankton cell abundances over the distance gradient sampled around each colony, we counted the number of instances when cell counts increased, decreased, or had no trend over the distance gradient surrounding each colony. We did this for each species as well as each picoplankton group. Using this metric, the percentage of colonies that had increasing, decreasing, or no trend in cell abundance was determined by species and picoplankton group.

DNA Extractions

Reef seawater (RSW) DNA was extracted from replicate samples taken at each reef. No duplicate samples were collected for the near-coral seawater (CSW) samples. DNA was extracted from the filters using two different DNA extraction protocols (Santoro et al. 2010; Urakawa et al. 2010). The sucrose–EDTA DNA extraction (Santoro et al. 2010) involves lysing the cells that are retained on the filter using a combination of chemical lysis (Sucrose – EDTA lysis buffer), bead-beating, a proteinase-K (25 µl of 20 mg/ml) (PK Solution, Promega, Madison, WI, USA) digestion, and a column-based separation of the DNA from the remaining cellular material. The original filter used during the first DNA extraction method was preserved at -80° C and used again for a second round of extractions using a modified phenol: chloroform: isoamyl alcohol (25:24:1) DNA extraction method (Urakawa et al., 2010). Purified DNA (yielded from the two different extraction methods) was combined for each sample using the Genomic DNA Clean and Concentrator kit (Zymo Research Corporation, Irvine, CA, USA). DNA concentrations were then quantified using the Qubit 2.0 high sensitivity dsDNA assay (ThermoFisher Scientific, Waltham, MA, USA) and inspected for quality using gel electrophoresis (1% Tris-Borate-Ethylenediaminetetracetic acid (TBE) agarose gel) and the HyperLadder^{TM} 1kb marker (Bioline, London, UK).

Fluidigm amplification

DNA extracts were amplified using Fluidigm microfluidic amplification according to Fluidigm protocols. Prior to the first amplification step, 2 ng (1 µl) of each DNA extract was added to 4 µl of a PCR mastermix (Roche High Fidelity Fast Start Kit) in a PCR plate for a total volume of 5 µl. PCR primers were added to a second plate (50 µM each) and diluted to a total volume of 100

µl with the Fluidigm loading reagent and water. Four µl of the sample and 4 µl of the primer were loaded into a primed Fluidigm 48.48 Access Array Integrated Fluidic Circuit (IFC) and the IFC was placed within an AX controller in order to combine the primers with the samples prior to PCR amplification. The Fluidigm Biomark HD PCR machine was used for amplification without imaging. The following amplification steps and cycle numbers were used: 50 °C for 2 minutes (1 cycle); 70 °C for 20 minutes (1 cycle); 95 °C for 10 minutes (1 cycle); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (10 cycles); 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (2 cycles); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (8 cycles), 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (2 cycles); 95 °C for 15 seconds, 55 °C for 30 seconds, and 72 °C for 1 minute (8 cycles); and 95 °C for 15 seconds, 80 °C for 30 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (5 cycles).

After the first amplification, 2 µl of Fluidigm Harvest Buffer was loaded into each sample inlet and the AX controller was used to collect the PCR products for each sample. PCR products were then diluted 1:100 in water and 1 µl of the diluted product was amplified using Illumina linkers and barcodes in 20 µl reactions (15 µl of PCR mastermix, 1 µl of diluted PCR product, and 4 µl of Illumina linker barcodes). The PCR reaction conditions included 95 °C for 10 minutes (1 cycle); 95 °C for 15 seconds, 60 °C for 30 seconds, and 72 °C for 1 minute (15 cycles); and an extension step at 72 °C for 3 minutes.

PCR products were quantified and amplicon regions and expected sizes were confirmed using a Fragment Analyzer (Advanced Analytics, Ames, IA). PCR products were pooled into equal ratios, pools were run on a gel for size selection, and bands of the expected size were extracted (Qiagen gel extraction kit). The size and profiles of the pooled and purified PCR products were then checked using a Bioanalyzer (Agilent).

Microbial community sequencing and analysis

Using the program mothur v.1.36.1 (Schloss et al. 2009), forward and reverse reads were united and the locus-specific forward and reverse primers were removed (*make.contigs)*. Reads with ambiguous positions or exceeding 275 bp in length were removed (*screen.seqs*), reads matching to unknown, mitochondrial, or eukaryotic sequences were identified (*classify.seqs*, method = 'knn') using the Silva database v119 (Quast et al. 2013) as a reference and removed (*remove.lineages*), and UCHIME (Edgar et al. 2011) was used to identify and remove chimeric reads (based on the command '*chimera.uchime'*, reference = self). Reads detected in the DNA extraction and pooling controls were removed from all samples (*remove.seqs)*, mock community samples were removed from the dataset prior to read clustering and analyzed separately, and data was subsampled to 8,500 reads per sample. These quality-filtering and processing steps omitted 14 samples and the remaining sequences were used to complete microbial community analysis.

The sequencing error rate (defined as the sum of mismatches to the reference file/ sum of bases within the query) was 0.0027 and was calculated using the '*seq.error'* command for the mock community samples using mothur (Schloss et al. 2009). Subsequent clustering analysis on the two sequenced mock community DNA samples yielded 21 and 17 MED nodes, matching closely with the 21 bacterial strains used to make the synthetic community.

For within site differential enrichment comparisons using DESeq2, geometric means were calculated separately for each MED node because zeros were present in the count data. The command "DESeq" was run with default parameters and a "local" fit smoothed trend-line to estimate gene dispersion. Cook's distance filtering was not applied because normalized count outliers could not be discerned as a result of low sample sizes. After manual inspection of the normalized counts, we determined that this filtering criteria was too conservative for this dataset and this approach has been taken in other studies (Pepe-Ranney and Hall 2015). These within site comparisons were made between all CSW and reef-depth seawater communities at reefs JR 1, 2, 5, and 6 and statistically significant differences (Wald test, Benjamini-Hochberg corrections) between specific contrasts (i.e. *P. astreodies* CSW vs. reef-depth SW) were determined. The R package 'phyloseq' (McMurdie and Holmes 2013) was used to generate bar plots of the relative abundances of the MED nodes.

Due to the nature of this field data, there are layers of environmental variables that contribute to microbial community similarity between samples. We completed nested PERMANOVA (Adonis) tests using the vegan R package (Oksanen et al. 2017) in order to control the permutations over the samples by specific factors. To complete the nested test for CSW collected from *P. strigosa*, *O. faveolata, M. cavernosa*, and *A. cervicornis*, the factor of sample type (CSW or RSW) was nested within the factor of reef site (e. g. JR 1). To complete the nested PERMANOVA test for CSW collected from *P. astreoides* as well as RSW, the factor of sample type was nested within the factor of reef site that was then nested within the factor of region (either JR or CAN). Lastly, the nested test for all of the CSW and RSW microbial communities sampled within JR was performed by nesting the factor of coral species (e. g. *O. faveolata*, *P. astreoides*, or RSW) within the factor of reef site. The R^2 value determined by these tests indicates the effect size and shows the percentage of variation explained by each specific factor.

Preparation of coral seawater and benthic seawater DNA for shotgun metagenomic sequencing

Seawater DNA was extracted from the four reef seawater metagenome filters using a modified cetyl-trimethylammonium bromide (CTAB)-phenol: chloroform: isoamyl alcohol extraction that was developed from two existing DNA extraction methods (Zhou et al. 1996; William et al. 2004). One half of each filter was exposed to physical, enzymatic, and chemical disruption via 3 freeze-thaw cycles, bead-beating, and proteinase-k (20 mg/ml) (PK Solution, Promega, Madison, WI, USA) and lysozyme (20 mg/ml) (Pierce™ Lysozyme, Thermo Scientific) digestions. CTAB, an effective surfactant used for purifying DNA in the presence of polysaccharides (Clarke 2009), was added to the sample, followed by a phenol: chloroform (24:1), phenol: chloroform: isoamyl alcohol (25:24:1), phenol: chloroform (24:1) rinsing series. DNA was precipitated using molecular grade isoproponal overnight at -20 °C and the DNA pellet was rinsed with 70% ethanol twice before it was eluted into 50 µl of TE buffer (10 mM Tris-Cl, pH 7.5; 1 mM EDTA). DNA concentrations for these samples were quantified and screened for quality using by using gel electrophoresis prior to sequencing.

After sequencing, the fastq files were demultiplexed, and library adaptors were trimmed from the 3' ends of the reads. Overall, 92,699,608 paired-end reads were generated with an average read number of 18,539,922 (+/- 9,882,964) paired-end reads per sample. The total number of paired-end reads sequenced in reef seawater samples were as follows: 28,934,702 reads for site JR 2; 35, 933, 406 for site JR 4; 27, 444, 129 reads for site JR 5; and 32, 748, 216

reads for site JR 6. The total number of paired-end reads obtained from CSW were as follows: 30, 095, 047 reads for *P. astreodies*; 7, 914, 591 reads for *P. strigosa*; 15, 345, 580 reads for *O. faveolata*, 11, 563, 389 reads for *M. cavernosa*, and 27, 781, 001 reads *for A. cervicornis*. DNA fragment size for the pooled *A. cervicornis* seawater samples ranged from 280-700 bp and fragment sizes for the other CSW samples ranged between 80-600 bp in length.

BBTools (Bushnell 2016) was used to quality-filter and prepare the raw metagenomic reads for functional analysis. Remnant sequencing adaptors were removed from the raw forward and reverse reads for each sample using *bbduk.sh*, the BBtools reference adaptors file (adapter.fa), and the following parameters: ktrim=r k=23 mink=11 hdist=1 tpe tbo (Bushnell 2016). Following removal of the adaptor sequences, the *bbduk.sh* script was implemented again to quality-trim the forward and reverse reads using the Phred algorithm and a Q score of 10 (qtrim=rl trimq=10).

After quality-filtering, the Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies (FMAP) (Kim et al. 2016) was used to determine functional differences between the CSW and RSW metagenomes. The example script (available here: https://qbrc.swmed.edu/FMAP/) was modified to suite this specific comparison. Reef seawater had a higher assignment of quality-filtered reads to KEGG Orthologies (KOs) compared to pooled coral seawater samples, ranging from 6.58 – 11.17 % of unmerged reads with an average read assignment of 8.61 %. Individually, 11.17% of all reads could be annotated for JR 2, 9.39% could be annotated for JR 4, 7.29% could be annotated for JR 5, and 6.58% could be annotated for JR 6.The pooled coral seawater samples had lower overall assignment, ranging from 1.59 – 12.83%, with an average read assignment of 4.48%. Reads from pooled *A. cervicornis* CSW had the highest level of annotation at 12.83%, followed by *P. strigosa* CSW (3.76%), *M. cavernosa* CSW (2.59%), *O. faveolata* CSW (1.66%), and *P. astreoides* CSW (1. 60%). The final files generated by FMAP provide the user with count data for the number of gene hits that are assigned to a specific Kyoto Encyclopedia of Genes and Genomes (KEGG) Ortholog identifier, as well as files revealing which KOs, KEGG pathways, KEGG modules, and KEGG operons are significantly different between the two sample groups (Kruskal – Wallis test, p-value < 0.05, FDR adjusted to control for false positives) .

To analyze the output from this pipeline, the KEGG KO abundance table file and significantly different (p adjusted < 0.05) KOs between the CSW and RSW file were merged using core R functions in R studio. The KO count data was converted into relative abundance using the sum of all genes that could be annotated in order to normalize changes and visually compare differences across samples. We also scaled the relative abundances using the 10^{th} and 90^{th} quantiles of the data to enhance visual comparison between the samples. Changes in gene abundance across the samples were visualized using the R package 'ComplexHeatmap' (Gu et al. 2016) (available from https://github.com/jokergoo/ComplexHeatmap) with scripts that were tailored for our data (script scaffolds were obtained from http://zuguang.de/supplementary/ComplexHeatmap-supplementary1-4/supplS2_scRNASeq/supplS2_scRNAseq.html). The dendrogram reflects hierarchical clustering of the samples using the 'hclust' function in R.

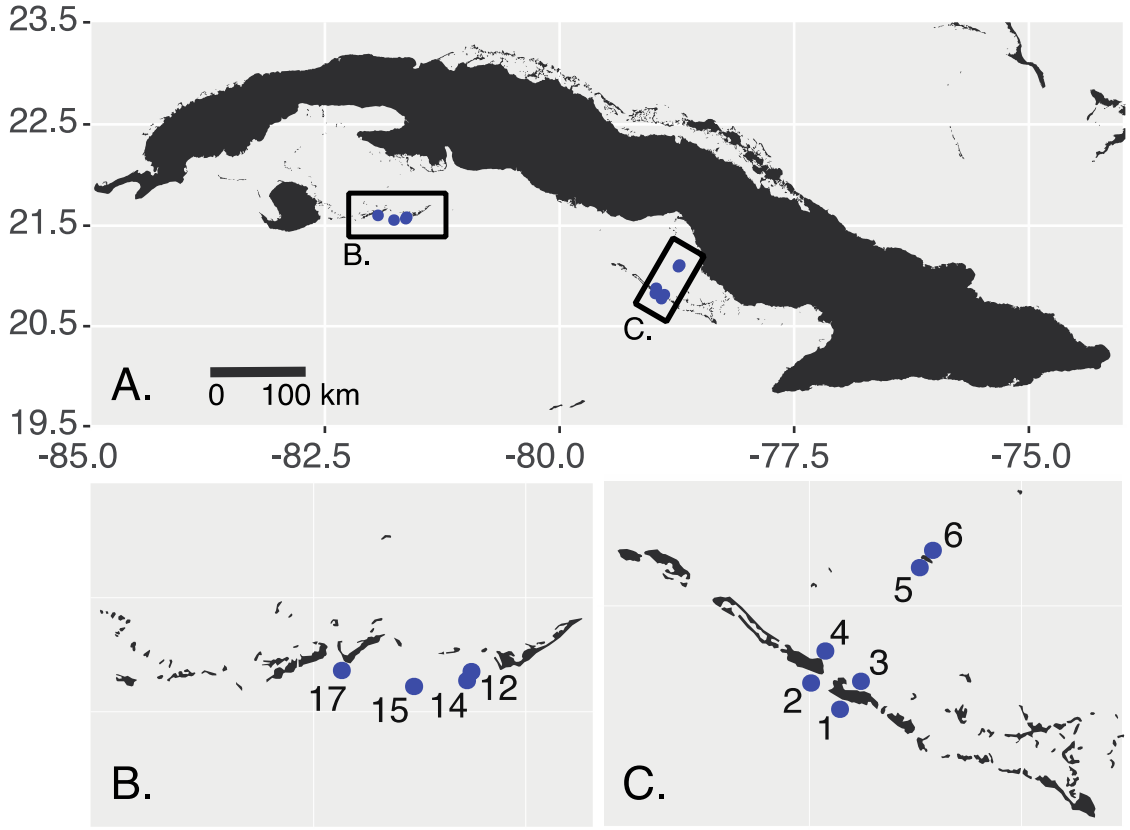

Figure S1. Overview map of the reef locations that were surveyed in this study. A) The location of the reefs relative to the island of Cuba. Reef-systems of Los Canarreos (B.) and Jardines de la Reina (C.) are contained within the black boxes B) Close-up of reef locations in the Canarreos reef-system. C) Close-up of reef locations in the Jardines de la Reina reef-system. The blue circles indicate the location of each reef.

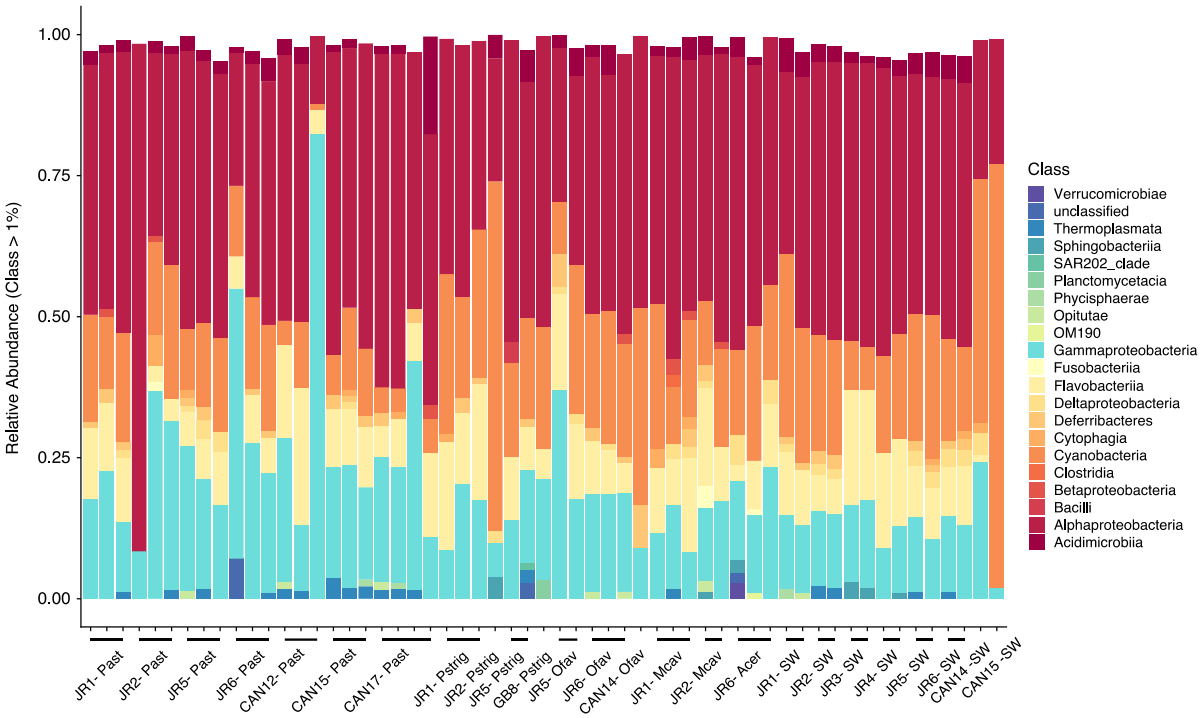

Figure S2. Relative abundance of bacterial and archaeal classes that comprise >1 % of the community across all coral seawater (CSW) and reef seawater (RSW) samples. Samples are grouped by coral species, reef location, and sample type. Colors indicate taxonomic class. JR = Jardines de la Reina, CAN = Los Canarreos. Past = *Porites astreoides* CSW, Pstrig = *Pseudodiploria strigosa* CSW, CSW, Ofav = *Orbicella faveolata* CSW, Mcav = *Montastraea cavernosa* CSW, Acer = *Acropora cervicornis* CSW, SW = reef-depth seawater.

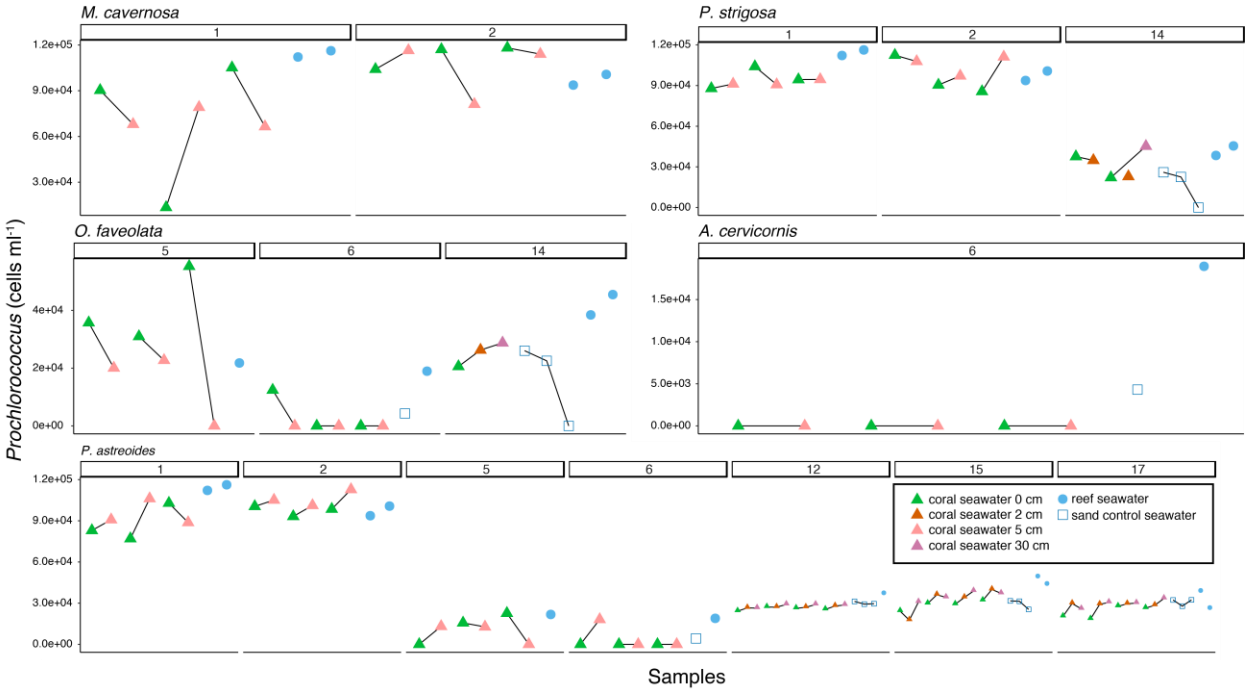

Figure S3. Abundances of *Prochlorococcus* within coral seawater (CSW), reef seawater (RSW), and sand control samples by coral species. The line connects samples that were obtained over a distance gradient from that colony and indicates the direction of the trend. RSW and sand control samples are not colored differently between surface and reef-depth or over the distance gradient sampled from the sand control samples (i.e. sites 12, 14, 15, 17; distances of 0, 2, and 30 cm away from the sand).

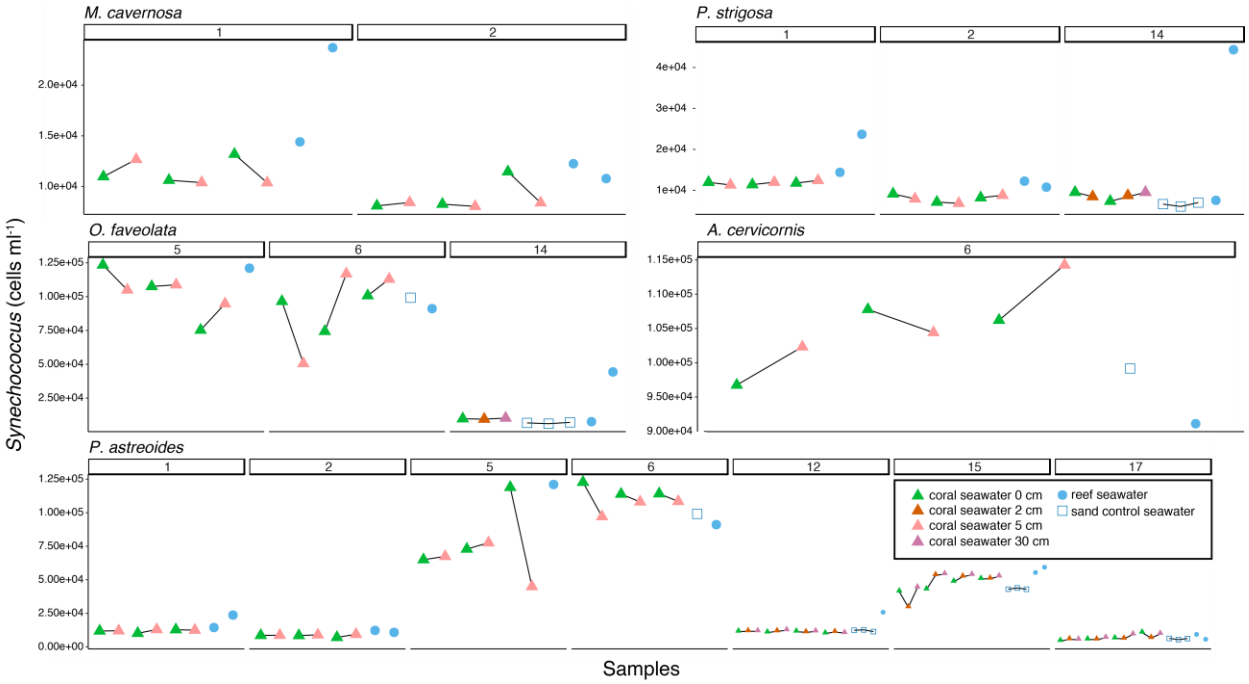

Figure S4. Abundances of *Synechococcus* within coral seawater (CSW), reef seawater (RSW), and sand control samples by coral species. The line connects samples that were obtained over a distance gradient from that colony and indicates the direction of the trend. RSW and sand control samples are not colored differently between surface and reef-depth or over the distance gradient sampled from the sand control samples (i.e. sites 12, 14, 15, 17; distances of 0, 2, and 30 cm away from the sand).

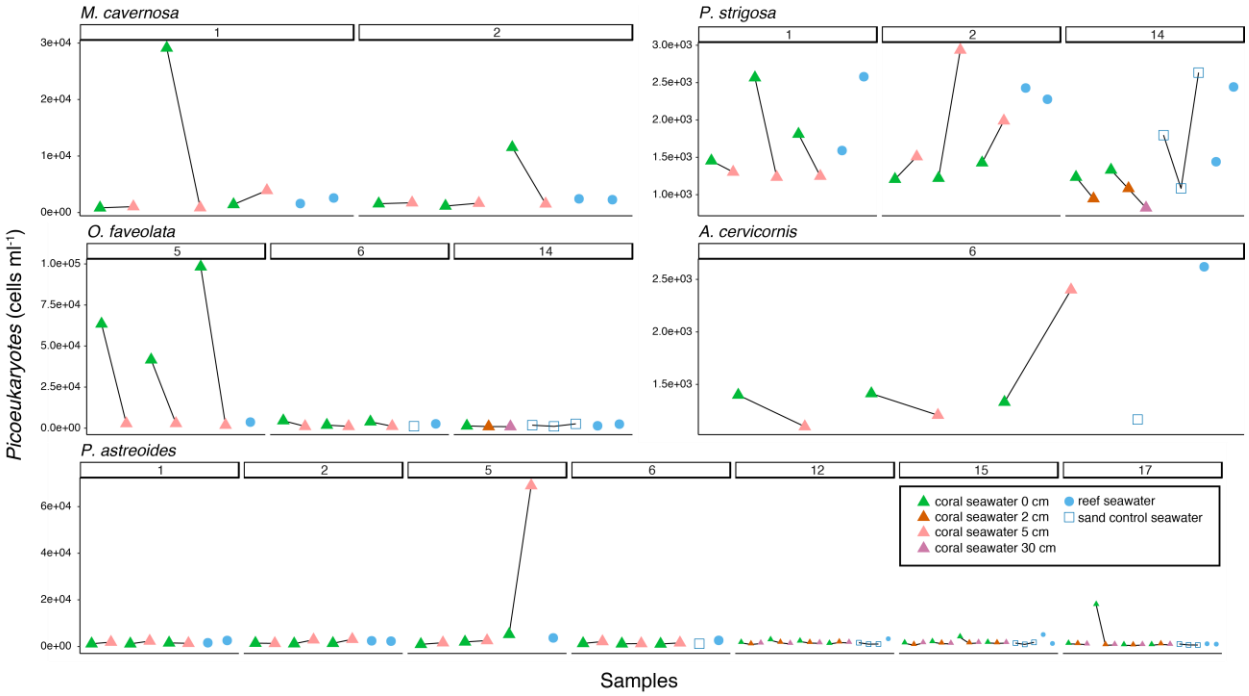

Figure S5. Abundances of picoeukaryotes within coral seawater (CSW), reef seawater (RSW), and sand control samples by coral species. The line connects samples that were obtained over a distance gradient from that colony and indicates the direction of the trend. RSW and sand control samples are not colored differently between surface and reef-depth or over the distance gradient sampled from the sand control samples (i.e. sites 12, 14, 15, 17; distances of 0, 2, and 30 cm away from the sand).

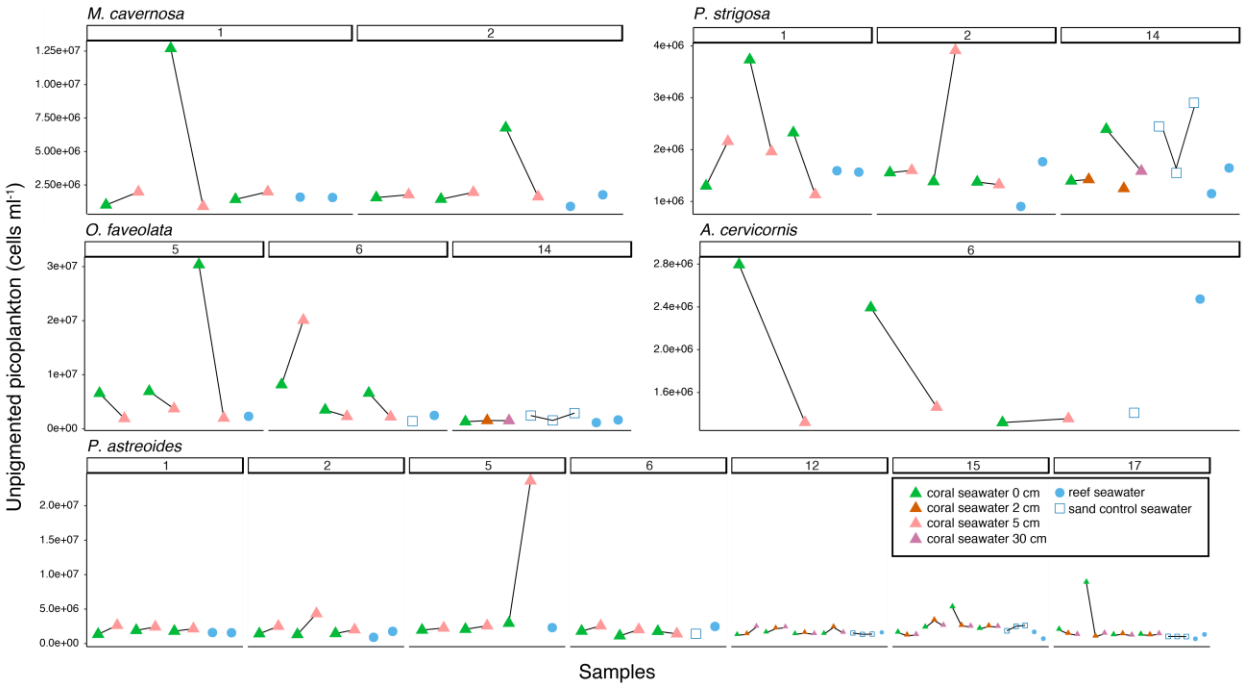

Figure S6. Abundances of unpigmented picoplankton within coral seawater (CSW), reef seawater (RSW), and sand control samples by coral species. The line connects samples that were obtained over a distance gradient from that colony and indicates the direction of the trend. RSW and sand control samples are not colored differently between surface and reef-depth or over the distance gradient sampled from the sand control samples (i.e. sites 12, 14, 15, 17; distances of 0, 2, and 30 cm away from the sand).

Table S1. Significantly enriched and depleted MED nodes detected in Jardines de la Reina (JR) coral seawater (CSW) compared to reef seawater (RSW) by reef site according to paired differential enrichment analysis using DESeq2[§].

Site	Coral	MED node	Log2 fold change	Padj*	Taxa
Enriched in Coral Seawater (CSW)					
JR1	*P. astreoides*	MED2613	21.88	4.28E-04	Gammaproteobacteria, Cellvibrionaceae
		MED3465	16.29	1.31E-03	Gammaproteobacteria, *Marinobacter*
		MED5309	15.55	1.31E-03	Gammaproteobacteria, *Alteromonas*
JR2	*P. astreoides*	MED56	27.54	4.83E-08	Alphaproteobacteria, SAR11 clade, Surface 1
		MED2424	23.61	5.96E-06	Gammaproteobacteria, *Vibrio*
		MED5299	21.23	4.83E-08	Gammaproteobacteria, *Vibrio*
		MED2524	21.18	4.83E-08	Fusobacteria, *Propionigenium*
		MED798	11.4	4.39E-10	Gammaproteobacteria, *Endozoicomonas*
		MED3416	11.32	2.98E-10	Gammaproteobacteria, *Endozoicomonas*
JR5	*P. astreoides*	MED3416	20.8	2.79E-06	Gammaproteobacteria, *Endozoicomonas*
		MED3450	17.7	3.02E-05	Firmicutes, Bacillales, unclassified
		MED5924	16.46	3.02E-05	Gammaproteobacteria, *Psychrobacter*
		MED2954	9.89	2.40E-03	Gammaproteobacteria, *Bermanella*
JR6	*P. astreoides*	MED3416	21.7	2.10E-04	Gammaproteobacteria, *Endozoicomonas*
		MED832	21.21	2.23E-04	Gammaproteobacteria, *Endozoicomonas*
		MED5123	18.93	1.03E-05	Gammaproteobacteria, *Psychrobacter*
		MED615	18.81	2.10E-04	Flavobacteria, *Mesoflavibacter*
		MED29181	18	3.10E-04	Gammaproteobacteria, *Pseudoalteromonas*
		MED20	10.67	2.88E-02	Gracilibacteria, unclassified
		MED2840	10.04	2.52E-02	Gammaproteobacteria, *Idiomarina*
		MED53094	7.03	1.95E-02	Gammaproteobacteria, *Alteromonas*
JR1	*P. strigosa*	MED1993	21.59	7.02E-05	Alphaproteobacteria, SAR11 clade
		MED4248	21.07	9.73E-05	Alphaproteobacteria, SAR11 clade, Surface 2
		MED5392	21.02	9.73E-05	Flavobacteria, unclassified
		MED4380	20.97	9.73E-05	Cyanobacteria, Subsection III, *Leptolyngbya*
		MED2918	20.54	1.39E-04	Gammaproteobacteria, *Pseudoalteromonas*
		MED2355	20.42	1.46E-04	Bacteroidetes, Cytophagia, Order III
		MED2577	20.33	2.12E-05	Gammaproteobacteria, *Marinobacter*
		MED5468	20.3	7.02E-05	Alphaproteobacteria, SAR 116 clade
		MED53091	19.72	4.53E-06	Gammaproteobacteria, *Alteromonas*
		MED34651	19.72	7.81E-06	Gammaproteobacteria, *Marinobacter*
		MED4860	7	4.58E-02	Alphaproteobacteria, SAR11 clade, Surface 1
JR2	*P. strigosa*	MED1408	12.01	1.09E-05	Bacteroidetes, Chitinophagaceae
		MED1451	11.53	5.39E-04	Gammaproteobacteria, *Endozoicomonas*

		MED3451	8.55	1.56E-02	Deltaproteobacteria, OM27 clade
		MED6080	2.86	7.11E-03	Alphaproteobacteria, SAR11, Surface 1 clade
		MED36901	2.63	3.95E-02	Actinobacteria, "*Candidatus* Actinomarina"
JR5	*P. strigosa*	MED2229	23.79	3.13E-10	Alphaproteobacteria, Rhodobacteraceae
		MED1731	23.52	1.50E-05	Bacteroidetes, *Owenweeksia*
		MED3330	22.73	2.16E-08	Alphaproteobacteria, "*Candidatus* Laris"
		MED34501	22.17	5.25E-08	Firmicutes, Bacillales, unclassified
		MED29541	7.97	1.86E-02	Gammaproteobacteria, *Bermanella*
JR1	*M. cavernosa*	MED2581	21.23	1.31E-04	Gammaproteobacteria, *Endozoicomonas*
		MED2356	20.1	3.41E-04	Bacteroidetes, Cytophagia
		MED42	20.05	1.31E-04	Betaproteobacteria, *Variovorax*
		MED1037	19.8	3.99E-04	Firmicutes, Clostridia, Halanaerobiales
		MED53092	19.49	3.02E-06	Gammaproteobacteria, *Alteromonas*
		MED34652	18.11	5.13E-05	Gammaproteobacteria, *Marinobacter*
		MED25771	15	5.30E-03	Gammaproteobacteria, *Marinobacter*
JR5	*O. faveolata*	MED59241	22.98	2.10E-09	Gammaproteobacteria, *Psychrobacter*
		MED2233	11.27	2.44E-03	Alphaproteobacteria, *Erythrobacter*
		MED1437	11.06	7.24E-03	Alphaproteobacteria, *Thalassospira*
		MED5923	10.02	7.24E-03	Gammaproteobacteria, *Psychrobacter*
		MED53093	8.44	1.14E-04	Gammaproteobacteria, *Alteromonas*
JR6	*A. cervicornis*	MED2482	23.32	1.49E-05	Gammaproteobacteria, Piscirickettsiaceae
		MED51231	21.73	1.35E-07	Gammaproteobacteria, *Psychrobacter*
		MED29182	21.53	1.43E-05	Gammaproteobacteria, *Pseudoalteromonas*
		MED1955	21.41	1.69E-04	Deltaproteobacteria, SAR324 Clade (marine group B)
		MED1266	21.37	1.69E-04	Parcubacteria (OD1), unclassified
		MED5184	21.22	4.53E-05	Gammaproteobacteria, OM60 (Nor5) clade
		MED4684	20.85	2.59E-04	Gammaproteobacteria, SAR86 clade
		MED710	20.18	4.57E-04	Verrucomicrobia, *Rubritalea*
		MED6177	1.81	4.87E-02	Alphaproteobacteria, SAR11 clade, Surface 1
		MED6119	1.23	1.69E-04	Cyanobacteria, *Synechococcus*
Depleted in Coral Seawater (CSW)					
JR1	*P. astreoides*	MED4826	-19.52	2.38E-04	Alphaproteobacteria, SAR11 clade, Surface 1
JR5	*P. astreoides*	MED2284	-17.34	3.14E-06	Alphaproteobacteria, Rhodobacteraceae
JR6	*P. astreoides*	MED1929	-15.64	1.37E-02	Bacteroidetes, Chitinophagaceae
JR1	*P. strigosa*	MED48261	-18.84	4.63E-05	Alphaproteobacteria, SAR11 clade, Surface 1
		MED4567	-18.24	1.43E-05	Bacteroidetes, NS9 marine group
		MED2968	-17.95	8.27E-06	Alphaproteobacteria, AEGEAN-169 marine group
		MED6081	-17.1	5.21E-04	Alphaproteobacteria, SAR11 clade, Surface 1
		MED1081	-8.87	3.48E-03	Actinobacteria, "*Candidatus* Actinomarina"
		MED3565	-8.77	2.93E-02	Alphaproteobacteria, SAR11 clade, Surface 4

		MED521	-8.62	5.25E-04	Alphaproteobacteria, Rhodobacteraceae
		MED5856	-6.67	4.58E-02	Alphaproteobacteria, SAR11 clade, Surface 1
		MED3690	-3.18	4.58E-02	Actinobacteria, "*Candidatus* Actinomarina"
JR2	*P. strigosa*	MED3636	-8.5	7.11E-03	Alphaproteobacteria, AEGEAN-169 marine group
		MED3857	-8.18	1.48E-02	Cyanobacteria, Subsection1, Family1
		MED5355	-7.76	8.24E-03	Bacteroidetes, NS5 marine group
		MED5238	-7.62	3.59E-02	Gammaproteobacteria, SAR86 clade
		MED3368	-7.42	3.95E-02	Gammaproteobacteria, SAR86 clade
		MED3501	-7.37	2.59E-02	Alphaproteobacteria, SAR116 clade
		MED4287	-6.98	3.95E-02	Alphaproteobacteria, SAR11 clade, Surface 2
		MED6124	-3.35	5.41E-03	Alphaproteobacteria, SAR11 clade
JR5	*P. strigosa*	MED2284	-15.4	8.12E-05	Alphaproteobacteria, Rhodobacteraceae
JR1	*M. cavernosa*	MED3620	-21.2	1.20E-07	Deferribacteres, SAR406 clade (Marine group A)
		MED1441	-9.93	1.31E-04	Gammaproteobacteria, SAR86 clade
		MED35651	-8.94	3.42E-02	Alphaproteobacteria, SAR11 clade, Surface 4
		MED5761	-5.8	3.42E-02	Alphaproteobacteria, SAR11 clade, Surface 1
JR6	*A. cervicornis*	MED3341	-8.65	1.36E-04	Gammaproteobacteria, SAR86 clade
		MED52381	-7.62	2.92E-02	Gammaproteobacteria, SAR86 clade

*padjust = adjusted p-value calculated using Benjamini-Hochberg corrections.
[§]Only site and species combinations with significantly enriched or depleted MED nodes are indicated and taxa are arranged in order from the highest to lowest $\log_2$ fold change within each species and site grouping.

Table S2. Relative abundance (%) of two-component system KEGG Orthologs (KO) across reef seawater (RSW) and coral seawater (CSW) metagenomes.

KO definition	JR 2 RSW	JR 4 RSW	JR 5 RSW	JR 6 RSW	Acer* CSW	Pstrig[&] CSW	Mcav[+] CSW	Ofav[#] CSW	Past[$] CSW
K02477; two-component system, LytTR family, response regulator	0.0070	0.0157	0.0080	0.0072	0.0658	0.0526	0.0366	0.0365	0.0263
K02478; two-component system, LytTR family, sensor kinase [EC:2.7.13.3]	0.0008	0.0029	0.0027	0.0018	0.0321	0.0697	0.0151	0.0133	0.0303
K02481; two-component system, NtrC family, response regulator	0.0108	0.0468	0.0407	0.0331	0.1015	0.0569	0.0452	0.1163	0.0850
K02483; two-component system, OmpR family, response regulator	0.0513	0.1251	0.1487	0.1280	0.3640	0.3529	0.3919	0.3489	0.2286
K02484; two-component system, OmpR family, sensor kinase [EC:2.7.13.3]	0.0025	0.0080	0.0160	0.0168	0.0440	0.1466	0.1055	0.0565	0.0283
pleD; two-component system, cell cycle response regulator [EC:2.7.7.65]	0.0017	0.0132	0.0027	0.0042	0.0131	0.0470	0.1012	0.0399	0.0121
pilH; twitching motility two-component system response regulator	0.0037	0.0055	0.0073	0.0024	0.0131	0.0313	0.0129	0.0166	0.0142
pilS, pehS; two-component system, NtrC family, sensor histidine kinase PilS [EC:2.7.13.3]	0.0004	0.0070	0.0060	0.0030	0.0354	0.0612	0.0194	0.0266	0.0364
cheA; two-component system, chemotaxis family, sensor kinase [EC:2.7.13.3]	0.0178	0.1043	0.0667	0.0625	0.4620	0.3685	0.3273	0.3057	0.1861
cheB; two-component system, chemotaxis family, response regulator [EC:3.1.1.61]	0.0033	0.0333	0.0173	0.0258	0.2554	0.1651	0.2541	0.1130	0.0829
cheY; two-component system, chemotaxis family, response regulator	0.0050	0.0516	0.0167	0.0319	0.1447	0.1181	0.1077	0.0797	0.0748
cheV; two-component system, chemotaxis family, response regulator	0.0037	0.0139	0.0040	0.0042	0.1545	0.0768	0.0431	0.0199	0.0829
phoQ; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0018	0.0047	0.0018	0.0457	0.0470	0.0086	0.0133	0.0344
rstB; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0037	0.0033	0.0012	0.0944	0.0825	0.0323	0.0598	0.0263
cpxA; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0033	0.0047	0.0030	0.0879	0.0370	0.0108	0.0233	0.0263
creC; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0022	0.0020	0.0006	0.0121	0.0455	0.0237	0.0233	0.0061
baeS, smeS; two-component system, OmpR family, sensor histidine kinase BaeS [EC:2.7.13.3]	0.0012	0.0026	0.0060	0.0066	0.0777	0.0455	0.0237	0.0199	0.0222
cusS, copS, silS; two-component system, OmpR family, heavy metal sensor histidine kinase CusS [EC:2.7.13.3]	0.0008	0.0018	0.0040	0.0024	0.0676	0.1138	0.0667	0.0432	0.0324

qseC; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0017	0.0015	0.0053	0.0054	0.0795	0.1039	0.0711	0.0365	0.0384
kdpD; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0015	0.0040	0.0024	0.0529	0.3088	0.2670	0.1229	0.0384
torS; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0026	0.0027	0.0048	0.0130	0.0100	0.0108	0.0066	0.0061
arcB; two-component system, OmpR family, aerobic respiration control sensor histidine kinase [EC:2.7.13.3]	0.0017	0.0051	0.0060	0.0066	0.0781	0.0228	0.0108	0.0266	0.0303
mtrB; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0012	0.0066	0.0113	0.0108	0.0119	0.0640	0.0301	0.0266	0.0364
phoP; two-component system, OmpR family, response regulator	0.0008	0.0190	0.0093	0.0072	0.0365	0.0256	0.0194	0.0332	0.0243
rstA; two-component system, OmpR family, response regulator	0.0025	0.0048	0.0027	0.0054	0.0735	0.0484	0.0129	0.0332	0.0222
cpxR; two-component system, OmpR family, response regulator	0.0004	0.0099	0.0073	0.0066	0.1146	0.0327	0.0129	0.0332	0.0425
cusR, copR, silR; two-component system, OmpR family, copper resistance phosphate regulon response regulator	0.0021	0.0048	0.0067	0.0024	0.0542	0.0953	0.0560	0.0332	0.0162
qseB; two-component system, OmpR family, response regulator	0.0029	0.0022	0.0020	0.0042	0.0564	0.0526	0.0172	0.0631	0.0162
kdpE; two-component system, OmpR family, KDP operon response regulator	0.0012	0.0026	0.0020	0.0024	0.0214	0.0953	0.1120	0.0399	0.0202
uhpB; two-component system, NarL family, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0004	0.0000	0.0000	0.0204	0.0398	0.0215	0.0233	0.0324
barA, gacS, varS; two-component system, NarL family, sensor histidine kinase BarA [EC:2.7.13.3]	0.0004	0.0106	0.0067	0.0084	0.2554	0.1537	0.0689	0.1362	0.0850
evgS, bvgS; two-component system, NarL family, sensor histidine kinase EvgS [EC:2.7.13.3]	0.0008	0.0026	0.0040	0.0024	0.0293	0.1138	0.0474	0.0465	0.0101
narL; two-component system, NarL family, nitrate/nitrite response regulator	0.0054	0.0088	0.0060	0.0024	0.0538	0.0341	0.0215	0.0100	0.0222
desR; two-component system, NarL family, response regulator	0.0021	0.0015	0.0040	0.0012	0.0422	0.0441	0.0086	0.0332	0.0162
glrR, qseF; two-component system, NtrC family, response regulator GlrR	0.0050	0.0124	0.0127	0.0156	0.0830	0.0427	0.0258	0.0299	0.0425
senX3; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0059	0.0073	0.0084	0.0155	0.0470	0.0366	0.0133	0.0202
arcA; two-component system, OmpR family, aerobic respiration control protein	0.0012	0.0022	0.0033	0.0048	0.0609	0.0242	0.0108	0.0133	0.0182
desK; two-component system, NarL family, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0007	0.0033	0.0006	0.0214	0.0228	0.0129	0.0365	0.0040

K07814; putative two-component system response regulator	0.0012	0.0172	0.0167	0.0126	0.0618	0.0470	0.0538	0.0465	0.0182
algZ; two-component system, LytTR family, sensor histidine kinase [EC:2.7.13.3]	0.0004	0.0055	0.0080	0.0036	0.0347	0.0413	0.0172	0.0266	0.0101
algR; two-component system, LytTR family, response regulator	0.0017	0.0066	0.0000	0.0030	0.0451	0.0384	0.0366	0.0332	0.0081
dctD; two-component system, NtrC family, C4-dicarboxylate transport response regulator	0.0004	0.0095	0.0080	0.0054	0.1543	0.0868	0.1249	0.0764	0.0344
algB; two-component system, NtrC family, response regulator	0.0008	0.0004	0.0000	0.0006	0.0122	0.0057	0.0108	0.0033	0.0040
divK; two-component system, cell cycle response regulator	0.0017	0.0062	0.0067	0.0066	0.0173	0.0128	0.0409	0.0266	0.0142
cpdR; two-component system, cell cycle response regulator	0.0004	0.0062	0.0020	0.0042	0.0057	0.0185	0.0237	0.0133	0.0101
cheBR; two-component system, chemotaxis family, CheB/CheR fusion protein [EC:2.1.1.80 3.1.1.61]	0.0012	0.0099	0.0073	0.0018	0.0308	0.0726	0.0409	0.0432	0.0162
fixL; two-component system, LuxR family, sensor kinase [EC:2.7.13.3]	0.0000	0.0011	0.0013	0.0006	0.0186	0.0413	0.0538	0.0199	0.0162
adeS; two-component system, OmpR family, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0000	0.0000	0.0000	0.0018	0.0100	0.0172	0.0100	0.0061
pfeS, pirS; two-component system, OmpR family, sensor histidine kinase PfeS [EC:2.7.13.3]	0.0012	0.0007	0.0027	0.0024	0.0287	0.0100	0.0065	0.0166	0.0162
pfeR, pirR; two-component system, OmpR family, response regulator	0.0000	0.0004	0.0000	0.0000	0.0091	0.0057	0.0086	0.0033	0.0101
NIK1, TCSC; osomolarity two-component system, sensor histidine kinase NIK1 [EC:2.7.13.3]	0.0008	0.0007	0.0013	0.0018	0.0023	0.0270	0.0258	0.0233	0.0222
sagS; two-component system, sensor histidine kinase [EC:2.7.13.3]	0.0012	0.0018	0.0007	0.0006	0.0113	0.0185	0.0108	0.0066	0.0142
K20974; two-component system, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0029	0.0033	0.0072	0.0599	0.0498	0.0194	0.0199	0.0243
K20975; two-component system, sensor histidine kinase [EC:2.7.13.3]	0.0000	0.0018	0.0060	0.0006	0.0851	0.0270	0.0280	0.0332	0.0121
hsbR; two-component system, HptB-dependent secretion and biofilm response regulator	0.0017	0.0055	0.0013	0.0030	0.0303	0.0185	0.0129	0.0399	0.0283

*Acer = *Acropora cervicornis*
[&]Pstrig = *Psuedoploria strigosa*
[+]Mcav = *Montastraea cavernosa*
[#]Ofav = *Orbicella faveolata*
[$]Past = *Porites astreoides*

Additional Supporting Methods: Seawater volume experiment

Sample collection and processing

Surface seawater samples of different volumes (60 mL, 1.5 L, and 2 L) were collected from two different reef sites in St. John, U.S. Virgin Islands in October 2016. Two 1.5 L samples and four 60 mL samples were collected from the 'Dock' location and two 2 L samples and five 60 mL samples were collected from Tektite reef. These samples were then filtered onto 0.2 μm Supor filters using a peristaltic pump, DNA was extracted using the Sucrose-EDTA lysis method (Santoro et al. 2010), and this DNA was prepared for Fluidigm® amplification of the SSU rRNA gene and 2x250 bp MiSeq Illumina sequencing at the Keck Center for Functional Genomics (University of Illinois, Urbana, IL) using the V4 primer pair 515F-Y (5'-GTGYCAGCMGCCGCGGTAA-3') and 806RB (5'-GGACTACNVGGGTWTCTAAT-3') (Apprill et al. 2015; Parada et al. 2016).

Microbial community analysis using MED and 97% similarity OTU clustering

After sequencing, the SSU rRNA gene amplicon sequences were processed using two different clustering methods: Minimum Entropy Decomposition (MED) (Eren et al. 2015) and 97% similarity Operational Taxonomic Unit (OTU) clustering. These clustering methods were chosen in order to investigate how different sampling volumes impacted the microbial diversity and composition using both methods of clustering. Prior to clustering, sequences were processed using the same pipeline that was used for analyzing the RSW and CSW samples collected in Cuba. MED analysis of these sequences was also conducted using the methods described in the methods section of this manuscript. We used mothur v.1.36.1 (Schloss et al. 2009) and guidance from the mothur MiSeq SOP webpage (https://www.mothur.org/wiki/MiSeq_SOP) to accomplish 97% similarity OTU clustering on this data. We also chose to complete clustering with non-subsampled and subsampled data (4940 sequences subsampled from each sample) to determine how subsampling impacted microbial community composition.

After clustering the sequences into either MED nodes or OTUs (97% similarity), a variety of methods were used to examine if differences in original seawater collection volume contributed to significant differences between the microbial communities. Microbial community composition data obtained using MED was analyzed with PhyloSeq, Vegan, and DESeq2 in the R environment with the same code used for analysis in this manuscript (McMurdie and Holmes 2013; Love et al. 2014; Oksanen et al. 2017). More specifically, we completed non-metric multidimensional scaling analysis (NMDS), tested differences in microbial community similarity using Permutational Multivariate Analysis of Variance tests using distance matrices (PERMANOVA/ADONIS) (Oksanen et al. 2017), and completed DESeq2 (Love et al. 2014) to test for impacts of different volumes on the microbial communities.

Additional Supporting Results: Seawater volume experiment

Microbial community alpha diversity

The 1.5 and 2 L seawater samples collected at the Dock and Tektite had greater species richness (absolute number of unique MED nodes) than the corresponding 60 mL samples that were collected at both of the sites (Figure S7). A Kruskal-Wallis rank sum test revealed statistically significant ($p = 0.04$) differences in MED node alpha diversity between the different sampling

volumes and sites, but post-hoc pairwise Dunn's tests did not detect significant differences between any of the individual sample types (Figure S8). Individual Kruskal-Wallis rank sum tests were also conducted for samples collected at each specific site (Figures S9 and S10). Both these tests revealed that there significant differences between the number of observed MED nodes as a factor of seawater sampling volume, although these differences were slight (Figures S9 and S10).

The impact of clustering method (97% OTU or MED) and depth of subsampling on alpha richness of the microbial community was also tested (Figure S11). OTU clustering of the non-subsampled and subsampled sequences yielded the highest alpha richness in samples collected from the Dock site compared to MED clustering that was preformed on the same sequences (Figure S11). In contrast, MED node clustering resulted in the highest alpha diversity in samples collected from the Tektite site (Figure S11). Within each site, larger volume samples had higher species richness, but there was also a larger discrepancy between the un-subsampled and subsampled dataset in comparison to 60 mL samples. A Kruskal-Wallis rank sum test indicated that there were significant differences by clustering type (97% OTU, subsampling OTU, or MED, $p < 0.05$), but no significant differences were found during pairwise post-hoc Dunn's testing with Bonferroni corrections.

Microbial community composition and beta diversity

Non-metric multidimensional scaling analysis (NMDS) revealed that microbial communities collected from the same site were more similar to each other than to samples collected at the other site (Figure S12). Within each site, larger volume samples clustered together, whereas the 60 mL samples were mostly evenly distributed from each other along the NMDS1 axis (Figure S12). An Adonis test revealed that site of collection significantly influenced microbial community composition whereas collection volume did not significantly contribute to these differences (Figure S13).

Lastly, no significant enrichment or depletion of MED nodes was detected between the 60 mL and 1.5 L or 2 L samples in DESeq2 comparisons.

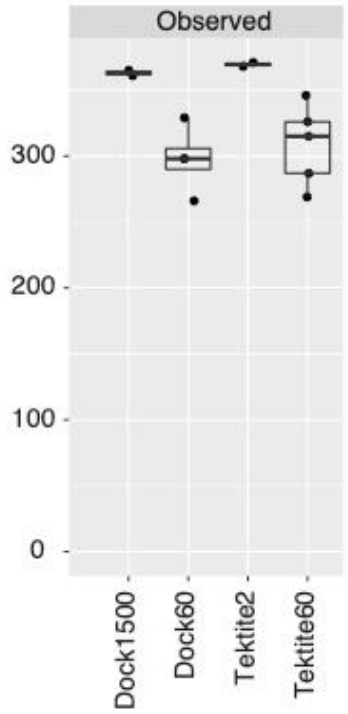

Figure S7. Boxplots of the number of observed MED nodes (absolute count of unique nodes within each sample) for seawater samples collected with different volumes of water. The absolute count of MED nodes is depicted on the y-axis and the sample type is indicated on the x-axis. Dock1500 = 1.5 L samples collected from the Dock, Dock60 = 60 mL samples collected from the Dock, Tektite2 = 2.0 L samples collected from Tektite, Tektite60 = 60 mL samples collected from Tektite. The lower and upper edges of the box correspond to the first and third quartiles and the middle black bar reflects the median. Points that fall outside of the whiskers extend beyond 1.5 X the interquartile range of the box plot.

```
  Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 8.119, df = 3, p-value = 0.04


                          Comparison of x by group
                                (Bonferroni)
Col Mean-|
Row Mean |   Dock1500     Dock60   Tektite2
---------+--------------------------------
  Dock60 |   1.781447
         |     0.2245
         |
Tektite2 |  -0.514259  -2.375262
         |     1.0000     0.0526
         |
Tektite6 |   1.567376  -0.344975   2.182034
         |     0.3511     1.0000     0.0873


List of pairwise comparisons: Z statistic (adjusted p-value)
----------------------------------------
Dock1500 - Dock60    :  1.781447 (0.2245)
Dock1500 - Tektite2  : -0.514259 (1.0000)
Dock60 - Tektite2    : -2.375262 (0.0526)
Dock1500 - Tektite60 :  1.567376 (0.3511)
Dock60 - Tektite60   : -0.344975 (1.0000)
Tektite2 - Tektite60 :  2.182034 (0.0873)

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure S8. Screenshot of results from a Kruskal-Wallis rank sum test that was performed on MED node alpha diversity within each sample. Both sites (Dock and Tektite) and sampling volumes (60 mL, 1.5 L, and 2 L) were tested.

```
  Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 3.5294, df = 1, p-value = 0.06


                          Comparison of x by group
                                (Bonferroni)
Col Mean-|
Row Mean |       1500
---------+-----------
      60 |   1.878672
         |     0.0301


List of pairwise comparisons: Z statistic (adjusted p-value)
-----------------------------
1500 - 60 :  1.878672 (0.0301)

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure S9. Screenshot of results from a Kruskal-Wallis rank sum test that was performed on MED node alpha diversity at samples collected from the Dock site.

```
  Kruskal-Wallis rank sum test

data: x and group
Kruskal-Wallis chi-squared = 3.75, df = 1, p-value = 0.05


                          Comparison of x by group
                                (Bonferroni)
Col Mean-|
Row Mean |          2
---------+-----------
      60 |   1.936491
         |     0.0264


List of pairwise comparisons: Z statistic (adjusted p-value)
---------------------------
2 - 60 :  1.936491 (0.0264)

alpha = 0.05
Reject Ho if p <= alpha/2
```

Figure S10. Screenshot of results from a Kruskal-Wallis rank sum test that was performed on MED node alpha diversity at samples collected from Tektite.

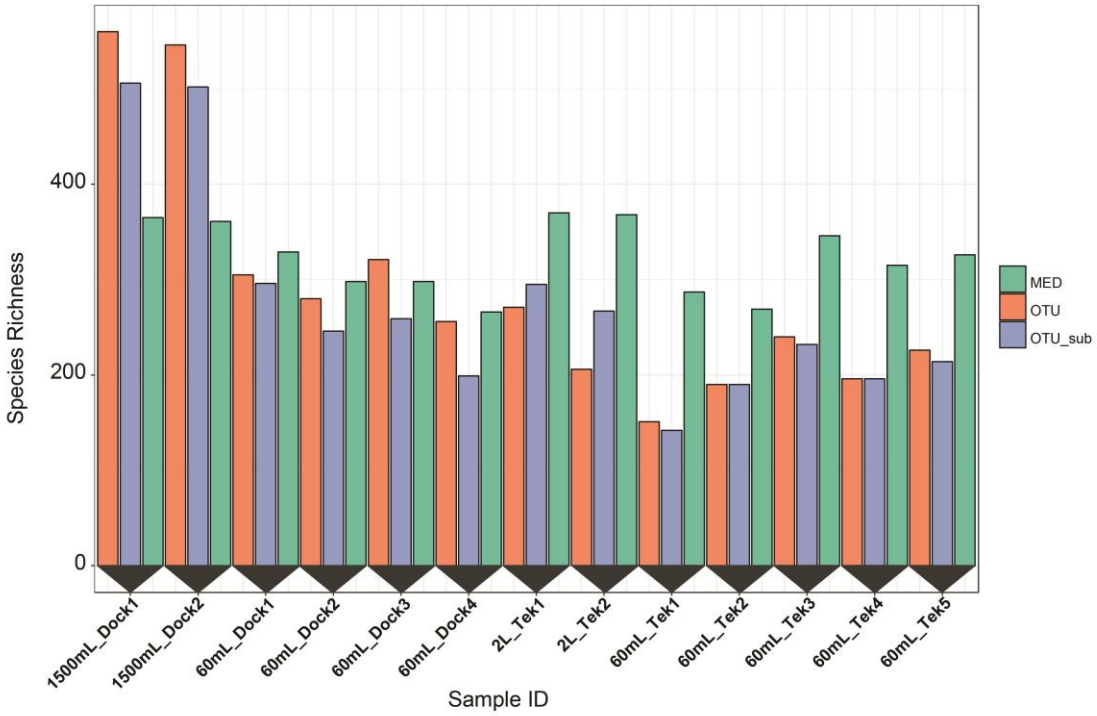

Figure S11. Comparison of microbial species richness (absolute count of unique species within each sample) by sample volume, site, and clustering method. Bars are colored by the clustering method used to analyze the sequences. MED = Minimum Entropy Decomposition, OTU = 97% OTU similarity clustering with no subsampling, and OTU_sub = 97% OTU similarity clustering with subsampling at 4940 sequences per sample.

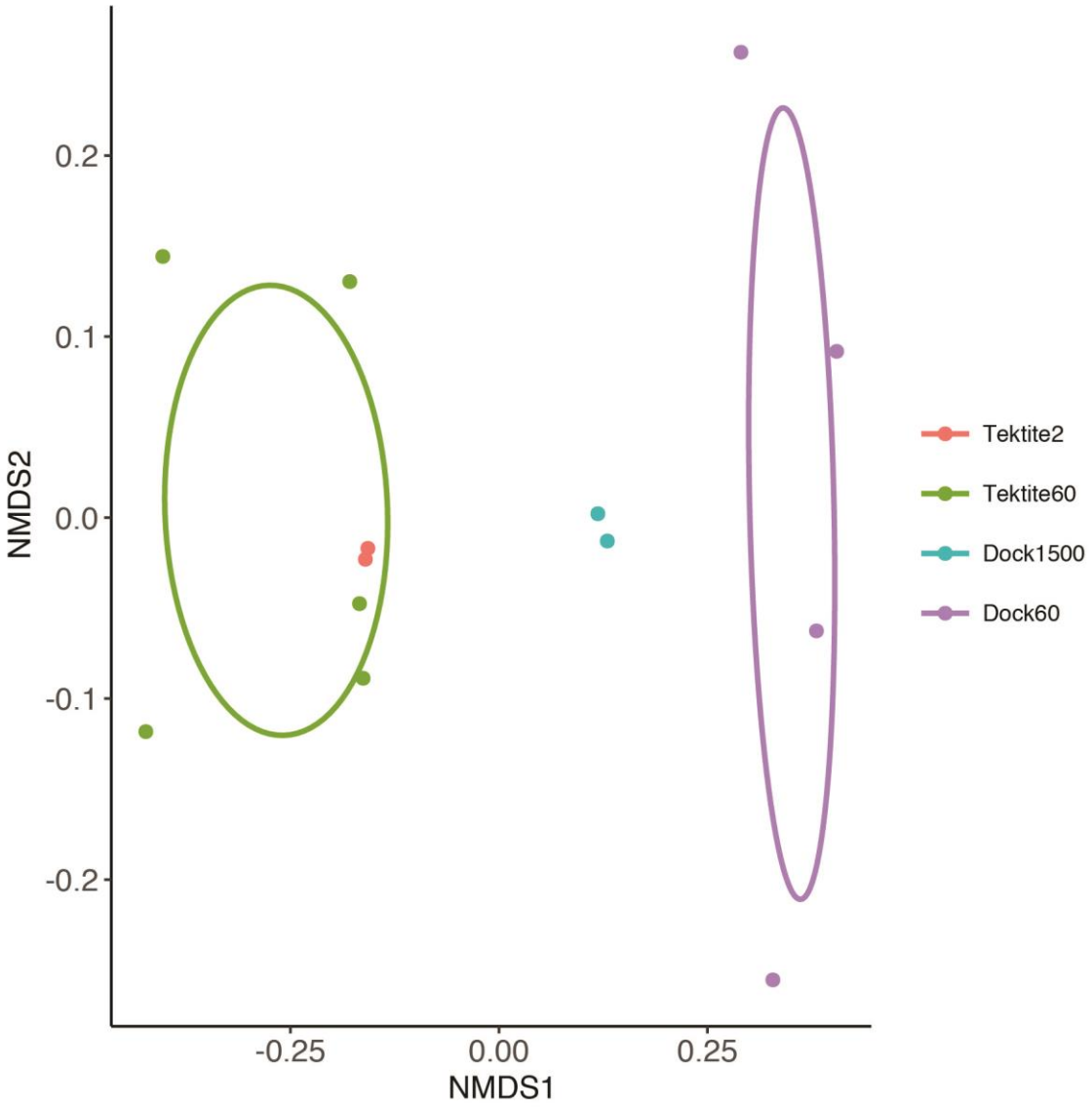

Figure S12. Non-metric multidimensional scaling (NMDS) was performed using a Bray-Curtis dissimilarity matrix that was obtained from square-root transformed microbial community SSU rRNA gene amplicon data. Ellipses are drawn using the group mean and covariance for each species (Eren et al., 2015). Different colors reflect the different sample types (site of collection and seawater volume). 2-D stress: 0.070.

```
Call:
adonis(formula = d_all ~ site + samplingvol, data = df_all)

Permutation: free
Number of permutations: 999

Terms added sequentially (first to last)

            Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
site         1   0.41303 0.41303  6.1322 0.36026  0.001 ***
samplingvol  1   0.05990 0.05990  0.8893 0.05224  0.550
Residuals   10   0.67355 0.06735         0.58749
Total       12   1.14648                 1.00000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure S13. Screenshot of PERMANOVA (Adonis) results from the test that was performed on the square-root transformed Bray-Curtis dissimilarity matrix.

References

Apprill, A., S. McNally, R. Parsons, and L. Weber. 2015. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* **75:** 129-137.

Bushnell, B. 2016. BBMap short read aligner.

Clarke, J. 2009. Cetyltrimethyl ammonium bromide (CTAB) DNA miniprep for plant DNA isolation. *Cold Spring Harb. Protoc.*

Edgar, R. C., B. J. Haas, J. C. Clemente, C. Quince, and R. Knight. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics.* **27:** 2194-2200.

Eren, A. M., H. G. Morrison, P. J. Lescault, J. Reveillaud, J. H. Vineis, and M. L. Sogin. 2015. Minimum entropy decomposition: unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *ISME J.* **9:** 968-979.

Gu, Z., R. Eils, and M. Schlesner. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* **32:** 2847-2849.

Kim, J., M. S. Kim, A. Y. Koh, Y. Xie, and X. Zhan. 2016. FMAP: Functional Mapping and Analysis Pipeline for metagenomics and metatranscriptomics studies. *BMC Bioinformatics.* **17:** 420.

Love, M. I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15:** 550.

Marie, D., F. Partensky, S. Jacquet, and D. Vaulot. 1997. Enumeration and cell cycle analysis of natural populations of marine picoplankton by flow cytometry using the nucleic acid stain SYBR Green I. *Appl. Environ. Microb.* **63:** 186-193.

McMurdie, P. J., and S. Holmes. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One.* **8:** e61217.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2017. vegan: community ecology package. R package version 2.4-5.

Parada, A. E., D. M. Needham, and J. A. Fuhrman. 2016. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* **18:** 1403-1414.

Pepe-Ranney, C., and E. K. Hall. 2015. The effect of carbon subsidies on marine planktonic niche partitioning and recruitment during biofilm assembly. *Front Microbiol.* **6:** 703.
Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glockner. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41:** D590-596.
Santoro, A. E., K. L. Casciotti, and C. A. Francis. 2010. Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ. Microbiol.* **12:** 1989-2006.
Schloss, P. D., S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, and C. F. Weber. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75:** 7537-7541.
Urakawa, H., W. Martens-Habbena, and D. A. Stahl. 2010. High abundance of ammonia-oxidizing Archaea in coastal waters, determined using a modified DNA extraction method. *Appl. Environ. Microbiol.* **76:** 2129-2135.
William, S., H. Feil, and A. Copeland. 2004. Bacterial DNA isolation CTAB protocol bacterial genomic DNA isolation using CTAB materials & reagents.
Zhou, J., M. A. Bruns, and J. M. Tiedje. 1996. DNA recovery from soils of diverse composition. *Appl. Environ. Microbiol.* **62:** 316-322.