

PNAS

www.pnas.org

Supplementary Information for

A multi-omic analysis of *in situ* coral-turf algal interactions

Ty N.F. Roach, Mark Little, Milou G.I. Arts, Joel Huckeba, Andreas F. Haas, Emma E. George, Robert Quinn, Ana G. Cobián-Güemes, Douglas S. Naliboff, Cynthia Silveira, Mark J.A. Vermeij, Linda Wegley Kelly, Pieter C. Dorrestein, and Forest Rohwer

Ty Roach

Email: smokinroachjr@gmail.com

This PDF file includes:

Supplementary text
Figures S1 to S20
Tables S1 to S5
SI References

Other supplementary materials for this manuscript include the following:

Supplementary Information Text

Supplementary Methods:

Sample collection

All samples were collected by divers on SCUBA in November 2015 around the island of Curaçao (12.1696° N, 68.9900° W). Samples for microscopy were taken by suctioning water and mucus directly off the surface of the coral, the algae, and the coral-algal interface using a 3 mL blunt tip syringe. Coral and algal samples were taken 10 cm away from the interface (Figure 1) by suctioning 100 µL of surface-associated water and mucus every 0.5 cm over a 10 cm transect parallel to the interface, yielding a total sample volume of 2 mL. A similar process was conducted along a 10 cm transect of the interface.

Biopsies for metagenomic and metabolomic analysis were taken using an underwater power drill (Nemo Power Tools, Santa Clara, CA, USA). Biopsies, 1 cm in diameter and 1 cm in length, were collected in a transect perpendicular to the coral-algal interface (Figure 1). The coral and algal samples were taken 10 cm away from the interface on their respective sides. The Interface samples were taken directly at the interface of the coral and turf algae. Biopsies for metabolomic analysis were placed into 10 mL of LCMS grade 70% methanol and 30% water for metabolite extraction and later analyzed via liquid chromatography tandem mass spectrometry (LC-MS/MS). Samples for Metagenomic analysis were placed in RNA Later (Thermo Fisher Scientific, Waltham, MA, USA) and stored at -80° C after 30 min.

Epifluorescence microscopy

Microscopy samples were divided into two aliquots in order to analyze both viral abundance and microbial size via SYBR and DAPI staining respectively. SYBR aliquots were fixed with microscopy grade paraformaldehyde at 1% final concentration, vacuum filtered onto a 0.02 µm Anodisc filter (Whatman Inc., Florham Park, NJ, USA), stained with SYBR Gold (5 X final concentration; Invitrogen, Carlsbad, CA, USA), and mounted on microscope slides. DAPI aliquots were fixed with microscopy grade glutaraldehyde at 2% final concentration, vacuum filtered onto a 0.2 µm Anodisc filter (Whatman Inc., Florham Park, NJ, USA), stained with DAPI (5 µg·mL⁻¹ final concentration; Invitrogen, Carlsbad, CA, USA), and mounted on microscope slides. The

stained filters were imaged using an epifluorescence microscope (excitation/emission: 358/461 nm) at 600x magnification and were quantified using *Image Pro* software (Media Cybernetics).

Calculation of metabolic power output

Metabolic power was calculated using the methods of McDole et al., 2012 (1). Briefly, whole organism metabolic rate (I), defined as the amount of energy per unit time that an individual organism requires, was calculated using Equation 1:

$$I = i_0 M^\alpha e^{-E/kT}$$

Where i_0 is the mass-independent normalization constant, M is the wet weight of the organism in grams, and α is the scaling exponent. The effects of temperature on metabolic rate are accounted for by $e^{-E/kT}$ where E is the activation energy, k is Boltzmann's constant ($8.62 \times 10^{-5} \text{ eV} \cdot \text{K}^{-1}$), and T is the water temperature (in Kelvin) at the site at the time of collection. Community-level metabolic rates were calculated by summing the individual metabolic rates (I) for all microbes in a sample.

Metagenome generation and analysis

Total DNA was extracted from coral punches using the AllPrep DNA/RNA kit (Qiagen). Metagenomic libraries were constructed using Illumina Nextera XT library preparation kits and sequenced at the SDSU sequencing facility via 96-plex sequencing with 600 cycles using the MiSeq platform with 2 x 300 pair-end read chemistry. Raw reads were quality-filtered by removing short reads (<60 bp), reads with quality scores <20, reads with >1% ambiguous bases, low complexity reads (entropy>70), and duplicate reads, using the program PRINSEQ (2). Quality filtered metagenomic libraries were aligned using SUPERFOCUS (3) and the SEED hierarchical database of BLASTX-translated protein orthologs classified according to putative functional families (4). Relative abundances of taxa and functional gene classifications within each metagenome were used as input to the multivariate statistical and distance-based analyses described in the *statistical analysis* section.

Shotgun sequence metagenomic libraries generated a total of 19,388,513 raw reads with an average of 1,077,140 reads per sample (\pm 148,402 raw reads per sample). This resulted in 18,819,251 quality-filtered reads with an average of 1,045,514 quality-filtered reads per sample (\pm

144,414 reads) of which an average of 12.3% ($\pm 3\%$) were bacterial and $< 0.5\%$ ($\pm 0.07\%$) were non-bacterial microbes. More information and statistics on metagenomic libraries can be found in Table S3.

Bioinformatics search for prophage in metagenomes

Unassembled metagenomic reads were queried using Fragment Recruit Assembly Purification (FRAP, <https://github.com/yinacobian/frap>) against a prophage protein database and bacterial genome reference dataset. FRAP uses the SMALT pairwise sequence alignment program where we chose an $\geq 80\%$ nucleotide identity for the complete query read against both databases (5). The prophage protein dataset consisted of 1.5 million contigs from PhiSpy predicted proteins in the NCBI RefSeq database and the bacterial genome database contained 66,000 complete bacterial genomes from NCBI RefSeq (6,7). Our metric for percent prophage was determined by dividing the number of hits to the PhiSpy predicted protein database by the hits to the NCBI bacterial genome reference set.

Ultra-Performance Liquid Chromatography – Tandem Mass Spectrometry

The extracted metabolites were separated with UltiMate 3000 Ultra-Performance Liquid Chromatography (UPLC) system (Thermo Scientific) using a Kinetex™ 1.7mm C18 reversed phase UHPLC column (50 mm x 2.1 mm). The gradient used for the chromatographic separation consists of two solvents, solvent A (2% acetonitrile and 98% of 0.1% formic acid in LC-MS grade water) and solvent B (98% acetonitrile and 2% of 0.1% formic acid in LC-MS grade water). The gradient started with 90%-10% of solvent A and B respectively for 1.5 minutes followed by a step wise gradient change of 10% every 30 sec. for 2 minutes. Then the 50%-50% mixture was held for 2 min., followed by the increase of solvent B from 50% (50%-50%) to 100% (0%-100%) in 6 min. The 100% solvent B was held for 30 sec. Within the next 30 sec., the mixture changed from 0%-100% to a 90%-10% mixture and was kept at this mixture for another 30 sec. Throughout the run, the flow rate was kept constant at $0.5 \text{ mL} \cdot \text{min}^{-1}$.

A Maxis Q-TOF mass spectrometer (MS) (Bruker Daltonics) was coupled to the UPLC system, directly measuring the compounds coming off the LC-column. The spectrometer was equipped with an electrospray ionization (ESI) source ($200 \text{ }^\circ\text{C}$). Positive ion mode acquired MS

spectra in the range of 50 – 2000 m/z. Prior to data collection, the spectrometer was externally calibrated with ESI-L Low Concentration Tuning Mix (Agilent technologies, Santa Clara, CA, USA). Throughout runs, Hexakis (1H,1H,3H-tetrafluoropropoxy) phosphazene (Synquest Laboratories, Alachua, FL, USA; m/z 922.0098) was used as the internal calibrant. Both MS1 and MS2 had a nebulizer gas pressure (nitrogen) of 2 bar, dry gas flow of 9 L·min⁻¹ source temperature, and a capillary voltage of 4500 V. MS1 had a spectral rate of 3 Hz, and MS2 a spectral rate of 10 Hz. To obtain MS/MS fragmentation, the ten most intense ions per MS1 scan were introduced into the MS2, where they were fragmented using collision-induced dissociation. Automatic exclusion was used where an ion would be ignored in more than 3 scans, but when intensity was 2.5x the previous scan it would be re-fragmented.

Feature Table Generation

The raw datafiles from the MS machine were converted into .mzXML files with the Bruker Data Analysis software version 4.1. The .mzXML files are available on the MassIVE database (massive.ucsd.edu) under number MSV000080597 and MSV000080632 (same dataset). The .mzXML files were imported into MZmine (8). Mass detection threshold for MS1 was 3.00E+03, and 1.00E+02 for MS2. For building the chromatogram a minimum peak height of 6.00E+03, with a minimum peak duration of 1.00E-2 was set, together with a mass error of 2.50E+01 ppm, and 5.00E-02 m/z. For deconvolution, we used a baseline cutoff, with the minimum peak height of 4.00E+03, peak duration range of 0.01 to 3 min., and a baseline level of 1.00E+03. The m/z range for MS2 scan paring was set to 5.00E-02 Da, and the retention time range was set to 0.2 min. Isotope peaks were grouped with the m/z tolerance set to 5.00E-02 or 25 ppm, and a retention time tolerance of 0.1 and a maximum charge of 4. The representative isotope would be the most intense one. Features of different samples are aligned with an m/z tolerance of 5.00E0.2 m/z or 25 ppm with a weight for m/z of 75. Retention time tolerance for alignment was set to 0.1 with a weight of 25. The peak list is filtered for a minimum of 2 peaks in a row, and 2 peaks in an isotope pattern. Only peaks with an MS2 scan were kept. Duplicated peaks were filtered out using the m/z tolerance of 5.00E-02 m/z or 25 ppm, and a retention time tolerance of 0.1. Gap filling occurred with an intensity tolerance of 10%, m/z tolerance of 5.00E-

02 m/z or 25 ppm, and a retention time tolerance of 0.15. The data was exported into a .CSV file (feature table) and a .mgf file for GNPS.

Molecular Network Generation

Molecular networks were created on GNPS using the molecular networking workflow with a cosine score above 0.65 and more than 4 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 20 most similar nodes. The spectra in the network were then searched against GNPS's spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 4 matched peaks. The GNPS buckettable was downloaded after network analysis and spectral intensities used to identify differential metabolites between coral, algae and interface. The molecular network used to identify the ceramide molecule is available here:

<https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=15cfb993ddb24c4986f40c62510b9661>.

Molecular Formula assignment and calculation of Nominal Oxidation State of Carbon

For formula assignment the .mgf file generated by MZMine2 (9) was imported into SIRIUS 4.0.1 (<https://bio.informatik.uni-jena.de/sirius/>) for molecular structure identification. SIRIUS 4.0.1 was used to generate putative molecular formulas with an allowed mass deviation of 0.0020 ppm. The formula with the highest probability (nr 1 predictor as identified by SIRIUS 4.0.1) was used to calculate the Nominal Oxidation State of Carbon (NOSC) as described in Graham et al., 2017 (10) using:

$$NOSC = - \left(\frac{-Z + 4a + b + 3c - 2d + 5e - 2f}{a} \right) + 4$$

where *a*, *b*, *c*, *d*, *e*, and *f* are the numbers of C, H, N, O, P, and S atoms respectively in a given organic molecule and *Z* is net charge of the organic molecule. These NOSCs were then used to calculate the Gibbs Free Energy of Carbon Oxidation ($\Delta G^{\circ}_{\text{CO}_x}$) of these compounds using the methods described in LaRowe and Van Cappellen, 2011 (11).

Statistical Analysis

All tests were conducted with an alpha of 0.05 (95% confidence level). A one-way analysis of variance (ANOVA) followed by a Tukey *post hoc* analysis were used to test for

significant differences in viral abundance, bacterial abundance, bacterial size, bacterial metabolic power output, virus to microbe ratio, bacterial taxa, functional genes, NOSCs, $\Delta G^{\circ}_{\text{COX}}$, and ceramide abundance by treatments (i.e., an effect of coral, algae, interface). Data were further analyzed with linear regression comparing log (viral abundance) and virus to microbe ratio to log (bacterial abundance) and comparing cell division genes and total bacterial biomass to the *Bacteroidetes* to *Firmicutes* ratio. All the aforementioned statistical analyses were performed using JMP 14 software (SAS Software). All dendrograms were produced using the Wards hierarchical clustering function in JMP 14. Two-way dendrogram heat maps, were produced by using the two-way clustering function in JMP 14.

All random forests analyses were performed in R using the 'rfPermute' in combination with the 'randomForest' package (12). In supervised random forests, the competition outcome (i.e., winning or losing) of each sample was given and used for the learning process to identify winning or losing samples based on the metabolomic, taxonomic, and functional gene data. Supervised random forests were done within a group of samples identified as either coral, algae, or interface. The variable importance plots (VIPs) from the random forests were used to identify the molecules, taxa, and functional genes that best distinguished winning and losing interactions within the three groups (i.e., coral, algae, interface). The top ten variables from each VIP was used to construct two-way dendrograms for distinguishing winning and losing interactions in each sample type in JMP 14.

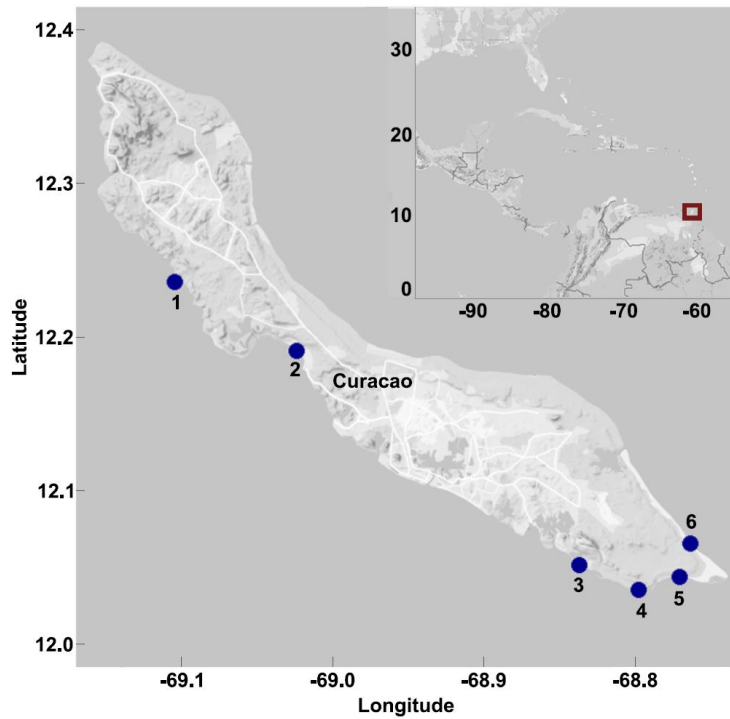


Figure S1: Map of sampling sites around the island of Curaçao. Surface-associated samples were taken at each site by suction from the surface of the coral, the algae, and the interface over a 10 cm transect parallel with the interface. Tissue samples were taken at each site using an underwater power drill taking a biopsy 1 cm in diameter and 1 cm in depth. All samples were taken from coral-algal interactions at a 10-15 m depth.

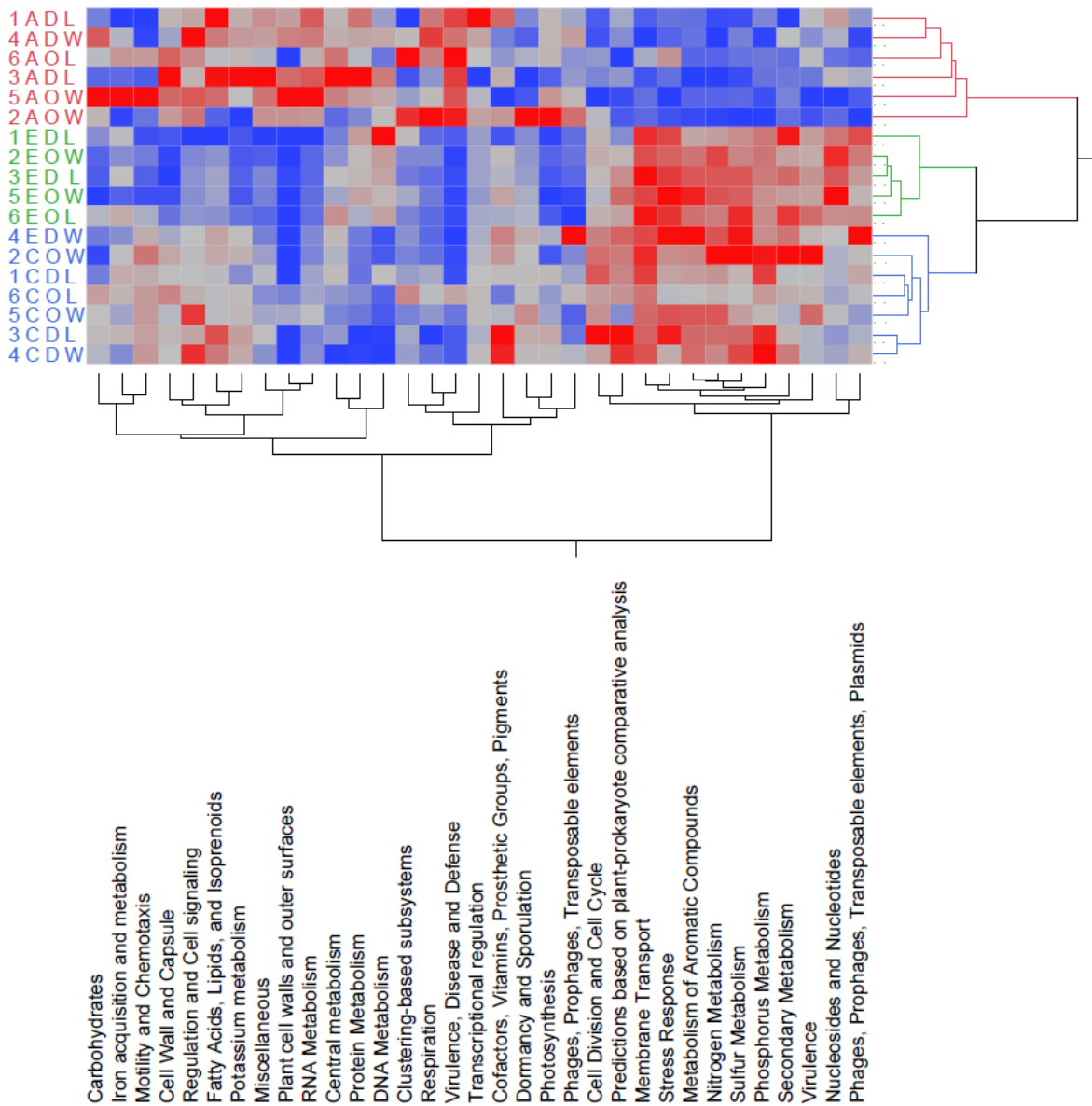


Figure S2. Two-way heat map constructed using functional genes at level 1 of the SEED hierarchical database. Green, red, and blue branches represent the three significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. stigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.

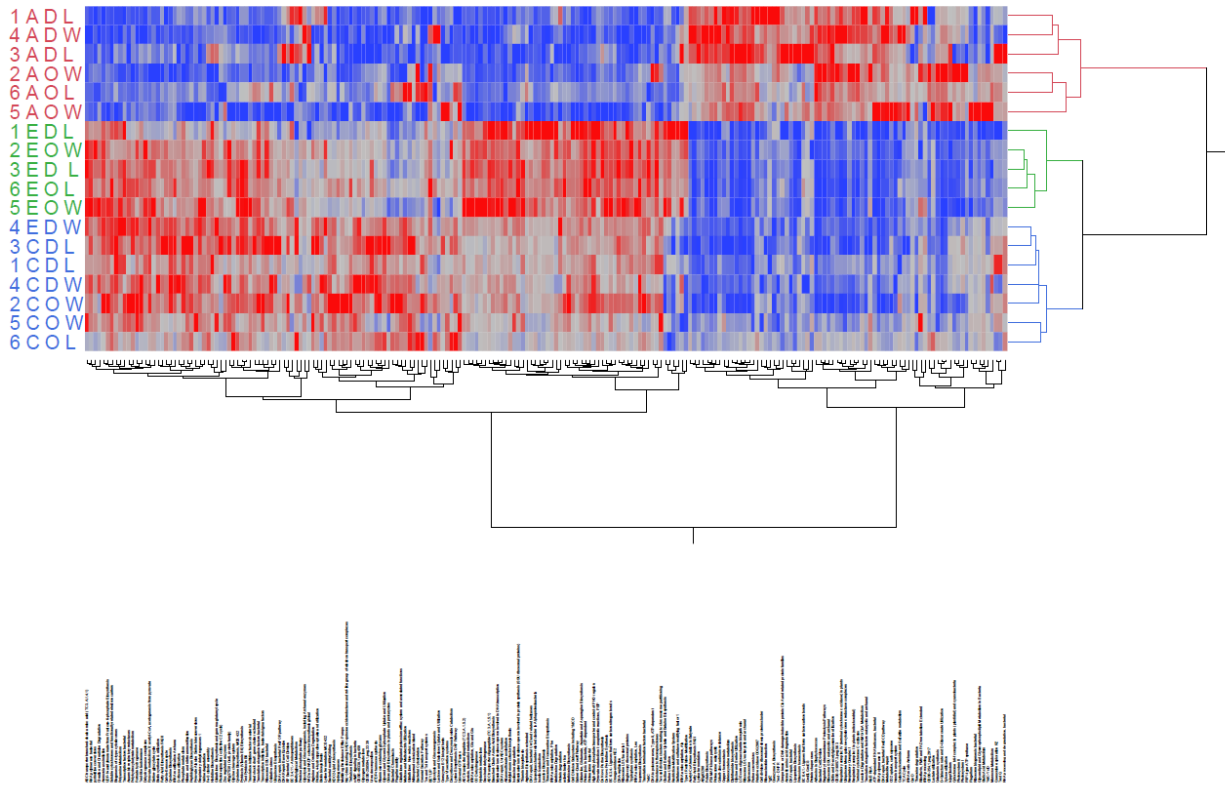


Figure S3. Two-way heat map constructed using functional genes at level 3 of the SEED hierarchical database. Green, red, and blue branches represent the three significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. stigosa*, O = *O. faveolata*), and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.

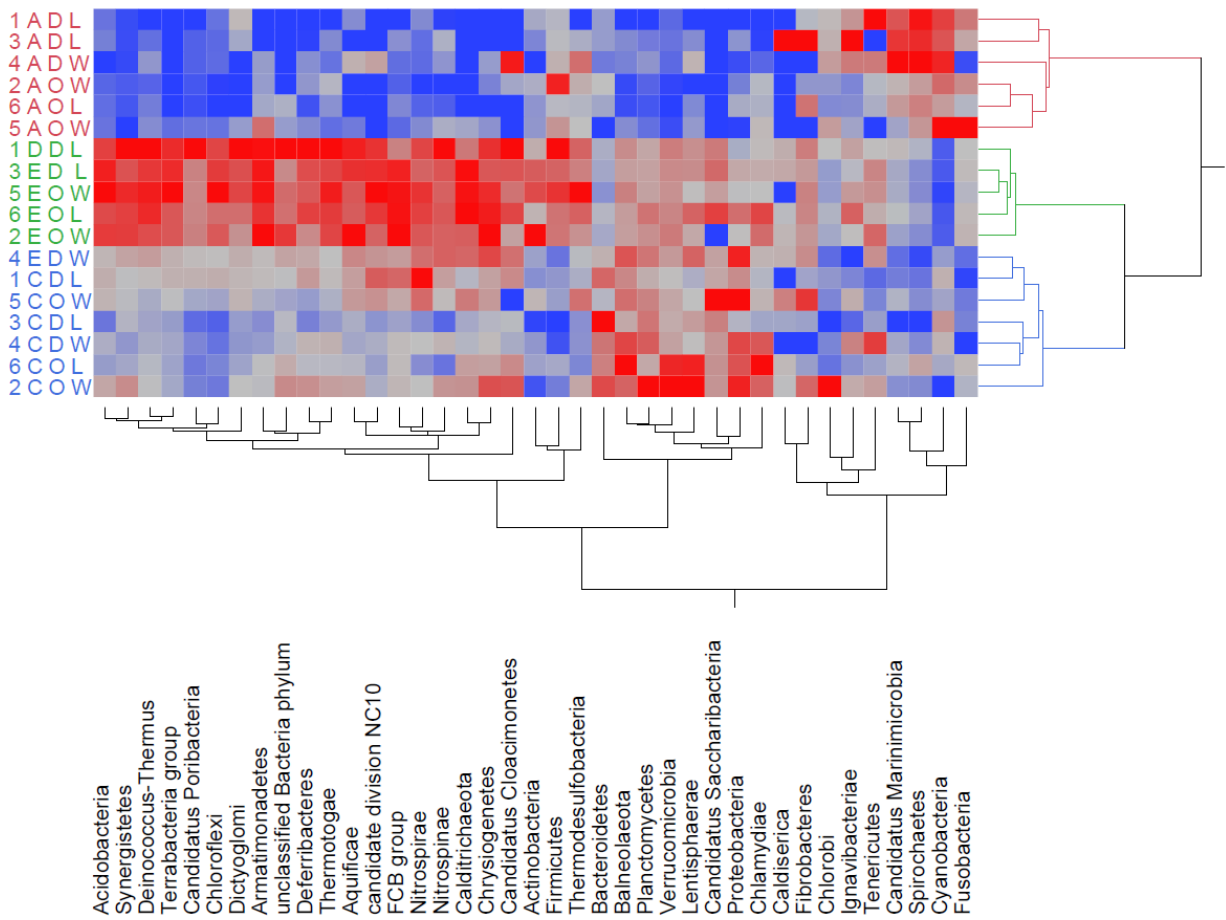


Figure S4. Two-way heat map constructed using bacterial phyla. Green, red, and blue represent branches the three significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*), and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.

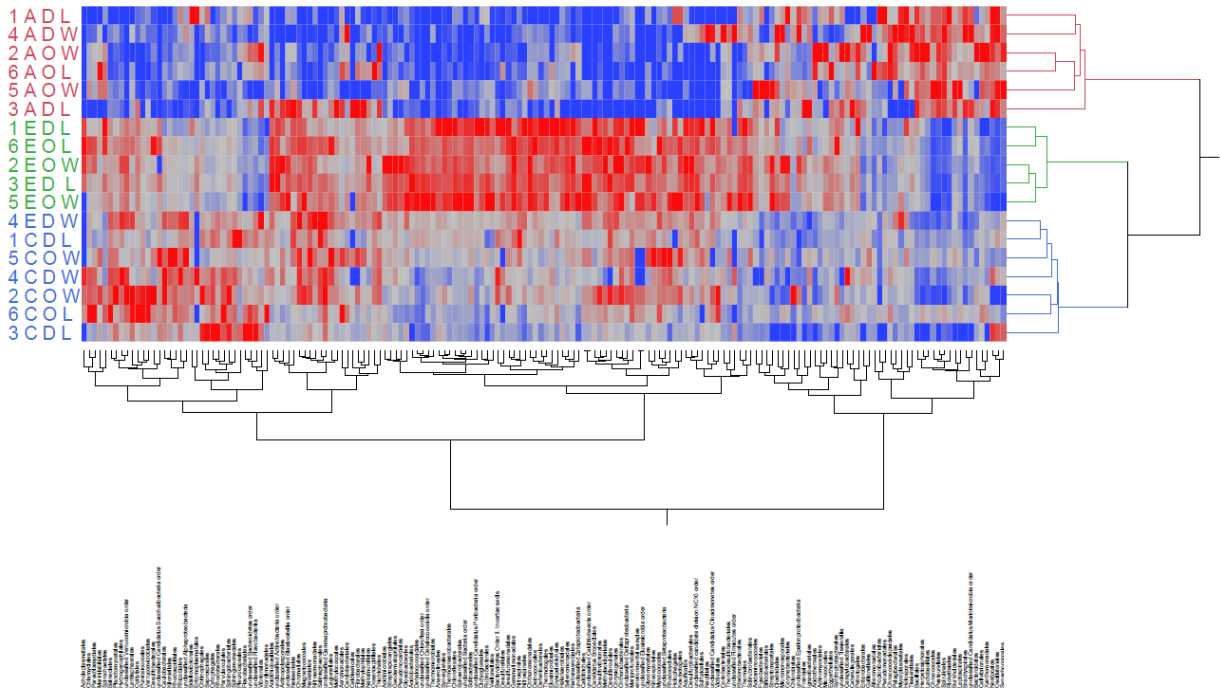


Figure S5. Two-way heat map constructed using bacterial orders. Green, red, and blue branches represent the three significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*), and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.

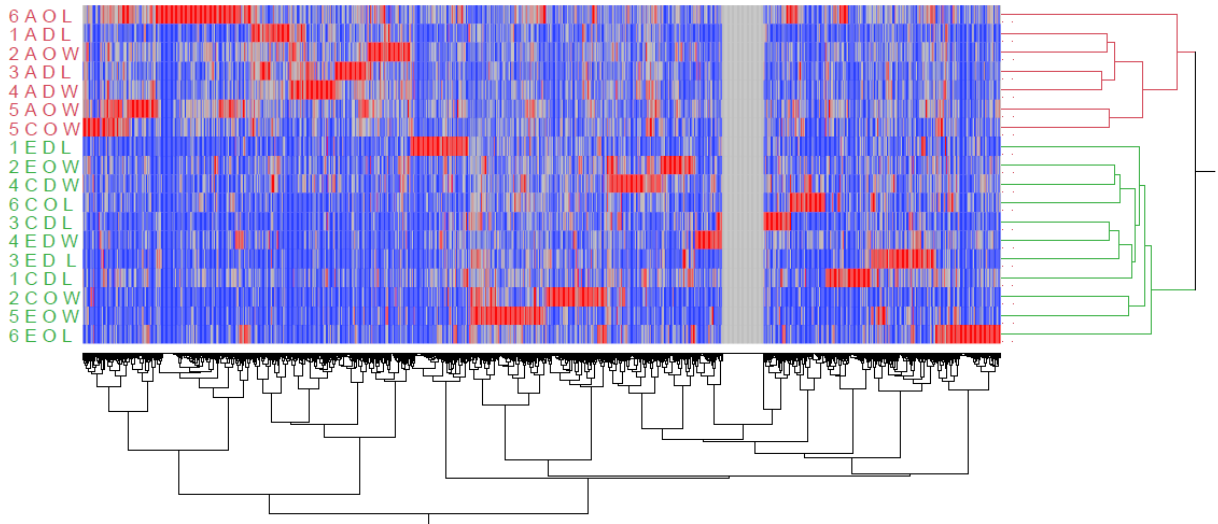


Figure S6. Two-way heat map constructed using metabolites. Green and red branches represent the two significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. stigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.

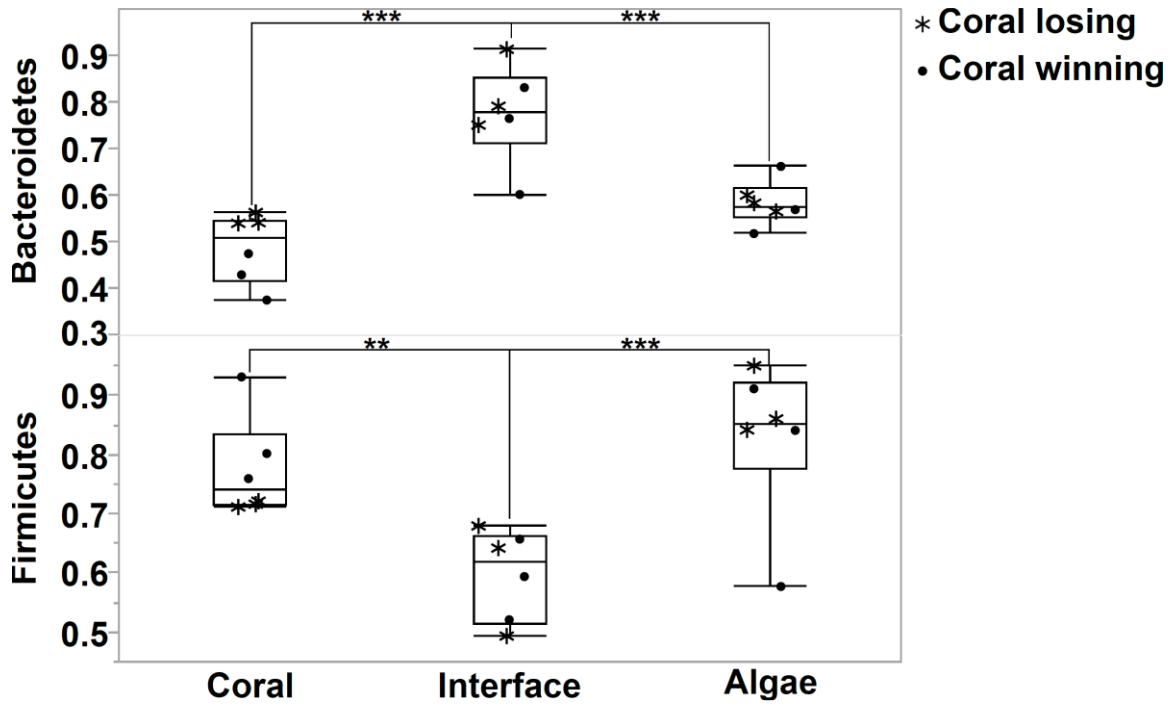


Figure S7. Box plots of the percent relative abundance of the two phyla where the interface samples were significantly different than both the coral and the algal samples. (N = 18, Tukey *post hoc* **p ≤ 0.05, ***p ≤ 0.01)

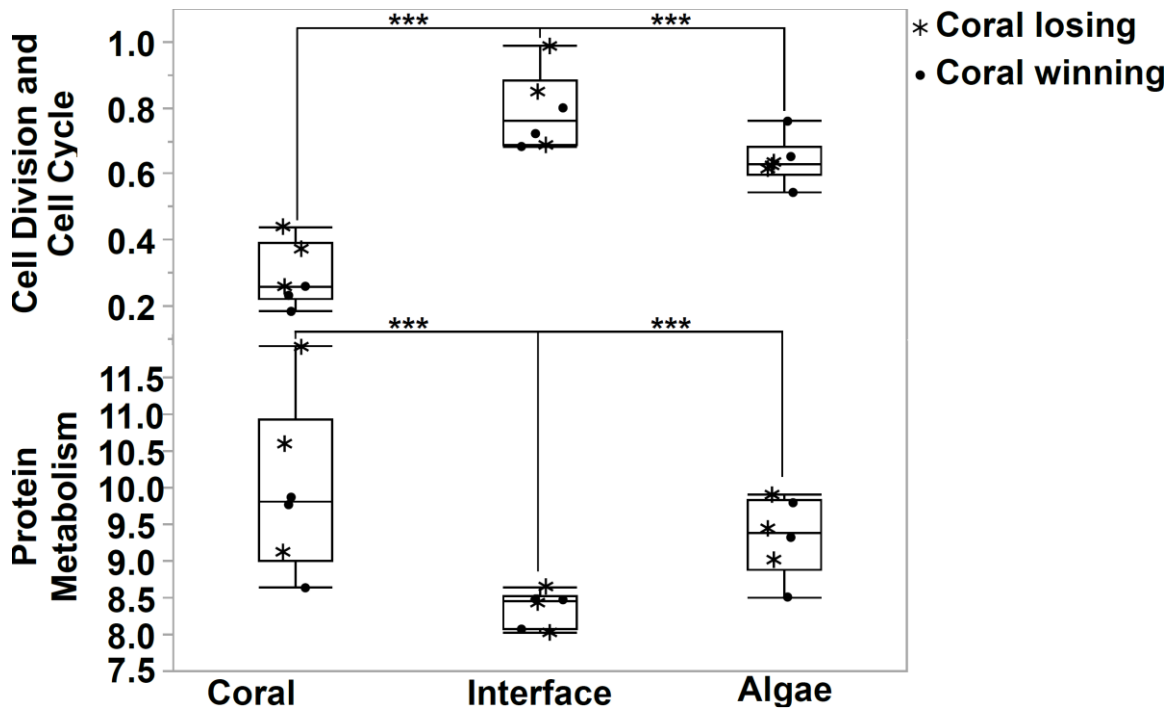


Figure S8. Box plots of the percent relative abundance of the two level 1 SEED subsystems where the interface samples were significantly different than both the coral and the algal samples. (N = 18, Tukey *post hoc* $p \leq 0.1$, $**p \leq 0.05$, $***p \leq 0.01$)

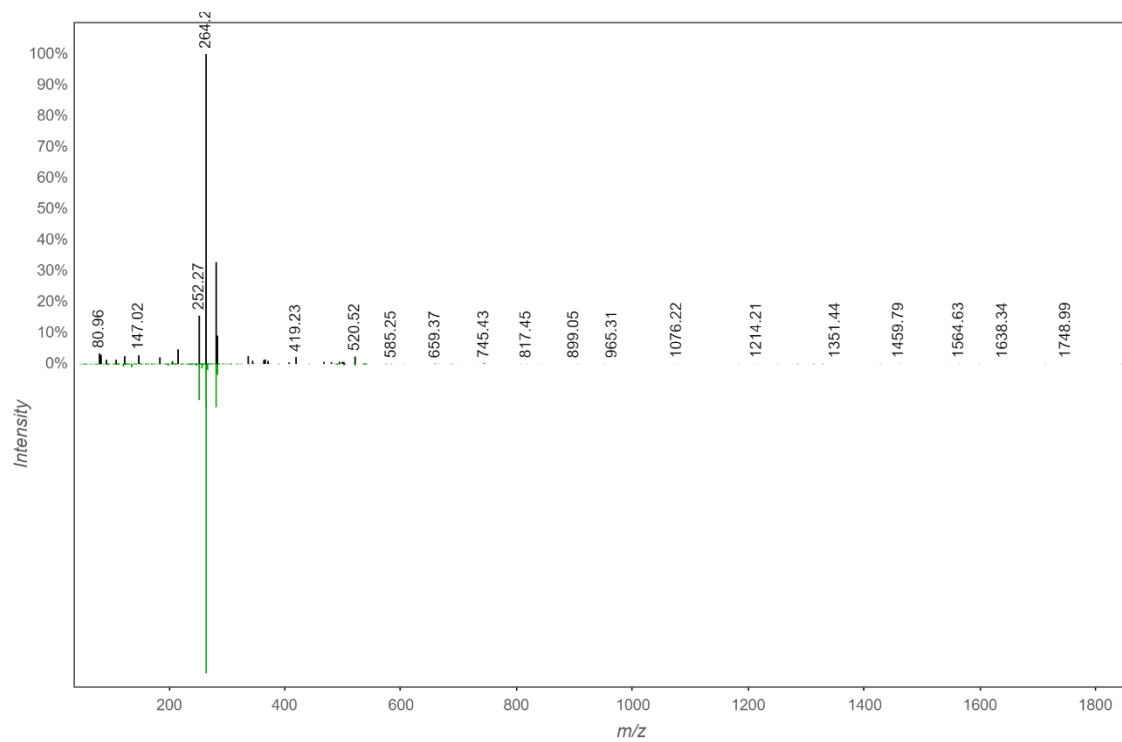


Figure S9. Mirror plot of spectral match for ceramide 18:1/16:0. The GNPS library reference is on the bottom in green with the spectra for the compound in the coral and interface samples is shown above in black. Note, the high intensity peak at 264.2 is decisive for the 18:0 backbone.

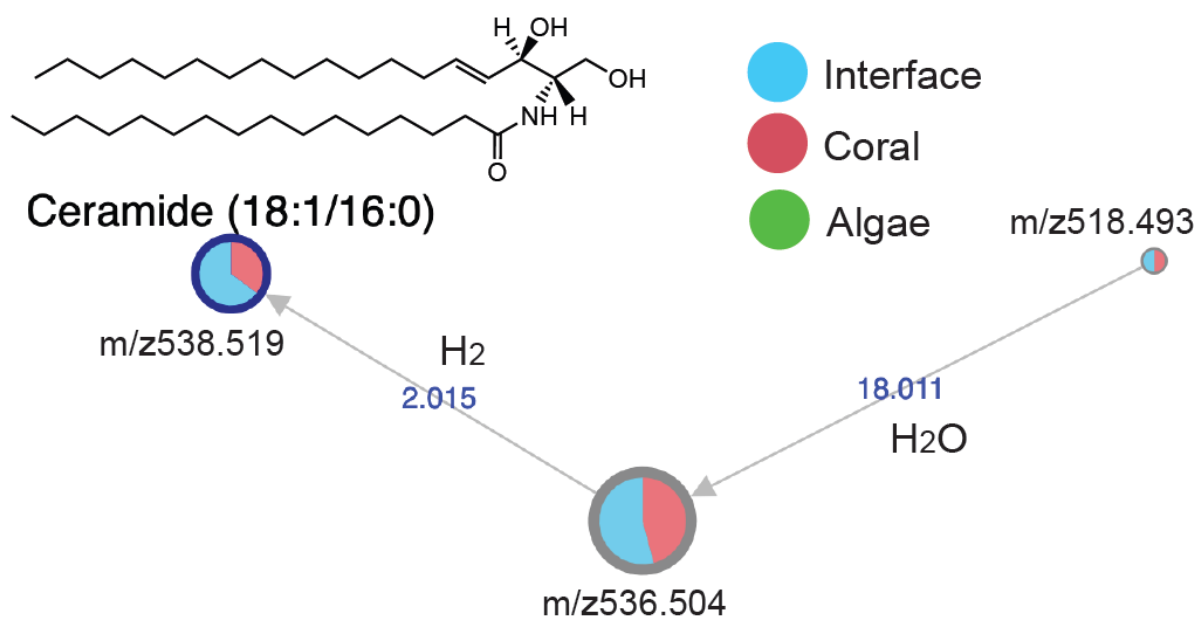


Figure S10. Molecular network of ceramides in coral, algal, and interface metabolomics data from GNPS. The structure of known ceramide is shown along with its less saturated form. Edges are labeled by the mass difference between related nodes and known biochemical transformations are highlighted. The nodes are colored in a pie chart based on the total spectral intensity between coral, algae, and interface according to the color legend.

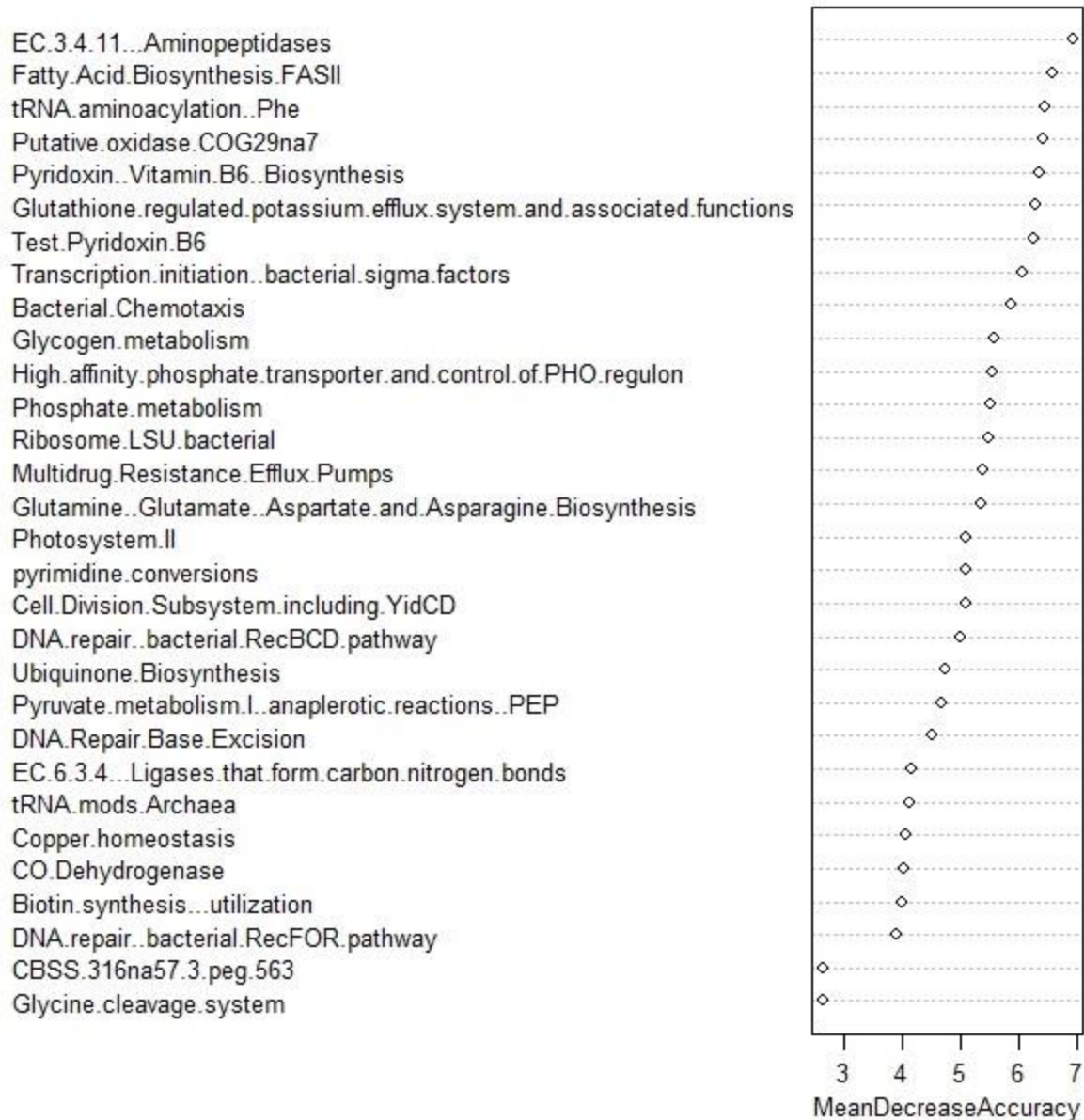


Figure S11. Variable importance plot of functional genes (SEED level 3) from random forest classification analysis based on winning and losing corals. Variable are ranked from highest to lowest according to their mean decrease in accuracy.

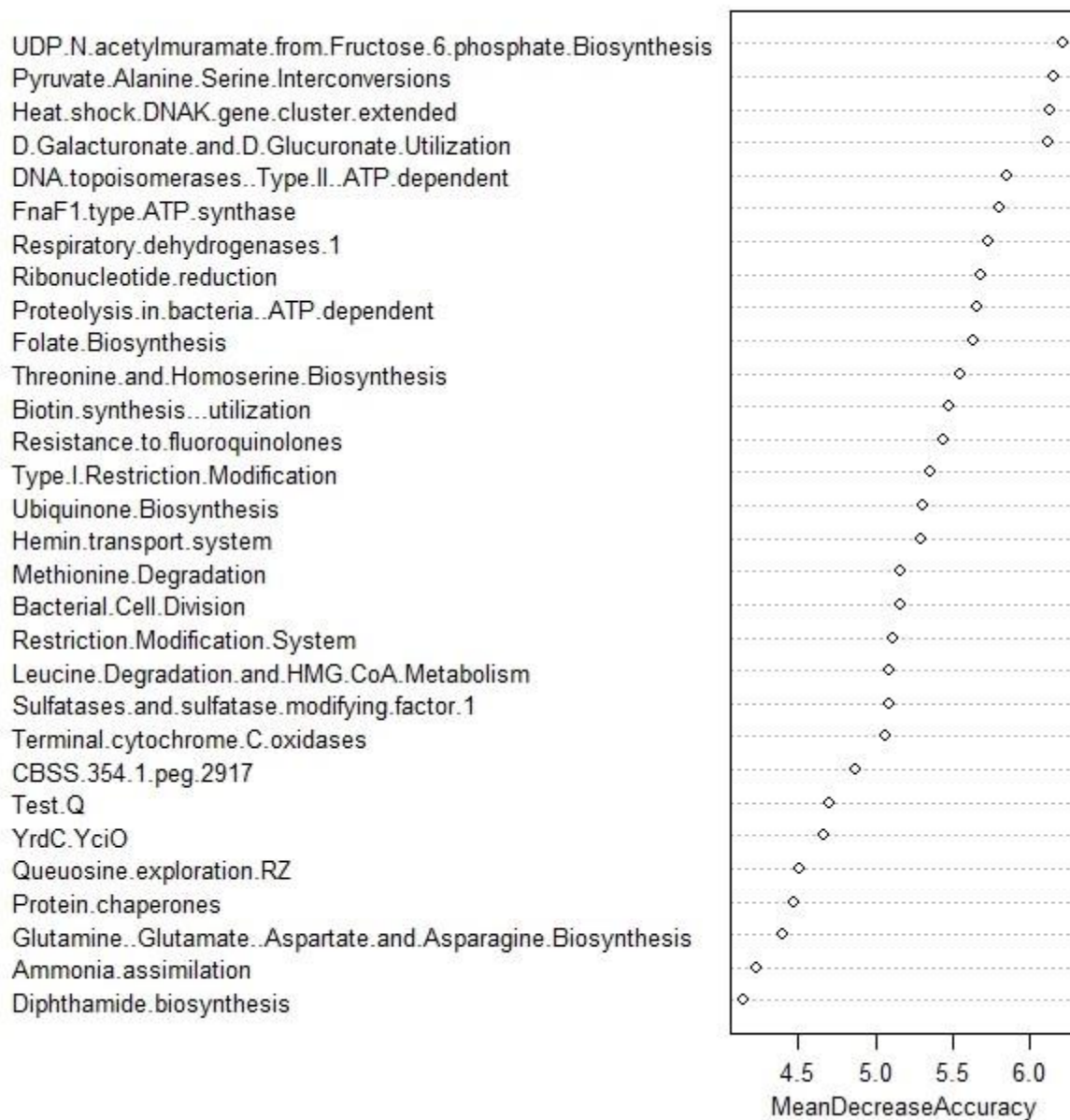


Figure S12. Variable importance plot of functional genes (SEED level 3) from random forest classification analysis based on winning and losing algae. Variable are ranked from highest to lowest according to their mean decrease in accuracy.

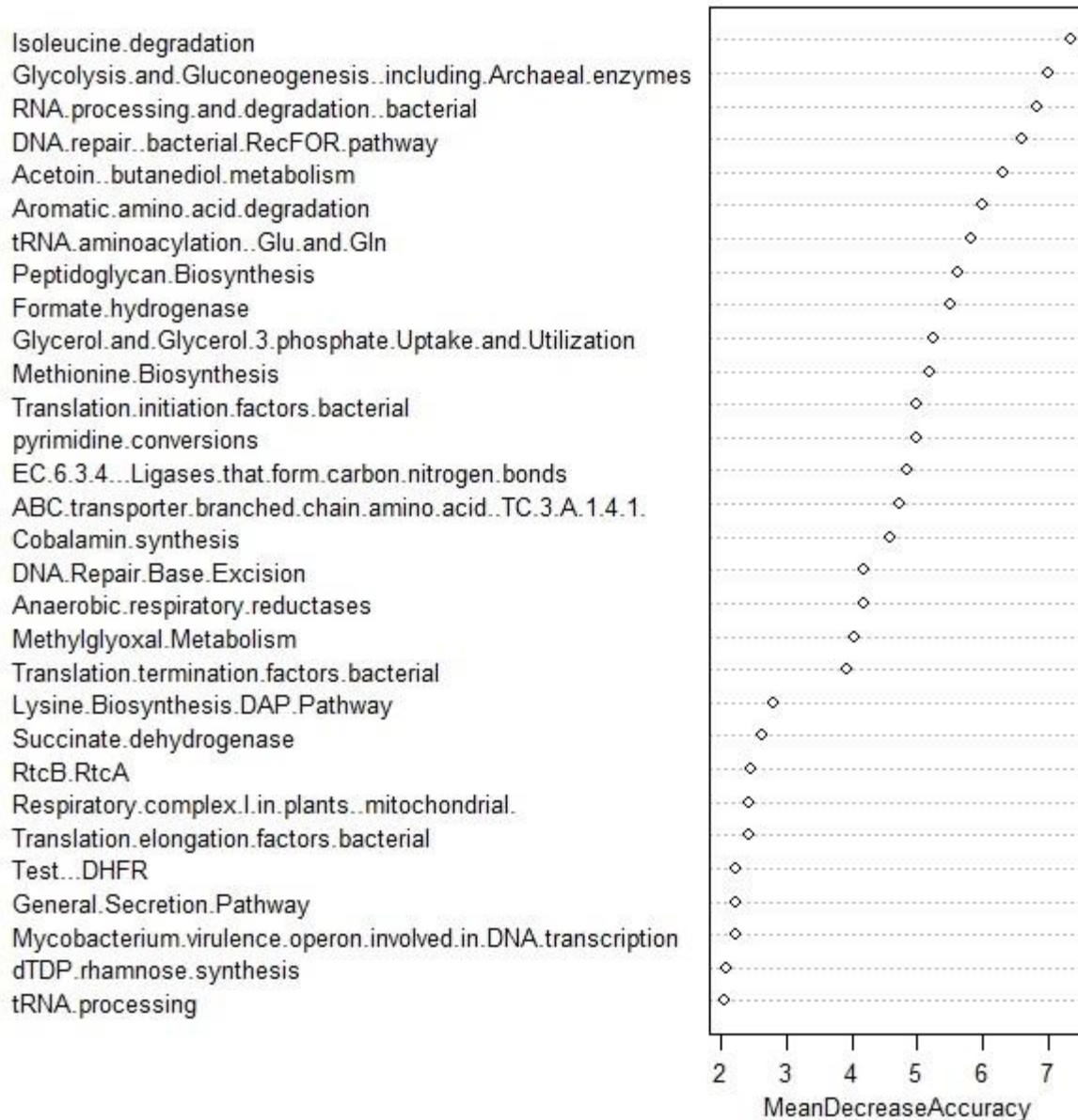


Figure S13. Variable importance plot of functional genes (SEED level 3) from random forest classification analysis based on winning and losing interfaces. Variable are ranked from highest to lowest according to their mean decrease in accuracy.

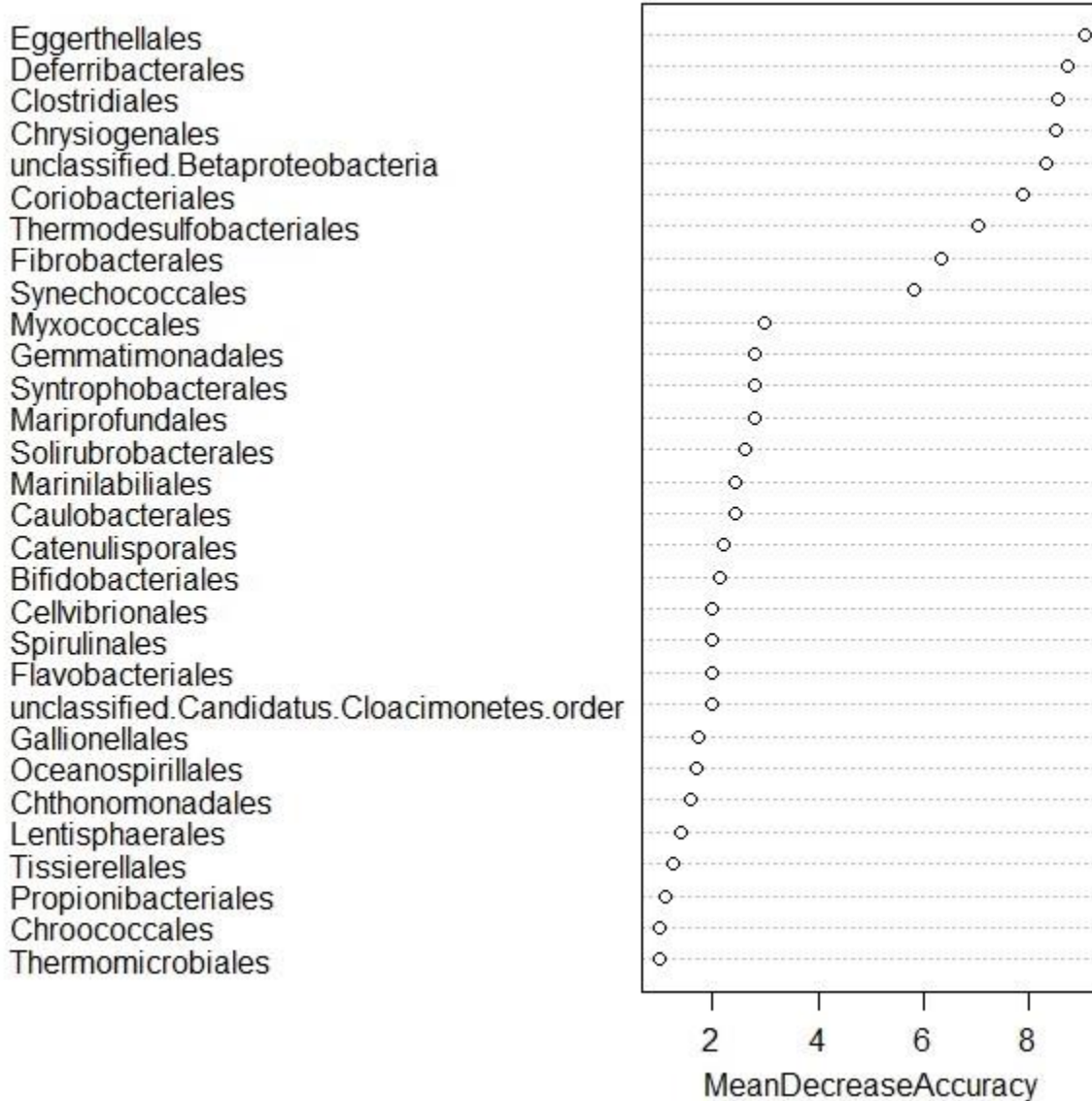


Figure S14. Variable importance plot of bacterial orders from random forest classification analysis based on winning and losing corals. Variable are ranked from highest to lowest according to their mean decrease in accuracy.

Tissierellales
 Nitrospinales
 unclassified.Candidatus.Saccharibacteria.order
 Fusobacteriales
 Bifidobacteriales
 unclassified.Actinobacteria.order
 Oceanospirillales
 Deferribacterales
 Pelagibacterales
 Desulfurobacteriales
 Desulfuromonadales
 Methyacidiphilales
 unclassified.Candidatus.Poribacteria.order
 unclassified.Alphaproteobacteria
 Spirochaetales
 Kordiimonadales
 Erysipelotrichales
 Elusimicrobiales
 Xnautiliales
 Balneolales
 Mariprofundales
 unclassified.Betaproteobacteria
 Mycoplasmatales
 Acidithiobacillales
 unclassified.Candidatus.Cloacimonetes.order
 Acholeplasmatales
 Vibrionales
 Cellvibrionales
 unclassified.Tissierellia
 Veillonellales

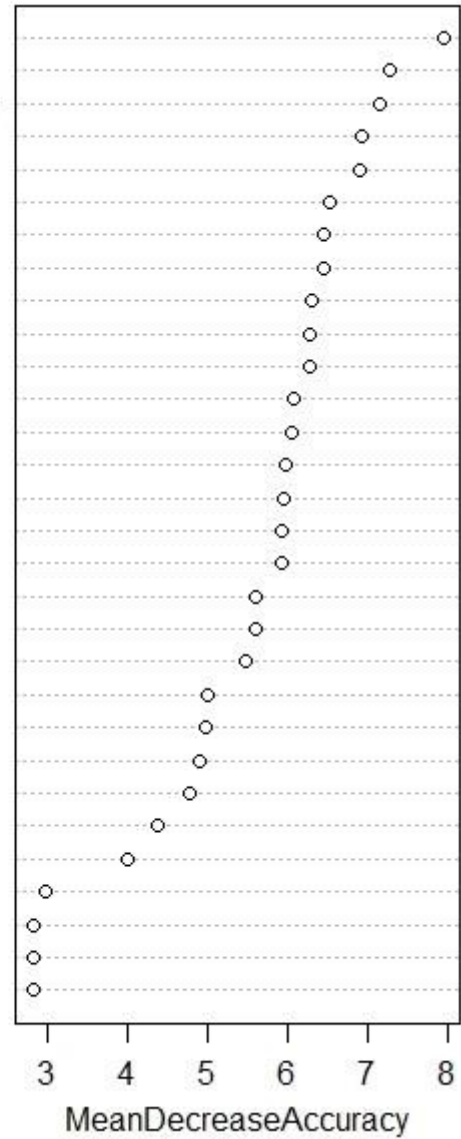


Figure S15. Variable importance plot of bacterial orders from random forest classification analysis based on winning and losing algae. Variable are ranked from highest to lowest according to their mean decrease in accuracy.

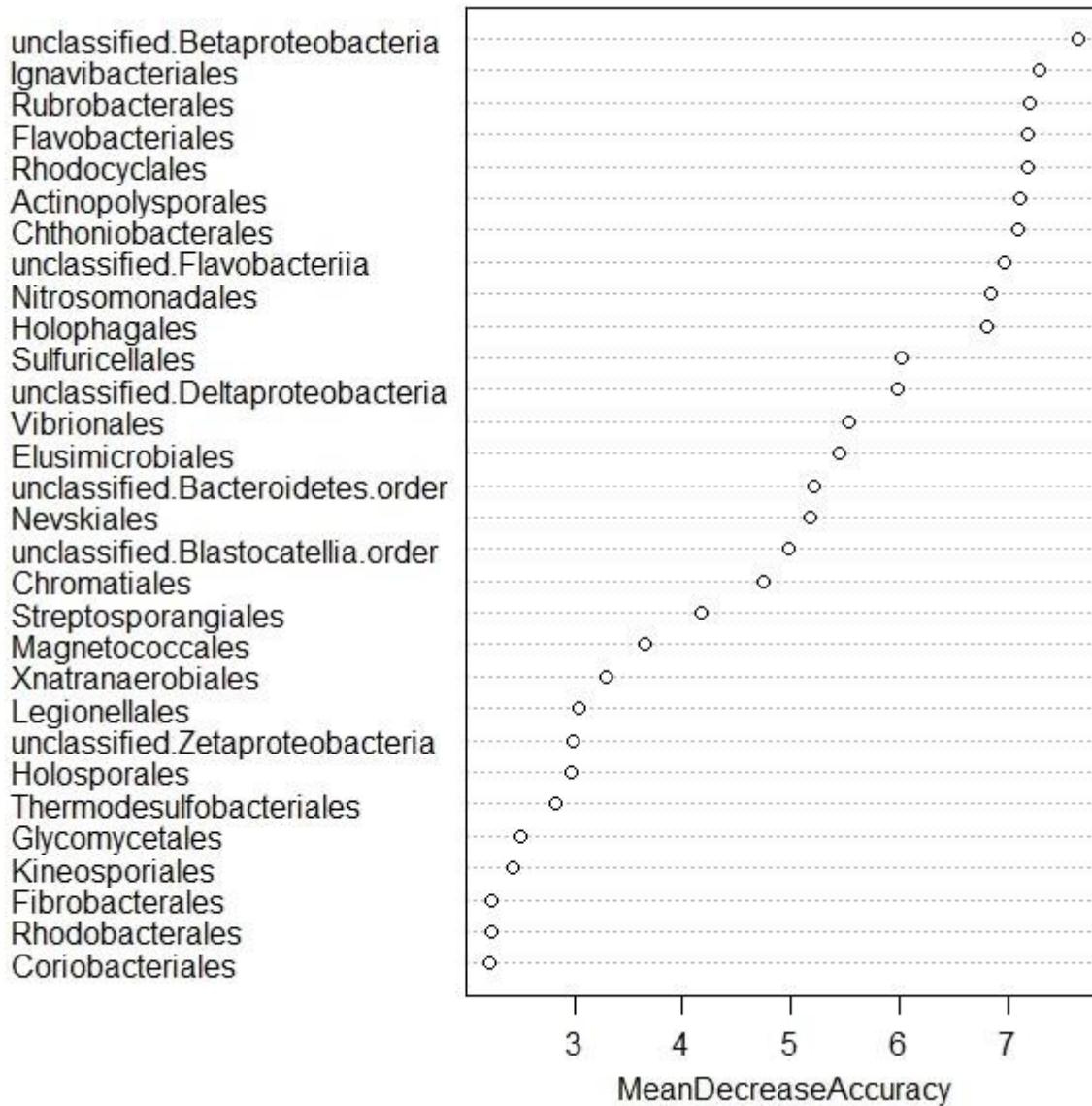


Figure S16. Variable importance plot of bacterial orders from random forest classification analysis based on winning and losing interfaces. Variable are ranked from highest to lowest according to their mean decrease in accuracy.

5116_371.2313384_7.190038456__107_C20H32N2O3
 7360_595.2630842_7.635572063__-1_C19H41N8O10P
 2197_429.2383972_6.342329167__-1_C13H32N8O8
 213_315.2281631_6.956382719_5(6)-EpETE methyl ester_61_C16H30N2O4
 7589_542.405199_5.657836407__-1_C20H51N11O6
 522_355.2232378_6.95063709__30_C18H30N2O5
 1320_180.0974364_0.344875895__87_C6H13N4O
 4169_180.0969331_0.265316042__87_C6H13N4O
 4189_497.3880564_9.584541735__377_C19H48N10O5
 5955_439.3198438_8.880832236__-1_C27H44O3
 7600_1083.811726_5.654099795__-1
 3655_673.8339512_3.483967623__-1_C11H18N5O7P11
 5916_335.2265579_6.390910314__-1_C13H28N8O
 7851_413.3391241_9.29725526__-1_C13H40N12O3
 927_409.2680391_6.906949932__-1_C17H36N4O7
 2965_639.2817381_8.496636488__173_C36H38N4O7
 7787_644.5331071_0.217858405__-1_C40H65N7
 4758_639.2843175_8.306427937__-1_C34H40N4O7
 6469_389.2569914_7.64953832__-1_C13H30N11O3
 5003_565.2946729_6.776451913__-1_C20H41N10O7P
 7688_607.2711463_7.826793443__-1_C34H34N6O5
 2535_317.2070489_6.101103825__349_C16H24N6O
 7268_385.2660327_7.775368716__51_C20H36N2O5
 2534_355.2427833_6.774946311__264_C18H36O5
 5910_305.209863_4.595807309__-1_C14H28N2O5
 3990_203.1092497_4.476542008__-1_C3H10N10O
 5802_290.6745458_0.185933019__-1
 8022_253.2140632_7.217129781__187_C12H28N3O
 419_371.2224085_6.418347131__110_C15H29N7O2P
 2167_1585.628376_4.540090632__-1

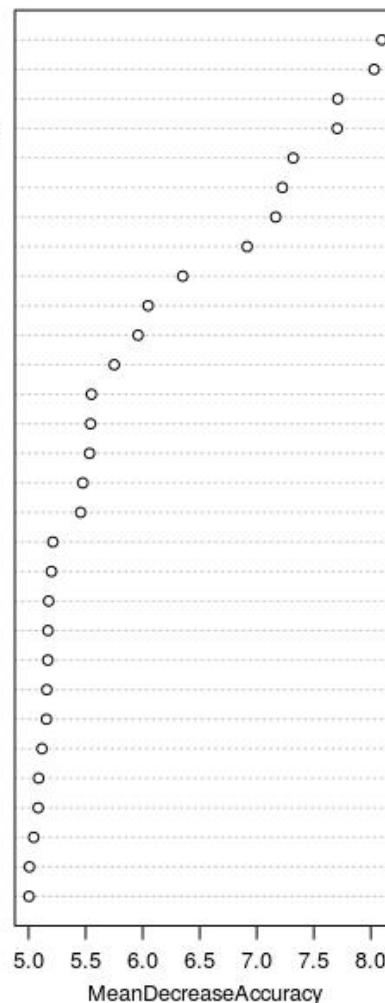


Figure S17. Variable importance plot of metabolites from random forest classification analysis based on winning and losing corals. Variable are ranked from highest to lowest according to their mean decrease in accuracy. The first number is the GNPS annotation number, the following number is the mass to charge ratio, the third number is the retention time, the fourth number is the network subcluster ID where -1 indicates a single looped compound. This is followed by the putative molecular formula for all molecules where annotation was possible.

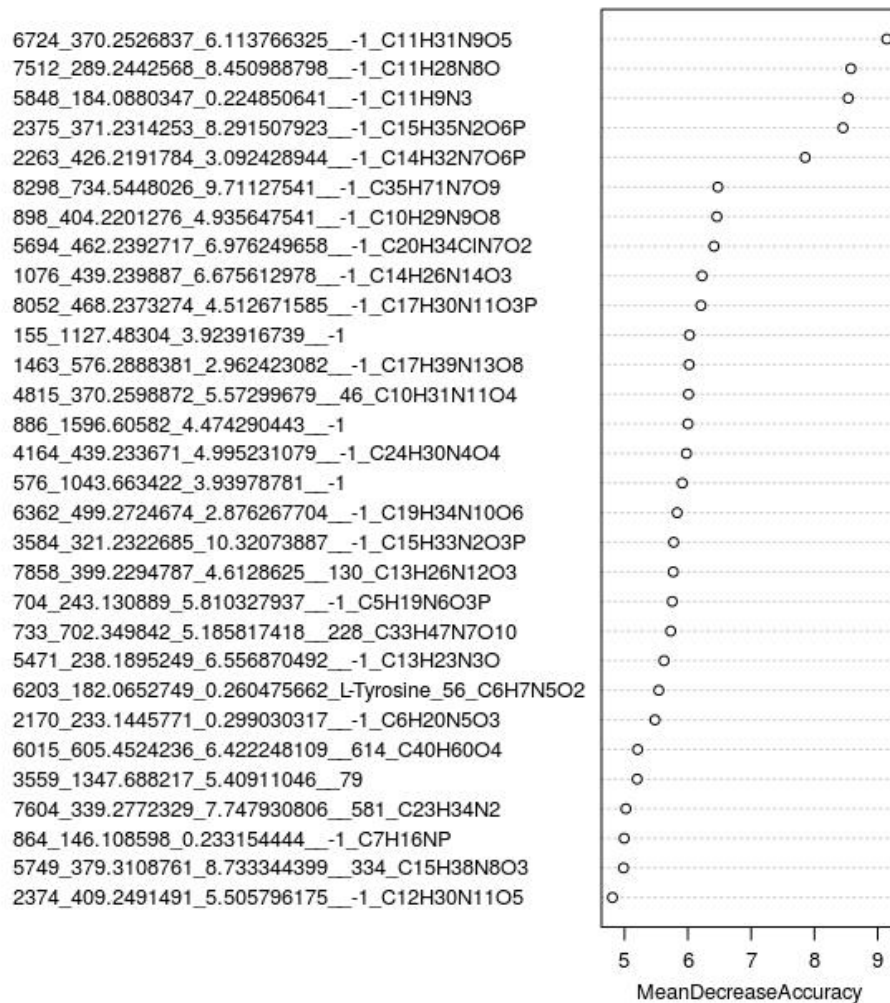


Figure S18. Variable importance plot of metabolites from random forest classification analysis based on winning and losing algae. Variable are ranked from highest to lowest according to their mean decrease in accuracy. The first number is the GNPS annotation number, the following number is the mass to charge ratio, the third number is the retention time, the fourth number is the network subcluster ID where -1 indicates a single looped compound. This is followed by the putative molecular formula for all molecules where annotation was possible.

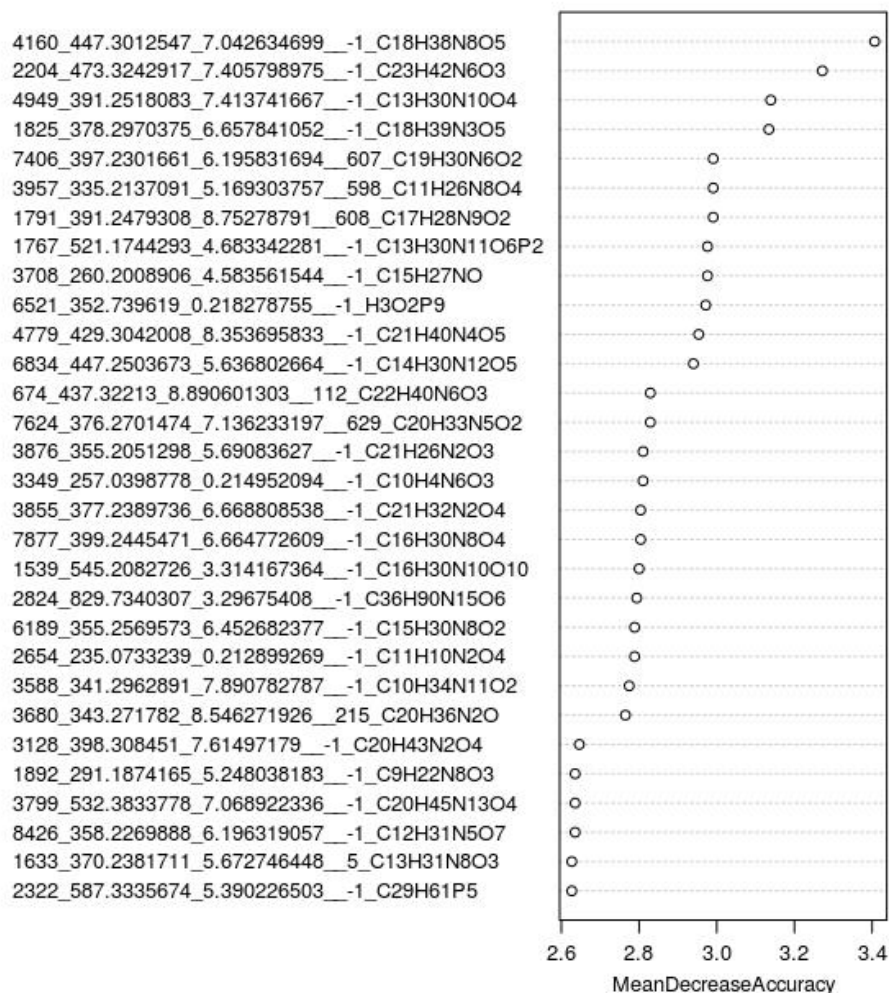


Figure S19. Variable importance plot of metabolites from random forest classification analysis based on winning and losing interfaces. Variable are ranked from highest to lowest according to their mean decrease in accuracy. The first number is the GNPS annotation number, the following number is the mass to charge ratio, the third number is the retention time, the fourth number is the network subcluster ID where -1 indicates a single looped compound. This is followed by the putative molecular formula for all molecules where annotation was possible.

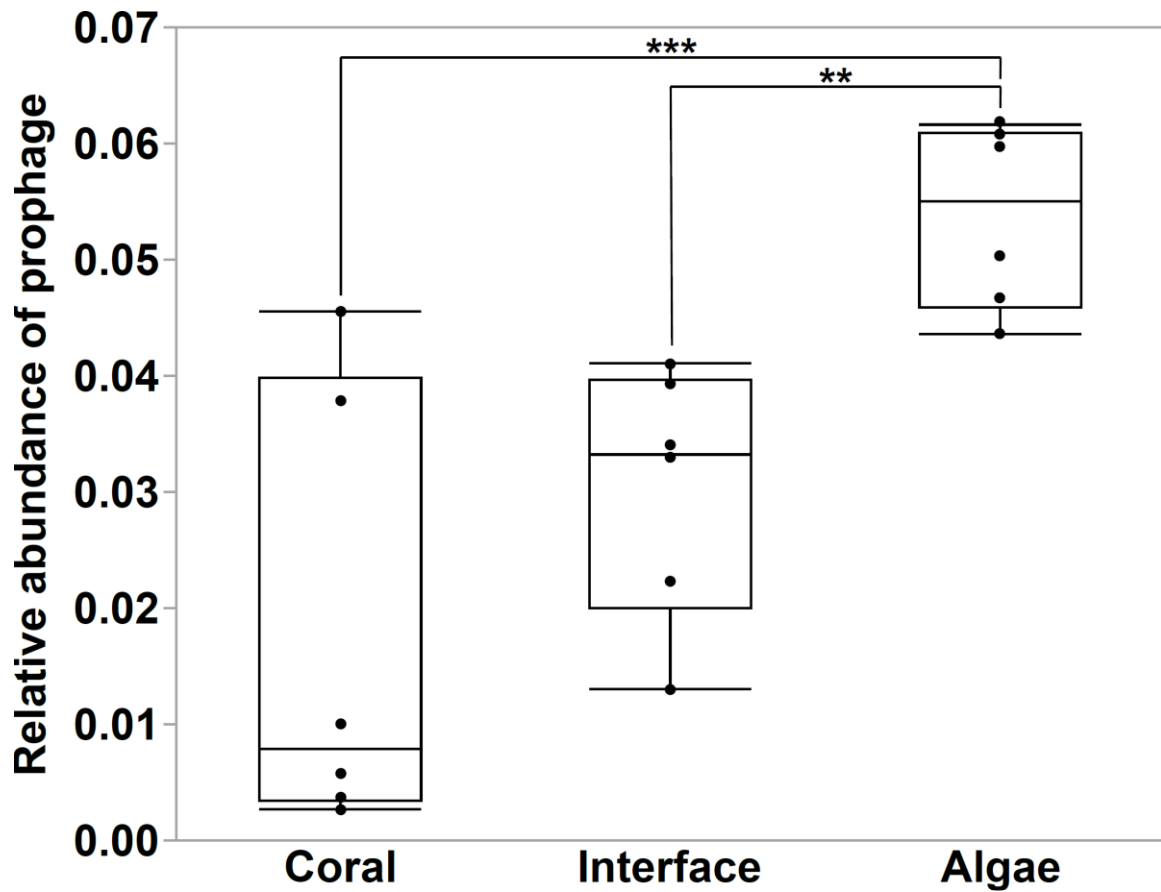


Figure S20. Relative abundance of prophages in metagenomes. (n = 18, **P ≤ 0.05, ***P ≤ 0.01)

Table S1: Sample metadata.

Sample Site	Latitude	Longitude	Coral	Algae	Competition type
1	12.06529	-68.7606812	<i>D.strigosa</i>	Turf	Coral losing
2	12.0431461	-68.767952	<i>O. faveolata</i>	Turf	Coral winning
3	12.035203	-68.795311	<i>D.strigosa</i>	Turf	Coral losing
4	12.050662	-68.8342896	<i>D.strigosa</i>	Turf	Coral winning
5	12.1903019	-69.0226288	<i>O. faveolata</i>	Turf	Coral winning
6	12.2352629	-69.1032791	<i>O. faveolata</i>	Turf	Coral losing

Table S2: Site-level water chemistry data. Site-level data including phosphate, nitrite, nitrite + nitrate, ammonium, and dissolved organic carbon (DOC) concentrations reported in micromolar (μM). Note, there were no significant differences between coral species, or between competition outcomes for any of the site level variables recorded.

Site	Coral type	Phosphate (μM)	Nitrite (μM)	Nitrite + Nitrate (μM)	Ammonium (μM)	DOC (μM)
1	<i>D. strigosa</i> losing	0.04495	0.17145	0.3715	1.029255	69.66
2	<i>O. faveolata</i> winning	0.0955	0.1813	0.466	0.33099	67.96
3	<i>D. strigosa</i> losing	0.13575	0.8605	9.22	1.1420925	79.08
4	<i>D. strigosa</i> winning	0.010011	0.08565	0.4925	0.934164	72.89
5	<i>O. faveolata</i> winning	0.0546	0.179	0.342	2.325	91.05
6	<i>O. faveolata</i> losing	0.123	0.1828	1.325	8.18537	
	ANOVA p-value for winning and losing	0.2722	0.3276	0.3168	0.4075	0.7828
	ANOVA p-value for coral species	0.5526	0.4785	0.491	0.3356	0.5883

Table S3: Metagenomic library details. Just as in the main text, sample names are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L).

Sample name	# of reads pre QC	# of reads post QC	Host (%)	Host (reads)	Bacteria (%)	Bacteria (reads)	Protist (%)	Protist (reads)	Archaea (%)	Archaea (reads)	Virus (%)	Virus (reads)	Unclassified (%)	Unclassified (reads)
1 C D L	1465793	1419816	98.360078	1441755	0.822430512	11677	0.309142045	4389	0.003168882	45	0.162685161	2310	0.342495077	4863
1 I D L	793526	758160	93.865282	744845	4.866123246	36893	0.268904149	2039	0.067390888	511	0.100276438	760	0.832023416	6308
1 A D L	230067	223401	74.958626	172455	23.56703864	52649	0.198268001	443	0.626580589	1400	0.028486241	64	0.621000057	1387
2 C O W	2572816	2501413	97.780174	2515704	0.931913283	23311	0.288824359	7225	0.012472415	312	0.454701441	11374	0.531914894	13305
2 I O W	653428	619257	86.592579	565820	12.15182065	75251	0.242190361	1500	0.216195263	1339	0.034545986	214	0.762669076	4723
2 A O W	965669	928129	62.153498	600197	36.21737927	336144	0.193882033	1799	0.570982588	5299	0.057399135	533	0.806859345	7489
3 C D L	1063274	1058088	98.043386	1042470	1.05662289	11180	0.385850215	4083	0.004347076	46	0.044718719	473	0.46507468	4921
3 I D L	879219	852962	71.925365	632381	26.49121532	225960	0.40033795	3415	0.061175764	522	0.053971793	460	1.067933677	9109
3 A D L	1088614	1050354	69.91405	761094	28.58255407	300218	0.23912003	2512	0.517934747	5440	0.029307445	308	0.717033287	7531
4 C D W	1656496	1610586	97.905338	1621798	0.802130405	12919	0.3141605	5060	0.003289964	53	0.448846928	7229	0.526234329	8475
4 I D W	429451	416244	91.112385	391283	7.135718473	29702	0.30433754	1267	0.151568262	631	0.09425388	392	1.201736964	5002
4 A D W	661053	644422	84.79583	560545	13.30339436	85730	0.397759934	2563	0.287926068	1855	0.298524867	1924	0.916564632	5907
5 C O W	1532535	1511462	98.335849	1507031	0.954373977	14425	0.309591059	4679	0.005093699	77	0.06931448	1048	0.325778055	4924
5 I O W	900394	874791	95.852206	863048	3.009175906	26324	0.279785127	2448	0.065031755	569	0.0569714	498	0.736830111	6446
5 A O W	144160	141209	64.66567	93222	33.55593482	47384	0.19826254	280	0.898554152	1269	0.044313146	63	0.637265246	900
6 C O L	1628373	1572820	98.262022	1600072	1.028725474	16180	0.279827568	4401	0.003051277	48	0.191559043	3013	0.234814311	3693
6 I O L	2141143	2072880	96.768028	2071942	2.207025974	45749	0.298195918	6181	0.007812993	162	0.059000918	1223	0.659936192	13680
6 A O L	582502	563257	74.603599	434567	24.48047315	137888	0.153719635	866	0.575649857	3242	0.006745203	38	0.179812922	1013
TOTAL	19388513	18819251		17620231		1489584		55149		22820		31923		109676
MEAN	1077139.611	1045513.944	86.43855	978901.7	12.286892	82755	0.2812311	3064	0.2265681	1268	0.1242012	1774	0.642554237	6093

Table S4: Microbial taxa that are significantly enriched in winning and losing interactions.
 Prokaryotic orders of bacteria that were significantly enriched as determined ANOVA are listed in the first column with the ANOVA p-value listed in the second column.

Enriched in winning corals	p-value (one-way ANOVA)	Notes
Deferribacterales	0.0004	Obligate anaerobes
Chrysiogenales	0.0006	Anoxic chemoautotrophs
Clostridiales	0.0131	Obligate anaerobes
Synechococcales	0.0210	Cyanobacteria
Enriched in losing corals	p-value (one-way ANOVA)	Notes
Unclassified Betaproteobacteria	0.0006	Mostly anaerobic
Fibrobacterales	0.0251	Super heterotroph
Eggerthellales	0.0324	Mostly anaerobic
Enriched in winning interfaces	p-value (one-way ANOVA)	Notes
Rubrobacterales	0.0083	
Ignavibacterales	0.0084	Anaerobic chemoautotrophs
Actinopolysporales	0.0139	
Unclassified Betaproteobacteria	0.0170	
Rhodocyclales	0.0216	Denitriphyers
Nitrosomonadales	0.0251	Chemoautotrophs
Chromatiales	0.0275	Anaerobic purple sulfur bacteria
Unclassified Deltaproteobacteria	0.0279	
Chthoniobacterales	0.0281	
Holophagales	0.0377	Anaerobes
Elusimicrobiales	0.0494	Obligate anaerobes
Enriched in losing interfaces	p-value (one-way ANOVA)	Notes
Unclassified Flavobacteriia	0.0126	Opportunistic pathogens
Flavobacteriales	0.0165	Opportunistic pathogens
Unclassified Bacteroidetes	0.0346	
Vibrionales	0.0430	Opportunistic pathogens
Enriched in algae when coral is winning	p-value (one-way ANOVA)	Notes
Balneodales	0.0235	
Oceanospirillales	0.0351	
Enriched in algae when coral is losing	p-value (one-way ANOVA)	Notes
Erysipelotrichales	0.0147	
Methylacidiphilales	0.0166	
Candidatus Saccharibacteria	0.0248	
Tissierellales	0.0362	
Unclassified Betaproteobacteria	0.0414	
Unclassified Actinobacteria	0.0432	
Elusimicrobiales	0.0488	

Table S5: Functional genes that are significantly enriched in winning and losing interactions. Functional genes from Level 3 of the SEED database that were significantly enriched as determined by ANOVA are listed in the first column with the ANOVA p-value listed in the second column.

Enriched in winning corals	p-value (one-way ANOVA)	Notes
DNA repair, bacterial RecBCD pathway	0.0381	
Enriched in losing corals	p-value (one-way ANOVA)	Notes
tRNA Aminoacylation, Phe	0.0002	
High affinity phosphate transporter and control of PHO	0.0010	
Putative oxidase COG2907	0.0021	
EC 3.4.11.-Aminopeptidases	0.0036	
Fatty acid biosynthesis FASII	0.0050	
Phosphate metabolism	0.0106	
Test Pyridoxin B6	0.0174	
Pyridoxin (B6) biosynthesis	0.0185	
Biotin synthesis and utilization	0.0411	
Enriched in winning interfaces	p-value (one-way ANOVA)	Notes
Glycolysis and gluconeogenesis	0.0008	
Isoleucine degradation	0.0060	
Acetoin and butenediol metabolism	0.0081	
Methionine biosynthesis	0.0143	
ABC transporter branched chain amino acid (TC 3.A.1.4.1)	0.0289	
Aromatic amino acid degradation	0.0433	
DNA repair base excision	0.0461	
Enriched in losing interfaces	p-value (one-way ANOVA)	Notes
RNA processing and degradation, bacterial	0.0247	
Peptidoglycan synthesis	0.0440	
Methylglyoxal metabolism	0.048	
Enriched in algae when coral is winning	p-value (one-way ANOVA)	Notes
Folate biosynthesis	0.0030	
UDP-N-acetylmuramate from fructose-6-phosphate	0.0045	
CBSS-354.1.peg.2917	0.0164	
pyruvate alanine serine interconversions	0.0272	
Bacterial cell division	0.0335	
respiratory dehydrogenase 1	0.0384	
YrdC-YciO	0.0456	
Ribonucleotide reduction	0.0470	

References

1. T. McDole, *et al.*, Assessing coral reefs on a Pacific-wide scale using the microbialization score. *PLoS One* **7**, e43233 (2012).
2. R. Schmieder, R. Edwards, Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
3. G. G. Z. Silva, K. T. Green, B. E. Dutilh, R. A. Edwards, SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* **32**, 354–361 (2016).
4. R. Overbeek, *et al.*, The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42**, D206–14 (2014).
5. H. Ponstingl, Z. Ning, SMALT-a new mapper for DNA sequencing reads. *F1000 Posters* **1** (2010).
6. S. Akhter, R. K. Aziz, R. A. Edwards, PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* **40**, e126 (2012).
7. K. D. Pruitt, T. Tatusova, D. R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–5 (2007).
8. T. Pluskal, S. Castillo, A. Villar-Briones, M. Oresic, MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395 (2010).
9. F. Olivon, G. Grelier, F. Roussi, M. Litaudon, D. Touboul, MZmine 2 data-preprocessing to enhance molecular networking reliability. *Anal. Chem.* **89**, 7836–7840 (2017).
10. E. B. Graham, *et al.*, Carbon inputs from riparian vegetation limit oxidation of physically bound organic carbon via biochemical and thermodynamic processes. *Journal of Geophysical Research: Biogeosciences* **122**, 3188–3205 (2017).
11. D. E. LaRowe, P. Van Cappellen, Degradation of natural organic matter: A thermodynamic analysis. *Geochim. Cosmochim. Acta* **75**, 2030–2042 (2011).
12. E. Archer, rfPermute: estimate permutation p-values for Random Forest importance metrics. *R package (Zenodo)*, Version, 2(1) (2016).