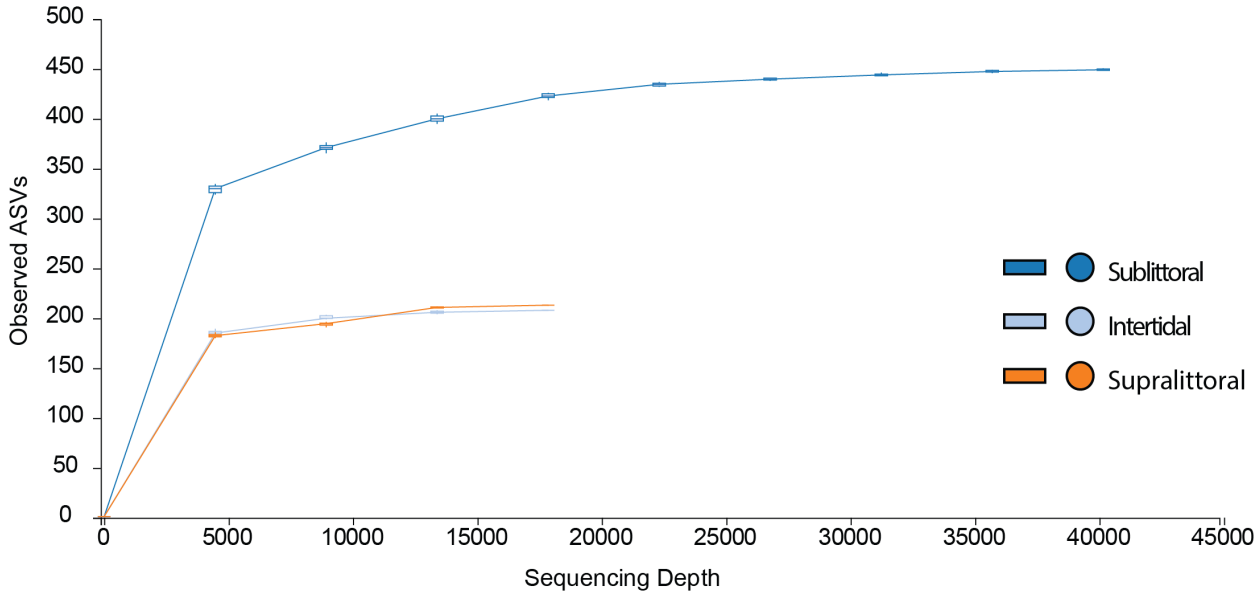
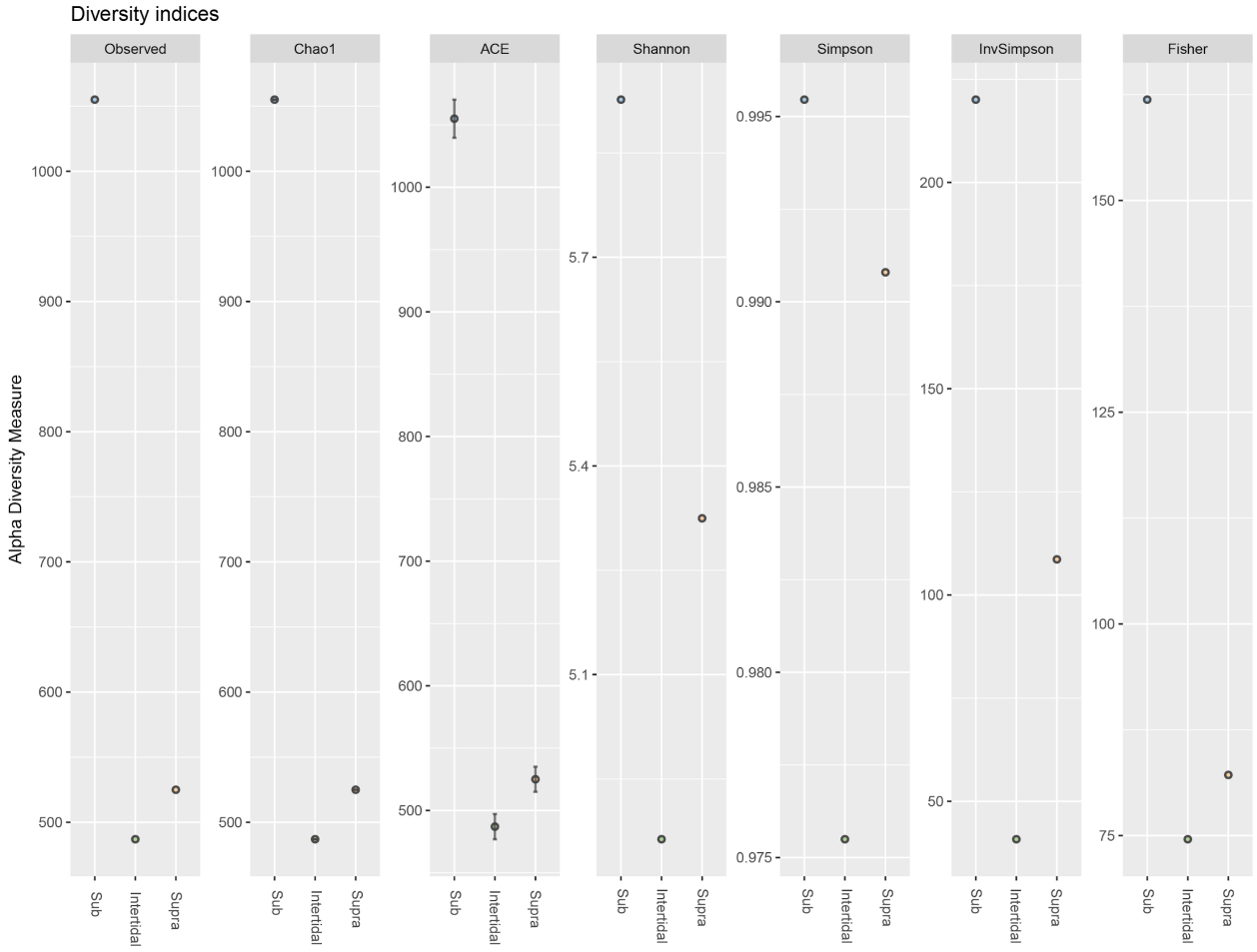
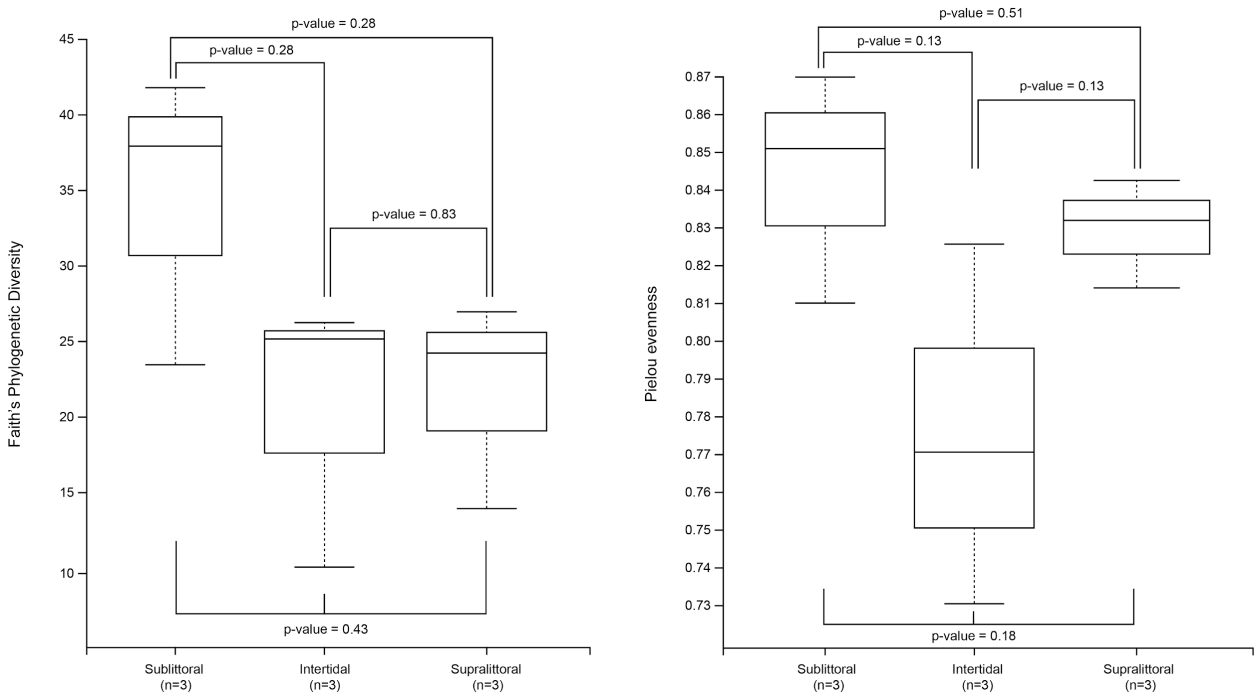
Supplemental Materials



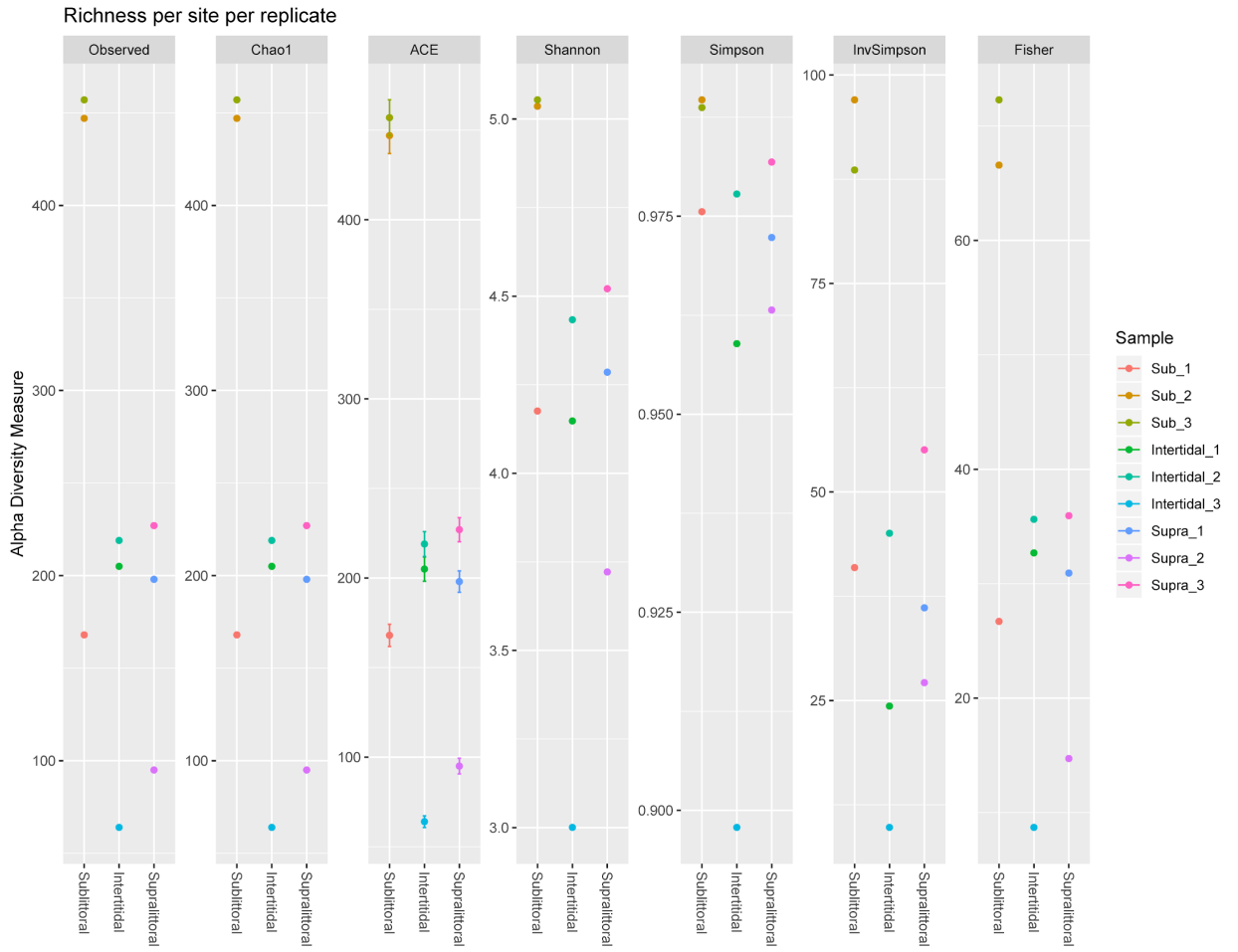
Supplemental Figure 1. Alpha-rarefaction. Progressive downsampling of sequences grouped by sampling site reveals early saturation for Intertidal and Supralittoral sediment sites but slow and incremental growth for the Sublittoral site.



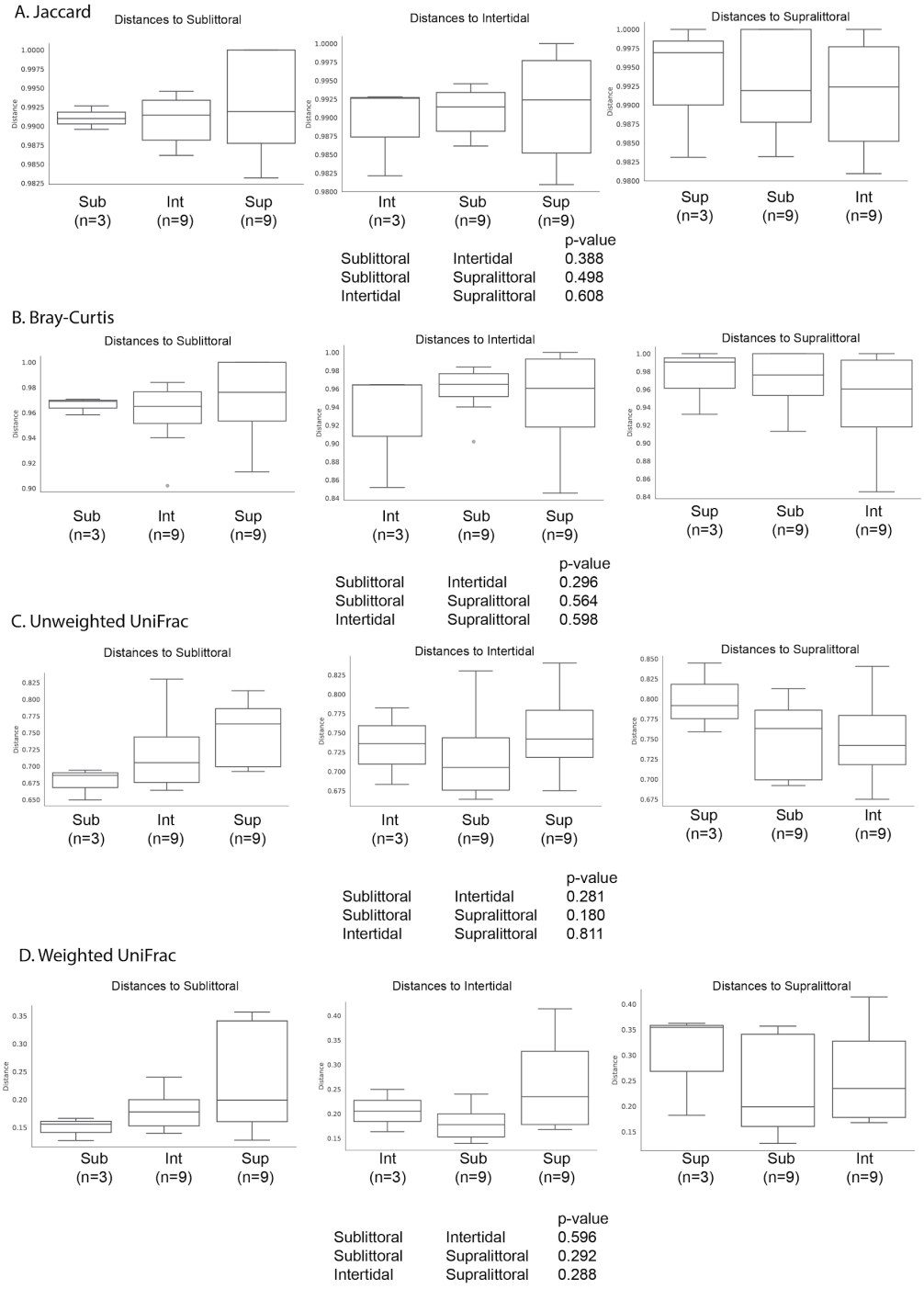
Supplemental Fig 2. **Alpha diversity by sampling site**. Alpha diversity (mean diversity of species per site) measured using the absolute number of Observed OTUs (**A**), Chao1 estimator (**B**), ACE (**C**), Shannon's diversity index (**D**), Simpson (**E**), InvSimpson (**F**), and Fisher (**G**).



Supplemental Figure 3. Statistical tests for Faith and Pielou evenness. Box plots showing the statistical results of alpha-diversity for Faith's phylogenetic diversity and Pielou's evenness, as calculated by QIIME2.

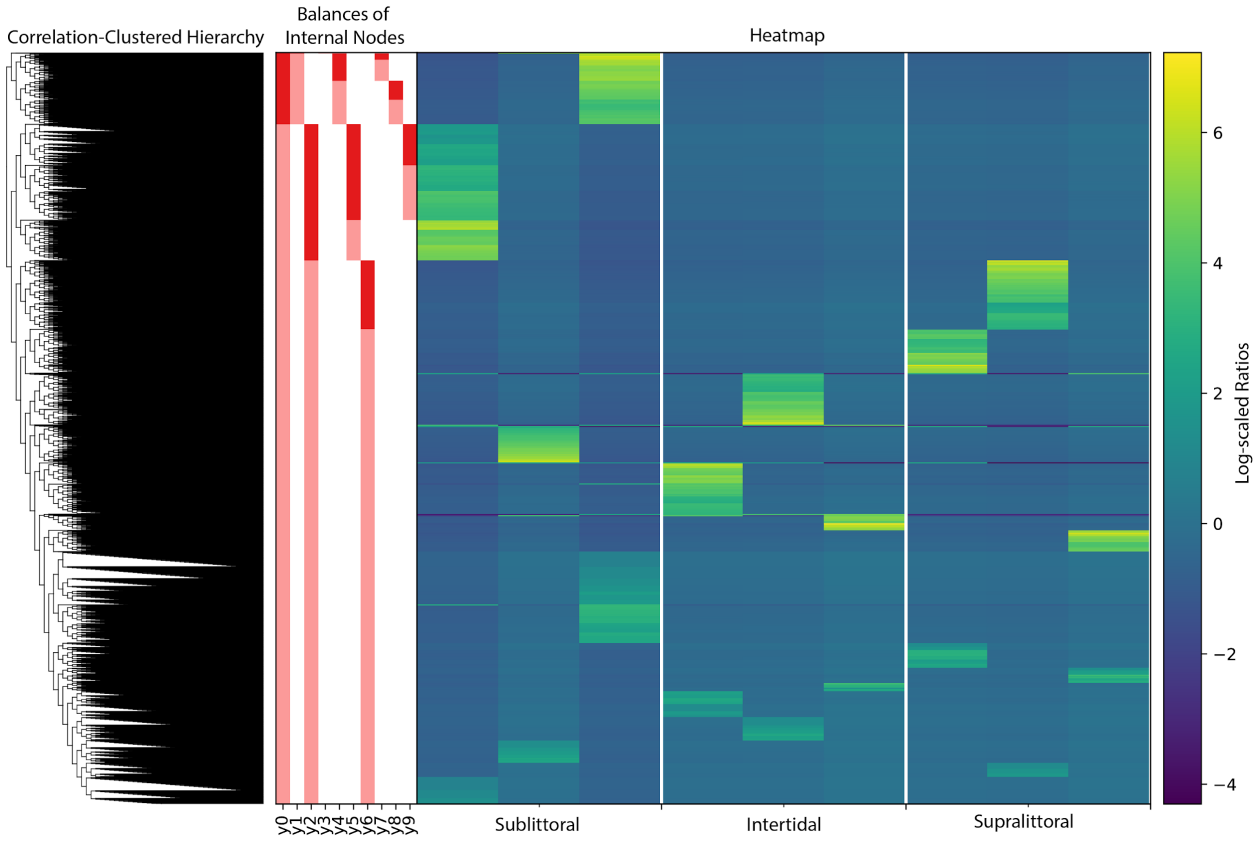


Supplemental Figure 4. Alpha-diversity by site by replicate. Alpha-diversity measures for each site with each replicate represented, as calculated by QIIME2.



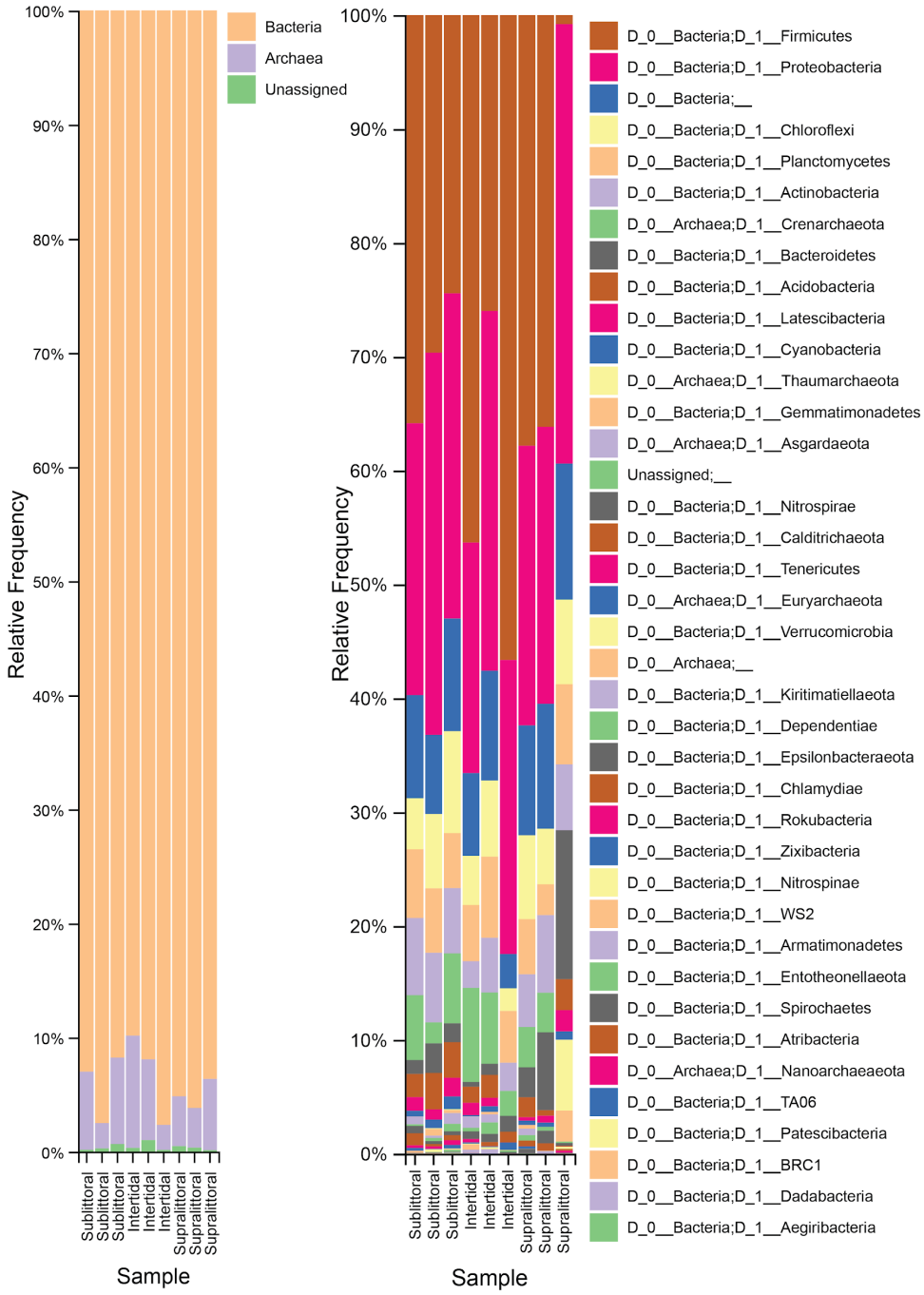
Supplemental Figure 5. Beta-diversity measures.

Here we present Jaccard (A), Bray-Curtis (B), and both the Unweighted (C) and Weighted (D) versions of UniFrac beta-diversity tests. In no instance is the diversity between sites significant.

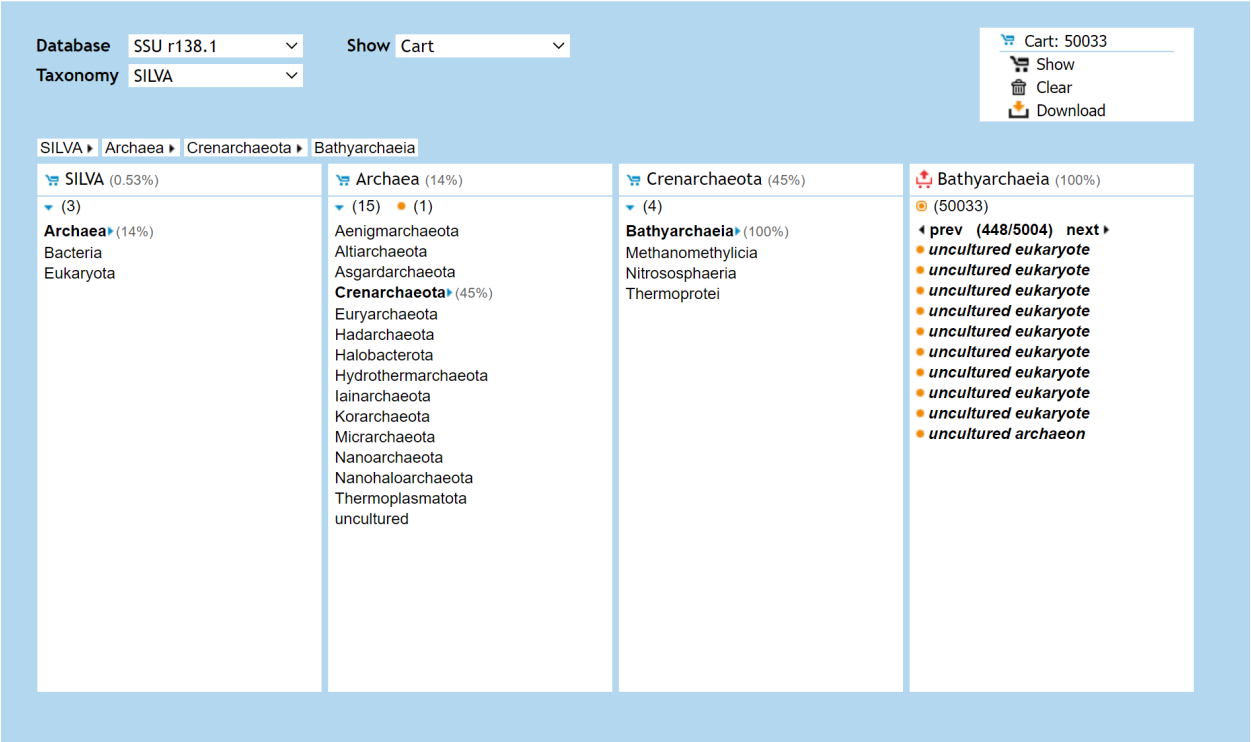


Supplemental Figure 6. Analysis of differential taxa abundances using QIIME2 gneiss.

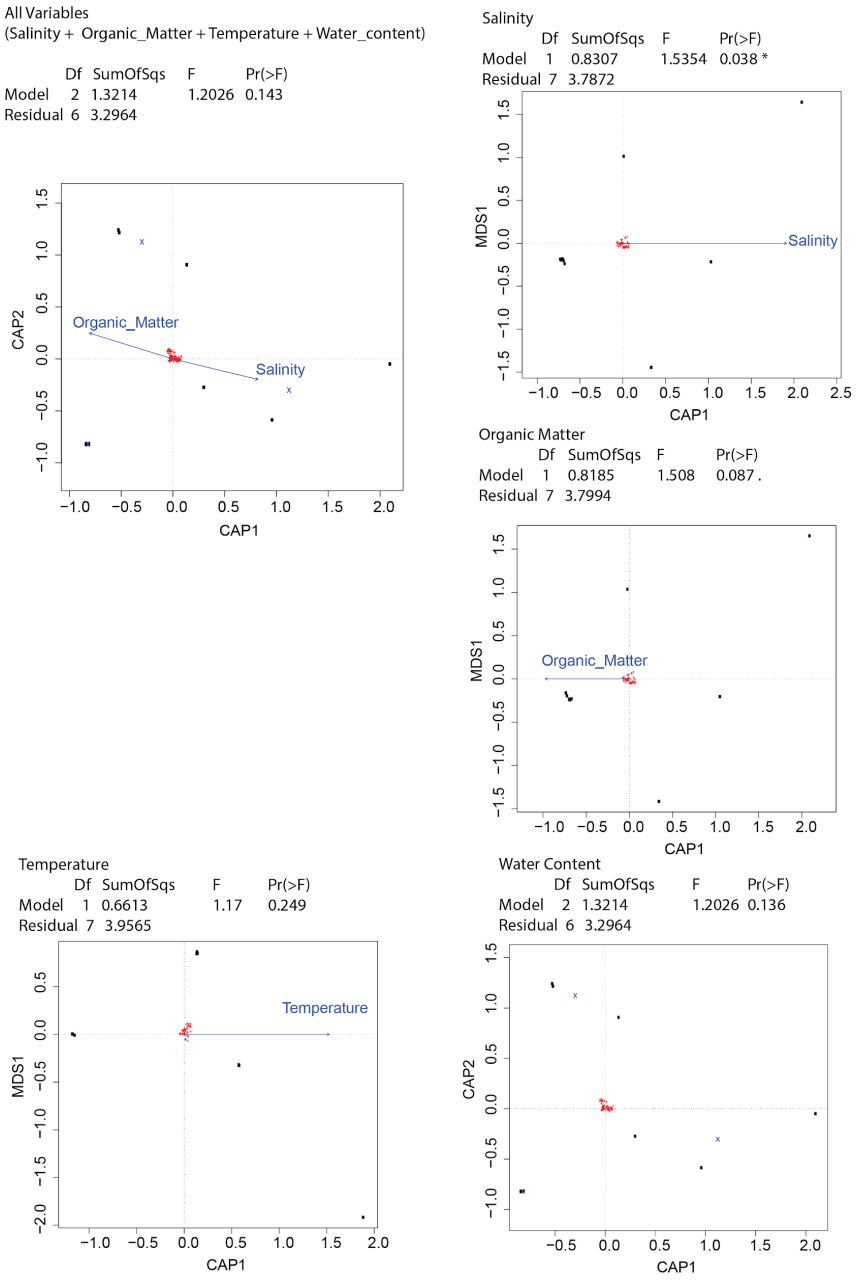
In addition to the beta-diversity metrics described above (Supplemental Figure 7[8]) we also used QIIME2's implementation of gneiss to perform correlation clustering across sites. However, we find that the balanced clusters only partially describe the distinct tidal zones (9 balances describe the correlation within Sublittoral).



Supplemental Figure 7. Bar charts of Domain and Phyla taxonomic proportion per site and replicate. Figures generated using the taxa-bar-plots.qzv (Supplemental File 2) in QIIME2 View.

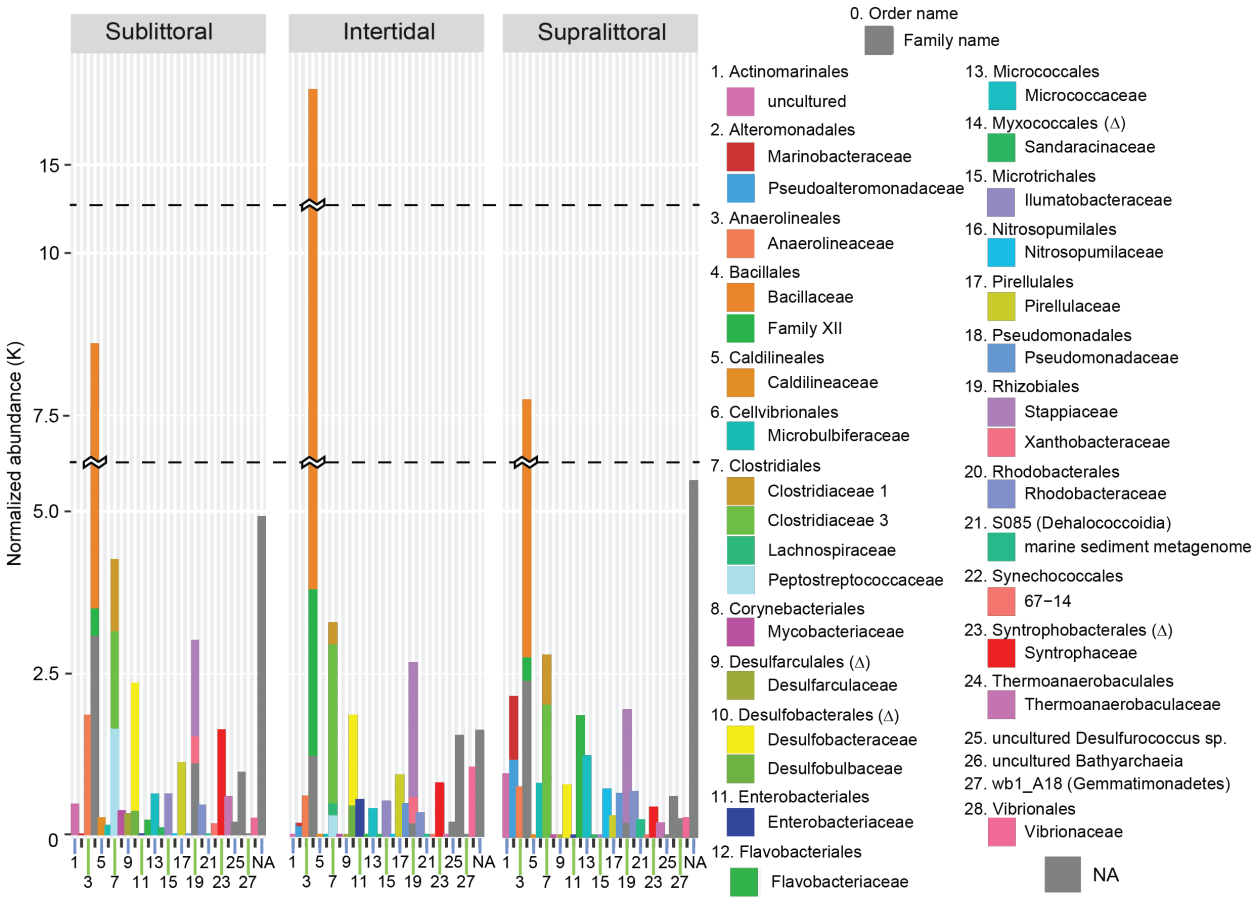


Supplemental Figure 8. Mislabelled taxa present in SILVA 138.1. The current version of SILVA has several mislabeled entries for ‘eukaryote’ under the *Bathyarchaeia* taxa. These are certainly not eukaryotes but mislabeled entries, of which there are several in the *Bathyarchaeia* taxa. To prevent confusion we have changed the text to refer to this group as “Uncultured Bathyarchaeia”.

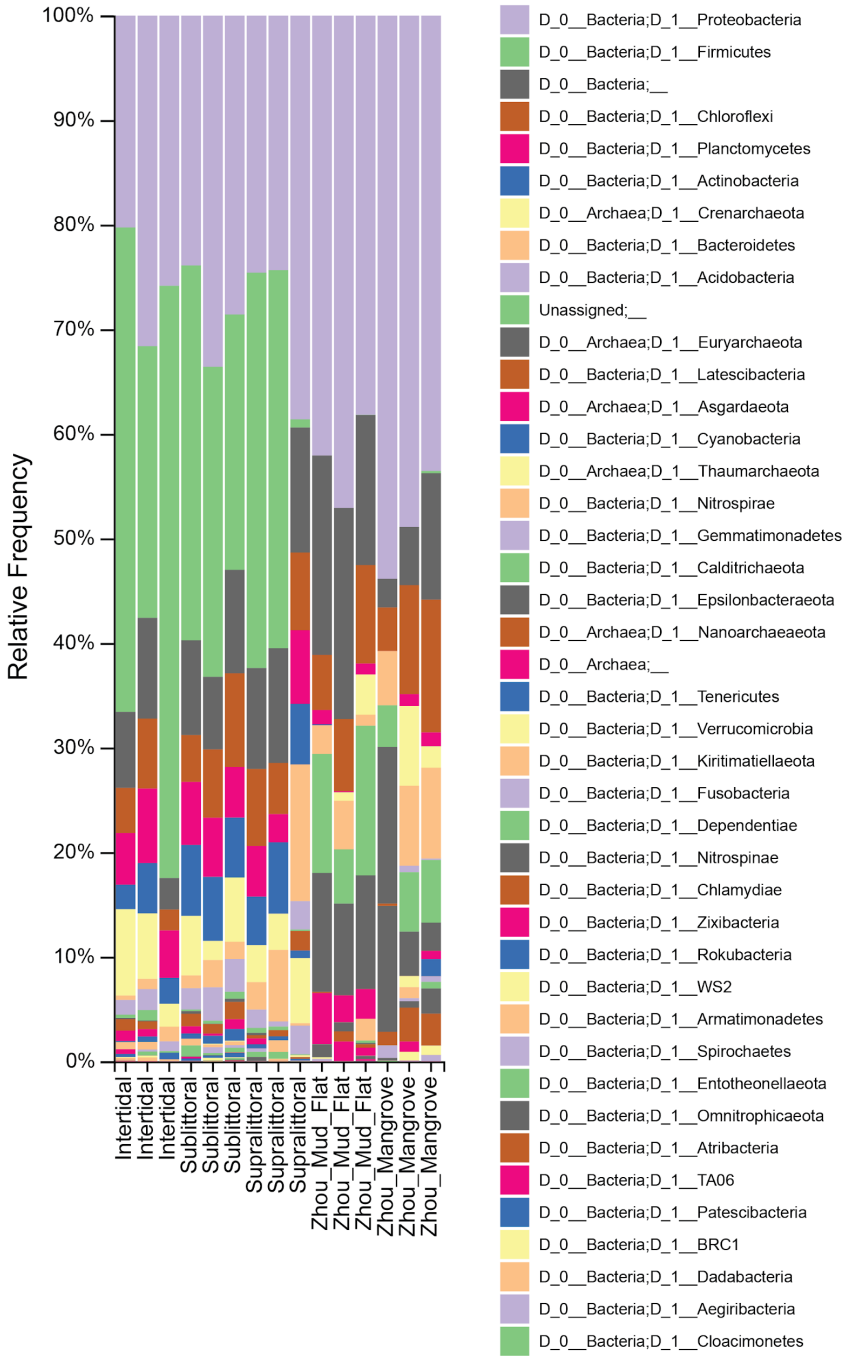


Supplemental Figure 9. Analysis of Environmental Variables using Vegan capscale.

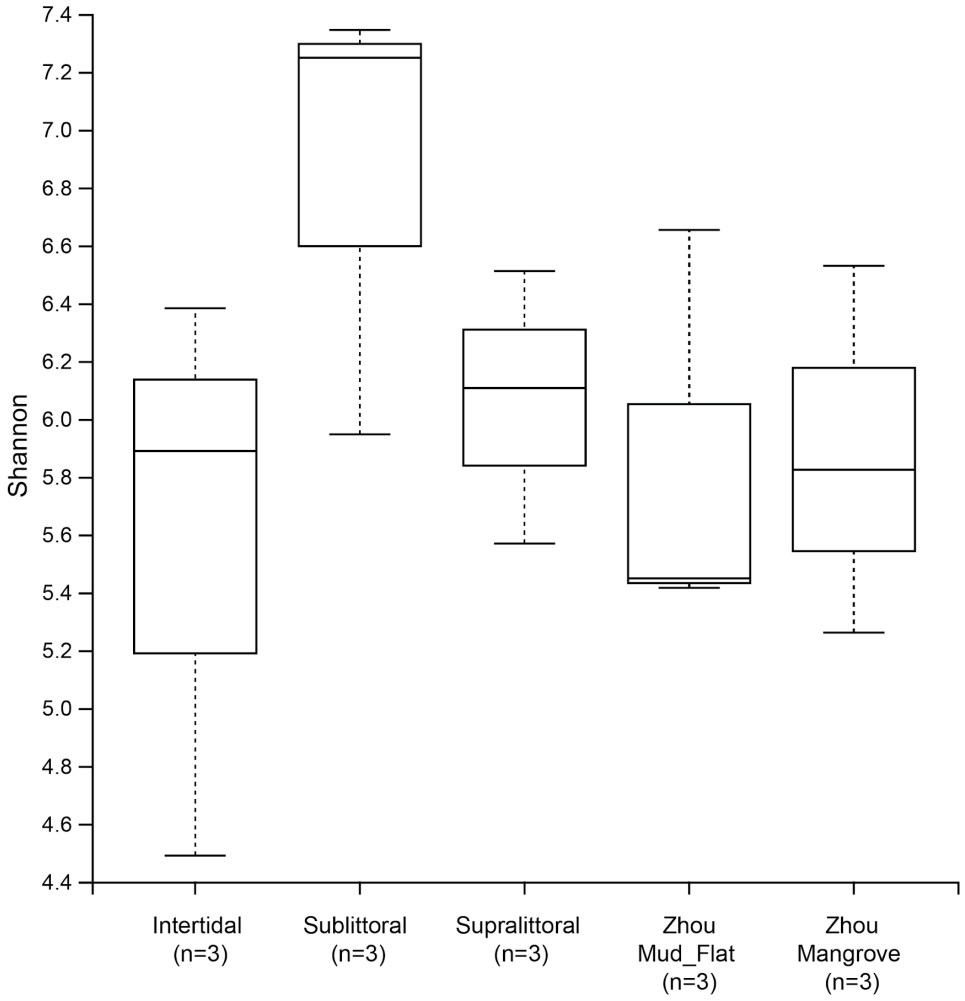
Here we show the results of Vegan's capscale, a constrained ordination analysis (SF6 Supplemental file 6, Vegan_capscale.r). The complete model is a basic linear module wherein all terms are additive and no terms are 'partialled out' (Oksanen, "Vegan: an introduction to ordination", https://cran.r-project.org/web/packages/vegan/vignettes/intro-vegan.pdf). Notably, we did not find this combined model to be statistically significant (Pr > 0.143). However, partialling out environmental variables resulted in increased model performance (Pr < 0.1 for Organic Matter and Pr < 0.05 for Salinity). These agree with the unconstrained ordination results using MetaMDS and Envfit.



Supplemental Figure 10. Normalized taxonomic abundances from each tidal zone. Taxa identified with abundance higher than 1% (post-normalization) are shown as stacked bar plots, the horizontal axis is the Order level while the stacked bars are the Family level. Taxa abundance has been normalized by downsampling sequence depth.

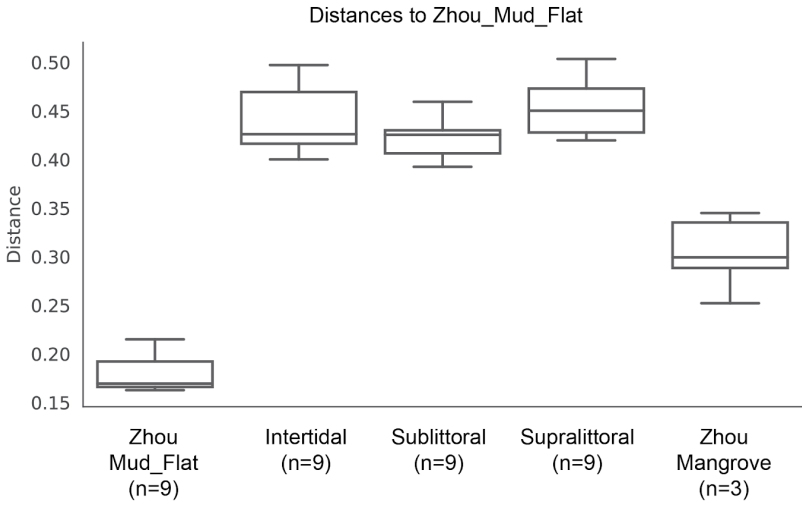


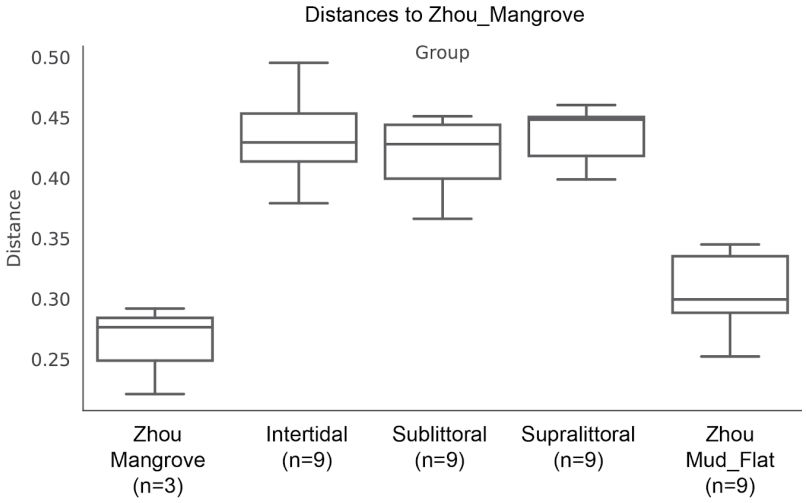
Supplemental Figure 11. Bar charts of Phyla taxonomic proportion per site and replicate compared to Zhou et al. 2017. We *Firmicutes* are notably absent from the Zhou data set (as well as our Supralittoral sample 3). Figures generated as described in methods section with one modification - to incorporate the most accurate V4-V5 regions amplicon sequencing of Zhou et al. 2017 we used the pre-trained classifier “Silva 138 99% OTUs full-length sequences” from QIIME2.



Group 1	Group 2	p-value
Intertidal (n=3)	Sublittoral (n=3)	0.127
Intertidal (n=3)	Supralittoral (n=3)	0.513
Intertidal (n=3)	Zhou_Mud_Flat (n=3)	0.827
Intertidal (n=3)	Zhou_Mangrove (n=3)	0.827
Sublittoral (n=3)	Supralittoral (n=3)	0.275
Sublittoral (n=3)	Zhou_Mud_Flat (n=3)	0.127
Sublittoral (n=3)	Zhou_Mangrove (n=3)	0.127
Supralittoral (n=3)	Zhou_Mud_Flat (n=3)	0.513
Supralittoral (n=3)	Zhou_Mangrove (n=3)	0.827
Zhou_Mud_Flat (n=3)	Zhou_Mangrove (n=3)	0.827

Supplemental Figure 12. Shannon alpha-diversity by site. Shannon alpha-diversity measures for each site as calculated by QIIME2. In no instance is the diversity between sites significant.





Group 1	Group 2	p-value
Intertidal	Sublittoral	0.599
Intertidal	Supralittoral	0.389
Intertidal	Zhou_Mud_Flat	0.1
Intertidal	Zhou_Mangrove	0.089
Sublittoral	Supralittoral	0.515
Sublittoral	Zhou_Mud_Flat	0.096
Sublittoral	Zhou_Mangrove	0.107
Supralittoral	Zhou_Mud_Flat	0.102
Supralittoral	Zhou_Mangrove	0.086
Zhou_Mud_Flat	Zhou_Mangrove	0.098

Supplemental Figure 13. Comparative beta-diversity measures.

Here we present Weighted UniFrac beta-diversity tests. In no instance is the diversity between sites significant.

Supplemental Methods section:

Read Trimming

Trimmomatic (Bolger, Lohse and Usadel 2014) was used to filter and trim demultiplexed sequences (ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:100). A minimum average read quality score of 15 was required for inclusion while the sliding window cuts any read at the point where the median quality score over a 4 nucleotide window is less than 15.

Read-pairing

This section summarizes the "Read_preprocessing.sh" (SF6 Supplemental file 6, Read_preprocessing.sh) and "jantar.py" (SF6 Supplemental file 6, jantar.py) files.

QIIME2 uses an implementation of vsearch for read pair joining ('vsearch join-pairs') which has a lower criteria of a minimum of 10 base-pairs of overlap. To extend vsearch with accurate read-pair joining below this cut-off we used a custom script (jantar.py) that would attempt to combine read-pairs that had failed to combined with vsearch. In order to be combined a read pair must have a "perfect" (starting from tails of reads, with no mismatches, in the correct orientation) 7bp overlap or 10bp with a single mismatch.

QIIME2

This section summarizes the "QIIME_run.sh" file (SF6 Supplemental file 6, QIIME_run.sh).

Reads were denoised using 'dada2 denoise-single', default settings using '--p-trim-left 3 and --p-trunc-len 0'. Phylogeny was determined using 'phylogeny align-to-tree-mafft-fasttree' with default settings. Minimum (9800) and maximum (40153) read depths were determined using 'feature_table.qzv'. Alpha-rarefaction was then calculated using the rooted MAFFT tree and the maximum read depth.

Alpha and Beta diversities were calculated using 'diversity core-metrics-phylogenetic' using the minimum read depth ('--p-sampling-depth') and the rooted MAFFT tree.

Taxonomy was calculated using 'feature-classifier classify-sklearn' and plotted using 'taxa barplot'.

Finally, gneiss was used to calculate the hierarchical correlation-clustering 'gneiss correlation-clustering' using a pseudo-count of 1. This was plotted using 'dendrogram-heatmap' for the meta-data column 'Site'.

PICRUSt2

This section summarizes the 'PICRUSt2_run.sh' file. (SF6 Supplemental file 6, PICRUSt2_run.sh)

PICRUSt2 was run twice using '--stratified' and '--per_sequence_contrib' options. One run was set to limit NSTI thresholds to 0.15, while the other (and data set subsequently used) was ran with the default NSTI threshold of 2.0.

Taxonomic abundance analysis
This section summarizes the use of the custom python script 'novembro.py' and 'Novembro_run.sh' (SF6 Supplemental file 6, novembro.py)

Novembro.py normalizes the ASV abundances generated by QIIME2 into their corresponding taxa.This requires feature-table.tsv and taxonomy.tsv files. While the taxonomy.tsv file an be found by unzipping the taxonomy.qza file the feature-table.tsv needs to be generated using a biom conversion command (eg. 'biom convert -i feature-table.biom -o feature-table.biom.txt --to-tsv')

Novembro.py iterates from the lowest to the highest taxonomic level combining ASV abundances into their corresponding taxa. Replicate normalizations are downsampled to match the lowest ASV abundant replicate. We use novembro.py to calculate zone specific taxonomic enrichment. This is accomplished by first using chi2 to compare the replicates of each zone against the others. To be deemed significant a zone must have an greater than a 5% effect size and p-value less than, or equal to, 0.05 relative to both other sites.

Taxa that are found to be significantly enriched are saved to a tab-delimited file along with their log10 transformed abundances.

Zone specific differential KO abundances
This section summarizes the use of ALDEx2, an R package, that uses GLM to identify differential expression abundances across sites using replicates (SF6 Supplemental file 6, sigilo.py).

First we generate Monte Carlo samples of each KO distribution of the unstratified metagenome predictions (eg. 'pred_metagenome_unstrat.tsv') using ALDEx2's centered log-transformed ('aldex.clr') module. We then use ALDEx2's GLM ANOVA test ('aldex.kw'), taking those KOs with an expected p-value ('glm.ep') of less than, or equal to, 0.05 to be significant.

Potential functional abundance analysis
This section summarizes the use of the custom python script 'sigilo.py' and 'Sigilo_run.sh'(SF6 Supplemental file 6, sigilo.py)

Sigilo.py performs several functions to aid in the visualization and analysis of PICRUSt2 and ALDEx2 generated KO data.

Generate heatmap: KO enrichment heatmaps can be generated using '--generate_heatmap', this function combines the predicted functional abundances of KOs

identified by ALDEx2 as significant into their corresponding KEGG Ortholog pathways, performs a log10 transform and plots them as heatmaps.

Correlate ASV with Functional Abundance: Using the '--asv2fa' command we first combine all ASVs into their corresponding taxa at a given level (for Family, level = 4) and then calculate the total functional abundance for those combined ASVs to derive the predicted functional abundance of a given taxa.

Correlate ASV with NSTI: Using the '--asv2nsti' command we first combine all ASVs into their corresponding taxa at a given level (for Family, level = 4) and then combine the associated NSTI values for those ASVs for each taxa. We can then use these to identify the mean, median, and standard deviation of NSTI scores for each given taxa.

Supplemental Tables

Taxa	Mean_nsti	Median_nsti	Standard_Deviation_nsti
Bacteria_Actinobacteria_Acidimicrobiia_Actinomarinales_uncultured	0.59	0.61	0.07
Bacteria_Actinobacteria_Actinobacteria_Micrococcales_Micrococcaceae	0.39	0.25	0.43
Bacteria_Bacteroidetes_Bacteroidia_Chitinophagales_ Saprospiraceae	0.41	0.38	0.14
Bacteria_Bacteroidetes_Bacteroidia_Cytophagales_ Cyclobacteriaceae	0.36	0.40	0.15
Bacteria_Bacteroidetes_Bacteroidia_Flavobacteriales_ Flavobacteriaceae	0.18	0.11	0.18
Bacteria_Calditrichaeota_Calditrichia_Calditrichales_ Calditrichaceae	0.35	0.33	0.14
Bacteria_Chloroflexi_Anaerolineae_Anaerolineales_ Anaerolineaceae	0.54	0.56	0.20
Bacteria_Firmicutes_Bacilli_Bacillales_ Bacillaceae	0.13	0.05	0.19
Bacteria_Firmicutes_Clostridia_Clostridiales_ Clostridiaceae 1	0.24	0.20	0.13
Bacteria_Firmicutes_Clostridia_Clostridiales_ Lachnospiraceae	0.28	0.27	0.10
Bacteria_Firmicutes_Clostridia_Clostridiales_ Peptostreptococcaceae	0.16	0.08	0.21
Bacteria_Planctomycetes_Planctomycetacia_Pirellulales_Pirellulaceae	0.43	0.42	0.19

Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales_Xanthobacteraceae	0.20	0.20	0.01
Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobacterales_Rhodobacteraceae	0.17	0.12	0.13
Bacteria_Proteobacteria_Deltaproteobacteria_Desulfobacterales_Desulfobacteraceae	0.39	0.28	0.28
Bacteria_Proteobacteria_Deltaproteobacteria_Syntrophobacterales_Syntrophaceae	0.31	0.36	0.16
Bacteria_Proteobacteria_Gammaproteobacteria_Vibrionales_Vibrionaceae	0.10	0.07	0.11

S1 table. Nearest sequenced taxa indices (NSTI) for significantly differentially abundant families. NSTI is a distance metric used by PICRUSt2 to compare the observed ASV to the nearest match in the functional database. Higher numbers mean lower similarity and a decreased confidence in the results. PICRUSt2 has a default cut-off of an NSTI of 2 but high similarity matches are 0.15 or less. Here we show the results of a custom script (sigilo.py --asv2nsti) to calculate the distribution of NSTI scores for families by combining the NSTI of all sub-family level taxa. Any family with a median NSTI within one standard deviation of 0.15 is given an asterisk. We remove from consideration families with median NSTI outside of this range.

Sampling Site	Water content	Salinity	Temperature (Celsius)	Organic Matter
Sublittoral	Submerged	665.9	25	3.71
Intertidal	Intertidal	677.6	29	3.86
Supralittoral	Dry	1047	29	1.4

S2 table. Environmental variables from collection sites.

#Summary of code

ord <- metaMDS(t(taxa_counts), engine = c("isoMDS"), try=1000, k = 3)

fit <- envfit(ord, my_varechem, perm = 999)

Vector	NMDS1	NMDS2	r2	Pr(>r)	Signif. codes
Salinity	0.59462	0.80400	0.5739	0.041	*
Temperature	0.90775	-0.41952	0.3588	0.271	ns
Organic_Matter	-0.55966	-0.82873	0.5720	0.039	*

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ns ’ 1

Permutation: free

Number of permutations: 999

Centroids	NMDS1	NMDS2	r2	Pr(>r)	Sig. codes
Dry	0.4671	0.1706			
Intertidal	-0.0246	-0.2237			
Submerged	-0.4425	0.0531			
Water_content			0.4563	0.07	.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ns ’ 1

Permutation: free

Number of permutations: 999

S3 table. Results of Vegan metaMDS and envfit.

sample	*Firmicutes*	*Gammaproteobacteria*
Sublittoral_1	3515.805763	196.874727
Sublittoral_2	2908.450707	804.947418
Sublittoral_3	2396.536942	632.375741
Intertidal_1	4543.672623	617.401501
Intertidal_2	2548.639765	1329.04881
Intertidal_3	5555.421202	335.829765
Supralittoral_1	3709.413567	360.789487
Supralittoral_2	3547	786
Supralittoral_3	77.63588391	2403.66786

Pearson's r	-0.840	p-value	0.004
R^2	0.706		
Spearman's *rho*	-0.7		

S4 table. Normalized reads of *Firmicutes* and *Gammaproteobacteria* across samples.
Previous work on contaminated mangroves in Brazil has suggested that Gammaproteobacteria is a bioindicator of an anthropogenically impacted mangrove (Andreote et al., 2012). Here we compare the normalized reads across our samples for all *Firmicutes* and *Gammaproteobacteria* and find that the two taxa are significantly negatively correlated (Pearson's R = -0.84, p-value = 0.004), consistent with *Firmicutes* as being a bioindicator of a pristine environment. However, *Firmicutes* were also found at elevated levels at several anthropogenically impacted sites (Torres et al., 2019), (Haldar and Nazareth, 2018), (Tiralerdpanich et al., 2018), suggesting that if it is a bioindicator it may not be a general bioindicator for all mangroves. While it could be that their presence is due to the pristine nature of the site, it may also be due to the local climate (Nogueira et al., 2015), an unmeasured environmental factor (Tong et al., 2019), or because of the presence of another organism such as a non-prokaryote or prokaryote which is difficult to detect using 16S rRNA amplicon sequencing (Eloe-Fadrosh et al., 2016), (Zhang et al., 2021).

References

Andreote, F. D., Jiménez, D. J., Chaves, D., Dias, A. C. F., Luvizotto, D. M., Dini-Andreote, F., Fasanella, C. C., Lopez, M. V., Baena, S., Taketani, R. G. and de Melo, I. S.: The microbiome of Brazilian mangrove sediments as revealed by metagenomics, PLoS One, 7(6), e38600, doi:10.1371/journal.pone.0038600, 2012.

Eloe-Fadrosh, E. A., Ivanova, N. N., Woyke, T. and Kyrpides, N. C.: Metagenomics uncovers gaps in amplicon-based detection of microbial diversity, Nat Microbiol, 1, 15032, doi:10.1038/nmicrobiol.2015.32, 2016.

Haldar, S. and Nazareth, S. W.: Taxonomic diversity of bacteria from mangrove sediments of Goa: metagenomic and functional analysis, 3 Biotech, 8(10), doi:10.1007/s13205-018-1441-6, 2018.

Nogueira, V. L. R., Rocha, L. L., Colares, G. B., Angelim, A. L., Normando, L. R. O., Cantão, M. E., Agnez-Lima, L. F., Andreote, F. D. and Melo, V. M. M.: Microbiomes and potential metabolic pathways of pristine and anthropized Brazilian mangroves, Regional Studies in Marine Science, 2, 56–64, doi:10.1016/j.rsma.2015.08.008, 2015.

Tiralerdpanich, P., Sonthiphand, P., Luepromchai, E., Pinyakong, O. and Pokethitiyook, P.: Potential microbial consortium involved in the biodegradation of diesel, hexadecane and phenanthrene in mangrove sediment explored by metagenomics analysis, Mar. Pollut. Bull., 133, 595–605, doi:10.1016/j.marpolbul.2018.06.015, 2018.

Tong, T., Li, R., Wu, S. and Xie, S.: The distribution of sediment bacterial community in mangroves across China was governed by geographic location and eutrophication, Mar. Pollut. Bull., 140, 198–203, doi:10.1016/j.marpolbul.2019.01.046, 2019.

Torres, G. G., Figueroa-Galvis, I., Muñoz-García, A., Polanía, J. and Vanegas, J.: Potential bacterial bioindicators of urban pollution in mangroves, Environ. Pollut., 255(Pt 2), 113293, doi:10.1016/j.envpol.2019.113293, 2019.

Zhang, J.-W., Dong, H.-P., Hou, L.-J., Liu, Y., Ou, Y.-F., Zheng, Y.-L., Han, P., Liang, X., Yin, G.-Y., Wu, D.-M., Liu, M. and Li, M.: Newly discovered Asgard archaea Hermodarchaeota potentially degrade alkanes and aromatics via alkyl/benzyl-succinate synthase and benzoyl-CoA pathway, ISME J., doi:10.1038/s41396-020-00890-x, 2021.