

ZOOMING IN AND OUT ON CORAL REEF MICROBIOMES:
MOLECULAR PATTERNS OVER SPACE AND TIME

by

Andrew James Maritan

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2025

©COPYRIGHT

by

Andrew James Maritan

2025

All Rights Reserved

DEDICATION

To my dear parents Susan and Jim. You've guided a distractable, energetic, and stubborn child to school, and let me try my hand at whatever I could (and couldn't) manage. I will be forever grateful for your support, love, listening, attention, affection, and time. In retrospect, we've always seen that (field) biology was my passion, but we didn't have a name yet.

After I got waylaid trying to be a doctor, studying chemistry, one animal behavior class chasing chickadees gave me a hazy sense for what I wanted to do. You both were so supportive as I chased the vague feeling of "mostly outside, not medical, biology". On the way I got to collect macroinvertebrates in rivers from Massachusetts, Vermont, and Montana, mist net birds and trap small mammals in Vermont, core 200-year-old trees, dissect birds, sharks, cats, rats, electrofishing across Vermont and Montana, culture bacteria from coastal algae, SCUBA dive in Massachusetts, Vermont, The Bahamas, and The Cayman Islands and snorkel in French Polynesia. I've been paid to work with corals (this dissertation and Brown et al. *in prep*), sea anemones, sea urchins (Vilanova-Cuevas et al., 2025), sea cucumbers and marine (Maritan et al. *accepted*) sediments, terrestrial sediments (Shoemaker et al., 2024), catfish (Brown et al. *in prep*), clownfish, rat, gerbils, birds, macroinvertebrates (Bertagnolli et al., 2023), algae (this dissertation), freshwater plants, trees, and freshwater fish (Rodriguez et al. *in prep*). I guess I'm a biologist now?

You sat with me and completed midnight homework in 3rd grade, witnessed me spin out about math in 7th grade, listened to me late-night rave about 9th grade biology, stayed on the phone when I need to cry in college. You've supported me in every way on this journey. You made this possible, and I love you dearly.

ACKNOWLEDGEMENTS

I have lots of people to thank. If you're on this list, I love and appreciate you and I owe my heartfelt and enduring thanks. If you aren't on the list, but you've supported me, I love you too and I'm sorry for the oversight. And now, in convenient order: Kaitlyn Maritan and William Maritan: for being companions, supporters, and a whole lot of fun. O. Berry, G. Duffy, J. Faro, S. Rodenhai, and K. Murphy: for showing me the power of supportive teaching. S. Johnson, W. Fedak, B. Shah, N. Antzperger: for enduring friendship, support and a whole lot of fun. T. Root, S. Sheldon, S. Trombulak, E. Eggleston, D. Allan, M. Spritzer, G. Spatafora, J. Ward, N. Dragone: for fostering a love of the natural world and introducing me to the beauty of field work, organismal biology, ecology, and microbiology. F. Napoli: through thick and thin, absolute gem. M. Borrmann, P. Ammons, J. Hulina, M. Sexias, A. Colgan, J. Hawkins, C. Howe: for enriching my work and play. F.J. Stewart and Z.A. Pratte: considerate, thoughtful, supportive, fun, patient, flexible, hands-off advisors. A. McDonnell: I would not have thought as deeply, as long, or as hard about the world without you nor would I have seen as much of Montana and the West without you. M. Brown, J. Crandall, A. Adams-Beyea, M. Rodriguez, W. Henriques, W. Christian, H. Dulay, J. Schafer, J. Schimetz, R. North, M. Schreve (AND MORE MCB FOLKS). B. Beavon, D. Foerster, H. Weas, K. Molina, E. Bolger, A. O'Connor, L. O'Connor, M. Hannan: for making my time in Bozeman a whole lot of fun. C.F.: thank you for your time, attention, and extremely useful self-reflection.

Additional thanks to for financial support from the U.S. National Science Foundation (grant no. OCE 1947522 to M.E.H. and F.J.S.), National Geographic Society (grant no. NGS-57078R-19 to M.E.H. and C.S.C.), Teasley Endowment to the Georgia Institute of Technology (to M.E.H.), Anna and Harry Teasley Gift Fund (to M.E.H.), and Simons Foundation (to F.J.S.). This work represents a contribution of the Mo'orea Coral Reef (MCR) LTER Site supported by the U.S. National Science Foundation (grant no. OCE 16-37396).

TABLE OF CONTENTS

1. VERTICALLY INTEGRATING ECOLOGICAL PERSPECTIVES TO GUIDE AND CONTEXTUALIZE MICROBIAL ECOLOGY (AND THE EFFECTIVE INTEGRATION OF ‘OMICS)	1
Purpose and Structure	1
Section 1: A Vertically Integrated Framework for Interrogating and Contextualizing the Ecology of Microorganisms	2
1.1: Ecology and Microbiology	2
1.1.1: A Historical Grounding	3
1.1.2: Ecological Philosophies and Microbiology	5
1.1.3: Vertical Integration: Examining a Single Process at Nested Ecological Scales	7
1.2 – A Vertically Integrated Workflow for Microbial Ecology	8
1.2.1 – Structure-Function (reductionism).....	10
Step 1 – Observe	10
Step 2 – Formulate Sequence-based Hypotheses.....	12
Step 3 – Test Hypothesis	14
Step 4 – Prove Function.....	16
1.2.2 – Contextualize (holism).....	17
Context A – Meta-analysis	17
Context B – Evolution	18
Context C – Biogeography.....	18
Context D – Modeling	19
Context E – Management.....	20
1.2.3 – Integrating Reductionist and Holist Perspectives	20
1.3 – Conclusions.....	21
Section 2: ‘Omics Data Products	21
2.1 – What is ‘Omics?.....	21
2.2 – Establishing Analytical Goals: Putting the Bio in Bioinformatics.....	23
2.3 – The Data Products.....	24
2.3.1 – Sequence Files	27
2.3.2 – Sequence Statistics (Length and Effort).....	30
2.3.3 – Taxonomy Table.....	30
2.3.4 – Function Table.....	30
2.3.5 – Count Matrix	32
2.3.6 – Putting it all Together: Merging Tables.....	32
2.4 – Conclusions.....	33
Section 3: Tools and Approaches	33
3.1 – Nucleic Acid Extraction and Sequencing	35
3.1.1 – Sequencing Technologies.....	35
3.1.2 – Nucleic Acid Extraction.....	37

TABLE OF CONTENTS CONTINUED

3.1.3 – Putting Nucleic Acid Extraction and Sequencing Together	38
3.2 – Sequence Statistics	38
3.3 – Quality Control	39
3.3.1 – Reads	39
3.3.2 – Contigs	40
3.3.3 – MAGs	41
3.4 – Community Description	42
3.4.1 – Ribosomal RNA Surveying	42
3.4.2 – Community Summaries	43
3.5 – Contaminant Removal	43
3.6 – Assembly	44
3.6.1 – Library Normalization	44
3.6.2 – Contig Assembly	44
3.7 – MAG Binning	45
3.7.1 – Binning Contigs	46
3.7.2 – MAG Improvement	46
3.8 – Taxonomic Classification	49
3.9 – Function Annotation	50
3.10 – Count Data	52
3.11 – Phylogeny	53
3.12 – Thoughts for Practitioners	55
 2. MICROHABITAT AND TIME-OF-DAY PREDICTABLY STRUCTURE DIVERSE REEF MICROBIOMES	 60
Contribution of Authors and Co-Authors	60
Manuscript Information	61
Abstract	62
Introduction	62
Materials and Methods	67
Experimental Design and General Sampling Description	67
Sample Collection and Processing	71
Microbiome DNA	71
Geophysical Measurements	71
Field Station Sample Processing	73
Nutrient Quantification	73
Microbiome Data Preparation – 16S and 18S rRNA Gene Sequencing	73
DNA Extraction	74
16S/18S rRNA Gene Amplification and Sequencing	74
Sequence Processing	75

TABLE OF CONTENTS CONTINUED

Statistical Testing and Plotting.....	76
Cartography.....	80
Results and Discussion	80
Microbiomes Across the Reef are Distinct and (with Notable Exceptions) Diverse	81
Time Structures Microbiomes Across the Reef	89
Microbial Abundances Exhibit Diel Periodicity Across Microhabitats	93
Conclusions.....	97
Permissions and Permits	98
 3. DOMINANT CORAL ENDOSYMBIONT EXPRESSES GENES FOR ACQUIRING HOST ATP, VITAMINS, SUGARS, AND MORE	99
Contribution of Authors and Co-Authors	99
Manuscript Information	100
Abstract.....	101
Introduction.....	102
Materials and Methods.....	106
Controlled Coral Preparation and Sample Collection.....	106
Reef Coral Sample Collection	109
Field Processing.....	110
Field Station Processing.....	110
Microbiome taxonomic analysis - DNA extraction, 16S rRNA gene amplification, sequencing, and analysis.....	110
DNA extraction	111
16S rRNA gene amplification and sequencing	111
Sequence Processing.....	112
Statistical Testing and Graphing	112
Holobiont Taxonomic and Functional Analysis - RNA Extraction and Sequencing	113
RNA Extraction and Sequencing	113
Holobiont taxonomic and functional analysis – transcriptome processing	114
PART 1: rRNA depletion and Taxonomic Survey	116
PART 2: Non-rRNA Transcript Abundances, Functional Annotation, and Taxonomic Classification.....	117
<i>Endozoicomonas</i> -specific Analysis.....	119
16S rRNA Phylogeny – Fragmentary and Assembled.....	119
<i>Endozoicomonas</i> Percentage of Holobiont Transcriptome.....	121
<i>Endozoicomonas</i> Functional Inference.....	122
Cartography.....	124

TABLE OF CONTENTS CONTINUED

Results and Discussion	124
<i>Endozoicomonas</i> Dominates <i>Pocillipora damicornis</i> 'Community and Holobiont Transcriptome	124
Bulk-tissue <i>Endozoicomonas</i> Transcriptome Captures Major Metabolic Processes.....	127
<i>Endozoicomonas</i> Expresses both Aerobic and Anaerobic Metabolisms Within the Coral Host.....	128
<i>Endozoicomonas</i> Steals Macromolecules and Nutrients from its Host	131
Central Carbon Metabolism.....	132
ATP Generation and Acquisition.....	134
DNA, Amino Acids, Vitamins, and Polyamines	134
Inorganic Compounds	136
Conclusions.....	138
4. SEA CUCUMBER GRAZING LINKED TO ENRICHMENT OF ANAEROBIC MICROBIAL METABOLISMS IN CORAL REEF SEDIMENTS	140
Contribution of Authors and Co-Authors	140
Manuscript Information	141
Abstract	142
Introduction.....	143
Materials and Methods.....	146
Cage Experiments and Sample Collection.....	146
DNA Extraction	148
Microbial Abundance Estimation - 16S rRNA Gene qPCR	149
Statistical Testing and Graphing	150
Microbiome Taxonomic Analysis - 16S rRNA Gene Amplicon Sequencing.....	150
16S rRNA Gene Sequencing.....	151
Sequence Processing.....	151
Statistical Testing and Graphing	152
Microbiome Metabolic Potential – Shotgun Metagenomics	153
Metagenomic Sequencing and Quality Control.....	153
Read-based Analysis – K-mer Profiling	154
Read-based Analysis – 18S rRNA Gene Screening	154
Read-based Analysis – Marker Gene Screening	154
Metagenome Assembled Genome (MAG) Analysis.....	156
Statistical Testing and Graphing	157
Results and Discussion	158
Sea Cucumber Grazing Depletes Sediment Microbial Abundance	159

TABLE OF CONTENTS CONTINUED

Sea cucumber Grazing Reshapes Sediment Microbiome Diversity	160
Sea Cucumber Grazing Significantly Depletes Phototrophs	163
Sea Cucumber Removal of Phototrophs may Influence Coral Health.....	165
Sea Cucumber Grazing Reshapes Microbiome Metabolic Potential, Enriching for Genes of Anaerobic Metabolisms.....	168
Conclusions.....	172
Data Availability	173
Acknowledgements.....	173
5. CONCLUSIONS AND SUGGESTIONS.....	175
Reef-scale Survey of Microhabitats.....	176
<i>Endozoicomonas</i> -Coral Interactions	177
Spatial Scales	178
Genomic Scales.....	178
Environmental Context.....	179
Sea Cucumber Grazing and Microbial Linkages to Coral Mortality.....	180
Microbial Quantification.....	181
Biogeochemistry and Toxin Production.....	181
Mechanism of Sediment Transformation for a Protective Effect	182
Site-Specific Conditions that Modulate Protective Effect	182
Final Remarks	183
CUMULATIVE REFERENCES CITED.....	184
APPENDICES	219
APPENDIX A: EXTENDED TABLES FROM CHAPTER TWO	220
APPENDIX B: EXTENDED TABLES FROM CHAPTER THREE.....	242
APPENDIX C: EXTENDED TABLES AND SUPPLEMENTARY FIGURES FROM CHAPTER FOUR.....	257

LIST OF TABLES

Table	Page
1. Table A1.: Rarefaction depths used for different analyses, broken up by 16S and 18S rRNA gene data. N, total samples; n, subset samples; raref., rarefaction; ret., retained; sam., samples; NA, not applicable.	221
2. Table A2.: Statistical testing for beta diversity differences between reef microbiomes (ASV-scale) depending on sample type. Comparison of Bray-Curtis dissimilarity for different reef microbiomes as tested by permutational multivariate analysis of variance (PERMANOVA) and permutational multivariate analysis of dispersion (PERMDISP), 999 permutations. Comparisons are bolded to identify statistically distinguishable contrasts (PERMANOVA or PERMDISP, $P < 0.05$). PDAM, Pocillipora damicornis, PRUS, Porites rus; perm., Permutations; n, sample size.	221
3. Table A3.: Statistical testing for beta diversity differences between reef microbiomes (ASV-scale) depending on phase of sampling (per Chapter 2 Figure 1I). Comparison of Bray-Curtis dissimilarity for different reef microbiomes as tested by permutational multivariate analysis of variance (PERMANOVA; 999 permutations), only displaying comparisons of Phase 1 vs. Phase 2 samples (Pocillipora damicornis, Porites rus, and water samples). Bolded comparisons identify statistically distinguishable contrasts (PERMANOVA, $P < 0.05$). PDAM, Pocillipora damicornis, PRUS, Porites rus, P1, Phase 1 (1800 h on 29 March 2022 to 1200 h on 2 April 2022, ~6 ft depth, coral rack); P2, Phase 2 (1200 h on 4 April 2022 to 0001h on 7 April 2022, ~3 ft depth, reef flat).	223
4. Table A4.: The core microbiomes of different reef microbiomes. The core microbiome is defined here as genera that were present in at least 80% of all samples for a given sample type and had a percent abundance greater than 1% in at least one of those samples and presented as average percentage abundances for each core prokaryote and non-host eukaryote. Hydr., Hydroclathrus sp.; Dict., Dictyiotia sp.; Hali., Halimeda sp.; Padi., Padina sp.; Sedi., sediment; Bact., Bacteria; Arch., Archaea; Euk., Eukaryota; Sam., Sample; Avg. Perc. Abd., average percentage abundance in the specified sample type ; UC, unclassified.	224

LIST OF TABLES CONTINUED

Table	Page
5. Table A5.: Statistical testing of cyclicity for measured geophysical features using RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods). Bolded <i>P</i> values indicate significance (RAIN, $P < 0.05$, Bonferroni correction).	226
6. Table A6.: Statistical testing of cyclicity for measured alpha diversity in different sample types using RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods). Bolded <i>P</i> values indicate significance (RAIN, $P < 0.05$, Bonferroni correction). Bf, Bonferroni correction; rRNA, ribosomal rRNA; PDAM, Pocillipora damicornis; PRUS, Porites rus.....	227
7. Table A7.: Statistical testing of differences in mean abundance over time for measured alpha diversity in different sample types using linear mixed effect models (corals) or ANOVA (water, sediment, and macroalgae). For each sample type and rRNA-type, all samples collected over time were compared to one another. If the model fit was initially significant, then post hoc test was performed to identify significant contrasts. In initial model testing significant (linear (mixed effect), <i>p</i> -values were bolded ($P < 0.05$, Tukey HSD) or underlined ($P > 0.05$ but < 0.10 , Tukey HSD). In post-hoc testing, groups with dissimilar letters in post hoc testing have <i>P</i> values < 0.05 , while any shared letter indicates <i>P</i> value > 0.05 . rRNA, ribosomal rRNA; PDAM, Pocillipora damicornis; PRUS, Porites rus.	228
8. Table A8.: Microbiome differences within a sample type by time of day as determined by permutational analysis of variance (PERMANOVA; 999 permutations). Comparisons are bolded to identify statistically significant ($P < 0.05$) and underlined for detectable contrasts ($P < 0.10$). <i>n</i> , sample size.	231
9. Table A9.: Taxa (amplicon sequence variants) identified as contributing to community differences by time in Similarity Percentage (SIMPER) analysis, only showing top 2 taxa per comparison. rRNA indicates amplicon source, 16S or 18S rRNA genes; sam. type, the reef sample class used in the comparison; group, the comparison being made (time A vs. time B) avg perc, the average percentage that the taxa is found in the sample type; Avg A/B perc., the average percentage abundance that the taxa is found in the focal sample type at time A/B; Abs Avg A MIN Avg B, the absolute value of the average abundance of the taxon in group A less the abundance in group B; lineage, the taxonomic lineage of the taxon per SILVA rRNA nomenclature (domain, phylum, class, order, family, genus).	233

LIST OF TABLES CONTINUED

Table	Page
10. Table A10.: Cyclic microbial taxa in each of the reef microbiomes sampled as identified statistically by RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods) and false discovery rate (FDR) corrected (RAIN, Adaptive Benjamini-Hochberg, p -value < 0.05) and filtered to only include the most abundant (>1-5%), and prevalent ($\geq 25\%$ of all samples) taxa, then placed into four unsupervised groups (visualized in Figure 9). To increase clarity for this table, we filtered the cyclic taxa from each microhabitat to only include five taxa from each Domain with the largest distance values (indicating more closely associated with their plotted cluster) for each cluster. Sam. type, sample type; clust., cluster; dist., distance; lineage (Domain, Phylum, Class, Order, Family, Genus).	236
11. Table B1.: Endozoicomonas 16S rRNA gene and amplicons resembled one another and point towards a couple dominant sequence variants. Percentage similarity Endozoicomonas 16S rRNA sequences sourced from 16S rRNA gene amplicon sequencing, 16S rRNA transcript fragments, and 16S rRNA assembled transcripts from an all-against-all BLASTn alignment. Each amplicon sequence variant or transcript fragment is listed with their most similar assembled transcript, with the alignment with the highest percent identity bolded.	243
12. Table B2.: Field preservation method for coral samples. Dry ice was used to snap freeze corals in the field, until COVID-based supply limitations required us to switch to RNA/ DNA preservation buffer (RNAlater) on ice.....	244
13. Table B3.: Reference genomes for initial read recruitment. All available genomes for Porites sp. and Pocillipora sp. as of February 2024.	245
14. Table B4.: Reference genomes for secondary read recruitment. 1-4 genomes accessions per taxa identified from rRNA screening (available per February 2024)	245
15. Table B5.: Taxonomic breakdown of reads after mapping to reference genomes. PDAM, Pocillipora damicornis; PRUS, Porites rus.	254
16. Table B6.: Reference Endozoicomonas genomes used in comparative analyses. (available per December 2024)	254
17. Table B7.: Expression levels of respiratory reductases in Endozoicomonas sp.	256

LIST OF TABLES CONTINUED

Table	Page
18. Table C1.: quantitative PCR (qPCR) and PCR master mix and thermocycler conditions	267
19. Table C3.: Common contaminant microbes filtered from amplicon sequence variant (ASV) table. All ASVs with taxonomy contained by the provided terms at the provided taxonomic rank and below were removed from the analysis. Taxonomic rank, the taxonomic level at which filtering was performed for the provided "taxon term"; taxon term, the name of the taxonomic group to be removed from the 16S gene count table	268
20. Table C4.: Marker Genes used for alignment with DIAMOND.....	269

LIST OF FIGURES

Figure	Page
1. Figure 1: Generalized environmental microbiology approaches for linking structure to function (“Structure-Function: go small”) and for contextualizing the importance of observed functions (“Contextualize: get big again”).....	9
2. Figure 2: Simplified examples of the major ‘omics data products	26
3. Figure 3: Generalized bioinformatics workflow for performing metagenomics, transcriptomics, and marker gene analyses.....	35
4. Figure 4: Sampling location, design, and sample types included in reef-scale, time-resolved microbiome sampling.....	70
5. Figure 5: Diversity and community composition of A,C,E) prokaryotic and B,D,F) eukaryotic reef microbiomes.	82
6. Figure 6: The core microbes of different reef microbiomes presented as percentage abundance for A) prokaryotes and B) eukaryotes.	85
7. Figure 7: Geophysical conditions during time-series microbiome sample collection between 29 March and 6 April 2022	88
8. Figure 8: Diversity and community composition for individual reef microbiomes sampled at 6-hour intervals (Phase 1: 1800h 29 March – 1200h 02 April 2022) using amplicon sequence variants	91
9. Figure 9: Time-series of cyclic, abundant, and prevalent genera (both prokaryotic and eukaryotic) for reef sample types collected at 6-hour intervals, clustered into four groups by temporal abundance trends. abundance	96
10. Figure 10: <i>Endozoicomonas</i> sp. dominates the community and transcriptome of one coral species on a reef in French Polynesia	108
11. Figure 11: coral bulk tissue RNAseq taxonomically classified and functionally annotated <i>Endozoicomonas</i> transcripts.....	115
12. Figure 12: Multiple sequencing methodologies all recapitulate <i>Endozoicomonas</i> sp. abundances	125

LIST OF FIGURES CONTINUED

Figure	Page
13. Figure 13: <i>Endozoicomonas</i> 16S rRNA transcripts from <i>Pocillipora damicornis</i> are phylogenetically consistent with reference genome sequences.	126
14. Figure 14: Coral-hosted <i>Endozoicomonas</i> operates a heterotrophic metabolism supplied by metabolites acquired from the host environment.....	131
15. Figure 15: Extended metabolic model - <i>Endozoicomonas</i> expresses transporters to acquire and process diverse carbon, sulfur, nitrogen, sulfur sources.....	134
16. Figure 16: Sediment microbial load decreases in the presence of the sea cucumber <i>Holothuria atra</i>	147
17. Figure 17: Sea cucumber grazing reshapes sediment microbiome diversity and depletes phototrophs.....	161
18. Figure 18: Abundant, high-quality cyanobacterial metagenome assembled genomes (MAGs) encode genes for cyanotoxin biosynthesis	166
19. Figure 19: Microbiome metabolic potential differs between ungrazed sediment (sed w/o) versus sediment grazed (sed w/ and feces) by sea cucumbers, with anaerobic metabolisms enriched in grazed sediments.....	169
20. Appendix C Figure 1: Rarefaction of 16S rRNA amplicons	258
21. Supplementary Figure 2: Diagnostic plots for linear model of genome equivalents per megabase (Mb) by sample type	259
22. Appendix C Figure 3: Ordination of grazed and ungrazed sediments using k-mer based distance matrices from unbinned metagenomes demonstrates that sample 06_in_wo_sediment deviates from all other samples.....	260
23. Appendix C Figure 4: Estimating microbial abundances before and after cucumber grazing shows that cucumber grazing reduces proxy measures for microbial load	260
24. Appendix C Figure 5: Beta diversity of sediment samples before, during and after digestion from 16S rRNA gene data show that sea cucumber grazing alters community composition, but not dispersion.	261

LIST OF FIGURES CONTINUED

Figure	Page
25. Appendix C Figure 6: Alpha diversity of sediment samples before, during and after sediment digestion from 16S rRNA gene data shows that cucumber GI tract and external sediment differs.....	262
26. Appendix C Figure 7: Depletion of cyanobacterial transformed relative abundances in grazed sediments	262
27. Appendix C Figure 8: Changes in cyanobacterial family abundances before and after grazing from 16S rRNA gene data.....	263
28. Appendix C Figure 9: Normalized microalgal abundance of 18S rRNA genes before and after cucumber grazing	264
29. Appendix C Figure 14: Inferred metabolic potential of the microbial community from unassembled metagenomes.....	266

ABSTRACT

The observable macro-scale diversity of coral reefs is mirrored in its microbiome. This diversity is exciting but overwhelming. Parsing this complexity to find and understand “*important*” microbes is a perennial challenge. To cut through this diversity I leveraged ecosystem (e.g. coral mortality, geophysical measures) and community-scale observations (e.g. microbes with high/ cyclic abundance and/ or prevalence) to identify tractable subjects for sequencing-based microbiome investigation. The framework for identifying research subjects is the focus of the Introduction. In the remainder of the dissertation, I will cover multiple angles of the coral reef microbiome using the framework from the introduction. Chapter 2 performs a marker gene survey of microbial taxa across reef microhabitats, identifying several microhabitats and taxa bearing further investigation. The next chapters will focus on microhabitats and microbes characterized in Chapter 2. Chapter 3 uses *in situ* marker gene surveying and bulk-tissue transcriptomics to zoom in on a high abundant and ubiquitous coral endosymbiont. I uncover gene expression evidence of unexpected, potentially coral parasitizing, metabolisms. Chapter 4 employs quantitative PCR, marker gene surveying, and metagenomics to dig into the reef sediment microbiome, searching for potential taxa and mechanisms that may explain reduced coral mortality after sea cucumber grazing. I find that the metabolites from phototrophs and sulfur reducers bear closer investigation. Collectively, this dissertation identifies several tractable microbial taxa and metabolisms that underlie reef function and follows-up on a couple in detail. The metabolic potential and taxa of interest identified bioinformatically provides clear future directions for research untangling microbial ecology across the coral reef.

INTRODUCTION

VERTICALLY INTEGRATING ECOLOGICAL PERSPECTIVES TO GUIDE AND CONTEXTUALIZE MICROBIAL ECOLOGY (AND THE EFFECTIVE INTEGRATION OF ‘OMICS)

“One should have a rather uncomfortable feeling that one is not covering the whole ground, that the problem is too big to tackle single handed...it would be worth getting to know a little about geology or the movement of the moon or of a dog’s tail, or the psychology of starlings...which are always turning out to be the basis of ecological problems” – Charles Elton Animal Ecology

“A jack of all trades is a master of none, but oftentimes better than a master of one”

Purpose and Structure

This dissertation is an extended love letter to ecology. I believe that all microbiology is at-its-core ecological, and this introduction articulates an ecological framework to guide and contextualize any microbiology. Though I believe that the best microbial ecology employs multiple tools, my expertise revolves around the analysis of biological sequences (‘omics), and for that reason my introduction will dive deeper into how ‘omics can be effectively leveraged in this microbial ecology framework. This ‘omics focus will necessarily persist through the three chapters of primary research, where I use an ecological perspective to interrogate coral reefs bioinformatically. The conclusion will use the ecological framework described in the introduction to suggest next steps for the primary research chapters. At the end of this dissertation, we will have outlined a roadmap for structuring microbiology studies ecologically

and covered three case studies of ecological inquiry using ‘omics, with non-omics next steps identified and discussed.

Practically, this introduction is intended to serve as a roadmap for current or prospective microbiologists aiming to ground their research ecologically (Section 1) and especially how to leverage ‘omics in this aim (Sections 2 & 3).

Section 1: A Vertically Integrated Framework for Interrogating and Contextualizing the Ecology of Microorganisms

Science is systematic, and increasingly accurate, storytelling.

1.1: Ecology and Microbiology

Ecology is the specific study of interactions between living and non-living entities. The ultimate goal is linking observed structure to emergent function and then contextualizing that result in our understanding of how the world operates. To do this, ecologists have spent nearly two centuries parsing global complexity into tractable chunks by delineating, acknowledging, and operating across biological, spatial, and temporal scales.

Ecology, by definition, includes microbes, but in-practice macroecology and microbiology have little overlap. Ecologists generally consider microbes as physiological features of their focal species (disease agents or digestion aids); while microbiologists often have little-to-no ecological foundation, trained with a heavy medical emphasis. This state of un-ecological microbiology may have been inevitable, as both disciplines use completely different tools and skills to make the same observations. For example, confidently seeing species diversity in macroecology uses the naked eye while microbiologists need florescence microscopy. One is the evolutionary capacity of humans while the other required thousands of years of scientific

advancement; in that time, the interests of the fields diverged. However, newfound technologies have allowed contemporary microbiologists to ask and answer increasingly sophisticated ecological questions. **To effectively capitalize on this opportunity to perform *microbial ecology*, microbiologists need to think and operate like macroecologists.**

In this review I define microbial ecology as the application of ecological perspectives to study microbes and is inclusive of all microbiology (e.g. environmental, medical, industrial, etc...). To provide a guiding framework for this microbial ecology, I will first summarize the historical progression of macroecology's research goals, describe the philosophical scales at which ecologists operate, and then integrate them into a useful workflow.

1.1.1: A Historical Grounding. To identify where microbiology and macroecology diverged and to identify the goals of modern ecology (macro and micro), we must first cover a brief history of ecology.

The history of macroecology can be roughly broken up into six phases ((J. G. T. Anderson, 2013; Begon et al., 1990; Marquet et al., 2014). **1)** Killing the scary things that kill us (management by self-defense), **2)** Cataloging the natural world (naturalists and natural historians), **3)** Describing interactions and articulating theory (the first ecologists), **4)** Managing the environment using ecological principles (rational/scientific management), **5)** Testing theory (experimentalists), **6)** Combining field observations, experimentation, and theory over space, time, and heterogenous landscapes (big data modelers).

In a macroscopic sense, humans depopulated most of the most dangerous megafauna between 25,000 and 50,000 years ago CEs (Phase 1; (Bergman et al., 2023)). After humans has achieved some breathing room and had more bandwidth, naturalists and natural historians began

cataloging global biodiversity in the 1800s (Phase 2; e.g. Alexander von Humboldt, (J. G. T. Anderson, 2013)). This paved the way for the first ecologists of the 1940s/1950s to begin synthesizing observations into unifying theories (Phase 3; e.g. Robert MacArthur, Robert Paine, Hutchinson; (MacArthur & Wilson, 1963; Paine, 1966)). This led into the beginning of ecologically informed ecosystem management and conservation of the 1950-1970s (Phase 4; e.g. Aldo Leopold; (Leopold, 1949)). After enough theory had been articulated, in the 1970-1980s experimental ecologists sought to test it with large and small-scale ecosystem experiments (Phase 5; e.g. Hubbard Brook, Canadian Experimental Lakes Area, Steven Carpenter; (Carpenter & Kitchell, 1988; Likens et al., 1970; Schindler, 1974)). We reach roughly modern day with ecology using long-term datasets across large spatial extents to understand landscape heterogeneity using complex models (Phase 6; e.g. NEON, LTER; (Marquet et al., 2014; Read et al., 2018; Schweiger & Laliberté, 2022)).

Microbiologists have lagged behind macroecology but are rapidly catching up. Just as prehistoric humans first sought to subdue nature, microbiologists first developed tools to subdue pathogens with vaccines in the 18th century (Hilleman, 1998), germ theory in the 19th (Pasteur, 1881), and antibiotics in the 20th (Hutchings et al., 2019). In the 19th century, culture-based approaches began the survey of microbial diversity (e.g. Sergei Winogradsky (Zavarzin, 2006). It took until the end of the 20th century before molecular surveys (Lane et al., 1985) and taxonomically resolved microscopy (Giovannoni et al., 1988) allowed broad taxonomic surveys of uncultured microbes. The early 21st century added functional inference to taxonomic surveys (Rondon et al., 2000; Tyson et al., 2004). Scientific microbial management manifested in the 19th

century (e.g. activated sludge wastewater treatment; (Sheik et al., 2014)), and was refined over time (e.g. anammox wastewater treatment; (Van Der Star et al., 2007)).

From a 25,000 year head start, microbiologists have quickly closed the gap with macroecologists, but the average environmental microbiologist still operates as an 1800s macroecologist, describing distributions of life and the characteristics of those habitats. Microbial ecology's next step is to integrate microbial functions and interactions into new and existing ecological theory, testing it, and building models, a step many have begun to take (D'Souza et al., 2018; Enke et al., 2018; Garren & Azam, 2012; Lui et al., 2021; Malik et al., 2020; Zaneveld et al., 2017).

1.1.2: Ecological Philosophies and Microbiology. The “holy grail” of modern ecology involves linking ecosystem structure (observable biotic and abiotic features) and emergent function (ecosystem, community, population, or organismal). In service to this goal, biologists have self-organized into different levels of reductionism to parse the structure-function question. To situate microbiology within an ecological framework, I will briefly summarize the ecological philosophies that cover increasingly reductionist biological scales, and then describe how microbes fit into these perspectives.

Ecosystem Ecology: is a very holist approach to ecology. The researcher treats a discrete volume of space as a closed (eco)system and is interested in the mass balance of that system (generally measuring indirect indicators of activity: material, energy flux, metabolic rates). The ecosystem is treated as a black box, and the “how” and “who” (biotic and abiotic) that mediate inputs and outputs is not the domain of ecosystem ecology (Weathers et al., 2021).

Community Ecology: aims to understand patterns in the composition and diversity of species that co-occur in space and time. This involves selecting a subset species and abiotic conditions from the ecosystem. The field is very interested in describing diversity and community composition, and how this community was formed, will persist, or might change (within the realm of: selection dispersion, drift, and speciation; (HilleRisLambers et al., 2012; Nemergut et al., 2013; Vellend, 2010)).

Population Ecology: is a specific facet of community ecology. In this context, the researcher is interested in a single taxon and is generally interested in the factors that alter the size, spread, and persistence of members of the community (Krebs, 2015). Generally, the metrics of study include demographic traits (population size, population growth rates: birth, death, emigration and immigration rates) and abiotic factors (food, niche space, shelter availability...). This discipline can quickly morph into community ecology when “food” is treated as an independent community member, and death rate acknowledges that a focal taxon being killed is often a biotic process (e.g. bacteria, viruses, wolves...) and the killer has its own ecology (e.g. disease ecology), and niche/ shelter availability is in-part dependent on competitive exclusion by other organisms (Sibly & Hone, 2002).

Behavioral and Physiological Ecology: are the most reductionist forms of ecology focusing on the external and internal manifestation of physiological processes respectively. Where physiology acts on other organisms (e.g. effectors, pheromones, quorum sensing, resource acquisition), it very quickly bleeds back into the realm of population and community ecology.

Microbiologists operate at each of these ecological/ biological scales. Biogeochemists concerned with mass balance are clearly ecosystem ecologists, 16S rRNA gene sequencing

specialists examining community structure are firmly community ecologists, microbiologists calculating growth rates and constructing bioreactors are population ecologists, while classical knockout microbiologists are physiological ecologists. Despite activity at each of these scales, there is limited crosstalk between them. This may be due to the siloing of each of these disciplines (e.g. biogeochemistry as the purview of earth sciences departments, molecular geneticists and biochemists generally trained in a medical context, and all generally separated from an ecology department) and necessary specialization of toolkits (e.g. isotopically labeled rate measurements, mathematical modeling of community ecology, and designing vectors and knockouts are extremely discrete tasks).

This segregation of disciplines has naturally led to many labs (that may be accurately considered microbial ecologists) operating at a single scale with a constrained toolkit, asking the same sort of question (e.g. using 16S rRNA gene sequencing to ask: how does community composition change in the face of some disturbance?). This fragmented environment relies on a vague outsourcing of ecological scaling to *someday* connect structure to function, and *maybe* ground this in a broader context. This outsourcing will be slow, the relevance of any observation unclear in the interim, and may never finish for esoteric systems. Drawing from the language of economics and supply chain efficiency (Coase, 1937; Hovenkamp, 2010) I term this fragmentation “horizontally integrated” ecology.

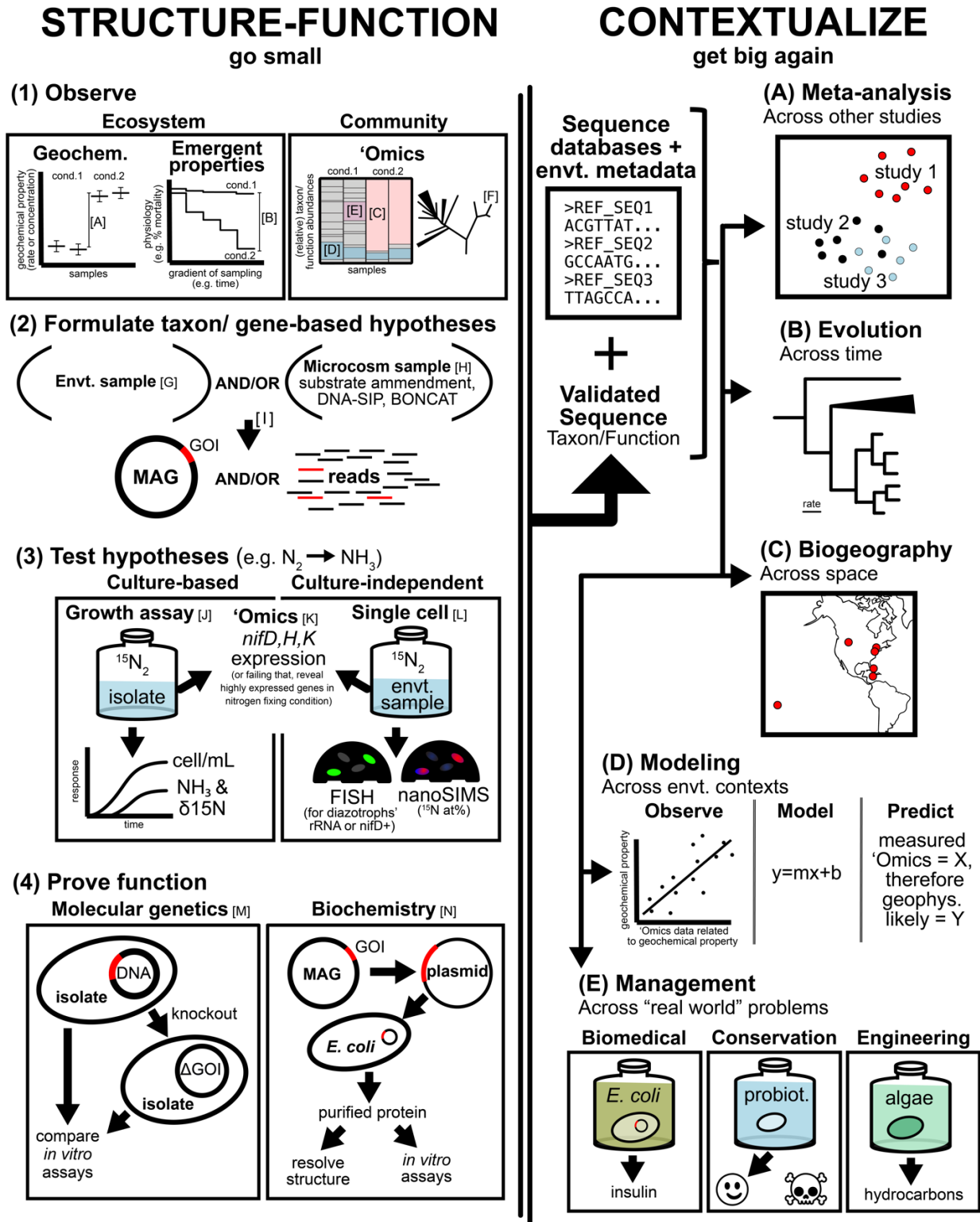
1.1.3: Vertical Integration: Examining a Single Process at Nested Ecological Scales.

“**Vertically integrated**” ecology (nesting scales within a paper or research program) convincingly links structure and function and allows contextualizing this function in our world. In practice, this vertical integration might look like a research team asking: 1) do we see

something interesting happening in the environment?, 2) what is the composition of the microbial community mediating this process?, 3) what microbes encode the genomic capacity to perform this task?, 4) can we prove that this microbe performs the hypothesized function?, 5) what is the global distribution of the microbe that mediates this process? This simplified example of “vertical integration” quickly highlights a need for a methodologically diverse workflow. Luckily many of these tools are widely disseminated, system-agnostic, have a small physical footprints, and/or are shared resources (e.g. microscopes, compute clusters sequencing platforms, chromatographers, spectrometers, GIS, etc...). **Vertically integrated ecology is attainable, efficient, and scalable.** In the following paragraphs, we will describe a workflow to apply vertically integration and contextualize the findings for microbial ecologists.

1.2 – A Vertically Integrated Workflow for Microbial Ecology

Using the vertical integration framework, we can categorize most environmental microbiology endeavors into two categories: **structure-function** and **contextualizing** (summarized in Figure 1). In a general sense, “structure-function” progressively zooms in on an observed ecosystem or community feature to identify and describe essential microbes and/or functions. Inversely, “contextualizing” uses an established link between genomic pattern and emergent process to infer how this gene (taxon or function) might matter in the wider world. Both these workflows integrate a variety of methodologies, and we’ll look at the concrete steps in more detail below.



(“Contextualize: get big again”). Structure-function generally includes these steps: **(Observation)** To identify features (functions/ taxa) worth pursuing often involves observing a process or taxa at **A,B**) ecosystem or **C-F**) community scale that may sustain further investigation. **(Formulate Taxon/ gene-based hypotheses)** To formulate specific hypotheses of function is often aided by genomic data sourced from the **G**) environment or **H**) microcosms amended with activity indicating or metabolism enriching compounds. These data can be used to **I**) generate genomic data (read and/ or genome-based) and formulate hypotheses linking taxa and function. **(Test hypotheses)** To test hypotheses of function, culture-based or culture-independent approaches can be used to make high-confidence inferences of realized function. For example, demonstrating taxon-specific nitrogen fixation may be performed by assessing **J**) isolate or **L**) single cell assimilation of labeled N₂, and/ or **K**) expression of NifH,D,K genes associated with a metagenome assembled genome. **(Prove function)** To definitively prove a sequence mediates a function requires **M**) molecular genetics and/or **N**) biochemical assays. This is illustrated with a workflow that might involve knockouts and add backs of genes of interest and/ or purification and assaying generated protein. After validating a gene’s relationship to a particular function, this sequence can be placed in one of four broader contexts: **(Meta-analysis)** The sequence can be contextualized against the results of other studies via meta-analysis. **(Evolution)** the sequence can be contextualized in time (development trajectory, co-occurrence with other taxa) using phylogenetic inference. **(Biogeography)** the distribution of the sequence can be shown over space and potentially used to scale associated processes over space. **(Modeling)** the presence and abundances of the focal sequence may be used to predict ecosystem-scale processes. at%, atomic percent; cond., condition, envt., environment; GOI, gene of interest; geophys., geophysical; geochem., geochemistry; MAG, metagenome assembled genome; probiot., probiotic.

1.2.1 – Structure-Function (reductionism). The process of causally linking a microbe and its metabolism to an emergent ecosystem property has four major steps: 1) observing, 2) formulating sequence-based hypotheses, 3) testing hypothesis, and 4) proving function. These steps are generally followed in this order, but depending on the specific question asked and methodologies employed, some of them can be skipped.

Step 1 – Observe:

At this step the environmental microbiologist observes (from primary data or secondary literature) high-level microbial-adjacent processes that inform hypotheses and future investigation. Effective observations at this point should give the general sense that: “there is

something interesting occurring, it is likely microbiological, and it bears further investigation”.

These observations are generally made at **ecosystem-scale** and/or **community**-scales.

Ecosystem-scale “black box” observations measure inputs and outputs from a defined system (e.g. water column, coral colony, sediment core, etc...). In the context of environmental microbiology, these measurements should be microbially relevant and provide evidence of microbial processes that can be dissected later using other methods. One class of observation is **geochemical** measurements. Geochemistry is very “black box”, as it generally measures a process (e.g. rates and/or standing concentrations of a compound of interest), but does not explicitly identify any microbial actors. Measures of standing concentrations might inform the question: “how might microbes respond to this: stressor/ nutrient/ redox potential?” (e.g. (Kumagai et al., 2018; Shoemaker et al., 2024; Steinsdóttir et al., 2022), Chapter 2), while rates show active metabolism and raise the question of “what/ who mediates this reaction?”. **Other emergent properties** is a catch-all for non-geochemical observations. This includes observing changes to host or adjacent species physiology (e.g. growth rate, photosynthetic efficiency, mortality, respiration rate,...; e.g. (Clements et al., 2024; Q. Wang, McDermott, et al., 2021), Chapter 3). If these processes appear microbially associated, they should raise the question of “what taxa and genes are responsible for this process?”.

Community-scale observations assess the taxonomic or functions composition of a defined space (e.g. water column, coral colony, sediment core, etc...) with the implicit assumption that understanding the function of individual community members will clarify collective function. This process requires identifying community features (taxa or functional potential) for further investigation. Tractable community features often exhibit some

combination of six noteworthy traits: **1) high abundance** (e.g. (Bayer et al., 2013; Patin et al., 2021), Chapter 4), **2) ubiquitous prevalence** (e.g. (Klinges et al., 2019), *Wolbachia* (Fromont et al., 2019), Chapter 4), **3) association with a condition of interest** (e.g. (Tschitschko et al., 2024), Chapter 2, Chapter 3), **4) *a priori* interest** (e.g. (Frias-Lopez et al., 2008; Padilla et al., 2016), Chapter 4), **5) divergence from expectations** (e.g. (Evans et al., 2015), Chapter 4), or **6) novelty** (e.g. Anammox metabolism (Van Niftrik & Jetten, 2012), CPR bacteria (Hug et al., 2016), Chapter 4). Any of these observations may raise the question “what is this [interesting] gene/taxon doing?”. **Exploratory ‘omics** (i.e. analysis based on nucleic acid sequencing; specifically metagenomics and transcriptomics here) are well suited for identifying these community features as they provide an accurate representation of the taxonomic and metabolic composition of a microbial community (Hauptfeld et al., 2024), which can be used to inform mechanistic hypotheses and future investigations. The questions raised by ecosystem or community-scale observations provide a framework to formulate hypotheses to inform further investigations.

Step 2 – Formulate Sequence-based Hypotheses:

Here, observations from Step 1 are turned into testable hypotheses. Generating these hypotheses is often facilitated by further ‘omics. These provide a sense of which microbes and metabolisms may be responsible for observed ecosystem (geophysical or physiological) process or identify the functional potential of taxa or genes of interest. It’s important to note that this step of sequencing for hypothesis generation can be skipped in a targeted culturing approach, where thoughtful amendments can produce isolates that demonstrate metabolisms of interest, which subsequently source genomes/transcriptomes (e.g. (Keller et al., 2025)). However, the cultivation

of novel microbes can benefit from a genomic understanding of auxotrophies (Tripp et al., 2008). As such, in both cultivation and culture-independent approaches, it's often worthwhile to sequence the microbiome to link taxonomy and function and/or identify sequences of interest. Sequencing for hypothesis generation can be performed in untargeted or targeted manners.

An **untargeted** approach uses samples collected directly from the environment, generating sequences from the entire community. This leaves the burden on the researcher to pare down sequences of interest using taxonomic and functional annotations (e.g. (Pratte et al., 2018)). A **targeted** approach simplifies bioinformatic inference by enriching and/or identifying taxa of interest before sequencing making these microbes more prominent in analyses. This is often performed using bioreactors or microcosm incubations of environmental samples that can be amended with substrates to 1) identify metabolically active microbes (e.g. DNA-SIP (Trubl et al., 2021), BONCAT (Trexler et al., 2023)) and/or 2) enrich microbes that perform a physiology of interest (e.g. (Baker & Girguis, 2024)), such that sequencing preferentially captures active microbes associated with the metabolism of interest, effectively cutting through microbiome complexity.

After sequences are produced, they can be analyzed utilizing metagenome assembled genomes (MAGs) and/or raw reads. MAG-resolved analyses are conceptually straightforward and appealing because genomes unify multiple sequences under a single taxonomic label, linking any metabolic potential to a specific microbe (Tyson et al., 2004). Though these stories are clear, MAG construction loses some sequence information (Hauptfeld et al., 2024), potentially missing useful taxonomic or metabolic information. Identifying taxonomy (e.g. Chapter 3,4) or metabolic potential (e.g. (Bertagnolli et al., 2023); Chapter 2) from unassembled read data allows a more

comprehensive, unbiased way to identify community-scale taxonomic and or metabolic potential (Hauptfeld et al., 2024). Reads can also be used to screen datasets for genes of interest before MAG generation (Speth & Orphan, 2018). If a tractable metabolism is identified, it can then be tested with a variety of experiments.

Step 3 – Test Hypothesis:

Here, genome-based hypotheses from Step 2 are tested with targeted assays and begin to provide concrete evidence that a particular sequence (function or taxon) is associated with the Step 1 observed process. In this step, we will highlight application of 1) culture-based and 2) culture-independent approaches to provide evidence of activity and how each approach utilizes ‘omics as direct evidence of activity of for method development.

Culture-based *in vitro* assays provide extremely clear and useful platforms for linking taxonomy and function. If an isolate can perform a function observed in Step 1, it provides convincing evidence that this microbe is (in-part) responsible for the ecosystem condition (e.g. (Alcolombri et al., 2021; Segev et al., 2016)). The controlled capacity to perform bioassays with isolates is extremely powerful but is not universally applicable as most microbes have not been cultivated and are challenging to cultivate (Park et al., 2022; Singh & Reddy, 2016)(Amann et al., 1995). Though MAGs have been used to improve cultivation of resistant microbes (e.g. (Tripp et al., 2008)), cultivation of all microbes is an unrealistic goal. Culture-independent approaches have been developed to study these “wild” microbes. **Culture-independent** tools include bulk geochemistry, ‘omics, and single cell approaches. Bulk geochemistry measures the integrated activity of all community members and cannot resolve individuals in a sample, which is not always appropriate for linking taxonomy and function. Both ‘omics (e.g. (Frias-Lopez et

al., 2008; Muratore et al., 2022)) and cell visualization (e.g. (Reichart et al., 2020)) can resolve taxon-specific activity. Here, I will focus on single cell visualization approaches (as supported by ‘omics) as they allow the overlay of taxonomy, morphology, and metabolism for a single cell.

Single cell taxonomy, gene content, and/ or transcriptional activity is often identified using different varieties of fluorescence in-situ hybridization (FISH) probing (with specific probes designed from MAGs of interest as identified in Steps 1 and 2; e.g. (Kitzinger et al., 2020)). The fluorescent signal from hybridized probes can be paired with epifluorescent microscopy or flow cytometry to quantify target cells in a mixed population (Kitzinger et al., 2020). Though cell counts are useful, FISH can be even more powerfully combined with substrate amendments to identify cell-specific metabolic activity. When uncultured microbes are incubated with hypothesized growth substrates labeled with heavy isotopes, metabolically active microbes will incorporate the heavy isotope into their biomass. This can be taxonomically and functionally resolved by first identifying taxa of interest with FISH staining and then subsequently subjecting them to RAMAN spectroscopy (non-destructive) or nanoSIMS (destructive), which can identify isotope assimilating cells (Schaible et al., 2022; Tschitschko et al., 2024). Depending on how these steps are strung together, a researcher could realistically identify and quantify cells from a taxa or function of interest with FISH, identify stable isotope incorporation with RAMAN, isolate select cells with microfluidics or optical tweezers, and generate single amplified genomes (SAGs) from the tagged cells, convincingly linking taxon and function (Ishoey et al., 2008; Sturm et al., 2023).

In both culture-dependent and culture-independent contexts it is possible to use ‘omics as a standalone or supporting data source. In all cases transcripts can be used to demonstrate the

metabolic “intention” of a microbe to perform a function (e.g. (Kohtz et al., 2024; Mitchell et al., 2024), Chapter 4).

Step 4 – Prove Function:

In this final step of “structure-function”, any ambiguity from Step 3 can be resolved by *proving* function using molecular genetics and biochemistry. This level of reductionism is not standard for environmental microbiology (realistically most researchers stop at Step 1) as protein purification and biochemical assays may be overkill if Steps 1-3 are highly compelling. But if relevant, biochemical evidence is often definitive proof of activity. To describe the general workflow of proving function, we will look again at molecular genetics and biochemistry approaches.

In a genetic context, if there is an environmental isolate that performs a function of interest *and* it is genetically tractable, it may be possible to knockout gene(s) of interest and assay changes in isolate biochemical function (Trotter et al., 2023). If a knockout stops the metabolism of interest, it is solid evidence of sequence-function association (e.g. (Q. Wang, Alowaiifeer, et al., 2021)). To definitively characterize the gene of interest and prove function, it would be necessary to perform biochemical assays with the purified protein (Dong et al., 2024). Purified protein could be sourced from an isolate with sufficient biomass. But if that is not possible, it could be produced heterologously from cultured or uncultured microbes in a competent host (e.g. (Tsementzi et al., 2016; Q. Wang, Alowaiifeer, et al., 2021)). The protein could then be used to perform definitive *in vitro* assays of function. If useful, the protein structure could be resolved using a combination of alphafold and cryoEM approaches (e.g. (Burman et al., 2024)).

Progressing through these steps results in increasingly confident linkage of observed structure and function by connecting a particular genomic sequence to ecosystem or community function. After this function-gene connection is established, this genomic sequence can be used to contextualize larger scale processes.

1.2.2 – Contextualize (holism). Clarifying a sequence’s function alone is not sufficient, it must be placed in a broader context to be *important*. Sequences (taxon or function) can be situated into a few larger contexts: **A)** identifying patterns in the distribution or variation in the target sequence(s) across datasets (i.e. metaanalysis), **B)** identifying how the sequence developed (i.e. evolution), **C)** identifying where the sequence occurs (and may operate) spatially (i.e. biogeography), **D)** and under what conditions the sequences may describe or predict ecosystem processes (i.e. modeling), **E)** applying sequence-based knowledge to solve real world problems (i.e. management; e.g. engineering, remediation, medicine, conservation). This contextualizing of sequences is a bioinformatics exercise requiring not only validated genomic sequences linked to function and/ or taxa, but also sequence databases with sufficient contextualizing metadata (geographic coordinates and geophysical conditions of sampling). The acquisition and use of reference data is known as data mining and will feature in all these contexts, but specific methodological details will not be discussed here.

Context A – Meta-analysis:

Meta-analysis contextualizes the results (i.e. sequences) of a focal study with the results of other studies to answer the question: “how do my results [writ large or small] fit into the existing literature?”. Meta-analyses might compare microbiome (genes or taxa) or geochemical results across biotic (e.g. (Clavere Graciette et al., 2023; Pratte et al., 2022; Thompson et al.,

2017)) or abiotic environments (e.g. (Shade et al., 2013; Thompson et al., 2017)). The results of these comparisons might show that the focal study aligned or diverged from current expectations in the literature. Alignment might support existing hypotheses or theories, while divergence might prompt the follow-up question of “what explains this difference?”, which may identify new taxa and functions to interrogate in Step 1 of the “structure-function” workflow.

Context B – Evolution:

An evolution context places a taxa or gene amongst other sequences to answer questions like: “how does microbe X fit into the current tree of life?”, “how closely related is microbe/ gene X to Y?”, and “when did gene Z originate?”. This uses publicly available datasets to identify shared sequences and calculate sequence similarity. The resulting phylogenetic tree can be used to contextualize the timing of biological (e.g. (Henriques et al., 2024; McCutcheon & Moran, 2012)) and biogeochemical processes (e.g. (Colman et al., 2018; Fischer et al., 2016)) or to suggest that related taxa may share unproven functions due to sequence similarity to other validated sequences (e.g. (Henriques et al., 2024; Porras et al., 2024)). This is particularly impactful when describing co-occurrence with major biological (e.g. origin of: life, eukaryotes, photosynthesis, symbioses) and geophysical (e.g. great oxidation event, asteroid impact) events.

Context C – Biogeography:

Biogeography contextualizes the distribution of taxa and functions across spatial extents. This involves screening publicly available datasets for similar genes (taxa or metabolisms) with coordinates and then mapping them. This microbial mapping can be an omics-only exercise across global (Choudoir et al., 2016; Thompson et al., 2017; Tschitschko et al., 2024) or regional (López-Pérez et al., 2020) extents, but the tiny size of microbes also allows *mapping* on

centimeter (e.g. body sites; (Pollock et al., 2018; The Human Microbiome Project Consortium, 2012)) or micron (e.g. biofilms; (Mark Welch et al., 2016; P. S. Stewart & Franklin, 2008)) scales. Biogeography may use sequence distributions to inform new data collection (Sogin et al., 2022), or may use spatial gene distributions to mathematically scale measured rates (Al-Haj & Fulweiler, 2020; Bianchi et al., 2018; Whitman et al., 1998).

Context D – Modeling:

A model is a structured way of describing a process that enables empirical testing of the relationship between input(s) and response(s). This provides a framework for generalizing, estimating, and predicting (Marquet et al., 2014). Model creation forces the researcher to explicitly state assumptions, define parameters of interest, and work out the consequences logically (Fox, 2011). Modeling generally begins conceptually (verbal or pictorial models) as a first pass defining important parameters (e.g. environmental, physiological, kinetic) and developing a consistent logic. These heuristic models may be sufficient for some applications (e.g. metabolic models in Section: “Formulate Sequence-based Hypotheses”), but may require mathematical articulation to work through interactions in a complex system (Fox, 2011). Numerical models may be mechanistic (describing relationships between theory and observation; (Marquet et al., 2014)) or predictive (using as many variables as needed to predict a response; (Zakem et al., 2020)). In all cases, model making and interpretation can identify relationships, missing data, or faulty assumptions, which can drive new research. Gene-based ecosystem modeling is a clear application for microbial ecology (framework for microbial ecology modeling in (Lui et al., 2021)). Gene-based modeling attempts to use gene or transcript abundances to predict ecosystem processes. This involves using new and/or publicly available

datasets of taxa and/or metabolisms (with attached ecosystem-scale metadata) to model gene-ecosystem association, test predictive power, and apply the model to predict outcomes. This prediction may be most familiar in molecular pathogen screening, where the presence of a pathogen often predicts infection (Sethuraman et al., 2020; Vilanova-Cuevas et al., 2025). But the presence of functional or taxon-specific genes can also be used to predict toxicological (Pacheco et al., 2016) and biogeochemical (Louca et al., 2016; Reed et al., 2014) outcomes.

Context E – Management:

Applying microbiology to solve non-academic problems is the practice of microbial management. The most ancestral, and not-particular scientific, means of microbial management was for food production (e.g. fermentation (Aouizerat et al., 2019)) and disease prevention (e.g. wastewater treatment, (Angelakis et al., 2022); vaccines, (Pasteur, 1881)). But, as in macroecological systems (Giakoumi et al., 2017; White et al., 2019), microbial management benefits from a solid foundation of research. Just such reductionist studies of microbial communities, taxa, and genetics have applied microbes to fantastic effect, featured in biomedical sciences (e.g. *E. coli* synthesizing insulin (Goeddel et al., 1979); fecal microbiota transplants, (Eiseman et al., 1958)), engineering (e.g. bioremediation (X. Li et al., 2021), industrial culture (Miller et al., 2024)), wastewater treatment optimization (Huang et al., 2023)), and conservation (e.g. disease attenuation (Rosado et al., 2019)). This application of microbiology is a context where its ecological basis is most often lost, as practitioners spawn whole new fields that become further and further removed from its ecological roots.

1.2.3 – Integrating Reductionist and Holist Perspectives. These reductionist and holist workflows operate complementarily. Contextualized structure-function relationships can identify

gaps in our understanding of taxa or functions in an ecosystem, spawning new reductionist research (Bianchi et al., 2018; Tschitschko et al., 2024). Further, this framework is accommodating, where researchers can integrate their data into these workflows wherever convenient.

1.3 – Conclusions

Ecology is predicated on cutting up the biosphere into tractable chunks. Each chunk requires description at subsidiary scales (ecosystem to molecular; (Garren & Azam, 2012)) and then pasting the ecosystem back into the rest of the world. We have articulated an ecologically informed workflow to both describe functions and contextualize the results.

Now, we will now dig into some specific tools underpinning this workflow. In the first section we identified multiple use-cases for sequence-based (‘omics) analyses and how they can bridge disciplines. The next two sections will describe the data products ‘omics can generate and the specific tools available to create these outputs.

Section 2: ‘Omics Data Products

A guiding experimental framework is essential to interrogating microbiomes, but an understanding of the available tools for inquiry is extremely helpful. In the following two sections we will pivot to discussing the uses of sequence-based analyses (‘omics) as they support the above workflows of microbial ecology.

2.1 – What is ‘Omics?

‘Omics describes a computational toolkit for the analysis of biological molecules. The prefix for each flavor of ‘omics is determined by the biological molecule in question (e.g.

genomics: DNA; transcriptomics: RNA; proteomics: proteins, metabolomics: catch-all for chemical outputs from metabolic processes). Here, our use of the term ‘omics refers to the nucleic acid sequence-based genomics and transcriptomics. The analysis of biological sequences allows the association of motifs with taxa and functions. This conserved context allows the comparison of new biological sequences against ever-expanding databases, enabling hypothesis generation, extrapolation and contextualizing of experimental data. In the next paragraphs, we’ll briefly summarize the history of ‘omics, and highlight current considerations for its use.

Nucleic acid sequence-based studies have progressed rapidly over the last 50 years. The identification of a third domain of life in the 1970s used tiny fragments of RNA compared to the size of the genomes that generated them (Woese & Fox, 1977). PCR of environmentally derived DNA in 1980s allowed the generation of large volumes of small fragments of DNA (Lane et al., 1985), but without genomic context. Concurrently, the publication of the first DNA genome (phiX) in 1977 (Sanger et al., 1977) ushered in the genomics age. Genome sequencing advanced quickly to achieve the first prokaryotic genome in 1995 (Fleischmann et al., 1995) and human genome in 2001 (International Human Genome Sequencing Consortium et al., 2001; Venter et al., 2001). The advent of mixed microbial community genome reconstruction followed quickly in 2004 with the first metagenome-assembled genome (MAG), opening the door to widespread reconstruction of genomes from non-model, un-cultured microbial communities (Tyson et al., 2004). The adoption of transcriptomics (sequencing RNA to map back to a known genome) began in mammalian cultures in 2006 (Bainbridge et al., 2006) but was quickly applied to mixed environmental microbiomes in 2008 (Frias-Lopez et al., 2008). This capacity to generate genomes and approximate metabolic activity from transcriptomes, has allowed researchers to

generate new hypotheses and make predictions of novel metabolic potential from un-observed communities. In the following paragraphs and sections, we will describe a vertically integrated microbial ecology workflow, highlighting when and how ‘omics fits into a broader methodological picture.

At this point the utility of ‘omics is not a question, as genomic and transcriptomic data is used at almost every step of environmental microbiology workflows for result generation or method development. Despite the utility of ‘omics, there is a sense of two-sided myopia between bioinformaticians and bench scientists. Many bioinformaticians are insulated from the generation of scientific questions, called-on to assemble and annotate metagenome assembled genomes (MAGs), but not necessarily involved in the nuances of project design or exposed to the breadth of the field in which their MAG will be situated. Conversely, wet-lab scientists may have a narrow view of bioinformatics, seeing it as a relatively simplistic black box for generating genomes, lacking perspective into the flexibility and utility of non-MAG approaches. The meeting of both perspectives seems to foster a relatively uncritical view of, and overreliance on, MAGs as a one-size-fits all solution for environmental microbiology. In the following sections we will describe the major data products of ‘omics (that include and extend beyond MAG generation) and their general considerations for use.

2.2 – Establishing Analytical Goals: Putting the Bio in Bioinformatics

‘Omics analyses (bioinformatics) is a technical skill like that found in wet-labs or fieldwork. Unlike the concrete physical challenges of lab- and fieldwork (limited quantities, expiration, specialized handling, and sample preservation), central bioinformatics challenges revolve around processing power (which is not the focus of this review and will not be addressed

here) and data storage. Bioinformatic analyses often run into the odd challenge of having *too much data* rather than not enough (Stephens et al., 2015). Large datasets combined with near-costless computing means that unguided bioinformatic analyses can generate near-infinite (and completely intractable) quantities of “results”. The clearest bioinformatic guidance comes from a well-articulated question (e.g. “what are the tradeoffs involved in conserving energy from light?” (Kumagai et al., 2018); “what pH did life originate in?” (Colman et al., 2024); “how does habitat specificity affect global patterns of speciation?” (Sriswasdi et al., 2017)). To be useful scientifically, this question must also address a gap in available non-bioinformatics literature. Addressing this twofold requirement of scientific rigor and (micro)biological importance is not immediately solved by informatics skill. Many bioinformaticians are skilled computer scientists, building and using analytical tools, but that does not necessarily overlap with biological expertise (Vincent & Charette, 2015).

Effective use of bioinformatics within the ecological frameworks described previously requires biological domain-specific expertise, and a minimum technical skill. We will address that technical skill below.

2.3 – The Data Products

Assuming a would-be bioinformatic practitioner has the required (micro)biological expertise and has formulated a scientifically meaningful guiding question, the next step is identifying tractable computation goals.

Guided by a question and data constraints, a bioinformatician can then identify computational endpoints (e.g. “do we need genomes?”, “Are we surveying genomic/ taxonomic content?”, “Do we need to contextualize the data phylogenetically?”, “Do we need

abundances?”...). This procedural endpoint allows the bioinformatician to work backwards to construct an analytical pipeline, identifying subsidiary questions along the way. To effectively outline goal checkpoints and endpoints, the practitioner needs to identify goal data products. Almost all sequence-based ‘omics data products fall into one of five classes **1) sequence files, 2) sequence statistics, 3) taxonomy tables, 4) function tables, and 5) count tables**. The generation of any of these products is a technical (not scientific) output and must be incorporated into a larger biological narrative to be useful. With that in-mind, we’ll now describe the general structure and uses for each of these data classes (summarized in **Figure 2**).

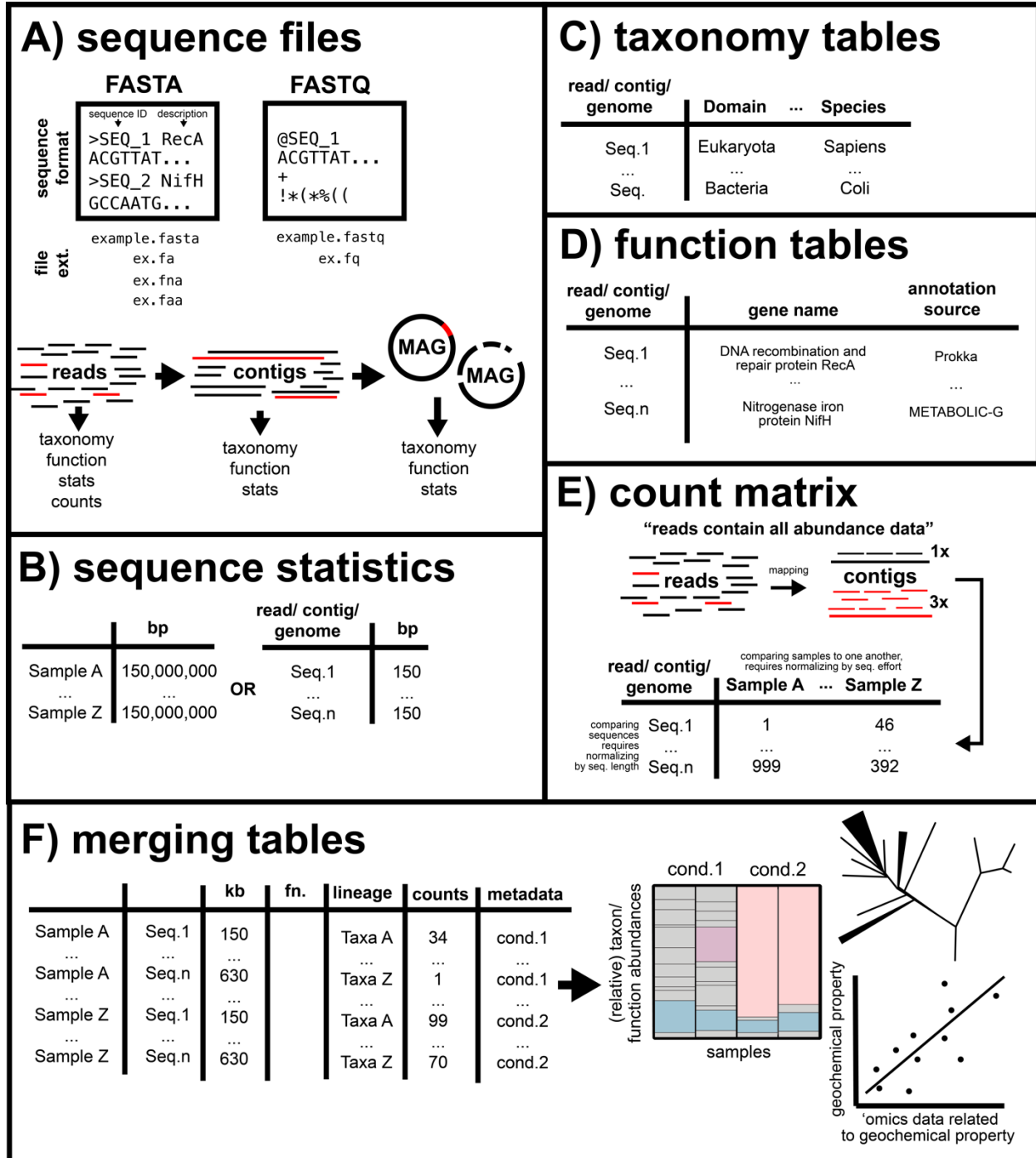


Figure 2: Simplified examples of the major ‘omics data products. **A)** sequence files, describes the major file formats for sequence data, and the data types that can be derived from each. **B)** sequence statistics, describes the general structure of descriptive statistics derived from different sequence data. **C)** taxonomy tables, describes the general structure of taxonomy tables. **D)** function tables, describes the general structure of function tables. **E)** count tables, describes the

general workflow for deriving count data and the structure of count tables. **F)** merging tables, describes the general structure of the result of merging tables, and mock data products of the from merging taxonomy, function, count, and statistics tables. cond., condition; ex., example; ext., extension; MAG, metagenome assembled genome; seq., sequence; stats, statistics.

2.3.1 – Sequence Files. Sequence files are the basis of all bioinformatic analyses.

Sequences are most commonly encoded in FASTA or FASTQ files. FASTA files use a sequence format that contains a sequence identifier and sequence data (nucleic acids or amino acid residues). FASTQ files are more specialized and are generally the output from nucleic acid sequencing. They contain a sequence identifier, the sequence, and quality scores (Q-scores). These quality scored bases allow the user to remove low-quality sequences and/or bases (which generally produce quality filtered FASTA files) before sequence analysis (**Figure 2A**).

There are three major classes of sequence files: 1) reads, 2) assembled contiguous sequences (contigs), and 3) genomes. Each of these classes are nested, with the previous sequence file used to derive the next (e.g. reads are used to make contigs, contigs to make genomes). We will break down the characteristics of each data type here.

Reads are the base unit of sequence analysis and are the raw product of sequencing platforms and are found as “short” or “long” depending on the sequencing technology used. Short reads are tens to hundreds of bases long, while it is not uncommon for long reads to be thousands to millions of bases long (Satam et al., 2023). Though read data is also the basis of marker gene (e.g. 16S rRNA gene) studies, we will focus here on reads generated from metagenomic and transcriptomic sequencing. Reads generated without targeted amplification are a relatively unbiased representation of the genomic composition of the sample (Hauptfeld et al., 2024). As such, these reads contain all the (relative) abundance information in a sequence library

(Gloor et al., 2017). Read sequences also contain fragments of taxonomically and functionally relevant marker genes. These marker genes can be used to survey taxonomic diversity in a metagenome or transcriptome, akin to rRNA gene surveys (Dragone et al., 2022). Fragments of metabolic and physiologically relevant genes can be used to identify and quantify genes of interest (Bertagnolli et al., 2023; Martínez-Pérez et al., 2022). Searching for functions in read data can also be used to sift through reference databases to pare down datasets to only those that include a metabolism of interest (Speth & Orphan, 2018). Though reads have a high degree of utility, the most common use of read data is the construction of contigs.

Reads can be assembled into longer contiguous sequences (**contigs**). This is used in metagenomics to create genomic scaffolds (Z. Li et al., 2012), and assembly-based transcriptomics to generate transcripts (Grabherr et al., 2011). Assembly decouples contig abundance in a sequence file from “true” environmental abundance. Thus, quantifying contig sequences in the sample, requires mapping reads to assembled contigs. These longer contig sequences are useful for community taxonomic and functional reconstruction because their longer length enables the search for complete genes of interest. Beyond identifying individual genes on contigs, if a contig contains multiple genes, this genomic neighborhood can be used to infer contig-wide shared function (Mihelčić et al., 2019) or taxonomy (Mirdita et al., 2021). Contig-derived genes are also potential inputs for phylogenetic reconstruction of marker gene or protein trees, enabling evolutionary and taxonomic inference of contigs. Though contig sequences have a high degree of utility, the most common use of contigs is the construction of genomes.

Genomes are made from contigs by identifying and grouping (binning) sequences that appear similar (e.g. nucleic acid coding qualities: GC content, oligonucleotide composition; and correlated sequence abundances). In an environmental microbiology context, most genomes are metagenome assembled genomes (MAGs), but both MAGs and isolate-derived genomes are created by the same general pipeline. Here we will focus on MAGs, though any discussion of genomes is generally applicable to isolates too. MAG-based analyses fundamentally allow the defensible splitting of a mixed-together set of reads and contigs into discrete genomes, even if those organisms have been never previously observed. Genomes are an extremely convenient type of sequence data because binning creates a biologically meaningful association between multiple contigs. In practice, if a taxonomic marker gene is found on one contig in a bin, and a relevant gene of function is found on a different contig in the same bin, though they are on different contigs, they can be defensibly attributed to a single organism and that taxon assigned metabolic potential. This quality of MAG analyses allows the characterization of metabolic potential in individuals and communities ((Shoemaker et al., 2024), **Chapter 3**). Overlaying MAGs and transcriptomic sequencing can identify “metabolic intention” in poorly understood individuals (Kitzinger et al., 2020; Porras et al., 2024; F. J. Stewart et al., 2012) and communities (Muratore et al., 2022). Beyond community description and reconstruction, MAGs (and transcriptomes) can also be used for microbial discovery and cultivation. Uncultivated microbes can be identified and novel functions in their genomes hypothesized (Evans et al., 2015; Wurch et al., 2016) and tested (Graf et al., 2021). The use of genomic (Tripp et al., 2008) and transcriptomic (Bomar et al., 2011) data can enable or enhance cultivation success.

2.3.2 – Sequence Statistics (Length and Effort). Sequence statistics are a derivative data product from sequence files. Sequence stats have two major purposes 1) contextualizing a narrative (e.g. describing dataset size and complexity and/or sampling effort) and 2) analytical refinement (e.g. normalizing count data). This generally comes down to generating tables of total bases in each sequence library or individual sequences (e.g. reads, contigs, or genomes) and are output as tables (**Figure 2B**).

2.3.3 – Taxonomy Table. Taxonomic classification of sequence data (reads, contigs, or genomes) compares a query sequence against that of taxonomically confident reference sequences. If the query sequence appears sufficiently similar to a reference sequence (Goris et al., 2007; Jain et al., 2018; Parks et al., 2020) the sequence is assumed to share the taxonomic lineage of the reference sequence. Any method of taxonomic classifications aims to generate tables that relates a sequence identifier to a taxonomic lineage (Figure 2C). There are two general approaches to assigning taxonomy to a sequence: 1) direct annotation and 2) inferential annotation. **Direct annotation** can be performed on any sequence, where it is directly compared to reference sequence databases (e.g. SILVA for rRNA sequences, NCBI nt/nr for nucleotide or protein sequences respectively). **Inferential annotation** is most common in MAGs, where the classification of a taxonomic marker gene in a MAG confers taxonomic annotation to all other sequences in the MAG. If a taxonomic classification is in-question or it is essential to prove, it is common to support database taxonomy with confident phylogenetic clustering amongst other sequences from that lineage ((Keller et al., 2025), Chapter 4).

2.3.4 – Function Table. Functional annotation of sequence data (reads, contigs, or genomes) compares a query sequence against annotated (and potentially experimentally

validated) reference sequences. Any method of function annotation aims to generate tables that relates a sequence identifier to a descriptor of function (Figure 2D). Function annotation generally serves one of two purposes: 1) targeted searches or 2) broad functional descriptions. **Targeted searches** can query reads or contigs (including MAGs and transcripts) and involves interrogating the presence and abundance of 1-100 genes of interest (e.g. (Leung & Greening, 2020)). This may require manually curating target gene databases if they are not common annotation targets. This approach can be used to identify marker genes in a focal dataset (e.g. (Bertagnolli et al., 2023; Martínez-Pérez et al., 2022) , **Chapter 3**) or to screen multiple datasets to find those containing a metabolism of interest (Speth & Orphan, 2018). **Broad description** generally uses longer sequences containing complete protein coding regions (contigs). It is common to use these protein coding regions to search against a default list of genes (e.g. (Huerta-Cepas et al., 2017; Seemann, 2014; Shaffer et al., 2020; Zhou et al., 2022)). Still unannotated genes can be annotated using a general database (e.g. NCBI nt/nr or UniProt kb, e.g. (Mitchell et al., 2024)). In any case, preliminary annotations of genes that are hard-to-annotate, novel, or pivotal to a manuscript's narrative may need validation against reference genes, searching for essential structural motifs (e.g. (Garber et al., 2020; Porras et al., 2024; Søndergaard et al., 2016)) and/or monophyletic placement within a reference tree (e.g. (Porras et al., 2024)). Two important caveats regarding using and interpreting functional annotations. **First**, it should always be noted that the quality of any annotation is tied to the completeness and applicability of the reference database. Sequences from taxa related to model organisms (*E. coli*, *Pseudomonas sp.*, etc...) are best represented in reference databases and provide useful annotations for conserved functions (ATP synthesis, glycolysis, etc...), but for more exotic taxa

and genes, annotations may be lacking (Goodacre et al., 2014). **Second**, genomically identified functions only indicate metabolic potential, while transcriptome identified functions indicate expression and are stronger evidence of true activity (“metabolic intention”), but neither genomics nor transcriptomics prove the metabolism is taken to completion.

2.3.5 – Count Matrix. Quantifying sequence data (reads, contigs, or genomes) aims to approximate the abundance of the sequence in the library. Any method of sequence quantification aims to generate tables that relates a sequence identifier to an estimate of relative abundance in a sample/ library (Figure 2E). All sequence quantification requires using reads, as they contain all the abundance data in a metagenomic or transcriptomic library. Read quantification involves simple counting, while quantifying longer sequences (contigs and genomes) requires mapping reads to the longer sequences (preprint: (Aroney et al., 2025)). Sequence counts numerically contextualize otherwise purely descriptive ‘omics. Counts can be used as a simple descriptor of community composition (e.g. **Chapter 2**), to test hypotheses of differences in function or taxa abundance (e.g. **Chapter 3, Chapter 4**), to identify associations between different taxa or metabolisms (e.g. (Mitchell et al., 2024)), or as in input in quantitative modeling (e.g. (Louca et al., 2016)).

2.3.6 – Putting it all Together: Merging Tables. These data products underlie all ‘omics analyses and are the midpoint and end goals of any bioinformatics pipeline. Once these data products are generated, they can be “put together” at both theoretical and pragmatic levels. Theoretically, they are ready for slotting into a question-driven environmental microbiology workflow. Pragmatically, they can be merged into a “master table”. If all samples, metadata, reads, contigs, and genomes have consistent naming, the generated tables can be merged for

subsequent filtering, plotting, phylogenetic inference, statistical tests, or other direct comparisons (Figure 2F; though merging all tables without a specific goal is not useful and creates unwieldy tables with millions of columns of rows).

2.4 – Conclusions

Identifying concrete ‘omics products should make the use of ‘omics less abstract while also highlighting its shortcomings. This should also make it clear why ‘omics is powerful but belongs in a workflow of diverse methodologies (Section 1). Thus far we have covered *why* someone might use ‘omics (framework for microbial ecology), *what* is generated from ‘omics (data products), and finally we will address *how* ‘omics is executed (specific tools).

Section 3: Tools and Approaches

The toolkit of sequence-based ‘omics involves the extraction and sequencing of nucleic acids and subsequent processing of generated sequences to produce the data products outlined in section 2. We will now dig into the methodologies available to practically do this (each major computational processing step is summarized in Figure 3).

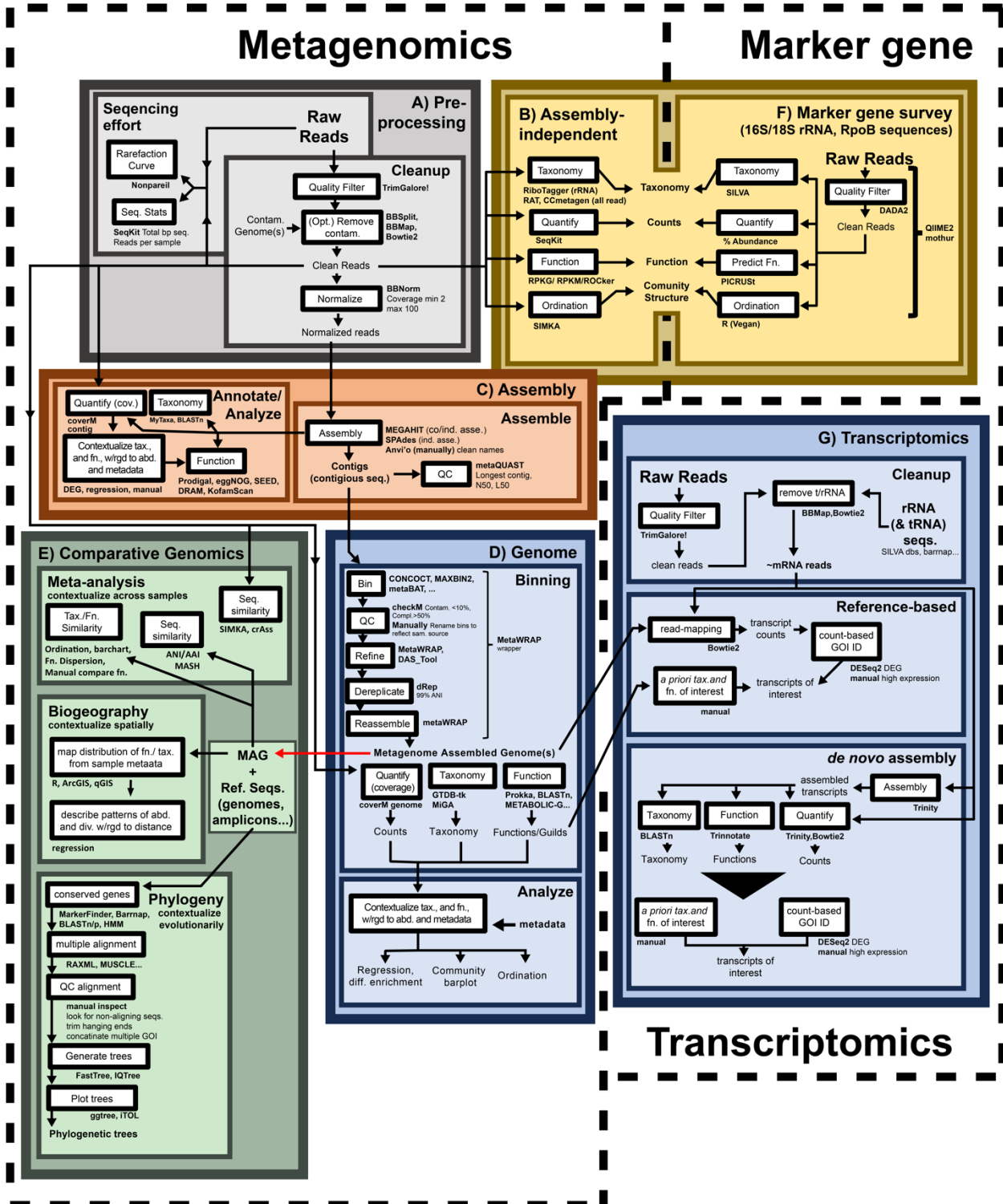


Figure 3: Generalized bioinformatics workflow for performing metagenomics, transcriptomics, and marker gene analyses. Major ‘omics disciplines are demarked by dashed lines. Within an ‘omics toolkit, major processing steps are grouped and labeled within colored boxes, with inset, colored, boxes indicating subsidiary steps. Discrete processing steps are represented as white boxes (with example tools for performing that task displayed alongside), colored boxed and arrows are used for clarity, but does not indicate importance. Specific inputs and outputs are connected to processing steps with arrows. abd., abundance; asse., assembly; compl., completion; contam., contamination; cov., coverage; db, database; DEG, differentially expressed genes diff., differential; div., diversity; fn., function; GOI, gene of interest; ind., individual; MAG, metagenome assembled genome; opt., optional; QC, quality control; ref., reference; sam., sample; seq(s), sequence(s); tax., taxon/ taxonomy; w/wgd., with regard to.

3.1 – Nucleic Acid Extraction and Sequencing

The basis of all sequence-based ‘omics is sequencing. Sequence quality, quantity, and length sets the tone for all subsequent analyses (increased sequence quality and length is generally useful). These quality parameters are largely determined by nucleic acid (xNA) extraction and sequencing technology, which we will discuss below.

3.1.1 – Sequencing Technologies. Nucleic acid sequencing has progressed through three generations of development; each of which are still in-use and have pros and cons (reviewed in (Satam et al., 2023; Slatko et al., 2018)). We will discuss each of these generations below.

First generation sequencing is often referred to as Sanger sequencing or “sequencing by termination”. This approach was used to sequence the human genome (International Human Genome Sequencing Consortium et al., 2001; Venter et al., 2001) and is currently an inexpensive approach to sequence amplicons and genomes from isolates. The major drawback of Sanger sequencing is that it cannot handle complex mixed communities as well and is not high-throughput (Slatko et al., 2018), a drawback not shared by second and third generation

sequencing. For this reason, Sanger sequencing has been largely supplanted by newer technologies for microbiome applications (Slatko et al., 2018)..

Second generation sequencing is also known as “sequencing by synthesis” or “short-read sequencing” (see “long-read sequencing” below). Today, this data product is synonymous with the most prominent producer of short-read sequencing platforms: Illumina, but short read data was also produced by 454 pyrosequencing and Ion Torrent platforms (Slatko et al., 2018). Short read sequencing generally entails spatially separating DNA fragments, and then observing the incorporation of individual nucleic acids (via fluorescence or pH change) in the physical proximity of the spatially isolated fragment, producing short (25-300 bp) reads (Slatko et al., 2018). These reads can be used as-is (e.g. RNAseq and 16S rRNA gene sequencing; **Chapter 2, Chapter 3, Chapter 4**) or assembled into larger contiguous sequences (Grabherr et al., 2011; D. Li et al., 2015; Prjibelski et al., 2020). These longer sequences can be used to construct genomes from mixed microbial communities. The major drawback of short read sequencing is that it is capped in maximum read length of ~500 bp (Slatko et al., 2018). This prevents the reconstruction of complex genomic regions that are longer than the produced reads (Mise & Iwasaki, 2022). Third generation sequencing solves this length limitation.

Third generation sequencing is also known as long read sequencing or by the trade names of the most prominent producers of long read sequencers: oxford nanopore (ONT) or Pacific Biosciences (PacBio). As the name suggests, long-read sequencing generates longer reads than second generation sequencing (10 Kb+; (Satam et al., 2023)) which allows each read to capture greater genomic context (e.g. full length 16S rRNA genes, near-complete genomes). This sequencing technology generally operates by passing (near-) full length molecules through a

fixed sequencing unit (e.g. conductive pore or modified DNA polymerase), which records the bases of the nucleic acid as it passes through. In addition to simple DNA sequences, third generation sequencing can also recover epigenetic motifs (Slatko et al., 2018) and amino acid sequences (Ouldali et al., 2020). A downside of long read sequencing is higher error rates than short read sequencing (Slatko et al., 2018), though it is possible to combine the higher quality reads from short-read sequencing and the greater genomic coverage of long-read sequencing *in silico* (see “hybrid assembly” below). It is important to note that the name “long read sequencing) speaks only to a technological capacity, not a guarantee. The length of sequences generated by long-read sequencing depends in large part on length of the DNA provided to the sequencer (fragmented DNA input for short read sequencing is less problematic as these platforms produce short reads anyways; (Zhang et al., 2022)). Different nucleic acid (xNA) extraction techniques have different tendencies to fragment xNAs and should be chosen to complement intended sequencing.

3.1.2 – Nucleic Acid Extraction. Nucleic acid (xNA) extraction aims to retrieve xNAs from host cells by cell lysis and non-xNA debris removal. There are two major kinds of extraction technique: mechanical (e.g. bead beating, freeze-thaw; (Hamilton et al., 2011)) or chemical (e.g. solvent: Phenol chloroform, Trizol, or enzymatic: lysozyme; (Trigodet et al., 2022)). Mechanical extraction techniques efficiently break open cells and retrieve xNAs (Ma et al., 2020), but fragment xNAs (Zhang et al., 2022) while chemical lysis can yield higher molecular weight xNAs (Zhang et al., 2022). Though these techniques are often used in tandem for maximal extraction efficiency (e.g. combining freeze-thaw, bead beating, and chemical lysis;

(Hamilton et al., 2011)), chemical lysis can be used alone to minimize fragmentation, as in long-read sequencing preparation (Trigodet et al., 2022).

3.1.3 – Putting Nucleic Acid Extraction and Sequencing Together. We'll quickly cover appropriate combinations of extraction and sequencing depending on data products and analytical goals. Optimizing long read sequencing (by minimizing fragmentation) is most useful when computational goals require high-quality assemblies (contigs, transcripts, and MAGs). Short read sequencing (not considering fragmentation) is appropriate when the outputs are either being used screen a microbial community (read-trawling or first-pass MAGs), or if reads are being aligned to a reference genome (alignment-based transcriptomics; (Chhangawala et al., 2015)). Though it is useful to optimize extraction technique and sequencing platform, extraction technique may also be dictated by historical precedent. To compare different datasets over time, a single extraction technique should be used to avoid biasing community recovery (Galla et al., 2024).

3.2 – Sequence Statistics

Sequence statistics are used for narrative or quantitative purposes. For **narrative purposes** it is helpful to include summary statistics for reads, contigs, and genomes (**Figure 3A**). Read statistics are used to show you did (or did not) collect enough data to tell a representative story. This can look like a simple count of total bp sequenced per sample (Liu et al., 2015) or could be used to construct rarefaction curves to show a saturating of genomic complexity in the dataset, indicating sufficient sample collection (e.g (Rodriguez-R et al., 2018)). For contig generation it is useful to provide summaries of individual and total assembly length to describe the success of assembly (e.g. N50, L50, longest contig, contig number; (Mikheenko et al., 2016)). Genomes

have similar statistics to describe binning success (e.g. completion and contamination, genome size, number of scaffolds; (Bowers et al., 2017)). These statistics should be generated any time that a sequence file is acquired or produced. For **quantitative purposes**, sequence statistics are often used to normalize count data. This can be used at the scales of comparing samples to one another, or when comparing sequences within a sample. When comparing samples, if two samples were sequenced to different depths (differential sequence sampling effort), a direct comparison between them for metrics like taxon presence-absence, evenness, or species richness may not be appropriate without normalization, as the sample that was sequenced less might have the same community features as the more thoroughly sequenced sample, but detail may have been missed in the more cursory exploration. In a simple application, sequence statistics are used to count total sequences (sequencing effort) and can be used to subset all samples' sequences to ensure even sampling effort, making cross-comparison more valid ((Gloor et al., 2017), Chapter 2, 3). This is needed when comparing the abundances of sequences of different lengths, as longer sequences are more likely to be sequenced, and will systematically appear more abundant than shorter sequences. For this purpose, it is necessary to measure the length of quantified sequences (contigs, transcripts, and genomes) and then generally counts by sequence length (e.g. RPKG, RPM; (Bertagnolli et al., 2023), Chapter 3), and TPM; (Porrás et al., 2024), Chapter 4)).

3.3 – Quality Control

There are multiple steps of quality control in sequencing 'omics, at each step in processing reads, contigs, and MAGs (**Figure 3A**).

3.3.1 – Reads. Read quality control is performed to assess: 1) are the sequences high quality? and 2) did we sequence enough? For **sequence quality**, there are two important

parameters: 1) sequence length and 2) base confidence. If a sequencer is expected to generate reads of a particular length (short read: 150-300 bp; long read: 10 kb+) a sequence substantially shorter than expectations may indicate a poorly sequenced molecule and it may be removed (Martin, 2011). For base confidence, the Q-score (i.e. Phred) can be used to remove sequences with bases (or a high percentage of bases) of low confidence (e.g. $Q < 25$, (Bertagnolli et al., 2023)). This produces a cleaned read set that can be confidently used in subsequent analyses.

For **sequencing effort**, this is often assessed with rarefaction curves. Rarefaction curves randomly subsample cleaned reads at set increments and at each step calculate a metric of novelty (e.g. number of unique sequences, number of unique k-mers, etc...), and plot these subsampled increments. This produces a smooth curve, where increasing the size of the subsample should initially increase sampled novelty until all diversity in the dataset is captured and novelty asymptotes. Saturated novelty indicates sufficient sampling effort (i.e. sequencing) to represent the community. If a sample's rarefaction curve does not asymptote, that indicates that there is uncaptured novelty. Insufficiently sequenced samples are still useful but are less representative of the community. If there are notable contaminating sequences, rarefaction curves should be generated after contaminants are removed.

3.3.2 – Contigs. Quality control for contigs assesses assembly quality by two metrics 1) assembly length and 2) representativeness of the reads.

For **assembly length**, it's standard practice to assess longest contig, number of contigs, L50, and N50. Goal values for the longest contig and number of contigs is context dependent, but large values for both the longest contig and number of contigs is generally a good sign for a successful assembly. N50 summarizes both the length of the contigs produced, and how many

contigs makes up half of the dataset. Large N50 values indicate that assembly is comprised of long contigs, which is generally good (International Human Genome Sequencing Consortium et al., 2001). L50 summarizes both the number of contigs produced, and how many contigs makes up half the dataset, a small value paired with long contigs (assessed above) indicates that the assembly is composed of long (likely data-rich) sequences (Bradnam, 2015). Length as a metric of quality is useful but requires caution. Long contigs are obviously useful, but short contigs are not automatically “bad”. Short contigs are potentially fragments of plasmids, viruses, mobile genetic elements or low-abundance microbes, and still contain useful information (Kieft et al., 2022; Maguire et al., 2020). Some contig analyses automatically remove contigs <1000-2000 bp with the justification that the average bacterial gene is ~1000 bp and sequences shorter than probably do not contain complete genes (L. Xu et al., 2006). This may be useful for simplifying a dataset but may discard data, proceed with caution. To assess well the assembly **represents the community**, it is possible to map the reads to the contigs they produced. The percent of reads confidently aligning to the contigs indicates how much of the community information was able to be incorporated into the contigs. If this value is very high (90%+) then the contigs are highly representative of the community while a low percent reads mapped value indicates the contigs are not representative of the community. In this case of low-representativeness, the contigs may still contain useful information, but it would be appropriate to compliment any assembly-based methods (contig or MAG analyses) with read-based analyses.

3.3.3 – MAGs. Genome quality is predominantly assessed by the metrics of completeness and contamination, for which the Joint Genome Institute has published quality standards (Bowers et al., 2017). Completeness and contamination metrics are calculated in a single

workflow. A focal genome is screened for the presence of a suite of single-copy genes (Parks et al., 2015). Because these genes are essential to (most) bacteria and (in cultured representatives) there is only copy of each of these genes in a genome, this allows the assumption that a complete genome should contain one copy of all the assessed genes. Contamination estimates assume that microbes should not have more than one copy of any of these assessed genes, and that the presence of duplicates indicates errors in binning or assembly. Completion is estimated by looking at the percent of these single-copy genes that are present in the focal genome (40/40 genes = 100% complete). It should be noted that “completion” does not mean the genome is present in its entirety (i.e. a closed circular genome), but rather that it contains a constrained number of essential genes. A closed circular genome is required to make any claims about true completeness (Bowers et al., 2017), and even then, there may be plasmids or phage that are important components of the microbe’s true genomic environment that are not captured in the metrics of completion or closed genomes, and requires SAG sequencing to resolve. In many analyses, it is common to only work with medium to high quality MAGs (e.g. (Colman et al., 2024; Shoemaker et al., 2024), Chapter 3), which improves confidence in genomic results, but may make the community-scale results less representative.

3.4 – Community Description

3.4.1 – Ribosomal RNA Surveying. It is possible to identify taxonomically useful ribosomal rRNA (rRNA) sequences from read or contig data to either survey a dataset for taxonomic composition, or (if count data is generated) to perform community-scale analyses analogous to targeted 16/18S amplicon sequencing. rRNA sequences (e.g. 16S or 18S rRNA

genes) can be identified by HMM search, with taxonomy assigned by alignment against a database or phylogenetic placement (e.g. (Gruber-Vodicka et al., 2020; Xie et al., 2016)).

3.4.2 – Community Summaries. Communities can be summarized by univariate or multivariate approaches. Univariate approaches can use taxonomic or oligonucleotide diversity to simply summarize a community (e.g. (Bertagnolli & Stewart, 2018)). Multivariate approaches capture a greater degree of complexity and collapse the most explanatory axes into a couple visualizable dimension. This can be done in a reference-agnostic fashion that does not rely on databases, comparing samples by oligonucleotide frequencies (k-mers). K-mer frequencies are an agnostic method for assessing similarity as it does not rely on databases and uses all read data. K-mer frequencies can be used to generate distance matrices, and plot ordinations that may show similarities or differences between samples at the highest level (e.g. (Benoit et al., 2016)). Genomes and functional coding potential can also be used in this ordination approach (e.g. (Fernandes-Martins et al., 2023)). If there are notable contaminating sequences, ordination should be generated after contaminants are removed.

3.5 – Contaminant Removal

Non-target sequences (contaminants) may be removed from the library by read mapping. Contaminating sequences fall into two categories: 1) defined (e.g. host-sequences, rRNA/ tRNA transcripts) and 2) incidental (e.g. bacterial contamination). Defined contaminants are known *a priori* and may be non-target sequences from a host (e.g. human gut, coral, Chapter 4) or are not information coding (e.g. t/rRNA for transcriptomics). Incidental contaminants can be identified from rRNA surveys (see 3.4.1 – Ribosomal RNA Surveying, Chapter 4). The removal of defined or incidental contamination are both performed the same way: a database of sequences with a

high degree of specificity to that of the contaminant is created (e.g. from NCBI genomes or SILVA t/rRNA database). Reads are mapped against the contaminant database. Reads that confidently align to reference contaminant sequences are likely contaminant derived and are then removed (e.g. (Porras et al., 2024; Sorek et al., 2018), Chapter 4)

3.6 – Assembly

Assembly generates longer contiguous sequences (contigs) from shorter reads. This process entails two steps: 1) library normalization and 2) assembly.

3.6.1 – Library Normalization. Normalization down samples over-represented sequences and removes low-abundance sequences in a library, simplifying the computational complexity of assembly. This is useful on both high-complexity (Chapter 2) and simplified microbial communities (Colman et al., 2024). Normalization can be performed using k-mers, where reads with high abundance k-mers (e.g. $> 100x$) are subsampled to not force an assembler to work with redundant data, while reads with low abundance k-mers (e.g. $< 2x$) are removed to not saturate the assembler with low abundance sequences that may not assemble well (e.g. (Marçais & Kingsford, 2011), bbnorm, “<https://sourceforge.net/projects/bbmap/>”).

3.6.2 – Contig Assembly. Assemblers use reads to reconstruct longer sequences (DNA or RNA). Describing the algorithms that perform assembly is beyond the scope of this review (benchmarked in (Z. Li et al., 2012)), but in a general sense assemblers look for overlaps between reads and use this overlap to create longer and longer sequences. There are two major classes of assembly: 1) guided and 2) *de novo*.

Guided assembly uses long sequences from a source similar to that of the reads to serve as a scaffold, guiding the placement of reads. The long sequences can come from a reference

genome (i.e. reference guided assembly; (Lischer & Shimizu, 2017)) or by pairing long and short-read sequencing, where short reads are aligned to long read to generate contigs, increasing assembly quality and length than either alone (i.e. hybrid assembly; (Antipov et al., 2016)).

De novo assembly has two varieties: 1) individual and 2) co-assembly. Individual assembly is the simplest conceptually, where a sequence library from a single sample is assembled into contigs, and all assembled contigs are by-definition derived from that sample. Co-assembly aims to generate contigs from low abundance organisms in the library. In this method, sequences from similar samples (e.g. all sediments from a single reef and timepoint; Chapter 3) are pooled and then assembled. The generated contigs are representative of the sequences across all samples, though not all contig may be “present” in every library. Disambiguating presence/ absence across samples requires quantification of contigs (or derived MAGs) in each sample. (If performing a co-assembly, normalize reads after concatenation, as this may retain low abundance k-mers that were removed in un-pooled libraries). It is possible to assemble using multiple approaches, construct MAGs with them, and later use quality controls to prevent duplicated sequences.

3.7 – MAG Binning

Binning (making MAGs) consists of splitting contigs into separate, taxonomically coherent, “bins”. These bins are draft genomes, and once quality is confirmed (e.g. quality controlled, dereplicated, refined, and/or reassembled), they are considered metagenome assembled genomes (MAGs). MAGs are a consensus population-level genome that cannot resolve strain differences (Meziti et al., 2021). Single amplified genomes (SAGs) can resolve strain differences between genomes but are far more expensive to generate (Zheng et al., 2022).

3.7.1 – Binning Contigs. Binnings divides contigs into bins using genomic features (GC content, k-mer clustering) and abundance correlation (read coverage), and each binner may use some or all of these parameters (Bowers et al., 2017). Binnings assume GC content and distribution of k-mer (often 4-mers) will be consistent across a single genome (Dick et al., 2009). This allows univariate (GC) and multivariate (k-mer composition) separation of contigs into initial clusters. These clusters can be refined with the additional assumption that in a single library, the abundances (as derived from read mapping) of contigs from a single genome will be correlated (Sharon et al., 2013). For instance, assume contig A and B are from Genome 1 and C and D are from Genome 2. If Contigs A and B both have a coverage of 30X and C and D have coverages of 2X, a binner might group A and B into one bin and C and D into another. Combining these approaches for contig partitioning, creates draft genomes which should be filtered by quality (completion and contamination) and can be further refined before analysis.

3.7.2 – MAG Improvement. The quality of generated MAGs can be improved by selecting the best MAGs from multiple bidders (refinement), removing redundant MAGs (dereplication) and reassembling high-quality MAGs (reassembly).

Refinement uses multiple bidders to generate multiple (somewhat redundant) bin sets. The resulting MAGs from each of these sets are compared to each other to find highest quality MAGs (based on completion and contamination). Each chosen MAG is placed in a final bin set and a final cleaning step may search for contigs shared amongst multiple bins, and remove them from all but the highest quality bin (e.g. (Uritskiy et al., 2018)), resulting in the best possible MAGs from multiple bidders.

Dereplication aims to represent multiple redundant MAGs by a single, highest quality, genome (Olm et al., 2017). This is often performed through pairwise sequence similarity comparisons, with two MAGs sharing 99% average nucleotide identity (ANI) considered practically identical (Jain et al., 2018). For a cluster of MAGs all sharing 99% ANI, the genome with highest completion, and lowest contamination is selected to serve as the representative sequence (Olm et al., 2017). This produces a set of highest quality representative MAGs from the dataset, reducing the downstream analytical burden of analyzing multiple extremely similar genomes.

MAG **reassembly** uses high quality (QC'd, refined, and/or dereplicated) MAGs to try to generate even longer contigs for each MAG. This involves identifying reads that align to each MAG, and assembling those reads using a non-metagenome assembler algorithm (potentially using the MAGs contigs for a guided assembly), generating more contigs than the metagenome assembler (e.g. SPAdes instead of metaSPAdes; (Uritskiy et al., 2018)). This non-metagenome approach is possible because read mapping bioinformatically isolated the reads of the genome. This process of re-assembly and read mapping can be repeated multiple times, until a “winning” MAG is achieved that displays the greatest completion and smallest contamination values (e.g. (Kitzinger et al., 2020)).

We have previously discussed the positive use cases of MAGs, but they do have some shortcomings. First, not all sequences bin equally well. Binning relies on high quality assemblies, and splitting these contigs based on sequence similarity (GC content and k-mer abundances), assuming that disparate parts of a single genome will have similar sequence characteristics (W. C. Nelson et al., 2020). When these assumptions are not true, binning breaks

down. These assumptions are definitely not true for the results of horizontal gene transfer (HGT). When an organism received DNA from another organism the donated DNA likely has the CG and K-mer signature of the donor, not the recipient, leading to potential mis-assembly and/or binning (Mise & Iwasaki, 2022), leading to **missing genomic content**. This problem is exacerbated for plasmids, where the DNA within the host is not contiguous to the genome. Assembly may reconstruct a plasmid, but because there is neither overlap nor sequence similarity to the chromosome, it is more likely to be unbinned or misbinned (Maguire et al., 2020). Additionally genomic regions with unclear sequence qualities (highly-repeated and complex; including rRNA genes; (Mise & Iwasaki, 2022)) are challenging to assemble or bin appropriately (Wilbanks et al., 2022). Next, Contigs and then MAGs are not representative of strains but rather easy-to-assemble and simple- to-bin sequences (Wilbanks et al., 2022). This results in intra-species “blurring”, obscuring genomic novelty, leading to **community homogenization**. Finally, not all reads and contigs are able to be assembled, and not all contigs can be confidently binned, leading to a progressively less accurate community representation from **un-assembled and unbinned sequences** (Hauptfeld et al., 2024). For some questions, genome-resolved studies may miss the presence of taxa and functions that do not bin. **Some of these shortcomings can be addressed**. Missing genomic features can be improved by using long-read sequencing technologies to sequence across ambiguous sequences, providing genomic anchors (Wilbanks et al., 2022). Species homogenization can be improved by using single amplified genomes (SAGs) to achieve single-cell genomic specificity. While the loss of unbinned reads can be balanced by using read and contig-based analyses to compliment MAGs, especially focusing on reads that did not assemble or bin (Chapter 3, e.g. preprint: (Ricci et al., 2025)).

3.8 – Taxonomic Classification

Though taxonomy is an imperfect method of classifying relationships between living things (Ward et al., 2008), it provides a useful framework for communication and analysis. Taxonomically classifying microbes generally involves comparing sequence similarity between two taxa (Jain et al., 2018). There are two major ways to determine the taxonomic basis of a sequence 1) direct sequence alignment and 2) guilt by association.

For **sequence alignment**, there are untargeted and targeted approaches. For an untargeted approach, sequences can be aligned to a database containing sequences with confident taxonomies (e.g. NCBI nt, SILVA, in-house datasets...). This is very common for genomes and contigs (especially for taxonomic marker genes: e.g. 16S/18S rRNA,...; (Dragone et al., 2022); Chapter 3,4), but can also be performed on reads (e.g. (Marcelino et al., 2020)), though ambiguous read mapping may require secondary processing for confident classification (see “guilt-by-association” (Hauptfeld et al., 2024)).

Taxonomic **guilt-by-association** is both extremely common and rarely articulated. The most common method of guilt-by-association is when we assume that all the contigs (and reads) in a MAG have the same taxonomic origin. This subsidiary property is leveraged by some tools to supplement alignment-based read taxonomy (e.g. creating MAGs and using that taxonomy to re-classify the constituent reads, iteratively improving read-based taxonomy; (Hauptfeld et al., 2024)). This MAG-contig-read taxonomic association is conceptually straightforward but is underpinned by graph-based binning methodologies.

Other graph-based sequence clustering approaches can also be used to taxonomically classify novel sequences against references, with any members of the cluster assigned the taxonomy of the reference sequences (i.e. guilty-by-association). This is very useful for

classification of sequences without clear taxonomic marker genes (e.g. viruses (Bin Jang et al., 2019)). Phylogenetic taxonomic inference also follows a guilt-by-association principle.

Phylogenetic placement may be used for principal taxonomic classification (e.g. (Chaumeil et al., 2020)) or as support for precise taxonomic placement (e.g. Chapter 4) or reassignment (e.g. (LaJeunesse et al., 2018)). This approach is very common for sequences from genomes but can also be applied to unbinned contigs (e.g. (Minch et al., 2023)) and reads (e.g. (Glöckner et al., 2017)). The results of such approaches can be used to describe the relationships between sequences graphically or categorically using hierarchical taxonomic lineages.

3.9 – Function Annotation

Functional annotation attempts to predict the potential uses and activities of these sequences in the genomes and environments they came from. Functional annotation of sequence data (reads, contigs, or genomes) compares a query sequence against annotated (and potentially experimentally validated) reference sequences. Here we will cover the executing the analytical goals of annotation and the tool classes available for annotation. There currently are two major approaches to sequence annotation 1) targeted and 2) untargeted searches.

Targeted searches are used to interrogate the presence of 1-100 genes of *a priori* interest (e.g. biogeochemical (Leung & Greening, 2020); secretion systems (Abby et al., 2024); prokaryote immune systems (Tesson et al., 2022)). For this purpose, reads may be used as-is (e.g. (Bertagnolli et al., 2023; Dragone et al., 2022)), while contigs (including MAGs and transcripts) are generally broken up into smaller putative protein coding sequences (i.e. open reading frames; ORFs). In a general sense, ORFs are identified by searching for start codons, stop codons with the sequence between each start and stop codons output into a FASTA file as a likely protein

coding region (e.g. (Hyatt et al., 2010)). These ORFs or reads are then queried (see tools for annotation below) for the presence of the genes of interest. **Untargeted** discovery often uses two steps: first-pass annotation against a broad suite of generally useful genes, and secondary annotation of initially un-annotated sequences. First-pass annotation attempts to assign functions to sequences using a constrained of database pre-determined broad-sweeping set of genes, using one or more annotation tools and/or databases (1,000+; e.g. (Aramaki et al., 2020; Huerta-Cepas et al., 2017; Seemann, 2014; Shaffer et al., 2020; Zhou et al., 2022)). Sequences that are still un-annotated by this broad approach may be annotated using a general gene database (e.g. NCBI nt/nr, UniProtKB...; per (Mitchell et al., 2024)). Still un-annotated genes may be assigned speculative annotations based on the functions of nearby sequences, using the guilt by association principle, assuming that genes in the same genomic neighborhood (close together) perform similar roles (Mihelčić et al., 2019). In either targeted or untargeted searches, preliminary annotations of genes that are known to be hard-to-annotate or that are pivotal to a manuscript may need validation against experimentally validated genes (Garber et al., 2020; Søndergaard et al., 2016).

There are two major tool classes for annotating sequences 1) alignment, 2) Hidden Markov models (HMMs), and 3) attention-based artificial intelligence. **Alignment** is theoretically simple, where a query sequence is compared to a reference sequence possessing confident functional annotation. If the two sequences are sufficiently similar, the query sequence can be assigned the functional annotation of the reference. The most common sequence aligners are the NCBI's BLAST suite (Camacho et al., 2009) which is able to precisely identify homologues (Al-Fatlawi et al., 2023), and have simple web integration with NCBI sequence

databases. **HMMs** are highly sensitive tools for sequence identification, and their exact functioning is beyond the scope of this paper (reviewed in (Mor et al., 2021)). HMMs for functional annotation are trained on high-confidence reference sequences, these models can identify distant sequence homologues more effectively than BLAST (Kirsip & Abroi, 2019). There are repositories of pre-trained HMMs for the identification of general (e.g. KOfam, TIGRFAM, Pfam...) or specific genes of interest.

Reads are well suited to alignment-based function annotation, while contigs (and genomes) are well suited to both alignment and HMMs annotation (e.g. (Anantharaman et al., 2016)). Notable, but still in its early days is the use **artificial intelligence** leveraging attention-based methods for genomic functional prediction (Jumper et al., 2021) and annotation (Hwang et al., 2024).

3.10 – Count Data

Quantifying sequence data (reads, contigs, or genomes) uses read abundances, as they contain all the abundance data in a sequence library. Read quantification involves a simple sequence counting (potentially with sequencing effort normalization, e.g. RPM), while quantifying contigs and genomes requires mapping reads to these longer sequences (with subsequent normalization for sequence length and dataset size, e.g. FPKM). Count data is used widely in genomic studies to estimate the abundance of genes ((Bertagnolli et al., 2023), Chapter 3) and microbes (e.g. (Shoemaker et al., 2024; Steinsdóttir et al., 2022)) in a sample. In transcriptomics, RNAseq-derived reads are aligned to a reference sequence (genome or assembled transcript) to estimate the expression of the gene.

There are some specific considerations when generating and interpreting count estimates addressing count overestimation, normalization methods, and the compositional nature of ‘omics count data. Alignment based quantification may **overestimate** sequence abundances, and there exist tools to account for this overestimation (e.g. TAD80; per (Viver et al., 2021)). As discussed previously, count data needs to be **normalized** against gene length (Kb) and library size (Mb) resulting in metrics like reads per kilobase per megabase sequenced (RPKM). Another method for library size normalization uses the abundance of single copy marker gene(s) in the dataset to normalize against microbial abundance in the dataset instead of bases sequenced. This results in a metric that still accounts for sampling effort, but also takes into account biological context (e.g. RPKG, (Bertagnolli et al., 2023), Chapter 3). It is also important to note that all sequence derived count data is **compositional** and only describes relative abundances (Gloor et al., 2017) unless paired with intentional quantification.

3.11 – Phylogeny

A phylogeny contextualizes a focal sequence(s) against other reference sequences to provide a sense of the evolutionary distance between the sequences. Though phylogeny was described earlier in the context of taxonomic classification and functional annotation, it has other uses which will be described here. Phylogenies are characterized by two key features: topology (shape of the tree) and branch length (distance to a tip).

Topology describes the general shape of a phylogenetic tree, especially branching patterns and clusters of sequences (Kapli et al., 2020). Topology is often used to support claims that a sequence belongs to a taxonomic or functional group, by illustrating that a sequence of interest (genome or gene) falls squarely within a clade of annotated reference sequences.

Monophyletic clustering increases confidence in sequence annotation and/or classification (e.g. (Porrás et al., 2024)). Alternatively, topology can show that a sequence is novel, if it clusters far from all other sequences. Observing divergent sequences may delineate new taxonomic groups at coarse (e.g. (Lane et al., 1985; Woese & Fox, 1977)) or fine phylogenetic scales (e.g. (Tsementzi et al., 2016)). **Branch length** is the distance from root to tip of the tree, and proxies the age of phylogenetic lineages. Time calibrated branch lengths (using dated fossils or geochemical evidence to contextualize node timing) can be used to provide insights into the timing of diversification (LaJeunesse et al., 2018). Beyond evolutionary description, branch lengths can be used to quantitatively test the phylogenetic basis of ecological association (Colman et al., 2024).

The creation of a phylogeny comprises of three main steps: 1) sequence curation, 2) multiple sequence alignment, 3) phylogenetic inference (all steps reviewed in (Kapli et al., 2020)). The first step of **sequence curation** involves identifying genes for phylogenetic reconstruction. A rough principle in phylogenetic reconstruction: the longer the available sequence, the better the tree represents evolutionary history. In this aim, as more data has become available, phylogenies have moved from comparing single marker-genes, to multiple concatenated genes, to whole genomes (Delsuc et al., 2005). To perform these phylogenetic comparisons, it is essential to first identify comparable sequences. For this, homologs must be identified and outliers removed. Initial homolog identification is relatively straightforward and can be performed in a manual (e.g. BLAST genomic sequences against a GOI reference database) or automated fashion (e.g. (Martinez-Gutierrez & Aylward, 2021)). In selecting nucleic acid or amino acid sequences for alignment for treeing, it is common practice to use nucleic acids to resolve closely related organisms (due to nucleic acid's higher variability than amino

acid sequences), and while amino acid sequences for deep divergences, however nucleic acids may provide similar resolution for these deep phylogenies (Kapli et al., 2023). Next, it is necessary to identify false positives to ensure that the sequences used are true homologs (Kapli et al., 2020). The result of this process is a FASTA file of sequences for gene(s) for alignment.

Multiple sequence alignment (MSA) uses the validated sequences to create a matrix of aligned sequences, where the residues in each column should correspond to an ancestral similarity. If multiple genes are used, they may be aligned together or separately (and then combined) to create a supermatrix. The quality of the MSA is the basis of all tree quality and should be inspected to ensure spurious sequences are removed. This may entail removing sequences that poorly align (potentially paralogous or non-target genes), sequences with high percentages of gaps, and/ or masking ambiguously aligned regions (Kapli et al., 2020). Finally, **phylogenetic inference** involves generating a tree that proxies true evolutionary relationships. This can be performed by creating multiple trees from each of the aligned genes, or a single tree from the supermatrix. The methods for constructing trees can search for a best tree amongst multiple attempts, and/ or use explicit models for sequence evolution, both of which improve relational reconstruction (Kapli et al., 2020). Though phylogenetic tree construction has been made highly accessible by straightforward tools, many applications require nuanced parameter selection that is beyond the scope of this review.

3.12 – Thoughts for Practitioners

Now that we have described the execution of nitty-gritty ‘omics and a framework to guide its application I have a few final suggestions and thoughts for a would-be practitioner (specifically aimed at an early-stage graduate student).

For **bioinformatics** advice. 1) First, get comfortable working in R, Linux, and/ or Python to look at and manipulate data tables and other relevant file formats (e.g. subsetting, filtering, merging; this is the computational equivalent of pipetting). 2) Make data organization a primary directive; get good at writing computational pipelines, organizing file directories, and naming files, and using electronic lab notebooks (the computational equivalent of keeping an organized lab notebook). 3) Before you have data, get familiar with the general workflows for creating and analyzing sequence data (genomes, transcriptomes, amplicons, and phylogenies, especially the uses and limitations of sequence alignment). You will be better able to use data you collect or mine. 4) Get a good sense of the major sequence databases and publicly accessible datasets that you could mine to contextualize your data, or to acquire data for re-analysis (e.g. Tara Oceans, NCBI nt/nr, EMBL, mgRAST, NEON, LTER...). A bioinformatician with databases is never without samples. 5) Finally, try to become independent at the entire sample to sequence pipeline [e.g. sample collection and preservation, nucleic acid extraction, sequencing prep (though I would suggest out-sourcing sequencing), and most ‘omics analyses (if you have done the work to get good at computational basics, learning new ‘omics pipelines will be relatively straightforward)

Now, **non-computational** advice. 1) First, get good at communicating (Hazelett, 2025). This is broadly applicable, but essential for describing ‘omics, because the purpose and utility of sequence-based analyses is not immediately obvious to most people. 2) Get comfortable with non-computational microbial ecology and microbiology literature (specifically your focal discipline, e.g. redox, stoichiometry, physiology, oceanography, pathogenicity, molecular biology...). If you talk to an expert in your field of biology, the goal is to be seen as “one of us”.

3) Get comfortable making informed guesses, both qualitative from your literature base, and quantitatively. (per (Fox, 2011; Milo & Phillips, 2015)). 4) Get acquainted with, and comfortable talking, about non-omics tools, especially how they fill the gaps left by ‘omics (e.g. rate measurements, FISH, nanoSIMS, enrichment cultures, isolation, knockouts, transformation, protein purification). This will help you plan research and identify tractable next steps (even if you never intend to do it yourself, it’ll help you find collaborators). 5) Related, try to add other relevant tools to your repertoire. Bioinformatics is (e.g. qPCR, microscopy, stable isotope incubations, microcosms, isolation, bioassays...). If you cannot, find collaborators (see below). 6) Finally, make friends (i.e. networking) in your field and beyond (e.g. conferences, workshops, fieldwork, your home department and adjacent departments: ecology, earth sciences, geography, biochemistry...). Use these connections to identify good colleagues, great collaborators, and people to avoid – it’s good to have options. It will be way easier to effectively find and nurture collaboration, if you understand non-omics methodologies and can effectively communicate what ‘omics do. Assuming you are a competent bioinformatician, you will be able to both articulate how ‘omics and non-omics can tell a coherent story, and then follow-through as a collaborator (in microbiology departments and beyond).

‘Omics, like all tools, has distinct strengths and weaknesses. ‘Omics capacity to compare new biological sequences against ever-expanding databases allows researchers to generate hypotheses, contextualize experimental data, and extrapolate findings across the world. But, *by itself*, ‘omics often stops short of *proving* function. At this point, the other methodological avenues available to the microbial ecologist are hopefully very clear.

Thus far, we have outlined a framework for guiding microbial ecology highlighting diverse methodologies (Section 1), and then focused on the applications and workflow for using ‘omics analyses (Sections 2 and 3). I will conclude with by highlighting how the primary research in my dissertation fits into the framework described in Section 1.

Section 4: Dissertation as Case Study

I describe a framework to guide microbial ecology in Section 1. In an ideal world, I would have been able to take the projects of my dissertation further through the workflow. However, time and travel (particularly COVID restrictions) constraints meant that this research revolved around sequence-based (‘omics) analyses. We used available microbiome sampling to begin to resolve ecological structure-function relationships (observing community and formulating hypothesis) and contextualize the results (phylogenetically and meta-analysis).

This dissertation covers multiple angles of the coral reef microbiome. Chapter 2 is a marker gene survey of microbial taxa across reef microhabitats, identifying several microhabitats and taxa bearing further investigation. The next chapters will focus on microhabitats and microbes sampled in Chapter 2. Chapter 3 is based on community-scale observations of noteworthy taxa from Chapter 2. We focus on one abundant and prevalent microbes, using *in situ* marker gene surveying and bulk-tissue transcriptomics to zoom in on a high abundant and ubiquitous coral endosymbiont (*Endozoicomonas sp.*). Here we formulate gene expression-based hypotheses of bacterial function, noting the expression evidence of unexpected, potentially coral parasitizing, metabolisms. Chapter 4 is based on the ecosystem-level observation that coral mortality decreases in the presence of sea cucumbers. To searching for potential taxa and mechanisms explaining this phenomenon, we employ quantitative PCR, marker gene surveying,

and metagenomics to dig into the reef sediment microbiome. I find genomic evidence that metabolites from phototrophs and sulfur reducers bear closer investigation. Collectively, this dissertation identifies several tractable microbial taxa and metabolisms and follows-up on a couple in detail. The metabolic potential and taxa of interest identified bioinformatically provides clear future directions for subsequent research to untangle microbe interactions across the coral reef, which are described in the dissertation's conclusion.

CHAPTER TWO

MICROHABITAT AND TIME-OF-DAY PREDICTABLY
STRUCTURE DIVERSE REEF MICROBIOMES

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: Andrew J. Maritan

Contributions: formulated guiding hypotheses, designed field experiments, collected samples, performed molecular microbiology analysis, devised bioinformatic strategy, performed bioinformatics, wrote manuscript

Author: Zoe A. Pratte

Contributions: designed field experiments, collected samples, assisted in sample processing and bioinformatics

Author: Cody S. Clements

Contributions: designed field experiments, constructed experimental apparatus, collected samples

Author: Mark E. Hay

Contributions: designed field experiments

Author: Frank J. Stewart

Contributions: formulated guiding hypotheses, designed field experiments, advised on experimental and analytical strategies, wrote manuscript

Manuscript Information

Andrew J. Maritan, Zoe A. Pratte, Cody S. Clements, Mark E. Hay, Frank J. Stewart

Status of Manuscript:

- ☒ Prepared for submission to a peer-reviewed journal
- ☐ Officially submitted to a peer-reviewed journal
- ☐ Accepted by a peer-reviewed journal
- ☐ Published in a peer-reviewed journal

Abstract

Light is both an energy source and selective pressure. Light's first and second order effects on microbes varies predictably with its daily availability, energizing and selecting communities on timescales that overlap with microbial reproductive rates. Microbes that persist in-the face of these stressors and/ or exhibit cyclic abundances may be active community members, responsive to the environment. To identify these responsive and/or ubiquitous microbes (both prokaryotic and eukaryotic) across coral reef microhabitats we performed hour-scale sampling across cyclic geophysical conditions *in situ* on a reef of Mo'orea, French Polynesia, using 16S and 18S rRNA gene sequencing to characterize the taxa in coral, macroalgae, water, and sediment microhabitats. Our results reveal that time-of-day consistently shapes microbiome composition across diverse reef microhabitats. Further, the abundances of individual taxa exhibit diel cyclicity, identifying several focal taxa for further investigation.

Introduction

In macroecology it would be unthinkable to sample at yearly intervals and treat this knowledge as truly representative. To understand the stochastic and deterministic factors that create end-state ecosystems requires sampling through seasonal cycles, months and days (Resasco et al., 2021). Yet, this is how we generally sample microbiomes. Microorganisms can reproduce in minutes and yet most studies measure communities once, or at monthly to yearly intervals. Scaling the “appropriate” macroecological sampling time to microbes calls for minute to hourly sampling. Over this microbially-relevant timeframe, daily (diel) oscillations in light levels and associated changes to UV irradiation, temperature, primary productivity, and oxygen

concentrations *should* collectively act as an ecological filter (HilleRisLambers et al., 2012) reshaping microbiomes over these short timescales. The net results of diel pressures have generally been captured through integrated measures of microbial activity (e.g. gas flux, cell counts, primary productivity), providing essential insights (Pelve et al., 2017) but obscuring the biotic context that created the measured outcome.

Sampling microbiomes across diel abiotic gradients should identify communities and individual taxa that persist or cycle in-time with their environment. These persisters and responders are fodder for identifying and following-up on meaningful ecological interactions (Park et al., 2022). Indeed, such hour-scale studies have shown that microbial communities respond to changing environmental conditions at hourly timescales (Nemergut et al 2013).

Microbial mats follow a day-night cycle characterized by oxygenic photosynthesis during the day and sulfur reduction at night (Jørgensen et al., 1979), with some microbes chasing the variable oxic-anoxic interface (Biddanda et al., 2023). The **open ocean** microbiome has light-dependent (Vislova et al., 2019) diel rhythms too. Phototrophs fix carbon during the day (Simó et al., 2002), some of which is leaked, stimulating cyclic heterotroph abundances (Uchimiya et al., 2022), linking their abundance (Aylward et al. 2015) and expression of nutrient transporters and carbon metabolism (Muratore et al. 2022) to diel rhythms. Phototrophs use stored carbon for nighttime cell division (Ribalet et al., 2015; L. Weber & Apprill, 2020) which co-occurs with evening predation of cyanobacteria (Ribalet et al., 2015). Host-associated microbes also exhibit diel variation. **In photosynthetic plants** the regular release of photosynthetic byproducts in the roots (e.g. O₂, fixed carbon; (Dai et al., 2022; Staley et al., 2017; Y. Xu et al., 2012; Zhao et al., 2021) (Herron et al., 2013; Staley et al., 2017) and respiratory wastes (e.g. CO₂; (Dai et al., 2022;

Staley et al., 2017; Y. Xu et al., 2012; Zhao et al., 2021) (Herron et al., 2013; Staley et al., 2017) alters the rhizosphere microbiome (Hubbard et al., 2018), particularly those microbes and metabolisms sensitive to oxygen (Smercina et al., 2019; Xu et al., 2012) and carbon (Staley et al., 2017) availability. Across each of these diverse systems, phototrophs responding to diel availability of light is the nexus of diel change which is then propagated through the rest of the microbiome via nutrient release, cross-feeding, and changes to redox conditions.

By these criteria, coral reefs are an ideal diel-responsive ecosystem, characterized by abundant light (Hoegh-Guldberg et al., 2017) and diverse phototrophs, foundational to an extensive heterotrophic food web (DeVantier, 2020; Littler & Littler, 2013). Reefs are notable for their macro-scale diversity, which is mirrored on microbial scales (Galand et al., 2023). This microbial diversity is exciting, but identifying meaningful ecological associations between microbes and their reef microhabitats is a perennial challenge (Park et al., 2022). Microbial interaction with **corals** is the focus of most reef microbiome study, with eukaryotes (LaJeunesse, 2020) and bacteria (Glasl et al., 2016) clearly essential to coral survival, but identifying and describing other ecologically important microbes is challenging (L. Richardson et al., 2009). Across other reef microhabitats, microbes also play essential roles in ecosystem function. Bacteria are key to **macroalgal** growth and survival (Singh & Reddy, 2016) but most macroalgae-microbial associations are hard to classify (Park et al., 2022). Reef sediments and water are the physical matrix of the reef, connecting all microhabitats (Garren & Azam, 2012). **Sediment** microbiomes cycle carbon, methane, and, nitrogen, oxygen, and sulfur (Alongi et al., 2008) while the **water** microbiome is characterized by low nutrient concentrations and rapid microbial turnover of dissolved organic matter (C. E. Nelson et al., 2023).

This microbial ecology becomes even more complex when microbes are transmitted between reef microhabitats (e.g. water-coral (Studivan et al., 2022), algae-corals (Clements & Hay, 2023), sediment-coral (Clements et al., 2024)). Across the reef, microbial complexity is challenging to parse, and often relies on seasonal or disturbance-based associative studies to link microbes to reef function (e.g. season (Sharp et al., 2017), disease (Clements et al., 2024), heatwaves (Vompe et al., 2024), nutrient addition (Thurber et al., 2014)), which requires time or misfortune to identify microbial candidates for ecologically relevant functions. A more readily available repeatable “disturbance” are diel changes. Daily changes to physical conditions on a reef can match the magnitude of seasonal variation (Caughman et al., 2021), and might be used to cut through reef-scale diversity to identify ecologically relevant microbes. Indeed, at an organismal-scale, corals with more robust responses to daily fluctuations in oxygen concentration are more resilient against bleaching (Alderdice et al., 2021).

Some microbiome studies have sought to identify important taxa on reefs by sampling over time. Most of these studies have sampled the water microbiome for bacteria (missing eukaryotes and other microhabitats). The water-borne microbiomes experience diel periodicity, at both near-coral (~ 5 cm; (L. Weber & Apprill, 2020)) or at the reef scale (≥ 1 m from a coral; (C. Becker et al., 2020; Sweet et al., 2010; L. Weber & Apprill, 2020)). At the reef scale, prokaryotic community diversity was significantly related to the time of day (Sweet et al., 2010), where reef and near-coral seawater microbial richness was significantly higher and more variable during the day than at night (L. Weber & Apprill, 2020). In these studies, photosynthetic microbes played a significant role in driving observed changes. Cyanobacterial relative and absolute abundances were 1.2 and 2 times higher at night compared to the day (L. Weber & Apprill, 2020),

photosynthetic microeukaryote abundance peaked after dusk (C. Becker et al., 2020) while most other microbial guilds did not change (C. Becker et al., 2020; L. Weber & Apprill, 2020). The results of these reef studies parallel that of diel microbial dynamics of the open ocean, where cyanobacteria are central to networked change (Aylward et al., 2015), with periodic primary productivity (K. W. Becker et al., 2021), gene expression (Aylward et al., 2017; Muratore et al., 2022; Ottesen et al., 2014) and abundances (Aylward et al., 2017; K. W. Becker et al., 2021; Ribalet et al., 2015; L. Weber & Apprill, 2020). Though informative for seawater ecology, these water microbiomes are distinct from that of corals (L. Weber & Apprill, 2020), which require specific attention. One study sampled coral tissues over time and found diel patterns in bacterial communities noting that dusk/ night microbiomes were more variable than that of the daytime, observing that *Vibrio* sp., *Aeromonas* sp., and *Endozoicomonas* sp. increased in relative abundance during the day, while *Porticoccus* sp. displayed increased midnight representation (Caughman et al., 2021). These available coral reef, time-series microbiome studies have only examined single reef microhabitats, focusing only on prokaryotes. Furthermore, these studies timecourses have been too short to decouple tidal influence from diel processes.

To fill this gap, we surveyed diverse coral reef microhabitats (i.e. coral, macroalgae, water, sediments) sampling both prokaryotic and eukaryotic communities at hour-scales *in situ* on a reef of Mo'orea, French Polynesia. We contextualized these microbiomes with reef geophysical conditions using a combination of dataloggers, nutrient sampling, and public data. This analysis aimed to identify ecologically relevant taxa based on the assumption that prevalence across steep environmental changes and/ or abundances that track (or trail) geophysical changes indicates a taxon responding to environmental gradients actively (e.g.

reproducing or resisting change) or passively (e.g. preyed upon or ceasing reproduction). Our results reveal that time-of-day consistently shapes microbiome composition across diverse reef microhabitats and the abundances of individual taxa exhibit diel cyclicity, highlighting several taxa for further investigation.

Materials and Methods

To capture daily variability in the microbiomes of diverse coral reef microhabitats, we sampled corals, macroalgae, sediments, and overlying water on a reef flat in Mo'orea, French Polynesia every 6 hours over multiple days. Community taxonomic profiles for both prokaryotes and eukaryotes were assessed by 16S and 18S rRNA gene sequencing, respectively. We contextualized diel changes to the microbial community alongside measurements of oxygen, light, temperature and key nutrients. We show that each reef microhabitat hosts a distinct microbial community and observe several microhabitat microbiomes structured by time of day, with individual microbes exhibiting diel-cyclic fluctuations in relative abundance.

Experimental Design and General Sampling Description

To allow hour-scale *in situ* sampling of reef microbiomes, we selected a site that was accessible both day and night (17°29'18.5"S 149°52'54.3"W; **Figure 4A**).

Microbiome sampling was split into two phases. Phase 1 aimed to collect microbiome samples at 6-hr resolution from coral genomic replicates and other reef microhabitats. In this phase we collected corals from an experimental platform (described in detail below; **Figure 4B**) pre-loaded with clonal replicates of two coral species. To contextualize coral microbiomes against other reef microhabitats, we concurrently collected macroalgae, water, and sediment

samples adjacent (2-3m) to the platform. Phase 2 aimed to contextualize the microbiomes of the platform corals from Phase 1 against free-living corals (across the reef-flat) to test for platform effects, sampling at coarser temporal resolution (12-hr). In this phase we collected samples of the same two coral species, three new macroalgae, and water (sampling details summarized in **Figure 4**).

Phase 1 collected microbiome samples of macroalgae, sediment, water, and coral (with controlled genomic replication; **Figure 4I**). To allow sampling of coral genotypic replicates (sourced from the same colony) *without* repeated sampling of contiguous tissue, we pre-fragmented coral colonies, allowed them to recover on a platform on the reef, and then sampled discontinuous fragments over time. In detail, we procured three colonies from each of two species of corals (*Porites rus* and *Pocillipora damicornis*; **Figure 4C,D**) from within 140 m of the experimental platform, fragmented them and randomly affixed the coral nubbins to a steel platform (8 X 8 m, ~2 m deep, suspended 0.3 m above the benthos, and 350 m from shore; **Figure 4A**) using a screw-on attachment methodology used by our group previously ((Clements & Hay, 2019, 2021); **Figure 4B**). These nubbins acclimated for 218 days before microbiome sampling in late-March 2022 (at that time they were ~5 cm tall; **Figure 4C,D**). Phase 1 sampling was performed every 6-hours beginning on 29 March 2022 at 1800 h and ran until 2 April 2022 at 1200 h (**Figure 4I**). At each sampling we collected (detailed below) microbiomes from each of the two coral species (n=3 per coral, with one exception; **Figure 4C,D,I**), one species of macroalgae (*Padina* sp.; n=5; **Figure 4E**), sediment (n=2; **Figure 4B**), and water (n=2; **Figure 4B**). We also collected water samples for nutrient analysis (n=3) and salinity (n=1; **Figure 4B**).

In Phase 2, we aimed to establish a baseline understanding of un-transplanted coral microbiomes with which to compare to our Phase 1 coral microbiomes (**Figure 4I**). To achieve this, we sampled *P. rus* and *P. damicornis* corals from the reef flat that sourced the platform corals from an area ~350 m away from the platform, approximately 8000 m², in water ~1 m deep, at 12-hour intervals, accompanied with samples of macroalgal and water microbiomes. Phase 2 sampling was performed every 12-hours, beginning on 4 April 2022 at 1200 h and running until 7 April 2022 at 0000 h (**Figure 4I**). At each sampling we collected microbiomes (described below) from each of the two coral species (n=3 per coral, 5m separating each colony with no colony resampling), macroalgae [*Hydroclathrus* sp. (n=3, with two exceptions; **Figure 4F,I**), *Dictyota* sp. (n=3; **Figure 4G**), *Halimeda* sp. (n=3; **Figure 4H**)], water (n=2). We also collected water samples for salinity (n=1; **Figure 4B**). This collection design also allowed comparison of corals and water microbiomes between two different depths and collection sites.

All samples were collected within 1 hour of the intended sampling time. In the field, all microbiome samples were preserved for transportation within 5 minutes of collection by either snap freezing on dry ice or in DNA stabilizing buffer (25 mM sodium citrate, 10 mM EDTA and 70 g ammonium sulfate per 100 ml solution, pH 5.2) with ice packs (sample preservation shown in **Figure 1I**). Samples were frozen at -40°C on DNA buffer within two hours of sampling.

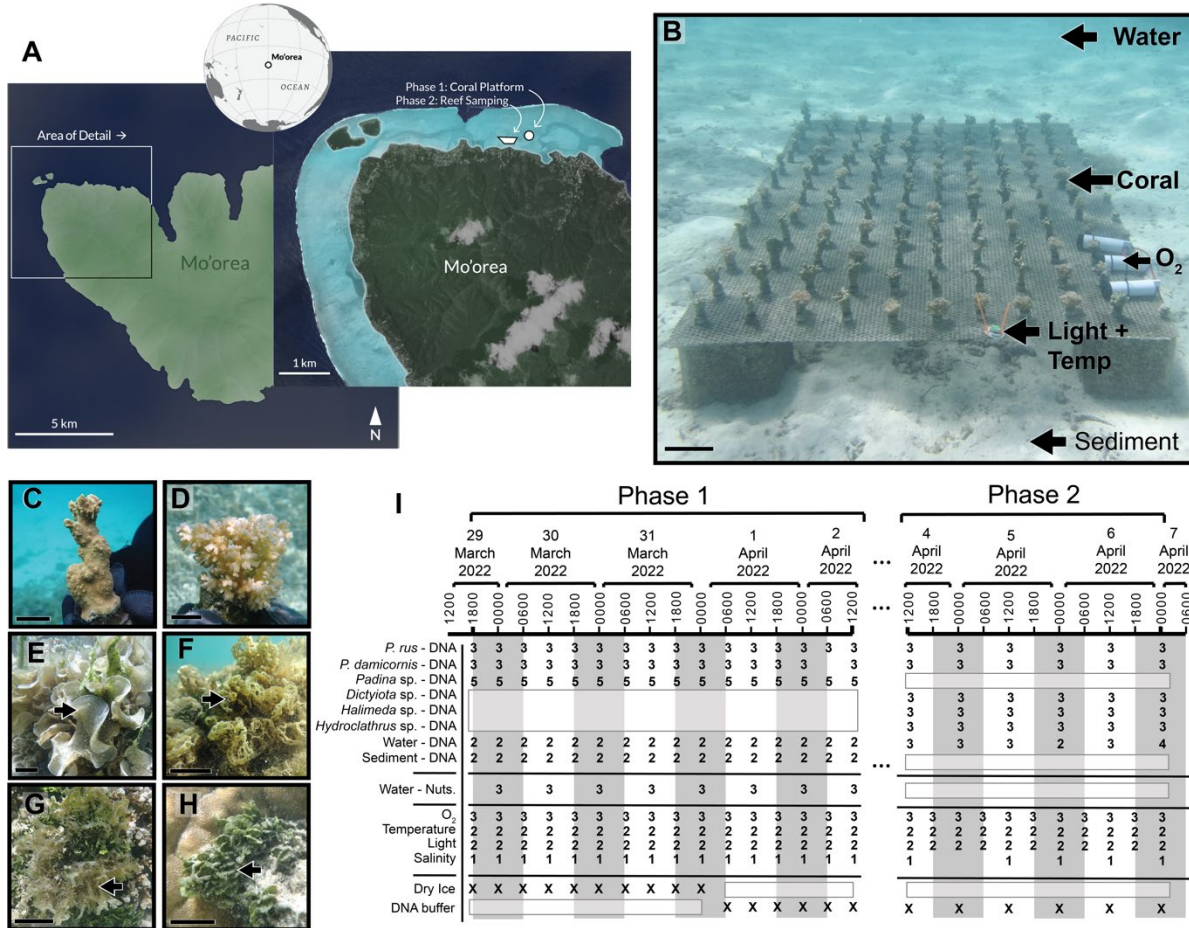


Figure 4: Sampling location, design, and sample types included in reef-scale, time-resolved microbiome sampling. **A)** Location of coral sampling site off the north coast of Mo’orea, French Polynesia. Cartographic data sourced from Copernicus Sentinel 2025; Shom, 2023, MNT topobathymétrie côtier de Moorea (Projet Météo-France), and Natural Earth. **B)** Experimental platform and genotypic coral replicates. Arrows and labels indicate key features of sample collection. “Coral” identifies one coral fragment on the rack, while “water” and “sediment” indicates the general location in-which these sample types were collected. “O₂” and “Temp + Light” identify locations of dataloggers measuring oxygen, temperature, and light. Oxygen dataloggers were clustered to allow locking and prevent theft. Scale bar 10 cm. **C-H)** Representative individuals of corals [**C)** *Pocillipora damicornis*, **D)** *Porites rus*] and macroalgae [**E)** *Padina* sp., **F)** *Hydroclathrus* sp., **G)** *Dictyota* sp., and **H)** *Halimeda* sp.] collected in this study, arrows indicate focal taxa, scale bar 1 cm. **I)** Sampling schematic for reef microbiome survey summarizing sample collection over time. The days and hours that were included in sampling Phases 1 and 2 are enclosed by brackets, with time of day shown graphically (white box as day and gray box as night) and by hour (e.g. 0600 h). Possible samples and field preservation methodologies represented as rows and timepoints as columns. Alphanumeric character indicates samples were collected, numbers describing replicates and “X” when replicates not applicable. Blanks indicate, and gray boxes emphasize, when samples not

collected. O₂, oxygen; nuts., nutrients; temp., temperature. Cartography: Max Borrmann, Photo credit: A. J. Maritan.

Sample Collection and Processing

Microbiome DNA. All sample collection was performed while wearing clean gardening gloves. Corals were sub-sampled by breaking off nubbins inside a sterile plastic wire-closure sample bag (Fisher, Waltham, MA) or with tin snips (Tactix, Centennial, CO) into sterile sample bag. Snips were rinsed thoroughly in ambient seawater between samplings. Coral sampling aimed to avoid epiphytes or endolithic algae while maximizing collected tissue. Coral nubbins (~1 cm³) were placed in a sterile sample bag. Macroalgae collection avoided algal tissue in-contact with corals and minimized epiphytic growth. Algal tissues were collected by hand and placed in sterile sample bags. Sediments were collected by scooping ~15-30 g sediment 1-2 cm deep with a sterile sample bag. For all sterile sample bags, excess water was immediately decanted, and sample preserved. Water samples were collected by syringe filtering 140 mL of seawater through a 0.22 µm Sterivex filter (Millipore, Darmstadt, DE), with the syringe (BD, Franklin Lakes, NJ) flushed twice before collection. After filtration, excess water was purged from the Sterivex with a syringe and backfilled with 2 mL DNA stabilizing buffer. In Phase 1, water samples were collected from 1 m above the platform, while in Phase 2, the water was collected adjacent to the first and last coral samples of that timepoint. Water microbiome samples were constrained to 140 mL filtrate to fit within constrained sampling time.

Geophysical Measurements. For geophysical context we assembled data from: 1) stationary rack-measures, 2) mobile sampling, and 3) public data retrieval. For stationary measures we established a datalogger monitoring site on the experimental platform that operated

over both phases of sampling. At this station, we measured temperature and light with two HOBO MX light/ temp data loggers (Onset Corp, Bourne, MA) and dissolved oxygen with three miniDOT data loggers (PME, Vista, CA), with all dataloggers sampling at 10-minute intervals. miniDOT dataloggers were mounted horizontally to prevent accumulation of air bubbles on the optodes, cleared regularly of biofouling during Phase 1 sampling, and clustered to allow locking to the platform for theft prevention. These data loggers provide geophysical measures at the experimental platform but were ~350 m away from the shallower sites in Phase 2, and as such they are not site-specific in Phase 2, but we expect the platform measurements to approximate the conditions on the reef flat. For mobile sampling, we collected water nutrient samples (n=3) from 1 m above the platform by capturing the flow-through from 0.22 μ m microbiome filtering (described above) in 30 mL HDPE scintillation vials. Nutrient samples were transported on ice before freezing at -40°C (Nalgene, Rochester, NY). Salinity (n=1) was measured 1 m above the platform in Phase 1, and adjacent to the last sample collected in Phase 2 using a handheld refractometer (Agriculture Solutions, Kingfield ME). We used public data to evaluate tide and lunar phase (<https://tides4fishing.com/pi/french-polynesia/tahiti>). Tide and lunar phase were converted into percent of maximum value (e.g. tide maxima to minima and new to full moon) at each sampling timepoint. Times of high and low tide were directly retrieved while intermediate tidal values were modeled using parabolic interpolation, assuming a non-linear rate of tidal change with more rapid changes near the midpoint between high and low tides and slower changes near the extrema. The model takes the form: $T(t) = 100 * (1 - t^2)$ for high to low tide transitions, and $T(t) = 100 * t^2$ for low to high tide transitions, where T is the interpolated tide level (0-100%) and t is the fractional time (0-1) between two consecutive tide extrema.

Field Station Sample Processing

Upon returning to the field station, samples were preserved for DNA analyses. Samples in sterile sample bags were transferred into 2 mL cryovials (Corning Inc., Corning, NY) containing DNA stabilizing buffer (except sediments which remained in sterile bags and has DNA preservation buffer added) and stored at -40°C . Negative controls for storage (DNA buffer) and sampling (sterilized forceps dipped into buffer) were also collected at this time and stored at -40°C . Nutrient water samples were frozen at -40°C and a storage blank was prepared by filling a scintillation vial with Ultrapure water. Sample preservation was performed sequentially and took less than 5 minutes per sample. All samples were in preservation buffer and frozen at -80°C within two hours of sampling. Samples were shipped to Montana State University on 14 April 2022 on dry ice and stored at -80°C until DNA extraction.

Nutrient Quantification

To quantify key reef nutrient concentrations in seawater over the sampling period we quantified nitrite, nitrate+nitrite, ammonium, and ortho-phosphate concentrations at the UC Santa Barbara Marine Science Institute Analytical Lab. There, water samples were subjected to Flow Injection Analysis using the QuikChem 8500 Series 2 (Lachat Instruments, Zellweger Analytics, Inc.). Nitrate concentrations were derived by subtracting Nitrite from Nitrate+Nitrite values.

Microbiome Data Preparation – 16S and 18S rRNA Gene Sequencing

We performed 16S and 18S rRNA gene amplicon sequencing to assess diversity in prokaryotic and eukaryotic microbiomes, respectively.

DNA Extraction. DNA was extracted from each sample in randomized order using the Qiagen PowerSoil Pro kit (Qiagen, Hilden, Germany). In all cases, flame-sterilized tools were used to process and transfer samples from cryovials or sample bags into the Qiagen bead beating tube. For coral and sediment samples, ~0.25g of sample were used, while algae tissue was first shredded into ~0.5 cm² pieces before extraction, while water Sterivex filter cartridges were opened gently with a hammer, the membrane excised with a scalpel (per (Cruaud et al., 2017)) and cut into ~0.5 cm² squares in a UV irradiated petri dish before transfer into the bead-beating tube. Extractions were performed following the manufacturer's protocol utilizing 20 minutes of bead beating. Extracted DNA was quantified using a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA). Extraction blanks were performed (reagents only) for each extraction kit (n=2) along with RNAlater storage blank (n=1) and was carried through the SSU rRNA gene amplification and sequencing.

16S/18S rRNA Gene Amplification and Sequencing. To generate taxonomic and percentage abundance profiles for each sample, we targeted the V4 region of the 16S rRNA gene and V9 region of the 18S rRNA gene for amplicon sequencing. For 16S V4 rRNA gene amplification, samples with DNA concentrations greater than 30 ng/μL were diluted to 10 ng/μL in PCR grade water. PCR amplification using barcoded universal primers 515 F (5'-GTGYCAGCMGCCGCGGTAA-3'; (Caporaso et al., 2011)) and 806R (5'-GGACTACNVGGGTWTCTAAT-3'; (Apprill et al., 2015)), following the thermocycler conditions of 95°C for 3 minutes and 35 cycles of 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds, 72°C for 10 minutes. For 18S V9 rRNA gene PCR amplification used the Earth Microbiome Project primers 1391F (5'-GTACACACCGCCCGTC-3'; (Stoeck et al.,

2010)) and EukBrR (5'- TGATCCTTCTGCAGGTTACCTAC-3'; (Stoeck et al., 2010)) following the thermocycler conditions of 95°C for 3 minutes and 30 cycles of 95°C for 30 seconds, 55°C for 45 seconds, 72°C for 90 seconds, 72°C for 10 minutes. All PCR reactions used the same components and volumes [1 µL template DNA, 0.5 µL F primer, 0.5 µL R primer, 0.25 µL Bovine Serum Albumin, 10.25 µL H₂O, 12.5 µL Go Taq Master Mix (Promega, Madison, WI)]. Samples were then sent to sequencing cores for dual index barcoding and sequencing.

16S rRNA gene sequencing was performed at the Georgia Genomics and Bioinformatics Core on a MiSeq (Illumina Inc., San Diego, CA) using a V2 2x250 bp cycle kit. Samples were split across three sequencing runs each with extraction and storage blanks included on each run. 18S rRNA gene sequencing was performed at Michigan State University Research Technology Support Facility where pooled samples were sequenced on one MiSeq run using a V2 2x150bp cycle kit. Both 16S and 18S amplicon sequencing used 15-25% PhiX to increase read diversity (Altman-Kurosaki et al., 2025).

Sequence Processing

To characterized prokaryotic and eukaryotic community using 16S and 18S rRNA reads to taxonomically classify and quantify amplicon sequence variants (ASVs). QIIME2 (v2022.8; (Bolyen et al., 2019)) was used for amplicon analysis. Raw fastq files were imported separately for each sequencing run, where reads were denoised, merged and filtered to remove chimeras using DADA2 (Callahan et al., 2016) with read trimming parameters 16S: `--P-trim-left-f 15 --P-trim-left-r 5 --P-trunc-len-f 165 --P-trunc-len-r 165`; 18S: `--P-trim-left-f 13 --P-trim-left-r 5 --P-trunc-len-f 137 --P-trunc-len-r 145`, then ASV tables from separate 16S rRNA sequencing runs and were merged into a single table.

Taxonomy was assigned to 16S rRNA ASVs using the Silva 138 515-806 database (Quast et al., 2013). ASVs annotated as mitochondria, chloroplast, and known lab-specific contaminants (Class: Thermoleophilia; Family: Sphingomonadaceae, Caldalkalibacillaceae, Pirellulaceae, Micrococcaceae, Orbaceae, Corynebacteriaceae; Genera: Methylobacteriaceae, Vulcaniibacterium, Bacillus) were removed. Further, all ASVs with a taxonomic designation of Domain Bacteria or Archaea, but with no further lineage detail were aligned against the NCBI nt eukaryote database (accessed January 2025; taxid:2759) with BLASTn (Camacho et al., 2009), with any ASVs that aligned to a eukaryote removed from further analysis. A final cleaning step to remove contaminating sequences was performed using the decontam package (v3.2.0; (Davis et al., 2018)) in R (v4.3.1; (R Core Team, 2023)), where extraction blanks (n=2-3) were used to estimate potential contaminants, with all putative contaminating taxa removed from subsequent analyses (Genus: *Caldalkalibacillus*, *Meiothermus*, *Nesterenkonia*, *Halomonas*, *Bacillus*, *Escherichia-Shigella*, *Pseudomonas*). Taxonomy was assigned to 18S rRNA ASVs by aligning each ASV to the NCBI eukaryotic small subunit rRNA database (accessed January 2025; (Sayers et al., 2024)). The taxonomic lineage from the top alignment (minimum percent identity $\geq 80\%$) was used to classify to each ASV. To remove host sequences, ASVs classified as cnidaria, *Hydroclathrus*, *Dictyoita*, *Halimeda*, or *Padina* were removed from the subsequent analyses. For samples lacking ASVs classified as a known host, the most abundant (n=6) ASVs from each sample type were aligned against the NCBI nt database (accessed January 2025) with BLASTn, and ASVs matching host taxonomy were removed from subsequent analysis.

Statistical Testing and Plotting

Analysis of 16S and 18S rRNA gene sequences aimed to 1) compare reef microbiomes among microhabitats and over site and water depth (e.g. Phase 1 vs. Phase 2) and 2) identify microbiome change over time, taking into account the potential for rhythmic variations from day to night. In all cases, QIIME2 microbiome abundance and taxonomy matrices were imported into R using qiime2R (v0.99.6 (Bisanz, 2018), and manipulated with the phyloseq (v1.44.0; (McMurdie & Holmes, 2013), microbiome, and tidyverse (v2.0.0; (Wickham et al., 2019)) packages.

To **describe and compare microbiomes** we asked four questions: 1) is microbiome composition distinct between microhabitats?, and 2) did phase of sampling affect microbiome composition within sample types?, 3) how does microbial diversity differ between different sample-types (e.g. water vs. sediment)?, and 4) what are the characteristic microbial taxa of each sample type? To do this, we directly compared all microbiomes to one another and did not directly consider temporal sampling. To allow sample comparison, we normalized sequencing effort for all samples to retain the samples with the fewest reads (gene-specific rarefaction depths in **Table A1**). These normalized data were used for all whole-reef analyses. To answer the **first** and **second** questions, we generated abundance-derived distance matrices in QIIME2 and tested clustering between variables of interest using PERMANOVA and PERMDISP as implemented in QIIME2. Sample clustering was visualized with total sum scaled (TSS) data using Non-metric Multidimensional Scaling (NMDS) as implemented in the R package vegan (v2.6.4; (Oksanen et al., 2022)). To answer the **third** question, we generated standard alpha diversity metrics in Phyloseq. These diversity metrics across sample type were visualized with violin plots as implemented in ggpirate (v0.1.2; (Braginsky, 2023)). Differences in mean alpha diversity across

sample types was tested with linear mixed effects model (LME) accounting for repeated sampling of the experimental coral genotypes (fixed effects: sample type, random effects: coral genotype), as implemented in the R packages lme4 (v1.1.34; (Bates et al., 2015) and emmeans (v1.8.8; (Lenth, 2023)). To answer the **fourth** question, we identified the core microbiome (clustered at genus-level) for each sample type with the criteria that a “core” microbe for a sample type must have at least 80% prevalence across all samples and 1% abundance in at least one sample (criteria adapted from (Hernandez-Agreda et al., 2018)).

To identify **microbiome change over time** we asked three questions: 1) do measured geophysical reef conditions change in a cyclic manner in-line with diel cycles?, 2) how consistent are each of the sample type’s microbiome diversity and structure over time?, 3) do individual microbes display diel patterns of abundance? We focused these analyses only on sample types that were collected at 6-hour intervals (Phase 1). To focus our analyses on intra-sample type comparisons (e.g. sediment-to-sediment and water-to-water but ***not*** water-to-sediment) and to use more of the reads per sample (but losing the capacity to compare between sample types), we selected rarefaction depths for each sample type separately, aiming to retain as many samples as possible (species and gene-specific rarefaction depths in **Table A1**). These microhabitat-specific normalized data were used for all time-series analyses. To answer this **first** question, we tested all geophysical time-series for diel cyclicity at the times of microbiome sampling. For continuous data, we pulled the 30-minute average during the sampling times (0001, 0600, 1200, 1800 h). For sampling timepoints, cyclicity was assessed with the R package RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods; v1.36.0; (Thaben & Westermark, 2014)), with multiple hypothesis testing accounted for using a Bonferroni FDR

correction. To answer this **second** question, we generated standard alpha (Shannon diversity) and beta diversity (Bray-Curtis distances) metrics in Phyloseq and QIIME2 respectively. Alpha diversity was plotted over time with line plots. Diversity time-series were tested for diel cyclicity using RAIN, with multiple hypothesis testing accounted for using Bonferroni FDR correction. Differences in mean alpha diversity for each timepoint for each sample type was tested with a linear mixed effects model (LME; fixed effect: timepoint, random effect: coral genotype) for corals and two-way ANOVA for sediment, water, and *Padina* sp., as implemented with the lmer function in the R packages lme4 (v1.1.34; (Bates et al., 2015) or the Anova function in the package car (v3.1-2; (Fox & Weisberg, 2019)) respectively, in both cases using Tukey HSD for post-hoc testing as implemented in the R package emmeans (v1.8.8; (Lenth, 2023)). Differences in beta diversity between times of day were identified with permutational multivariate analysis of variance (PERMANOVA) and permutational multivariate analysis of dispersion (PERMDISP) as implemented in QIIME2, and for all significant time contrasts in a sample type ($P < 0.05$), taxa contributing to community dissimilarity were identified using Similarity Percentage (SIMPER) as implemented in vegan. Sample clustering by time of day was visualized with TSS data using NMDS as implemented in vegan. To answer the **third** question, we identified taxa with cyclic abundance patterns with and then grouped them with unsupervised clustering based on cyclic abundance patterns. To do this, abundance matrices were center log-ratio (CLR) transformed, and detrended. Taxa with diel abundances were identified using the package RAIN, with P values with multiple hypothesis testing accounted for with an adaptive Benjamini-Hochberg FDR correction (Benjamini & Hochberg, 1995). All statistical testing for cyclicity was performed with scripts modified from (Coenen et al., 2020). For plotting, CLR and detrended

abundances for all taxa (cyclic and not) were Z-score transformed (scaled) and then filtered to only include taxa that were cyclic, abundant (greater than 1% for sediment 16S/18S and *P. rus* 18S or 5% for all other sample types in at least one sample), and prevalent (present in 25% of all samples). Scaled 16S and 18S abundances were combined within microhabitats and this matrix was subject to k-medoid clustering (per (Muratore et al., 2022)) to group taxa with similar abundance patterns. The scaled abundances for each taxon were averaged at each timepoint and were plotted alongside the median value for the cluster. The time of peak abundance was assigned qualitatively.

Cartography

Mapping was performed in qGIS (QGIS Association, 2025), where the Mo'orea coastline was traced from Sentinel-2 Imagery (NIR, band 8; (Copernicus Data Space Ecosystem, 2025)), the base map (ocean, land, and graticules) was sourced from Natural Earth Data (Natural Earth, 2025), and elevation data acquired from Shom (Shom, 2023). Final image processing was performed in Adobe Illustrator and Photoshop (Adobe, San Jose, CA).

Results and Discussion

We assessed spatial and short-term (hours to days) variation in microbiomes across diverse habitats at a coral reef in Mo'orea, French Polynesia. Our results show that different reef microhabitats host distinct microbial communities. Furthermore, in-line with cyclic geophysical conditions, we observe several reef microbial communities being structured by time of day, with all microbiomes (excluding those dominated by a single taxon) containing taxa with significant cyclic fluctuations over the diel period. These results identify specific microbiome members

whose functional contributions likely also vary significantly over the span of hours, underscoring a need to account for time of day as predictor of microbiome contributions to reef ecology.

Microbiomes Across the Reef are Distinct and (with Notable Exceptions) Diverse

Prokaryotic and eukaryotic microbiomes varied among reef microhabitats. In both the 16S and 18S rRNA-based analyses, microbiome taxonomic composition differed significantly based on sample type (PERMANOVA, $P < 0.008$; **Figure 5A-D, Table A2**). Sequencing blanks were distinct from all other sample types (PERMANOVA, $P < 0.008$; **Figure 5A,B, Table A2**) and were thus removed for all analyses (**Figure 5C-F**). Microhabitat-specific partitioning is not surprising given the presumed substantial differences in biotic and abiotic factors among microhabitats (e.g., algae vs. animal host), with such factors being well-accepted drivers of microbiome structure (Thompson et al., 2017). In-line with expectations, we observed algae microbiomes clustering (though each macroalgal genus is statistically distinct; PERMANOVA, $P < 0.001$, **Table A2**), while coral prokaryotic microbiomes did not cluster together. Indeed, the *P. damicornis* prokaryotic microbiome is distinctive not just from all other communities, but also appears to be less variable (lower dispersion) than other microenvironments (PERMDISP, $P < 0.001$, **Table A2**).

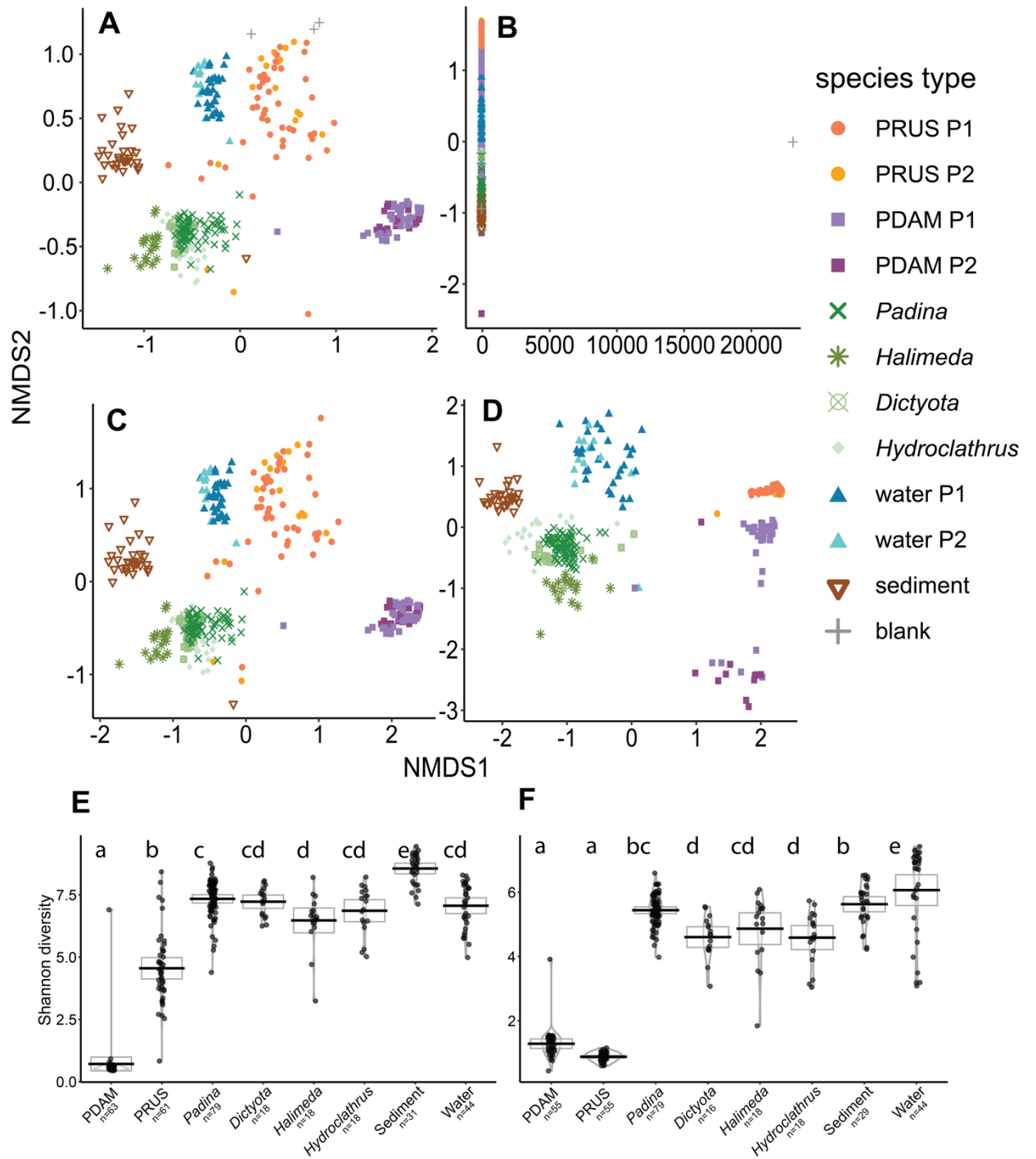


Figure 5: Diversity and community composition of A,C,E) prokaryotic and B,D,F) eukaryotic reef microbiomes. A-D) Beta dissimilarity (Bray-Curtis) between reef microbiomes presented as a NMDS, both A,B) including and C,D) excluding extraction blanks. Each point represents one sample and is colored and drawn by sample type. E,F) Alpha diversity (Shannon diversity) of

reef microbiomes presented as violin plots (mean as estimate of center; box bounded by 95% confidence interval; density estimator used to draw violin; observations as individual points). Dissimilar letters indicate $P < 0.05$ in pairwise comparison between sample type, while sharing of any letter indicates $P > 0.05$ (P values were obtained from linear mixed effects model; Tukey-Kramer HSD). PDAM, *Pocillipora damicornis*, PRUS, *Porites rus*, P1, Phase 1 (1800 h on 29 March 2022 to 1200 h on 2 April 2022, ~6 ft depth, coral rack); P2, Phase 2 (1200 h on 4 April 2022 to 0001h on 7 April 2022, ~3 ft depth, reef flat).

Sample site and depth of microbiome collection also influence the composition of reef microbiomes. When comparing microbiome structure between Phase 1 (~2 m water depth) and Phase 2 (~1m water depth, ~350 m away from experimental rack), we found that microbiomes grouped by sample type (**Figure 5C,D**), but separate based on sample site and depth (proxied by sampling phase). Both *P. rus* and water prokaryotes and eukaryotes communities and *P. damicornis* eukaryotes were distinct between sample sites (PERMANOVA, $P < 0.007$; **Table A3**), while *P. damicornis* prokaryotes were indistinguishable across sites (PERMANOVA, $P = 0.323$; **Table A3**). Qualitatively, it appears that microhabitat type is a more important determinant of microbiome structure than water depth or site (except *P. damicornis* eukaryotes, which cluster with those of *P. rus* in Phase 1; **Figure 5C,D**).

Microbiome alpha diversity also varies significantly with microhabitat type. Based on Shannon diversity, corals were less diverse than any of the other sample types for both prokaryotes and eukaryotes, while *P. damicornis* has a less diverse prokaryotic community than *P. rus* (LME, $P < 0.05$, **Figure 5E,F**). For prokaryotes, water and most algae genera had the same diversity, and all greater than corals, while sediments have the highest diversity (LME, $P < 0.05$, **Figure 5E**). For eukaryotes, *Dictyota sp.*, *Halimeda sp.*, and *Hydroclathrus sp.* algae all have indistinguishable diversity levels, greater than corals, but generally less than that of *Padina sp.* *Padina sp.* and sediments have similar diversity levels, being higher than those of other

macroalgae species but lower than that of the seawater community (LME, $P < 0.05$, **Figure 5F**).

These community-scale patterns raise several questions, including: What is the cause of low diversity and low variability in some microhabitats, notably within the *P. damicornis* prokaryotic microbiome and the eukaryotic microbiome of both corals? And which microbial taxa drive the community compositional differences among microhabitats?

The core microbiome of each reef microhabitat comprised a variable proportion of the community, ranging from 6.6 to 98.4% (**Table A4**). Examining the prokaryotic core immediately reveals that *P. damicornis*' is dominated by *Endozoicomonas* sp., which average 98.14% of all sequences in samples of this coral species (**Figure 6A**). *Endozoicomonas*' noteworthy abundance in *P. damicornis* has been observed in other coral studies (Bayer et al., 2013; Beatty et al., 2018, 2022; Caughman et al., 2021; Meyer et al., 2014; Pogoreutz et al., 2018). The *in situ* behavior of *Endozoicomonas* in corals is unclear, though many studies hypothesize a mutualistic relationship based on *in vitro* and genomic inferences (Pogoreutz & Ziegler, 2024). In contrast, *Endozoicomonas* is found at much lower abundances (<1%) in *P. rus*. In this coral host, a single genus belonging to *Halomonas* sp. is the sole member of the core microbiome, averaging 6.6% of the community across *P. rus* samples. Halomonads have been recovered in other anthozoa and contain genomes encoding genes for managing oxidative stress, dimethylsulfoniopropionate (DMSP) metabolism, degrading aromatic compounds, secretion systems, and flagella (Villela et al., 2023). *Halomonas* spp. have also been identified as a beneficial microbe for corals (Rosado et al., 2023). The core prokaryotic taxa of both sampled corals (*Halomonas* sp. and *Endozoicomonas* sp.) are in the order Oceanospirillales. The coral eukaryotic microbiome consists nearly entirely of *Symbiodinium* sp. at an average of 96%; *Symbiodinium* sp. can also

be found in association with the macroalgal species (*Padina sp.*, *Halimeda sp.*, and *Dictyota sp.*) but at lower average abundances (6-19.5%; **Figure 6B**, **Table A4**). *Symbiodinium* dominance in corals is unsurprising as these dinoflagellates are the iconic symbionts of corals, reaching densities of 3×10^5 cells/mg coral tissue homogenate (Putnam & Edmunds, 2011) and providing the coral with fixed carbon (Sorek et al., 2018).

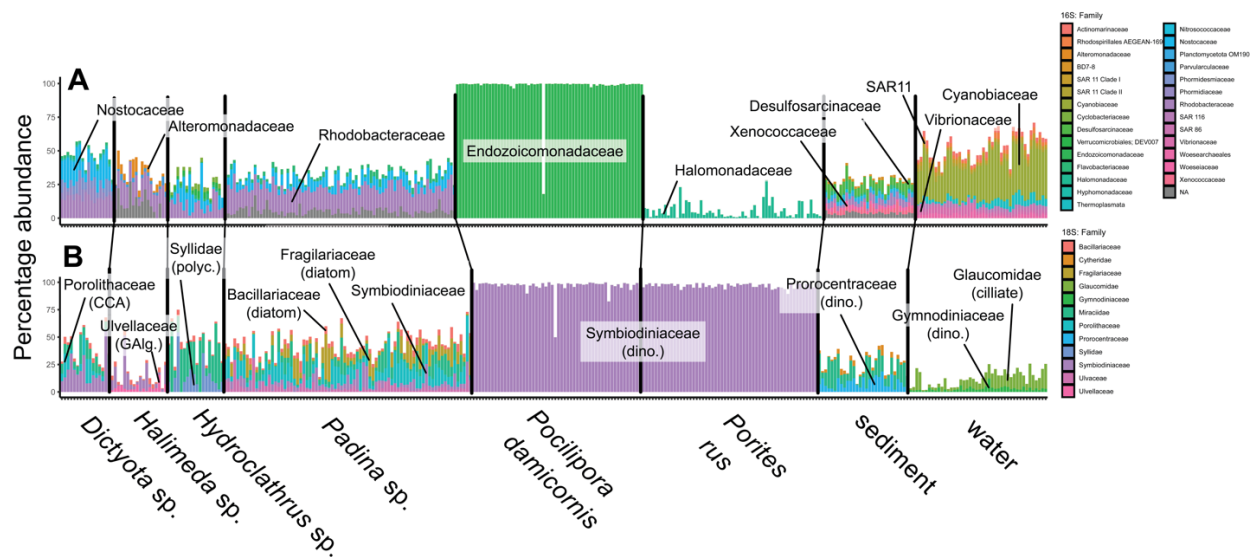


Figure 6: The core microbes of different reef microbiomes presented as percentage abundance for A) prokaryotes and B) eukaryotes. Core microbes are defined here as genera that were present in at least 80% of all samples for a given sample type and had a percentage abundance greater than 1% in at least one of those samples. Key taxa are identified with lines and labels. CCA, crustose coralline algae; Dino., dinoflagellate; GAlg, green algae; Polyc., polychaetae.

The macroalgae core microbiome was predominantly epiphytes (diatoms, dinoflagellates, cyanobacteria, green algae, and red algae), which has been observed previously (Cyanobacteria (Kuba et al., 2021) and diatoms (Busetti et al., 2017)). Bacteria are important algal symbionts, altering growth and providing nutrients (Park et al., 2022; Singh & Reddy, 2016), while algae interact with bacteria by producing secondary metabolites and organics that alter bacterial growth (Park et al., 2022). For core prokaryotes, all sampled macroalgae hosted

Rhodobacteraceae (average 10-17%), a common feature in macroalgal microbiomes (Lu et al., 2023; Park et al., 2022) with representatives forming opportunistically mutualistic symbioses with macro- (Spoerner et al., 2012) and microalgae (Seyedsayamdost et al., 2011).

Flavobacteraceae (average ~2%) are common symbionts and have been reported to degrade macroalgal products (Lu et al., 2023; Nahor et al., 2024). The analysis of core taxa also detected macroalgae-associated metazoans (e.g. Syllidae polychaetes and arthropods). Macroalgae-associated polychaete communities (syllid dominated) have been previously described, with community-scale guilds shifting from carnivory to omnivory over seasons (Craveiro & Rosa Filho, 2024).

The reef water core microbiome contained prokaryotic (*Synechococcus* sp. and *Prochlorococcus* sp.) and eukaryotic (*Lepidodinium* sp.) phototrophs, *Vibrio* sp., Rhodobacteraceae, and small genome bacteria and archaea (SAR11 Clades Ia and II, SAR86, SAR116, *Ca. Actinomarina*, and Woese archaeae; **Figure 6A,B, Table A4**). This microbiome is consistent with other studies of reef (L. Weber & Apprill, 2020), coastal (C. Becker et al., 2020; Patin et al., 2021), and pelagic microbes (Aylward et al., 2015; López-Pérez et al., 2020).

The sediment microbiome is diverse with the core microbiome comprising ~25% of the prokaryotic and eukaryotic communities (**Figure 6A,B, Table A4**). This core microbiome includes phototrophs (cyanobacteria and dinoflagellates) and potential sulfur reducers (Desulfosarcinaceae, average abundance ~6%) consistent with other studies photic marine sediments (Billerbeck et al., 2007; Jørgensen, 1977).

Biologically-relevant Reef Geochemistry Exhibits Diel, not Tidal, Cyclicity

Light, temperature, and oxygen levels all exhibit diel periodicity with daytime maxima (RAIN, $P < 0.018$, adjusted Benjamini-Hochberg, **Figure 7A-C; Table A5**). Light intensity change is a first order diel process, directly linked to the 24hr rotation of the Earth, causing second order effects. Increased temperature is caused by radiative heating from light exposure (Jimenez et al., 2011), while reef oxygen changes are also due to light exposure, where coral photosymbionts supersaturate reef-scale oxygen levels (reviewed in (H. R. Nelson & Altieri, 2019)). Because oxygen concentrations are highest in coral tissues, and decrease as it diffuses out of the coral (de Beer et al., 2000; Kühl et al., 1995; Pacherres et al., 2020; Shashar et al., 1993) (Shashar et al., 1993, 1996) (Pernice et al., 2020; Ricci et al., 2019), our measured oxygen (5-10 cm away from coral fragments) is likely an underestimation of coral tissue oxygen concentrations. Nutrients (**Figure 7D**) did not exhibit cyclic concentrations, and are in-line with measurements on other coral reefs (C. Becker et al., 2020; L. Weber & Apprill, 2020). In prior studies, cyclic nutrient concentrations have been described (e.g. phosphate and silica; (L. Weber & Apprill, 2020)). Average salinity was 37 ppt over the course of sampling with variation of 3 ppt, aligning with other reef time-series measurements (C. Becker et al., 2020); but did not exhibit diel (24 hr) or tidal (~12hr) periodicity (**Figure 7E**).

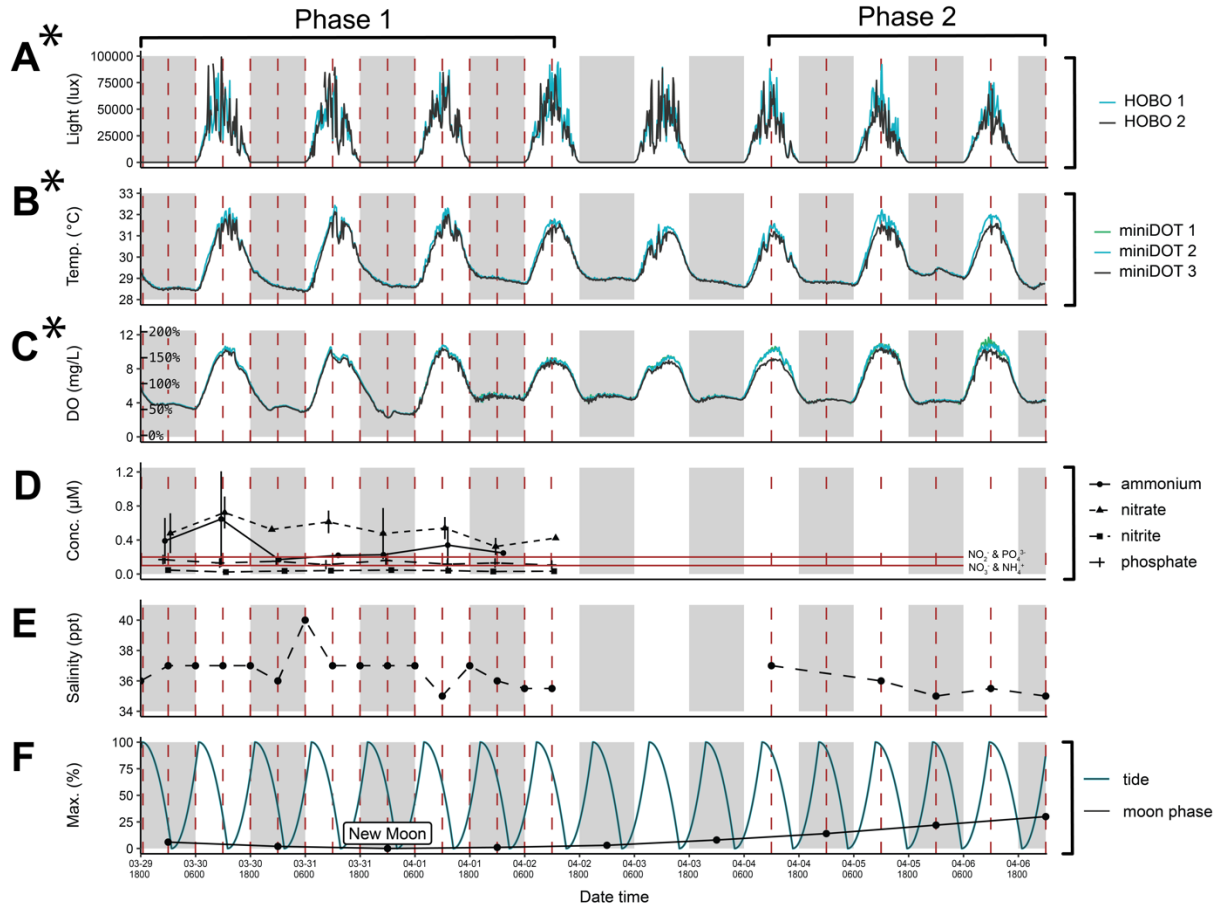


Figure 7: Geophysical conditions during time-series microbiome sample collection between 29 March and 6 April 2022. Only Phase 1 samples are used in time-series microbiome analyses, but entire sampling period presented for completeness. Dates and time indicated on the x-axis, with phase of sampling indicated by brackets and each discrete sampling event indicated with dashed red vertical lines. Continuous samples were collected with dataloggers, sampling at 10-minute intervals, with each replicate measure in a different color. Time-series with statistically detectable diel (24 hr) cyclicity indicated with an asterisk adjacent to the panel's identifying letter. (Rhythmicity Analysis Incorporating Non-parametric Methods, $P < 0.05$, Bonferroni correction). **A**) Light intensity and **B**) temperature were measured by HOBO dataloggers (Onset Corp, Bourne, MA; $n=2$). while **C**) dissolved oxygen (concentration and percent saturation) was measured with optode-based miniDOT data loggers (PME, Vista, CA; $n=3$). Discrete samples were collected at the time of microbiome sampling, and only capture 6- or 12-hour resolution. **D**) Key nutrients (nitrate, nitrite, ammonium, and phosphate) were measured by flow injection analysis, and plotted as jittered points around the sampling time with error bars displaying standard deviation ($n=3$). Limit of detection for different analytes displayed as solid red lines at the base of the panel and labeled with limited analytes (ammonium and nitrate were only analytes above limit of detection). **E**) Salinity was measured with a handheld refractometer ($n=1$). For discrete sampling, timepoints are connected with dashed lines to aid visibility and do not indicate predicted trends. **F**) Lunar phase and timing of high and low tides were acquired

from online resources (<https://tides4fishing.com/pi/french-polynesia/tahiti>), plotting tide and lunar phase as percent of maximum. Points for lunar phase were plotted at midnight but does not indicate time of lunar transit. Intermediate (not-high or low) tidal values were interpolated parabolically at 10-minute intervals assuming a non-linear rate of tidal change. Gray boxes indicate nighttime period, per light measurements in panel **A**). Conc., concentration; DO, dissolved oxygen; max., maximum; ppt, parts per thousand; sat., saturation; temp, temperature.

In this study, the relationship between light intensity and time-of-day are simple as we sampled over the spring equinox, meaning that 0000, 0600, 1200, and 1800 h are true midnight, dawn, noon, and dusk (**Figure 7A**). Additionally, Phase 1 was performed during the new moon meaning incident light at night was negligible (**Figure 7G**). Furthermore, in this study, unlike other reef time-series studies (e.g. (C. Becker et al., 2020; Caughman et al., 2021; L. W. Kelly et al., 2019; L. Weber & Apprill, 2020)), the sampling duration was sufficiently long to decouple tidal variation diel processes (Phase 1; **Figure 7F**), which allows 24-hour cyclicity be interpreted as diel, not tidal.

Because oxygen (H. R. Nelson & Altieri, 2019), temperature (Gillooly et al., 2001), and UV irradiation (Boeing et al., 2004) are primary determinants of biological structure and function, we would expect host physiology (reviewed in (Heath-Heckman, 2016)) and microbiome physiology (e.g. (Sorek et al., 2018)) and structure to co-vary with these parameters (e.g. (Vislova et al., 2019)).

Time Structures Microbiomes Across the Reef

Examining microbiome diversity across microhabitats sampled at 6-hour scales (Phase 1; **Figure 8A-E**) reveals wide temporal fluctuations in microbiome structure. While microbiome alpha diversity (Shannon index) did not vary in a repeated cyclic pattern in any habitat type (RAIN, $P > 0.330$; **Table A6, Chapter Figure 5**), values varied significantly over time in a

stochastic pattern in a few microbiome habitats (16S: water; 18S: water, *P. rus*, *Padina sp.*, and sediment; LME, $P < 0.05$, **Table A7, Figure 8A,C-E**). For example, on multiple occasions during the sampling period, mean alpha diversity of the prokaryotic microbiome in *P. rus* increased by 1/3 in a span of 6 hrs (**Figure 8A**). The eukaryotic microbiome of the seawater was even more variable, increasing 2-fold in the span of 24 hrs during the sampling period (**Figure 8E**). Such large fluctuation highlights that single timepoint sampling may inaccurately estimate community diversity in some microhabitats.

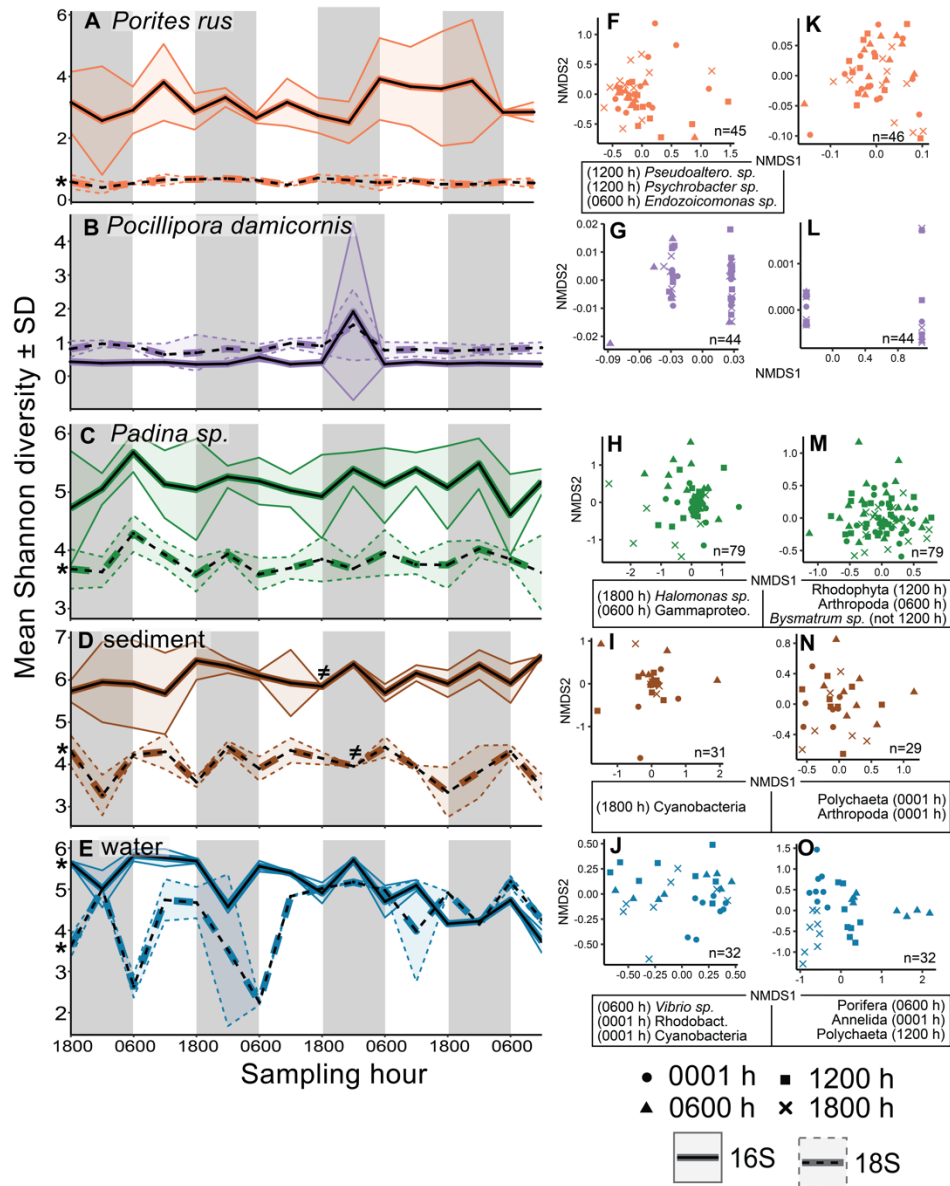


Figure 8: Diversity and community composition for individual reef microbiomes sampled at 6-hour intervals (Phase 1: 1800h 29 March – 1200h 02 April 2022) using amplicon sequence variants. Microbiome samples for **A,F,K)** *Porites rus*, **B,G,L)** *Pocillipora damicornis*, **C,H,M)** *Padina sp.*, **D,I,N)** sediment, and **E,J,O)** water were rarified separately to maximize read recovery but does not permit direct comparison between sample types. **A-E)** Mean alpha diversity for prokaryotes and non-host eukaryotes over time were plotted as solid and dashed traces respectively with standard deviation shown as background shadow. Timepoints with only one replicate are indicated with an “#”. Time-series with at least one sample with distinct diversity values will be annotated an asterisk (LME/ ANOVA p-value ≤ 0.056) to the left of the first value of the trace (significant contrasts detailed in **Table A7**). Gray boxes indicate nighttime (per light measurements in **Figure 7A**). Discrete sampling intervals are connected with lines for

legibility, but do not indicate projections. **F-O**) Microbiome beta diversity (Bray-Curtis) presented as Non-metric Multidimensional Scaling (NMDS) for **F-J**) prokaryotes and **K-O**) non-host eukaryotes. Sample types with at least one time of day differing from one other (PERMANOVA, $P < 0.05$) are shown with box below the NMDS displaying taxa contributing to the dissimilarity and associated time of day (Similarity Percentage, $P < 0.05$; list of temporally distinct samples and explanatory taxa in **Chapter 2 Tables 8,9**). For NMDS, each sample is represented by a single point, colored by species and drawn by sampling hour, replicates shown in bottom right corner. Outlier sample 'D3-T2-EC5-dna' (visible in panel **B** 0001 h) was removed from all *P. damicornis* ordinations to aid visualization and avoid disproportionate impact on statistics. Gammaproteo., Gammaproteobacteria; *Pseudoaltero.*, *Pseudoalteromonas*; Rhodobact., Rhodobacteraceae; SD, standard deviation; 16S, 16S rRNA gene; 18S, 18S rRNA gene; PDAM, *Pocillipora damicornis*; PRUS, *Porites rus*.

A similarly large temporal fluctuation in beta diversity was also observed in most microbiome habitats (16S: *P. rus*, *Padina sp.*, sediment, water; 18S: *Padina sp.*, sediment, water; PERMANOVA, $P < 0.05$; **Table A8**), with communities clustering by time (**Figure 8F,H-O**), especially clear qualitatively in water microbiomes (**Figure 8J,O**).

Specific taxa contributed disproportionately to microbiome compositional differences over time. Specifically, light responsive prokaryotic phototrophs drive microbiome changes in abiotic (sediments and water) microhabitats, with relative abundances peaking in the evening (1800 to 0001 h (SIMPER, $P < 0.05$; **Table A9**, **Figure 8I,J**). This evening peaking of cyanobacterial abundances aligns with studies examining absolute abundances of cyanobacteria in reef (C. Becker et al., 2020; L. Weber & Apprill, 2020) and open ocean (Ribalet et al., 2015) habitats. Heterotrophs (*Vibrio sp.* and Rhodobacteriales) also drive community change in the water (SIMPER, $P < 0.05$; **Table A9**, **Figure 8J**), broadly paralleling open-ocean studies phototrophs propagate diel responses into heterotrophs (Aylward et al., 2015; K. W. Becker et al., 2021).

Just as cyanobacteria and dinoflagellates are the center of diel change in sediments and water, phototrophic coral and algae drive changes to their associated heterotrophs. In both the coral (*P. rus*) and macroalgal (*Padina sp.*) prokaryotic communities, microbes driving change are predominantly heterotrophs (e.g. coral: *Endozoicomonas sp.*, *Psychrobacter sp.*, and *Pseudosalteromonas sp.*; macroalgae: *Halomonas sp.*; **Table A9, Figure 8F,H**). These taxa have been previously described as cyclic taxa in coral microbiomes (*Psychrobacter*, (L. W. Kelly et al., 2019); *Endozoicomonas sp.*, and *Pseudosalteromonas sp.* (Caughman et al., 2021)) and, as discussed previously, may play beneficial roles in corals (Rosado et al., 2019) and algae (Spoerner et al., 2012).

The coral microbiomes dominated by a single microbe (eukaryotic *Symbiodinium sp.* in both corals and prokaryotic *Endozoicomonas sp.* in *P. damicornis*) did not have detectable community change over time (**Table A9, Figure 8G,K,L**).

Though we identified community-scale compositional changes, the marker gene sequencing employed in this study cannot detect changes to absolute abundances of microbes, which may conceal other diel patterns. With this caveat, our results show significant community-scale variation over short timescales across multiple reef habitats. Phototrophs both as host (coral and algae) and microbiome member (cyanobacteria and dinoflagellates) seem to propagate diel changes to heterotrophs. Though we observe community-scale changes over time, this raises the question: which individual microbes have cyclic patterns of abundance?

Microbial Abundances Exhibit Diel Periodicity Across Microhabitats

Individual genera across the coral reef exhibit diel periodicity in relative abundance, with phototrophs (e.g. cyanobacteria, diatoms, and dinoflagellates) being the most consistently cyclic

microbes. However, multiple putative heterotrophs also undergo cyclic fluctuations (**Figure 9, Table A10**).

Water-associated taxa have the strongest diel signals (**Figure 9D**), with abundances peaking at noon, dusk, dawn/noon, and midnight. The noon-peaking community included cyanobacteria (*Prochlorococcus* sp.), SAR11, dinoflagellates, and diatoms. Different cyanobacterial genera peaked at each of the times of day (dusk: *Schizothrix* sp., dawn/noon: Gastranaerophilales). Interestingly, the noon peaking group also included fish (Actinopteri) eukaryotic sequences, potentially reflecting daytime metazoan activity (**Figure 9D, Table A10**).

In the coral *P. rus*, the microbes with clearest cyclic abundances peaked somewhere between dawn and noon, and included *Alteromonas* sp., *Pseudoalteromonas* sp., Flavobacteraceae, and *Mycobacterium* sp. (**Figure 9A**). A prior study of *Porites* spp. showed similar patterns of daytime (dawn-noon) peak abundances, but with different taxa peaking at these times (Caughman et al., 2021). These daytime abundance maxima may be driven by coral-metabolite production. Corals exude particulate organic carbon and nitrogen (Kurihara et al., 2018) and specifically DMSP-derived compounds (Masdeu-Navarro et al., 2022) during the day. Coral-associated *Alteromonas* sp. and *Pseudoalteromonas* sp. have been described as dimethyl sulfide degraders (a DMSP-byproduct; (Raina et al., 2009)), which may be replicating in-response to nutrient inputs, as observed in the mammalian gut (Schmalle & Lorentz, 2020). *Mycobacterium* is generally considered a parasite, with clear instances of pathogenesis in fish (reviewed in (Jacobs et al., 2009)) and invertebrates (reviewed in (Davidovich et al., 2020)). In this apparently healthy coral, it appears that *Mycobacterium* is responsive to diel cues (coral-mediated or abiotic), and its role in this setting is unclear.

In the macroalgae *Padina* sp., dinoflagellates had the clearest cyclic abundances, peaking around midnight (**Figure 9B**). Also peaking around midnight were bacteria (Alteromonaceae and Rhizobiales). It is possible that macroalgae-associated rhizobiales may be involved in cyclic nitrogen fixation in this low dissolved nitrogen environment (**Figure 7D**) as has recently been demonstrated in pelagic diatom-rhizobiales (Tschitschko et al., 2024) and seagrass-*Ca. Celerinatantimonas neptuna* (Mohr et al., 2021) diazotrophic symbioses.

Sediments also contain cyclic taxa, with the clearest cyclic signals originating from polychaete worms, with abundances peaking at midnight (**Chapter 6 Figure 6C**). This parallels coastal observations of nocturnal polychaetes emergence at the sediment surface for carrion-feeding (C.-G. Lee et al., 2004). There is also evidence for cyclic abundances in sediment-hosted potential predatory bacteria (Myxococcota), archaea (Thermoplasmata) and sulfur reducers (Desulfosarcinaceae). Though presently unclear, variable abundances of functional guilds (e.g. predators or sulfur reducers) may play cascading roles in hour-scale population dynamics or geochemical cycling (as has been observed in hour-scale changes to methane production (De Groot et al., 2024)).

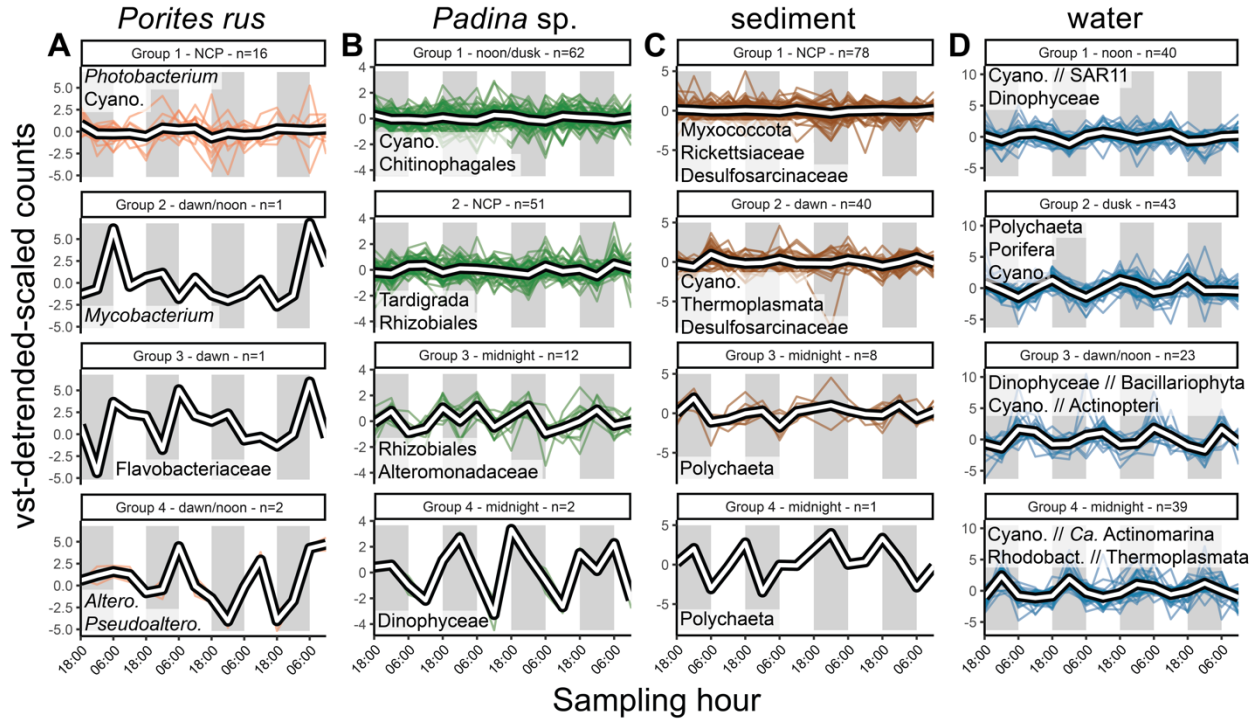


Figure 9: Time-series of cyclic, abundant, and prevalent genera (both prokaryotic and eukaryotic) for reef sample types collected at 6-hour intervals, clustered into four groups by temporal abundance trends. Taxa with cyclic abundances were identified statistically (RAIN, Adaptive Benjamini-Hochberg, $P < 0.05$), filtered to only include the most abundant ($\geq 1\text{-}5\%$) and prevalent ($\geq 25\%$ of all samples) taxa, then placed into four unsupervised groups. Percentage abundances of each taxon was detrended, variance stabilized transformed (vst), and then scaled (Z-scored) to allow inter-comparability between taxa of different abundances (0 is average taxon abundance, positive is above average and vice versa, 1 scaled unit is 1 standard deviation from the mean). At each timepoint, replicate measures of any taxon's abundance were collapsed into a single value and plotted as a trace colored by sample type with focal taxa noted in the sub-panel. Each group is labeled with arbitrary group number, time of median peak abundance, and number of unique taxa in each group. The white trace is the median value for all taxa at each timepoint in the cluster. Discrete timepoints are connected for legibility, but do not indicate projections, gray boxes indicate nighttime on the reef (per light measurements in **Figure 7A**). This was performed for **A) Porites rus**, **B) Padina sp.**, **C) sediment**, and **D) water** samples. *Pocillipora damicornis* was excluded from this visualization because its microbiome (dominated by *Endozoicomonas* sp. and *Symbiodinium* sp.) had no cyclic taxa that met the abundance and prevalence criteria. *Altero.* *Alteromonas* sp.; Cyano., Cyanobacteria; NCP, no clear pattern; *Pseudoaltero.*, *Pseudoalteromonas*; Rhodobact., Rhodobacteraceae; vst, variance stabilized transform.

Conclusions

Diverse reef microhabitats host distinct microbiomes and all sampled microbiomes that were not dominated by a single taxon exhibited significant change over the sampling period, either in a stochastic or cyclic pattern. Across microhabitats, these short term changes were driven by phototrophs, as has been observed in other free-living (e.g. (Aylward et al., 2015)) and host-associated (e.g. (Hubbard et al., 2018; Sorek et al., 2018)) habitats. Identifying taxa with the most pronounced abundance fluctuations can help estimate the potential for fluctuation in biochemical flux or ecological interaction, presuming that these microbial taxa can be linked to a functional activity (metabolic output, predation, migration, etc.). While our analysis of 16S and 18S rRNA genes provides only a first-pass identification of taxa and cannot unambiguously link taxa to function, the results nonetheless suggest a strong potential for both cyclic and stochastic change in microbiome ecology over short timescales in all reef microhabitats. Further investigation into microbiome change over hourly to daily scales would benefit from genomic, transcriptomic, or metabolomic datasets. Such methods are valuable but are challenging or cost-prohibitive to implement in certain reef microhabitats, notably for host-associated systems. Other measurements would also be valuable in assessing ecological implications of short-term microbiome change. These include measurements of absolute cell numbers, nutrient concentration, or chemical transformation rates. Our results help identify those reef microhabitats or microbiome systems (e.g., *P. rus* prokaryotic microbiomes, water column eukaryotic microbiomes) in need for sampling across multiple hourly timepoints in order to most accurately assess microbiome diversity and functional significance. The results also identify those core microbiome members potentially of high consequence in the reef system. These

include taxa that are prevalent across time and responding to diverse environmental conditions, including presumed day-night changes in the biochemistry of the host or surrounding environment (e.g., *Endozoicomonas sp.* bacteria in coral microbiomes; Rhizobiales bacteria in macroalgae microbiomes; dinoflagellates and polychaetes in sediment). Such taxa are clear targets for future research to more accurately predict the cost-benefit of the microbiome, and individual microbiome members, to reef health

Permissions and Permits

All experiments in French Polynesia were conducted in accordance with relevant permissions (**Protocole d'Accueil 2017–2021**) from the Délégation à la Recherche and the Haut-commissariat de la République en Polynésie Française (DTRT). Coral samples utilized in microbiome analyses were exported from French Polynesia to the United States in accordance with CITES permit FR2298700091-E.

CHAPTER THREE

DOMINANT CORAL ENDOSYMBIONT EXPRESSES GENES
FOR ACQUIRING HOST ATP, VITAMINS, SUGARS, AND
MORE

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Andrew J. Maritan

Contributions: formulated guiding hypotheses, designed field experiments, collected samples, performed molecular microbiology analysis, devised bioinformatic strategy, performed bioinformatics, wrote manuscript

Author: Zoe A. Pratte

Contributions: designed field experiments, collected samples, assisted in sample processing and bioinformatics

Author: Cody S. Clements

Contributions: designed field experiments, constructed experimental apparatus, collected samples

Author: Mark E. Hay

Contributions: designed field experiments

Author: Frank J. Stewart

Contributions: formulated guiding hypotheses, designed field experiments, advised on experimental and analytical strategies, wrote manuscript

Manuscript Information

Andrew J. Maritan, Zoe A. Pratte, Cody S. Clements, Mark E. Hay, Frank J. Stewart

Status of Manuscript:

- ☒ Prepared for submission to a peer-reviewed journal
- ☐ Officially submitted to a peer-reviewed journal
- ☐ Accepted by a peer-reviewed journal
- ☐ Published in a peer-reviewed journal

Abstract

Endozoicomonas bacteria are prevalent and abundant in coral microbiomes. Although *Endozoicomonas* is often assumed to be a mutualist, its functional contributions to coral health and physiology are incompletely characterized and likely vary among hosts or over time with changes in host physiology or environmental conditions. Prior studies of *Endozoicomonas* function have primarily evaluated genomic potential or *in vitro* gene expression or metabolomics. Data capturing *in situ* physiology are limited. We evaluate *in situ* transcriptome patterns of coral-associated *Endozoicomonas*, focusing on a population of the reef-building coral *Pocillopora damicornis* in Mo'orea (French Polynesia) with microbiomes dominated (>85%) by *Endozoicomonas*. Bulk tissue transcriptomics from *P. damicornis* collected over day-night periods yielded 1,241 functionally distinct transcripts aligning to *Endozoicomonas* sequences. While temporal patterns were not identified in the *Endozoicomonas* transcriptome, *Endozoicomonas* transcribed diverse genes for aerobic heterotrophy, including genes for nearly the entire glycolytic pathway, tricarboxylic acid cycle, and oxygen-reducing electron transport chain, with *Endozoicomonas* appearing to fuel its central carbon metabolism with diverse host or algal-derived sugars. *Endozoicomonas* also transcribes genes to acquire ATP, vitamins B₁ and B₁₂, TCA-intermediates, amino acids, nucleotides, polyamines, phosphate, and trace metals from the host environment. Few transcripts could be clearly linked to pathways by which *Endozoicomonas* provides substrates to the coral host. While the costs or benefits of *Endozoicomonas* to coral health remain unclear, our results suggest that *Endozoicomonas*, under certain circumstances, may function primarily as a coral parasite.

Introduction

Corals are globally in-decline and their resident microbes have the potential either contribute to this mortality (Shilling et al., 2021) or actively resist death (Rosado et al., 2019). One microbe of interest in this work is *Endozoicomonas* spp.

Endozoicomonas has garnered attention for its high **abundance and wide distribution** in coral microbiomes. *Endozoicomonas* can be found dominating coral microbiomes at abundances upwards of 80% (Bayer et al., 2013; Beatty et al., 2018, 2022; Caughman et al., 2021; Meyer et al., 2014; Pogoreutz et al., 2018). Endozoicomonads (Gammaproteobacteria) and their close relatives often associate with marine animals (hard (Caughman et al., 2021) and soft corals (Kellogg & Pratte, 2021); sponges (Fiore et al., 2015); fish (Katharios et al., 2015), and mussels (Porrás et al., 2024)). Though some strains are invertebrate and fish pathogens (Bennion et al., 2021; Katharios et al., 2015), the valence of the coral-*Endozoicomonas* relationship is almost always portrayed as mutualistic, based on correlative observations, genomic predictions, and *in vitro* bioassays (reviewed by (Pogoreutz & Ziegler, 2024)).

Associative ecological evidence can be interpreted to place *Endozoicomonas* in a positive or negative light. In a positive sense, *Endozoicomonas* is highly abundant in putatively healthy corals in marine protected areas (Beatty et al., 2019), and the genus is found in high abundance in un-lesioned corals (Meyer et al., 2014). *Endozoicomonas* abundances decrease when lesions develop (S. T. M. Lee et al., 2015) or when corals become stressed (Pogoreutz & Ziegler, 2024).

In a negative sense, it has been noted that lower *Endozoicomonas* abundances in stressed corals may indicate a symbiont “fleeing” a sinking ship, and might not necessarily a loss of a

mutualistic bacterium (Pogoreutz & Ziegler, 2024). Further it has been noted that some coral microbiomes dominated by *Endozoicomonas* do not change when disturbed, which might indicate metabolic inflexibility, potentially lowering coral resilience to these stressors (Pogoreutz et al., 2018). In all cases, it should be noted that reported abundances are relative, and it is unclear if absolute populations of *Endozoicomonas* change in any of these cases.

Endozoicomonas' community dominance, and association with coral stress raises the question: **what does *Endozoicomonas* really do inside its coral host?**

In corals, Endozoicomonads are often intracellularly localized, found as the dominant member of dense multi-species microbial aggregates (Bayer et al., 2013; Maire et al., 2023; Neave, Rachmawati, et al., 2017; Wada et al., 2022). *Endozoicomonas* inhabits separate tissues than corals' dinoflagellate photosymbionts (Zooxanthellae), suggesting *Endozoicomonas* interacts predominantly with the coral host (Maire et al., 2023). Analyses of *Endozoicomonas* genomes have provided insight into its metabolic potential. *Endozoicomonas* genomes are large (~5-7 Mb) and are metabolically diverse (Pogoreutz & Ziegler, 2024). Multiple studies have shown that Endozoicomonads are heterotrophs, with genes for both glycolysis and the tricarboxylic acid cycle (TCA; (Maire et al., 2023; Wada et al., 2022)). Though the papers rarely discuss it specifically, *Endozoicomonas* genomes encode terminal cytochromes for aerobic respiration (**Figure 14A**), and possess genes coding for denitrification (*napAB*; (Tandon et al., 2022)). Several studies highlight the metabolic potential for a diverse array of transporters for import [sugars (glucose, maltose, maltodextrin, galactose, maltose/ maltodextrin), amino acids, polyamines, phospholipids, phosphate, trace metals (molybdate, iron, and zinc)] and export (sodium, heme, lipoprotein, LPS, and heme siderophores; (Maire et al., 2023; Wada et al.,

2022)). Several papers highlight the presence of genes for types 1,2,3, and 6 secretion systems (Ding et al., 2016; Maire et al., 2023; Wada et al., 2022). While other genomes encode polyphosphate generation and breakdown (Wada et al., 2022), indicating possible phosphorous storage. Some *Endozoicomonas* also possess genes for B vitamin synthesis (Neave, Michell, et al., 2017; Pogoreutz et al., 2022) which is generally interpreted as a potential source of B vitamins for the coral host (Pogoreutz et al., 2022; Pogoreutz & Ziegler, 2024). *Endozoicomonas*' associative ecological observations and diverse metabolic capacities has led to *Endozoicomonas* being hailed as a coral mutualist, assuming its hypothesized benefits to the holobiont (e.g. vitamin provision) outweigh the costs of supporting this abundant intracellular bacterium (Pogoreutz & Ziegler, 2024).

Other studies have gone beyond genomic predictions to understand the *Endozoicomonas*-coral relationship using isolates with ***in vitro* transcriptomics, proteomics, and bioassays**. Some of these experiments pursued specific questions. One such study identified genes for DMSP degradation, and confirmed expression and activity with rt-qPCR and gas chromatography (Tandon et al., 2020). These results were extended using transcriptomics and stable isotope probing to definitively demonstrate that *Endozoicomonas* could degrade DMSP (a coral holobiont product; (Masdeu-Navarro et al., 2022)) into DMS for use as a sulfur and carbon source (Chiou et al., 2023). Other studies were more exploratory, often heating and/or adding crude coral tissue extracts to *Endozoicomonas* cultures. One such exploration examined the transcriptional and proteomic response of *Endozoicomonas* to the addition of coral tissue extract (Pogoreutz et al., 2022), observing that coral metabolites decreased *Endozoicomonas* flagellar gene expression and increased expression of diverse transporters, while proteomics showed an

increase in glycolytic-associated proteins (Pogoreutz et al., 2022). Heating *Endozoicomonas* isolates to coral-stressing temperatures (31-33°C), increased heat shock protein levels and decreased antioxidant defense proteins (Chan et al., 2022). Combining coral-tissue metabolomics, metagenomics, and *in vitro* assays showed that *Endozoicomonas* isolates can convert cholesterol into testosterone and use it in their central carbon metabolism (Preprint: (Ochsenkühn et al., 2023)). These studies generally demonstrate that *Endozoicomonas* is responsive to, and ready to uptake and use, coral metabolites to fuel its metabolism.

Though illuminating, *in vitro* and genomic metabolic predictions are limited in their applicability because they do not demonstrate the function of *Endozoicomonas* **within the coral host**. To more closely approximate *in hospite* function of *Endozoicomonas* we performed transcriptomic sequencing of *Endozoicomonas*-rich coral tissues. We find that *Endozoicomonas* expresses genes involved in aerobic heterotrophy, ROS detoxification, cell division, secretion systems, importing organic (ATP, vitamins, nucleosides, diverse sugars, TCA intermediates, amino acids, polyamines, heme, ...) and inorganic (ammonia, phosphorous, ...) compounds from its host milieu, all without clear synthesis of compounds hypothesized for coral benefit (e.g. vitamins).

The hypothesized *Endozoicomonas*-coral mutualism relies on the assumption that *Endozoicomonas* benefits the coral more than it harms it. In this study we identify some very real potential downsides for corals hosting this bacterium, and question if some of the hypothesized benefits materialize. This suggests parasitism. *In situ* transcriptomics is highly suggestive of realized metabolism, these results require further validation. That being said, the possibility of

Endozoicomonas-coral pariticism has been raised before (Pogoreutz & Ziegler, 2024), this study provides *in situ* evidence that *Endozoicomonas* may take more than it gives the coral holobiont.

Materials and Methods

The DNA and RNA samples used to interrogate *Endozoicomonas* abundance and function were collected *in situ* from two species of corals on a reef flat in Mo'orea, French Polynesia. These corals were collected at sub-daily intervals over multiple days from controlled replicates (DNA and RNA) and opportunistic sampling (DNA only). The DNA samples were used for 16S rRNA gene sequencing to characterize the coral microbiome, while RNA samples were used for bulk-tissue transcriptomic sequencing to identify the expressed metabolic intention in coral tissues. These samples reveals that *Endozoicomonas* makes up a large fraction of a coral species' microbiome and transcriptome across sites and time.

Controlled Coral Preparation and Sample Collection

We prepared a field site with pre-fragmented corals to ensure species representation and replication while sampling over short timescales (6-to-12-hour intervals). This enabled rapid sampling and RNA preservation in the field as is often performed by our group (Clements & Hay, 2019, 2021). To do this, a steel mesh platform (8x8m; **Figure 10B**) was established in water ~2 m deep, suspended 30 cm off the benthos, and ~350 m from shore (17°29'18.5"S 149°52'54.3"W; **Figure 10A**). Briefly, upturned bottle caps were affixed to the platform, and coral nubbins (described below) were epoxied into the sawed-off and upturned neck of a water bottle. After the epoxy hardened overnight, the bottle cap and neck were screwed together, temporarily affixing the corals to the platform allowing sampling without disturbing nearby

corals by unscrewing the caps (**Figure 10B**). To do this, on 14 August 2021, three replicate coral colonies of both *Porites rus* and *Pocillopora damicornis* were collected off the coast of Mo'orea, French Polynesia from the reef flat, inland from the reef crest in water ~2 m deep, all within 140 m of the experimental platform. Each colony was fragmented into 20 nubbins (each considered a clonal replicate), resulting in a total of 120 nubbins. On 15 August 2021 coral nubbins from both species were arranged in randomized configuration on the experimental platform and allowed to acclimate for 218 days before sampling in the end of March 2022.

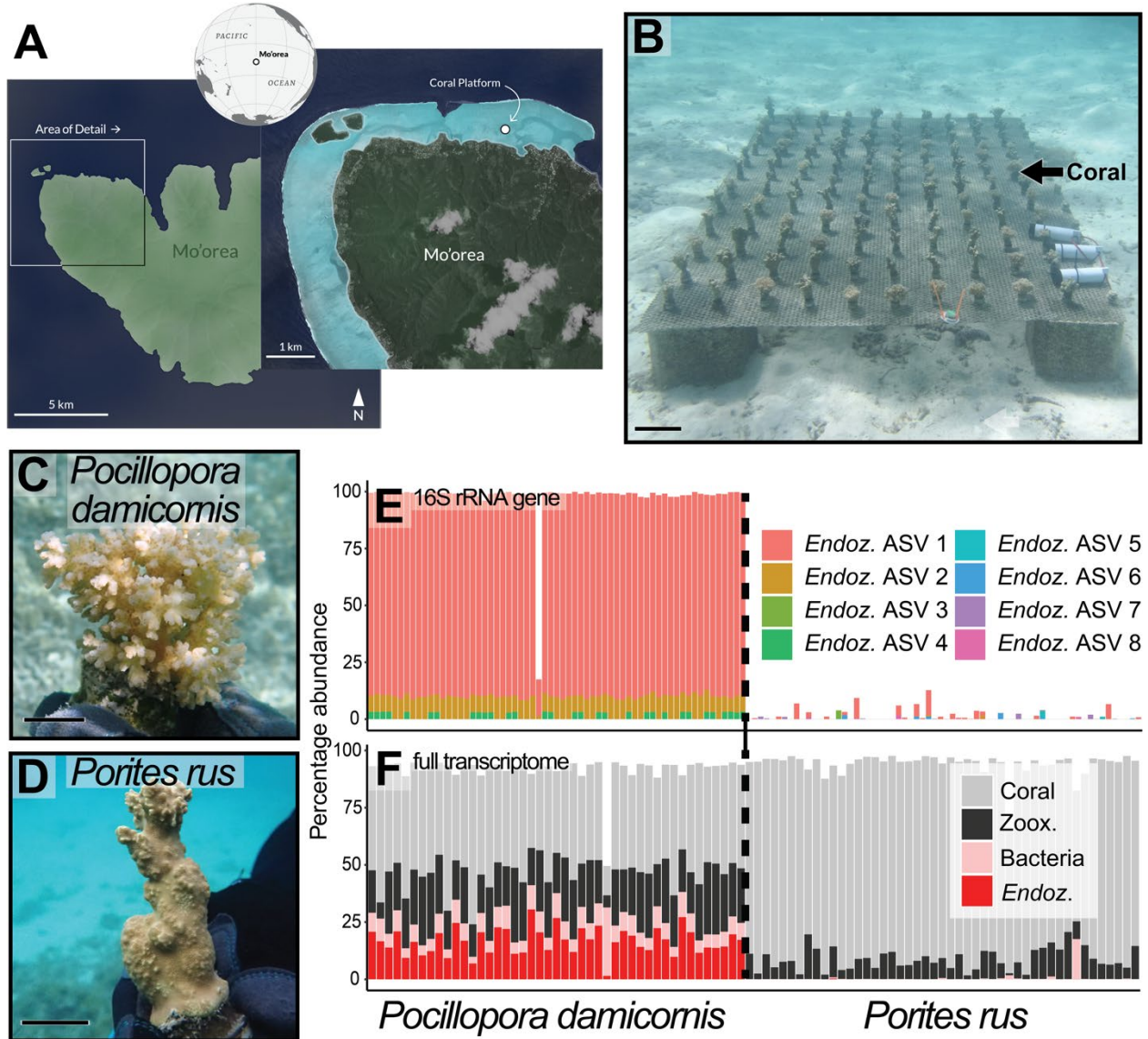


Figure 10: *Endozoicomonas* sp. dominates the community and transcriptome of one coral species on a reef in French Polynesia. Sampling location, design, and sample types included in reef-scale, time-resolved microbiome sampling. A) Location of coral sampling site off the north coast of Moorea, French Polynesia. Cartographic data sourced from Copernicus Sentinel 2025; Shom, 2023, MNT topo-bathymétrie côtier de Moorea (Projet Météo-France), and Natural Earth. B) Experimental platform and genotypic coral replicates. Arrows and labels indicate key features of sample collection. "Coral" identifies one coral fragment on the rack. Scale bar 10 cm. C-D) Representative individuals of corals [C) *Pocillopora damicornis*, D) *Porites rus*] collected in this study, scale bar 1 cm. Percentage abundance of E) *Endozoicomonas* sp. 16S rRNA V4 gene amplicon sequence variants (ASVs) across coral samples (only ASVs with maximum abundance >1.5% in one sample) and F) taxonomic composition of transcriptome (length normalized by transcripts per million) transcripts recovered from bulk-tissue RNAseq across coral samples.

ASV, amplicon sequence variant; *Endoz.*, *Endozoicomonas*; *Zoox.*, *Zooxanthellae*. Cartography: Max Borrmann, Photo credit: A. J. Maritan.

To comprehensively sample microbiome composition and function, corals were sampled over four consecutive days at six-hour intervals (beginning at 1800 h on 29 March 2022 and ending at 1200 h on 2 April 2022), resulting in 96 hours of repeated, uninterrupted sampling. Sample collection took a maximum of 30 minutes to complete, and all sampling occurred within one hour of the intended collection time (0001, 0600, 1200, and 1800 h $\pm$ 1 hr). At each sampling, one coral nubbin from each replicate colony was randomly sampled. Platform corals were sub-sectioned and processed using the same methods as the opportunistically sampled corals (see “coral field processing” below). After coral sub-sectioning, remaining coral was returned to the experimental platform and at the end of the experiment, coral nubbins were outplanted on the reef flat. During our sampling, peak daytime water temperatures at the site were $\sim 32^{\circ}\text{C}$, and the coral nubbins were approximately ~ 5 cm tall. On the experimental platform and on the surrounding reef flat, *P. damicornis* showed early signs of bleaching, while *P. rus* was not visibly bleaching.

Reef Coral Sample Collection

Following the controlled coral sampling, we sampled unmanipulated *P. rus* and *P. damicornis* colonies to assess if fragmentation and platform acclimation altered the microbiome. These corals were collected at 12-hour intervals (0001 h and 1200 h) over three consecutive days. These corals were sampled ~ 350 m from the experimental platform over an area of approximately $8,000\text{m}^2$, 50-80 m from shore in water ~ 1 m deep. In this collection coral colonies were not re-sampled, and species replicates were each $> 5\text{m}$ away from one another.

Field Processing

All sample collection was performed while wearing clean gardening gloves. Coral nubbins were sub-sampled by breaking off fragments inside a sterile plastic wire-closure sample bag (Fisher, Waltham, MA) or with tin snips (Tactix, Centennial, CO) into a sterile sample bag. Snips were rinsed thoroughly in ambient seawater between samplings. Coral sampling aimed to avoid epiphytes or endolithic algae while maximizing collected tissue. Coral tissues were preserved for transportation within 5 minutes of collection by either snap freezing on dry ice or immersion in RNA/DNA stabilizing buffer (25 mM sodium citrate, 10 mM EDTA and 70 g ammonium sulfate per 100 ml solution, pH 5.2) on ice (sample preservation method by time and day, **Table B2**).

Field Station Processing

Upon returning to the field station, coral fragments were preserved for molecular analyses by placing coral fragments (~1 cm³) into paired 2 mL cryovials (one each for DNA or RNA; Corning Inc.; Corning, NY) containing RNA/DNA stabilizing buffer and freezing at -40°C. Sample preservation was performed sequentially by sample and took less than 5 minutes per sample. All samples were in preservation buffer and frozen within two hours of sampling. Samples were shipped to Montana State University on 14 April 2022 on dry ice and stored at -80°C until DNA or RNA extraction.

Microbiome taxonomic analysis - DNA extraction, 16S rRNA gene amplification, sequencing, and analysis

We used DNA-based analyses to profile the coral prokaryotic community, characterizing the abundance of *Endozoicomonas* sp. across two sampled coral species.

DNA extraction: DNA was extracted from each sample in randomized order by transferring ~0.25 g coral (N=129 total samples – controlled and opportunistic sampling) into the bead beating tube of the PowerSoil Pro kit (Qiagen, Hilden, Germany) with flame sterilized forceps. Extractions were performed following the manufacturer's protocol utilizing 20 minutes of bead beating. Extracted DNA was quantified using a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA). Extraction blanks were performed (reagents only) for each extraction kit (n=2) and carried through the 16S rRNA gene amplification and sequencing.

16S rRNA gene amplification and sequencing: We used the 16S rRNA gene to profile the coral bacterial community. Before amplification, samples with DNA concentrations greater than 30 ng/ μ L were diluted to 10 ng/ μ L in PCR grade water. PCR amplification targeted the V4 region of the 16S rRNA gene using barcoded universal primers 515 F (5'-GTGYCAGCMGCCGCGGTAA-3')(Caporaso et al., 2011) and 806R (5'-GGACTACNVGGGTWTCTAAT-3')(Apprill et al., 2015), following the thermocycler conditions of 95°C for 3 minutes; 35 cycles of: 95°C for 30 seconds, 55°C for 30 seconds, 72°C for 30 seconds, and a final 72°C for 10 minutes [1 μ L template DNA, 0.5 μ L F primer, 0.5 μ L R primer, 0.25 μ L Bovine Serum Albumin, 10.25 μ L H₂O, 12.5 μ L Go Taq Master Mix (Promega, Madison, WI)]. Samples were then sent to the Georgia Genomics and Bioinformatics Core for dual index barcoding and sequencing. Samples were split across two sequencing runs each with extraction and storage blanks included on each run with 15 or 25% PhiX used to increase read diversity (Altman-Kurosaki et al., 2025) and sequenced on a MiSeq (Illumina Inc., San Diego, CA) using a V2 2x250 bp cycle kit.

Sequence Processing: QIIME2 (v2022.8; (Bolyen et al., 2019)) was used for amplicon analysis. Raw fastq files were imported, reads denoised, merged and filtered to remove chimeras using DADA2 (Callahan et al., 2016) with read trimming parameters `--p-trim-left-f 15 --p-trim-left-r 15 --p-trunc-len-f 165 --p-trunc-len-r 165`. Taxonomy was assigned to each Amplicon Sequence Variant (ASV) using the Silva 138 515-806 database (Quast et al., 2013). ASVs annotated as mitochondria, chloroplast, and known lab-specific contaminants (Class: Thermoleophilia; Family: Sphingomonadaceae, Caldalkalibacillaceae, Pirellulaceae, Micrococcaceae, Orbaceae, Corynebacteriaceae; Genera: Methylobacterium, Bacillus) were removed. To remove off-target eukaryotic amplification, all ASVs with a taxonomic designation of Domain “Bacteria” or “Archaea”, but with no other lineage detail were aligned against the NCBI nt database (Accessed December 2024; (Sayers et al., 2024)) targeting eukaryotes (taxid:2759) using BLASTn (Camacho et al., 2009). Any ASVs that aligned to a eukaryotic sequence were removed. A final cleaning step to remove contaminating sequences was performed using the decontam package (Davis et al., 2018) in R (R Core Team, 2023), where extraction blanks were used to estimate potential contaminants, and all identified contaminating taxa were removed from subsequent analyses (Genus: *Caldalkalibacillus*, *Meiothermus*, *Nesterenkonia*, *Halomonas*, *Bacillus*, *Escherichia-Shigella*, *Pseudomonas*). These cleaned data were used for all subsequent analyses.

Statistical Testing and Graphing: Analysis of 16S rRNA gene sequences sought to answer two questions: 1) what percentage of the bacterial community is *Endozoicomonas* sp.? 2) what is the sequence similarity of *Endozoicomonas* sp. ASVs to assembled rRNA transcripts (see “RNAseq” below). The first question was answered qualitatively by plotting the relative

abundances of ASVs across the two coral species. QIIME2 microbiome abundance and taxonomy matrices were imported into R (v4.3.1; [27]) using qiime2R (v0.99.6 [28]) and manipulated with phyloseq (v1.44.0; [25]) and tidyverse (v2.0.0; [29]). The second question was answered by performing an all-against-all sequence alignment of ASV sequences of interest against transcript sequences using BLASTn with percentage identity used to identify pairwise sequences with the highest sequence similarity.

Holobiont Taxonomic and Functional Analysis - RNA Extraction and Sequencing

To understand the activities of the different members of the coral holobiont (including coral, photosymbiotic dinoflagellates, and other eukaryotic and prokaryotic microbes – particularly *Endozoicomonas* sp.), we performed bulk tissue transcriptomics.

RNA Extraction and Sequencing: RNA was extracted from coral samples (N=93; only controlled coral sampling) in randomized order using the RNeasy PowerFecal Pro kit (Qiagen, Hilden, Germany) per the manufacturer protocol. To extract RNA, bead beating tubes were pre-filled with lysis buffer [Qiagen Reagent CD1 and phenol-chloroform-isoamyl alcohol (25:24:1; PC:IA)]. Then, up to 6 coral samples were thawed on ice, rice grain sized coral fragments were placed into lysis buffer using flame sterilized forceps, vortexed to homogenize, and placed back on ice until all samples were in lysis buffer. Samples were then vortexed at maximum speed for ten minutes, with all subsequent extraction steps per the manufacturer's protocol. RNA concentrations were quantified with a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA).

RNA processing and sequencing was performed at the Georgia Genomics and Bioinformatics Core. RNA was prepped for library construction through first removing

ribosomal RNA using a QIAseq FastSelect –rRNA Worm Kit (Qiagen, Hilden, Germany). Library construction used the KAPA Stranded RNA-Seq Library Preparation Kit (Roche, Basel, Switzerland) following manufacturer's instructions. rRNA depleted cDNA was pooled in equimolar concentrations and sequenced on a NextSeq 2000 using a P3 300 cycle reagent kit (Illumina Inc., San Diego, CA), yielding 308.6 gigabases (Gb; with an average of 3.35 Gb per library with a standard deviation of 0.74 Gb).

Holobiont taxonomic and functional analysis – transcriptome processing

To resolve *Endozoicomonas* expression in the coral holobiont, we performed *de-novo* transcriptome assembly to generate transcripts that were used for quantification, taxonomic classification, and functional annotation. This was selected because we lacked metagenome assembled genomes (MAGs) for genome-based read mapping. Our approach can be separated into two major parts: 1) ribosomal RNA (rRNA) depletion and taxonomic survey, 2) non-rRNA transcript assembly, quantification, taxonomic classification, and functional annotation (simplified workflow in **Figure 11**).

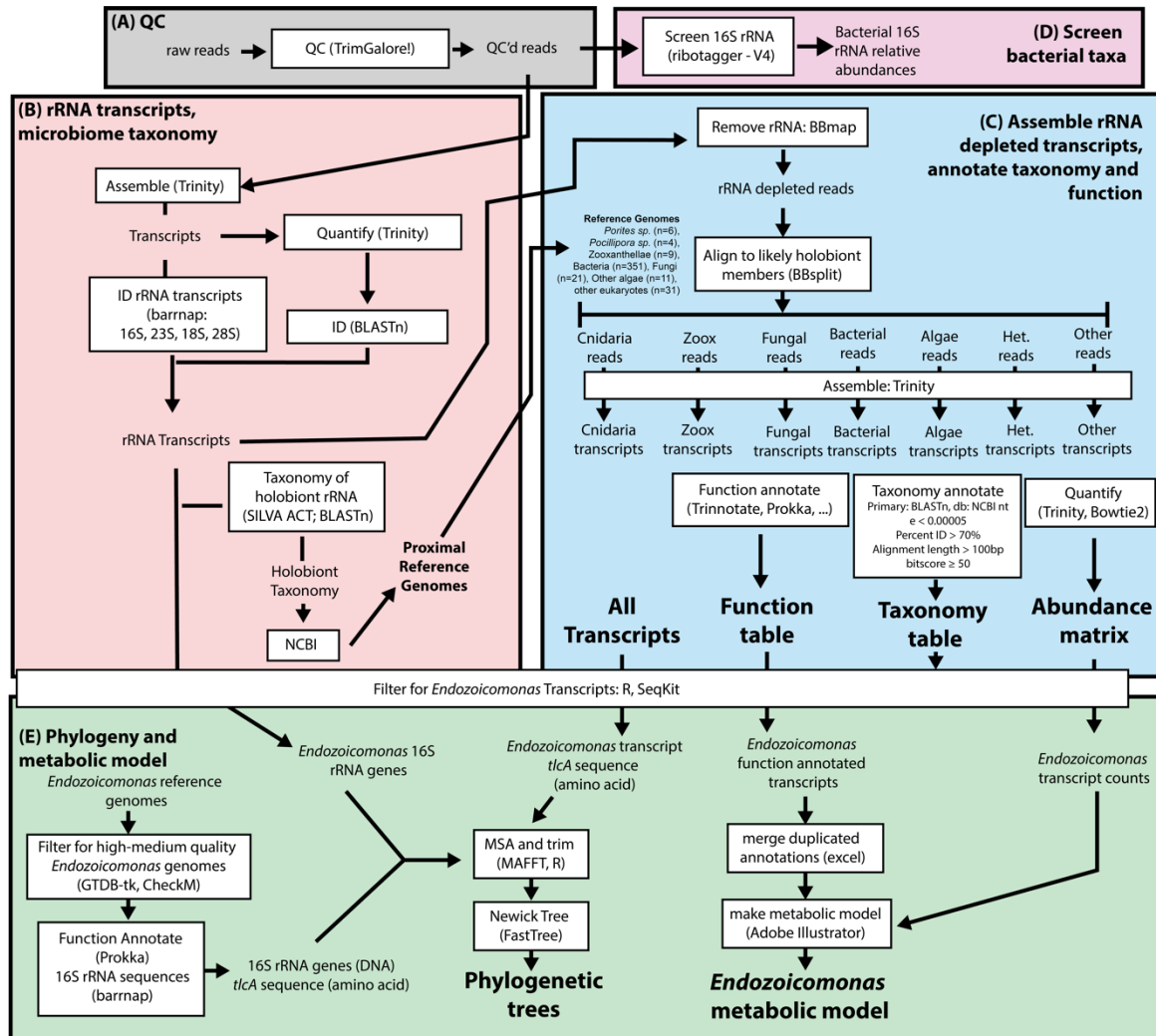


Figure 11: coral bulk tissue RNAseq taxonomically classified and functionally annotated *Endozoicomonas* transcripts. Simplified bioinformatic workflow of transcriptome analysis including major processing steps and software used. Some steps have been consolidated for legibility (especially sequential read mapping in B, see Methods for full details). A) Quality control (QC) of sequence data. B) assembly, identification, and taxonomic classification of ribosomal RNA (rRNA) transcripts. This step was performed to both survey the specific taxa present in the coral samples and identify rRNA sequences for later removal. C) RNAseq reads were depleted of rRNA, assigned rough taxonomy, assembled into transcripts, then taxonomically classified, functionally annotated, and quantified. D) RNAseq reads were used to survey the bacterial community. E) Transcripts were filtered to only include sequences classified to the bacterial genus *Endozoicomonas*. *Endozoicomonas* transcripts were used to construct contextualizing phylogenies of both 16S rRNA transcripts and genes of interest against sequences acquired from reference genomes, and to construct a metabolic model for *Endozoicomonas* activity in a coral host.

PART 1: rRNA depletion and Taxonomic Survey. This section had the sequential goals of first identifying rRNA transcripts (for use in removing rRNA reads from a second-pass analysis). Second, we taxonomically classified the identified rRNAs to identify taxa represented in the sequence data for genome-based read-mapping to assign high-level taxonomy.

To simplify transcript assembly, we first separated our reads into broad taxonomic classes based on our *a priori* knowledge of the coral holobiont composition (containing coral and zooxanthellae sequences). This was performed using a sequential read-mapping approach against reference genomes of representative zooxanthellae and all unmapped reads next to congeners of the corals in this study (per (Sorek et al., 2018); reference genomes listed in **Chapter 3 Table 3**). This separated reads into one of three categories: “coral”, “zooxanthellae”, and “other” (not mapped to any genomes; **Figure 11B**). These reads were then filtered by quality score (Phred >30) and length (>100 bp) using TrimGalore! (v0.6.6; (Martin, 2011)). These three quality-controlled buckets of reads were then co-assembled using Trinity ((Haas et al., 2013), v2.15.1) into transcripts for coral, zooxanthellae, and “other”. rRNA was then identified by Barrnap (<https://github.com/tseemann/barrnap>, v0.9), permissive of transcripts of a minimum of 10% of the full-length sequence (n=16,458). The transcripts identified as rRNA in each assembly were separated from all other transcripts and concatenated into a “rRNA file”. The rRNA depleted transcripts were then quantified by mapping the reads of a single sample (coral, zooxanthellae, and other) against its respective transcripts (e.g. coral reads vs. coral transcripts...) using Trinity as a wrapper for Bowtie2 ((Langmead & Salzberg, 2012), v2.2.5), samtools ((Danecek et al., 2021), v1.6), and RSEM ((B. Li & Dewey, 2011), v1.2.28). The top 20 most abundant transcripts for each taxonomic bucket (*P. rus*, *P. damicornis*, zooxanthellae, and other, n=80) were then

aligned against the NCBI nt (Accessed November 2024) using BLASTn, and any transcript that were identified as rRNA (n=75) in this approach was added to the list of rRNA transcripts.

These rRNA transcripts were then taxonomically classified to identify reference genomes for use in a second-pass (rRNA depleted) read-mapping taxonomic partitioning approach to give the “other” bucket finer taxonomic resolution. The rRNA transcripts were assigned taxonomy by first screening with ribotagger (v0.8.1; (Xie et al., 2016)) to identify the V4-7 hypervariable regions and then submission to the **SILVA ACT Server** for alignment against and taxonomy from the SILVA, RDP, GTDB, LTP, and EMBL-EBI databases (Accessed November 2024, minimum 95% percent identity). When taxonomy was assigned by multiple sources, the finest resolution available was selected. Still un-annotated rRNA sequences were aligned against the NCBI nt database (Accessed November 2024) with BLASTn. This census of the coral holobiont was used to identify and download related genomes from the NCBI. Two to five genomes were acquired for each taxon identified from rRNA, when available, to the finest taxonomic rank possible (list of accessions in **Table B4**). These genomes were clustered into six taxonomic groups: anthozoan (n=10 reference genomes), zooxanthellae (n=9), bacteria (n=351), fungi (n=21), non-zooxanthellae algae (n=11), other (non-coral nor algae) eukaryotes (n=31). Genomes within each of these groups were concatenated and used as databases for read mapping in the next section (see *non-rRNA transcript*).

PART 2: Non-rRNA Transcript Abundances, Functional Annotation, and Taxonomic Classification. This section is a second pass of quality control and assembly, informed by dataset-specific taxonomic composition of the coral holobiont. In this case, the reference

genomes used in read mapping include not only self-evident coral and zooxanthellae, but also bacteria, fungi, and other eukaryotes (**Figure 11C**).

To do this, raw sequence data was quality filtered by length (>100bp) and quality (Phred >25). These quality-controlled reads were rRNA depleted by alignment against our rRNA transcripts (from “Part 1”) using bbmap (<https://sourceforge.net/projects/bbmap/>, v38.90). Reads were deemed “non-rRNA” if at-least one read in the read pair did not align to the rRNA database. To assign domain or phylum-level taxonomy to all reads, rRNA-depleted reads were then simultaneously aligned to the six taxonomic databases (i.e. anthozoan, zooxanthellae, bacteria, fungi, non-zooxanthellae algae, and other eukaryotes) using bbsplit (<https://sourceforge.net/projects/bbmap/>, v38.90). This alignment created seven buckets of reads that included all above taxon categories and “unmapped” (percentages reads aligned in **Table B5**). These seven buckets of reads were co-assembled into transcripts and quantified with Trinity.

The assembled non-rRNA transcripts were then assigned a taxonomic lineage. The read mapping to taxon-specific databases provided each transcript a minimum domain/phylum-level taxonomy. As such, if the subsequent classification steps did not provide higher-resolution classification, the mapping-based annotation would stand. For higher resolution taxonomy, all transcripts were aligned against the NCBI nt database (Accessed December 2024) using BLASTn. The taxonomy of the database subject sequence was assigned to the query transcript only if the alignment had a percent ID $\geq 70\%$, alignment length ≥ 100 bp, bit-score ≥ 50 and an e-value $\leq 1 \times 10^{-5}$.

Function for each transcript was predicted using taxon-specific tools. Eukaryotic and unmapped transcripts had open reading frames (ORFs) identified by TransDecoder

(<https://github.com/TransDecoder/TransDecoder>, v2.6.3) and prokaryotes by Prodigal ``-meta`` ((Hyatt et al., 2010), 2.6.3). We used Trinotate ((Bryant et al., 2017), v5.7.1), eggNOG ((Cantalapiedra et al., 2021), v5.0.2), KofamScan ((Aramaki et al., 2020), v1.3.0), Prokka ((Seemann, 2014), v1.14.5), and DRAM ((Shaffer et al., 2020), v1.5.0) to assign putative functions. When a transcript was annotated by multiple tools, an annotation from a single tool was assigned (in order of preference: Trinotate, Prokka, DRAMv, DRAMP, KofamScan, eggNOG). In all cases, annotations were only accepted with an e-value $< 1 \times 10^{-5}$.

Endozoicomonas-specific Analysis

To examine the metabolic activity of *Endozoicomonas*, we focused on transcripts classified as genus *Endozoicomonas*. First, we estimated *Endozoicomonas* abundance in the prokaryotic community by identifying 16S rRNA transcript fragments in the raw reads and full length 16S rRNA transcripts from rRNA transcripts. Both fragments and transcripts were used to estimate abundance, while the full-length transcripts were phylogenetically contextualized by 16S rRNA sequences from reference genomes. Next, we determined *Endozoicomonas* contribution to holobiont gene expression by quantifying all transcripts and plotting percentage of transcriptome classified as *Endozoicomonas*. Finally, we used transcripts aligning to *Endozoicomonas* to reconstruct its metabolic activity in a coral host.

16S rRNA Phylogeny – Fragmentary and Assembled. To supplement the 16S rRNA gene-based abundance estimates, we identified bacterial 16S rRNA transcript V4 fragments using Ribotagger from the minimally processed quality filtered reads (no mapping performed). Using default taxonomy and count outputs, we plotted the percentage abundance of *Endozoicomonas* 16S transcript fragments using ggplot in R.

To look beyond fragmentary estimates of *Endozoicomonas* abundance we used assembled rRNA transcripts to plot normalized abundances and contextualize our sequences phylogenetically. To acquire longer 16S rRNA transcripts, we used the first-pass assembled rRNA transcripts (see above “*rRNA depletion and taxonomic survey*”) to enrich for rRNA reads for a more targeted assembly. We did this by aligning quality-controlled reads from each sample against the first-pass rRNA transcripts using Bowtie2. These sample-specific reads were then co-assembled into rRNA transcripts using Trinity and quantified by mapping sample-specific rRNA enriched reads against rRNA transcripts using Trinity. Longer 16S rRNA transcripts (>50% of the full 16S rRNA gene length) were identified with Barrnap. 16S rRNA transcripts were assigned taxonomy by alignment each against the SILVA 138 small subunit rRNA database with BLASTn. Any high-abundance transcripts (>30% of the sample) that were not annotated as genus *Endozoicomonas* were investigated further by individual alignment against the NCBI nt database (Accessed December 2024) using BLASTn. Any transcripts that were identified as *Endozoicomonas* in this step, they were reclassified as genus *Endozoicomonas* (Family: “Ectothiorhodospiraceae” and Genus: “Alcanivorax”). These *Endozoicomonas* transcripts were used to plot normalized percentage abundance, and for phylogenetic placement. Percentage abundance estimates for 16S rRNA transcripts were normalized by transcript length (alignment-based hits divided by transcript length; bp), then percentage abundance was calculated by dividing normalized abundance for each transcript by the total normalized abundance in the sample. Percentage abundance for *Endozoicomonas* was converted into a phyloseq object and plotted in ggplot2 in R.

Phylogenetic placement was performed by aligning rRNA transcripts against contextualizing reference 16S rRNA sequences. Reference *Endozoicomonas* 16S rRNA sequences were acquired from reference genomes (Accession number, taxonomy, completeness, and contamination in **Table B6**). Genomes were acquired using NCBI toolkit (<https://github.com/ncbi/sra-tools>, v3.1.1), retrieving all NCBI deposited genomes classified as Genus *Endozoicomonas* (n=48; as of January 2025). These genomes' taxonomy was checked by GTDB-tk (v2.1.1; (Chaumeil et al., 2020)) were screened to only include high-to-medium quality genomes (estimated completeness >50%, estimated contamination <10%; (Bowers et al., 2017)) using checkM (v1.1.3; (Parks et al., 2015)), resulting in 43 *Endozoicomonas* genomes. 16S rRNA genes in were identified by barrnap (n=39 genomes). Outgroup 16S rRNA sequences were acquired from the NCBI SRA. A multiple sequence alignment (MSA) of the 16S rRNA sequences was performed by MAFFT using default parameters ((Katoh, 2002), v7.505). The alignment was trimmed and filtered using the R (v4.3.1; (R Core Team, 2023)) package Biostrings ((Pagès et al., 2024), v3.2.0). Trimming prioritized retaining the rRNA transcripts, subsequently, all sequences with $\geq 30\%$ gaps was removed. In this trimmed alignment, all sites with gaps were removed, leaving a trimmed 492 bp MSA. A Maximum Likelihood tree was constructed from the MSA using FastTree (Price et al., 2010, v2.1.11) with default parameters with a generalized time-reversible (GTR) model with gamma-distributed rate heterogeneity. Phylogeny was plotted in the R package ggtree ((Yu et al., 2017), v3.2.0).

Endozoicomonas Percentage of Holobiont Transcriptome: *Endozoicomonas*' contribution to total non-rRNA transcriptional activity in a coral host was assessed qualitatively by converting each transcript's length-normalized abundance (transcript per million; TPM) into a percentage of

the library. Transcript taxonomy was used to sum abundances within taxonomic groups (*Endozoicomonas*, bacteria, coral, and zooxanthellae) yielding percentage expression across all samples by taxonomy. Percentage transcription was plotted in R with ggplot.

Endozoicomonas Functional Inference: To infer the metabolic activity of *Endozoicomonas* in a coral host, functionally annotated, non-rRNA transcripts, classified as genus *Endozoicomonas*, and found in *P. damicornis* libraries were used in all subsequent analyses (n = 5,506 unique transcripts).

Our confidence in *Endozoicomonas* transcript classification is twofold. First, we applied stringent alignment cutoffs (ID $\geq$ 70%, alignment length $\geq$ 100 bp, bit-score $\geq$ 50 and an e-value $\leq 1 \times 10^{-5}$) for taxonomic classification. Second, the bacterial fraction of *P. damicornis* samples is estimated to be >80% *Endozoicomonas* from 16S rRNA gene and transcript data, meaning any bacterial transcript (even in the absence of confident taxonomic classification) is likely to be *Endozoicomonas*.

Normalized abundances of *Endozoicomonas* transcripts were generated by first filtering holobiont raw count matrices to only include *Endozoicomonas* transcripts. Next, counts were normalized by dividing the counts of each transcript by the total abundance of transcripts annotated as “DNA recombination protein RecA” (samples lacking *recA* transcripts were removed from subsequent analyses, retaining 29/45 samples). *recA* normalized counts were used to calculate TPM for all transcripts.

All *Endozoicomonas* transcript annotations were manually curated to merge identical annotations from different programs and identify major metabolic pathways expressed. To contextualize the prevalence of major functions identified in our *Endozoicomons* transcriptome

across *Endozoicomonas* genomes, we calculated the prevalence of these functions across our curated *Endozoicomonas* reference genomes (**Table B6**). Prevalence was determined by annotating reference genome with Prokka and manually searching for homologues or parallel functions.

Transcripts of particular interest were contextualized phylogenetically. For ADP,ATP antiporters (*tlcA*), homologues were identified in the transcriptome and reference genomes from Prokka annotation. For *Endozoicomonas* reference genomes, coral-associated *Aquarickettsiales*, and non-coral associated *Rickettsiales* outgroups (accessions in **Figure 14B**), the amino acid sequences were used to phylogenetically contextualize transcript-derived sequences. For the transcriptome sequences, the nucleic acid sequences for *tlcA* were translated into all possible frames, and the coding frame selected by aligning all frames to the reference genome-derived amino acid sequences for *tlcA* using BLASTp, retaining only the single aligning translated sequence. To construct the tree, an amino acid multiple sequence alignment (MSA) was constructed using MAFFT, the MSA was trimmed to retain all transcriptome derived amino acid sequences. Then, all reference sequences with >30% gaps were removed before a Maximum Likelihood tree was constructed from the MSA (297 residues) using FastTree with default parameters with a generalized time-reversible (GTR) model with gamma-distributed rate heterogeneity. Phylogeny was plotted with ggtree.

To summarize the major predicted metabolic functions of *Endozoicomonas* in a coral host, we constructed a population-scale *Endozoicomonas* metabolic model incorporating identified proteins and their abundance and prevalence in reference genomes. Furthermore,

community-scale holobiont interactions with *Endozoicomonas* were inferred using functional annotations from all non-*Endozoicomonas* transcripts.

Cartography

Mapping was performed in qGIS (QGIS Association, 2025), where the Mo’orea coastline was traced from Sentinel-2 Imagery (NIR, band 8; (Copernicus Data Space Ecosystem, 2025)), the base map (ocean, land, and graticules) was sourced from Natural Earth Data (Natural Earth, 2025), and elevation data acquired from Shom (Shom, 2023). Final image processing was performed in Adobe Illustrator and Photoshop (Adobe, San Jose, CA).

Results and Discussion

Corals collected *in situ*, on a reef in Mo’orea, French Polynesia, across multiple days, contained *Endozoicomonas* 16S rRNA genes and transcripts. Ribosomal profiling revealed that *Endozoicomonas* dominated the *Pocillipora damicornis* bacterial microbiome. Bulk-tissue transcriptomic sequencing of *P. damicornis* shows *Endozoicomonas* is transcriptionally active *in hospite*, contributed to upwards of 15% of coral transcriptional activity, and expresses genes indicative of parasitizing its coral host. Though these samples were collected over time, we could not detect robust temporal trends in *Endozoicomonas* gene expression, and thus time of expression is not explored here.

Endozoicomonas Dominates *Pocillipora damicornis*’ Community and Holobiont Transcriptome

Though both *Porites rus* and *P. damicornis* harbored *Endozoicomonas* sequences, *Endozoicomonas* dominated the *P. damicornis* microbiome. Average *Endozoicomonas* 16S rRNA

percentage abundances in *P. damicornis* ranged from 86.6-98.1% depending on the sequencing and analysis (86.6% in assembled 16S rRNA transcripts, 98.1% in 16S rRNA gene amplicons; **Figure 10E, Figure 12A-C**), while in *P. rus*, average percentage abundances were 1.0-3.6% (1.0% in assembled 16S rRNA transcripts, 3.6% in RNAseq 16S V4 reads). The drivers of differences between taxa are not entirely clear, but branching and sensitive Pocilloporid and Acroporid corals have been the coral genera consistently reported to have these notably high Endozoicomonad abundances (Bayer et al., 2013; Caughman et al., 2021).

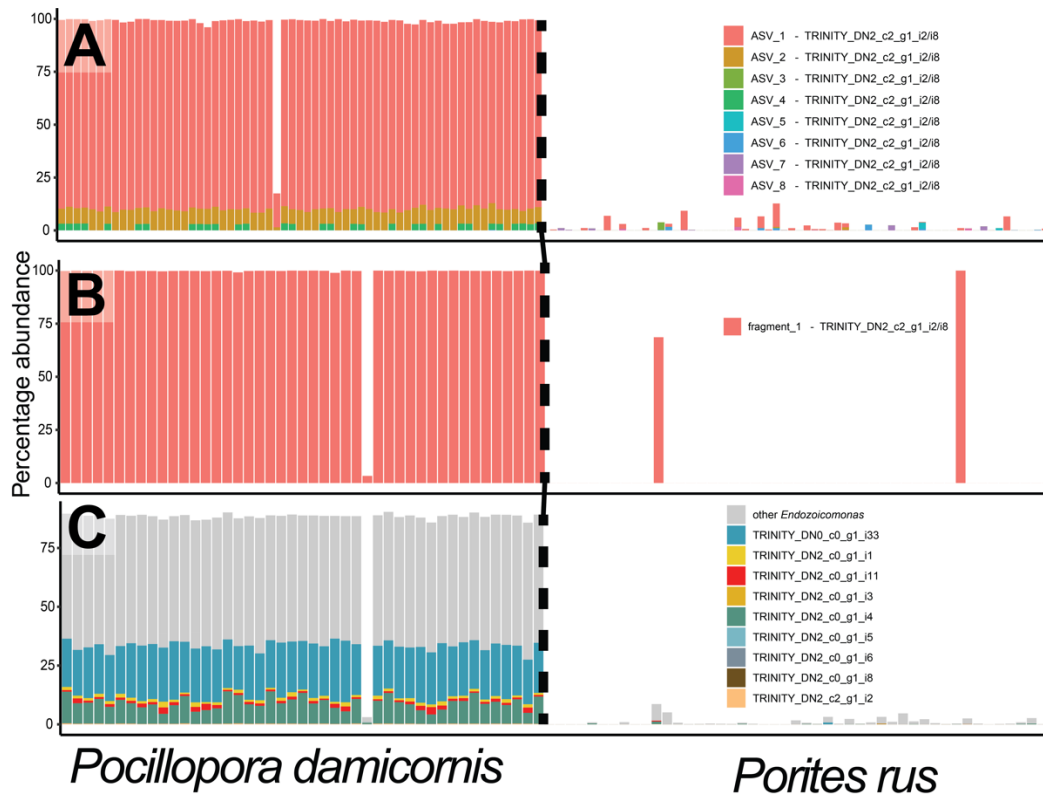
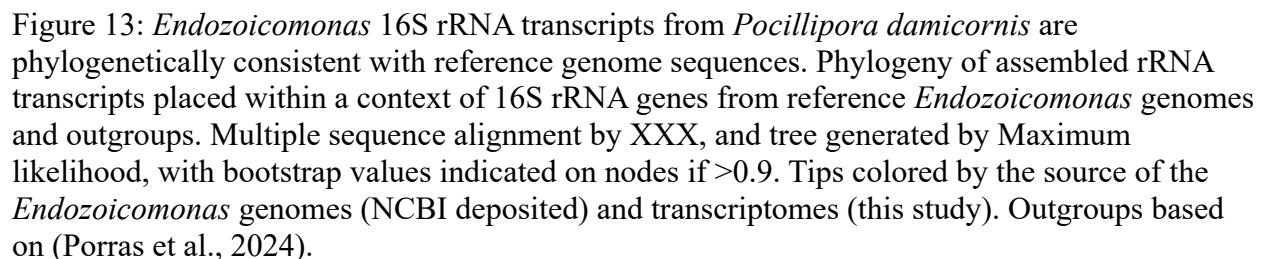


Figure 12: Multiple sequencing methodologies all recapitulate *Endozoicomonas* sp. dominance. E-G) Percentage abundance of prokaryotes in each coral as assessed with E) 16S rRNA V4 region gene amplicon sequencing (>1.5%), F) 16S rRNA V4 region transcript fragments (>0.5%), G) 16S rRNA assembled transcripts (percentage abundance of transcript counts normalized by transcript length), displaying only *Endozoicomonas* sp. sequences. Each column represents one sample, and the coral species from which they originate is indicated at the bottom of all three plots. 16S V4 rRNA gene amplicons and transcript fragments IDs are labeled with the 16S rRNA assembled transcript ID to which they most closely align (per **Table B1**).



Beyond ribosomal profiling, *Endozoicomonas* sp. represents a major fraction of the *P. damicornis* holobiont gene expression, with *Endozoicomonas* transcripts comprising an average of $16.8\% \pm 5.9$ (standard deviation) of all expression in *P. damicornis* samples and $68.6\% \pm 9.86$ of total bacterial expression (**Figure 10F**) and less than 1% in *P. rus*. Though corals are widely recognized as a fusion of metazoan and microbe (i.e. holobiont; (Peixoto et al., 2017; Rohwer et al., 2002)) the main players are generally coral and zooxanthellae, with other microbes playing important, but less central roles. In this study, we show that in *P. damicornis* bacterial gene expression rivals that of zooxanthellae, with *Endozoicomonas* contributing the most bacterial transcripts (**Figure 10F**). This is a particularly striking example of the coral holobiont in-action, where a coral's transcriptional activity equals that of its microbiome. Because *Endozoicomonas* dominated the *P. damicornis* microbiome and transcriptome, we focused our analysis only on *P. damicornis*-associated *Endozoicomonas* for all subsequent analyses.

Bulk-tissue *Endozoicomonas* Transcriptome Captures Major Metabolic Processes

After merging annotations from different sources, the final *Endozoicomonas* non-rRNA transcriptome contained 1,241 functionally unique transcripts. This transcriptome was thorough in its physiological scope. *Endozoicomonas* is metabolically active and replicating in its coral host, expressing tRNA ligases for 14 amino acids, DNA polymerases (e.g. *dnaX*) and structural cell division genes (e.g. *smc* and *ftsZ*; **Figure 14A**). *Endozoicomonas* also expresses flagellar genes (**Figure 14A**), in-line with electron micrographs of peritrichiously flagellated *Endozoicomonas* *in vitro* (Chiou et al., 2023).

Further, *Endozoicomonas* seems to be actively responding to cellular stressors, expressing genes for repairing protein (e.g. HSP70, *dnaJ*; HSP 100, *clpA*; (Rosic et al., 2011)) and DNA

damage (e.g. *mutL*; *uvrA*; *radA*) and resisting oxidative damage (e.g. *sodB*; *trxB*; *ahpF*).

Expression of these stress and damage response genes are among the highest in the transcriptome (**Figure 14A**). This corroborates the expectation that coral tissues would be stressful microhabitats. Corals (and their microbiome) are exposed to intense sunlight, requiring repairs to inflicted DNA damage, daytime heating (Jimenez et al., 2011), and oxygen supersaturation from oxygenic photosynthesis (Kühl et al., 1995).

Endozoicomonas Expresses both Aerobic and Anaerobic Metabolisms Within the Coral Host

For energy conservation, *Endozoicomonas* transcribed the entire oxygen-reducing electron transport chain (ETC; Complexes I-V; **Figure 15C**). *Endozoicomonas* expressed three distinct cytochrome oxidases: Cytochrome C *aa3*-type oxidase (*ctaD*), Cytochrome C *cbb3* oxidase (*ccoN*), and Cytochrome *bd* ubiquinol oxidase (*cydA,B*; **Figure 14A,6C**). Expression of higher oxygen affinity *cbb3*- and *bd*-ubiquinol type cytochromes (D'mello et al., 1996; Kalvelage et al., 2015; Kumagai et al., 2018) may indicate that some *Endozoicomonas* experience microaerophilic conditions and are using these enzymes to recover scarcer oxygen.

In line with potential oxygen limitation, *Endozoicomonas* expressed genes for anaerobic respiration of dimethyl sulfoxide (DMSO; *dmsA,B*) and nitrate (*napF*) as a terminal electron acceptors (**Figure 14A**). DMSO reductases (e.g. *dmsA,B*) reduce DMSO into dimethyl sulfide (DMS), generating proton motive force from NADH dehydrogenase proton translocation ((Bogachev et al., 1996); **Figure 15C**). In the ocean, DMSO is generally synthesized by algae directly (P. A. Lee & De Mora, 1999), or from the multi-step conversion of algal-derived DMSP to DMSO by bacteria or algae (Thume et al., 2018), or by bacterial oxidation of DMS (Hatton et al., 2012). Zooxanthellae production of DMSP (Jackson et al., 2020) and other microbial

transformation is the likely source of the reduced sulfur for *Endozoicomonas*' DMSO reductase (we did not recover genes for *Endozoicomonas*-mediated DMSP oxidation). In the *Endozoicomonas* transcriptome *dmsA,B* had had equal-to or higher expression levels than any of the cytochrome oxidases (**Table B7; Figure 14A**), and 62.5% of *Endozoicomonas* reference genomes encoded homologues for *dmsA,B*. *In vitro*, DMSO reductase synthesis is repressed with oxygen (Bogachev et al., 1996), indicating that some *Endozoicomonas* are performing anaerobic respiration in the coral holobiont, and this may be a widespread trait in coral-*Endozoicomonas* symbioses.

Though *napF* expression hints at denitrifying activity of *Endozoicomonas*, it alone is insufficient to argue for active denitrification in the coral host. This being said, *Endozoicomonas* reference genomes encode *napA,B,C,D,F,G* (prevalence of 19-73% across all genes) and future studies might examine *Endozoicomonas* denitrification in the coral host.

Collectively, the expression of high-affinity cytochromes and anaerobic DMSO reductases provides evidence that some fraction of the *Endozoicomonas* population may have limited oxygen access and adjusting their respiration accordingly. We did not observe a temporal trend to this process, which is surprising as surprising as corals tissue oxygen concentrations vary over the day with daytime supersaturation (Kühl et al., 1995). This population-scale simultaneous aerobic and anaerobic respiration may be due to micro-scale heterogeneity in coral tissues (per (Ricci et al., 2019)) or within an *Endozoicomonas* aggregate, in either case oxygen consumption may limit access deeper in the structure, as observed in biofilms (P. S. Stewart & Franklin, 2008).

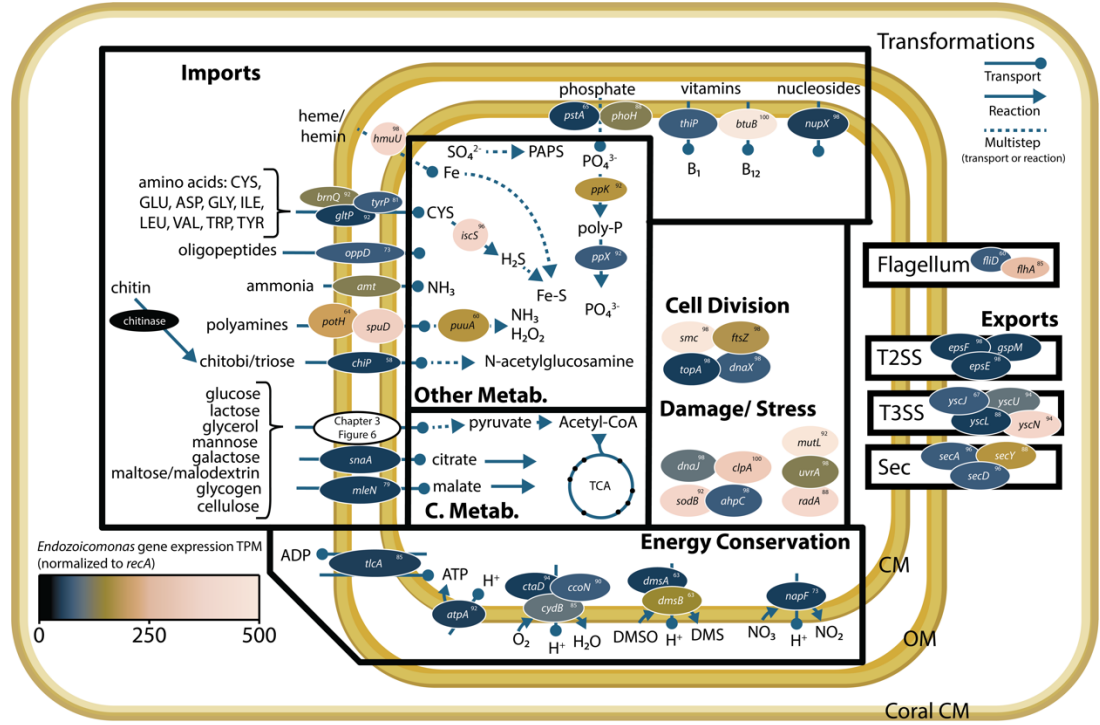
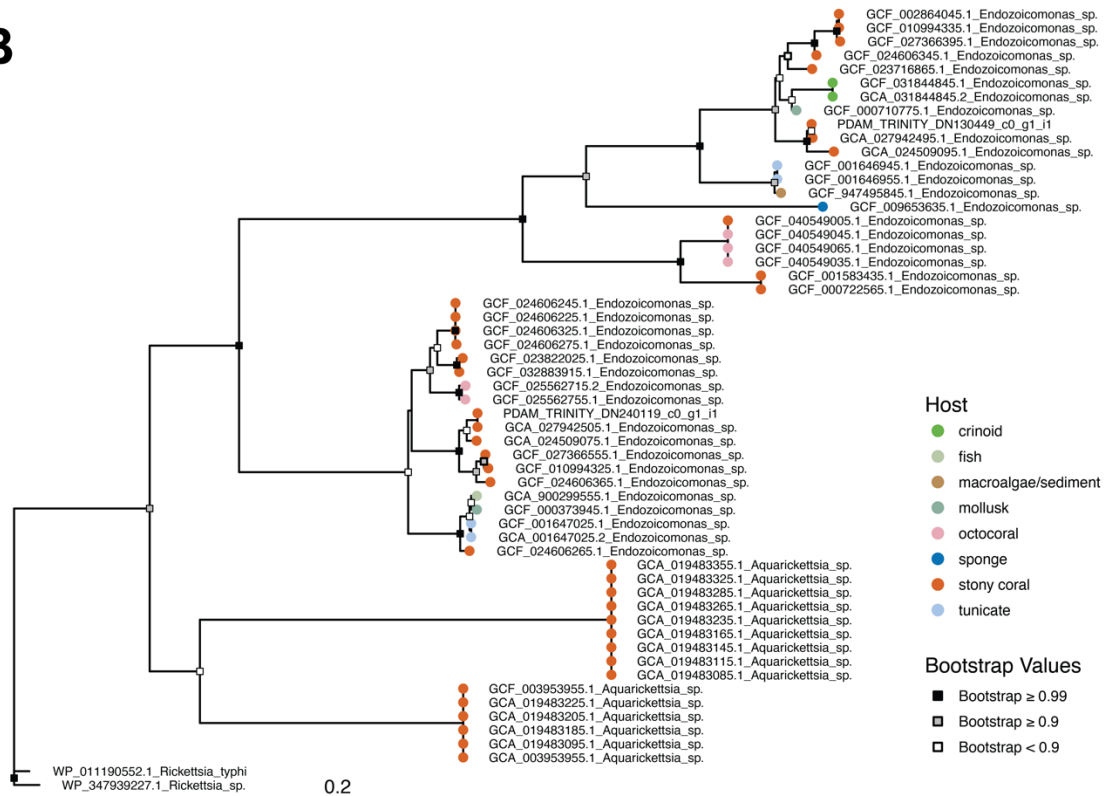
A**B**

Figure 14: Coral-hosted *Endozoicomonas* operates a heterotrophic metabolism supplied by metabolites acquired from the host environment. A) Metabolic model of genus *Endozoicomonas* reconstructed from bulk tissue transcriptomes from transcripts aligning (BLASTn, NCBI nt) to *Endozoicomonas* sequences. Key proteins are represented by ovals, with gene abbreviations overlaid, and colored by abundance (maximum transcripts per million, TPM). Black enzymes indicate a gene found in reference *Endozoicomonas* genomes, but not in transcriptome. Percent prevalence of transcript across # reference *Endozoicomonas* genomes indicated numerically in superscript if prevalence is greater than 49%. *ahpC*, Alkyl hydroperoxide reductase C; *amt*, Ammonia channel; *atpA*, ATP synthase alpha chain; *brnQ*, transport system for branched-chain amino acids; *btuB*, Vitamin B12 transporter BtuB; *ccoN*, *cbb3*-type Cytochrome C oxidase subunit; *chiP*, chitoporin; *ctaD*, *aa3*-type Cytochrome C oxidase subunit 1; *cydB*, Cytochrome *bd*-I ubiquinol oxidase subunit 2; *dmsB*, Anaerobic dimethyl sulfoxide reductase chain B; *dnaJ*, Chaperone protein DnaJ; *dnaX*, DNA polymerase III subunit tau; *epsE*, Type II secretion system protein E; *epsF*, Type II secretion system protein F; *flhA*, Flagellar biosynthesis protein; *fliD*, B-type flagellar hook-associated protein 2; *ftsZ*, Cell division protein FtsZ; *gltP*, Proton/glutamate-aspartate symporter; *gspM*, Putative type II secretion system protein M; *hmuU*, Hemin transport system permease protein HmuU; *iscS*, Cysteine desulfurase; *mleN*, Malate-2H(+)/Na(+)-lactate antiporter; *mutL*, DNA mismatch repair protein; *napF*, Ferredoxin-type protein NapF; *nupX*, concentrative nucleoside transporter; *oppD*, Oligopeptide transport ATP-binding protein OppD; *phoH*, phosphate starvation-inducible protein; *potH*, Putrescine transport system permease protein; *ppK*, Polyphosphate kinase; *ppX*, Exopolyphosphatase; *pstA*, phosphate transporter; *puuA*, Gamma-glutamylputrescine synthetase; *radA*, DNA repair protein RadA; *secA*, Protein translocase subunit SecA; *secD*, Protein translocase subunit SecD; *secY*, Protein translocase subunit SecY; *smc*, Chromosome partition protein Smc; *snaA*, Citrate transporter; *sodB*, Superoxide dismutase [Fe]; *spuD*, Putrescine-binding periplasmic protein; *thiP*, thiamine transport system permease protein; *tlcA*, ADP,ATP carrier protein 1; *topA*, DNA topoisomerase 1; *tyrP*, Tyrosine-specific transport protein; *uvrA*, UvrABC system protein A; *yscJ*, Yop proteins translocation lipoprotein J; *yscL*, Yop proteins translocation protein L; *yscN*, putative ATP synthase YscN; *yscU*, Yop proteins translocation protein U. For legibility, only key features and representative enzymes of *Endozoicomonas* metabolism and physiology are shown; the complete list of expressed genes in transcriptomic is available in Supplementary Tables X and X. Transportation or reactions without enzyme notation are represented in Figure 15A-E (including carbon metabolism, energy conservation, imports, and other metabolisms). B): Phylogeny of *Endozoicomonas* ADP,ATP antiporter recovered from bulk-tissue transcriptome contextualized by homologous sequences retrieved from reference genomes. CM, cell membrane; OM, outer membrane.

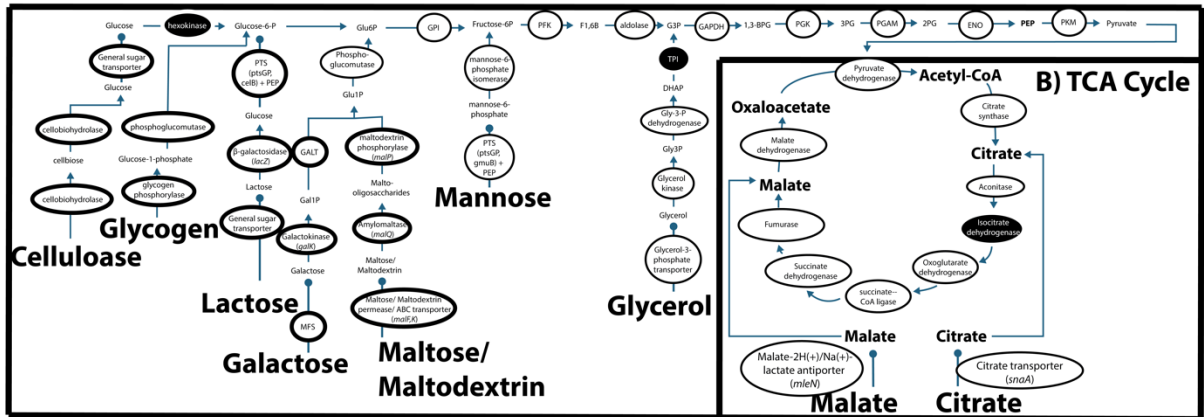
Endozoicomonas Steals Macromolecules and Nutrients from its Host

Endozoicomonas expression indicates a heterotopic metabolism, utilizing diverse complex molecules acquired from the coral host

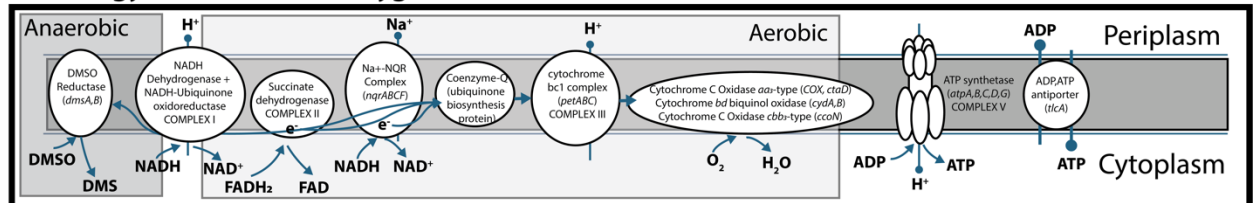
Central Carbon Metabolism

We recovered transcripts for nearly the complete tricarboxylic cycle (TCA; missing only isocitrate dehydrogenase; **Figure 15B**) while also expressing transporters for TCA-intermediates malate (*mleN*) and citrate (*snaA*). *Endozoicomonas* appears to be supplying its TCA cycle with diverse sugars, expressing near complete pathways for acquiring and utilizing seven different carbon sources (glycerol, glucose, lactose, maltose/ maltodextrin, glycogen, cellulose, and mannose; **Figure 14A,6A**) in glycolysis. As the dominant photosynthetic Zooxanthellae is the likely ultimate source of these sugars (Muscatine & Cernichiari, 1969), but the specific holobiont member that directly provisions *Endozoicomonas* is not clear from these data.

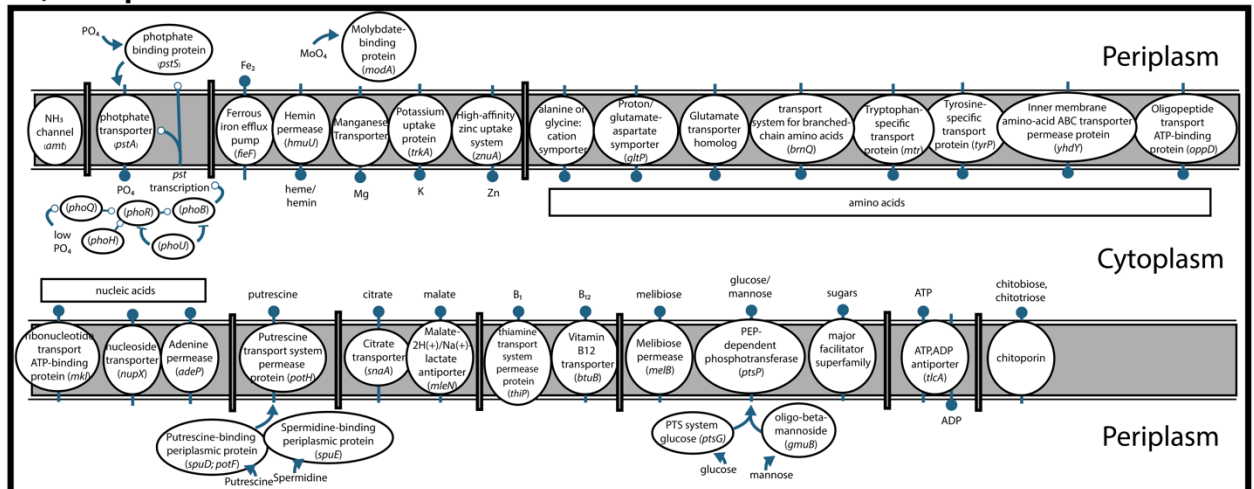
A) Glycolysis



C) Energy Conservation - Oxygen Reduction



D) Transporters



E) Additional Metabolisms

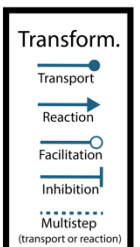
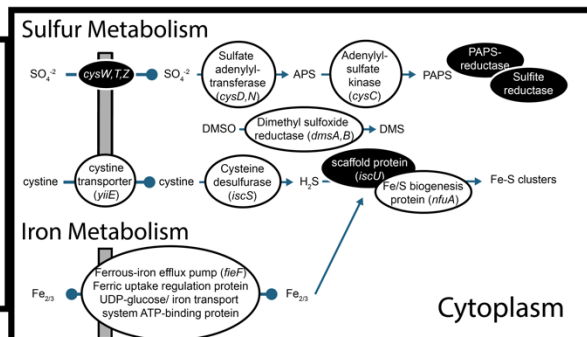
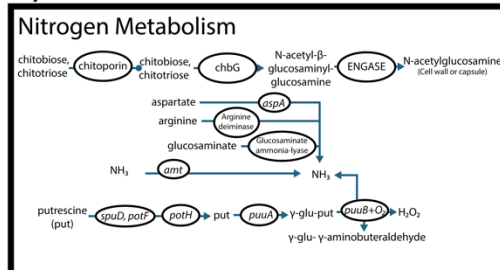


Figure 15: Extended metabolic model - *Endozoicomonas* expresses transporters to acquire and process diverse carbon, sulfur, nitrogen, sulfur sources. A) Glycolysis: *Endozoicomonas* expression indicates acquisition and integration of diverse sugars into their glycolytic pathway. LIST ABBREVIATIONS. B) Tricarboxylic acid cycle (TCA) cycle: *Endozoicomonas*' TCA cycle appears to be supplies by both glycolysis end-products and acquisition of TCA-intermediates from transporters. C) Energy Conservation - Oxygen Reduction: *Endozoicomoans* expresses the entire electron transport chain for utilizing oxygen as a terminal electron acceptor. Additionally, *Endozoicomonas* expresses ATP,ADP antiporters for acquiring ATP from the host environment. D) Transporters: *Endozoicomonas* expresses transporters for acquiring ATP, ammonia, phosphate, metals, amino acids, nucleotides, polyamines, TCA-intermediates, vitamins, sugars, chitin breakdown products. E) Additional Metabolisms: *Endozoicomonas* expresses genes for acquiring and metabolizing nitrogen, sulfur, and iron. APS, adenosine-5'-phosphosulfate; PAPS, 3'-Phosphoadenosine-5'-phosphosulfate; Tranform., transformation.

ATP Generation and Acquisition

Endozoicomonas can and does synthesize its own ATP (as demonstrated by expression of *atpA-D,G* presence in >90% of *Endozoicomonas* sp. reference genomes; **Figure 14A,6C**).

However, it appears that *Endozoicomonas* is also acquiring ATP from its host milieu.

Endozoicomonas expresses ADP,ATP antiporters (*tlcA*; **Figure 14A,6D**) that are ubiquitous across *Endozoicomonas* genomes (85.4%; **Figure 14A**). The *tlcA* transcript is monophyletic with *tlcA* sequences from *Endozoicomonas* reference genomes, and distinct from that of *tlcA* sequences derived from other ATP parasites from coral and non-coral contexts (**Figure 14B**).

ADP,ATP antiporters are diagnostic features of parasites, and have been found in other coral parasitizing microbes (e.g. *Ca. Aquarickettsia rohweri* (Klinges et al., 2019)). This suggests that *Endozoicomonas* may be actively exploiting its coral host, acquiring ATP in exchange for ADP.

DNA, Amino Acids, Vitamins, and Polyamines

Beyond ATP, *Endozoicomonas* expresses transporters for vitamins, nucleotides, amino acids, and polyamines.

Endozoicomonas has been reported to encode genes to synthesize vitamins B₁ and B₆, a potential source of B vitamins for a coral host (Maire et al., 2023; Pogoreutz et al., 2022), has been genomically classified a B₁₂ auxotroph (Preprint: (Lim et al., 2025)(Ding et al., 2016)). From our transcriptional evidence, *Endozoicomonas* appears to acquire vitamins from the holobiont, expressing transporters for vitamin B₁ (*thiP*; **Figure 14A**; (Chimento et al., 2003)) and B₁₂ (*btuB*; **Figure 14A**). We did not recover any transcripts for to B₆ synthesis and detected expression of two genes potentially involved in B₁ (thiamine) synthesis (*thiL* and *thiC*). These genes can be involved in biosynthesis (*thiL* and *thiC*; (X. Xu et al., 2024)) or salvage (*thiL*; (Kim et al., 2020)) of B₁. For *Endozoicomonas* to make successfully make B₁ requires *thiL*, *thiC*, *thiG*, and *thiE* (X. Xu et al., 2024). We detected expression of only two of these genes and none of the queried *Endozoicomonas* reference genomes encoded *thiE*, suggesting that *Endozoicomonas* is not synthesizing vitamin B₁. As such, expression of *thiL* may be involved in converting thiamine monophosphate acquired from import into the active form thiamine pyrophosphate (X. Xu et al., 2024). This finding runs counter to the hypothesized mutualist role of *Endozoicomonas*, where it appears that *Endozoicomonas* is instead performing the opposite function, acquiring multiple B vitamins from the host environment.

Endozoicomonas expresses nucleoside transporters, acquiring (e.g. ribonucleotide transport ATP-binding protein, *mkl* 81.2% reference genome prevalence; concentrative nucleoside transporter, *nupX*, 97.9%; and adenine permease, *adeP*, 75%; **Figure 14A,6D**) and salvaging (e.g. adenine deaminase, *add*, 96%; adenine phosphoribosyltransferase, *apt*, 100%) the building blocks of DNA and nitrogen from its host environment (Martins et al., 2024; Nygaard et

al., 1996). With nucleic acid-derivative exploitation characteristic of parasitic microbes (Major et al., 2017).

Endozoicomonas also acquires amino acids and polyamines as potential nitrogen, sulfur, and amino acids sources. For amino acids, *Endozoicomonas* expresses eight amino acid transporters (**Figure 14A,6D**). For cystine specifically, *Endozoicomonas* expresses cysteine desulfurase (*iscS*; **Figure 14A,6D**) that cleaves hydrogen sulfide off cystine and may be used to make iron sulfur clusters with Fe/S biogenesis protein NfuA (*nfuA*; **Figure 14A,6E**).

Endozoicomonas expression for the important of polyamine putrescine and conversion to ammonia are among its highest expressed genes (**Figure 14A,6E**).

The *Endozoicomonas* transcriptome also expresses transporters and metabolic processes for degrading chitin breakdown products (chitopodin, *chiP*; chitoooligosaccharide deacetylase, *chbG*; beta-N-acetylglucosaminidase, ENGASE; **Figure 14A,6D**). *Endozoicomonas* does not seem to be the microbe producing expensive extracellular enzymes to break down chitin, but rather playing the well-established role of scavenger of other organism's efforts (Datta et al., 2016). The primary degrader of chitin appears to be **the coral host**, based on the recovery of chitinase transcripts taxonomically classified to Scleractinia.

Inorganic Compounds

Endozoicomonas expresses multiple other transporters for acquiring essential elements from the holobiont. Within corals, its transcripts encode heme/hemin transporters (*hmuU*; **Figure 14A,6D**) for iron acquisition, and **pho/ pst** phosphate transporter systems (*pstA,S* and *phoB,H,Q,R,U*; **Figure 14A,6D**) for the regulation and execution of phosphate acquisition (Qi et al., 1997). Furthermore, *Endozoicomonas* expresses genes for polyphosphate formation

(phosphate kinase, *ppK*) and degradation (pholyphosphotase, *ppX*) indicating potential storage and retrieval of phosphorus (Srivastava et al., 2022).

From the expression of importers of ATP, vitamins, nucleosides, amino acids, multiple sugars, chitin derivatives, iron, and phosphate, it appears that *Endozoicomonas* is actively exploiting the biosynthetic products of the holobiont. This is in-line with an *in vitro* observation that *Endozoicomonas* increases expression of transporters and protein concentrations for glycolytic-associated enzymes after contact with crude coral extracts (Pogoreutz et al., 2022). This is also in-line with observed activity of closely related endosymbionts in clams. A member of the shared family Endozoicomonaceae (*Ca. Endonuclobacter*) has recently been demonstrated to invade their clam host and acquire host metabolites by expressing diverse transporters, all while subverting the host's apoptotic response (Porrás et al., 2024). It appears that in this case, *Endozoicomonas* too is exploiting its coral host.

For all these major metabolisms, there is no clear temporal expression pattern. The nature of bulk tissue sampling and de-novo transcriptomics may have homogenized distinct populations of *Endozoicomonas* into an integrated whole that has lost nuances of fine spatial and genomic scale. For instance, the simultaneous expression of high affinity cytochrome *cbb₃* oxidase and “normoxic” cytochrome *aa₃* oxidase indicates that different cells might be experiencing oxygen scarcity and plenty simultaneously. Other studies of fine scale physical properties of corals have shown that the oxygen concentration in coral tissues is variable over space and time (Kühl et al., 1995), and this micro-scale variation is likely experienced and responded-to by the coral microbiome. For this reason, these results may be considered a comprehensive survey of

Endozoicomonas activity, but not appropriate for disentangling micron or strain-scale variation in *Endozoicomonad* function.

Defining parasitism is hard to pin down, as it requires the harm against a host to outweigh any marginal benefit from the relationship. It should be noted that this study only examines metabolic intent through gene expression and does not prove physiological reality; as such, *Endozoicomonas*' potential parasitic behavior should be validated with controlled experiments in live corals. With that caveat, *Endozoicomonas* shows limited metabolic intention that benefit the coral host, while expressing exploitative metabolisms. As previously suggested (Pogoreutz & Ziegler, 2024), *Endozoicomonas* may opportunistically parasitize its coral host.

Conclusions

Endozoicomonas is a dominant member of some coral microbiomes, and in this dominant position, *Endozoicomonas* expresses genes that demonstrate a heterotrophic and aerobic lifestyle. *Endozoicomonas* may also be performing anaerobic respiration, but from this study, it is difficult to parse if this is a detoxifying response or energy generating. Furthermore, while in symbiosis with corals, *Endozoicomonas* expresses genes for exploiting the host milieu, to steal and utilize ATP, vitamins, nucleosides, amino acids, sugars and more. This transcription-based parasitic picture of *Endozoicomonas* flies in the face of the broadly promoted belief that *Endozoicomonas* is a “key mutualist” (reviewed by (Pogoreutz & Ziegler, 2024)). Most favorable discussions of *Endozoicomonas* arise from associative ecological observations and *in silico* genomic potential, and *in vitro* experimentation. Though *Endozoicomonas* possesses a relatively large and versatile genome, studies highlight its metabolic potential for biosynthesis of useful compounds and assuming that it might supply the host with these needed compounds. One favorite mutualist

hypothesis is that *Endozoicomonas* might synthesize and supply the host with B vitamins. In the coral tissues, *Endozoicomonas* did not express these B vitamin synthesis genes, instead expressed vitamin B₁ and B₁₂ transporters for acquiring these compounds from the host. *Endozoicomonas* expressed other genes for host exploitation. *Endozoicomonas* scavenges multiple key molecules from the host, notably ATP. Other coral-hosted ATP scavengers are parasites, and are believed to cause chronic stress to their coral hosts (Klinges et al., 2019), raising the same prospect for *Endozoicomonas*. *Endozoicomonas* expressed transporters for acquire yet other host metabolites, a trait identified in *Endozoicomonas* genomes (Maire et al., 2023; Pogoreutz et al., 2022; Wada et al., 2022).

Coral-associated bacteria have the potential to reduce host mortality. *Endozoicomonas* is a fixture in coral microbiomes due to its tendency to dominate coral microbiomes. This study provides a view into the metabolic activity of *Endozoicomonas* that shows this prominent microbe is expressing exploitative genes. This contradicts the dominant view of this bacterium as a mutualist and suggests a reassessment of the potential chronic harms of coral-association with this bacterium.

CHAPTER FOUR

SEA CUCUMBER GRAZING LINKED TO ENRICHMENT OF
ANAEROBIC MICROBIAL METABOLISMS IN CORAL REEF
SEDIMENTS

Contribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: Andrew J. Maritan

Contributions: formulated guiding hypotheses, performed molecular microbiology analysis,
devised bioinformatic strategy, performed bioinformatics, wrote manuscript

Co-Author: Cody S. Clements

Contributions: formulated guiding hypotheses, designed field experiments, collected samples

Co-Author: Zoe A. Pratte

Contributions: formulated guiding hypotheses, assisted in sample processing and bioinformatics

Co-Author: Mark E. Hay

Contributions: [formulated guiding hypotheses, designed field experiments, collected samples]

Co-Author: Frank J. Stewart

Contributions: formulated guiding hypotheses, advised on experimental and analytical strategies,
wrote manuscript

Manuscript Information

Andrew J. Maritan, Cody S. Clements, Zoe A. Pratte, Mark E. Hay, Frank J. Stewart

ISMEJ

Status of Manuscript:

- ☐ Prepared for submission to a peer-reviewed journal
- ☐ Officially submitted to a peer-reviewed journal
- ☒ Accepted by a peer-reviewed journal
- ☐ Published in a peer-reviewed journal

Oxford University Press

23 January 2025

Abstract

Sea cucumbers have been overharvested world-wide, making assessments of their ecological effects challenging, but recent research demonstrated that sea cucumbers increase coral survival via disease suppression and were therefore important for facilitating reef health. The mechanisms underpinning the sea cucumber-coral interaction therefore are not well understood but are likely mediated through sea cucumber grazing of microbes from reef sediments. We explored how sea cucumber grazing alters the sediment microbiome by leveraging a healthy sea cucumber population on a reef in French Polynesia. We used quantitative PCR, 16S rRNA gene sequencing, and shotgun metagenomics to compare the sediment microbiome in cages placed in situ with or without sea cucumbers. We hypothesized that grazing would lower microbial biomass, change sediment microbiome composition, and deplete sediment metagenomes of anaerobic metabolisms, likely due to aeration of the sediments. Sea cucumber grazing resulted in a 75% reduction in 16S rRNA gene abundances and reshaped microbiome composition, causing a significant decrease of cyanobacteria and other phototrophs relative to ungrazed sediments. Grazing also resulted in a depletion of genes associated with cyanotoxin synthesis, suggesting a potential link to coral health. In contrast to expectations, grazed sediment metagenomes were enriched with marker genes of diverse anaerobic or microaerophilic metabolisms, including those encoding high oxygen affinity cytochrome oxidases. This enrichment differs from patterns linked to other bioturbating invertebrates. We hypothesize that grazing enriches anaerobic processes in sediment microbiomes through removal of oxygen-producing autotrophs, fecal deposition of sea cucumber gut-associated anaerobes, or modification of sediment diffusibility. These results suggest that sea

cucumber harvesting influences biogeochemical processes in reef sediments, potentially mediating coral survival by altering the sediment microbiome and its production of coral-influencing metabolites.

Introduction

Corals are the foundation species of tropical reef ecosystems but have rapidly declined worldwide from diverse and compounding stressors including disease, pollution, overfishing, and ocean warming and acidification (Harborne et al., 2017; Hughes et al., 2018; Rasher & Hay, 2010). Given that many human-mediated stressors are predicted to increase in frequency and severity, there is increasing interest in leveraging biotic interactions to increase coral survival and provide more tractable and effective management options (Shaver & Silliman, 2017). For example, sustaining healthy and diverse communities of herbivorous fishes can positively affect corals by removing harmful seaweeds (Burkepile & Hay, 2008). However, the coral-seaweed-fish relationship can change from exerting positive to negative effects on corals depending on environmental conditions (Hixon, 2015) or the presence-absence of third-party consumers, e.g., sea stars that eat corals but cannot find corals under high seaweed cover (Clements & Hay, 2015). The diversity of reef communities suggests that coral health could be influenced by many other multi-species interactions that remain largely uncharacterized.

Sea cucumber-coral interactions have recently been identified as important for coral health and potential targets for coral reef management (Clements et al., 2024; Grayson et al., 2022). Sea cucumbers (Class Holothuroidea) are important cleaners of reef sediments by removing sediment organic material and microbes during feeding (Uthicke, 1999). These grazers were once commonplace on reefs but have been globally overharvested for decades to centuries,

with recent harvest estimates of more than 1 billion individuals per year (Purcell et al., 2023). This overharvesting may have serious overlooked implications for reefs if sea cucumbers influence coral health, either directly or indirectly. In some less populated areas of the world, sea cucumbers populations remain strong, either because they are inaccessible (Clements et al., 2024) or undesirable for human consumption (Purcell et al., 2023) and can be used in manipulative experiments to test the sea cucumber-coral relationship. Recent such experiments on reefs of Mo‘orea, French Polynesia and Palmyra Atoll revealed that whole colony mortality of reef-building acroporid corals increased up to 15-times in the absence versus presence of sea cucumbers, with tissue death consistently beginning at the sediment-coral interface and progressing up the coral branch (2024). When corals were separated from direct contact with the sediments (e.g., by having turf algae at their bases to act as a living barrier to sediment contact), colony tissue death decreased by 94-100%. This prior work also confirmed that sea cucumber grazing altered the taxonomic composition of sediment microbiomes and that coral-associated microbiomes differed between diseased and healthy corals (Clements et al., 2024). Together, these results link coral mortality to contact with sediments and identify sea cucumber grazing as a significant influence on coral mortality, potentially by affecting sediment microbiome composition and function.

Prior study of the sea cucumber-sediment-coral relationship did not identify a pathogen(s) linking coral disease to the altered sediment microbiome, raising the question: how exactly do cucumber-driven changes to the sediment microbiome influence coral health? We hypothesized that in the absence of sea cucumber grazing, sediment microbiomes negatively affect corals through diverse metabolic mechanisms. Such mechanisms may influence coral exposure to

anoxia (from the absence of mechanical aeration of sediments from sea cucumber grazing; (Uthicke, 1999)), acidity, nutrients, or harmful chemicals. The latter may include toxic products of metabolism, such as sulfide generated by sulfate reduction in sediment biofilms (M. Weber et al., 2012), or toxins from pathogens or benthic cyanobacteria (Ford et al., 2018). The effect of such products on corals, as well as the potential for sediment oxygen depletion, will presumably vary with microbial load, which is expected to be lower in sediments grazed by sea cucumbers due to microbial consumption (Uthicke, 1999) but was not measured in prior work. It is also possible that grazing removes coral-relevant microbes from sediment during passage through the sea cucumber GI tract or enriches coral-relevant microbes in sediment upon deposition of feces.

To explore potential mechanisms by which sea cucumber grazing affects sediment microbiomes and biogeochemical processes, we quantified 1) microbiome abundance, taxonomic composition, and metabolic potential in sediments exposed to or protected from sea cucumber grazing, and 2) microbiome composition dynamics during passage through the sea cucumber gut. We used cages with and without sea cucumbers to experimentally simulate sea cucumber presence and absence *in situ* on a reef of Mo'orea, French Polynesia. Cages without sea cucumbers simulated conditions following overharvest of sea cucumbers. Using this framework, we sampled ungrazed sediment (sed w/o sea cucumbers), grazed sediment (sed w/), and partially digested sediment from the sea cucumber GI tract. We used DNA from these samples to estimate prokaryotic microbial load via quantitative PCR (qPCR) counts of 16S rRNA genes and microbial taxonomic and functional gene composition via 16S rRNA gene and shotgun metagenomic sequencing. The results confirm that grazing significantly reshapes the biochemical

landscape of the sediment microbiome, notably in ways affecting the potential for oxygen utilization and cyanotoxin production.

Materials and Methods

Cage Experiments and Sample Collection

We examined sediment microbiomes from cages with and without *Holothuria atra* sea cucumbers and from areas external to the cages (Figure 16A,B,C). For cage experiments, we erected ten 50x50x12 cm (LWH) cages (1 cm mesh size) on 11 June 2021 on sandy surfaces at a depth of ~2 m in the north shore lagoon of Mo'orea, French Polynesia (-17.489340, -149.885769). This site supports a population of *H. atra* at an average density of 7.1/m² (Grayson et al., 2022). For five days, sea cucumbers were excluded from cages to allow sediment microbiomes to develop without sea cucumber grazing. On day six, a second smaller cage (12x12x12 cm) was inset in one corner of each larger cage, and one *H. atra* (9-14 cm length) was added to the larger portion of each cage and allowed to graze for 20.5-22 hours.

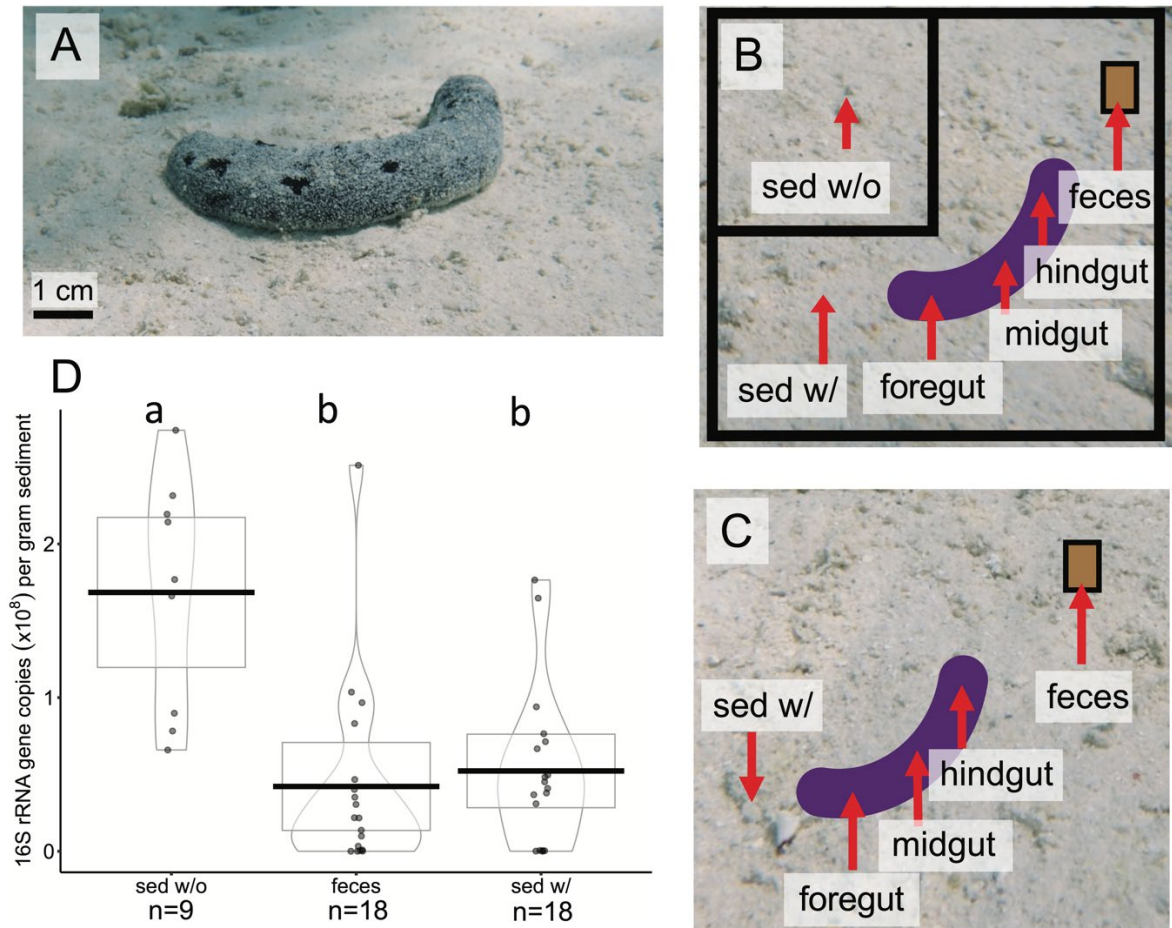


Figure 16: Sediment microbial load decreases in the presence of the sea cucumber *Holothuria atra*. **A)** Representative *H. atra* individual at the field site in Mo'orea, French Polynesia. **B, C)** Experimental design for testing effects of sea cucumber grazing on reef sediment microbiomes (diagram not to scale for clarity). **B)** Wire mesh cages (n=9; 50x50x12cm) with inset cages (12x12x12cm) were used to exclude grazing for 14 days and then to enclose a sea cucumber for 20.5-22 hours before collection of ungrazed sediment (sed w/o), grazed sediment (sed w/), and fecal samples. Foregut, midgut, and hindgut samples were also dissected from collected individuals. **C)** Free-roaming (outside cages) sea cucumbers (n=9) were collected concurrently with caged sea cucumbers and used for sampling of sed w/, foregut, midgut, hindgut, and fecal microbiomes to assess possible artifacts of caging. **D)** Quantitative PCR (qPCR) of 16S rRNA gene copies normalized to grams of sediment used in DNA extraction, presented as violin plots (mean as estimate of center; box bounded by 95% confidence interval; density estimator used to draw violin; observations as individual points). Microbiome samples from caged (B) versus free-ranging sea cucumbers (C) did not differ in qPCR microbial abundance ($P>0.811$; Appendix C Figure 4A) so those samples are pooled for this figure. Dissimilar letters indicate $P<0.0001$ in pairwise comparison between groups, while sharing of any letter indicates $P>0.05$ (P values were obtained from linear mixed effects model; Tukey-Kramer HSD). Photo credit: A. J. Maritan.

On 18 June 2021, we collected: i) sediment samples (surface 1-3 mm) from within the larger portion (grazed; in-front of the cucumbers' feeding tentacles) and inset portion (ungrazed) of each cage, ii) fresh *H. atra* fecal pellets proximal to the sea cucumber's anus, and iii) sediment dissected from the sea cucumber's foregut, midgut, and hindgut (GI tract) in the field (Fig 1B). An enclosed sea cucumber escaped from one of the cages before the experiment, reducing our grazed samples to n=9. On the same day we collected the samples from cages, we also collected samples external to, but nearby, the cages. These included sediments grazed by free-ranging (not caged) *H. atra*, fecal pellets proximal to the anus of free-ranging *H. atra*, and GI tract samples from nine free-ranging *H. atra* of similar size to those in the experimental population (Fig 1C). GI tract samples were collected at the same time as paired fecal and sediment samples. Individual sea cucumbers were brought to the water surface, sliced longitudinally, and GI contents were sampled from the foregut, midgut, and hindgut of each specimen. Implements used in dissection and collection were rinsed thoroughly in ambient seawater between samplings.

All sample collection was completed between 1130-1300 h, with samples immediately preserved in RNA/DNA stabilizing buffer (25 mM sodium citrate, 10 mM EDTA and 70 g ammonium sulfate per 100 ml solution, pH 5.2) on ice in the field, and then stored at -40°C after return to the lab at 1700 h. Samples were shipped to Montana State University on dry ice and stored at -80°C until processing.

DNA Extraction

We estimated the abundance, taxonomy, and function of the sediment microbiome using DNA-based analyses. DNA was extracted from each sample by transferring 0.18 – 0.33 g sediment/feces (N=99 total samples) into the bead beating tube of the Qiagen PowerSoil Pro kit

(Qiagen, Hidaigen, Germany) with flame sterilized forceps. Extractions were performed following the manufacturer's protocol, with extracted DNA quantified using a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA). An extraction blank (reagents only, no sample) was included in the workflow and used to evaluate contamination in the 16S rRNA gene amplification and sequencing process (below).

Microbial Abundance Estimation - 16S rRNA Gene qPCR

To quantify impacts of sea cucumbers on microbial abundance, we used qPCR to count 16S rRNA gene copies per gram of fecal (n=18), grazed sediment (n=18), and ungrazed sediment (n=9) samples. A standard curve relating 16S rRNA gene copy number to bacterial abundance was generated using DNA extracted from *Escherichia coli* K12. *E. coli* was grown to log phase OD₆₀₀ of 0.6 at 4.2×10^8 cells / mL (Sezonov et al., 2007), with each cell containing seven copies of the 16S rRNA gene (Ellwood & Nomura, 1980). Serial 1:10 dilutions from this culture were used to construct a standard curve relating 16S rRNA gene copies to the critical threshold (CT) of amplification. Aliquots (2 mL) from each step in the dilution series were pelleted by centrifuging at 16,000 x g for 10 minutes. Supernatant was removed and cellular DNA extracted using the Qiagen Power Soil Pro with modifications. Briefly, 270 µL of lysis buffer and 270 µL of bead beating beads were added to the *E. coli* cell pellet and vortexed for 3 minutes. The full volume of the microcentrifuge tubes was transferred into the Qiagen bead beating tubes, an additional 530 µL of lysis buffer was added, and the tube was vortexed for 10 minutes. Subsequent extraction steps proceeded per manufacturer's protocols. *E. coli* DNA was stored at 4°C, and all qPCR assays were completed within a week of extraction.

qPCR was performed using SYBR Green chemistry on a QuantStudio™ 5 Real-Time PCE system (Applied Biosystems, Waltham, MA) using the same primers as in the 16S rRNA gene sequencing described below [515 F (5'-GTGYCAGCMGCCGCGGTAA-3'; (Caporaso et al., 2011)) and 806R (5'-GGACTACNVGGGTWTCTAAT-3'; (Apprill et al., 2015)); conditions in Supplementary Table 1]. All samples were run over two days with each sample run in triplicate and standards run in duplicate. Averaged technical replicate CT values for each sample were converted to copies of 16S rRNA genes based on the standard curve and then normalized to grams of sediment used in each sample's DNA extraction.

Statistical Testing and Graphing: Our qPCR analyses sought to address the question of how sea cucumber grazing affected sediment microbial abundance. To answer this question, differences in mean gene copy number across sample types was compared with a linear mixed effects model (LME) accounting for the clustering of samples associated with the same sea cucumber and the effect of caging (Fixed: sample type [feces, grazed sediment...]; Random: cage status [in, out], cucumber replicate [1-18]), multiple comparisons were adjusted using Tukey-Kramer HSD, as implemented in the R packages lme4 (v1.1.34; (Bates et al., 2015)) and emmeans (v1.8.8; (Lenth, 2023)). 16S rRNA gene copy number per gram of extracted sediment by sample type was visualized with violin plots as implemented in ggpirate (v0.1.2; (Braginsky, 2023)).

Microbiome Taxonomic Analysis - 16S rRNA Gene Amplicon Sequencing

We used 16S rRNA gene amplicon sequencing to assess the impact of sea cucumber grazing on microbiome diversity and composition in grazed and ungrazed sediments and in sea cucumber foregut, midgut, hindgut, and fecal samples.

16S rRNA Gene Sequencing: PCR amplification targeted the V4 region of the 16S rRNA gene using barcoded universal primers 515 F (5'-GTGYCAGCMGCCGCGGTAA-3'; (Caporaso et al., 2011)) and 806R (5'-GGACTACNVGGGTWTCTAAT-3';(Apprill et al., 2015)).

Amplicons were generated with the thermocycler protocol of (1) 95°C for 3 minutes, (2) 25 cycles of 95°C for 0:45 minutes, 55°C for 0:45 minutes, 72°C for 1:30 minutes, and (3) 72°C for 10 minutes (reaction components and volumes in Supplementary Table 1). Each amplicon was appended with unique 10 bp barcode identifiers for dual indexing (25 cycles), then were pooled in equimolar amounts and sequenced at Georgia Genomics and Bioinformatics Core on an Illumina MiSeq (Illumina Inc. San Diego, CA) using a paired-end 500 cycle kit with V2 chemistry, with 10% PhiX to increase read diversity.

Sequence Processing: QIIME2 (v2022.8; (Bolyen et al., 2019) was used for amplicon analysis. Raw fastq files were imported into QIIME2, and reads were denoised, merged and filtered to remove chimeras using DADA2 (Callahan et al., 2016) with read trimming parameters `--p-trim-left-f 20 --p-trim-left-r 20 --p-trunc-len-f 245 --p-trunc-len-r 245`` (quality control statistics in Supplementary Table 2). Taxonomy was assigned to each Amplicon Sequence Variant (ASV) using the Silva 138 515-806 database (Quast et al., 2013). ASVs annotated as mitochondria, chloroplast, and known lab-specific contaminants (Supplementary Table 3) were removed, which removed all the ASVs in the extraction blank. To normalize sequencing effort, we rarified to 16,490 reads per sample without replacement (Appendix C Figure 1). This filtering and rarefaction collectively resulted in the removal of 11 (including the extraction blank) out of 100 samples (99 sediment/feces samples plus the extraction blank). These cleaned and normalized data were used for all subsequent analyses.

Statistical Testing and Graphing: Analysis of 16S rRNA gene amplicon data addressed four questions: 1) How does sea cucumber grazing affect alpha diversity in sediment microbiomes?, 2) How does sea cucumber grazing affect beta diversity in sediment microbiomes?, 3) Are microbiome differences between grazed and ungrazed sediments linked to sediment microbiome change during passage through the sea cucumber GI tract?, 4) Which microbial taxa are enriched or depleted by sea cucumber grazing? To answer the first question, we generated alpha diversity metrics (Observed features, Faith's phylogenetic diversity, Shannon diversity) in QIIME2 and tested for differences across sample types using a linear mixed effects model (LME) accounting for clustering of samples associated with the same sea cucumber and the effect of caging as implemented in the R packages lme4 (v1.1.34; (Bates et al., 2015)) and emmeans (v1.8.8; (Lenth, 2023)). Alpha diversity was visualized with violin plots using ggpirate. To answer the second and third questions, we generated phylogenetic and abundance resolved distance matrices and tested clustering between samples with PERMANOVA and dispersion within samples using PERMDISP as implemented in QIIME2. Sample clustering was visualized with NMDS as implemented in the R package vegan (v2.6.4; (Oksanen et al., 2022)), and dispersion was visualized with violin plots using ggpirate. To answer the fourth question, taxa of interest were initially identified by qualitative examination and the percentage abundances of these taxa were center log ratio (CLR) transformed in the R package microbiome (v1.23.1; (McMurdie & Holmes, 2013)) to satisfy the LME assumptions of equal variance and normal distribution, and visualized with pirate plots. Differences in transformed percentage abundance were assessed with LME. To visualize changes to composition, percentage abundances were plotted by sample as a stacked barplot using the R package ggplot2 (v3.4.3; (Wickham, 2016)) or

using CLR values on violin plots generated in ggpirate. In this analysis, LME model structure consisted of Fixed: sample type (feces, grazed sediment...); Random: cage status (in, out), cucumber replicate (1-18), multiple comparisons were adjusted using Tukey-Kramer HSD. In all cases, QIIME2 microbiome abundance matrices were imported into R (v4.3.1; (R Core Team, 2023)) using qiime2R (v0.99.6 (Bisanz, 2018)) and manipulated with phyloseq (v1.44.0; (McMurdie & Holmes, 2013)) and tidyverse (v2.0.0; (Wickham et al., 2019)).

Microbiome Metabolic Potential – Shotgun Metagenomics

We used shotgun metagenomics to test how sea cucumber grazing affects the metabolic potential of the sediment microbiome. Both read-based (no assembly) and metagenome-assembled genome (MAG)-based approaches were used to compare ungrazed and grazed sediment metagenomes and to explore whether the grazed sediment microbiome resembles that of the fecal microbiome.

Metagenomic Sequencing and Quality Control: Twenty (20) samples yielding the highest DNA concentrations were selected for shotgun sequencing, representing ungrazed (n=6) and grazed (n=7) sediment and fecal (n=7) samples. Sequencing was done at the Harvard Bauer Core using 9 ng of DNA per sample. DNA was simultaneously fragmented and tagged with Illumina adapter sequences (tagmentation) using the Illumina DNA library preparation kit (Illumina Inc. San Diego, CA) creating ~350 bp inserts. Tagged inserts were enriched and 10 bp barcodes were appended with 12 cycles of amplification to allow multiplexing. After barcoding, libraries were purified with magnetic bead-based cleanup using Illumina Sample Purification Beads (Illumina Inc. San Diego, CA). Resulting libraries were assessed using a 4200 TapeStation (Agilent Technologies, Santa Clara, CA) and quantified by qPCR (Roche Sequencing, Indianapolis, IN).

Libraries were pooled and sequenced (paired 150 bp reads) on one lane of an Illumina NovaSeq S4. Raw fastq files for the 20 metagenomes were filtered by quality score (Phred >25) and length (>100 bp) with TrimGalore! (v0.6.6; (Martin, 2011)), resulting in a cleaned dataset of 761 Gigabases.

Read-based Analysis – K-mer Profiling: We used k-mer abundance patterns to agnostically identify differences in metagenomic composition among sample types. We used the program Simka (v1.5.1; (Benoit et al., 2016)) to construct Bray-Curtis distance matrices based on 21-mer abundances in subsamples of reads (14,429,400 reads; accommodating the sample with the shallowest sequencing depth) from each metagenome. These matrices were used for the graphing described below.

Read-based Analysis – 18S rRNA Gene Screening: We used the abundance of 18S rRNA genes in each metagenome to describe the abundance and diversity of microeukaryotes in grazed and ungrazed sediments. Reads matching the V4 region of 18S rRNA genes were identified and quantified using RiboTagger (v0.8.1; (Xie et al., 2016)) with the functions ribotagger.pl and biom.pl. Each matching sequence was assigned taxonomy by alignment against the NCBI nt database using BLASTn (Camacho et al., 2009) with assignment based on the annotation of the top matching reference sequence with at-least phylum-level resolution.

Read-based Analysis – Marker Gene Screening: To test the effect of sea cucumber grazing on microbial biogeochemical processes, we analyzed 59 marker genes of microbial metabolism. This database contained 39 genes used in prior studies by our group and others to screen for ecologically relevant metabolisms, including trace gas (e.g., methane) metabolism, dissimilatory sulfur and nitrogen metabolism, carbon fixation, photosynthesis, and aerobic

respiration (Bertagnolli et al., 2023; Dragone et al., 2022; Leung & Greening, 2020), and supplemented with an additional 20 marker genes for other biogeochemical or ecological strategies (e.g., sulfur reduction, secretion, flagellation). The latter 20 genes were obtained from UniProtKB (The UniProt Consortium, 2023); Supplementary Table 4 lists all genes in our database. Gene relative abundances were estimated by aligning reads from each unassembled metagenome to each marker gene protein database with DIAMOND BLASTx (v2.0.15; (Buchfink et al., 2015)), applying a query coverage cutoff of 80% and amino acid identity cutoff of 50%. We estimated the proportion of community members encoding each gene by dividing the number of hits for each gene of interest by the gene's average length in the database and again by the number of genomes (genome equivalents) in the metagenome as estimated in MicrobeCensus (v1.1.0 (Nayfach & Pollard, 2015)), resulting in values as reads per kilobase genome equivalent (RPKG).

To assess the distribution of genome equivalents across the metagenomes to detect statistical outliers, we divided genome equivalents by sequencing effort (Mb) for each sample, and then fit a linear regression of GenomeEquivPerMb by Sample Type. After accounting for sample type, our data violated the assumptions of normality and equal variance, with one sample ('06_in_wo_sediment'; ungrazed sediment) exhibiting a Cook's Distance of 0.955 and identifying this sample as an influential outlier that could skew downstream statistical tests. Removing this sample resulted in normally distributed residuals and equal variances, satisfying the assumptions of the model (before and after diagnostic plots in Appendix C Figure 2A,B). Further support for removing this sample came from the agnostic numerical classification of all samples based on the 21-mer frequencies, which showed this sample clustering separately from

all other samples (NMDS and hierarchical clustering; Appendix C Figure 3A,B). For these reasons, sample `06\_in\_wo\_sediment` was excluded from all subsequent metagenomic analyses.

Metagenome Assembled Genome (MAG) Analysis: We used MAGs to assess the abundance of cyanobacteria in grazed and ungrazed sediments, requiring contig assembly, contig binning into MAGs, dereplication, and read mapping to estimate abundances. For contig assembly, reads were first normalized using BBnorm (<https://sourceforge.net/projects/bbmap/>), where kmer coverage was filtered to less than 100x and greater than 2x (“target=100 mindepth=2”). Metagenomes were individually assembled and then co-assembled into contigs using MEGAHIT (v1.2.9; (D. Li et al., 2015)) with “meta-large” presets. Contigs were binned into MAGs using the MetaWRAP (v1.3.2; (Uritskiy et al., 2018)) platform to map reads to the contigs using the BWA aligner (v0.7.17-1088; (H. Li & Durbin, 2009)). MetaWRAP was then used to generate MAGs with the binning programs MetaBAT 2 (v2.2.15; (D. D. Kang et al., 2019)) and MaxBin 2 (v2.2.7; (Wu et al., 2016)) utilizing contigs >2,500 bp. MetaWRAP then identified the best quality MAGs from the two binning strategies based on contamination and completion estimates from CheckM (v1.1.3; (Parks et al., 2015)) and the ‘bin_refinement’ module. Only MAGs with completeness >75% and contamination <10% (consistent with medium to high quality genomes; (Bowers et al., 2017)) were retained for further analysis. MAGs were then dereplicated at 99% average nucleotide identity using dRep (v3.4.3; (Olm et al., 2017)) resulting in 154 high-to-medium quality dereplicated MAGs (Bowers et al., 2017). MAGs were assigned taxonomy using GTDB-Tk (v2.1.1; (Chaumeil et al., 2020)), and relative abundance was estimated by mapping reads back to MAGs using BWA-mem in CoverM genome (v0.6.1; <https://github.com/wwood/CoverM>).

We focused our MAG-based analyses on cyanobacteria because they appeared to be markedly depleted after sea cucumber grazing in the 16S rRNA gene analyses. Only MAGs identified as cyanobacteria (n=18; binning stats in Supplementary Table 5) were functionally annotated. We identified open reading frames (ORFs) with Prodigal (Hyatt et al., 2010) and annotated these genes using METABOLIC-G (v4.0; (Zhou et al., 2022); with criteria specified in Supplementary Table 4). We also screened these MAGs for the presence of cyanotoxin genes to assess the potential for cyanotoxin production. To do so, ORFs from each MAG (amino acid translation) were aligned against custom cyanotoxin biosynthesis protein databases (Downloaded from UniProtKB and listed in Supplementary Table 4) using DIAMOND BLASTp in “ultra-sensitive” mode (v2.0.15; (Buchfink et al., 2015)), applying a query coverage cutoff of >80%, an amino acid identity cutoff of >40% and an e-value cutoff of $<10^{-50}$.

Statistical Testing and Graphing: Our metagenomic comparisons addressed three questions: 1) Does sea cucumber grazing influence the composition of sediment metagenomes inferred from k-mer frequency? 2) Does sea cucumber grazing influence the metabolic composition of sediment metagenomes inferred from marker genes? 3) Does sea cucumber grazing influence the abundance of phototrophs?

To answer the first question, we used hierarchical clustering and non-metric multidimensional scaling (NMDS) to qualitatively assess sample clustering based on 21-mer-based distance matrices generated in Simka (see above). Bray-Curtis distances were square root-transformed and clustered using Ward’s cluster algorithm via the *agnes* function in the R package *cluster* (v2.1.4; (Maechler et al., 2022)). Hierarchical clustering plots were generated with the base R package *stats*, and NMDS plots were generated with *vegan* (v2.6.4; (Oksanen et al.,

2022)). To answer the second question, differences in mean RPKG across sample types were assessed with LME while accounting for clustering due to sampling from the same sea cucumber and caged-uncaged status (Fixed: sample type [feces, grazed sediment...]; Random: cage status [in, out], cucumber replicate [1-18]), as implemented in the R packages lme4 (v1.1.34; (Bates et al., 2015)) and emmeans (v1.8.8; (Lenth, 2023)). RPKG for each gene of interest was plotted with violin plots using ggpirate. To answer the third question, counts of 18S rRNA reads were normalized by dividing count values by the size of respective metagenome libraries, resulting in values as reads per megabase sequenced (RPM). MAG relative abundances were center log ratio (CLR)-transformed in the R package microbiome (v1.23.1; (McMurdie & Holmes, 2013)) to satisfy the LME assumptions of equal variance and normal distribution. Differences in mean abundance for both metrics (18S RPM and MAG abundance) across sample types were assessed with LME (Fixed: sample type [feces, grazed sediment...], phototroph taxa [e.g. diatom, dinoflagellate, Xenococcaceae...], sample type * phototroph taxa; Random: cage status [in, out], cucumber replicate [1-18]), with abundance estimates visualized with violin plots using ggpirate.

Results and Discussion

Because prior research found that coral mortality increased following sea cucumber removal and was exacerbated by direct contact with sediment (Clements et al., 2024), we hypothesized that sea cucumber grazing restructured the sediment microbiome in a manner that lessened coral susceptibility to disease or other agents of mortality. We predicted that, in comparison to ungrazed sediment, grazed sediment would have: 1) lower microbial load, 2) altered microbial community composition, and 3) relative depletion of anaerobic metabolisms in sediment metagenomes due to mechanical aeration from sea cucumber grazing. Our results

confirm predictions #1 and #2 but show that grazing is linked to sediment metabolic potential in an unexpected pattern. Compared to ungrazed sediments, sea cucumbers were associated with enrichment of anaerobic metabolisms and depletion of genes for cyanotoxin production. We discuss these results below.

Sea Cucumber Grazing Depletes Sediment Microbial Abundance

The presence of *H. atra* reduced prokaryotic microbial load in coral reef sediment. The mean abundance of 16S rRNA genes in grazed sediments was ~75% lower than in ungrazed sediments, as estimated by qPCR (Figure 16D; LME, $P < 0.0001$). Sea cucumber feces were similarly depleted in 16S rRNA gene copies, suggesting that cells are removed as sediment passes through the sea cucumber GI tract and that cell abundance does not immediately rebound after fecal deposition. This pattern was recapitulated in normalized extracted DNA, suggesting broader taxonomic depletion (Appendix C Figure 4B). It is also possible that the physical churning of sediments by sea cucumbers allows for removal of microbes by water flow, facilitates grazing by other consumers, or alters the environment in a manner that suppresses microbial growth, although these mechanisms were not explored. We also note that variation in 16S rRNA gene copy number among prokaryotic taxa (Větrovský & Baldrian, 2013) could confound interpretation of qPCR data. However, for the purpose of this analysis, we assume that average gene copy number per cell is relatively consistent at the community level across samples, which is reasonable given that each sample is composed of hundreds or thousands of diverse taxa (Figure 17A). Despite this caveat, our results suggest that sea cucumber presence substantially reduces sediment microbial load, as is consistent with studies showing a decrease in

organics (S. Lee et al., 2018; Purcell et al., 2016) and microalgae (Uthicke, 1999) from sea cucumber feeding.

Sea cucumber Grazing Reshapes Sediment Microbiome Diversity

The presence of *H. atra* significantly influenced sediment microbiome diversity. Comparing 16S rRNA gene amplicons revealed significant differences in multiple beta diversity metrics in sediments with versus without sea cucumbers (Bray-Curtis and Weighted UniFrac; PERMANOVA $P < 0.05$; Appendix C Figure 5B,D), suggesting that sea cucumber grazing changes the taxonomic composition of sediment microbiomes. However, neither community dispersion (PERMDISP, Bray-Curtis: $p = 0.157$ and Weighted UniFrac $p = 0.254$; Appendix C Figure 5A,C) nor alpha diversity (LME, Observed features $p = 0.837$, Shannon diversity $p = 0.744$, Faith's phylogenetic diversity $p = 0.937$; Figure 17A, Appendix C Figure 6A,B) varied significantly in sediments with versus without sea cucumbers. The results suggest that the presence of sea cucumbers predominantly affects the taxonomic composition but not overall richness of the sediment microbiome, potentially via selective removal or enrichment of microbial taxa during passage of sediment through the GI tract.

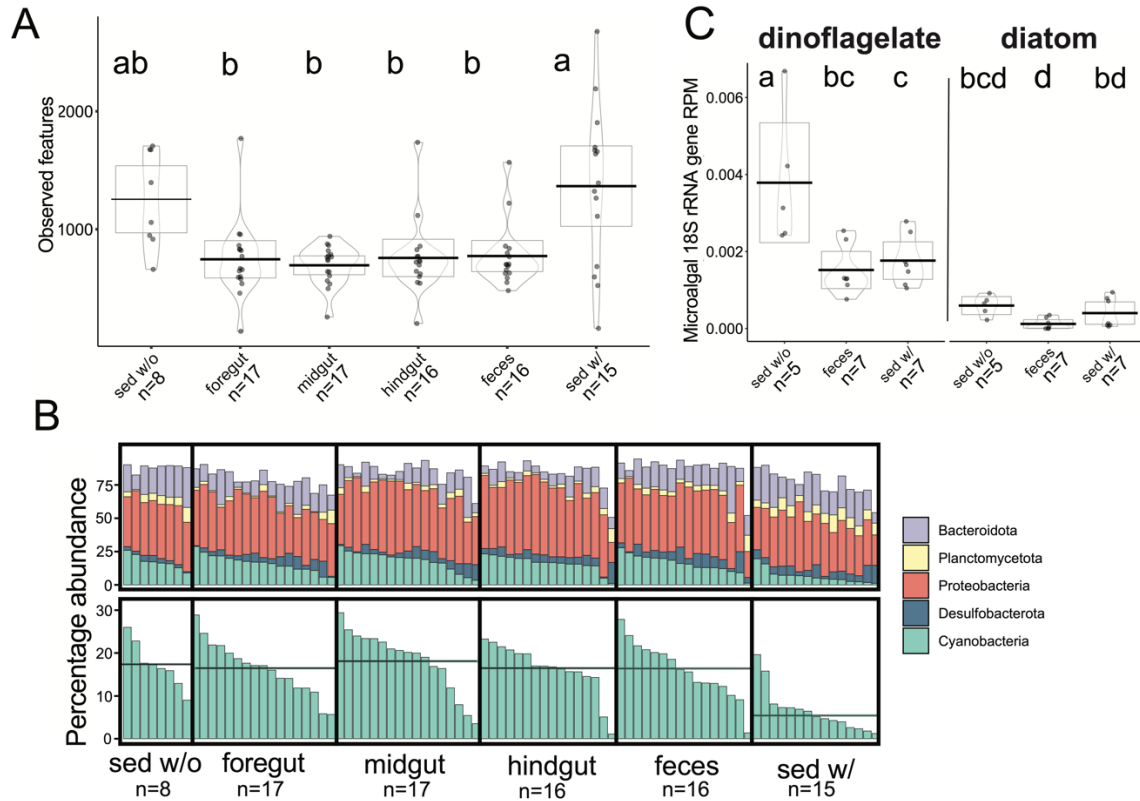


Figure 17: Sea cucumber grazing reshapes sediment microbiome diversity and depletes phototrophs. **A)** Alpha diversity (Observed features) in microbiomes of ungrazed (sed w/o) and grazed (sed w/) sediment compared to those of the sea cucumber gastrointestinal (GI) tract and feces. **B)** Upper: Phylum-level community composition, including only phyla with minimum percentage abundances >0.25%. Lower: Same data as in upper graphs, but only including Cyanobacteria, average percent abundances for each sample type shown as horizontal bar. Center log ratio (CLR)-transformed abundances are in Appendix C Figure 7A. **C)** Relative abundance of microalgal 18S rRNA genes in un-assembled metagenomes, with genes assigned phylum-level taxonomy based on top hits in the NCBI nt-nucleotide database. Normalized 18S rRNA gene abundances were calculated by dividing 18S rRNA gene reads by megabase of sequencing effort for each sample, resulting in reads per megabase (RPM). **A, C)** Alpha diversity and normalized microalgal abundances presented as violin plots (mean as estimate of center; box bounded by 95% confidence interval; density estimator used to draw violin; observations as individual points). Dissimilar letters indicate $P < 0.05$ in pairwise comparison between groups, while sharing of any letter indicates $P > 0.05$ (p values were obtained from linear mixed effects model; Tukey-Kramer HSD).

Comparison of microbiomes in ungrazed and grazed sediment and digesta demonstrates that sediment microbiomes change during passage through the sea cucumber GI tract.

Microbiome composition varied significantly at each stage of sediment digestion, from before [ungrazed, sed w/o], to during [foregut, midgut, hindgut, feces], and after [grazed, sed w/]) grazing (Bray-Curtis and Weighted UniFrac; PERMANOVA, $P < 0.05$; Appendix C Figure 5B,D). Microbiome alpha diversity also changed with digestion, being relatively high in ungrazed sediment, immediately decreasing in the foregut and staying depressed throughout the GI tract and feces, then rebounding in the grazed sediment to a level indistinguishable from ungrazed sediments (LME, $P < 0.05$, Figure 17A: Observed features, Appendix C Figure 6B: Faith's phylogenetic diversity).

Both beta and alpha diversity differed significantly between grazed sediments (sed/ w) and sediments conglomerated into fecal pellets (Figure 17A, Supplementary Figs. 5B,D 6B). Furthermore, of 11,071 16S rRNA ASVs detected across feces and grazed sediment samples, only 23% are shared in both sample types, while 20% and 57% are unique to feces and grazed sediment, respectively.

These results suggest that passage of sediments through sea cucumbers influences but does not fully explain microbiome diversity in grazed sediments. The microbiome in grazed sediments is distinct from that of ungrazed sediments but does not simply reflect the sea cucumber GI microbiome spread across the reef. This pattern is consistent with recent work on another *Holothuria* species, which identified a gut microbiome distinct from that of surrounding sediments both in taxonomic composition and in having lower alpha diversity, suggestive of removal of microbial taxa during digestion or adaptation to the intestinal niche (Pan et al., 2024). These and our results suggest that activity of sea cucumbers likely influences the sediment microbiome through a combination of mechanisms, potentially including: i) selective

consumption of microbes whose metabolisms or behavior disproportionately shape sediment microbiome composition (e.g., oxygen-producing phototrophs, bacterivorous protists), ii) alteration of the diversity and concentration of chemical substrates (e.g., particulate organic carbon) supporting microbial metabolism in sediments, iii) physical disruption of the sediment biofilm to facilitate access by other consumers or microbial dispersal out of sediments, and iv) release of a fecal microbiome that serves as a distinct starting point (inoculum) for successional regrowth of the sediment microbiome. Our analysis of microbiome gene content (below) provides preliminary insight into some of these hypotheses.

Sea Cucumber Grazing Significantly Depletes Phototrophs

A diverse set of microbial taxa differed in relative abundance between ungrazed and grazed sediments, with Cyanobacteria as the Phylum showing the largest variation. The CLR-transformed average relative abundance of cyanobacterial 16S rRNA gene amplicons was significantly higher in ungrazed (17.2%) compared to grazed sediment (6.5%; LME, $P < 0.05$; Figure 17B, Appendix C Figure 7A), with sequences matching Microcystaceae being particularly depleted in the presence of sea cucumbers (from 3.1% to 1.2% without versus with grazing, respectively; LME, $P < 0.05$; Appendix C Figure 8). This decline in amplicon-based relative abundance was observed alongside the qPCR-based decline (mean 75%) in total 16S rRNA genes per gram of sediment under grazing, suggesting that the absolute numerical abundance of cyanobacteria decreased by an estimated 90% under grazing. A substantial depletion was also observed in metagenomic sequences mapped to cyanobacterial MAGs (Appendix C Figure 7B). Interestingly, the relative abundance of cyanobacterial amplicons was also significantly higher in feces compared to grazed sediments (Figure 17B). This suggests that cyanobacterial depletion

from grazed sediments is not due solely to removal by digestion, but rather to some other mechanism, potentially including changes in substrate (e.g., nutrient) availability or physical reworking of the grazed sediment environment.

The relative abundance of eukaryotic microalgae also decreased under sea cucumber grazing, based on the frequency of 18S rRNA genes in metagenomes (Figure 17C). Notably, dinoflagellate 18S rRNA genes were ~50% less common in metagenomes from both grazed sediments and feces compared to metagenomes from ungrazed sediments (LME, $P < 0.05$). Although less abundant across all samples compared to dinoflagellate genes, diatom 18S rRNA genes were also depleted (~50%) in feces compared to ungrazed sediment metagenomes, consistent with literature suggesting that *H. atra* digests eukaryotic microalgae (Uthicke, 1999) and indeed contains a microbiome specifically adapted for digestion of microalgae and bacteria (Pan et al., 2024). However, diatom gene abundances did not significantly differ between grazed and ungrazed sediment metagenomes (Figure 17C). These patterns suggest that even in the presence of sea cucumber grazing, microalgal taxa such as diatoms may rebound quickly in reef sediments, while other taxa such as dinoflagellates remain depleted.

These taxonomic data link sea cucumber grazing to declines in the relative abundance of both cyanobacterial and eukaryotic phototrophs in sediment microbiomes. Given the qPCR results reported above, the biomass of cyanobacteria in surface reef sediments presumably also declines substantially under grazing. Eukaryotic phototroph biomass likely also declines but was not measured here. This removal of photosynthetic microbes is consistent with tank experiments showing decreased algal growth with versus without sea cucumbers (e.g. (Michio et al., 2003)) and field observations showing an increase in green or rust-colored pigmentation (Grayson et al.,

2022) and presumed cyanobacterial mat growth after sea cucumber exclusion (Clements et al., 2024; Purcell et al., 2016).

Sea Cucumber Removal of Phototrophs may Influence Coral Health

Corals may be harmed if the absence of sea cucumbers leads to an increased abundance of sediment phototrophs. Cyanobacteria, in particular, are known for producing compounds allelopathic to animals. Cyanotoxins such as microcystin, as well as other cyanobacteria-derived metabolites, have been found in cyanobacteria-dominated coral lesions and suggested to contribute to the progression of black band disease (BBD) (Meyer et al., 2023; L. L. Richardson et al., 2007). Here, several known cyanotoxin-producing families, including the Microcystaceae, Phormidiaceae and Cyanobacteriaceae (Huisman et al., 2018; J. Kang et al., 2022; W. Li et al., 2024), were present in our 16S rRNA gene datasets in the absence of sea cucumbers (Appendix C Figure 8). To follow-up on this association genomically we searched our 18 cyanobacterial MAGs (Figure 18A) for evidence of cyanotoxin genes. Many cyanotoxins are produced through non-ribosomal protein synthesis pathways encoded by well-described biosynthetic gene clusters, with the presence of these genes being an indicator of cyanotoxins in the environment (Pacheco et al., 2016; Zupančič et al., 2021).

We detected several near-complete cyanotoxin gene clusters in diverse cyanobacterial MAGs from feces and sediment samples (Figure 18B). Six of these MAGs (2 *Crocospaera* sp. [Microcystaceae], 2 *Xenococcaceae*, 1 *Prochloraceae*, and 1 *Trichodesmium* sp. [Microcoleaceae]) contain all or nearly all the queried gene cluster for synthesis of anatoxin (*anaA-G*). Anatoxin is neurotoxic to humans and many vertebrates (Lovin & Brooks, 2019), and recent literature has shown this alkaloid toxin is variably toxic to aquatic invertebrates (B.

Anderson et al., 2018; L. T. Kelly et al., 2020). The anatoxin-encoding Prochloraceae MAG also contains most of the genes for synthesizing microcystin (*mcyA-H*), lacking only *mcyI* and *mcyJ*. Gene clusters for the cyanotoxins cylindrospermopsin and saxitoxin were also present but were incomplete in all MAGs (Figure 18B).

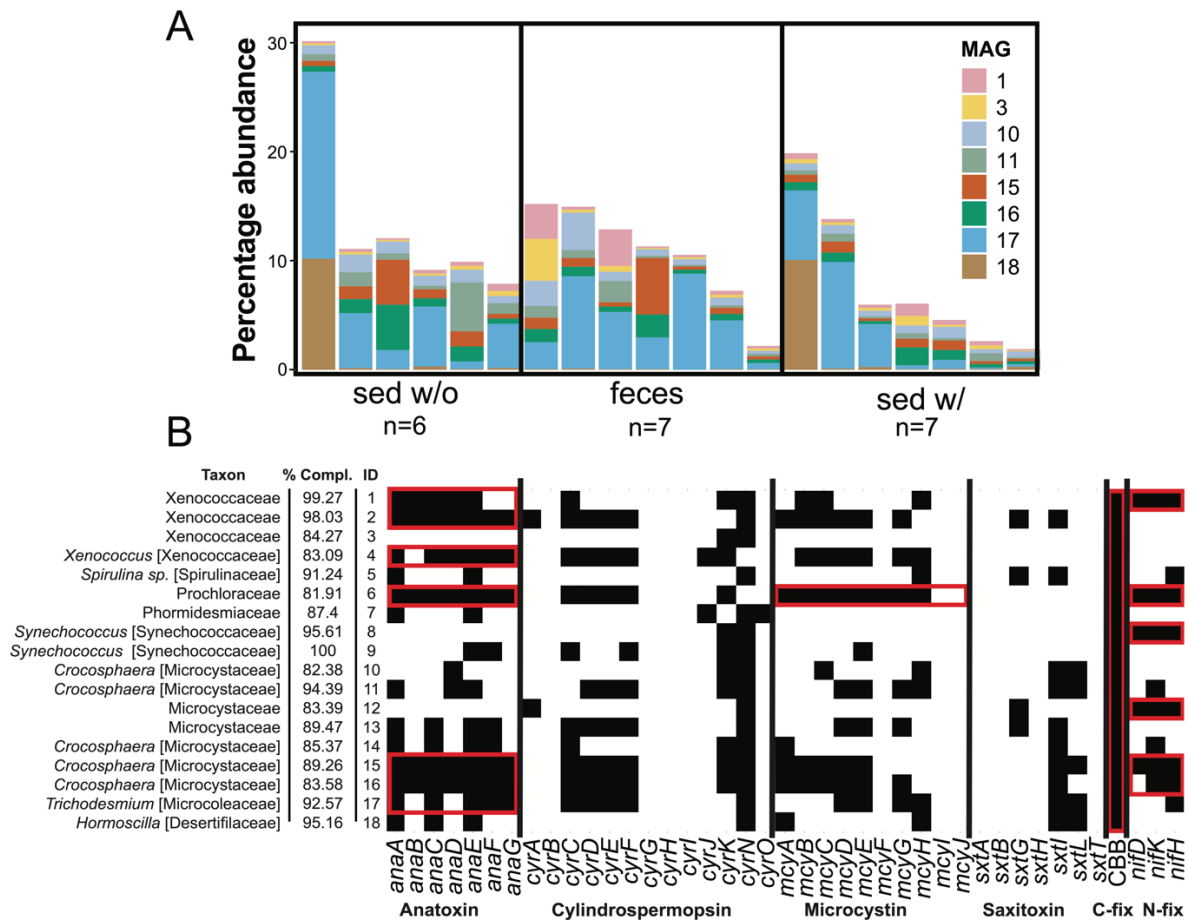


Figure 18: Abundant, high-quality cyanobacterial metagenome assembled genomes (MAGs) encode genes for cyanotoxin biosynthesis. **A)** Percentage abundance of the most abundant cyanobacterial MAGs. Abundance was estimated by first recovering MAGs from sea cucumber feces and sediment samples using individual and co-assembly approaches, then percentage abundances of cyanobacterial MAGs were estimated by mapping metagenomic reads to all dereplicated MAGs (99% ANI, >75% completion, <10% contamination) and plotted the most abundant MAGs (max abundance >3%). **B)** Cyanobacterial MAGs contain diverse cyanotoxin biosynthesis genes and carbon and nitrogen fixation genes. Cyanotoxin genes were identified by

querying (via DIAMOND BLASTp) open reading frames (amino acid translation) to custom databases for each gene of interest (see Supplementary Table 4). The thresholds for positive gene detection were >40% sequence identity, >80% query coverage, and $<10^{-50}$ e-value. MAGs possessing all but 1-2 genes for any queried cyanotoxin gene cluster are indicated with a red bounding box. Carbon and nitrogen fixation genes were identified via Hidden Markov models as implemented in METABOLIC-G. MAG binning statistics are in Supplementary Table 5. Taxon, finest taxonomic resolution available – brackets used to specify family if genus is available; ID, unique cyanobacterial MAG identifier; % Compl., estimated MAG percentage completion; *anaA-G*, anatoxin; *cyrA-K,N,O*, cylindrospermopsin toxin; *mcyA-J*, microcystin toxin; *sxtA,B,G,H,I,L,T*, saxitoxin; CBB, Calvin-Benson-Basham cycle genes; *nifD,K,H*, nitrogenase; C-fix, carbon fixation; N-fix, nitrogen fixation.

Cyanobacteria may also contribute to diminished coral health through mechanisms that do not involve known cyanotoxins. For example, cyanobacteria of the genus *Pleurocapsa* (Xenococcaceae) are not known for cyanotoxin production but have been associated with low recruitment of acroporid and pocilloporid coral larvae (Kegler et al., 2017; Turnlund et al., 2023). Here, *Pleurocapsa* 16S rRNA genes were more abundant in 16S rRNA gene datasets when sea cucumbers were absent. However, a mechanism(s) by which *Pleurocapsa* may interact (negatively or positively) with corals has not been confirmed.

Coral health may also be affected by other phototrophs detected in our datasets, including dinoflagellates. Dinoflagellate blooms have been linked to coral death (e.g., (Bauman et al., 2010)), and the potential for dinoflagellates to harm corals is well known in coral aquaculture (F.J.S. personal communication). Dinoflagellate blooms likely harm corals primarily by influencing oxygen, nutrient, or light availability. However, dinoflagellates also produce toxins, some of which can affect coral physiology, for example by altering calcium ion flux across membranes (McConnaughey, 2012). While our sediment metagenomes did not contain dinoflagellate contigs sufficient for detecting toxin genes, we observed a relative depletion of dinoflagellate 18S rRNA gene fragments from metagenomes when sea cucumbers are present

(Figure 17C). Collectively, our results raise the hypothesis that sea cucumbers benefit corals by removing potentially harmful cyanobacteria, dinoflagellates, or other phototrophs.

Sea Cucumber Grazing Reshapes Microbiome Metabolic Potential, Enriching for Genes of Anaerobic Metabolisms

Sea cucumbers may also alter the metabolic properties of the sediment microbiome, potentially in ways that influence coral health. To explore this potential, we analyzed sediment metagenomes for marker genes of biogeochemically relevant metabolisms. Notably, sea cucumber grazing significantly reduced the relative abundance of genes for oxygenic photosynthesis, including *psaA* and *psbA* encoding components of photosystem I (PSI) and photosystem II (PSII) respectively, responsible for the splitting of water and production of oxygen during light-dependent reactions (Figure 19A, Appendix C Figure 14C). The relative abundance of *rbcL*, encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase responsible for Calvin cycle-based carbon fixation did not vary. Similarly, *nifH* encoding nitrogenase for nitrogen fixation also decreased in abundance under grazing, most likely due to removal of diazotrophic cyanobacteria, examples of which were detectable in the MAG data (Figure 18B, Appendix C Figure 14B). This depletion of genes associated with photosynthesis or phototrophs is consistent with the taxonomic trends described above and predictions that sea cucumber grazing removes primary producers.

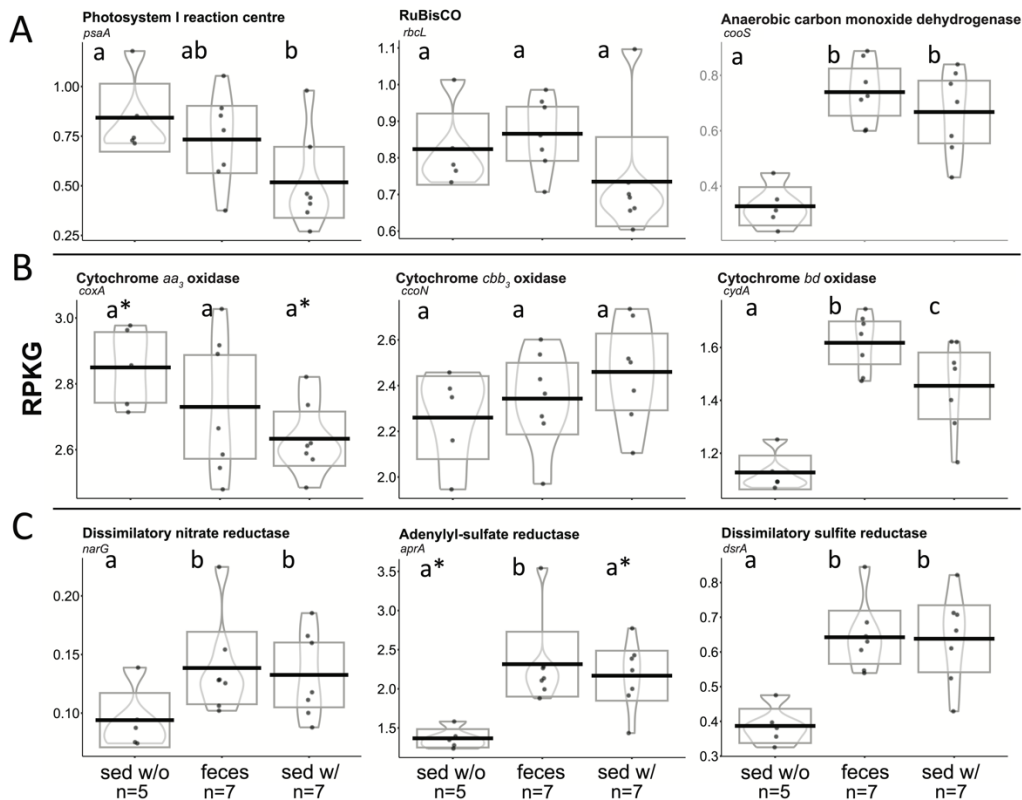


Figure 19: Microbiome metabolic potential differs between ungrazed sediment (sed w/o) versus sediment grazed (sed w/ and feces) by sea cucumbers, with anaerobic metabolisms enriched in grazed sediments. Relative abundance of marker genes shown as reads per kilobase genome equivalent (RPKG). **A)** Photosynthesis and carbon fixation genes, showing depletion of phototrophs and enrichment in anaerobic carbon fixation genes. *psaA*, subunit of photosystem I reaction center (photosynthesis); *rbcL*, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) large subunit (Calvin-Benson-Basham carbon fixation), *cooS*, carbon monoxide dehydrogenase (anaerobic Wood Ljungdahl carbon fixation). **B)** Cytochrome oxidase genes, showing enrichment of high O₂ affinity (*ccb₃* and *bd*-type) oxidase genes with sea cucumber grazing. **C)** Oxygen-sensitive reductive energy metabolism genes, showing increase in potential use of non-O₂ electron acceptors with sea cucumber grazing. *narG*, dissimilatory nitrate reductase alpha subunit (nitrate reduction); *aprA*, adenylyl-sulfate reductase alpha subunit (sulfur reduction); *dsrA*, dissimilatory sulfite reductase alpha subunit (sulfur reduction). Gene abundance presented as violin plots (mean as estimate of center; box bounded by 95% confidence interval; density estimator used to draw violin; observations as individual points). Dissimilar letters indicate $P < 0.05$ in pairwise comparison between groups, while sharing of any letter indicates $P > 0.05$; shared letters with an asterisk indicates $P < 0.10$ between contrasts (P values were obtained from linear mixed effects model; Tukey-Kramer HSD).

In contrast to our expectations, sea cucumber grazing resulted in a significant enrichment of diverse genes linked to anaerobic or microaerophilic metabolism. Concurrent with removal of oxygenic phototrophs, marker genes of the anaerobic Wood-Ljungdahl (*cooS* encoding carbon monoxide dehydrogenase Figure 19A; *acsB* encoding acetyl-CoA synthase, Appendix C Figure 14D) and reductive-TCA (*aclB* encoding ATP citrate synthase, Appendix C Figure 14D) carbon fixation pathways became significantly more abundant. Markers of dissimilatory anaerobic metabolism also increased significantly. These included genes of denitrification (*narG* encoding nitrate reductase, Figure 19c; *nirS* and *norB* encoding nitrite and nitric oxide reductase, respectively, Appendix C Figure 14B), anaerobic ammonium oxidation (anammox; *hzsA* encoding hydrazine synthase; Appendix C Figure 14A), and sulfur reduction (*dsrA* encoding dissimilatory sulfite reductase; Figure 19C; *aprA* encoding adenylyl sulfate reductase, *asrA* encoding anaerobic sulfite reductase, *sat* encoding sulfate adenylyltransferase, Appendix C Figure 14B). Genes encoding terminal oxidases with differing affinities for oxygen also changed in abundance consistent with a decrease in oxygen availability under grazing (Figure 19B). In contrast to ungrazed sediments, sediments under grazing were depleted of genes encoding lower O₂ affinity *aa₃*-type cytochrome c oxidase genes typically associated with respiration under normoxic conditions. Rather, grazed sediments were enriched in genes for higher affinity *cbb₃*- and *bd*-type cytochrome oxidases, with this enrichment being significant and largest for *bd* variants. Both *cbb₃* and *bd* oxidases are associated with microorganisms conducting microaerobic respiration under limited oxygen availability (D'mello et al., 1996; Kalvelage et al., 2015; Kumagai et al., 2018).

The representation of oxygen-sensitive marker genes in fecal metagenomes differed from that of grazed or ungrazed sediment, but not consistently for all genes (Figure 19B, Appendix C Figure 14). For some genes, their relative abundance in feces was similar to, or in some cases indistinguishable from (e.g., *narG*, *aprA*, *dsrA*, *cooS*), values in grazed sediment. However, other genes were intermediate in abundance between values in grazed and ungrazed sediment (e.g., *coxA*, *psaA*, *ccoN*), while still others were significantly elevated in feces compared to both sediment types (e.g., *napA*, *cydA*, *asrA*, *sat*; Appendix C Figure 14B,D,C). These complex functional gene patterns are consistent with the taxonomic trends reported above, suggesting that grazed sediment microbiomes may be influenced by, but are nonetheless distinct from, fecal microbiomes.

Collectively, the marker gene data indicate that sea cucumber grazing alters the functional potential of marine sediments, notably enriching for anaerobic or microaerobic metabolisms suggestive of a decrease in local oxygen concentrations. This pattern is seemingly not consistent with the hypothesis that sea cucumber bioturbation aerates reef sediments. The pattern is instead consistent with 1) the observed removal of phototrophs by grazing, which would likely remove a local oxygen source (oxygenic photosynthesis) from sediment biofilms, and 2) addition of gut-associated and low-oxygen adapted microbial communities to sediment via sea cucumber defecation. While the data provide support for both mechanisms, other mechanisms may also be contributing. For example, it is likely that grazed sediment represents some combination of ungrazed sediment grains and sediment grains held together by fecal casing/mucus, which may serve to limit oxygen diffusion from overlying waters relative to ungrazed sediment. However, we stress that we did not measure sediment oxygen concentration

in this study but that future measurements of oxygen are essential for validating the predictions based solely on gene data, as well as the potential mechanisms contributing to oxygen changes.

Conclusions

Sea cucumber grazing significantly transforms the abundance and biochemical potential of reef sediment microbiomes. Sea cucumber presence notably decreases the abundances of all bacteria, particularly oxygenic phototrophs, and by extension the abundance of oxygen and cyanotoxin production pathways. While we initially predicted that sea cucumber grazing, like that of many other bioturbating freshwater and marine invertebrates (e.g., freshwater oligochaetes (Saaltink et al., 2018), marine crabs (Kristensen, 2008), marine polychaetes (Gutiérrez & Jones, 2006), wetland oligochaetes and insects (Michael et al., 2023)), would increase oxygen flux into sediments, our data suggest that the presence of sea cucumbers instead results in an elevated relative potential for sediment anaerobic processes (including sulfide production from sulfate reduction or desulfurylation or the removal of oxygen for coral respiration (e.g., (L. Richardson et al., 2009; M. Weber et al., 2012))), at least within the timeframe of our *in situ* experiments. Collectively, these results suggest sea cucumbers influence the potential for cyanotoxin, oxygen, and anaerobic metabolite production, all of which have potential to affect coral survival.

However, we caution that our results reflect metabolic potential. Measurements of sediment microbial activity and metabolite and oxygen concentration were not obtained but could help identify potential chemical mechanisms by which the sea cucumber-sediment interaction influences coral health. Indeed, it is possible that the absolute concentration of metabolites associated with anaerobiosis (e.g., sulfide) may still decline under grazing, given the

coinciding and substantial reduction in sediment microbial load. Chemical measurements are justified given that prior work (e.g., (Clements et al., 2024)) has not identified a pathogen(s) explaining the increased coral mortality in the absence of sea cucumbers. The current study suggests that metabolites of cyanobacterial or microalgal origin might be particularly likely to vary with sea cucumber abundance, given the pronounced effect of sea cucumber grazing on sediment phototrophs, and therefore potential candidates for having health effects on corals.

Sea cucumbers exert diverse, significant, and potentially cascading effects on both microbial and macroorganismal (e.g., coral) components of coral reefs. The mechanisms by which these effects are connected remain uncertain but are likely linked to a transformation of sediment microbial biogeochemistry under sea cucumber grazing. Identifying these mechanisms may be critical for mitigating the detrimental effects of historic and ongoing overharvesting of sea cucumbers from reefs worldwide.

Data Availability

All sequence data (16S rRNA gene amplicons, metagenomes, and cyanobacterial MAGs) were deposited at the National Center for Biotechnology Information under BioProject PRJNA1136881.

Acknowledgements

We thank the French Polynesian Government (Délégation à la Recherche) and the Haut-commissariat de la République en Polynésie Française (DTRT) for permits (Protocole d'Accueil 2017–2021) to work in French Polynesia. N. Altman-Kurosaki, and A. Caughman assisted in the

field, A. Bertagnolli provided useful bioinformatics advice, S. Walk provided access to, and T. Wolfe provided training on, qPCR.

Financial support came from the U.S. National Science Foundation (grant no. OCE 1947522 to M.E.H. and F.J.S.), National Geographic Society (grant no. NGS-57078R-19 to M.E.H. and C.S.C.), Teasley Endowment to the Georgia Institute of Technology (to M.E.H.), Anna and Harry Teasley Gift Fund (to M.E.H.), and Simons Foundation (to F.J.S.). This work represents a contribution of the Mo‘orea Coral Reef (MCR) LTER Site supported by the U.S. National Science Foundation (grant no. OCE 16–37396).

CHAPTER FIVE

CONCLUSIONS AND SUGGESTIONS

Coral reefs are diverse macroecological ecosystems (DeVantier, 2020; Littler & Littler, 2013), and this diversity is recapitulated at the microbiome scale (Chapter 2, (Galand et al., 2023)). Reef building corals are a foundational species on eponymous coral reefs, the basis of reef biodiversity. These essential corals are high high-priority targets for conservation in the face of global anthropogenic stressors (Harborne et al., 2017; Hughes et al., 2018; Rasher & Hay, 2010). The coral microbiome is an essential facet to coral survival where coral-microbial interactions include pathogens (Chapter 3, (Clements et al., 2024; L. Richardson et al., 2009)) and mutualists (Glasl et al., 2016; Rosado et al., 2019). This microbial diversity makes it relatively challenging to choose a target to parse coral-microbe-interaction. This is especially challenging if utilizing untargeted sequencing tools which accurately renders *in situ* community complexity bioinformatically.

In this dissertation, my introduction begins by laying out a systematic framework for selecting ecologically relevant targets and investigating function. The next three chapters of original research implement this framework, focusing on sequence-based analyses. In Chapter Two, I survey prokaryotes and eukaryotes across the coral reef and in Chapters Three and Four I dive deeper into two distinct coral-microbial interactions.

These interactions were chosen for investigation based on two criteria outlined in the introduction: 1) identifying a prevalent, high-abundance microbe from 16S rRNA survey data and 2) identifying a microbially-mediated ecosystem-scale phenomena. Chapters Two and Three utilize transcriptomics (Chapter 3) and metagenomics (Chapter 4) to reconstruct microbial

physiology across the reef, identifying potentially key taxa and metabolisms for follow-up study. These investigations reveal that a dominant coral endosymbiont is expressing genes indicative of a parasitic lifestyle (a departure from its hypothesized function from prior genomic and *in vitro* studies; Chapter 3) and that coral-sea cucumber macroecological facilitation may be mediated through non-infectious alterations to a shared sediment microbiome (Chapter 4). These results, enrich our understanding of the ecology of coral reefs, highlighting microbial and metazoan beneficial and harmful roles for coral health. Though highly informative, these results bear further investigation to elucidate how these coral-microbial interactions operate over space and time, eventually shaping coral physiology and survival.

Reef-scale Survey of Microhabitats

The survey of reef microbiomes (Chapter 2) revealed several taxa of interest. Some of these were addressed in other chapters of this dissertation (e.g. coral-associated *Endozoicomonas*, Chapter 3; sediment-associated dinoflagellates, Chapter 4). Two outstanding microbes identified in Chapter 2 was the coral-associated *Halomonas* sp. and the macroalgae-associated rhizobiales.

Coral-associated *Halomonas* spp. have been characterized genomically (Rosado et al., 2023; Villela et al., 2023) and live microbes applied to coral mesocosms, alleviating bleaching severity (Rosado et al., 2019). Though there is room to investigate the molecular basis of taxa's beneficial role for corals, but I personally would not want to attempt more complex molecular biology on field or aquarium-based corals.

The macroalgae-associated rhizobiales is a different story. This Order is best known for associating with plants in mutualistic nitrogen fixing symbioses and animals as pathogens (S.

Wang et al., 2020), with their function in marine systems poorly explored. Recently it was demonstrated that marine rhizobiales fix nitrogen for diatoms (Tschitschko et al., 2024). This adds to a growing literature of diazotrophs associating with marine phototrophs (e.g. seagrasses (Mohr et al., 2021)). I would want to pursue characterizing the metabolism of marine rhizobiales in-association with macroalgae (particularly nitrogen fixation). I would structure this in two stages. First, I would sequence the macroalgae-associated microbiome with metagenomics (and then, if warranted, transcriptomics). Using this sequence data to reconstruct individual and community metabolic potential, I would perform a specific search for diazotrophy genes (e.g. *nif*; (S. Wang et al., 2020)). If rhizobiales (or another microbiome member) possessed genes for nitrogen cycling, I would then seek to demonstrate function.

To demonstrate microbial nitrogen fixation *and* provisioning to the algal host, it would be advantageous to perform stable isotope probing of algal-microbiome exchange. I would incubate macroalgae mesocosms with heavy nitrogen gas, light, and carbon dioxide, and then measure algal tissue for N¹⁵ incorporation. Simultaneous transcriptomic sampling of the macroalgae microbiome would likely identify the diazotrophic (expression of nitrogen fixing genes). Assuming that a microbe was fixing nitrogen and providing it to the host, the taxon could be contextualized by screening public datasets for macroalgae microbiomes (e.g. (Nahor et al., 2024)) to identify other species and locations where this process may occur.

Endozoicomonas-Coral Interactions

Results from Chapter Three suggest that the ubiquitous and abundant coral endosymbiont *Endozoicomonas* parasitizes its host by stealing ATP, vitamins, sugars, amino acids, nucleic

acids, and trace elements. This expression-based hypothesis contradicts the popular genomic and *in vitro*-based model that considers *Endozoicomonas* a coral mutualist. This chapter provides a useful foundation for future dissection of the valence of coral-*Endozoicomonas* relationship. For next steps, I would suggest similar transcriptomic *in hospite* investigations of *Endozoicomonas* function at finer grained scales (both spatial and genomic) and under other environmental conditions.

Spatial Scales

This study used bulk-tissue transcriptomics, which was effective for acquiring enough reads to assemble *Endozoicomonas* transcripts but obscures any differences in *Endozoicomonas* expression over spatial scales. *Endozoicomonas* forms intracellular aggregates (Pogoreutz & Ziegler, 2024), and expression may differ depending on the location of these aggregates within coral tissues, influenced by exposure to differential microscale oxygen, UV, nutrient, and nutrient conditions (as suggested by the sampled *Endozoicomonas* population expressing aerobic and anaerobic metabolism simultaneously). The effect of spatial differences may be resolved with field sampling (*Endozoicomonas* is lost in aquaria; (Pogoreutz & Ziegler, 2024)) of coral tissues for DNA and RNA analyses, followed whole mount cryo-sectioning of corals (Wada et al. 2016, Marine Biology), and laser microdissection (LMD) of *Endozoicomonas* aggregates (Wada et al., 2022) to recover metagenomes for use with transcriptomics. It may also be possible to use CARD-FISH on whole mount coral to label mRNA transcripts and identify heterogeneity between and within aggregates.

Genomic Scales

The present study also homogenized genomic scales. Transcripts for *Endozoicomonas* activity were acquired by using all sequences that aligned to any *Endozoicomonas* sequence in the NCBI. This was effective at creating a population-level transcriptome, but not appropriate for resolving species or strain-specific differences. Though our samples were dominated by a single 16S rRNA gene/ transcript fragment (Chapter 3), there may be strain specific differences within *Endozoicomonas* genomes that underlie this apparent homogeneity. Furthermore, these strains may act differently within the coral microbiome, resulting in divergent coral fates depending on which strain it is colonized by. It may be possible to disentangle strain specific differences by mapping reads from bulk-transcriptome sequencing to *Endozoicomonas* single amplified genomes (SAGs; (Ide et al., 2022)), potentially distinguishing parasitic strains from mutualistic ones.

Environmental Context

This study is a snapshot of a single coral host species under one set of conditions. *Endozoicomonas* relative abundance varies by host health (Pogoreutz & Ziegler, 2024) and coral host (Chapter 3), and the nature of *Endozoicomonas* symbiosis may differ in other hosts or stress regime.

Chaper Three used samples from the coral *Pocillipora damicornis* that were beginning to bleach, a visible indicator of coral stress. Stress may play a role in the valence of the *Endozoicomonas*-coral symbiosis, as observed in other eukaryote-bacterial interactions. When the coccolithophore *Gephyrocapsa huxleyi* is stressed, its *Roseobacter* symbiont switches from a mutualist into an opportunistic pathogen (Segev et al., 2016; Seyedsayamdost et al., 2011). *Endozoicomonas* may also be an opportunist, evidenced by its large genome and capacity for

free-living (Pogoreutz & Ziegler, 2024). Repeating a transcriptomic (either *de novo* assembly or SAG/MAG alignment) analysis of *Endozoicomonas* function with less-stressed corals may help tease apart how this interaction may shift under stress.

To extend these results past *P. damicornis* it would be informative to perform any of these *in hospite* transcriptomic approaches in other coral species. Host species is broadly predictive of *Endozoicomonas* relative abundance (Chapter 3, (Caughman et al., 2021)), and these two variables may play interacting roles determining *Endozoicomonas* function, requiring testing an array of conditions to disambiguate this.

Sea Cucumber Grazing and Microbial Linkages to Coral Mortality

Sea cucumber grazing of sediments reduces coral mortality (Clements et al., 2024). Results from Chapter Four suggest multiple microbial mechanisms that could underpin this observation, specifically: sea cucumber removal of toxin producing phototrophs, depletion of sulfide producing bacteria, and the depletion of the microbial community that may remove unobserved pathogens or parasites. These mechanisms are based on metabolic potential and further experiments could move our understanding closer to metabolic reality. A mechanistic understanding of the sea cucumber-microbiome-coral interaction would be useful in an informed use of sea cucumbers (or taxa that share pertinent functions) for the proactive management and conservation of coral reefs.

Further investigation could easily leverage the experimental design of Chapter Four, using sea cucumber inclusion/ exclusion cages to isolate the effects of cucumber grazing, working with a constrained area. In all cases, it would be useful to perform these experiments in the presence of coral fragments, using coral mortality as a direct response variable, linking

sediment manipulation and associated microbiome changes to the key response variable: coral survival. To elaborate on our microbial hypotheses, I would focus on clarifying four features of this hypothesized sea cucumber-microbial-coral interaction: enhanced microbial quantification, measurements of predicted metabolites and associated transcripts, the exact mechanism of sediment transformation, and the site-specific conditions that may modulate the protective effect of sea cucumbers.

Microbial Quantification

In this chapter, we inferred changes in microbial abundances using qPCR and DNA concentrations. Though this approach paints an accurate picture of microbial abundances, it would be informative to quantify the the guilds underpinning this change through direct cell observation. I would propose collection of samples for flow cytometry (or failing that, epifluorescent microscopy). Utilizing a general DNA stain (e.g. DAPI) and photosynthetic autofluorescence we would be able to determine absolute changes to heterotroph and autotroph abundances before and after grazing. And, by simultaneously collecting cell size, we would be able to further disambiguate changes to microeukaryote vs. prokaryote abundances (L. Weber & Apprill, 2020).

Biogeochemistry and Toxin Production

In this chapter, we predict changes to sediment biogeochemical flux based on genomic potential. A useful extension of these predictions would be to use *in situ* transcriptomics and geophysical measurements of the target metabolites to confirm genomic predictions. For biogeochemical evidence, we would measure sulfide in sediment porewater (Fogo & Popowsky, 1949), and microprofile oxygen content (Jørgensen et al., 1979). Further, we would measure

cyanotoxin presence and concentration using enzyme-linked immunosorbent assays (ELISA; per. (Glas et al., 2010; L. L. Richardson et al., 2007)). In all cases, if metabolites of interest were produced, reference-based transcriptomics could be used to identify the culpable microbes. For transcription evidence, sediment collected simultaneously with the biogeochemical measures could be extracted for DNA to generate MAGs and RNA to generate reads that could be mapped back to the MAGs to identify which microbes are expressing genes that might mediate sediment biogeochemistry and/ or toxin production.

Mechanism of Sediment Transformation for a Protective Effect

From the available literature, it is unclear how sea cucumbers alter the sediments to reduce coral mortality. This may be modulated by microbial selection via digestion, deposition of sea cucumber gut microbes or mucosa, or mechanical disruption. If mechanical disruption is the key feature of this interaction, it is unclear if other bioturbating organisms perform a similar function. To disambiguate the importance of sediment disruption, it would be useful for researchers to manually disrupt the reef sediment to a similar depth to *Holothuria atra* grazing (0.5 cm; (Uthicke, 1994)), and to include conditions that include other reef bioturbators/ sand-sifters (e.g. sea stars/ bristle stars, fiddler crabs, conchs, whelks, and snails). If the coral survival benefit can be repeated without sea cucumbers, it may be more broadly applicable to reef management, as sea cucumbers have been depleted worldwide, allowing functionally redundant taxa to pick up the slack, protecting corals.

Site-Specific Conditions that Modulate Protective Effect

Finally, it is unclear if and how site-specific conditions modulate the protective effect of sea cucumber grazing, which may limit or expand the scope of sea cucumbers for coral conservation. To begin to understand potential geographic constraints/ applicability, it would be useful to test the protective effect of cucumbers on sites with different nutrient loading regimes (high vs. low nutrients) and temperatures (proxied by season or latitude) to identify broad geographic regions that may benefit from more intensive sea cucumber management.

Final Remarks

Coral reefs are complex ecosystems at macro and micro-ecological scales. The interactions between reef-members are variable across spatial, genomic, and temporal scales. Because corals are an essential and threatened macrofaunal species on reefs, the goal and result of this dissertation was to survey the microbial milieu of the coral reef, identify and investigate tractable coral-microbial interactions, and describe the features underlying these interactions. All of this serves to understanding how microbes integrate into coral ecology and physiology, mediating coral health and ultimately persistence.

CUMULATIVE REFERENCES CITED

- Abby, S. S., Denise, R., & Rocha, E. P. C. (2024). Identification of Protein Secretion Systems in Bacterial Genomes Using MacSyFinder Version 2. In L. Journet & E. Cascales (Eds.), *Bacterial Secretion Systems* (Vol. 2715, pp. 1–25). Springer US.
https://doi.org/10.1007/978-1-0716-3445-5_1
- Alcolombri, U., Peaudecerf, F. J., Fernandez, V. I., Behrendt, L., Lee, K. S., & Stocker, R. (2021). Sinking enhances the degradation of organic particles by marine bacteria. *Nature Geoscience*, 14(10), Article 10. <https://doi.org/10.1038/s41561-021-00817-x>
- Alderdice, R., Suggett, D. J., Cárdenas, A., Hughes, D. J., Kühl, M., Pernice, M., & Voolstra, C. R. (2021). Divergent expression of hypoxia response systems under deoxygenation in reef-forming corals aligns with bleaching susceptibility. *Global Change Biology*, 27(2), 312–326. <https://doi.org/10.1111/gcb.15436>
- Al-Fatlawi, A., Menzel, M., & Schroeder, M. (2023). Is Protein BLAST a thing of the past? *Nature Communications*, 14(1), 8195. <https://doi.org/10.1038/s41467-023-44082-5>
- Al-Haj, A. N., & Fulweiler, R. W. (2020). A synthesis of methane emissions from shallow vegetated coastal ecosystems. *Global Change Biology*, 26(5), 2988–3005.
<https://doi.org/10.1111/gcb.15046>
- Alongi, D. M., Trott, L. A., & Pfitzner, J. (2008). Biogeochemistry of inter-reef sediments on the northern and central Great Barrier Reef. *Coral Reefs*, 27(2), 407–420.
<https://doi.org/10.1007/s00338-007-0347-2>
- Altman-Kurosaki, N. T., Pratte, Z. A., Stewart, F. J., & Hay, M. E. (2025). Coral–algal competition: Allelopathy, temporal variance, and effects on coral microbiomes. *Coral Reefs*, 44(1), 49–62. <https://doi.org/10.1007/s00338-024-02585-7>
- Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*, 59(1), 143–169. <https://doi.org/10.1128/mr.59.1.143-169.1995>
- Anantharaman, K., Brown, C. T., Hug, L. A., Sharon, I., Castelle, C. J., Probst, A. J., Thomas, B. C., Singh, A., Wilkins, M. J., Karaoz, U., Brodie, E. L., Williams, K. H., Hubbard, S. S., & Banfield, J. F. (2016). Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nature Communications*, 7(1), 13219.
<https://doi.org/10.1038/ncomms13219>
- Anderson, B., Voorhees, J., Phillips, B., Fadness, R., Stancheva, R., Nichols, J., Orr, D., & Wood, S. A. (2018). Extracts from benthic anatoxin-producing *Phormidium* are toxic to 3 macroinvertebrate taxa at environmentally relevant concentrations. *Environmental Toxicology and Chemistry*, 37(11), 2851–2859. <https://doi.org/10.1002/etc.4243>
- Anderson, J. G. T. (2013). *Deep Things Out of Darkness: A History of Natural History*. University of California Press.
- Angelakis, A. N., Capodaglio, A. G., & Dialynas, E. G. (2022). Wastewater Management: From Ancient Greece to Modern Times and Future. *Water*, 15(1), 43.
<https://doi.org/10.3390/w15010043>
- Antipov, D., Korobeynikov, A., McLean, J. S., & Pevzner, P. A. (2016). HYBRID SPA DES: An algorithm for hybrid assembly of short and long reads. *Bioinformatics*, 32(7), 1009–1015.
<https://doi.org/10.1093/bioinformatics/btv688>
- Aouizerat, T., Gutman, I., Paz, Y., Maeir, A. M., Gadot, Y., Gelman, D., Szitenberg, A., Drori, E., Pinkus, A., Schoemann, M., Kaplan, R., Ben-Gedalya, T., Copenhagen-Glazer, S.,

- Reich, E., Saragovi, A., Lipschits, O., Klutstein, M., & Hazan, R. (2019). Isolation and Characterization of Live Yeast Cells from Ancient Vessels as a Tool in Bio-Archaeology. *mBio*, 10(2), e00388-19. <https://doi.org/10.1128/mBio.00388-19>
- Apprill, A., McNally, S., Parsons, R., & Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.*, 75, 129–137.
- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, 36(7), 2251–2252. <https://doi.org/10.1093/bioinformatics/btz859>
- Aroney, S. T. N., Newell, R. J. P., Nissen, J. N., Camargo, A. P., Tyson, G. W., & Woodcroft, B. J. (2025). *CoverM: Read alignment statistics for metagenomics* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2501.11217>
- Aylward, F. O., Boeuf, D., Mende, D. R., Wood-Charlson, E. M., Vislova, A., Eppley, J. M., Romano, A. E., & DeLong, E. F. (2017). Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proceedings of the National Academy of Sciences*, 114(43), 11446–11451.
- Aylward, F. O., Eppley, J. M., Smith, J. M., Chavez, F. P., Scholin, C. A., & DeLong, E. F. (2015). Microbial community transcriptional networks are conserved in three domains at ocean basin scales. *Proceedings of the National Academy of Sciences*, 112(17), 5443–5448. <https://doi.org/10.1073/pnas.1502883112>
- Bainbridge, M. N., Warren, R. L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., Mardis, E. R., Sadar, M. D., Siddiqui, A. S., Marra, M. A., & Jones, S. J. (2006). Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics*, 7(1), 246. <https://doi.org/10.1186/1471-2164-7-246>
- Baker, I. R., & Girguis, P. R. (2024). Sulfur cycling likely obscures dynamic biologically-driven iron redox cycling in contemporary methane seep environments. *Environmental Microbiology Reports*, 16(3), e13263. <https://doi.org/10.1111/1758-2229.13263>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bauman, A. G., Burt, J. A., Feary, D. A., Marquis, E., & Usseglio, P. (2010). Tropical harmful algal blooms: An emerging threat to coral reef communities? *Marine Pollution Bulletin*, 60(11), 2117–2122. <https://doi.org/10.1016/j.marpolbul.2010.08.015>
- Bayer, T., Neave, M. J., Alsheikh-Hussain, A., Aranda, M., Yum, L. K., Mincer, T., Huguen, K., Apprill, A., & Voolstra, C. R. (2013). The Microbiome of the Red Sea Coral *Stylophora pistillata* Is Dominated by Tissue-Associated Endozoicomonas Bacteria. *Applied and Environmental Microbiology*. <https://doi.org/10.1128/AEM.00695-13>
- Beatty, D. S., Clements, C. S., Stewart, F. J., & Hay, M. E. (2018). Intergenerational effects of macroalgae on a reef coral: Major declines in larval survival but subtle changes in microbiomes. *Marine Ecology Progress Series*, 589, 97–114. <https://doi.org/10.3354/meps12465>

- Beatty, D. S., Clements, C. S., Valayil, J. M., Jarvis, S. Y., Ritchie, K. B., Stewart, F. J., & Hay, M. E. (2022). Variance of coral anti-pathogen defense in response to transplantation between coral- and macroalgal-dominated reefs. *Coral Reefs*, 41(5), 1417–1431. <https://doi.org/10.1007/s00338-022-02291-2>
- Beatty, D. S., Valayil, J. M., Clements, C. S., Ritchie, K. B., Stewart, F. J., & Hay, M. E. (2019). Variable effects of local management on coral defenses against a thermally regulated bleaching pathogen. *Science Advances*, 5(10), eaay1048. <https://doi.org/10.1126/sciadv.aay1048>
- Becker, C., Weber, L., Suca, J., Llopiz, J., Mooney, T., & Apprill, A. (2020). Microbial and nutrient dynamics in mangrove, reef, and seagrass waters over tidal and diurnal time scales. *Aquatic Microbial Ecology*, 85, 101–119. <https://doi.org/10.3354/ame01944>
- Becker, K. W., Harke, M. J., Mende, D. R., Muratore, D., Weitz, J. S., DeLong, E. F., Dyhrman, S. T., & Van Mooy, B. A. S. (2021). Combined pigment and metatranscriptomic analysis reveals highly synchronized diel patterns of phenotypic light response across domains in the open oligotrophic ocean. *The ISME Journal*, 15(2), 520–533. <https://doi.org/10.1038/s41396-020-00793-x>
- Begon, M., Harper, J., & Townsend, C. (1990). *Ecology: Individuals, Populations, Communities* (2nd ed.). Blackwell Science.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bennion, M., Ross, P., Howells, J., McDonald, I., & Lane, H. (2021). Characterisation and distribution of the bacterial genus *Endozoicomonas* in a threatened surf clam. *Diseases of Aquatic Organisms*, 146, 91–105. <https://doi.org/10.3354/dao03626>
- Benoit, G., Peterlongo, P., Mariadassou, M., Drezen, E., Schbath, S., Lavenier, D., & Lemaitre, C. (2016). Multiple comparative metagenomics using multiset k-mer counting. *PeerJ Computer Science*, 2, e94. <https://doi.org/10.7717/peerj-cs.94>
- Bergman, J., Pedersen, R. Ø., Lundgren, E. J., Lemoine, R. T., Monsarrat, S., Pearce, E. A., Schierup, M. H., & Svenning, J.-C. (2023). Worldwide Late Pleistocene and Early Holocene population declines in extant megafauna are associated with *Homo sapiens* expansion rather than climate change. *Nature Communications*, 14(1), 7679. <https://doi.org/10.1038/s41467-023-43426-5>
- Bertagnolli, A. D., Maritan, A. J., Tumolo, B. B., Fritz, S. F., Oakland, H. C., Mohr, E. J., Poole, G. C., Albertson, L. K., & Stewart, F. J. (2023). Net-spinning caddisflies create denitrifier-enriched niches in the stream microbiome. *ISME Communications*, 3(1), Article 1. <https://doi.org/10.1038/s43705-023-00315-8>
- Bertagnolli, A. D., & Stewart, F. J. (2018). Microbial niches in marine oxygen minimum zones. *Nature Reviews Microbiology*, 16(12), 723–729. <https://doi.org/10.1038/s41579-018-0087-z>
- Bianchi, D., Weber, T. S., Kiko, R., & Deutsch, C. (2018). Global niche of marine anaerobic metabolisms expanded by particle microenvironments. *Nature Geoscience*, 11(4), 263–268. <https://doi.org/10.1038/s41561-018-0081-0>

- Biddanda, B. A., Weinke, A. D., & Stone, I. P. (2023). Extant mat microbes synchronize vertical migration to a diel tempo. *Journal of Great Lakes Research*, 49(1), 220–228. <https://doi.org/10.1016/j.jglr.2022.10.006>
- Billerbeck, M., Røy, H., Bosselmann, K., & Huettel, M. (2007). Benthic photosynthesis in submerged Wadden Sea intertidal flats. *Estuarine, Coastal and Shelf Science*, 71(3–4), 704–716. <https://doi.org/10.1016/j.ecss.2006.09.019>
- Bin Jang, H., Bolduc, B., Zablocki, O., Kuhn, J. H., Roux, S., Adriaenssens, E. M., Brister, J. R., Kropinski, A. M., Krupovic, M., Lavigne, R., Turner, D., & Sullivan, M. B. (2019). Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nature Biotechnology*, 37(6), 632–639. <https://doi.org/10.1038/s41587-019-0100-8>
- Bisanz, J. E. (2018). *qiime2R: Importing QIIME2 artifacts and associated data into R sessions*. <https://github.com/jbisanz/qiime2R>
- Boeing, W. J., Leech, D. M., Williamson, C. E., Cooke, S., & Torres, L. (2004). Damaging UV radiation and invertebrate predation: Conflicting selective pressures for zooplankton vertical distribution in the water column of low DOC lakes. *Oecologia*, 138(4), 603–612. <https://doi.org/10.1007/s00442-003-1468-0>
- Bogachev, A. V., Murtazina, R. A., & Skulachev, V. P. (1996). H⁺/e⁻ stoichiometry for NADH dehydrogenase I and dimethyl sulfoxide reductase in anaerobically grown *Escherichia coli* cells. *Journal of Bacteriology*, 178(21), 6233–6237. <https://doi.org/10.1128/jb.178.21.6233-6237.1996>
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., ... Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, 37(8), Article 8. <https://doi.org/10.1038/s41587-019-0209-9>
- Bomar, L., Maltz, M., Colston, S., & Graf, J. (2011). Directed Culturing of Microorganisms Using Metatranscriptomics. *mBio*, 2(2), e00012-11. <https://doi.org/10.1128/mBio.00012-11>
- Bowers, R. M., Kyrpides, N. C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T. B. K., Schulz, F., Jarett, J., Rivers, A. R., Eloie-Fadrosch, E. A., Tringe, S. G., Ivanova, N. N., Copeland, A., Clum, A., Becraft, E. D., Malmstrom, R. R., Birren, B., Podar, M., Bork, P., ... Woyke, T. (2017). Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology*, 35(8), Article 8. <https://doi.org/10.1038/nbt.3893>
- Bradnam, K. (2015, June 11). *L50 vs N50: That's another fine mess that bioinformatics got us into*. <http://www.acgt.me/blog/2015/6/11/l50-vs-n50-thats-another-fine-mess-that-bioinformatics-got-us-into>
- Braginsky, M. (2023). *ggpirate: Pirate plotting for ggplot2*. <https://github.com/mikabr/ggpirate>
- Bryant, D. M., Johnson, K., DiTommaso, T., Tickle, T., Couger, M. B., Payzin-Dogru, D., Lee, T. J., Leigh, N. D., Kuo, T.-H., Davis, F. G., Bateman, J., Bryant, S., Guzikowski, A. R., Tsai, S. L., Coyne, S., Ye, W. W., Freeman, R. M., Peshkin, L., Tabin, C. J., ... Whited, J. L. (2017). A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of

- Limb Regeneration Factors. *Cell Reports*, 18(3), 762–776. <https://doi.org/10.1016/j.celrep.2016.12.063>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- Burkepile, D. E., & Hay, M. E. (2008). Herbivore species richness and feeding complementarity affect community structure and function on a coral reef. *Proceedings of the National Academy of Sciences*, 105(42), 16201–16206. <https://doi.org/10.1073/pnas.0801946105>
- Burman, N., Belukhina, S., Depardieu, F., Wilkinson, R. A., Skutel, M., Santiago-Frangos, A., Graham, A. B., Livenskyi, A., Chechenina, A., Morozova, N., Zahl, T., Henriques, W. S., Buyukyoruk, M., Rouillon, C., Saudemont, B., Shyrokov, L., Kurata, T., Haurlyuk, V., Severinov, K., ... Isaev, A. (2024). A virally encoded tRNA neutralizes the PARIS antiviral defence system. *Nature*, 634(8033), 424–431. <https://doi.org/10.1038/s41586-024-07874-3>
- Busetti, A., Maggs, C. A., & Gilmore, B. F. (2017). Marine macroalgae and their associated microbiomes as a source of antimicrobial chemical diversity. *European Journal of Phycology*, 52(4), 452–465. <https://doi.org/10.1080/09670262.2017.1376709>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 421. <https://doi.org/10.1186/1471-2105-10-421>
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Molecular Biology and Evolution*, 38(12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., & Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108(supplement_1), 4516–4522. <https://doi.org/10.1073/pnas.1000080107>
- Carpenter, S. R., & Kitchell, J. F. (1988). Consumer Control of Lake Productivity. *BioScience*, 38(11), 764–769. <https://doi.org/10.2307/1310785>
- Caughman, A. M., Pratte, Z. A., Patin, N. V., & Stewart, F. J. (2021). Coral microbiome changes over the day–night cycle. *Coral Reefs*, 40(3), 921–935. <https://doi.org/10.1007/s00338-021-02097-8>
- Chan, Y.-F., Chen, C.-Y., Lu, C.-Y., Tu, Y.-C., Tandon, K., Shikina, S., & Tang, S.-L. (2022). A First Insight Into the Heat-Induced Changes in Proteomic Profiles of the Coral Symbiotic Bacterium *Endozoicomonas montiporae*. *Frontiers in Marine Science*, 9, 808132. <https://doi.org/10.3389/fmars.2022.808132>
- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, 36(6), 1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>

- Chhangawala, S., Rudy, G., Mason, C. E., & Rosenfeld, J. A. (2015). The impact of read length on quantification of differentially expressed genes and splice junction detection. *Genome Biology*, 16(1), 131. <https://doi.org/10.1186/s13059-015-0697-y>
- Chimento, D. P., Mohanty, A. K., Kadner, R. J., & Wiener, M. C. (2003). Substrate-induced transmembrane signaling in the cobalamin transporter BtuB. *Nature Structural & Molecular Biology*, 10(5), 394–401. <https://doi.org/10.1038/nsb914>
- Chiou, Y.-J., Chan, Y.-F., Yu, S.-P., Lu, C.-Y., Hsiao, S. S.-Y., Chiang, P.-W., Hsu, T.-C., Liu, P.-Y., Wada, N., Lee, Y., Jane, W.-N., Lee, D.-C., Huang, Y.-W., & Tang, S.-L. (2023). Similar but different: Characterization of *dddD* gene-mediated DMSP metabolism among coral-associated *Endozoicomonas*. *Science Advances*, 9(47), eadk1910. <https://doi.org/10.1126/sciadv.adk1910>
- Choudoir, M. J., Doroghazi, J. R., & Buckley, D. H. (2016). Latitude delineates patterns of biogeography in terrestrial *Streptomyces*. *Environmental Microbiology*, 18(12), 4931–4945. <https://doi.org/10.1111/1462-2920.13420>
- Clavere Graciette, A. G., Hoopes, L. A., Clauss, T., Stewart, F. J., & Pratte, Z. A. (2023). The microbiome of African penguins (*Spheniscus demersus*) under managed care resembles that of wild marine mammals and birds. *Scientific Reports*, 13(1), 16679. <https://doi.org/10.1038/s41598-023-43899-w>
- Clements, C. S., & Hay, M. E. (2015). Competitors as accomplices: Seaweed competitors hide corals from predatory sea stars. *Proceedings of the Royal Society B: Biological Sciences*, 282(1814), 20150714. <https://doi.org/10.1098/rspb.2015.0714>
- Clements, C. S., & Hay, M. E. (2019). Biodiversity enhances coral growth, tissue survivorship and suppression of macroalgae. *Nature Ecology & Evolution*, 3(2), 178–182. <https://doi.org/10.1038/s41559-018-0752-7>
- Clements, C. S., & Hay, M. E. (2021). Biodiversity has a positive but saturating effect on imperiled coral reefs. *Science Advances*, 7(42), eabi8592. <https://doi.org/10.1126/sciadv.abi8592>
- Clements, C. S., & Hay, M. E. (2023). Disentangling the impacts of macroalgae on corals via effects on their microbiomes. *Frontiers in Ecology and Evolution*, 11, 1083341. <https://doi.org/10.3389/fevo.2023.1083341>
- Clements, C. S., Pratte, Z. A., Stewart, F. J., & Hay, M. E. (2024). Removal of detritivore sea cucumbers from reefs increases coral disease. *Nature Communications*, 15(1), 1338. <https://doi.org/10.1038/s41467-024-45730-0>
- Coase, R. H. (1937). The nature of the firm. *Economica*, 4(16), 386–405.
- Coenen, A. R., Hu, S. K., Luo, E., Muratore, D., & Weitz, J. S. (2020). A Primer for Microbiome Time-Series Analysis. *Frontiers in Genetics*, 11. <https://doi.org/10.3389/fgene.2020.00310>
- Colman, D. R., Keller, L. M., Arteaga-Pozo, E., Andrade-Barahona, E., St. Clair, B., Shoemaker, A., Cox, A., & Boyd, E. S. (2024). Covariation of hot spring geochemistry with microbial genomic diversity, function, and evolution. *Nature Communications*, 15(1), 7506. <https://doi.org/10.1038/s41467-024-51841-5>
- Colman, D. R., Poudel, S., Hamilton, T. L., Havig, J. R., Selensky, M. J., Shock, E. L., & Boyd, E. S. (2018). Geobiological feedbacks and the evolution of thermoacidophiles. *The ISME Journal*, 12(1), 225–236. <https://doi.org/10.1038/ismej.2017.162>

- Copernicus Data Space Ecosystem. (2025). *Sentinel-2 MSI: MultiSpectral Instrument, Level-1C* [Dataset]. European Space Agency.
<https://documentation.dataspace.copernicus.eu/Data/SentinelMissions/Sentinel2.html#sentinel-2-level-1c-top-of-atmosphere-toa>
- Craveiro, N., & Rosa Filho, J. S. (2024). Macroalgae traits and seasonality as drivers of polychaete assemblages on macroalgae of tropical sandstone reefs. *Estuarine, Coastal and Shelf Science*, 297, 108619. <https://doi.org/10.1016/j.ecss.2024.108619>
- Cruaud, P., Vigneron, A., Fradette, M., Charette, S. J., Rodriguez, M. J., Dorea, C. C., & Culley, A. I. (2017). Open the Sterivex™ casing: An easy and effective way to improve DNA extraction yields. *Limnology and Oceanography: Methods*, 15(12), 1015–1020.
<https://doi.org/10.1002/lom3.10221>
- Cullen, A., Pearson, L. A., Mazmouz, R., Liu, T., Soeriyadi, A. H., Ongley, S. E., & Neilan, B. A. (2019). Heterologous expression and biochemical characterisation of cyanotoxin biosynthesis pathways. *Natural Product Reports*, 36(8), 1117–1136.
<https://doi.org/10.1039/C8NP00063H>
- Dai, H., Wu, B., Chen, B., Ma, B., & Chu, C. (2022). Diel Fluctuation of Extracellular Reactive Oxygen Species Production in the Rhizosphere of Rice. *Environmental Science & Technology*, 56(12), 9075–9082. <https://doi.org/10.1021/acs.est.2c00005>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- Datta, M. S., Sliwerska, E., Gore, J., Polz, M. F., & Cordero, O. X. (2016). Microbial interactions lead to rapid micro-scale successions on model marine particles. *Nature Communications*, 7(1), Article 1. <https://doi.org/10.1038/ncomms11965>
- Davidovich, N., Morick, D., & Carella, F. (2020). Mycobacteriosis in Aquatic Invertebrates: A Review of Its Emergence. *Microorganisms*, 8(8), 1249.
<https://doi.org/10.3390/microorganisms8081249>
- Davis, N. M., Proctor, D. M., Holmes, S. P., Relman, D. A., & Callahan, B. J. (2018). Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*, 6(1), 226. <https://doi.org/10.1186/s40168-018-0605-2>
- de Beer, D., Kühl, M., Stambler, N., & Vaki, L. (2000). A microsensor study of light enhanced Ca²⁺ uptake and photosynthesis in the reef-building hermatypic coral *Favia* sp.†. *Marine Ecology Progress Series*, 194, 75–85. <https://doi.org/10.3354/meps194075>
- De Groot, T. R., Menoud, M., Van Bleijswijk, J., Van Leeuwen, S. M., Van Der Molen, J., Hernando-Morales, V., Czerski, H., Maazallahi, H., Walter, S., Rush, D., Röckmann, T., & Niemann, H. (2024). Tidal and seasonal influence on cold seep activity and methanotroph efficiency in the North Sea. *Communications Earth & Environment*, 5(1), 368. <https://doi.org/10.1038/s43247-024-01483-8>
- Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, 6(5), 361–375. <https://doi.org/10.1038/nrg1603>
- DeVantier, L. (2020). Species Richness and Abundance of Reef-Building Corals in the Indo-West Pacific: The Local–Regional Relation Revisited. *Frontiers in Marine Science*, 7.

- Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A. P., & Banfield, J. F. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biology*, 10(8), R85. <https://doi.org/10.1186/gb-2009-10-8-r85>
- Ding, J.-Y., Shiu, J.-H., Chen, W.-M., Chiang, Y.-R., & Tang, S.-L. (2016). Genomic Insight into the Host–Endosymbiont Relationship of *Endozoicomonas montiporae* CL-33T with its Coral Host. *Frontiers in Microbiology*, 7. <https://doi.org/10.3389/fmicb.2016.00251>
- Dittmann, E., Fewer, D. P., & Neilan, B. A. (2013). Cyanobacterial toxins: Biosynthetic routes and evolutionary roots. *FEMS Microbiology Reviews*, 37(1), 23–43. <https://doi.org/10.1111/j.1574-6976.2012.12000.x>
- D’mello, R., Hill, S., & Poole, R. K. (1996). The cytochrome bd quinol oxidase in *Escherichia coli* has an extremely high oxygen affinity and two oxygen-binding haems: Implications for regulation of activity in vivo by oxygen inhibition. *Microbiology*, 142(4), 755–763. <https://doi.org/10.1099/00221287-142-4-755>
- Dong, X., Zhang, T., Wu, W., Peng, Y., Liu, X., Han, Y., Chen, X., Gao, Z., Xia, J., Shao, Z., & Greening, C. (2024). A vast repertoire of secondary metabolites potentially influences community dynamics and biogeochemical processes in cold seeps. *Science Advances*, 10(17), ead12281. <https://doi.org/10.1126/sciadv.adl2281>
- Dragone, N. B., Henley, J. B., Holland-Moritz, H., Diaz, M., Hogg, I. D., Lyons, W. B., Wall, D. H., Adams, B. J., & Fierer, N. (2022). Elevational Constraints on the Composition and Genomic Attributes of Microbial Communities in Antarctic Soils. *mSystems*, 7(1), e01330-21. <https://doi.org/10.1128/msystems.01330-21>
- D’Souza, G., Shitut, S., Preussger, D., Yousif, G., Waschina, S., & Kost, C. (2018). Ecology and evolution of metabolic cross-feeding interactions in bacteria. *Natural Product Reports*, 35(5), 455–488. <https://doi.org/10.1039/C8NP00009C>
- Eiseman, B. W., Silen, B. G., & Kauvar, A. (1958). Fecal enema as an adjunct in the treatment of pseudomembranous enterocolitis. *Surgery*, 44(5), 854–859.
- Ellwood, M., & Nomura, M. (1980). Deletion of a ribosomal ribonucleic acid operon in *Escherichia coli*. *Journal of Bacteriology*, 143(2), 1077–1080. <https://doi.org/10.1128/jb.143.2.1077-1080.1980>
- Enke, T. N., Leventhal, G. E., Metzger, M., Saavedra, J. T., & Cordero, O. X. (2018). Microscale ecology regulates particulate organic matter turnover in model marine microbial communities. *Nature Communications*, 9(1), Article 1. <https://doi.org/10.1038/s41467-018-05159-8>
- Evans, P. N., Parks, D. H., Chadwick, G. L., Robbins, S. J., Orphan, V. J., Golding, S. D., & Tyson, G. W. (2015). Methane metabolism in the archaeal phylum Bathyarchaeota revealed by genome-centric metagenomics. *Science*, 350(6259), 434–438. <https://doi.org/10.1126/science.aac7745>
- Fernandes-Martins, M. C., Colman, D. R., & Boyd, E. S. (2023). Relationships between fluid mixing, biodiversity, and chemosynthetic primary productivity in Yellowstone hot springs. *Environmental Microbiology*, 1462-2920.16340. <https://doi.org/10.1111/1462-2920.16340>
- Fiore, C. L., Labrie, M., Jarett, J. K., & Lesser, M. P. (2015). Transcriptional activity of the giant barrel sponge, *Xestospongia muta* Holobiont: Molecular evidence for metabolic interchange. *Frontiers in Microbiology*, 6. <https://doi.org/10.3389/fmicb.2015.00364>

- Fischer, W. W., Hemp, J., & Johnson, J. E. (2016). Evolution of Oxygenic Photosynthesis. *Annual Review of Earth and Planetary Sciences*, 44(1), 647–683. <https://doi.org/10.1146/annurev-earth-060313-054810>
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Liu, L., Glodek, A., ... Venter, J. C. (1995). Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd. *Science*, 269(5223), 496–512. <https://doi.org/10.1126/science.7542800>
- Fogo, J. K., & Popowsky, M. (1949). Spectrophotometric Determination of Hydrogen Sulfide. *Analytical Chemistry*, 21(6), 732–734. <https://doi.org/10.1021/ac60030a028>
- Ford, A. K., Bejarano, S., Nugues, M. M., Visser, P. M., Albert, S., & Ferse, S. C. A. (2018). Reefs under Siege—The Rise, Putative Drivers, and Consequences of Benthic Cyanobacterial Mats. *Frontiers in Marine Science*, 5, 18. <https://doi.org/10.3389/fmars.2018.00018>
- Fox, J. (2011, June 17). Zombie ideas in ecology. *Oikos Blog*. <http://www.oikosjournal.org/blog/zombie-ideas-ecology>
- Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third edition). SAGE.
- Frias-Lopez, J., Shi, Y., Tyson, G. W., Coleman, M. L., Schuster, S. C., Chisholm, S. W., & DeLong, E. F. (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences*, 105(10), 3805–3810. <https://doi.org/10.1073/pnas.0708897105>
- Fromont, C., Adair, K. L., & Douglas, A. E. (2019). Correlation and causation between the microbiome, *Wolbachia* and host functional traits in natural populations of drosophilid flies. *Molecular Ecology*, 28(7), 1826–1841. <https://doi.org/10.1111/mec.15041>
- Galand, P. E., Ruscheweyh, H.-J., Salazar, G., Hochart, C., Henry, N., Hume, B. C. C., Oliveira, P. H., Perdereau, A., Labadie, K., Belser, C., Boissin, E., Romac, S., Poulain, J., Bourdin, G., Iwankow, G., Moulin, C., Armstrong, E. J., Paz-García, D. A., Ziegler, M., ... Planes, S. (2023). Diversity of the Pacific Ocean coral reef microbiome. *Nature Communications*, 14(1), 3039. <https://doi.org/10.1038/s41467-023-38500-x>
- Galla, G., Praeg, N., Rzehak, T., Sprecher, E., Colla, F., Seeber, J., Illmer, P., & Hauße, H. C. (2024). Comparison of DNA extraction methods on different sample matrices within the same terrestrial ecosystem. *Scientific Reports*, 14(1), 8715. <https://doi.org/10.1038/s41598-024-59086-4>
- Garber, A. I., Nealson, K. H., Okamoto, A., McAllister, S. M., Chan, C. S., Barco, R. A., & Merino, N. (2020). FeGenie: A Comprehensive Tool for the Identification of Iron Genes and Iron Gene Neighborhoods in Genome and Metagenome Assemblies. *Frontiers in Microbiology*, 11, 37. <https://doi.org/10.3389/fmicb.2020.00037>
- Garren, M., & Azam, F. (2012). New directions in coral reef microbial ecology. *Environmental Microbiology*, 14(4), 833–844. <https://doi.org/10.1111/j.1462-2920.2011.02597.x>
- Giakoumi, S., Scianna, C., Plass-Johnson, J., Micheli, F., Grorud-Colvert, K., Thiriet, P., Claudet, J., Di Carlo, G., Di Franco, A., Gaines, S. D., García-Charton, J. A., Lubchenco, J., Reimer, J., Sala, E., & Guidetti, P. (2017). Ecological effects of full and partial protection in the crowded Mediterranean Sea: A regional meta-analysis. *Scientific Reports*, 7(1), 8940. <https://doi.org/10.1038/s41598-017-08850-w>

- Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., & Charnov, E. L. (2001). Effects of Size and Temperature on Metabolic Rate. *Science*, 293(5538), 2248–2251. <https://doi.org/10.1126/science.1061967>
- Giovannoni, S. J., DeLong, E. F., Olsen, G. J., & Pace, N. R. (1988). Phylogenetic group-specific oligodeoxynucleotide probes for identification of single microbial cells. *Journal of Bacteriology*, 170(2), 720–726. <https://doi.org/10.1128/jb.170.2.720-726.1988>
- Glas, M. S., Motti, C. A., Negri, A. P., Sato, Y., Froschio, S., Humpage, A. R., Krock, B., Cembella, A., & Bourne, D. G. (2010). Cyanotoxins are not implicated in the etiology of coral black band disease outbreaks on Pelorus Island, Great Barrier Reef. *FEMS Microbiology Ecology*, 73(1), 43–54. <https://doi.org/10.1111/j.1574-6941.2010.00874.x>
- Glasl, B., Herndl, G. J., & Frade, P. R. (2016). The microbiome of coral surface mucus has a key role in mediating holobiont health and survival upon disturbance. *The ISME Journal*, 10(9), 2280–2292. <https://doi.org/10.1038/ismej.2016.9>
- Glöckner, F. O., Yilmaz, P., Quast, C., Gerken, J., Beccati, A., Ciuprina, A., Bruns, G., Yarza, P., Peplies, J., Westram, R., & Ludwig, W. (2017). 25 years of serving the community with ribosomal RNA gene reference databases and tools. *Journal of Biotechnology*, 261, 169–176. <https://doi.org/10.1016/j.jbiotec.2017.06.1198>
- Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V., & Egozcue, J. J. (2017). Microbiome Datasets Are Compositional: And This Is Not Optional. *Frontiers in Microbiology*, 8, 2224. <https://doi.org/10.3389/fmicb.2017.02224>
- Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., & Riggs, A. D. (1979). Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences*, 76(1), 106–110. <https://doi.org/10.1073/pnas.76.1.106>
- Goodacre, N. F., Gerloff, D. L., & Uetz, P. (2014). Protein Domains of Unknown Function Are Essential in Bacteria. *mBio*, 5(1), e00744-13. <https://doi.org/10.1128/mBio.00744-13>
- Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., & Tiedje, J. M. (2007). DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *International Journal of Systematic and Evolutionary Microbiology*, 57(1), 81–91. <https://doi.org/10.1099/ijs.0.64483-0>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., Di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., ... Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- Graf, J. S., Schorn, S., Kitzinger, K., Ahmerkamp, S., Woehle, C., Huettel, B., Schubert, C. J., Kuypers, M. M. M., & Milucka, J. (2021). Anaerobic endosymbiont generates energy for ciliate host by denitrification. *Nature*, 591(7850), 445–450. <https://doi.org/10.1038/s41586-021-03297-6>
- Grayson, N., Clements, C. S., Towner, A. A., Beatty, D. S., & Hay, M. E. (2022). Did the historic overharvesting of sea cucumbers make coral more susceptible to pathogens? *Coral Reefs*. <https://doi.org/10.1007/s00338-022-02227-w>

- Gruber-Vodicka, H. R., Seah, B. K. B., & Pruesse, E. (2020). phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems*, 5(5), e00920-20. <https://doi.org/10.1128/mSystems.00920-20>
- Gutiérrez, J. L., & Jones, C. G. (2006). Physical Ecosystem Engineers as Agents of Biogeochemical Heterogeneity. *BioScience*, 56(3), 227–236. [https://doi.org/10.1641/0006-3568\(2006\)056\[0227:PEEAAO\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)056[0227:PEEAAO]2.0.CO;2)
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., ... Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, 8(8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>
- Hamilton, T. L., Boyd, E. S., & Peters, J. W. (2011). Environmental Constraints Underpin the Distribution and Phylogenetic Diversity of nifH in the Yellowstone Geothermal Complex. *Microbial Ecology*, 61(4), 860–870. <https://doi.org/10.1007/s00248-011-9824-9>
- Harborne, A. R., Rogers, A., Bozec, Y.-M., & Mumby, P. J. (2017). Multiple Stressors and the Functioning of Coral Reefs. *Annual Review of Marine Science*, 9(1), 445–468. <https://doi.org/10.1146/annurev-marine-010816-060551>
- Hatton, A. D., Shenoy, D. M., Hart, M. C., Mogg, A., & Green, D. H. (2012). Metabolism of DMSP, DMS and DMSO by the cultivable bacterial community associated with the DMSP-producing dinoflagellate *Scrippsiella trochoidea*. *Biogeochemistry*, 110(1–3), 131–146. <https://doi.org/10.1007/s10533-012-9702-7>
- Hauptfeld, E., Pappas, N., Van Iwaarden, S., Snoek, B. L., Aldas-Vargas, A., Dutilh, B. E., & Von Meijenfeldt, F. A. B. (2024). Integrating taxonomic signals from MAGs and contigs improves read annotation and taxonomic profiling of metagenomes. *Nature Communications*, 15(1), 3373. <https://doi.org/10.1038/s41467-024-47155-1>
- Hazelett, D. J. (2025). An open letter to graduate students and other procrastinators: It's time to write. *Nature Biotechnology*, 43(3), 447–450. <https://doi.org/10.1038/s41587-025-02584-1>
- Heath-Heckman, E. A. C. (2016). The Metronome of Symbiosis: Interactions Between Microbes and the Host Circadian Clock. *Integrative and Comparative Biology*, 56(5), 776–783. <https://doi.org/10.1093/icb/icw067>
- Henriques, W. S., Young, J. M., Nemudryi, A., Nemudraia, A., Wiedenheft, B., & Malik, H. S. (2024). The Diverse Evolutionary Histories of Domesticated Metaviral Capsid Genes in Mammals. *Molecular Biology and Evolution*, 41(4), msae061. <https://doi.org/10.1093/molbev/msae061>
- Hernandez-Agreda, A., Leggat, W., Bongaerts, P., Herrera, C., & Ainsworth, T. D. (2018). Rethinking the Coral Microbiome: Simplicity Exists within a Diverse Microbial Biosphere. *mBio*, 9(5), e00812-18. <https://doi.org/10.1128/mBio.00812-18>
- Herron, P., Gage, D., Arango Pinedo, C., Haider, Z., & Cardon, Z. (2013). Better to light a candle than curse the darkness: Illuminating spatial localization and temporal dynamics of rapid microbial growth in the rhizosphere. *Frontiers in Plant Science*, 4. <https://www.frontiersin.org/articles/10.3389/fpls.2013.00323>

- Hilleman, M. R. (1998). Six decades of vaccine development—A personal history. *Nature Medicine*, 4(S5), 507–514. <https://doi.org/10.1038/nm0598supp-507>
- HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M., & Mayfield, M. M. (2012). Rethinking Community Assembly through the Lens of Coexistence Theory. *Annual Review of Ecology, Evolution, and Systematics*, 43(1), 227–248. <https://doi.org/10.1146/annurev-ecolsys-110411-160411>
- Hixon, M. A. (2015). Reef Fishes, Seaweeds, and Corals: A Complex Triangle. In *Coral Reefs in the Anthropocene*. Springer Netherlands. <https://doi.org/10.1007/978-94-017-7249-5>
- Hoegh-Guldberg, O., Poloczanska, E. S., Skirving, W., & Dove, S. (2017). Coral Reef Ecosystems under Climate Change and Ocean Acidification. *Frontiers in Marine Science*, 4, 158. <https://doi.org/10.3389/fmars.2017.00158>
- Hovenkamp, H. (2010). The Law of Vertical Integration and the Business Firm: 1880-1960. *Iowa Law Review*, 95, 863–918.
- Huang, S., Fu, Y., Zhang, H., Wang, C., Zou, C., & Lu, X. (2023). Research progress of novel bio-denitrification technology in deep wastewater treatment. *Frontiers in Microbiology*, 14, 1284369. <https://doi.org/10.3389/fmicb.2023.1284369>
- Hubbard, C. J., Brock, M. T., van Diepen, L. T., Maignien, L., Ewers, B. E., & Weinig, C. (2018). The plant circadian clock influences rhizosphere community structure and function. *The ISME Journal*, 12(2), Article 2. <https://doi.org/10.1038/ismej.2017.172>
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8), 2115–2122. <https://doi.org/10.1093/molbev/msx148>
- Hug, L. A., Baker, B. J., Anantharaman, K., Brown, C. T., Probst, A. J., Castelle, C. J., Butterfield, C. N., Hernsdorf, A. W., Amano, Y., Ise, K., Suzuki, Y., Dudek, N., Relman, D. A., Finstad, K. M., Amundson, R., Thomas, B. C., & Banfield, J. F. (2016). A new view of the tree of life. *Nature Microbiology*, 1(5), 16048. <https://doi.org/10.1038/nmicrobiol.2016.48>
- Hughes, T. P., Anderson, K. D., Connolly, S. R., Heron, S. F., Kerry, J. T., Lough, J. M., Baird, A. H., Baum, J. K., Berumen, M. L., Bridge, T. C., Claar, D. C., Eakin, C. M., Gilmour, J. P., Graham, N. A. J., Harrison, H., Hobbs, J.-P. A., Hoey, A. S., Hoogenboom, M., Lowe, R. J., ... Wilson, S. K. (2018). Spatial and temporal patterns of mass bleaching of corals in the Anthropocene. *Science*, 359(6371), 80–83. <https://doi.org/10.1126/science.aan8048>
- Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, 16(8), 471–483. <https://doi.org/10.1038/s41579-018-0040-1>
- Hutchings, M. I., Truman, A. W., & Wilkinson, B. (2019). Antibiotics: Past, present and future. *Current Opinion in Microbiology*, 51, 72–80. <https://doi.org/10.1016/j.mib.2019.10.008>
- Hwang, Y., Cornman, A. L., Kellogg, E. H., Ovchinnikov, S., & Girguis, P. R. (2024). Genomic language model predicts protein co-regulation and function. *Nature Communications*, 15(1), 2880. <https://doi.org/10.1038/s41467-024-46947-9>
- Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11(1), 119. <https://doi.org/10.1186/1471-2105-11-119>

- Ide, K., Nishikawa, Y., Maruyama, T., Tsukada, Y., Kogawa, M., Takeda, H., Ito, H., Wagatsuma, R., Miyaoka, R., Nakano, Y., Kinjo, K., Ito, M., Hosokawa, M., Yura, K., Suda, S., & Takeyama, H. (2022). Targeted single-cell genomics reveals novel host adaptation strategies of the symbiotic bacteria *Endozoicomonas* in *Acropora tenuis* coral. *Microbiome*, 10(1), 220. <https://doi.org/10.1186/s40168-022-01395-9>
- International Human Genome Sequencing Consortium, Whitehead Institute for Biomedical Research, Center for Genome Research, Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., ... Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409(6822), 860–921. <https://doi.org/10.1038/35057062>
- Ishoey, T., Woyke, T., Stepanauskas, R., Novotny, M., & Lasken, R. S. (2008). Genomic sequencing of single microbial cells from environmental samples. *Current Opinion in Microbiology*, 11(3), 198–204. <https://doi.org/10.1016/j.mib.2008.05.006>
- Jackson, R. L., Gabric, A. J., Cropp, R., & Woodhouse, M. T. (2020). Dimethylsulfide (DMS), marine biogenic aerosols and the ecophysiology of coral reefs. *Biogeosciences*, 17(8), 2181–2204. <https://doi.org/10.5194/bg-17-2181-2020>
- Jacobs, J. M., Stine, C. B., Baya, A. M., & Kent, M. L. (2009). A review of mycobacteriosis in marine fish. *Journal of Fish Diseases*, 32(2), 119–130. <https://doi.org/10.1111/j.1365-2761.2008.01016.x>
- Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, 9(1), 5114. <https://doi.org/10.1038/s41467-018-07641-9>
- Jimenez, I. M., K hl, M., Larkum, A. W. D., & Ralph, P. J. (2011). Effects of flow and colony morphology on the thermal boundary layer of corals. *Journal of The Royal Society Interface*, 8(65), 1785–1795. <https://doi.org/10.1098/rsif.2011.0144>
- J rgensen, B. B. (1977). The sulfur cycle of a coastal marine sediment (Limfjorden, Denmark)1. *Limnology and Oceanography*, 22(5), 814–832. <https://doi.org/10.4319/lo.1977.22.5.0814>
- J rgensen, B. B., Revsbech, N. P., Blackburn, T. H., & Cohen, Y. (1979). Diurnal Cycle of Oxygen and Sulfide Microgradients and Microbial Photosynthesis in a Cyanobacterial Mat Sediment. *Applied and Environmental Microbiology*, 38(1), 46–58. <https://doi.org/10.1128/aem.38.1.46-58.1979>
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R.,  idek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
- Kalvelage, T., Lavik, G., Jensen, M. M., Revsbech, N. P., L scher, C., Schunck, H., Desai, D. K., Hauss, H., Kiko, R., Holtappels, M., LaRoche, J., Schmitz, R. A., Graco, M. I., & Kuypers, M. M. M. (2015). Aerobic Microbial Respiration In Oceanic Oxygen Minimum Zones. *PLOS ONE*, 10(7), e0133526. <https://doi.org/10.1371/journal.pone.0133526>

- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359. <https://doi.org/10.7717/peerj.7359>
- Kang, J., Mohamed, H. F., Liu, X., Pei, L., Huang, S., Lin, X., Zheng, X., & Luo, Z. (2022). Combined Culture and DNA Metabarcoding Analysis of Cyanobacterial Community Structure in Response to Coral Reef Health Status in the South China Sea. *Journal of Marine Science and Engineering*, 10(12), Article 12. <https://doi.org/10.3390/jmse10121984>
- Kapli, P., Kotari, I., Telford, M. J., Goldman, N., & Yang, Z. (2023). DNA Sequences Are as Useful as Protein Sequences for Inferring Deep Phylogenies. *Systematic Biology*, 72(5), 1119–1135. <https://doi.org/10.1093/sysbio/syad036>
- Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, 21(7), 428–444. <https://doi.org/10.1038/s41576-020-0233-0>
- Katharios, P., Seth-Smith, H. M. B., Fehr, A., Mateos, J. M., Qi, W., Richter, D., Nufer, L., Ruetten, M., Guevara Soto, M., Ziegler, U., Thomson, N. R., Schlapbach, R., & Vaughan, L. (2015). Environmental marine pathogen isolation using mesocosm culture of sharpsnout seabream: Striking genomic and morphological features of novel *Endozoicomonas* sp. *Scientific Reports*, 5(1), 17609. <https://doi.org/10.1038/srep17609>
- Katoh, K. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Kegler, P., Kegler, H. F., Gärdes, A., Ferse, S. C. A., Lukman, M., Alfiansah, Y. R., Hassenrück, C., & Kunzmann, A. (2017). Bacterial Biofilm Communities and Coral Larvae Settlement at Different Levels of Anthropogenic Impact in the Spermonde Archipelago, Indonesia. *Frontiers in Marine Science*, 4, 270. <https://doi.org/10.3389/fmars.2017.00270>
- Keller, L. M., Colman, D. R., & Boyd, E. S. (2025). Simultaneous aerobic and anaerobic respiration in hot spring chemolithotrophic bacteria. *Nature Communications*, 16(1), 1063. <https://doi.org/10.1038/s41467-025-56418-4>
- Kellmann, R., Mihali, T. K., Jeon, Y. J., Pickford, R., Pomati, F., & Neilan, B. A. (2008). Biosynthetic Intermediate Analysis and Functional Homology Reveal a Saxitoxin Gene Cluster in Cyanobacteria. *Applied and Environmental Microbiology*, 74(13), 4044–4053. <https://doi.org/10.1128/AEM.00353-08>
- Kellogg, C., & Pratte, Z. (2021). Unexpected diversity of *Endozoicomonas* in deep-sea corals. *Marine Ecology Progress Series*, 673, 1–15. <https://doi.org/10.3354/meps13844>
- Kelly, L. T., Puddick, J., Ryan, K. G., Champeau, O., & Wood, S. A. (2020). An ecotoxicological assessment of the acute toxicity of anatoxin congeners on New Zealand *Deleatidium* species (mayflies). *Inland Waters*, 10(1), 101–108. <https://doi.org/10.1080/20442041.2019.1626151>
- Kelly, L. W., Nelson, C. E., Haas, A. F., Naliboff, D. S., Calhoun, S., Carlson, C. A., Edwards, R. A., Fox, M. D., Hatay, M., Johnson, M. D., Kelly, E. L. A., Lim, Y. W., Macherla, S., Quinlan, Z. A., Silva, G. G. Z., Vermeij, M. J. A., Zgliczynski, B., Sandin, S. A., Smith, J. E., & Rohwer, F. (2019). Diel population and functional synchrony of microbial communities on coral reefs. *Nature Communications*, 10(1), Article 1. <https://doi.org/10.1038/s41467-019-09419-z>

- Kieft, K., Adams, A., Salamzade, R., Kalan, L., & Anantharaman, K. (2022). vRhyme enables binning of viral genomes from metagenomes. *Nucleic Acids Research*, 50(14), e83–e83. <https://doi.org/10.1093/nar/gkac341>
- Kim, H. J., Lee, H., Lee, Y., Choi, I., Ko, Y., Lee, S., & Jang, S. (2020). The ThiL enzyme is a valid antibacterial target essential for both thiamine biosynthesis and salvage pathways in *Pseudomonas aeruginosa*. *Journal of Biological Chemistry*, 295(29), 10081–10091. <https://doi.org/10.1074/jbc.RA120.013295>
- Kirsip, H., & Abroi, A. (2019). Protein Structure-Guided Hidden Markov Models (HMMs) as A Powerful Method in the Detection of Ancestral Endogenous Viral Elements. *Viruses*, 11(4), 320. <https://doi.org/10.3390/v11040320>
- Kitzinger, K., Marchant, H. K., Bristow, L. A., Herbold, C. W., Padilla, C. C., Kidane, A. T., Littmann, S., Daims, H., Pjevac, P., Stewart, F. J., Wagner, M., & Kuypers, M. M. M. (2020). Single cell analyses reveal contrasting life strategies of the two main nitrifiers in the ocean. *Nature Communications*, 11(1), 767. <https://doi.org/10.1038/s41467-020-14542-3>
- Klinges, J. G., Rosales, S. M., McMinds, R., Shaver, E. C., Shantz, A. A., Peters, E. C., Eitel, M., Wörheide, G., Sharp, K. H., Burkepile, D. E., Silliman, B. R., & Vega Thurber, R. L. (2019). Phylogenetic, genomic, and biogeographic characterization of a novel and ubiquitous marine invertebrate-associated Rickettsiales parasite, *Candidatus Aquarickettsia rohweri*, gen. Nov., sp. Nov. *The ISME Journal*, 13(12), 2938–2953. <https://doi.org/10.1038/s41396-019-0482-0>
- Kohtz, A. J., Petrosian, N., Krukenberg, V., Jay, Z. J., Pilhofer, M., & Hatzenpichler, R. (2024). Cultivation and visualization of a methanogen of the phylum Thermoproteota. *Nature*, 632(8027), 1118–1123. <https://doi.org/10.1038/s41586-024-07631-6>
- Krebs, C. J. (2015). One hundred years of population ecology: Successes, failures and the road ahead. *Integrative Zoology*, 10(3), 233–240. <https://doi.org/10.1111/1749-4877.12130>
- Kristensen, E. (2008). Mangrove crabs as ecosystem engineers; with emphasis on sediment processes. *Journal of Sea Research*, 59(1–2), 30–43. <https://doi.org/10.1016/j.seares.2007.05.004>
- Kuba, G. M., Spalding, H. L., Hill-Spanik, K. M., & Fullerton, H. (2021). Microbiota-Macroalgal Relationships at a Hawaiian Intertidal Bench Are Influenced by Macroalgal Phyla and Associated Thallus Complexity. *mSphere*, 6(5), e00665-21. <https://doi.org/10.1128/mSphere.00665-21>
- Kühl, M., Cohen, Y., Dalsgaard, T., Jørgensen, B. B., & Revsbech, N. P. (1995). Microenvironment and photosynthesis of zooxanthellae in scleractinian corals studied with microsensors for O₂, pH and light. *Marine Ecology-Progress Series*, 117(1–3), 159–172. <https://doi.org/10.3354/meps117159>
- Kumagai, Y., Yoshizawa, S., Nakajima, Y., Watanabe, M., Fukunaga, T., Ogura, Y., Hayashi, T., Oshima, K., Hattori, M., Ikeuchi, M., Kogure, K., DeLong, E. F., & Iwasaki, W. (2018). Solar-panel and parasol strategies shape the proteorhodopsin distribution pattern in marine Flavobacteriia. *The ISME Journal*, 12(5), 1329–1343. <https://doi.org/10.1038/s41396-018-0058-4>

- Kurihara, H., Ikeda, N., & Umezawa, Y. (2018). Diurnal and seasonal variation of particle and dissolved organic matter release by the coral *Acropora tenuis*. *PeerJ*, 6, e5728. <https://doi.org/10.7717/peerj.5728>
- LaJeunesse, T. C. (2020). Zooxanthellae. *Current Biology*, 30(19), R1110–R1113. <https://doi.org/10.1016/j.cub.2020.03.058>
- LaJeunesse, T. C., Parkinson, J. E., Gabrielson, P. W., Jeong, H. J., Reimer, J. D., Voolstra, C. R., & Santos, S. R. (2018). Systematic Revision of Symbiodiniaceae Highlights the Antiquity and Diversity of Coral Endosymbionts. *Current Biology*, 28(16), 2570–2580.e6. <https://doi.org/10.1016/j.cub.2018.07.008>
- Lane, D. J., Pace, B., Olsen, G. J., STAHLt, D. A., SOGINt, M. L., & Pace, N. R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc. Natl. Acad. Sci. USA*.
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lee, C.-G., Huettel, M., Hong, J.-S., & Reise, K. (2004). Carrion-feeding on the sediment surface at nocturnal low tides by the polychaete *Phyllodoce mucosa*. *Marine Biology*, 145(3). <https://doi.org/10.1007/s00227-004-1334-6>
- Lee, P. A., & De Mora, S. J. (1999). INTRACELLULAR DIMETHYLSULFOXIDE (DMSO) IN UNICELLULAR MARINE ALGAE: SPECULATIONS ON ITS ORIGIN AND POSSIBLE BIOLOGICAL ROLE. *Journal of Phycology*, 35(1), 8–18. <https://doi.org/10.1046/j.1529-8817.1999.3510008.x>
- Lee, S., Ford, A. K., Mangubhai, S., Wild, C., & Ferse, S. C. A. (2018). Effects of sandfish (*Holothuria scabra*) removal on shallow-water sediments in Fiji. *PeerJ*, 6, e4773. <https://doi.org/10.7717/peerj.4773>
- Lee, S. T. M., Davy, S. K., Tang, S.-L., Fan, T.-Y., & Kench, P. S. (2015). Successive shifts in the microbial community of the surface mucus layer and tissues of the coral *Acropora muricata* under thermal stress. *FEMS Microbiology Ecology*, 91(12), fiv142. <https://doi.org/10.1093/femsec/fiv142>
- Lenth, R. V. (2023). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. <https://CRAN.R-project.org/package=emmeans>
- Leopold, A. (1949). *A Sand County Almanac*. Oxford University Press.
- Leung, P. M., & Greening, C. (2020). *Greening lab metabolic marker gene databases* [Dataset]. *Greening lab metabolic marker gene databases*.
- Li, B., & Dewey, C. N. (2011). RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1), 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics*, 31(10), 1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

- Li, W., Baliu-Rodriguez, D., Premathilaka, S. H., Thenuwara, S. I., Kimbrel, J. A., Samo, T. J., Ramon, C., Kiledal, E. A., Rivera, S. R., Kharbush, J., Isailovic, D., Weber, P. K., Dick, G. J., & Mayali, X. (2024). Microbiome processing of organic nitrogen input supports growth and cyanotoxin production of *Microcystis aeruginosa* cultures. *The ISME Journal*, 18(1), wrac082. <https://doi.org/10.1093/ismejo/wrac082>
- Li, X., Li, J., Qu, C., Yu, T., & Du, M. (2021). Bioremediation of clay with high oil content and biological response after restoration. *Scientific Reports*, 11(1), 9725. <https://doi.org/10.1038/s41598-021-88033-w>
- Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H., Gan, J., Li, N., Hu, X., Liu, B., Yang, B., & Fan, W. (2012). Comparison of the two major classes of assembly algorithms: Overlap-layout-consensus and de-bruijn-graph. *Briefings in Functional Genomics*, 11(1), 25–37. <https://doi.org/10.1093/bfpg/elr035>
- Li, Z., Zhu, X., Wu, Z., Sun, T., & Tong, Y. (2023). Recent Advances in Cyanotoxin Synthesis and Applications: A Comprehensive Review. *Microorganisms*, 11(11), 2636. <https://doi.org/10.3390/microorganisms11112636>
- Likens, G. E., Bormann, F. H., Johnson, N. M., Fisher, D. W., & Pierce, R. S. (1970). Effects of Forest Cutting and Herbicide Treatment on Nutrient Budgets in the Hubbard Brook Watershed-Ecosystem. *Ecological Monographs*, 40(1), 23–47. <https://doi.org/10.2307/1942440>
- Lim, S. L., Chin, C.-H., Chiou, Y.-J., Hsu, M.-T., Chiang, P.-W., Chen, H.-J., Tu, Y.-C., & Tang, S.-L. (2025). *Unusual Ecofunctional Traits of Endozoicomonas: A Pan-Genomic Perspective*. <https://doi.org/10.1101/2025.03.06.641946>
- Lischer, H. E. L., & Shimizu, K. K. (2017). Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC Bioinformatics*, 18(1), 474. <https://doi.org/10.1186/s12859-017-1911-6>
- Littler, M. M., & Littler, D. S. (2013). The Nature of Macroalgae and Their Interactions on Reefs. In *Research and Discoveries: The Revolution of Science through Scuba* (Vol. 39). Smithsonian Institution Scholarly Press. <https://doi.org/10.5479/si.1943667X.39>
- Liu, X., Zhao, B., Zheng, H.-J., Hu, Y., Lu, G., Yang, C.-Q., Chen, J.-D., Chen, J.-J., Chen, D.-Y., Zhang, L., Zhou, Y., Wang, L.-J., Guo, W.-Z., Bai, Y.-L., Ruan, J.-X., Shangguan, X.-X., Mao, Y.-B., Shan, C.-M., Jiang, J.-P., ... Chen, X.-Y. (2015). *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. *Scientific Reports*, 5(1), 14139. <https://doi.org/10.1038/srep14139>
- López-Pérez, M., Haro-Moreno, J. M., Iranzo, J., & Rodriguez-Valera, F. (2020). Genomes of the “*Candidatus Actinomarinales*” Order: Highly Streamlined Marine Epipelagic Actinobacteria. *mSystems*, 5(6), e01041-20. <https://doi.org/10.1128/mSystems.01041-20>
- Louca, S., Hawley, A. K., Katsev, S., Torres-Beltran, M., Bhatia, M. P., Kheirandish, S., Michiels, C. C., Capelle, D., Lavik, G., Doebeli, M., Crowe, S. A., & Hallam, S. J. (2016). Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. *Proceedings of the National Academy of Sciences*, 113(40). <https://doi.org/10.1073/pnas.1602897113>

- Lovin, L. M., & Brooks, B. W. (2019). Global scanning of anatoxins in aquatic systems: Environment and health hazards, and research needs. *Marine and Freshwater Research*, 71(5), 689–700.
- Lu, D.-C., Wang, F.-Q., Amann, R. I., Teeling, H., & Du, Z.-J. (2023). Epiphytic common core bacteria in the microbiomes of co-located green (*Ulva*), brown (*Saccharina*) and red (*Grateloupia*, *Gelidium*) macroalgae. *Microbiome*, 11(1), 126. <https://doi.org/10.1186/s40168-023-01559-1>
- Lui, L. M., Majumder, E. L.-W., Smith, H. J., Carlson, H. K., Von Netzer, F., Fields, M. W., Stahl, D. A., Zhou, J., Hazen, T. C., Baliga, N. S., Adams, P. D., & Arkin, A. P. (2021). Mechanism Across Scales: A Holistic Modeling Framework Integrating Laboratory and Field Studies for Microbial Ecology. *Frontiers in Microbiology*, 12, 642422. <https://doi.org/10.3389/fmicb.2021.642422>
- Ma, Z. Y., Zhang, X. M., Wang, R., Wang, M., Liu, T., & Tan, Z. L. (2020). Effects of Chemical and Mechanical Lysis on Microbial DNA Yield, Integrity, and Downstream Amplicon Sequencing of Rumen Bacteria and Protozoa. *Frontiers in Microbiology*, 11, 581227. <https://doi.org/10.3389/fmicb.2020.581227>
- MacArthur, R. H., & Wilson, E. O. (1963). An Equilibrium Theory of Insular Zoogeography. *Evolution*, 17(4), 373. <https://doi.org/10.2307/2407089>
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2022). *cluster: Cluster Analysis Basics and Extensions*. <https://CRAN.R-project.org/package=cluster>
- Maguire, F., Jia, B., Gray, K. L., Lau, W. Y. V., Beiko, R. G., & Brinkman, F. S. L. (2020). Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic Islands. *Microbial Genomics*, 6(10). <https://doi.org/10.1099/mgen.0.000436>
- Maire, J., Tandon, K., Collingro, A., Van De Meene, A., Damjanovic, K., Gotze, C. R., Stephenson, S., Philip, G. K., Horn, M., Cantin, N. E., Blackall, L. L., & Van Oppen, M. J. H. (2023). Colocalization and potential interactions of *Endozoicomonas* and chlamydiae in microbial aggregates of the coral *Pocillopora acuta*. *Science Advances*, 9(20), eadg0773. <https://doi.org/10.1126/sciadv.adg0773>
- Major, P., Embley, T. M., & Williams, T. A. (2017). Phylogenetic Diversity of NTT Nucleotide Transport Proteins in Free-Living and Parasitic Bacteria and Eukaryotes. *Genome Biology and Evolution*, 9(2), 480–487. <https://doi.org/10.1093/gbe/evx015>
- Malik, A. A., Martiny, J. B. H., Brodie, E. L., Martiny, A. C., Treseder, K. K., & Allison, S. D. (2020). Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *The ISME Journal*, 14(1), Article 1. <https://doi.org/10.1038/s41396-019-0510-0>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- Marcelino, V. R., Clausen, P. T. L. C., Buchmann, J. P., Wille, M., Iredell, J. R., Meyer, W., Lund, O., Sorrell, T. C., & Holmes, E. C. (2020). CCMetagen: Comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data. *Genome Biology*, 21(1), 103. <https://doi.org/10.1186/s13059-020-02014-2>

- Mark Welch, J. L., Rossetti, B. J., Rieken, C. W., Dewhirst, F. E., & Borisy, G. G. (2016). Biogeography of a human oral microbiome at the micron scale. *Proceedings of the National Academy of Sciences*, 113(6). <https://doi.org/10.1073/pnas.1522149113>
- Marquet, P. A., Allen, A. P., Brown, J. H., Dunne, J. A., Enquist, B. J., Gillooly, J. F., Gowaty, P. A., Green, J. L., Harte, J., Hubbell, S. P., O'Dwyer, J., Okie, J. G., Ostling, A., Ritchie, M., Storch, D., & West, G. B. (2014). On Theory in Ecology. *BioScience*, 64(8), 701–710. <https://doi.org/10.1093/biosci/biu098>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, 17(1), Article 1. <https://doi.org/10.14806/ej.17.1.200>
- Martinez-Gutierrez, C. A., & Aylward, F. O. (2021). Phylogenetic Signal, Congruence, and Uncertainty across Bacteria and Archaea. *Molecular Biology and Evolution*, 38(12), 5514–5527. <https://doi.org/10.1093/molbev/msab254>
- Martínez-Pérez, C., Greening, C., Bay, S. K., Lappan, R. J., Zhao, Z., De Corte, D., Hulbe, C., Ohneiser, C., Stevens, C., Thomson, B., Stepanauskas, R., González, J. M., Logares, R., Herndl, G. J., Morales, S. E., & Baltar, F. (2022). Phylogenetically and functionally diverse microorganisms reside under the Ross Ice Shelf. *Nature Communications*, 13(1), 117. <https://doi.org/10.1038/s41467-021-27769-5>
- Martins, F. H., Rosay, T., Rajan, A., Carter, H. E., Turocy, T., Mejia, A., Crawford, J. M., Maresso, A. W., & Sperandio, V. (2024). Enterococcus faecalis-derived adenine enhances enterohaemorrhagic Escherichia coli Type 3 Secretion System-dependent virulence. *Nature Microbiology*, 9(9), 2448–2461. <https://doi.org/10.1038/s41564-024-01747-1>
- Masdeu-Navarro, M., Mangot, J.-F., Xue, L., Cabrera-Brufau, M., Gardner, S. G., Kieber, D. J., González, J. M., & Simó, R. (2022). Spatial and diel patterns of volatile organic compounds, DMSP-derived compounds, and planktonic microorganisms around a tropical scleractinian coral colony. *Frontiers in Marine Science*, 9, 944141. <https://doi.org/10.3389/fmars.2022.944141>
- McConnaughey, T. (2012). Zooxanthellae that open calcium channels: Implications for reef corals. *Marine Ecology Progress Series*, 460, 277–287. <https://doi.org/10.3354/meps09776>
- McCutcheon, J. P., & Moran, N. A. (2012). Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology*, 10(1), 13–26. <https://doi.org/10.1038/nrmicro2670>
- McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE*, 8(4), e61217.
- Meyer, J. L., Gunasekera, S. P., Brown, A. L., Ding, Y., Miller, S., Teplitski, M., & Paul, V. J. (2023). Cryptic Diversity of Black Band Disease Cyanobacteria in Siderastrea siderea Corals Revealed by Chemical Ecology and Comparative Genome-Resolved Metagenomics. *Marine Drugs*, 21(2), 76. <https://doi.org/10.3390/md21020076>
- Meyer, J. L., Paul, V. J., & Teplitski, M. (2014). Community Shifts in the Surface Microbiomes of the Coral Porites astreoides with Unusual Lesions. *PLoS ONE*, 9(6), e100316. <https://doi.org/10.1371/journal.pone.0100316>
- Meziti, A., Rodriguez-R, L. M., Hatt, J. K., Peña-Gonzalez, A., Levy, K., & Konstantinidis, K. T. (2021). The Reliability of Metagenome-Assembled Genomes (MAGs) in Representing Natural Populations: Insights from Comparing MAGs against Isolate Genomes Derived

- from the Same Fecal Sample. *Applied and Environmental Microbiology*, 87(6), e02593-20. <https://doi.org/10.1128/AEM.02593-20>
- Michael, T. C., Costello, D. M., Fitzgibbon, A. S., & Kinsman-Costello, L. E. (2023). Invertebrate Activities in Wetland Sediments Influence Oxygen and Nutrient Dynamics at the Sediment-water Interface. *Wetlands*, 43(8), 96. <https://doi.org/10.1007/s13157-023-01737-9>
- Michio, K., Kengo, K., Yasunori, K., Hitoshi, M., Takayuki, Y., Hideaki, Y., & Hiroshi, S. (2003). Effects of deposit feeder *Stichopus japonicus* on algal bloom and organic matter contents of bottom sediments of the enclosed sea. *Marine Pollution Bulletin*, 47(1–6), 118–125. [https://doi.org/10.1016/S0025-326X\(02\)00411-3](https://doi.org/10.1016/S0025-326X(02)00411-3)
- Mihelčić, M., Šmuc, T., & Supek, F. (2019). Patterns of diverse gene functions in genomic neighborhoods predict gene function and phenotype. *Scientific Reports*, 9(1), 19537. <https://doi.org/10.1038/s41598-019-55984-0>
- Mikheenko, A., Saveliev, V., & Gurevich, A. (2016). MetaQUAST: Evaluation of metagenome assemblies. *Bioinformatics*, 32(7), 1088–1090. <https://doi.org/10.1093/bioinformatics/btv697>
- Miller, I. R., Bui, H., Wood, J. B., Fields, M. W., & Gerlach, R. (2024). Understanding phycosomal dynamics to improve industrial microalgae cultivation. *Trends in Biotechnology*, 42(6), 680–698. <https://doi.org/10.1016/j.tibtech.2023.12.003>
- Milo, R., & Phillips, R. (2015). *Cell Biology by the Numbers* (0 ed.). Garland Science. <https://doi.org/10.1201/9780429258770>
- Minch, B., Akter, S., Weinheimer, A., Rahman, M. S., Parvez, M. A. K., Rezwana Rahman, S., Ahmed, M. F., & Moniruzzaman, M. (2023). Phylogenetic diversity and functional potential of large and cell-associated viruses in the Bay of Bengal. *mSphere*, 8(6), e00407-23. <https://doi.org/10.1128/msphere.00407-23>
- Mirdita, M., Steinegger, M., Breitwieser, F., Söding, J., & Levy Karin, E. (2021). Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics*, 37(18), 3029–3031. <https://doi.org/10.1093/bioinformatics/btab184>
- Mise, K., & Iwasaki, W. (2022). Unexpected absence of ribosomal protein genes from metagenome-assembled genomes. *ISME Communications*, 2(1), 118. <https://doi.org/10.1038/s43705-022-00204-6>
- Mitchell, J. H., Freedman, A. H., Delaney, J. A., & Girguis, P. R. (2024). Co-expression analysis reveals distinct alliances around two carbon fixation pathways in hydrothermal vent symbionts. *Nature Microbiology*, 9(6), 1526–1539. <https://doi.org/10.1038/s41564-024-01704-y>
- Mohr, W., Lehnen, N., Ahmerkamp, S., Marchant, H. K., Graf, J. S., Tschitschko, B., Yilmaz, P., Littmann, S., Gruber-Vodicka, H., Leisch, N., Weber, M., Lott, C., Schubert, C. J., Milucka, J., & Kuypers, M. M. M. (2021). Terrestrial-type nitrogen-fixing symbiosis between seagrass and a marine bacterium. *Nature*, 600(7887), 105–109. <https://doi.org/10.1038/s41586-021-04063-4>
- Mor, B., Garhwal, S., & Kumar, A. (2021). A Systematic Review of Hidden Markov Models and Their Applications. *Archives of Computational Methods in Engineering*, 28(3), 1429–1448. <https://doi.org/10.1007/s11831-020-09422-4>

- Muratore, D., Boysen, A. K., Harke, M. J., Becker, K. W., Casey, J. R., Coesel, S. N., Mende, D. R., Wilson, S. T., Aylward, F. O., Eppley, J. M., Vislova, A., Peng, S., Rodriguez-Gonzalez, R. A., Beckett, S. J., Virginia Armbrust, E., DeLong, E. F., Karl, D. M., White, A. E., Zehr, J. P., ... Weitz, J. S. (2022). Complex marine microbial communities partition metabolism of scarce resources over the diel cycle. *Nature Ecology & Evolution*, 6(2), 218–229. <https://doi.org/10.1038/s41559-021-01606-w>
- Muscatine, L., & Cernichiar, E. (1969). ASSIMILATION OF PHOTOSYNTHETIC PRODUCTS OF ZOOXANTHELLAE BY A REEF CORAL. *The Biological Bulletin*, 137(3), 506–523. <https://doi.org/10.2307/1540172>
- Nahor, O., Israel, Á., Barger, N., Rubin-Blum, M., & Luzzatto-Knaan, T. (2024). Epiphytic microbiome associated with intertidal seaweeds in the Mediterranean Sea: Comparative analysis of bacterial communities across seaweed phyla. *Scientific Reports*, 14(1), 18631. <https://doi.org/10.1038/s41598-024-69362-y>
- Natural Earth. (2025). *1:10m Physical Vectors* [Dataset]. <https://www.naturalearthdata.com/downloads/10m-physical-vectors/>
- Nayfach, S., & Pollard, K. S. (2015). Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biology*, 16(1), 51. <https://doi.org/10.1186/s13059-015-0611-7>
- Neave, M. J., Michell, C. T., Apprill, A., & Voolstra, C. R. (2017). Endozoicomonas genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Scientific Reports*, 7(1), 40579. <https://doi.org/10.1038/srep40579>
- Neave, M. J., Rachmawati, R., Xun, L., Michell, C. T., Bourne, D. G., Apprill, A., & Voolstra, C. R. (2017). Differential specificity between closely related corals and abundant *Endozoicomonas* endosymbionts across global scales. *The ISME Journal*, 11(1), 186–200. <https://doi.org/10.1038/ismej.2016.95>
- Nelson, C. E., Wegley Kelly, L., & Haas, A. F. (2023). Microbial Interactions with Dissolved Organic Matter Are Central to Coral Reef Ecosystem Function and Resilience. *Annual Review of Marine Science*, 15(1), 431–460. <https://doi.org/10.1146/annurev-marine-042121-080917>
- Nelson, H. R., & Altieri, A. H. (2019). Oxygen: The universal currency on coral reefs. *Coral Reefs*, 38(2), 177–198. <https://doi.org/10.1007/s00338-019-01765-0>
- Nelson, W. C., Tully, B. J., & Mobberley, J. M. (2020). Biases in genome reconstruction from metagenomic data. *PeerJ*, 8, e10119. <https://doi.org/10.7717/peerj.10119>
- Nemergut, D. R., Schmidt, S. K., Fukami, T., O'Neill, S. P., Bilinski, T. M., Stanish, L. F., Knelman, J. E., Darcy, J. L., Lynch, R. C., Wickey, P., & Ferrenberg, S. (2013). Patterns and Processes of Microbial Community Assembly. *Microbiology and Molecular Biology Reviews*, 77(3), 342–356. <https://doi.org/10.1128/MMBR.00051-12>
- Nygaard, P., Duckert, P., & Saxild, H. H. (1996). Role of adenine deaminase in purine salvage and nitrogen metabolism and characterization of the ade gene in *Bacillus subtilis*. *Journal of Bacteriology*, 178(3), 846–853. <https://doi.org/10.1128/jb.178.3.846-853.1996>
- Ochsenkühn, M. A., Mohamed, A. R., Haydon, T. D., Coe, L. S. Y., Abrego, D., & Amin, S. A. (2023). *Endozoicomonas provides corals with steroid hormones during thermal stress*. <https://doi.org/10.1101/2023.09.19.558257>

- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., ... Weedon, J. (2022). *vegan: Community Ecology Package*. <https://CRAN.R-project.org/package=vegan>
- Olm, M. R., Brown, C. T., Brooks, B., & Banfield, J. F. (2017). dRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, 11(12), Article 12. <https://doi.org/10.1038/ismej.2017.126>
- Ottesen, E. A., Young, C. R., Gifford, S. M., Eppley, J. M., Marin, R., Schuster, S. C., Scholin, C. A., & DeLong, E. F. (2014). Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science*, 345(6193), 207–212. <https://doi.org/10.1126/science.1252476>
- Ouldali, H., Sarthak, K., Ensslen, T., Piguet, F., Manivet, P., Pelta, J., Behrends, J. C., Aksimentiev, A., & Oukhaled, A. (2020). Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nature Biotechnology*, 38(2), 176–181. <https://doi.org/10.1038/s41587-019-0345-2>
- Pacheco, A., Guedes, I., & Azevedo, S. (2016). Is qPCR a Reliable Indicator of Cyanotoxin Risk in Freshwater? *Toxins*, 8(6), 172. <https://doi.org/10.3390/toxins8060172>
- Pacherres, C. O., Ahmerkamp, S., Schmidt-Grieb, G. M., Holtappels, M., & Richter, C. (2020). Ciliary vortex flows and oxygen dynamics in the coral boundary layer. *Scientific Reports*, 10(1), 7541. <https://doi.org/10.1038/s41598-020-64420-7>
- Padilla, C. C., Bristow, L. A., Sarode, N., Garcia-Robledo, E., Gómez Ramírez, E., Benson, C. R., Bourbonnais, A., Altabet, M. A., Girguis, P. R., Thamdrup, B., & Stewart, F. J. (2016). NC10 bacteria in marine oxygen minimum zones. *The ISME Journal*, 10(8), 2067–2071. <https://doi.org/10.1038/ismej.2015.262>
- Pagès, H., Aboyoun, P., Gentleman, R., & DebRoy, S. (2024). *Biostrings: Efficient manipulation of biological strings. R package version 2.74.1* [R]. Bioconductor. <https://doi.org/10.18129/B9.BIOC.BIOSTRINGS>
- Paine, R. T. (1966). Food Web Complexity and Species Diversity. *The American Naturalist*, 100(910), 65–75. <https://doi.org/10.1086/282400>
- Pan, W., Wang, X., Ren, C., Jiang, X., Gong, S., Xie, Z., Wong, N.-K., Li, X., Huang, J., Fan, D., Luo, P., Yang, Y., Ren, X., Yu, S., Qin, Z., Wu, X., Huo, D., Ma, B., Liu, Y., ... Chen, T. (2024). Sea cucumbers and their symbiotic microbiome have evolved to feed on seabed sediments. *Nature Communications*, 15(1), 8825. <https://doi.org/10.1038/s41467-024-53205-5>
- Park, J., Davis, K., Lajoie, G., & Parfrey, L. W. (2022). Alternative approaches to identify core bacteria in *Fucus distichus* microbiome and assess their distribution and host-specificity. *Environmental Microbiome*, 17(1), 55. <https://doi.org/10.1186/s40793-022-00451-z>
- Parks, D. H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A. J., & Hugenholtz, P. (2020). A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology*, 38(9), 1079–1086. <https://doi.org/10.1038/s41587-020-0501-8>
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and

- metagenomes. *Genome Research*, 25(7), 1043–1055.
<https://doi.org/10.1101/gr.186072.114>
- Pasteur, L. (1881). On the Germ Theory. *Science*, os-2(62), 420–422.
<https://doi.org/10.1126/science.os-2.62.420>
- Patin, N. V., Dietrich, Z. A., Stancil, A., Quinan, M., Beckler, J. S., Hall, E. R., Culter, J., Smith, C. G., Taillefert, M., & Stewart, F. J. (2021). Gulf of Mexico blue hole harbors high levels of novel microbial lineages. *The ISME Journal*, 15(8), 2206–2232.
<https://doi.org/10.1038/s41396-021-00917-x>
- Peixoto, R. S., Rosado, P. M., Leite, D. C. de A., Rosado, A. S., & Bourne, D. G. (2017). Beneficial Microorganisms for Corals (BMC): Proposed Mechanisms for Coral Health and Resilience. *Frontiers in Microbiology*, 8, 341.
<https://doi.org/10.3389/fmicb.2017.00341>
- Pelve, E. A., Fontanez, K. M., & DeLong, E. F. (2017). Bacterial Succession on Sinking Particles in the Ocean's Interior. *Frontiers in Microbiology*, 8.
<https://www.frontiersin.org/articles/10.3389/fmicb.2017.02269>
- Pernice, M., Raina, J.-B., Rädcker, N., Cárdenas, A., Pogoreutz, C., & Voolstra, C. R. (2020). Down to the bone: The role of overlooked endolithic microbiomes in reef coral health. *The ISME Journal*, 14(2), Article 2. <https://doi.org/10.1038/s41396-019-0548-z>
- Pogoreutz, C., Oakley, C. A., Rädcker, N., Cárdenas, A., Perna, G., Xiang, N., Peng, L., Davy, S. K., Ngugi, D. K., & Voolstra, C. R. (2022). Coral holobiont cues prime *Endozoicomonas* for a symbiotic lifestyle. *The ISME Journal*, 16(8), 1883–1895.
<https://doi.org/10.1038/s41396-022-01226-7>
- Pogoreutz, C., Rädcker, N., Cárdenas, A., Gärdes, A., Wild, C., & Voolstra, C. R. (2018). Dominance of *Endozoicomonas* bacteria throughout coral bleaching and mortality suggests structural inflexibility of the *Pocillopora verrucosa* microbiome. *Ecology and Evolution*, 8(4), 2240–2252. <https://doi.org/10.1002/ece3.3830>
- Pogoreutz, C., & Ziegler, M. (2024). Frenemies on the reef? Resolving the coral–*Endozoicomonas* association. *Trends in Microbiology*, 32(5), 422–434.
<https://doi.org/10.1016/j.tim.2023.11.006>
- Pollock, F. J., McMinds, R., Smith, S., Bourne, D. G., Willis, B. L., Medina, M., Thurber, R. V., & Zaneveld, J. R. (2018). Coral-associated bacteria demonstrate phyllosymbiosis and cophylogeny. *Nature Communications*, 9(1), Article 1. <https://doi.org/10.1038/s41467-018-07275-x>
- Porras, M. Á. G., Assié, A., Tietjen, M., Violette, M., Kleiner, M., Gruber-Vodicka, H., Dubilier, N., & Leisch, N. (2024). An intranuclear bacterial parasite of deep-sea mussels expresses apoptosis inhibitors acquired from its host. *Nature Microbiology*.
<https://doi.org/10.1038/s41564-024-01808-5>
- Pratte, Z. A., Besson, M., Hollman, R. D., & Stewart, F. J. (2018). The Gills of Reef Fish Support a Distinct Microbiome Influenced by Host-Specific Factors. *Applied and Environmental Microbiology*, 84(9), e00063-18. <https://doi.org/10.1128/AEM.00063-18>
- Pratte, Z. A., Perry, C., Dove, A. D. M., Hoopes, L. A., Ritchie, K. B., Hueter, R. E., Fischer, C., Newton, A. L., & Stewart, F. J. (2022). Microbiome structure in large pelagic sharks with distinct feeding ecologies. *Animal Microbiome*, 4(1), 17. <https://doi.org/10.1186/s42523-022-00168-x>

- Prijbelski, A., Antipov, D., Meleshko, D., Lapidus, A., & Korobeynikov, A. (2020). Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics*, 70(1), e102. <https://doi.org/10.1002/cpbi.102>
- Purcell, S. W., Conand, C., Uthicke, S., & Byrne, M. (2016). Ecological Roles of Exploited Sea Cucumbers. In *Oceanography and Marine Biology: An Annual Review*. CRC Press. <https://doi.org/10.1201/9781315368597>
- Purcell, S. W., Lovatelli, A., Gonzalez Wanguemert, M., Solis Marin, F. A., Samyn, Y., & Conand, C. (2023). *Commercially important sea cucumbers of the world*. Food & Agriculture Org.
- Putnam, H. M., & Edmunds, P. J. (2011). The physiological response of reef corals to diel fluctuations in seawater temperature. *Journal of Experimental Marine Biology and Ecology*, 396(2), 216–223. <https://doi.org/10.1016/j.jembe.2010.10.026>
- QGIS Association. (2025). *QGIS Geographic Information System* (Version 3.40) [Computer software]. <http://www.qgis.org>
- Qi, Y., Kobayashi, Y., & Hulett, F. M. (1997). The pst operon of *Bacillus subtilis* has a phosphate-regulated promoter and is involved in phosphate transport but not in regulation of the pho regulon. *Journal of Bacteriology*, 179(8), 2534–2539. <https://doi.org/10.1128/jb.179.8.2534-2539.1997>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, 41(Database issue), D590–D596. <https://doi.org/10.1093/nar/gks1219>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Raina, J.-B., Tapiolas, D., Willis, B. L., & Bourne, D. G. (2009). Coral-Associated Bacteria and Their Role in the Biogeochemical Cycling of Sulfur. *Applied and Environmental Microbiology*. <https://doi.org/10.1128/AEM.02567-08>
- Rantala, A., Fewer, D. P., Hisbergues, M., Rouhiainen, L., Vaitomaa, J., Börner, T., & Sivonen, K. (2004). Phylogenetic evidence for the early evolution of microcystin synthesis. *Proceedings of the National Academy of Sciences*, 101(2), 568–573. <https://doi.org/10.1073/pnas.0304489101>
- Rasher, D. B., & Hay, M. E. (2010). Chemically rich seaweeds poison corals when not controlled by herbivores. *Proceedings of the National Academy of Sciences*, 107(21), 9683–9688. <https://doi.org/10.1073/pnas.0912095107>
- Read, Q. D., Grady, J. M., Zarnetske, P. L., Record, S., Baiser, B., Belmaker, J., Tuanmu, M.-N., Strecker, A., Beaudrot, L., & Thibault, K. M. (2018). Among-species overlap in rodent body size distributions predicts species richness along a temperature gradient. *Ecography*, 41(10), 1718–1727. <https://doi.org/10.1111/ecog.03641>
- Reed, D. C., Algar, C. K., Huber, J. A., & Dick, G. J. (2014). Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proceedings of the National Academy of Sciences*, 111(5), 1879–1884. <https://doi.org/10.1073/pnas.1313713111>
- Reichart, N. J., Jay, Z. J., Krukenberg, V., Parker, A. E., Spietz, R. L., & Hatzenpichler, R. (2020). Activity-based cell sorting reveals responses of uncultured archaea and bacteria to

- substrate amendment. *The ISME Journal*, 14(11), Article 11.
<https://doi.org/10.1038/s41396-020-00749-1>
- Resasco, J., Chacoff, N. P., & Vázquez, D. P. (2021). Plant–pollinator interactions between generalists persist over time and space. *Ecology*, 102(6), e03359.
<https://doi.org/10.1002/ecy.3359>
- Ribalet, F., Swalwell, J., Clayton, S., Jiménez, V., Sudek, S., Lin, Y., Johnson, Z. I., Worden, A. Z., & Armbrust, E. V. (2015). Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. *Proceedings of the National Academy of Sciences*, 112(26), 8008–8012. <https://doi.org/10.1073/pnas.1424279112>
- Ricci, F., Leung, P. M., Hutchinson, T., Nguyen-Dinh, T., Frank, A. H., Hood, A. v.S., Salazar, V. W., Eate, V., Wong, W. W., Cook, P. L. M., Greening, C., & McClelland, H. (2025). *Chemosynthesis enhances carbon fixation and nutrient cycling in an active microbialite ecosystem*. <https://doi.org/10.1101/2024.12.18.629060>
- Ricci, F., Rossetto Marcelino, V., Blackall, L. L., Kühl, M., Medina, M., & Verbruggen, H. (2019). Beneath the surface: Community assembly and functions of the coral skeleton microbiome. *Microbiome*, 7(1), 159. <https://doi.org/10.1186/s40168-019-0762-y>
- Richardson, L. L., Sekar, R., Myers, J. L., Gantar, M., Voss, J. D., Kaczmarsky, L., Remily, E. R., Boyer, G. L., & Zimba, P. V. (2007). The presence of the cyanobacterial toxin microcystin in black band disease of corals. *FEMS Microbiology Letters*, 272(2), 182–187. <https://doi.org/10.1111/j.1574-6968.2007.00751.x>
- Richardson, L., Miller, A., Broderick, E., Kaczmarsky, L., Gantar, M., Stanic, D., & Sekar, R. (2009). Sulfide, microcystin, and the etiology of black band disease. *Diseases of Aquatic Organisms*, 87, 79–90. <https://doi.org/10.3354/dao02083>
- Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R., & Konstantinidis, K. T. (2018). Nonpareil 3: Fast Estimation of Metagenomic Coverage and Sequence Diversity. *mSystems*, 3(3), e00039-18. <https://doi.org/10.1128/mSystems.00039-18>
- Rohwer, F., Seguritan, V., Azam, F., & Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series*, 243, 1–10.
<https://doi.org/10.3354/meps243001>
- Rondon, M. R., August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Loiacono, K. A., Lynch, B. A., MacNeil, I. A., Minor, C., Tiong, C. L., Gilman, M., Osburne, M. S., Clardy, J., Handelsman, J., & Goodman, R. M. (2000). Cloning the Soil Metagenome: A Strategy for Accessing the Genetic and Functional Diversity of Uncultured Microorganisms. *Applied and Environmental Microbiology*, 66(6), 2541–2547. <https://doi.org/10.1128/AEM.66.6.2541-2547.2000>
- Rosado, P. M., Cardoso, P. M., Rosado, J. G., Schultz, J., Nunes Da Rocha, U., Keller-Costa, T., & Peixoto, R. S. (2023). Exploring the Potential Molecular Mechanisms of Interactions between a Probiotic Consortium and Its Coral Host. *mSystems*, 8(1), e00921-22.
<https://doi.org/10.1128/msystems.00921-22>
- Rosado, P. M., Leite, D. C. A., Duarte, G. A. S., Chaloub, R. M., Jospin, G., Nunes da Rocha, U., P. Saraiva, J., Dini-Andreote, F., Eisen, J. A., Bourne, D. G., & Peixoto, R. S. (2019). Marine probiotics: Increasing coral resistance to bleaching through microbiome manipulation. *The ISME Journal*, 13(4), 921–936. <https://doi.org/10.1038/s41396-018-0323-6>

- Rosic, N. N., Pernice, M., Dove, S., Dunn, S., & Hoegh-Guldberg, O. (2011). Gene expression profiles of cytosolic heat shock proteins Hsp70 and Hsp90 from symbiotic dinoflagellates in response to thermal stress: Possible implications for coral bleaching. *Cell Stress and Chaperones*, 16(1), 69–80. <https://doi.org/10.1007/s12192-010-0222-x>
- Saaltink, R. M., Honingh, E., Dekker, S. C., Griffioen, J., van Riel, M. C., Verdonschot, P. F. M., Vink, J. P. M., Winterwerp, J. C., & Wassen, M. J. (2018). Respiration and aeration by bioturbating Tubificidae alter biogeochemical processes in aquatic sediment. *Aquatic Sciences*, 81(1), 13. <https://doi.org/10.1007/s00027-018-0610-3>
- Sanger, F., Air, G. M., Barrell, B. G., Brown, N. L., Coulson, A. R., Fiddes, C., & Hutchison, C. A. (1977). *Nucleotide sequence of bacteriophage*. 265.
- Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R. P., Banday, S., Mishra, A. K., Das, G., & Malonia, S. K. (2023). Next-Generation Sequencing Technology: Current Trends and Advancements. *Biology*, 12(7), 997. <https://doi.org/10.3390/biology12070997>
- Sayers, E. W., Beck, J., Bolton, E. E., Brister, J. R., Chan, J., Comeau, D. C., Connor, R., DiCuccio, M., Farrell, C. M., Feldgarden, M., Fine, A. M., Funk, K., Hatcher, E., Hoepfner, M., Kane, M., Kannan, S., Katz, K. S., Kelly, C., Klimke, W., ... Sherry, S. T. (2024). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 52(D1), D33–D43. <https://doi.org/10.1093/nar/gkad1044>
- Schaible, G. A., Kohtz, A. J., Cliff, J., & Hatzenpichler, R. (2022). Correlative SIP-FISH-Raman-SEM-NanoSIMS links identity, morphology, biochemistry, and physiology of environmental microbes. *ISME Communications*, 2(1), 52. <https://doi.org/10.1038/s43705-022-00134-3>
- Schindler, D. W. (1974). Eutrophication and Recovery in Experimental Lakes: Implications for Lake Management. *Science*, 184(4139), 897–899. <https://doi.org/10.1126/science.184.4139.897>
- Schmalle, V., & Lorentz, A. (2020). Role of the microbiota in circadian rhythms of the host. *Chronobiology International*, 37(3), 301–310. <https://doi.org/10.1080/07420528.2020.1726374>
- Schweiger, A. K., & Laliberté, E. (2022). Plant beta-diversity across biomes captured by imaging spectroscopy. *Nature Communications*, 13(1), 2767. <https://doi.org/10.1038/s41467-022-30369-6>
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Segev, E., Wyche, T. P., Kim, K. H., Petersen, J., Ellebrandt, C., Vlamakis, H., Barteneva, N., Paulson, J. N., Chai, L., Clardy, J., & Kolter, R. (2016). Dynamic metabolic exchange governs a marine algal-bacterial interaction. *eLife*, 5, e17473. <https://doi.org/10.7554/eLife.17473>
- Sethuraman, N., Jeremiah, S. S., & Ryo, A. (2020). Interpreting Diagnostic Tests for SARS-CoV-2. *JAMA*, 323(22), 2249. <https://doi.org/10.1001/jama.2020.8259>
- Seyedsayamdost, M. R., Case, R. J., Kolter, R., & Clardy, J. (2011). The Jekyll-and-Hyde chemistry of *Phaeobacter gallaeciensis*. *Nature Chemistry*, 3(4), 331–335. <https://doi.org/10.1038/nchem.1002>

- Sezonov, G., Joseleau-Petit, D., & D'Ari, R. (2007). *Escherichia coli* Physiology in Luria-Bertani Broth. *Journal of Bacteriology*, 189(23), 8746–8749. <https://doi.org/10.1128/jb.01368-07>
- Shade, A., Gregory Caporaso, J., Handelsman, J., Knight, R., & Fierer, N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time. *The ISME Journal*, 7(8), 1493–1506. <https://doi.org/10.1038/ismej.2013.54>
- Shaffer, M., Borton, M. A., McGivern, B. B., Zayed, A. A., La Rosa, S. L., Solden, L. M., Liu, P., Narrowe, A. B., Rodríguez-Ramos, J., Bolduc, B., Gazitúa, M. C., Daly, R. A., Smith, G. J., Vik, D. R., Pope, P. B., Sullivan, M. B., Roux, S., & Wrighton, K. C. (2020). DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Research*, 48(16), 8883–8900. <https://doi.org/10.1093/nar/gkaa621>
- Sharon, I., Morowitz, M. J., Thomas, B. C., Costello, E. K., Relman, D. A., & Banfield, J. F. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research*, 23(1), 111–120. <https://doi.org/10.1101/gr.142315.112>
- Sharp, K. H., Pratte, Z. A., Kerwin, A. H., Rotjan, R. D., & Stewart, F. J. (2017). Season, but not symbiont state, drives microbiome structure in the temperate coral *Astrangia poculata*. *Microbiome*, 5(1), 120. <https://doi.org/10.1186/s40168-017-0329-8>
- Shashar, N., Cohen, Y., & Loya, Y. (1993). Extreme Diel Fluctuations of Oxygen in Diffusive Boundary Layers Surrounding Stony Corals. *The Biological Bulletin*, 185(3), 455–461. <https://doi.org/10.2307/1542485>
- Shaver, E. C., & Silliman, B. R. (2017). Time to cash in on positive interactions for coral restoration. *PeerJ*, 5, e3499. <https://doi.org/10.7717/peerj.3499>
- Sheik, A. R., Muller, E. E. L., & Wilmes, P. (2014). A hundred years of activated sludge: Time for a rethink. *Frontiers in Microbiology*, 5. <https://doi.org/10.3389/fmicb.2014.00047>
- Shilling, E. N., Combs, I. R., & Voss, J. D. (2021). Assessing the effectiveness of two intervention methods for stony coral tissue loss disease on *Montastraea cavernosa*. *Scientific Reports*, 11(1), 8566. <https://doi.org/10.1038/s41598-021-86926-4>
- Shoemaker, A., Maritan, A., Cosar, S., Nupp, S., Menchaca, A., Jackson, T., Dang, A., Baxter, B. K., Colman, D. R., Dunham, E. C., & Boyd, E. S. (2024). Wood–Ljungdahl pathway encoding anaerobes facilitate low-cost primary production in hypersaline sediments at Great Salt Lake, Utah. *FEMS Microbiology Ecology*, 100(8), fiae105. <https://doi.org/10.1093/femsec/fiae105>
- Shom. (2023). *MNT topo-bathymétrique côtier de Moorea (Projet Météo-France)* (http://www.shom.fr/BATHYMETRIE/MNT_COTIER_MOOREA_METEO-FRANCE_20m_WGS84) [Dataset].
- Sibly, R. M., & Hone, J. (2002). Population growth rate and its determinants: An overview. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 357(1425), 1153–1170. <https://doi.org/10.1098/rstb.2002.1117>
- Simó, R., Archer, S. D., Pedrós-Alió, C., Gilpin, L., & Stelfox-Widdicombe, C. E. (2002). Coupled dynamics of dimethylsulfoniopropionate and dimethylsulfide cycling and the microbial food web in surface waters of the North Atlantic. *Limnology and Oceanography*, 47(1), 53–61. <https://doi.org/10.4319/lo.2002.47.1.0053>

- Singh, R. P., & Reddy, C. R. K. (2016). Unraveling the Functions of the Macroalgal Microbiome. *Frontiers in Microbiology*, 6. <https://doi.org/10.3389/fmicb.2015.01488>
- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology*, 122(1), e59. <https://doi.org/10.1002/cpmb.59>
- Smercina, D. N., Evans, S. E., Friesen, M. L., & Tiemann, L. K. (2019). To Fix or Not To Fix: Controls on Free-Living Nitrogen Fixation in the Rhizosphere. *Applied and Environmental Microbiology*, 85(6), e02546-18. <https://doi.org/10.1128/AEM.02546-18>
- Sogin, E. M., Michellod, D., Gruber-Vodicka, H. R., Bourceau, P., Geier, B., Meier, D. V., Seidel, M., Ahmerkamp, S., Schorn, S., D'Angelo, G., Procaccini, G., Dubilier, N., & Liebeke, M. (2022). Sugars dominate the seagrass rhizosphere. *Nature Ecology & Evolution*, 6(7), 866–877. <https://doi.org/10.1038/s41559-022-01740-z>
- Søndergaard, D., Pedersen, C. N. S., & Greening, C. (2016). HydDB: A web tool for hydrogenase classification and analysis. *Scientific Reports*, 6(1), 34212. <https://doi.org/10.1038/srep34212>
- Sorek, M., Schnytzer, Y., Waldman Ben-Asher, H., Caspi, V. C., Chen, C.-S., Miller, D. J., & Levy, O. (2018). Setting the pace: Host rhythmic behaviour and gene expression patterns in the facultatively symbiotic cnidarian *Aiptasia* are determined largely by *Symbiodinium*. *Microbiome*, 6(1), 83. <https://doi.org/10.1186/s40168-018-0465-9>
- Speth, D. R., & Orphan, V. J. (2018). Metabolic marker gene mining provides insight in global *mcrA* diversity and, coupled with targeted genome reconstruction, sheds further light on metabolic potential of the *Methanomassiliicoccales*. *PeerJ*, 6, e5614. <https://doi.org/10.7717/peerj.5614>
- Spoerner, M., Wichard, T., Bachhuber, T., Stratmann, J., & Oertel, W. (2012). Growth and Thallus Morphogenesis of *Ulva mutabilis* (Chlorophyta) Depends on A Combination of Two Bacterial Species Excreting Regulatory Factors. *Journal of Phycology*, 48(6), 1433–1447. <https://doi.org/10.1111/j.1529-8817.2012.01231.x>
- Sriswasdi, S., Yang, C., & Iwasaki, W. (2017). Generalist species drive microbial dispersion and evolution. *Nature Communications*, 8(1), 1162. <https://doi.org/10.1038/s41467-017-01265-1>
- Srivastava, S., Anand, V., Kaur, J., Ranjan, M., Bist, V., Asif, M. H., & Srivastava, S. (2022). Functional Genetic Diversity and Plant Growth Promoting Potential of Polyphosphate Accumulating Bacteria in Soil. *Microbiology Spectrum*, 10(1), e00345-21. <https://doi.org/10.1128/spectrum.00345-21>
- Staley, C., Ferrieri, A. P., Tfaily, M. M., Cui, Y., Chu, R. K., Wang, P., Shaw, J. B., Ansong, C. K., Brewer, H., Norbeck, A. D., Markillie, M., do Amaral, F., Tuleski, T., Pellizzaro, T., Agtuca, B., Ferrieri, R., Tringe, S. G., Paša-Tolić, L., Stacey, G., & Sadowsky, M. J. (2017). Diurnal cycling of rhizosphere bacterial communities is associated with shifts in carbon metabolism. *Microbiome*, 5(1), 65. <https://doi.org/10.1186/s40168-017-0287-1>
- Steinsdóttir, H. G. R., Gómez-Ramírez, E., Mhatre, S., Schauburger, C., Bertagnolli, A. D., Pratte, Z. A., Stewart, F. J., Thamdrup, B., & Bristow, L. A. (2022). Anaerobic methane oxidation in a coastal oxygen minimum zone: Spatial and temporal dynamics. *Environmental Microbiology*, 24(5), 2361–2379. <https://doi.org/10.1111/1462-2920.16003>

- Stephens, Z. D., Lee, S. Y., Faghri, F., Campbell, R. H., Zhai, C., Efron, M. J., Iyer, R., Schatz, M. C., Sinha, S., & Robinson, G. E. (2015). Big Data: Astronomical or Genomical? *PLOS Biology*, 13(7), e1002195. <https://doi.org/10.1371/journal.pbio.1002195>
- Stewart, F. J., Ulloa, O., & DeLong, E. F. (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology*, 14(1), 23–40. <https://doi.org/10.1111/j.1462-2920.2010.02400.x>
- Stewart, P. S., & Franklin, M. J. (2008). Physiological heterogeneity in biofilms. *Nature Reviews Microbiology*, 6(3), 199–210. <https://doi.org/10.1038/nrmicro1838>
- Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner, H., & Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, 19(s1), 21–31. <https://doi.org/10.1111/j.1365-294X.2009.04480.x>
- Studivan, M. S., Baptist, M., Molina, V., Riley, S., First, M., Soderberg, N., Rubin, E., Rossin, A., Holstein, D. M., & Enochs, I. C. (2022). Transmission of stony coral tissue loss disease (SCTLD) in simulated ballast water confirms the potential for ship-born spread. *Scientific Reports*, 12(1), 19248. <https://doi.org/10.1038/s41598-022-21868-z>
- Sturm, G., Mojarrad, M., & Kaster, A.-K. (2023). Targeted Cell Labeling and Sorting of Prokaryotes for Cultivation and Omics Approaches. *Microbial Physiology*, 33(1), 63–84. <https://doi.org/10.1159/000532088>
- Sweet, M. J., Croquer, A., & Bythell, J. C. (2010). Temporal and spatial patterns in waterborne bacterial communities of an island reef system. *Aquatic Microbial Ecology*, 61(1), 1–11. <https://doi.org/10.3354/ame01433>
- Tandon, K., Chiou, Y.-J., Yu, S.-P., Hsieh, H. J., Lu, C.-Y., Hsu, M.-T., Chiang, P.-W., Chen, H.-J., Wada, N., & Tang, S.-L. (2022). Microbiome Restructuring: Dominant Coral Bacterium *Endozoicomonas* Species Respond Differentially to Environmental Changes. *mSystems*, 7(4), e00359-22. <https://doi.org/10.1128/msystems.00359-22>
- Tandon, K., Lu, C.-Y., Chiang, P.-W., Wada, N., Yang, S.-H., Chan, Y.-F., Chen, P.-Y., Chang, H.-Y., Chiou, Y.-J., Chou, M.-S., Chen, W.-M., & Tang, S.-L. (2020). Comparative genomics: Dominant coral-bacterium *Endozoicomonas acroporae* metabolizes dimethylsulfoniopropionate (DMSP). *The ISME Journal*, 14(5), 1290–1303. <https://doi.org/10.1038/s41396-020-0610-x>
- Tesson, F., Hervé, A., Mordret, E., Touchon, M., d’Humières, C., Cury, J., & Bernheim, A. (2022). Systematic and quantitative view of the antiviral arsenal of prokaryotes. *Nature Communications*, 13(1), 2561. <https://doi.org/10.1038/s41467-022-30269-9>
- Thaben, P. F., & Westermark, P. O. (2014). Detecting Rhythms in Time Series with RAIN. *Journal of Biological Rhythms*, 29(6), 391–400. <https://doi.org/10.1177/0748730414553029>
- The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207–214. <https://doi.org/10.1038/nature11234>
- The UniProt Consortium. (2023). UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Research*, 51(D1), D523–D531. <https://doi.org/10.1093/nar/gkac1052>
- Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., Prill, R. J., Tripathi, A., Gibbons, S. M., Ackermann, G., Navas-Molina, J. A., Janssen, S., Kopylova,

- E., Vázquez-Baeza, Y., González, A., Morton, J. T., Mirarab, S., Zech Xu, Z., Jiang, L., ... Zhao, H. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*, 551(7681), 457–463. <https://doi.org/10.1038/nature24621>
- Thume, K., Gebser, B., Chen, L., Meyer, N., Kieber, D. J., & Pohnert, G. (2018). The metabolite dimethylsulfoxonium propionate extends the marine organosulfur cycle. *Nature*, 563(7731), 412–415. <https://doi.org/10.1038/s41586-018-0675-0>
- Thurber, R. L. V., Burkepile, D. E., Fuchs, C., Shantz, A. A., McMinds, R., & Zaneveld, J. R. (2014). Chronic nutrient enrichment increases prevalence and severity of coral disease and bleaching. *Global Change Biology*, 20(2), 544–554. <https://doi.org/10.1111/gcb.12450>
- Trexler, R. V., Van Goethem, M. W., Goudeau, D., Nath, N., Malmstrom, R. R., Northen, T. R., & Couradeau, E. (2023). BONCAT-FACS-Seq reveals the active fraction of a biocrust community undergoing a wet-up event. *Frontiers in Microbiology*, 14, 1176751. <https://doi.org/10.3389/fmicb.2023.1176751>
- Trigodet, F., Lolans, K., Fogarty, E., Shaiber, A., Morrison, H. G., Barreiro, L., Jabri, B., & Eren, A. M. (2022). High molecular weight DNA extraction strategies for long-read sequencing of complex metagenomes. *Molecular Ecology Resources*, 22(5), 1786–1802. <https://doi.org/10.1111/1755-0998.13588>
- Tripp, H. J., Kitner, J. B., Schwalbach, M. S., Dacey, J. W. H., Wilhelm, L. J., & Giovannoni, S. J. (2008). SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature*, 452(7188), 741–744. <https://doi.org/10.1038/nature06776>
- Trotter, V. V., Shatsky, M., Price, M. N., Juba, T. R., Zane, G. M., De León, K. B., Majumder, E. L.-W., Gui, Q., Ali, R., Wetmore, K. M., Kuehl, J. V., Arkin, A. P., Wall, J. D., Deutschbauer, A. M., Chandonia, J.-M., & Butland, G. P. (2023). Large-scale genetic characterization of the model sulfate-reducing bacterium, *Desulfovibrio vulgaris* Hildenborough. *Frontiers in Microbiology*, 14, 1095191. <https://doi.org/10.3389/fmicb.2023.1095191>
- Trubl, G., Kimbrel, J. A., Lique-Gonzalez, J., Nuccio, E. E., Weber, P. K., Pett-Ridge, J., Jansson, J. K., Waldrop, M. P., & Blazewicz, S. J. (2021). Active virus-host interactions at sub-freezing temperatures in Arctic peat soil. *Microbiome*, 9(1), 208. <https://doi.org/10.1186/s40168-021-01154-2>
- Tschitschko, B., Esti, M., Philippi, M., Kidane, A. T., Littmann, S., Kitzinger, K., Speth, D. R., Li, S., Kraberg, A., Tienken, D., Marchant, H. K., Kartal, B., Milucka, J., Mohr, W., & Kuypers, M. M. M. (2024). Rhizobia–diatom symbiosis fixes missing nitrogen in the ocean. *Nature*, 630(8018), 899–904. <https://doi.org/10.1038/s41586-024-07495-w>
- Tsementzi, D., Wu, J., Deutsch, S., Nath, S., Rodriguez-R, L. M., Burns, A. S., Ranjan, P., Sarode, N., Malmstrom, R. R., Padilla, C. C., Stone, B. K., Bristow, L. A., Larsen, M., Glass, J. B., Thamdrup, B., Woyke, T., Konstantinidis, K. T., & Stewart, F. J. (2016). SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature*, 536(7615), 179–183. <https://doi.org/10.1038/nature19068>
- Turnlund, A. C., Vanwonderghem, I., Botté, E. S., Randall, C. J., Giuliano, C., Kam, L., Bell, S., O'Brien, P., Negri, A. P., Webster, N. S., & Lurgi, M. (2023). Linking differences in microbial network structure with changes in coral larval settlement. *ISME Communications*, 3(1), 114. <https://doi.org/10.1038/s43705-023-00320-x>

- Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., & Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature*, 428(6978), 37–43. <https://doi.org/10.1038/nature02340>
- Uchimiya, M., Schroer, W., Olofsson, M., Edison, A. S., & Moran, M. A. (2022). Diel investments in metabolite production and consumption in a model microbial system. *The ISME Journal*, 16(5), 1306–1317. <https://doi.org/10.1038/s41396-021-01172-w>
- Uritskiy, G. V., DiRuggiero, J., & Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, 6(1), 158. <https://doi.org/10.1186/s40168-018-0541-1>
- Uthicke, S. (1994). Distribution patterns and growth of two reef flat holothurians, *Holothuria atra* and *Stichopus chloronotus*. In *Echinoderms Through Time*. CRC Press.
- Uthicke, S. (1999). Sediment bioturbation and impact of feeding activity of *Holothuria (halodeima) atra* and *Stichopus chloronotus*, two sediment feeding holothurians, at Lizard Island, Great Barrier Reef. *Bulletin of Marine Science*, 64(1).
- Van Der Star, W. R. L., Abma, W. R., Blommers, D., Mulder, J.-W., Tokutomi, T., Strous, M., Picoreanu, C., & Van Loosdrecht, M. C. M. (2007). Startup of reactors for anoxic ammonium oxidation: Experiences from the first full-scale anammox reactor in Rotterdam. *Water Research*, 41(18), 4149–4163. <https://doi.org/10.1016/j.watres.2007.03.044>
- Van Niftrik, L., & Jetten, M. S. M. (2012). Anaerobic Ammonium-Oxidizing Bacteria: Unique Microorganisms with Exceptional Properties. *Microbiology and Molecular Biology Reviews*, 76(3), 585–596. <https://doi.org/10.1128/MMBR.05025-11>
- Vellend, M. (2010). Conceptual Synthesis in Community Ecology. *The Quarterly Review of Biology*, 85(2), 183–206. <https://doi.org/10.1086/652373>
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., ... Zhu, X. (2001). The Sequence of the Human Genome. *Science*, 291(5507), 1304–1351. <https://doi.org/10.1126/science.1058040>
- Větrovský, T., & Baldrian, P. (2013). The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE*, 8(2), e57923. <https://doi.org/10.1371/journal.pone.0057923>
- Vilanova-Cuevas, B., Philipp, K., Altera, A., Apprill, A., Becker, C., Behringer, D., Brandt, M., Breitbart, M., Budd, K., DeRito, C., Duermit-Moreau, E., Evans, J., Hopson-Fernandes, M., Fleischer, J., Gittens, S., Henson, M., Hylkema, A., Kellogg, C., Maritan, A., ... Hewson, I. (2025). Detection of the *Diadema antillarum* scuticociliatosis *Philaster* clade on sympatric metazoa, plankton, and abiotic surfaces and assessment for its potential reemergence. *Marine Ecology Progress Series*, 753, 19–35. <https://doi.org/10.3354/meps14763>
- Villela, H., Modolon, F., Schultz, J., Delgadillo-Ordoñez, N., Carvalho, S., Soriano, A. U., & Peixoto, R. S. (2023). Genome analysis of a coral-associated bacterial consortium highlights complementary hydrocarbon degradation ability and other beneficial

- mechanisms for the host. *Scientific Reports*, 13(1), 12273.
<https://doi.org/10.1038/s41598-023-38512-z>
- Vincent, A. T., & Charette, S. J. (2015). Who qualifies to be a bioinformatician? *Frontiers in Genetics*, 6. <https://doi.org/10.3389/fgene.2015.00164>
- Vislova, A., Sosa, O. A., Eppley, J. M., Romano, A. E., & DeLong, E. F. (2019). Diel Oscillation of Microbial Gene Transcripts Declines With Depth in Oligotrophic Ocean Waters. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.02191>
- Viver, T., Conrad, R. E., Orellana, L. H., Urdiain, M., González-Pastor, J. E., Hatt, J. K., Amann, R., Antón, J., Konstantinidis, K. T., & Rosselló-Móra, R. (2021). Distinct ecotypes within a natural haloarchaeal population enable adaptation to changing environmental conditions without causing population sweeps. *The ISME Journal*, 15(4), 1178–1191.
<https://doi.org/10.1038/s41396-020-00842-5>
- Vompe, A. D., Epstein, H. E., Speare, K. E., Schmeltzer, E. R., Adam, T. C., Burkepile, D. E., Sharpton, T. J., & Vega Thurber, R. (2024). Microbiome ecological memory and responses to repeated marine heatwaves clarify variation in coral bleaching and mortality. *Global Change Biology*, 30(1), e17088. <https://doi.org/10.1111/gcb.17088>
- Wada, N., Hsu, M.-T., Tandon, K., Hsiao, S. S.-Y., Chen, H.-J., Chen, Y.-H., Chiang, P.-W., Yu, S.-P., Lu, C.-Y., Chiou, Y.-J., Tu, Y.-C., Tian, X., Chen, B.-C., Lee, D.-C., Yamashiro, H., Bourne, D. G., & Tang, S.-L. (2022). High-resolution spatial and genomic characterization of coral-associated microbial aggregates in the coral *Stylophora pistillata*. *Science Advances*, 8(27), eabo2431. <https://doi.org/10.1126/sciadv.abo2431>
- Wang, Q., Alowaifeer, A., Kerner, P., Balasubramanian, N., Patterson, A., Christian, W., Tarver, A., Dore, J. E., Hatzenpichler, R., Bothner, B., & McDermott, T. R. (2021). Aerobic bacterial methane synthesis. *Proceedings of the National Academy of Sciences*, 118(27), e2019229118. <https://doi.org/10.1073/pnas.2019229118>
- Wang, Q., McDermott, T. R., & Walk, S. T. (2021). A Single Microbiome Gene Alters Murine Susceptibility to Acute Arsenic Exposure. *Toxicological Sciences*, 181(1), 105–114.
<https://doi.org/10.1093/toxsci/kfab017>
- Wang, S., Meade, A., Lam, H.-M., & Luo, H. (2020). Evolutionary Timeline and Genomic Plasticity Underlying the Lifestyle Diversity in *Rhizobiales*. *mSystems*, 5(4), e00438-20. <https://doi.org/10.1128/mSystems.00438-20>
- Ward, D. M., Cohan, F. M., Bhaya, D., Heidelberg, J. F., Kühl, M., & Grossman, A. (2008). Genomics, environmental genomics and the issue of microbial species. *Heredity*, 100(2), 207–219. <https://doi.org/10.1038/sj.hdy.6801011>
- Weathers, K. C., Strayer, D. L., & Likens, G. E. (2021). Introduction to Ecosystem Science. In *Fundamentals of ecosystem science* (pp. 3–26). Academic Press.
- Weber, L., & Apprill, A. (2020). Diel, daily, and spatial variation of coral reef seawater microbial communities. *PLOS ONE*, 15(3), e0229442.
<https://doi.org/10.1371/journal.pone.0229442>
- Weber, M., De Beer, D., Lott, C., Polerecky, L., Kohls, K., Abed, R. M. M., Ferdelman, T. G., & Fabricius, K. E. (2012). Mechanisms of damage to corals exposed to sedimentation. *Proceedings of the National Academy of Sciences*, 109(24).
<https://doi.org/10.1073/pnas.1100715109>

- White, E. R., Cox, K., Melbourne, B. A., & Hastings, A. (2019). Success and failure of ecological management is highly variable in an experimental test. *Proceedings of the National Academy of Sciences*, 116(46), 23169–23173.
<https://doi.org/10.1073/pnas.1911440116>
- Whitman, W. B., Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences*, 95(12), 6578–6583.
<https://doi.org/10.1073/pnas.95.12.6578>
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
<https://ggplot2.tidyverse.org>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
<https://doi.org/10.21105/joss.01686>
- Wilbanks, E. G., Doré, H., Ashby, M. H., Heiner, C., Roberts, R. J., & Eisen, J. A. (2022). Metagenomic methylation patterns resolve bacterial genomes of unusual size and structural complexity. *The ISME Journal*, 16(8), 1921–1931.
<https://doi.org/10.1038/s41396-022-01242-7>
- Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74(11), 5088–5090. <https://doi.org/10.1073/pnas.74.11.5088>
- Wu, Y.-W., Simmons, B. A., & Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, 32(4), 605–607. <https://doi.org/10.1093/bioinformatics/btv638>
- Wurch, L., Giannone, R. J., Belisle, B. S., Swift, C., Utturkar, S., Hettich, R. L., Reysenbach, A.-L., & Podar, M. (2016). Genomics-informed isolation and characterization of a symbiotic Nanoarchaeota system from a terrestrial geothermal environment. *Nature Communications*, 7(1), 12115. <https://doi.org/10.1038/ncomms12115>
- Xie, C., Goi, C. L. W., Huson, D. H., Little, P. F. R., & Williams, R. B. H. (2016). RiboTagger: Fast and unbiased 16S/18S profiling using whole community shotgun metagenomic or metatranscriptome surveys. *BMC Bioinformatics*, 17(19), 508.
<https://doi.org/10.1186/s12859-016-1378-x>
- Xu, L., Chen, H., Hu, X., Zhang, R., Zhang, Z., & Luo, Z. W. (2006). Average Gene Length Is Highly Conserved in Prokaryotes and Eukaryotes and Diverges Only Between the Two Kingdoms. *Molecular Biology and Evolution*, 23(6), 1107–1108.
<https://doi.org/10.1093/molbev/msk019>
- Xu, X., Li, C., Cao, W., Yan, L., Cao, L., Han, Q., Gao, M., Chen, Y., Shen, Z., Jiang, J., & Chen, C. (2024). Bacterial growth and environmental adaptation via thiamine biosynthesis and thiamine-mediated metabolic interactions. *The ISME Journal*, 18(1), wrae157.
<https://doi.org/10.1093/ismejo/wrae157>
- Xu, Y., Ma, K., Huang, S., Liu, L., & Lu, Y. (2012). Diel cycle of methanogen mcrA transcripts in rice rhizosphere. *Environmental Microbiology Reports*, 4(6), 655–663.
<https://doi.org/10.1111/j.1758-2229.2012.00392.x>

- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. (2017). GGTREE: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36.
<https://doi.org/10.1111/2041-210X.12628>
- Zakem, E. J., Polz, M. F., & Follows, M. J. (2020). Redox-informed models of global biogeochemical cycles. *Nature Communications*, 11(1), 5680.
<https://doi.org/10.1038/s41467-020-19454-w>
- Zaneveld, J. R., McMinds, R., & Vega Thurber, R. (2017). Stress and stability: Applying the Anna Karenina principle to animal microbiomes. *Nature Microbiology*, 2(9), Article 9.
<https://doi.org/10.1038/nmicrobiol.2017.121>
- Zavarzin, G. A. (2006). Winogradsky and modern microbiology. *Microbiology*, 75(5), 501–511.
<https://doi.org/10.1134/S0026261706050018>
- Zhang, L., Chen, T., Wang, Y., Zhang, S., Lv, Q., Kong, D., Jiang, H., Zheng, Y., Ren, Y., Huang, W., Liu, P., & Jiang, Y. (2022). Comparison Analysis of Different DNA Extraction Methods on Suitability for Long-Read Metagenomic Nanopore Sequencing. *Frontiers in Cellular and Infection Microbiology*, 12, 919903.
<https://doi.org/10.3389/fcimb.2022.919903>
- Zhao, K., Ma, B., Xu, Y., Stirling, E., & Xu, J. (2021). Light exposure mediates circadian rhythms of rhizosphere microbial communities. *The ISME Journal*, 15(9), Article 9.
<https://doi.org/10.1038/s41396-021-00957-3>
- Zheng, W., Zhao, S., Yin, Y., Zhang, H., Needham, D. M., Evans, E. D., Dai, C. L., Lu, P. J., Alm, E. J., & Weitz, D. A. (2022). High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science*, 376(6597), eabm1483.
<https://doi.org/10.1126/science.abm1483>
- Zhou, Z., Tran, P. Q., Breister, A. M., Liu, Y., Kieft, K., Cowley, E. S., Karaoz, U., & Anantharaman, K. (2022). METABOLIC: High-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome*, 10(1), 33. <https://doi.org/10.1186/s40168-021-01213-8>
- Zupančič, M., Kogovšek, P., Šter, T., Remec Rekar, Š., Cerasino, L., Baebler, Š., Krivograd Klemenčič, A., & Eleršek, T. (2021). Potentially Toxic Planktic and Benthic Cyanobacteria in Slovenian Freshwater Bodies: Detection by Quantitative PCR. *Toxins*, 13(2), 133. <https://doi.org/10.3390/toxins13020133>

APPENDICES

APPENDIX A

EXTENDED TABLES FROM CHAPTER TWO

Table A1.: Rarefaction depths used for different analyses, broken up by 16S and 18S rRNA gene data. N, total samples; n, subset samples; raref., rarefaction; ret., retained; sam., samples; NA, not applicable.

Sample Type	Whole reef comparison				Time-series comparison			
	16S rRNA		18S rRNA		16S rRNA		18S rRNA	
	Raref. depth (reads)	Ret. (n/N) sam.	Raref. depth (reads)	Ret. (n/N) sam.	Raref. depth (reads)	Ret. (n/N) sam.	Raref. depth (reads)	Ret. (n/N) sam.
<i>P. damicornis</i>	3,415	63/63	1,822	55/62	21,251	45/45	1,047	45/45
<i>P. rus</i>		61/66		55/63	3,415	45/48	911	46/47
<i>Halimeda</i> sp.		18/18		18/18	41,533	NA	NA	NA
<i>Padina</i> sp.		79/79		79/79	18,328	79/79	4,329	79/79
<i>Hydroclathrus</i> sp.		18/18		18/18	15,108	NA	NA	NA
<i>Dictyota</i> sp.		18/18		16/16	12,756	NA	NA	NA
Sediment		31/31		29/29	28,195	31/31	12,524	29/29
Water		44/44		44/44	28,012	32/32	12,819	32/32

Table A2.: Statistical testing for beta diversity differences between reef microbiomes (ASV-scale) depending on sample type. Comparison of Bray-Curtis dissimilarity for different reef microbiomes as tested by permutational multivariate analysis of variance (PERMANOVA) and permutational multivariate analysis of dispersion (PERMDISP), 999 permutations. Comparisons are bolded to identify statistically distinguishable contrasts (PERMANOVA or PERMDISP, $P < 0.05$). PDAM, *Pocillipora damicornis*, PRUS, *Porites rus*; perm., Permutations; n, sample size.

rRNA	Group 1	Group 2	n	PERMANOVA P value	PERMDISP P value
16S	blank	dictyota	21	0.001	0.895
16S	blank	padina	82	0.001	0.586
16S	blank	sediment	34	0.001	0.013
16S	blank	water	47	0.001	0.696
16S	blank	hydroclathrus	21	0.002	0.462
16S	blank	halimeda	21	0.003	0.965
16S	dictyota	halimeda	36	0.001	0.168
16S	dictyota	hydroclathrus	36	0.001	0.001
16S	dictyota	padina	97	0.001	0.249
16S	dictyota	sediment	49	0.001	0.006
16S	dictyota	water	62	0.001	0.397
16S	halimeda	hydroclathrus	36	0.001	0.001
16S	halimeda	padina	97	0.001	0.782
16S	halimeda	sediment	49	0.001	0.205
16S	halimeda	water	62	0.001	0.031

Table A2 Continued.

rRNA	Group 1	Group 2	n	PERMANOVA <i>P</i> value	PERMDISP <i>P</i> value
16S	hydroclathrus	padina	97	0.001	0.141
16S	hydroclathrus	sediment	49	0.001	0.328
16S	hydroclathrus	water	62	0.001	0.001
16S	padina	sediment	110	0.001	0.324
16S	padina	water	123	0.001	0.009
16S	PDAM	PRUS	124	0.001	0.001
16S	PDAM	blank	66	0.001	0.001
16S	PDAM	dictyota	81	0.001	0.001
16S	PDAM	halimeda	81	0.001	0.001
16S	PDAM	hydroclathrus	81	0.001	0.001
16S	PDAM	padina	142	0.001	0.001
16S	PDAM	sediment	94	0.001	0.001
16S	PDAM	water	107	0.001	0.001
16S	PRUS	blank	64	0.001	0.077
16S	PRUS	dictyota	79	0.001	0.001
16S	PRUS	halimeda	79	0.001	0.001
16S	PRUS	hydroclathrus	79	0.001	0.003
16S	PRUS	padina	140	0.001	0.001
16S	PRUS	sediment	92	0.001	0.001
16S	PRUS	water	105	0.001	0.001
16S	sediment	water	75	0.001	0.002
18S	blank	padina	81	0.001	0.784
18S	blank	sediment	31	0.002	0.924
18S	blank	halimeda	20	0.005	0.897
18S	blank	hydroclathrus	20	0.005	0.916
18S	blank	water	46	0.007	0.872
18S	blank	dictyota	18	0.008	0.933
18S	dictyota	halimeda	34	0.001	0.002
18S	dictyota	hydroclathrus	34	0.001	0.261
18S	dictyota	padina	95	0.001	0.848
18S	dictyota	sediment	45	0.001	0.328
18S	dictyota	water	60	0.001	0.005
18S	halimeda	hydroclathrus	36	0.001	0.001
18S	halimeda	padina	97	0.001	0.001
18S	halimeda	sediment	47	0.001	0.006
18S	halimeda	water	62	0.001	0.384
18S	hydroclathrus	padina	97	0.001	0.221
18S	hydroclathrus	sediment	47	0.001	0.744
18S	hydroclathrus	water	62	0.001	0.002
18S	padina	sediment	108	0.001	0.341

Table A2 Continued.

rRNA	Group 1	Group 2	n	PERMANOVA <i>P</i> value	PERMDISP <i>P</i> value
18S	Padina	water	123	0.001	0.001
18S	PDAM	PRUS	110	0.001	0.532
18S	PDAM	blank	57	0.001	0.392
18S	PDAM	dictyota	71	0.001	0.073
18S	PDAM	halimeda	73	0.001	0.216
18S	PDAM	hydroclathrus	73	0.001	0.035
18S	PDAM	padina	134	0.001	0.001
18S	PDAM	sediment	84	0.001	0.134
18S	PDAM	water	99	0.001	0.001
18S	PRUS	dictyota	71	0.001	0.001
18S	PRUS	halimeda	73	0.001	0.001
18S	PRUS	hydroclathrus	73	0.001	0.001
18S	PRUS	padina	134	0.001	0.001
18S	PRUS	sediment	84	0.001	0.001
18S	PRUS	water	99	0.001	0.001
18S	PRUS	blank	57	0.002	0.001
18S	sediment	water	73	0.001	0.001

Table A3.: Statistical testing for beta diversity differences between reef microbiomes (ASV-scale) depending on phase of sampling (per Figure 4I). Comparison of Bray-Curtis dissimilarity for different reef microbiomes as tested by permutational multivariate analysis of variance (PERMANOVA; 999 permutations), only displaying comparisons of Phase 1 vs. Phase 2 samples (Pocillipora damicornis, Porites rus, and water samples). Bolded comparisons identify statistically distinguishable contrasts (PERMANOVA, $P < 0.05$). PDAM, Pocillipora damicornis, PRUS, Porites rus, P1, Phase 1 (1800 h on 29 March 2022 to 1200 h on 2 April 2022, ~6 ft depth, coral rack); P2, Phase 2 (1200 h on 4 April 2022 to 0001h on 7 April 2022, ~3 ft depth, reef flat).

rRNA	Group 1	Group 2	sample size	Pseudo <i>F</i>	<i>P</i> value
16S	PDAM P1	PDAM P2	63	0.504	0.323
16S	water P1	water P2	44	2.902	0.001
16S	PRUS P1	PRUS P2	61	3.255	0.001
18S	PDAM P1	PDAM P2	55	31.293	0.001
18S	water P1	water P2	44	2.671	0.001
18S	PRUS P1	PRUS P2	55	4.280	0.007

Table A4.: The core microbiomes of different reef microbiomes. The core microbiome is defined here as genera that were present in at least 80% of all samples for a given sample type and had a percent abundance greater than 1% in at least one of those samples and presented as average percentage abundances for each core prokaryote and non-host eukaryote. Hydr., Hydroclathrus sp.; Dict., Dictyosphaera sp.; Hali., Halimeda sp.; Padi., Padina sp.; Sedi., sediment; Bact., Bacteria; Arch., Archaea; Euk., Eukaryota; Sam., Sample; Avg. Perc. Abd., average percentage abundance in the specified sample type; UC, unclassified.

Sam. type	Domain	Phylum	Family	Genus	Avg. Perc. Abd.
PDA M	Bact.	Proteobacteria	Endozoicomonadaceae	Endozoicomonas	98.14
PDA M	Euk.	UC. dinoflagellate	Symbiodiniaceae	Symbiodinium	96.43
PRU S	Bact.	Proteobacteria	Halomonadaceae	Halomonas	6.6
PRU S	Euk.	UC. dinoflagellate	Symbiodiniaceae	Symbiodinium	96.82
Dict.	Bact.	Proteobacteria	Rhodobacteraceae		15.06
Dict.	Bact.	Cyanobacteria	Phormidiaceae	Trichodesmium	6.94
Dict.	Bact.	Cyanobacteria	Nostocaceae	Rivularia	6
Dict.	Bact.	Cyanobacteria	Nostocaceae		5.29
Dict.	Bact.	Cyanobacteria	Nostocaceae	Cylindrospermopsis	5.21
Dict.	Bact.	Cyanobacteria	Phormidesmiaceae	Acrophormium	3.33
Dict.	Bact.	Cyanobacteria	Phormidesmiaceae	Phormidesmis	3.15
Dict.	Bact.	Bacteroidota	Flavobacteriaceae	Muricauda	2.21
Dict.	Bact.	Proteobacteria	Hyphomonadaceae	uncultured	1.62
Dict.	Euk.	UC. dinoflagellate	Symbiodiniaceae	Symbiodinium	19.47
Dict.	Euk.	Arthropoda	Miraciidae	Paramphiascella	9.94
Dict.	Euk.	Rhodophyta	Porolithaceae	Porolithon	9.14
Dict.	Euk.	Bacillariophyta	Bacillariaceae	Psammodictyon	3.03
Hali.	Bact.	Proteobacteria	Rhodobacteraceae		15.66
Hali.	Bact.	Proteobacteria	Alteromonadaceae	Aestuariusbacter	6.07
Hali.	Bact.	Proteobacteria	Parvularculaceae		2.63
Hali.	Euk.	UC. dinoflagellate	Symbiodiniaceae	Symbiodinium	11.8
Hali.	Euk.	Chlorophyta	Ulvellaceae	Ulvella	4.58
Hali.	Euk.	Bacillariophyta	Bacillariaceae	Psammodictyon	3.22
Hydr.	Bact.	Cyanobacteria	Nostocaceae	Cylindrospermopsis	11.5
Hydr.	Bact.	Proteobacteria	Rhodobacteraceae		10.47
Hydr.	Bact.	Bacteroidota	Cyclobacteriaceae	Ekhidna	3.41

Table A4 Continued.

Sam. type	Domain	Phylum	Family	Genus	Avg. Perc. Abd.
Hydr.	Bact.	Verrucomicrobiota	DEV007	DEV007	3.08
Hydr.	Bact.	Bacteroidota	Flavobacteriaceae	Muricauda	1.89
Hydr.	Euk.	Arthropoda	Miraciidae	Paramphiascella	29.38
Hydr.	Euk.	Annelida	Syllidae	Salvatoria	15.97
Hydr.	Euk.	Bacillariophyta	Bacillariaceae	Psammodictyon	2.81
Padi.	Bact.	Proteobacteria	Rhodobacteraceae		17.16
Padi.	Bact.	Cyanobacteria	Nostocaceae	Rivularia	3.98
Padi.	Bact.	Cyanobacteria	Nostocaceae	Cylindrospermopsis	3.6
Padi.	Bact.	Bacteroidota	Flavobacteriaceae	Muricauda	3.1
Padi.	Euk.	Rhodophyta	Porolithaceae	Porolithon	12.75
Padi.	Euk.	Bacillariophyta	Fragilariaceae	Diatoma	9.82
Padi.	Euk.	Arthropoda	Miraciidae	Paramphiascella	9.21
Padi.	Euk.	unclassified	Symbiodiniaceae	Symbiodinium	5.84
Padi.	Euk.	Chlorophyta	Ulvaceae	Ulva	5.15
Padi.	Euk.	Bacillariophyta	Bacillariaceae	Psammodictyon	4.08
Sedi.	Bact.	Desulfobacterota	Desulfosarcinaceae	uncultured	5.38
Sedi.	Bact.	Proteobacteria	Rhodobacteraceae		4.67
Sedi.	Bact.	Cyanobacteria	Xenococcaceae	Xenococcus	3.36
Sedi.	Bact.	Planctomycetota	OM190	OM190	2.53
Sedi.	Bact.	Proteobacteria	Woeseiaceae	Woeseia	2.51
Sedi.	Bact.	Bacteroidota	Flavobacteriaceae		1.91
Sedi.	Bact.	Proteobacteria	BD7-8	BD7-8	1.67
Sedi.	Bact.	Desulfobacterota	Desulfosarcinaceae	Sva0081_sediment_group	1.65
Sedi.	Bact.	Proteobacteria	Nitrosococcaceae		1.62
Sedi.	Euk.	Arthropoda	Miraciidae	Paramphiascella	15.57
Sedi.	Euk.	UC. dinoflagellate	Prorocentraceae	Prorocentrum	8.66
Sedi.	Euk.	Arthropoda	Cytheridae	Cythere	3.31
water	Bact.	Cyanobacteria	Cyanobiaceae	Synechococcus	19.06
water	Bact.	Cyanobacteria	Cyanobiaceae	Prochlorococcus	10.19
water	Arch.	Thermoplasmata			4.63
water	Bact.	Proteobacteria	Vibrionaceae	Vibrio	2.99
water	Bact.	Actinobacteriota	Actinomarinales	Candidatus_Actinomarina	2.79
water	Arch.	Nanoarchaeota	Woesearchaeales	Woesearchaeales	2.72

Table A4 Continued.

Sam. type	Domain	Phylum	Family	Genus	Avg. Perc. Abd.
water	Bact.	Proteobacteria	SAR11 Clade I	SAR11 Clade Ia	2.59
water	Bact.	Proteobacteria	AEGEAN-169 mar. grp	AEGEAN-169 mar. grp	2.37
water	Bact.	Proteobacteria	SAR116 clade	SAR116 clade	2.32
water	Bact.	Proteobacteria	SAR86 clade	SAR86 clade	2.31
water	Bact.	Proteobacteria	SAR11 Clade II	SAR11 Clade II	2.14
water	Euk.	Ciliophora	Glaucomidae	Glaucoma	9.7
water	Euk.	UC. dinoflagellate	Gymnodiniaceae	Lepidodinium	3.1

Table A5.: Statistical testing of cyclicity for measured geophysical features using RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods). Bolded *P* values indicate significance (RAIN, $P < 0.05$, Bonferroni correction).

Measured feature	Bonferroni corrected <i>P</i> value
Dissolved Oxygen mg/L	< 0.001
Temperature (°C)	< 0.001
Light (lux)	< 0.001
Salinity (ppt)	1
Phosphate (μM) – below limit of detection	0.018
Nitrite (μM) – below limit of detection	1
Nitrate (μM)	0.56
Ammonium (μM)	0.99

Table A6.: Statistical testing of cyclicity for measured alpha diversity in different sample types using RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods). Bolded *P* values indicate significance (RAIN, $P < 0.05$, Bonferroni correction). Bf, Bonferroni correction; rRNA, ribosomal rRNA; PDAM, *Pocillipora damicornis*; PRUS, *Porites rus*.

Sample type	rRNA	Bf corrected <i>P</i> value
PRUS	16S	0.813
	18S	0.794
PDAM	16S	0.706
	18S	0.831
<i>Padina</i> sp.	16S	0.754
	18S	0.330
Sediment	16S	0.830
	18S	0.856
Water	16S	0.806
	18S	0.860

Table A7.: Statistical testing of differences in mean abundance over time for measured alpha diversity in different sample types using linear mixed effect models (corals) or ANOVA (water, sediment, and macroalgae). For each sample type and rRNA-type, all samples collected over time were compared to one another. If the model fit was initially significant, then post hoc test was performed to identify significant contrasts. In initial model testing significant (linear (mixed effect), p-values were bolded ($P < 0.05$, Tukey HSD) or underlined ($P > 0.05$ but < 0.10 , Tukey HSD). In post-hoc testing, groups with dissimilar letters in post hoc testing have P values < 0.05 , while any shared letter indicates P value > 0.05 . rRNA, ribosomal rRNA; PDAM, Pocillipora damicornis; PRUS, Porites rus.

Sample type	rRNA	Initial model ANOVA <i>P</i> value	Contrasts with Tukey HSD (distinct group)
PRUS	16S	0.9045	NA
	18S	<u>0.056</u>	2022-03-30 00:01:00 (A) 2022-03-31 12:00:00 (AB) 2022-04-02 00:01:00 (AB) 2022-04-01 18:00:00 (AB) 2022-03-30 06:00:00 (AB) 2022-04-02 12:00:00 (AB) 2022-04-01 06:00:00 (AB) 2022-03-29 18:00:00 (AB) 2022-04-02 06:00:00 (AB) 2022-03-30 18:00:00 (AB) 2022-04-01 12:00:00 (AB) 2022-04-01 00:01:00 (AB) 2022-03-31 06:00:00 (AB) 2022-03-30 12:00:00 (AB) 2022-03-31 00:01:00 (AB) 2022-03-31 18:00:00 (B)
PDAM	16S	0.471	NA
	18S	0.471	NA

Table A7 Continued.

Sample type	rRNA	Initial model ANOVA <i>P</i> value	Contrasts with Tukey HSD (distinct group)
<i>Padina</i> sp.	16S	0.2785	NA
	18S	0.0365	2022-03-30 18:00:00 (AB) 2022-03-31 06:00:00 (A) 2022-04-02 12:00:00 (AB) 2022-03-30 00:01:00 (AB) 2022-03-29 18:00:00 (AB) 2022-04-01 00:01:00 (AB) 2022-03-31 12:00:00 (AB) 2022-04-01 18:00:00 (AB) 2022-04-01 12:00:00 (AB) 2022-03-31 18:00:00 (AB) 2022-04-02 06:00:00 (AB) 2022-03-30 12:00:00 (AB) 2022-03-31 00:01:00 (AB) 2022-04-01 06:00:00 (AB) 2022-04-02 00:01:00 (AB) 2022-03-30 06:00:00 (B)
Sediment	16S	0.873	NA
	18S	0.0355	No significant contrasts after HSD

Table A7 Continued.

Sample type	rRNA	Initial model ANOVA <i>P</i> value	Contrasts with Tukey HSD (distinct group)
Water	16S	< 0.0001	2022-04-02 12:00:00 (A) 2022-04-01 18:00:00 (AB) 2022-04-02 00:01:00 (ABC) 2022-03-31 00:01:00 (BCD) 2022-04-01 06:00:00 (BCDE) 2022-04-02 06:00:00 (BCDE) 2022-03-31 18:00:00 (BCDEF) 2022-03-30 00:01:00 (CDEFG) 2022-04-01 12:00:00 (DEFGH) 2022-03-31 12:00:00 (EFGH) 2022-03-31 06:00:00 (FGH) 2022-03-29 18:00:00 (FGH) 2022-03-30 18:00:00 (FGH) 2022-04-01 00:01:00 (GH) 2022-03-30 12:00:00 (GH) 2022-03-30 06:00:00 (H)
	18S	0.0019	2022-03-31 06:00:00 (A) 2022-03-30 06:00:00 (AB) 2022-03-31 00:01:00 (ABC) 2022-03-29 18:00:00 (ABC) 2022-04-01 12:00:00 (ABC) 2022-04-02 00:01:00 (ABC) 2022-04-02 12:00:00 (ABC) 2022-03-30 18:00:00 (BC) 2022-03-30 12:00:00 (BC) 2022-03-31 12:00:00 (BC) 2022-04-01 18:00:00 (BC) 2022-04-01 06:00:00 (BC) 2022-03-30 00:01:00 (C) 2022-03-31 18:00:00 (C) 2022-04-01 00:01:00 (C) 2022-04-02 06:00:00 (C)

Table A8.: Microbiome differences within a sample type by time of day as determined by permutational analysis of variance (PERMANOVA; 999 permutations). Comparisons are bolded to identify statistically significant ($P < 0.05$) and underlined for detectable contrasts ($P < 0.10$). n, sample size.

rRNA	species	Group 1	Group 2	n	PERMANOVA <i>P</i> value
16S	padina	0:01:00	6:00:00	40	0.025
16S	padina	6:00:00	18:00:00	39	<u>0.07</u>
16S	padina	0:01:00	18:00:00	39	0.12
16S	padina	6:00:00	12:00:00	40	0.134
16S	padina	0:01:00	12:00:00	40	0.245
16S	padina	12:00:00	18:00:00	39	0.413
16S	PDAM	12:00:00	18:00:00	33	0.325
16S	PDAM	0:01:00	12:00:00	42	0.351
16S	PDAM	6:00:00	18:00:00	21	0.414
16S	PDAM	6:00:00	12:00:00	30	0.787
16S	PDAM	0:01:00	6:00:00	30	0.884
16S	PDAM	0:01:00	18:00:00	33	0.929
16S	PRUS	12:00:00	18:00:00	31	0.002
16S	PRUS	0:01:00	6:00:00	30	0.003
16S	PRUS	0:01:00	12:00:00	40	0.009
16S	PRUS	6:00:00	12:00:00	30	<u>0.094</u>
16S	PRUS	6:00:00	18:00:00	21	<u>0.099</u>
16S	PRUS	0:01:00	18:00:00	31	0.316
16S	sediment	6:00:00	18:00:00	15	<u>0.06</u>
16S	sediment	0:01:00	6:00:00	16	0.166
16S	sediment	6:00:00	12:00:00	16	0.402
16S	sediment	0:01:00	18:00:00	15	0.793
16S	sediment	12:00:00	18:00:00	15	0.896
16S	sediment	0:01:00	12:00:00	16	0.936
16S	water	0:01:00	12:00:00	28	0.001
16S	water	0:01:00	18:00:00	22	0.001
16S	water	0:01:00	6:00:00	22	0.004
16S	water	12:00:00	18:00:00	22	0.018
16S	water	6:00:00	18:00:00	16	<u>0.096</u>
16S	water	6:00:00	12:00:00	22	0.17
18S	padina	0:01:00	12:00:00	40	0.002
18S	padina	12:00:00	18:00:00	39	0.004
18S	padina	6:00:00	12:00:00	40	0.005
18S	padina	6:00:00	18:00:00	39	0.013
18S	padina	0:01:00	6:00:00	40	<u>0.081</u>
18S	padina	0:01:00	18:00:00	39	0.864

Table A8 Continued.

rRNA	species	Group 1	Group 2	n	PERMANOVA <i>P</i> value
18S	PDAM	0:01:00	6:00:00	29	0.15
18S	PDAM	6:00:00	12:00:00	30	0.169
18S	PDAM	0:01:00	18:00:00	32	0.332
18S	PDAM	12:00:00	18:00:00	33	0.371
18S	PDAM	6:00:00	18:00:00	21	0.744
18S	PDAM	0:01:00	12:00:00	41	0.899
18S	PRUS	6:00:00	18:00:00	23	0.332
18S	PRUS	6:00:00	12:00:00	32	0.346
18S	PRUS	0:01:00	6:00:00	30	0.351
18S	PRUS	0:01:00	12:00:00	38	0.656
18S	PRUS	0:01:00	18:00:00	29	0.785
18S	PRUS	12:00:00	18:00:00	31	0.851
18S	sediment	0:01:00	6:00:00	15	0.032
18S	sediment	6:00:00	18:00:00	14	0.118
18S	sediment	0:01:00	12:00:00	15	0.161
18S	sediment	12:00:00	18:00:00	14	0.281
18S	sediment	6:00:00	12:00:00	16	0.346
18S	sediment	0:01:00	18:00:00	13	0.403
18S	water	0:01:00	6:00:00	22	0.001
18S	water	0:01:00	12:00:00	28	0.001
18S	water	0:01:00	18:00:00	22	0.001
18S	water	6:00:00	12:00:00	22	0.001
18S	water	6:00:00	18:00:00	16	0.001
18S	water	12:00:00	18:00:00	22	0.001

Table A9.: Taxa (amplicon sequence variants) identified as contributing to community differences by time in Similarity Percentage (SIMPER) analysis, only showing top 2 taxa per comparison. rRNA indicates amplicon source, 16S or 18S rRNA genes; sam. type, the reef sample class used in the comparison; group, the comparison being made (time A vs. time B) avg perc, the average percentage that the taxa is found in the sample type; Avg A/B perc., the average percentage abundance that the taxa is found in the focal sample type at time A/B; Abs Avg A MIN Avg B, the absolute value of the average abundance of the taxon in group A less the abundance in group B; lineage, the taxonomic lineage of the taxon per SILVA rRNA nomenclature (domain, phylum, class, order, family, genus).

rRNA	sample type	group	avg perc.	Avg A perc.	Avg B perc.	Abs Avg A MIN Avg B	lineage
16S	PRUS	0:01 to 12:00	4.92	1.84	10.26	8.42	d__Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Pseudoalteromonadaceae; Pseudoalteromonas
16S	PRUS	0:01 to 12:00	1.00	0.55	1.99	1.44	d__Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Moraxellaceae; Psychrobacter
16S	PRUS	0:01 to 6:00	1.51	0.82	2.69	1.87	d__Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; Endozoicomonadaceae; Endozoicomonas
16S	PRUS	0:01 to 6:00	1.00	0.18	1.98	1.80	d__Bacteria; Actinobacteriota; Actinobacteria; Corynebacteriales; Mycobacteriaceae; Mycobacterium
16S	PRUS	18:00 to 12:00	4.98	1.05	10.26	9.21	d__Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Pseudoalteromonadaceae; Pseudoalteromonas
16S	PRUS	18:00 to 12:00	2.18	4.37	0.00	4.37	d__Bacteria; Proteobacteria; Gammaproteobacteria; Cellvibrionales; Cellvibrionaceae
16S	PRUS	18:00 to 6:00	0.98	0.12	1.98	1.85	d__Bacteria; Actinobacteriota; Actinobacteria; Corynebacteriales; Mycobacteriaceae; Mycobacterium
16S	PRUS	18:00 to 6:00	1.53	0.93	2.69	1.76	d__Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; Endozoicomonadaceae; Endozoicomonas
16S	PRUS	6:00 to 12:00	5.19	5.92	10.26	4.34	d__Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Pseudoalteromonadaceae; Pseudoalteromonas
16S	PRUS	6:00 to 12:00	3.33	4.16	6.09	1.92	d__Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae; Alteromonas
16S	padi na	0:01 to 6:00	0.95	0.73	1.86	1.13	d__Bacteria; Proteobacteria; Gammaproteobacteria
16S	padi na	0:01 to 6:00	0.52	0.37	1.21	0.84	d__Bacteria; Bacteroidota; Bacteroidia; Flavobacteriales; Flavobacteriaceae; Muricauda
16S	padi na	18:00 to 6:00	0.92	0.03	1.86	1.83	d__Bacteria; Proteobacteria; Gammaproteobacteria
16S	padi na	18:00 to 6:00	0.85	1.59	0.20	1.39	d__Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; Halomonadaceae; Halomonas
16S	sediment	18:00 to 6:00	0.41	1.13	0.39	0.74	d__Bacteria; Proteobacteria; Gammaproteobacteria
16S	sediment	18:00 to 6:00	0.20	0.54	0.15	0.39	d__Bacteria; Cyanobacteria; Cyanobacteriia; Synechococcales; Cyanobiaceae
16S	water	0:01 to 12:00	5.99	23.38	13.25	10.13	d__Bacteria; Cyanobacteria; Cyanobacteriia; Synechococcales; Cyanobiaceae; Synechococcus CC9902
16S	water	0:01 to 12:00	0.75	2.23	0.72	1.51	d__Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; HIMB11

Table A9 Continued.

rR N A	sam. type	group	avg perc .	Avg A perc.	Avg B perc.	Abs Avg A MIN Avg B	lineage
16 S	wate r	0:01 to 6:00	0.67	2.23	0.88	1.34	d__Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; HIMB11
16 S	wate r	0:01 to 6:00	0.33	1.31	0.72	0.58	d__Bacteria; Proteobacteria; Alphaproteobacteria; Puniceispirillales; SAR116 clade; SAR116 clade
16 S	wate r	18:00 to 0:01	5.96	16.44	23.38	6.95	d__Bacteria; Cyanobacteria; Cyanobacteriia; Synechococcales; Cyanobiaceae; Synechococcus CC9902
16 S	wate r	18:00 to 0:01	1.00	0.60	2.53	1.93	d__Bacteria; Actinobacteriota; Acidimicrobiia; Actinomarinales; Actinomarinaceae; Candidatus Actinomarina
16 S	wate r	18:00 to 12:00	0.75	1.78	0.29	1.49	d__Bacteria; Proteobacteria; Gammaproteobacteria
16 S	wate r	18:00 to 12:00	0.37	0.05	0.78	0.73	d__Bacteria; Proteobacteria; Gammaproteobacteria
16 S	wate r	18:00 to 6:00	0.87	1.78	0.03	1.75	d__Bacteria; Proteobacteria; Gammaproteobacteria
16 S	wate r	18:00 to 6:00	0.28	0.80	0.29	0.51	d__Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Vibrio
18 S	padi na	0:01 to 12:00	1.51	1.48	3.44	1.96	Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Porolithaceae; Porolithon
18 S	padi na	0:01 to 12:00	0.95	2.43	0.61	1.82	Eukaryota; unclassified Eukaryota phylum; Dinophyceae; Peridiniales; Peridiniaceae; Bysmatrum
18 S	padi na	0:01 to 6:00	0.68	0.23	1.46	1.23	Eukaryota; Arthropoda; Hexanauplia; Harpacticoida; Miraciidae; Paramphiascella
18 S	padi na	0:01 to 6:00	0.27	0.03	0.52	0.49	Eukaryota; Arthropoda; Malacostraca; Tanaidacea; Tanaididae; Tanais
18 S	padi na	18:00 to 12:00	1.00	2.48	0.61	1.87	Eukaryota; unclassified Eukaryota phylum; Dinophyceae; Peridiniales; Peridiniaceae; Bysmatrum
18 S	padi na	18:00 to 12:00	1.38	1.79	3.44	1.65	Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Porolithaceae; Porolithon
18 S	padi na	18:00 to 6:00	0.67	0.34	1.46	1.11	Eukaryota; Arthropoda; Hexanauplia; Harpacticoida; Miraciidae; Paramphiascella
18 S	padi na	18:00 to 6:00	0.36	0.76	0.10	0.66	Eukaryota; Bacillariophyta; Bacillariophyceae; Bacillariales; Bacillariaceae; Psammodictyon
18 S	padi na	6:00 to 12:00	4.85	7.96	13.90	5.94	Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Porolithaceae; Porolithon
18 S	padi na	6:00 to 12:00	1.46	1.05	3.44	2.40	Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Porolithaceae; Porolithon
18 S	sedi ment	0:01 to 6:00	2.23	4.46	0.00	4.46	Eukaryota; Annelida; Polychaeta; unclassified Polychaeta order; Maldanidae; Praxillella
18 S	sedi ment	0:01 to 6:00	1.58	3.25	0.27	2.98	Eukaryota; Arthropoda; Hexanauplia; Harpacticoida; Cancrincolidae; Cancrincola
18 S	wate r	0:01 to 12:00	4.24	8.69	0.26	8.43	Eukaryota; Annelida; Sipuncula; Aspidosiphonidormes; Aspidosiphonidae; Aspidosiphon
18 S	wate r	0:01 to 12:00	2.24	0.02	4.49	4.47	Eukaryota; unclassified Eukaryota phylum; Phaeophyceae; Ectocarpales; Chordariaceae; Leathesia

Table A9 Continued.

rR N A	sam. type	group	avg perc .	Avg A perc.	Avg B perc.	Abs Avg A MIN Avg B	lineage
18 S	wate r	0:01 to 6:00	13.9 3	0.16	27.92	27.76	Eukaryota; Porifera; Demospongiae; Clionaida; Clionaidae; Cliona
18 S	wate r	0:01 to 6:00	4.13	8.69	1.36	7.33	Eukaryota; Annelida; Sipuncula; Aspidosiphonidormes; Aspidosiphonidae; Aspidosiphon
18 S	wate r	18:00 to 0:01	4.32	0.05	8.69	8.64	Eukaryota; Annelida; Sipuncula; Aspidosiphonidormes; Aspidosiphonidae; Aspidosiphon
18 S	wate r	18:00 to 0:01	3.02	6.04	0.00	6.04	Eukaryota; Annelida; Polychaeta; Phyllodocida; Syllidae; Typosyllis
18 S	wate r	18:00 to 12:00	3.02	6.04	0.00	6.04	Eukaryota; Annelida; Polychaeta; Phyllodocida; Syllidae; Typosyllis
18 S	wate r	18:00 to 12:00	2.62	5.30	0.06	5.24	Eukaryota; unclassified Eukaryota phylum; Phaeophyceae; Dictyotales; Dictyotaceae; Zonaria
18 S	wate r	18:00 to 6:00	13.9 4	0.06	27.92	27.86	Eukaryota; Porifera; Demospongiae; Clionaida; Clionaidae; Cliona
18 S	wate r	18:00 to 6:00	3.02	6.04	0.00	6.03	Eukaryota; Annelida; Polychaeta; Phyllodocida; Syllidae; Typosyllis
18 S	wate r	6:00 to 12:00	13.9 1	27.92	0.42	27.50	Eukaryota; Porifera; Demospongiae; Clionaida; Clionaidae; Cliona
18 S	wate r	6:00 to 12:00	2.49	0.13	4.91	4.78	Eukaryota; Annelida; Polychaeta; Spionida; Spionidae; Prionospio

Table A10.: Cyclic microbial taxa in each of the reef microbiomes sampled as identified statistically by RAIN (Rhythmicity Analysis Incorporating Non-parametric Methods) and false discovery rate (FDR) corrected (RAIN, Adaptive Benjamini-Hochberg, p-value < 0.05) and filtered to only include the most abundant (>1-5%), and prevalent ($\geq 25\%$ of all samples) taxa, then placed into four unsupervised groups (visualized in Figure 9). To increase clarity for this table, we filtered the cyclic taxa from each microhabitat to only include five taxa from each Domain with the largest distance values (indicating more closely associated with their plotted cluster) for each cluster. Sam. type, sample type; clust., cluster; dist., distance; lineage (Domain, Phylum, Class, Order, Family, Genus).

sam. type	clust.	dist.	lineage
padina	1	0.066	Bacteria; Cyanobacteria; Cyanobacteriia; Phormidesmiales; Phormidesmiaceae; Phormidesmis ANT.LACV5.1
padina	1	0.062	Bacteria; Cyanobacteria; Cyanobacteriia; Synechococcales; Synechococcales Incertae Sedis; Schizothrix LEGE_07164
padina	1	0.056	Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales; Hyphomonadaceae; uncultured
padina	1	0.054	Bacteria; Proteobacteria; Gammaproteobacteria; Granulosicoccales; Granulosicoccaceae; Granulosicoccus
padina	1	0.053	Bacteria; Cyanobacteria; Cyanobacteriia; Phormidesmiales; Phormidesmiaceae; NA
padina	1	0.053	Eukaryota; Chlorophyta; Ulvophyceae; Ulvales; Ulvellaceae; Ulvella
padina	1	0.047	Eukaryota; Ciliophora; Colpodea; Colpodida; Sandmanniellidae; Sandmanniella
padina	1	0.041	Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Porolithaceae; Porolithon
padina	1	0.028	Eukaryota; Bacillariophyta; Bacillariophyceae; Naviculales; Naviculaceae; Navicula
padina	1	0.028	Eukaryota; unclassified Eukaryota phylum; Cryptophyceae; Pyrenomonadales; Chroomonadaceae; Komma
padina	2	- 0.011	Archaea; Nanoarchaeota; Nanoarchaeia; Woesearchaeales; Woesearchaeales
padina	2	0.057	Bacteria; Chloroflexi; Anaerolineae; SBR1031; SBR1031; SBR1031
padina	2	0.054	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Actibacterium
padina	2	0.048	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Methylobacteriaceae; uncultured
padina	2	0.045	Bacteria; Myxococcota; Polyangia; Polyangiales; Eel-36e1D6; Eel-36e1D6
padina	2	0.042	Bacteria; Bacteroidota; Bacteroidia; Flavobacteriales; Flavobacteriaceae; Algitalia
padina	2	0.046	Eukaryota; unclassified Eukaryota phylum; Dinophyceae; Gymnodiniales; Kareniaceae; Karenia

Table A10 Continued.

sam. type	clust.	dist.	lineage
padina	2	0.046	Eukaryota; unclassified Eukaryota phylum; Pelagophyceae; Sarcinochrysidales; Sarcinochrysidaceae; Sargassococcus
padina	2	0.042	Eukaryota; Platyhelminthes; Rhabditophora; Rhabdocoela; Cystioplanidae; Cystiplex
padina	2	0.041	Eukaryota; unclassified Eukaryota phylum; Dinophyceae; Peridinales; Kryptoperidiniaceae; Durinskia
padina	2	0.031	Eukaryota; Tardigrada; Eutardigrada; Parachela; Macrobiotidae; Macrobiotus
padina	3	0.100	Bacteria; Proteobacteria; Gammaproteobacteria; Oceanospirillales; Saccharospirillaceae; Reinekea
padina	3	0.087	Bacteria; Cyanobacteria; Cyanobacteriia; Cyanobacteriales; Oscillatoriaceae; Planktothricoides SR001
padina	3	0.070	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae; Agaribacter
padina	3	0.049	Bacteria; Bacteroidota; Bacteroidia; Cytophagales; Cyclobacteriaceae; Fabibacter
padina	3	0.035	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae; Aestuariibacter
padina	3	- 0.011	Eukaryota; Ciliophora; Spirotrichea; Euplotida; Aspidiscidae; Aspidisca
padina	3	- 0.045	Eukaryota; unclassified Eukaryota phylum; unclassified Eukaryota class; unclassified Eukaryota order; Colpodellaceae; Colpodella
padina	4	0.435	Eukaryota; unclassified Eukaryota phylum; Dinophyceae; Peridinales; Peridiniaceae; Bysmatrum
padina	4	0.378	Eukaryota; unclassified Eukaryota phylum; Dinophyceae; Peridinales; Amphidiniopsidaceae; Herdmania
PRUS	1	0.408	Bacteria; Firmicutes; Bacilli; Staphylococcales; Staphylococcaceae; Staphylococcus
PRUS	1	0.389	Bacteria; Bacteroidota; Bacteroidia; Bacteroidales; Tannerellaceae; Macellibacteroides
PRUS	1	0.362	Bacteria; Bacteroidota; Bacteroidia; Cytophagales; Cyclobacteriaceae; Ekhidna
PRUS	1	0.354	Bacteria; Bacteroidota; Bacteroidia; Chitinophagales; Saprospiraceae; Lewinella
PRUS	1	0.353	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Thaumasiovibrio
PRUS	2	0.000	Bacteria; Actinobacteriota; Actinobacteria; Corynebacteriales; Mycobacteriaceae; Mycobacterium

Table A10 Continued.

sam. type	clust.	dist.	lineage
PRUS	3	0.000	Bacteria; Bacteroidota; Bacteroidia; Flavobacteriales; Flavobacteriaceae; NA
PRUS	4	0.319	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Alteromonadaceae; Alteromonas
PRUS	4	0.250	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Pseudoalteromonadaceae; Pseudoalteromonas
sediment	1	0.047	Archaea; Nanoarchaeota; Nanoarchaeia; Woeseearchaeales; Woeseearchaeales; Woeseearchaeales
sediment	1	0.028	Archaea; Nanoarchaeota; Nanoarchaeia; Woeseearchaeales; GW2011_GWC1_47_15; GW2011_GWC1_47_15
sediment	1	0.160	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; NA; NA
sediment	1	0.156	Bacteria; Proteobacteria; Gammaproteobacteria; KI89A_clade; KI89A_clade; KI89A_clade
sediment	1	0.153	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; Actibacterium
sediment	1	0.145	Bacteria; Myxococcota; bacteriap25; bacteriap25; bacteriap25; bacteriap25
sediment	1	0.136	Bacteria; Verrucomicrobiota; Verrucomicrobiae; Verrucomicrobiales; Rubritaleaceae; Roseibacillus
sediment	1	0.135	Eukaryota; Ciliophora; Colpodea; Colpodida; Marynidae; Maryna
sediment	1	0.134	Eukaryota; Chlorophyta; Ulvophyceae; Ulvales; Kornmanniaceae; Pseudendoclonium
sediment	1	0.128	Eukaryota; Blastocladiomycota; Blastocladiomycetes; Blastocladiiales; Catenariaceae; Catenophlyctis
sediment	1	0.124	Eukaryota; Discosea; Flabellinia; Dactylopodida; Paramoebidae; Paramoeba
sediment	1	0.107	Eukaryota; Ciliophora; Heterotrichea; Heterotrichida; Peritromidae; Peritromus
sediment	2	0.038	Archaea; Thermoplasmatota; Thermoplasmata; Marine_Benthic_Group_D_and_DHVEG-1; Marine_Benthic_Group_D_and_DHVEG-1; Marine_Benthic_Group_D_and_DHVEG-1
sediment	2	0.027	Archaea; Nanoarchaeota; Nanoarchaeia; Woeseearchaeales; NA; NA
sediment	2	0.023	Archaea; Euryarchaeota; Thermococci; Methanofastidiosales; uncultured; uncultured
sediment	2	0.013	Archaea; Crenarchaeota; Bathyarchaeia; Bathyarchaeia; Bathyarchaeia; Bathyarchaeia
sediment	2	- 0.001	Archaea; Altiarchaeota; Altiarchaeia; Altiarchaeales; Altiarchaeaceae; Candidatus Altiarchaeum

Table A10 Continued.

sam. type	clust.	dist.	lineage
sediment	2	0.061	Bacteria; Marinimicrobia_(SAR406_clade); Marinimicrobia_(SAR406_clade); Marinimicrobia_(SAR406_clade); Marinimicrobia_(SAR406_clade); Marinimicrobia_(SAR406_clade)
sediment	2	0.038	Bacteria; Spirochaetota; Spirochaetia; Spirochaetales; Spirochaetaceae; NA
sediment	2	0.035	Bacteria; Desulfobacterota; Desulfobacteria; Desulfobacterales; Desulfosarcinaceae; LCP-80
sediment	2	0.026	Bacteria; Desulfobacterota; Desulfobacteria; Desulfobacterales; Desulfosarcinaceae; NA
sediment	2	0.023	Bacteria; Chloroflexi; Anaerolineae; Anaerolineales; Anaerolineaceae; uncultured
sediment	2	0.020	Eukaryota; Chlorophyta; Ulvophyceae; Cladophorales; Cladophoraceae; Cladophora
sediment	2	0.009	Eukaryota; Chlorophyta; Ulvophyceae; Ulvales; Ulvaceae; Ulva
sediment	2	0.006	Eukaryota; Ciliophora; Litostomatea; Pleurostomatida; Litonotidae; Epiphyllum
sediment	2	0.004	Eukaryota; Bacillariophyta; Coscinodiscophyceae; Chaetocerotales; Chaetocerotaceae; Chaetoceros
sediment	2	0.000	Eukaryota; Nemertea; Palaeonemertea; Archinemertea; Cephalotrichidae; Cephalothrix
sediment	3	0.126	Bacteria; Chloroflexi; Anaerolineae; SBR1031; NA; NA
sediment	3	0.089	Bacteria; Proteobacteria; Gammaproteobacteria; Gammaproteobacteria_Incertae_Sedis; Unknown_Family; Unknown_Family
sediment	3	0.046	Bacteria; Firmicutes; NA; NA; NA; NA
sediment	3	- 0.004	Bacteria; Bacteroidota; Bacteroidia; Sphingobacteriales; NS11- 12_marine_group; NS11-12_marine_group
sediment	3	- 0.024	Bacteria; Proteobacteria; Gammaproteobacteria; OM182_clade; OM182_clade; OM182_clade
sediment	3	0.102	Eukaryota; unclassified Eukaryota phylum; Dinophyceae; Peridinales; Amphidiniopsidaceae; Herdmania
sediment	3	0.094	Eukaryota; Euglenozoa; Diplonemea; unclassified Diplonemea order; Diplonemidae; Diplonema
sediment	4	0.000	Eukaryota; Annelida; Polychaeta; Phyllodocida; Syllidae; Sphaerosyllis
water	1	0.092	Archaea; Nanoarchaeota; Nanoarchaeia; Woesearchaeales; SCGC_AAA011-D5; SCGC_AAA011-D5
water	1	0.218	Bacteria; Cyanobacteria; Cyanobacteriia; Synechococcales; Cyanobiaceae; Prochlorococcus_MIT9313
water	1	0.203	Bacteria; Proteobacteria; Alphaproteobacteria; SAR11_clade; Clade_II; Clade_II

Table A10 Continued.

sam. type	clust.	dist.	lineage
Water	1	0.203	Bacteria; Proteobacteria; Alphaproteobacteria; SAR11_clade; Clade_I; NA
water	1	0.181	Bacteria; Proteobacteria; Alphaproteobacteria; SAR11_clade; Clade_I; Clade_Ia
water	1	0.159	Bacteria; Verrucomicrobiota; Verrucomicrobiae; Opitutales; Puniceicoccaceae; Lentimonas
water	1	0.143	Eukaryota; Ascomycota; Eurotiomycetes; Eurotiales; Trichocomaceae; Talaromyces
water	1	0.124	Eukaryota; Bacillariophyta; Bacillariophyceae; Bacillariales; Bacillariaceae; Psammodictyon
water	1	0.122	Eukaryota; Chlorophyta; Ulvophyceae; Ulvales; Ulvaceae; Ulva
water	1	0.104	Eukaryota; Ciliophora; Phyllopharyngea; Dysteriida; Dysteriidae; Dysteria
water	1	0.095	Eukaryota; Ciliophora; Spirotrichea; unclassified Spirotrichea order; Tontoniidae; Pseudotontonia
water	2	0.141	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales; Vibrionaceae; Thaumasiobacter
water	2	0.088	Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales; Shewanellaceae; Ferrimonas
water	2	0.046	Bacteria; Cyanobacteria; Cyanobacteriia; Synechococcales; Synechococcales Incertae Sedis; Schizothrix LEGE 07164
water	2	0.046	Bacteria; Bacteroidota; Bacteroidia; Flavobacteriales; Flavobacteriaceae; NA
water	2	0.044	Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales; Hyphomonadaceae; uncultured
water	2	0.185	Eukaryota; Porifera; Calcarea; Clathrinida; Leucettidae; Pericharax
water	2	0.176	Eukaryota; Porifera; Demospongiae; Poecilosclerida; Microcionidae; Clathria
water	2	0.164	Eukaryota; Rhodophyta; Florideophyceae; Bonnemaisoniales; Bonnemaisoniaceae; Bonnemaisonia
water	2	0.161	Eukaryota; Cnidaria; Hydrozoa; Limnomedusae; Olindiidae; Craspedacusta
water	2	0.144	Eukaryota; Rhodophyta; Florideophyceae; Corallinales; Porolithaceae; Porolithon
water	3	- 0.018	Bacteria; Patescibacteria; Gracilibacteria; Candidatus_Peribacteria; Candidatus_Peribacteria; Candidatus_Peribacteria
water	3	- 0.019	Bacteria; Cyanobacteria; Vampirivibrionia; Gastranaerophilales; Gastranaerophilales; Gastranaerophilales
water	3	- 0.038	Bacteria; Planctomycetota; Phycisphaerae; Phycisphaerales; Phycisphaeraceae; Urania-1B-19 marine sediment group
water	3	- 0.064	Bacteria; Myxococcota; Polyangia; Haliangiales; Haliangiaceae; Haliangium

Table A10 Continued.

sam. type	clust.	dist.	lineage
water	3	- 0.091	Bacteria; Acidobacteriota; Thermoanaerobaculia; Thermoanaerobaculales; Thermoanaerobaculaceae; Subgroup_10
water	3	0.086	Eukaryota; Ciliophora; Litostomatea; Cyclotrichida; Mesodiniidae; Mesodinium
water	3	0.079	Eukaryota; Chordata; Actinopteri; Pleuronectiformes; Soleidae; Aesopia
water	3	0.068	Eukaryota; unclassified Eukaryota phylum; unclassified Eukaryota class; Apusomonadida; Apusomonadidae; Amastigomonas
water	3	0.024	Eukaryota; Porifera; Demospongiae; Clionaida; Clionaidae; Cliona
water	3	- 0.011	Eukaryota; unclassified Eukaryota phylum; Phaeophyceae; Ectocarpales; Chordariaceae; Leathesia
water	4	0.205	Archaea; Thermoplasmatota; Thermoplasmata; Marine_Group_II; Marine_Group_II; Marine_Group_II
water	4	0.290	Bacteria; Proteobacteria; Alphaproteobacteria; Puniceispirillales; SAR116 clade; SAR116 clade
water	4	0.273	Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales; Rhodobacteraceae; HIMB11
water	4	0.259	Bacteria; Bacteroidota; Bacteroidia; Flavobacteriales; Flavobacteriaceae; NS5 marine group
water	4	0.228	Bacteria; Firmicutes; Clostridia; Lachnospirales; Lachnospiraceae; NA
water	4	0.217	Bacteria; Verrucomicrobiota; Verrucomicrobiae; Arctic97B-4 marine group; Arctic97B-4 marine group; Arctic97B-4 marine group
water	4	0.245	Eukaryota; Chlorophyta; Mamiellophyceae; Mamiellales; Bathycoccaceae; Ostreococcus
water	4	0.221	Eukaryota; Endomyxa; Ascetosporea; Haplosporida; Haplosporidiidae; Minchinia
water	4	0.204	Eukaryota; Annelida; Polychaeta; unclassified Polychaeta order; Chaetopteridae; Phyllochaetopterus
water	4	0.201	Eukaryota; Annelida; Polychaeta; unclassified Polychaeta order; Opheliidae; Ophelia
water	4	0.201	Eukaryota; Chlorophyta; Mamiellophyceae; Mamiellales; Mamiellaceae; Micromonas

APPENDIX B

EXTENDED TABLES FROM CHAPTER THREE

Table B1.: Endozoicomonas 16S rRNA gene and amplicons resembled one another and point towards a couple dominant sequence variants. Percentage similarity Endozoicomonas 16S rRNA sequences sourced from 16S rRNA gene amplicon sequencing, 16S rRNA transcript fragments, and 16S rRNA assembled transcripts from an all-against-all BLASTn alignment. Each amplicon sequence variant or transcript fragment is listed with their most similar assembled transcript, with the alignment with the highest percent identity bolded.

ASV or Fragment ID	Most similar transcript(s)	BLASTn Percent ID
ASV_1	TRINITY_DN2_c2_g1_i8 & TRINITY_DN2_c2_g1_i2	100.0
ASV_1	TRINITY_DN0_c0_g1_i2	89.7
ASV_2	TRINITY_DN2_c2_g1_i8 & TRINITY_DN2_c2_g1_i2	99.6
ASV_2	TRINITY_DN0_c0_g1_i2	89.7
ASV_3	TRINITY_DN2_c2_g1_i8 & TRINITY_DN2_c2_g1_i2	95.8
ASV_3	TRINITY_DN0_c0_g1_i2	89.7
ASV_4	TRINITY_DN2_c2_g1_i8 & TRINITY_DN2_c2_g1_i2	99.6
ASV_4	TRINITY_DN0_c0_g1_i2	89.7
ASV_5	TRINITY_DN0_c0_g1_i26 & TRINITY_DN0_c0_g1_i10	73.5
ASV_6	TRINITY_DN2_c2_g1_i8 & TRINITY_DN2_c2_g1_i2	95.4
ASV_6	TRINITY_DN0_c0_g1_i2	89.7
ASV_7	TRINITY_DN2_c2_g1_i8 & TRINITY_DN2_c2_g1_i2	95.8
ASV_7	TRINITY_DN0_c0_g1_i2	88.9
ASV_8	TRINITY_DN2_c2_g1_i8 & TRINITY_DN2_c2_g1_i2	96.6
ASV_8	TRINITY_DN0_c0_g1_i2	88.8
fragment_1	TRINITY_DN2_c2_g1_i8 & TRINITY_DN2_c2_g1_i2	100.0

Table B2.: Field preservation method for coral samples. Dry ice was used to snap freeze corals in the field, until COVID-based supply limitations required us to switch to RNA/ DNA preservation buffer (RNAlater) on ice.

date (yymmdd)	hour	field Preservation method	Sample collected	Sampling type
220329	1800 h	Dry ice	DNA+RNA	platform
220330	0001 h	Dry ice	DNA+RNA	platform
220330	0600 h	Dry ice	DNA+RNA	platform
220330	1200 h	Dry ice	DNA+RNA	platform
220330	1800 h	Dry ice	DNA+RNA	platform
220331	0001 h	Dry ice	DNA+RNA	platform
220331	0600 h	Dry ice	DNA+RNA	platform
220331	1200 h	Dry ice	DNA+RNA	platform
220331	1800 h	Dry ice	DNA+RNA	platform
220401	0001 h	Dry ice	DNA+RNA	platform
220401	0600 h	RNAlater + ice	DNA+RNA	platform
220401	1200 h	RNAlater + ice	DNA+RNA	platform
220401	1800 h	RNAlater + ice	DNA+RNA	platform
220402	0001 h	RNAlater + ice	DNA+RNA	platform
220402	0600 h	RNAlater + ice	DNA+RNA	platform
220402	1200 h	RNAlater + ice	DNA+RNA	platform
220404	1200 h	RNAlater + ice	DNA	reef
220405	0001 h	RNAlater + ice	DNA	reef
220405	1200 h	RNAlater + ice	DNA	reef
220406	0001 h	RNAlater + ice	DNA	reef
220406	1200 h	RNAlater + ice	DNA	reef
220407	0001 h	RNAlater + ice	DNA	reef

Table B3.: Reference genomes for initial read recruitment. All available genomes for *Porites* sp. and *Pocillopora* sp. as of February 2024.

simple name	species	accession
coral	<i>Pocillopora damicornis</i>	GCA_003704095.1
coral	<i>Pocillopora verrucosa</i>	GCA_030620025.1
coral	<i>Pocillopora verrucosa</i>	GCA_014529365.1
coral	<i>Pocillopora meandrina</i>	GCA_942486045.1
coral	<i>Porites australiensis</i>	GCA_022179025.1
coral	<i>Porites lobata</i>	GCA_942486035.1
coral	<i>Porites evermanni</i>	GCA_942486025.1
coral	<i>Porites rus</i>	GCA_900290455.1
coral	<i>Porites lutea</i>	GCA_958299795.1
coral	<i>Porites lutea</i>	GCA_958299805.1
zooxanthellae	<i>Symbiodinium microadriaticum</i>	GCA_001939145.1
zooxanthellae	<i>Symbiodinium</i> sp. clade A Y106	GCA_003297005.1
zooxanthellae	<i>Symbiodinium kawagutii</i>	GCA_009767595.1
zooxanthellae	<i>Symbiodinium natans</i>	GCA_905221605.1
zooxanthellae	<i>Symbiodinium</i> sp. CCMP2592	GCA_905221615.1
zooxanthellae	<i>Symbiodinium pilosum</i>	GCA_905231905.1
zooxanthellae	<i>Symbiodinium necroappetens</i>	GCA_905231915.1
zooxanthellae	<i>Symbiodinium</i> sp. CCMP2456	GCA_905221635.1
zooxanthellae	<i>Symbiodinium microadriaticum</i>	GCA_018327485.1

Table B4.: Reference genomes for secondary read recruitment. 1-4 genomes accessions per taxa identified from rRNA screening (available per February 2024)

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Bacteria	Acidobacteriota	<i>Acanthopleuribacter</i>	genus	GCA_017377855.1//
Bacteria	Acidobacteriota	<i>Thermoanaerobaculum</i>	genus	GCA_000687145.1// GCA_025998615.1// GCA_937934115.1//
Bacteria	Actinomycetota	Urania-1B-19 marine sediment group	family	GCA_000023265.1// GCA_000949255.1// GCA_000949295.1// GCA_023130475.1
Bacteria	Actinomycetota	<i>Ilumatobacter</i>	genus	GCA_000348785.1// GCA_004364865.1// GCA_000350145.1//
Bacteria	Actinomycetota	<i>Micrococcus</i>	genus	GCA_020097155.1// GCA_019890915.1// GCA_023573625.1// GCA_000350145.1
Bacteria	Actinomycetota	<i>Mycobacterium</i>	genus	GCA_000195955.2// GCA_000157895.2// GCA_010727605.1//

Table B4 Continued.

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Bacteria	Bacillota/Firmicutes	Staphylococcus	genus	GCA_000013425.1// GCA_006094375.1
Bacteria	Bacillota/Firmicutes	Alkalibacterium	genus	GCA_007991315.1// GCA_007988845.1
Bacteria	Bacillota/Firmicutes	Clostridium	genus	GCA_018223745.1// GCA_001484725.1// GCA_000010265.1// GCA_003606265.1
Bacteria	Bacillota/Firmicutes	Pir4 lineage	family	GCA_009831375.1// GCA_010669205.1// GCA_033096465.1// GCA_034356035.1
Bacteria	Bacteroidota	Balneola	genus	GCA_000375465.1//
Bacteria	Bacteroidota	Candidatus Amoebophilus	genus	GCA_000020565.1// GCA_001897825.1
Bacteria	Bacteroidota	Ekhidna	genus	GCA_900188325.1// GCA_963691765.1
Bacteria	Bacteroidota	Flexithrix	genus	GCA_000379765.1//
Bacteria	Bacteroidota	Marinoscillum	genus	GCA_024714495.1// GCA_003386095.1// GCA_902506655.1//
Bacteria	Bacteroidota	Roseivirga	genus	GCA_034376045.1// GCA_004340805.1// GCA_001747105.1//
Bacteria	Bacteroidota	Hymenobacter	genus	GCA_022919135.1// GCA_022919215.1// GCA_022921055.1//
Bacteria	Bacteroidota	Microscilla	genus	GCA_000169175.1//
Bacteria	Bacteroidota	Crocinitomix	genus	GCA_001661595.2// GCA_000621625.1
Bacteria	Bacteroidota	Cryomorpha	genus	GCA_010686655.1//
Bacteria	Bacteroidota	Vicingus	genus	GCA_007993035.1//
Bacteria	Bacteroidota	Aquimarina	genus	GCA_000430645.1//
Bacteria	Bacteroidota	Aureicoccus	genus	GCA_002954325.1//
Bacteria	Bacteroidota	Croceivirga	genus	GCA_002072105.2// GCA_012272815.1// GCA_014235875.1//
Bacteria	Bacteroidota	Dokdonia	genus	GCA_001653755.1// GCA_003688895.1
Bacteria	Bacteroidota	Enhygromyxa	genus	GCA_002994635.1//
Bacteria	Bacteroidota	Gilvibacter	genus	GCA_028561565.1//
Bacteria	Bacteroidota	Lutibacter	genus	GCA_001543325.1// GCA_003426875.1// GCA_003260195.1//
Bacteria	Bacteroidota	Muricauda	genus	GCA_028464105.1// GCA_013407935.1
Bacteria	Bacteroidota	Tenacibaculum	genus	GCA_900198195.1// GCA_027941775.1// GCA_004803875.1//
Bacteria	Bacteroidota	Winogradskyella	genus	GCA_013394655.1// GCA_013394595.1// GCA_013394675.1//

Table B4 Continued.

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Bacteria	Bacteroidota	Owenweeksia	genus	GCA_000236705.1//
Bacteria	Bacteroidota	Portibacter	genus	GCA_021533195.1// GCA_030160415.1
Bacteria	Bacteroidota	Aureispira	genus	GCA_026000115.1// GCA_031326345.1
Bacteria	Bacteroidota	Lewinella	genus	GCA_000379805.1//
Bacteria	Bdellovibrionota	Peredibacter	genus	GCA_034259205.1//
Bacteria	Bdellovibrionota	NA	family	GCA_034259205.1// GCA_006874605.1// GCA_035013165.1//
Bacteria	Candidatus Melainabacteria	Vampirovibrio	genus	GCA_001858525.1// GCA_024998755.1
Bacteria	Chlamydiota	Simkania	genus	GCA_000237205.1// GCA_017116025.1
Bacteria	Chlorobi	NA	order	GCA_000020465.1// GCA_000012485.1// GCA_000006985.1// GCA_000020525.1
Bacteria	Chloroflexi	NA	family	GCA_003966975.1// GCA_000199675.1// GCA_900184705.1// GCA_001050275.1
Bacteria	Chloroflexi	Litorilinea	genus	GCA_006569185.2// GCA_026003675.1
Bacteria	Cyanobacteriota	Acaryochloris	genus	GCA_000018105.1// GCA_003231495.1
Bacteria	Cyanobacteriota	Chroococcus	genus	GCA_017312365.1// GCA_019359635.1
Bacteria	Cyanobacteriota	Cyanobacterium	genus	GCA_000317675.1// GCA_947331815.1
Bacteria	Cyanobacteriota	Crustifilum	family	GCA_000317515.1// GCA_024178385.1// GCA_000155555.1//
Bacteria	Cyanobacteriota	Hormoscilla	genus	GCA_963921275.2// GCA_014323675.1
Bacteria	Cyanobacteriota	Roseofilum	genus	GCA_030068455.1// GCA_030068505.1// GCA_030068475.1//
Bacteria	Cyanobacteriota	Leptodesmis	genus	GCA_021379005.1//
Bacteria	Cyanobacteriota	Leptolyngbya	genus	GCA_002142475.1// GCA_002215035.1// GCA_015207065.1//
Bacteria	Cyanobacteriota	Stenomitos	genus	GCA_003003795.1//
Bacteria	Cyanobacteriota	Halomicronema	genus	GCA_002075285.3// GCA_016888165.1
Bacteria	Cyanobacteriota	Halotia	genus	GCA_029953635.1//
Bacteria	Cyanobacteriota	Mastigocoleus	genus	GCA_001456025.1// GCA_030065675.1
Bacteria	Cyanobacteriota	Richelia	genus	GCA_019056575.1// GCA_000350105.1
Bacteria	Cyanobacteriota	Rivularia	genus	GCA_000316665.1// GCA_963665115.1

Table B4 Continued.

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Bacteria	Cyanobacteriota	Mastigocladopsis	genus	GCA_000315565.1//
Bacteria	Cyanobacteriota	Planktothricoides	genus	GCA_014698535.1//
Bacteria	Cyanobacteriota	Trichodesmium	genus	GCA_018457135.1// GCA_913057915.1// GCA_022448615.1//
Bacteria	Cyanobacteriota	Lyngbya	genus	GCA_000817775.3// GCA_000478195.2
Bacteria	Cyanobacteriota	Oscillatoria	genus	GCA_000317105.1// GCA_000317475.1// GCA_020144665.1//
Bacteria	Cyanobacteriota	Phormidium	genus	GCA_023983615.1// GCA_001904775.1// GCA_001904725.1//
Bacteria	Cyanobacteriota	Chroococcidiopsis	genus	GCA_000317125.1// GCA_003991895.1
Bacteria	Cyanobacteriota	Pleurocapsa	genus	GCA_963665995.1// GCA_000317025.1
Bacteria	Cyanobacteriota	Xenococcus	genus	GCA_000332055.1// GCA_030065615.1
Bacteria	Cyanobacteriota	Spirulina	genus	GCA_001890765.1// GCA_000314005.1
Bacteria	Cyanobacteriota	Cyanobium	genus	GCA_000316515.1// GCA_003011885.1
Bacteria	Cyanobacteriota	Synechococcus	genus	GCA_000012525.1// GCA_002018015.1// GCA_029582805.1// GCA_024345595.1
Bacteria	Gemmatimonadetes	BD2-11 terrestrial group	class	GCA_000695095.2// GCA_000522985.1// GCA_000010305.1// GCA_013004105.1
Bacteria	Lentisphaerota	Lentisphaerae	genus	GCA_001804815.1// GCA_020683205.1// GCA_001804885.1//
Bacteria	Myxococcota	P3OB-42	family	GCA_002305895.1// GCA_000331735.1// GCA_000280925.3//
Bacteria	Myxococcota	Nannocystis	genus	GCA_026965555.1// GCA_002343915.1// GCA_026626665.1//
Bacteria	Myxococcota	Sandaracinus	genus	GCA_000737325.2// GCA_020633175.1
Bacteria	Myxococcota	Eel-36e1D6	family	GCA_021404885.1// GCA_035063565.1// GCA_002085155.1//
Bacteria	Myxococcota	Haliangium	genus	GCA_000024805.1// GCA_020724705.1
Bacteria	Myxococcota	Pseudenhymyxa	genus	GCA_011083025.2//
Bacteria	Planctomycetota	Phycisphaera	genus	GCA_000284115.1// GCA_025916675.1

Table B4 Continued.

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Bacteria	Planctomycetota	NA	family	GCA_018398935.1// GCA_013128195.2// GCA_008254045.1// GCA_007747215.1
Bacteria	Planctomycetota	Blastopirellula	genus	GCA_000153105.1// GCA_007859755.1// GCA_020966695.1//
Bacteria	Planctomycetota	Pirellula	genus	GCA_000025185.1// GCA_001610875.1
Bacteria	Planctomycetota	Rhodopirellula	genus	GCA_000196115.1// GCA_000346505.1
Bacteria	Planctomycetota	Gimesia	genus	GCA_009720525.1// GCA_007747015.1// GCA_007748015.1//
Bacteria	Planctomycetota	Planctomicrobium	genus	GCA_900113665.1// GCA_945890625.1
Bacteria	Planctomycetota	Planctomyces	genus	GCA_001610835.1// GCA_001610895.1// GCA_016795105.1//
Bacteria	Pseudomonadota	Filomicrobium	genus	GCA_000981545.1// GCA_900104305.1// GCA_000981565.1// GCA_963669235.1
Bacteria	Pseudomonadota	Rhodobium	genus	GCA_003258835.1// GCA_025961475.1
Bacteria	Pseudomonadota	Labrenzia	genus	GCA_009363435.1// GCA_009363595.1// GCA_006517275.1//
Bacteria	Pseudomonadota	Hyphomonas	genus	GCA_000013025.1// GCA_017795365.1// GCA_000682755.1//
Bacteria	Pseudomonadota	Kordiimonas	genus	GCA_015240255.1// GCA_000375545.1// GCA_014656395.1//
Bacteria	Pseudomonadota	Parvularcula	genus	GCA_000152825.2// GCA_000733125.1// GCA_003399445.1//
Bacteria	Pseudomonadota	Rhodovulum	genus	GCA_001633145.1// GCA_003265965.1// GCA_004342445.1//
Bacteria	Pseudomonadota	Pseudoceanicola	genus	GCA_003590145.2// GCA_014899185.1// GCA_017795185.1//
Bacteria	Pseudomonadota	Roseobacter	genus	GCA_000154785.2// GCA_012932215.1// GCA_002983865.1// GCA_019375555.1
Bacteria	Pseudomonadota	Roseovarius	genus	GCA_008728195.1// GCA_025639885.1// GCA_009762325.1//
Bacteria	Pseudomonadota	Rubricella	genus	GCA_014199445.1//

Table B4 Continued.

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Bacteria	Pseudomonadota	Ruegeria	genus	GCA_000011965.2// GCA_000192475.2// GCA_005938215.1//
Bacteria	Pseudomonadota	Sagittula	genus	GCA_000169415.1// GCA_014196795.1
Bacteria	Pseudomonadota	Pseudoroseomonas	genus	GCA_000164635.1// GCA_003604215.1
Bacteria	Pseudomonadota	Fodinicurvata	genus	GCA_000420625.1// GCA_000686045.1
Bacteria	Pseudomonadota	Fodinicurvata	genus	GCA_000420625.1// GCA_000686045.1
Bacteria	Pseudomonadota	Pelagibius	genus	GCA_014925385.1// GCA_011683915.1
Bacteria	Pseudomonadota	Rhodovibrio	genus	GCA_000515255.1// GCA_016583645.1
Bacteria	Pseudomonadota	Defluviicoccus	genus	GCA_014672695.1// GCA_013414705.1
Bacteria	Pseudomonadota	Defluviicoccus	genus	GCA_014672695.1// GCA_013414705.1
Bacteria	Pseudomonadota	Thalassospira	genus	GCA_000300235.2// GCA_003403095.1// GCA_002844375.1//
Bacteria	Pseudomonadota	Terasakiella	genus	GCA_014641495.1// GCA_000688235.1
Bacteria	Pseudomonadota	Candidatus Lariskella	family	GCA_000219355.1// GCA_029981845.1// GCA_003072485.1// GCA_000815465.1
Bacteria	Pseudomonadota	Erythrobacter	genus	GCA_001719165.1// GCA_023823125.1// GCA_001663175.1//
Bacteria	Pseudomonadota	Candidatus Alysiosphaera	family	GCA_000264455.2// GCA_000427665.1
Bacteria	Pseudomonadota	Micropelagos	genus	GCA_000469155.1//
Bacteria	Pseudomonadota	Nitratireductor	genus	GCA_007922615.2// GCA_019891395.1// GCA_000300335.1// GCA_019891355.1
Bacteria	Pseudomonadota	Aeromonas	genus	GCA_016127195.1// GCA_900637545.1// GCA_016026875.1// GCA_008693705.1
Bacteria	Pseudomonadota	Aestuariibacter	genus	GCA_000429145.1// GCA_020687825.1
Bacteria	Pseudomonadota	Alteromonas	genus	GCA_001562195.1// GCA_000730385.1// GCA_000213655.1//
Bacteria	Pseudomonadota	Glaciecola	genus	GCA_000226565.1// GCA_000428905.1// GCA_001756615.1//
Bacteria	Pseudomonadota	Thalassomonas	genus	GCA_000948985.2// GCA_000948975.2// GCA_028657945.1//

Table B4 Continued.

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Bacteria	Pseudomonadota	Thalassotalea	genus	GCA_030295975.1// GCA_030295995.1// GCA_030295935.1//
Bacteria	Pseudomonadota	Algicola	genus	GCA_000374485.1//
Bacteria	Pseudomonadota	Pseudoalteromonas	genus	GCA_001444405.1// GCA_002310835.1// GCA_000238395.4// GCA_000238255.4
Bacteria	Pseudomonadota	Porticoccus	genus	GCA_000744735.1// GCA_030676955.1
Bacteria	Pseudomonadota	Pseudohalica	genus	GCA_000764025.1//
Bacteria	Pseudomonadota	Immundisolibacter	genus	GCA_001697225.1//
Bacteria	Pseudomonadota	Coxiella	genus	GCA_000007765.2// GCA_018107685.1// GCA_907164955.1// GCA_000815025.1
Bacteria	Pseudomonadota	Marinicella	genus	GCA_004363335.1// GCA_009671245.1// GCA_002000055.1//
Bacteria	Pseudomonadota	Endozoicomonas	genus	GCA_025562755.1// GCA_001583435.1// GCA_002864045.1// GCA_000710775.1
Bacteria	Pseudomonadota	Halomonas	genus	GCA_000196875.2// GCA_034479615.1// GCA_011399095.1// GCA_030406665.1
Bacteria	Pseudomonadota	Marinomonas	genus	GCA_000192865.1// GCA_000214215.1// GCA_014805825.1//
Bacteria	Pseudomonadota	Neptuniibacter	genus	GCA_030295765.1// GCA_000153345.1// GCA_001597735.1//
Bacteria	Pseudomonadota	Acinetobacter	genus	GCA_000191145.1// GCA_000196795.1// GCA_018324105.1//
Bacteria	Pseudomonadota	Psychrobacter	genus	GCA_001444505.1// GCA_000012305.1// GCA_002208765.2//
Bacteria	Pseudomonadota	Pseudomonas	genus	GCA_000006765.1// GCA_900215245.1// GCA_018394375.1//
Bacteria	Pseudomonadota	Woeseia	genus	GCA_001677435.1//
Bacteria	Pseudomonadota	Caedibacter	genus	GCA_002803295.2//
Bacteria	Pseudomonadota	Fangia	genus	GCA_000379445.1//
Bacteria	Pseudomonadota	Beggiatoa	genus	GCA_001305575.3// GCA_000245015.1
Bacteria	Pseudomonadota	Echinimonas	genus	GCA_023703465.1//
Bacteria	Pseudomonadota	Vibrio	genus	GCA_003391375.1// GCA_001558415.2// GCA_001456255.1//

Table B4 Continued.

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Bacteria	Spirochaetota	Spirochaeta	genus	GCA_000184345.2// GCA_000242595.3// GCA_000426705.1//
Bacteria	Thermodesulfobacteriota	Desulfobacter	genus	GCA_004319545.1// GCA_000233695.3// GCA_013389365.1//
Bacteria	Thermodesulfobacteriota	Desulfocurvus	genus	GCA_000519125.1// GCA_020692325.1
Bacteria	Thermodesulfobacteriota	Desulfohalovibrio	genus	GCA_000711295.1//
Bacteria	Thermodesulfobacteriota	Geopsychrobacter	genus	GCA_000384395.1// GCA_012974425.1
Bacteria	Verrucomicrobiota	Cerasicoccus	genus	GCA_016595555.1// GCA_031296915.1// GCA_031296935.1//
Bacteria	Verrucomicrobiota	Roseibacillus	genus	GCA_014651675.1// GCA_016595415.1
Bacteria	Verrucomicrobiota	Rubritalea	genus	GCA_000378105.1// GCA_002954445.1
Bacteria	Verrucomicrobiota	DEV007	family	GCA_000172155.1// GCA_010681825.2// GCA_012913485.1// GCA_018137965.1
Bacteria	Verrucomicrobiota/ Kiritimatiellota	WCHB1-41	class	GCA_029269125.1//
Eukaryota	Foraminifera	NA	phylum	GCA_000211355.2// GCA_000512085.1// GCA_003354225.1//
Eukaryota	Ascomycota	NA	family	GCA_026367775.1// GCA_014334815.1// GCA_002952155.1//
Eukaryota	Ascomycota	Komagataella	genus	GCA_000027005.1// GCA_029632965.1// GCA_023213065.1//
Eukaryota	Ascomycota	Pichia	genus	GCA_030570135.1//
Eukaryota	Ascomycota	Protocreopsis	family	GCA_026958235.1// GCA_001599755.1
Eukaryota	Ascomycota	NA	family	GCA_022664585.2//
Eukaryota	Basidiomycota	Schizonella	family	GCA_005498985.1// GCA_001010845.1// GCA_900080155.1//
Eukaryota	Chytridiomycota	Pendulichytrium	family	GCA_027604725.1// GCA_027596945.1// GCA_002104985.1//
Eukaryota	Mucoromycota	Endogone	genus	GCA_003990785.1//
Eukaryota	Olpidiomycota	Olpidium	genus	GCA_017911155.1//
Eukaryota	Zoopagomycota	Coemansia	genus	GCA_025766245.1// GCA_027478445.1// GCA_027579525.1//
Eukaryota	Mollusca	Elysia	genus	GCA_003991915.1// GCA_033675545.1// GCA_019649035.1//

Table B4 Continued.

domain	phylum	genus	NCBI Taxonomy level used	NCBI accession(s)
Eukaryota	Platyhelminthes	Macrostomum	genus	GCA_002269645.1// GCA_950097015.1// GCA_950096795.1//
Eukaryota	Oomycota	Eurychasma	phylum	GCA_025722995.1// GCA_030027945.1// GCA_030324255.1// GCA_026184515.1
Eukaryota	NA	Aplanochytrium	family	GCA_000818945.1// GCA_014084085.1// GCA_001462505.1//
Eukaryota	Apicomplexa	Symbiont group	sub-order	GCA_000208865.2// GCA_000006565.2// GCA_963920595.1// GCA_031892065.1
Eukaryota	Ciliophora	NA	phylum	GCA_000189635.1// GCA_000165425.1// GCA_000711775.1// GCA_000220395.1
Eukaryota	Ciliophora	Frontonia	order	GCA_000165425.1// GCA_016759055.1// GCA_000733375.1// GCA_000733385.1
Eukaryota	Ciliophora	Peritromus	order	GCA_001970955.1// GCA_914767885.1// GCA_949751835.1//
Eukaryota	Dinoflagellata	Amphidinium	genus	GCA_019702695.1//
Eukaryota	Dinoflagellata	NA	family	GCA_963575745.1//
Eukaryota	Chlorophyta	Derbesia	order	GCA_030272585.1// GCA_905146915.1
Eukaryota	Chlorophyta	Ostreobium	genus	GCA_905146915.1//
Eukaryota	Chlorophyta	Phaeophila	class	GCA_023078555.1// GCA_024500015.1// GCA_030272585.1// GCA_900538255.1
Eukaryota	Chlorophyta	Ulvella	family	GCA_032359135.1//
Eukaryota	Rhodophyta	Polysiphonia	family	GCA_032358235.1//

Table B5.: Taxonomic breakdown of reads after mapping to reference genomes. PDAM, *Pocillipora damicornis*; PRUS, *Porites rus*.

	F+R total reads	F+R total Gb	percent of total reads	percent of total bases
PDAM	169,242,166	23.60	38.87	38.53
PRUS	138,222,658	19.69	31.74	32.16
zooxanthelle	32,820,294	4.65	7.54	7.59
bacteria	58,068,580	8.05	13.34	13.14
other algae	1,943,024	0.28	0.45	0.45
other heterotrophs	924,966	0.13	0.21	0.21
fungi	43,094	0.0059	0.01	0.01
unmapped	34,159,928	4.84	7.85	7.91

Table B6.: Reference *Endozoicomonas* genomes used in comparative analyses. (available per December 2024)

Genome ID	checkM Completeness	checkM Contamination
GCF_001647025.1 <i>Endozoicomonas</i> sp.	98.92	3.43
GCA_001729985.1 <i>Endozoicomonas</i> sp.	94.85	10.52
GCA_031844845.2 <i>Endozoicomonas</i> sp.	96.17	2.87
GCF_031844845.1 <i>Endozoicomonas</i> sp.	96.17	2.87
GCA_900299555.1 <i>Endozoicomonas</i> sp.	98.61	5.11
GCF_947495845.1 <i>Endozoicomonas</i> sp.	98.69	1.99
GCF_000373945.1 <i>Endozoicomonas</i> sp.	98.98	3.43
GCF_000710775.1 <i>Endozoicomonas</i> sp.	98.98	3.43
GCF_025562715.2 <i>Endozoicomonas</i> sp.	99.14	2.19
GCF_025562755.1 <i>Endozoicomonas</i> sp.	98.71	1.81
GCF_040549035.1 <i>Endozoicomonas</i> sp.	98.78	2.18
GCF_040549045.1 <i>Endozoicomonas</i> sp.	98.56	1.97

Table B6 Continued.

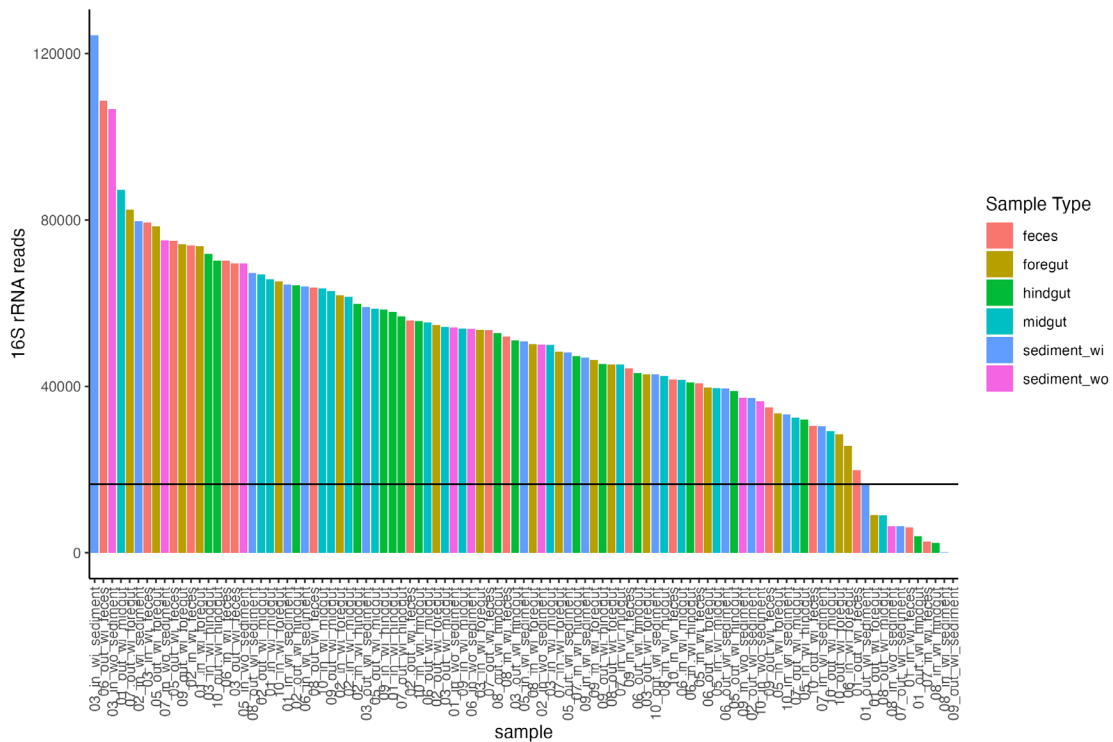
Genome ID	checkM Completeness	checkM Contamination
GCF_040549065.1 Endozoicomonas sp.	98.91	2.18
GCA_001562005.1 Endozoicomonas sp.	85.02	4.97
GCF_000722635.1 Endozoicomonas sp.	99.14	1.66
GCF_001562015.1 Endozoicomonas sp.	99.14	2.05
GCF_009653635.1 Endozoicomonas sp.	99.14	1.94
GCA_024509095.1 Endozoicomonas sp.	88.53	1.8
GCA_027942495.1 Endozoicomonas sp.	89.85	3.13
GCA_027942505.1 Endozoicomonas sp.	86.46	2.73
GCF_002864045.1 Endozoicomonas sp.	98.56	1.39
GCF_010994325.1 Endozoicomonas sp.	98.99	2.25
GCF_010994335.1 Endozoicomonas sp.	98.99	2.25
GCF_024606345.1 Endozoicomonas sp.	98.52	1.3
GCF_024606365.1 Endozoicomonas sp.	87.97	1.73
GCF_027366395.1 Endozoicomonas sp.	98.46	2.25
GCF_027366555.1 Endozoicomonas sp.	97.29	2.25
GCA_024509075.1 Endozoicomonas sp.	90.89	1.9
GCF_000722565.1 Endozoicomonas sp.	98.99	1.72
GCF_001583435.1 Endozoicomonas sp.	98.99	1.72
GCF_023716865.1 Endozoicomonas sp.	98.71	2.81
GCF_023822025.1 Endozoicomonas sp.	97.52	0.98
GCF_024606225.1 Endozoicomonas sp.	98.49	1.06
GCF_024606245.1 Endozoicomonas sp.	81.65	1.59
GCF_024606255.1 Endozoicomonas sp.	59.41	2.05
GCF_024606265.1 Endozoicomonas sp.	95.02	1.79
GCF_024606275.1 Endozoicomonas sp.	49.22	0.26
GCF_024606325.1 Endozoicomonas sp.	90.73	1.49
GCF_032883915.1 Endozoicomonas sp.	99.14	0.54
GCF_040549005.1 Endozoicomonas sp.	98.99	2.18
GCF_001646945.1 Endozoicomonas sp.	98.28	1.81
GCF_001646955.1 Endozoicomonas sp.	97.63	2.02
GCA_001647025.2 Endozoicomonas sp.	98.92	3.43

Table B7.: Expression levels of respiratory reductases in *Endozoicomonas* sp.

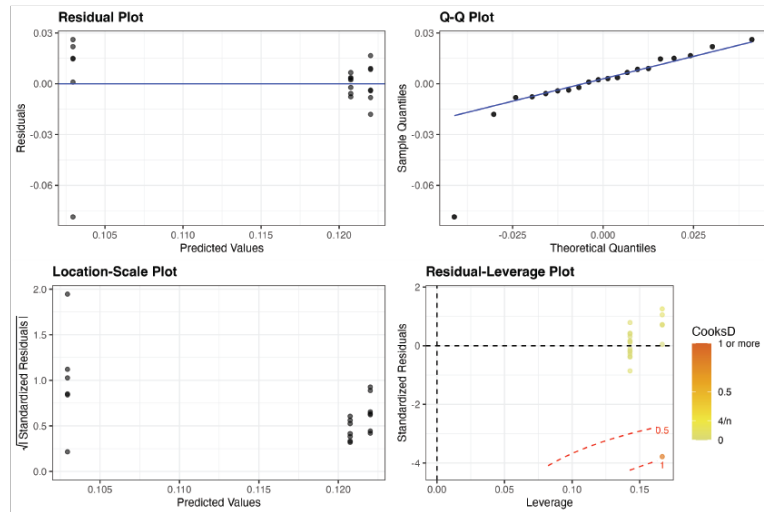
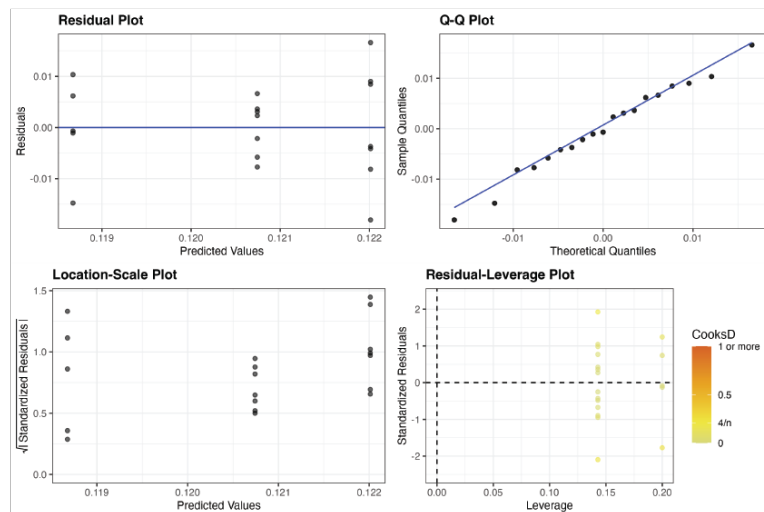
	Maximum TPM	Average TPM
<i>dmsA</i>	124.85	13.28
<i>dmsB</i>	253.36	16.19
<i>cydA</i>	129.90	16.78
<i>cydB</i>	67.22	5.94
<i>ctaD</i>	72.02	3.34
<i>ccoN</i>	180.95	15.49

APPENDIX C

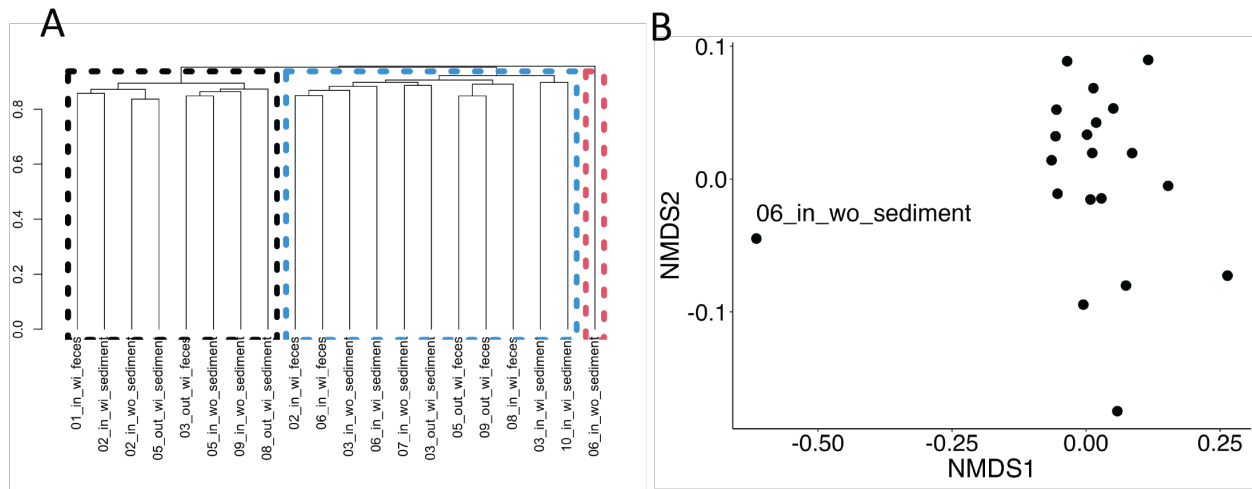
EXTENDED TABLES AND SUPPLEMENTARY FIGURES
FROM CHAPTER FOUR



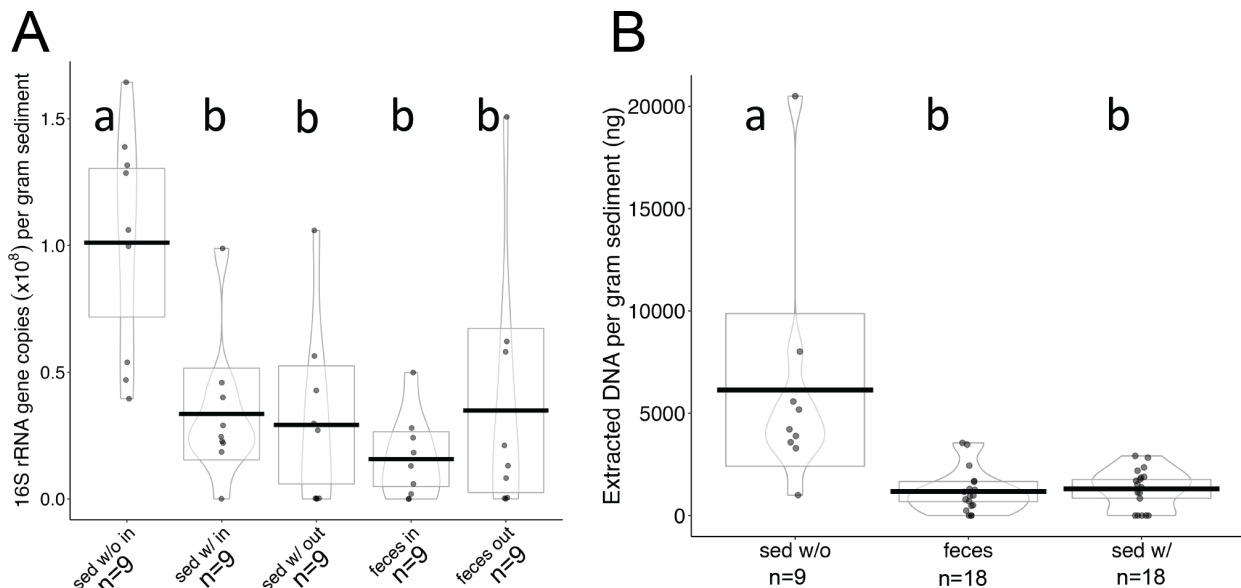
Appendix C Figure 1: Rarefaction of 16S rRNA amplicons. 16S rRNA reads per sample were normalized by rarefaction to a sequencing depth of 16,490 reads per sample (horizontal line) removing 9 samples. Sample type, categorical classification of samples from collection; feces, *Holothuria atra* fecal samples collected proximal to the sea cucumber's anus; foregut, *H. atra* foregut samples; midgut, *H. atra* midgut samples; hindgut, *H. atra* hindgut samples; sediment_wi, sediment samples collected proximal to *H. atra* mouthparts; sediment_wo, sediment collected from within the caged *H. atra* enclosure.

A**B**

Appendix C Figure 2: Diagnostic plots for linear model of genome equivalents per megabase (Mb) by sample type [lm(GenomeEquivPerMb~sample_type)]. **A)** Plots with sample 06_in_wo_sediment included. In the residual vs. fitted plot there is strong evidence against the assumption of equal variance, in the QQ plot, there is strong evidence against the assumption of normal distribution, and in the Residual-Leverage plot, there is a likely influential outlier with a Cooks distance of 0.955. **B)** Plots when sample 06_in_wo_sediment is excluded. In the residual vs. fitted plot there is weak evidence against the assumption of equal variance, in the QQ plot, there is weak evidence against the assumption of normal distribution, and in the Residual-Leverage plot, there are no influential outliers.

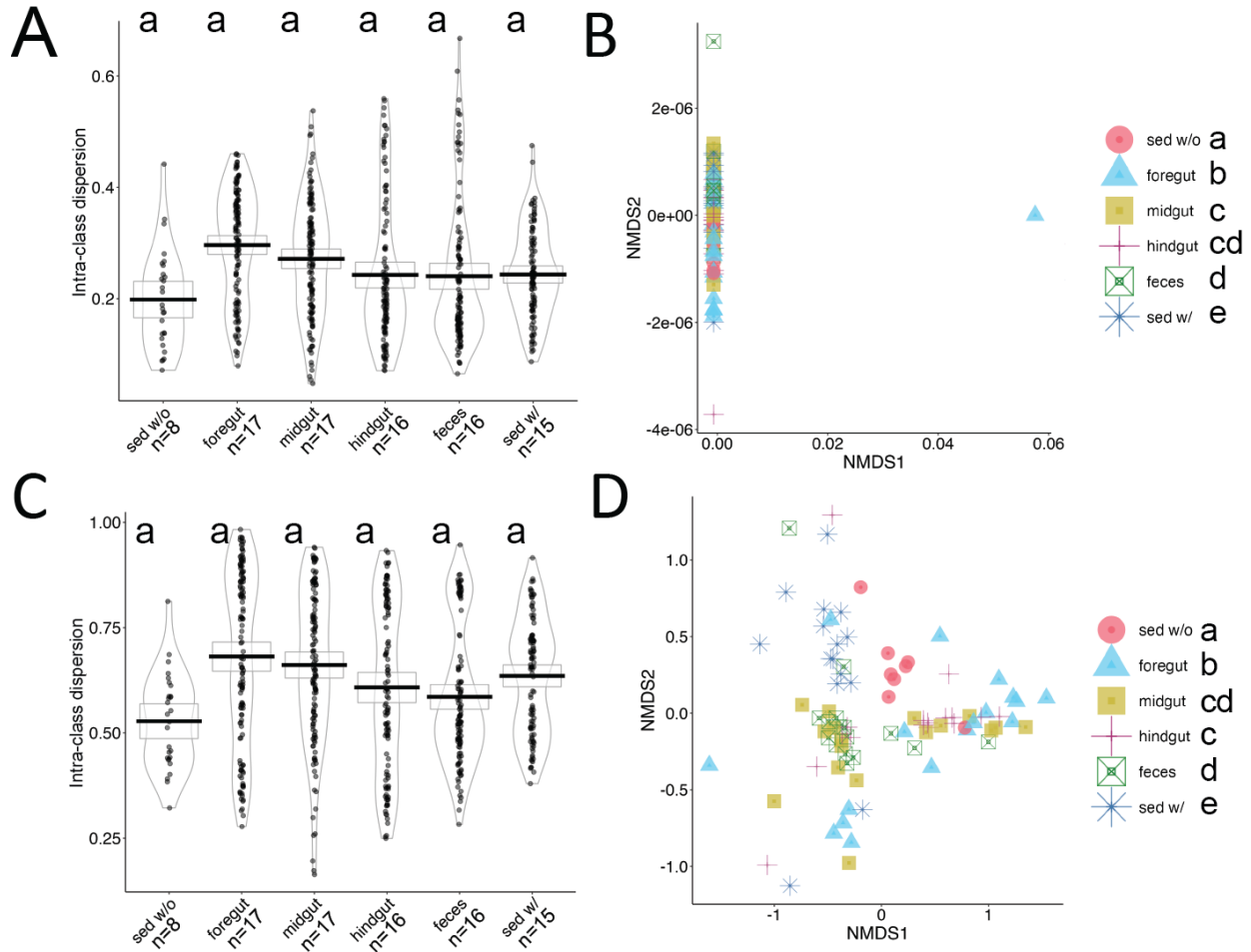


Appendix C Figure 3: Ordination of grazed and ungrazed sediments using k-mer based distance matrices from unbinned metagenomes demonstrates that sample 06_in_wo_sediment deviates from all other samples. A) hierarchical clustering of square root transformed Bray-Curtis distance matrices derived from 21-mer abundances using Ward's Clustering Algorithm as implemented in the agnes function in the cluster package in R. **B)** Non-Metric Multidimensional Scaling (NMDS) plots of non-metric Bray-Curtis distance matrices based on 21-mer abundances of available samples.

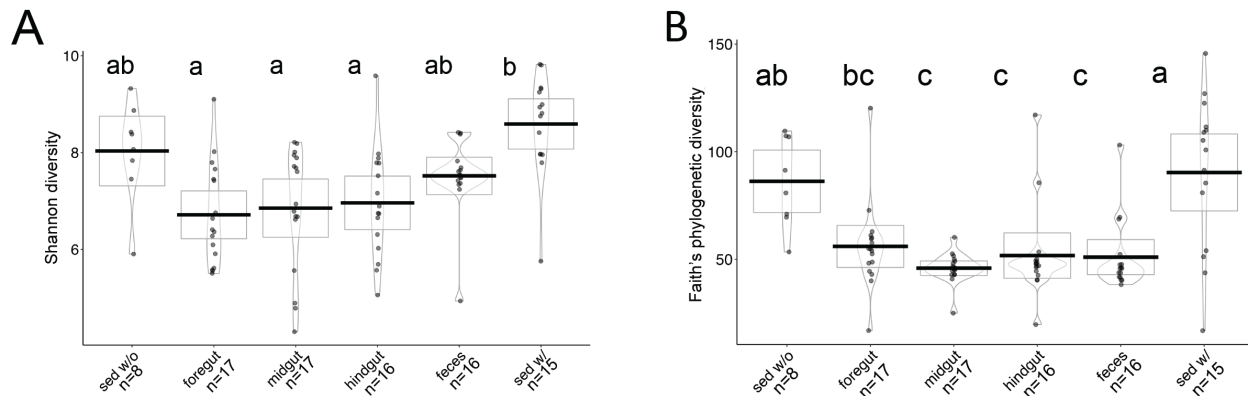


Appendix C Figure 4: Estimating microbial abundances before and after cucumber grazing shows that cucumber grazing reduces proxy measures for microbial load. A) Quantitative PCR of 16S rRNA gene copies normalized to grams of sediment used in DNA extraction. **B)** Extracted DNA per gram of sediment from each sample type normalized to grams of sediment used in DNA extraction. Quantitative measures presented as violin plots (mean as estimate of center, box

bounded by 95% Confidence interval (CI), density estimator to draw violin, observations as individual points). Dissimilar letters indicate $P < 0.01$ between contrasts (p values were obtained from linear mixed effects model; Tukey-Kramer HSD).

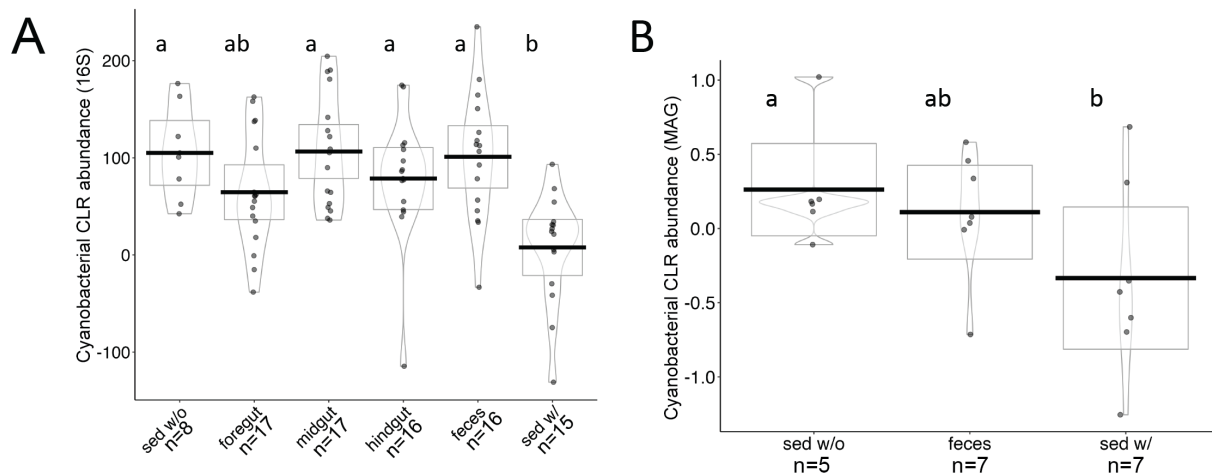


Appendix C Figure 5: Beta diversity of sediment samples before, during and after digestion from 16S rRNA gene data show that sea cucumber grazing alters community composition, but not dispersion. Displaying **AC**) intra-class dispersion (Permutational Analyses Of Multivariate Dispersions; PERMDISP) and **BD**) Non-Metric Multidimensional Scaling (NMDS) of communities using **A,B**) Weighted UniFrac and **C,D**) Bray-Curtis distance matrices. Intra-class dispersion presented as violin plots (mean as estimate of center, box bounded by 95% Confidence interval (CI), density estimator to draw violin, observations as individual points). Dissimilar letters indicate $P < 0.05$ between contrasts. Statistically detectable differences in community composition presented as dissimilar letters on the respective keys (Permutational multivariate analysis of variance; PERMANOVA, $P < 0.05$).



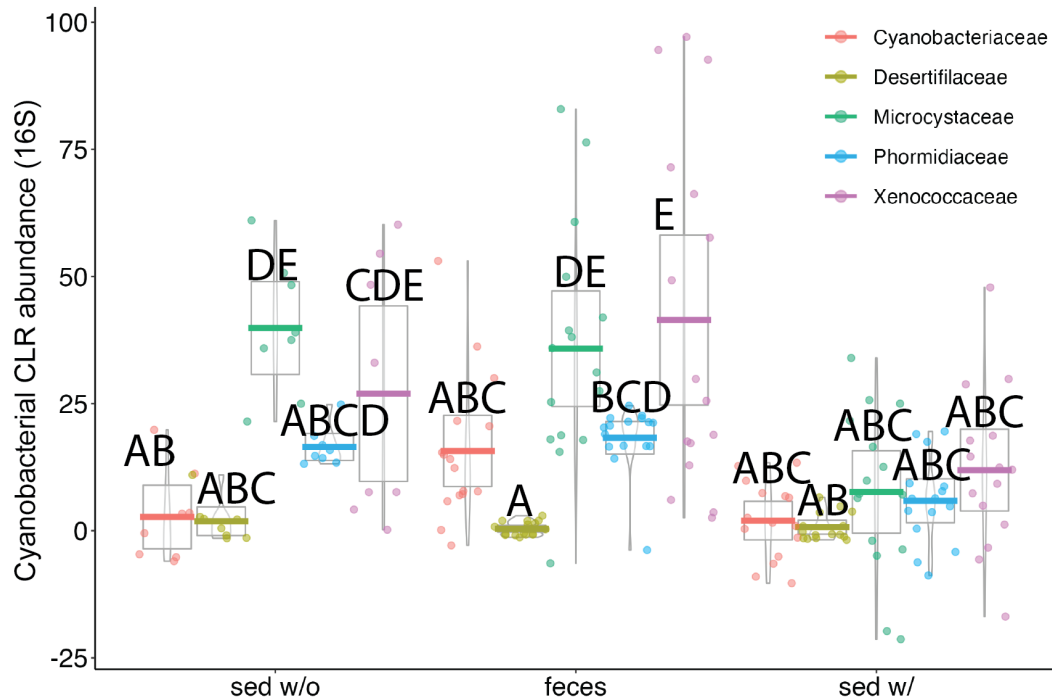
Appendix C Figure 6: Alpha diversity of sediment samples before, during and after sediment digestion from 16S rRNA gene data shows that cucumber GI tract and external sediment differs.

A) Shannon's Diversity and B) Faith's phylogenetic diversity. Diversity metrics presented as violin plots (mean as estimate of center, box bounded by 95% Confidence interval (CI), density estimator to draw violin, observations as individual points). Dissimilar letters indicate $P < 0.05$ between contrasts (p values were obtained from linear mixed effects model; Tukey-Kramer HSD).

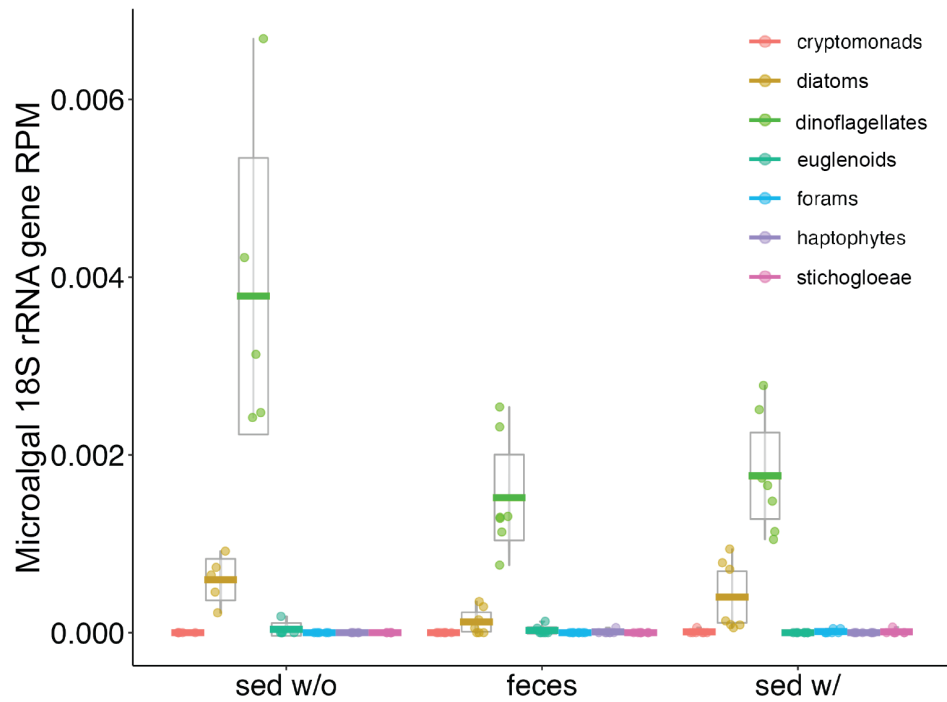


Appendix C Figure 7: Depletion of cyanobacterial transformed relative abundances in grazed sediments. Inferred from **A) 16S rRNA genes** and **B) metagenome assembled genomes (MAGs).**

A) 16S rRNA gene abundances were estimated from the percent of cyanobacterial reads relative to the total reads. **B) Cyanobacterial MAG abundances** were estimated by mapping metagenomic reads to all dereplicated MAGs (99% ANI, >75% completion, <10% contamination). Abundance presented as violin plots (mean as estimate of center, box bounded by 95% Confidence interval (CI), density estimator to draw violin, observations as individual points). Dissimilar letters indicate $P < 0.05$ between contrasts (p values were obtained from linear mixed effects model; Tukey-Kramer HSD). Cyanobacterial percentage abundances were center-log ratio (CLR) transformed to satisfy mixed effects model assumptions of normal distribution and equal variance.

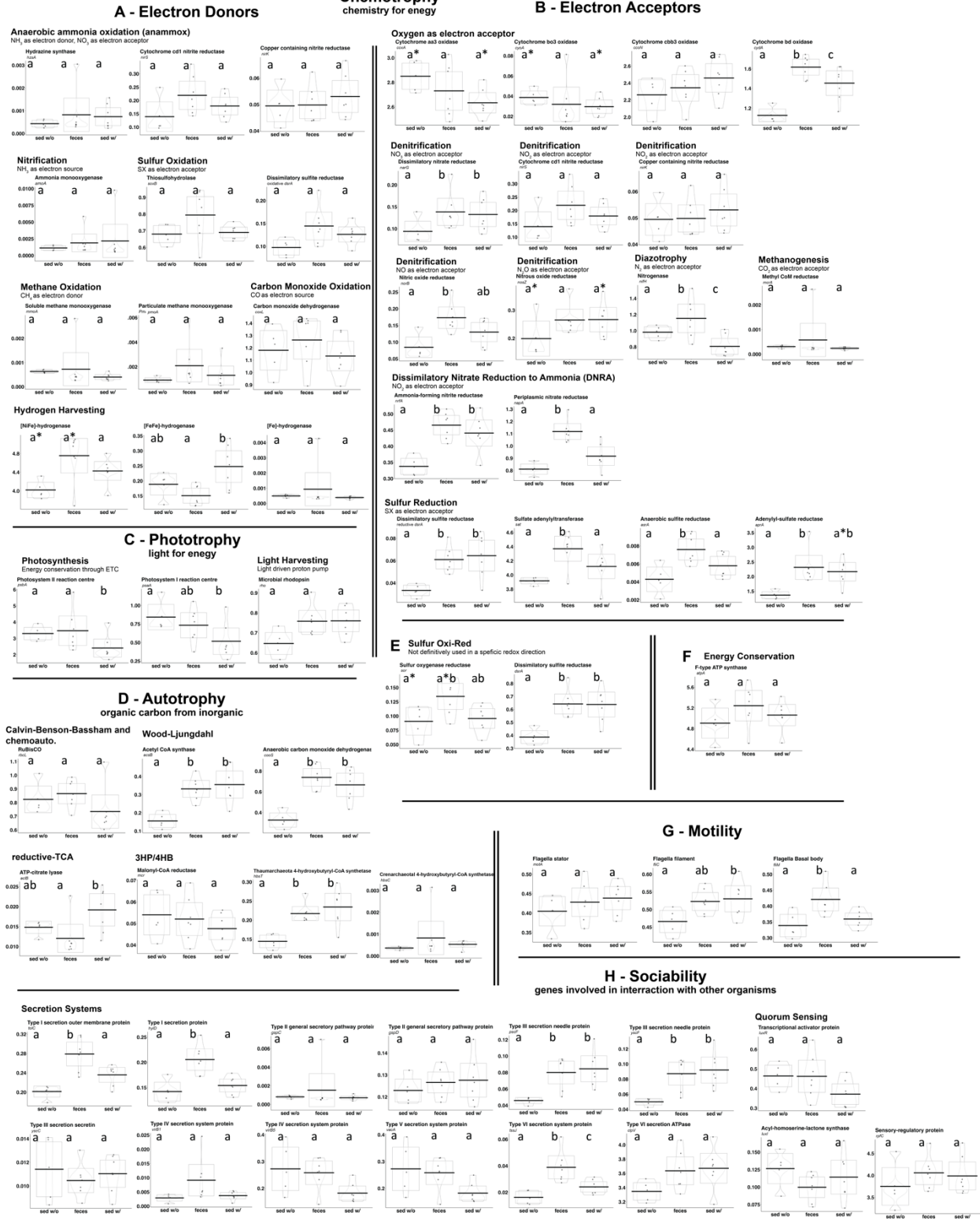


Appendix C Figure 8: Changes in cyanobacterial family abundances before and after grazing from 16S rRNA gene data. Only cyanobacterial families with maximum percentage abundances of >1.5% are shown. Cyanobacterial center-log ratio (CLR) transformed abundances presented as violin plots (mean as estimate of center, box bounded by 95% Confidence interval (CI), density estimator to draw violin, observations as individual points). Dissimilar letters indicate $P < 0.05$ between contrasts (p values were obtained from linear mixed effects model; Tukey-Kramer HSD). CLR transform was implemented to meet mixed effects model assumptions of normal distribution and equal variance.



Appendix C Figure 9: Normalized microalgal abundance of 18S rRNA genes before and after cucumber grazing. Normalized abundance (reads per megabase; RPM) presented as violin plots (mean as estimate of center, box bounded by 95% Confidence interval (CI), density estimator to draw violin, observations as individual points).

Chemotrophy chemistry for energy



Appendix C Figure 14: Inferred metabolic potential of the microbial community from unassembled metagenomes. Relativized abundance of genes as reads per kilobase genome equivalent (RPKG) of diagnostic genes for ungrazed sediment (sed w/o), feces, grazed sediment (sed w/). Genes grouped by functional categories: **A) Electron donors.** Anaerobic ammonia oxidation (anammox): *hzsA*, Hydrazine synthetase; *nirS*, cytochrome cd1 nitrite reductase; *nirK*, copper containing nitrite reductase. Nitrification: *amoA*, ammonia monooxygenase. Sulfur oxidation: *soxB*, Thiosulfohydrolase; oxidative *dsrA*, dissimilatory sulfite reductase. Carbon monoxide oxidation: *coxL*, carbon monoxide dehydrogenase. Methane oxidation: *mmoA*, soluble methane monooxygenase; *pmoA*, particulate methane monooxygenase. Hydrogen harvesting: [NiFe]-Hydrogenase; [FeFe]-hydrogenase; [Fe]-hydrogenase. **B) Electron acceptors.** Oxygen reduction: *coxA*, cytochrome *aa3* oxidase; *cyoA*, cytochrome *bo3* oxidase; *ccoN*, cytochrome *cbb3* oxidase; *cydA*, cytochrome *bd* oxidase. Denitrification: *narG*, dissimilatory nitrate reductase; *nirS*, cytochrome cd1 nitrite reductase; *nirK*, copper containing nitrite reductase; *norB*, nitric oxide reductase; *nosZ*, nitrous oxide reductase. Diazotrophy: *nifH*, nitrogenase. Methanogenesis: *mcrA*, methyl CoM reductase. Dissimilatory nitrate reduction to ammonia: *nrfA*, ammonia-forming nitrite reductase; *napA*, periplasmic nitrate reductase. Sulfur reduction: Reductive *dsrA*, dissimilatory sulfite reductase; *sat*, sulfate adenylyltransferase; *asrA*, anaerobic sulfite reductase; *aprA*, adenylyl-sulfate reductase. **C) Phototrophy.** Photosynthesis: *psbA*, Photosystem II reaction center; *psaA*, Photosystem I reaction center. Light driven proton pump: *rho*, microbial rhodhopsin. **D) Autotrophy.** Calvin-Benson-Bassham and Chemolithoautotrophy: *rbcL*, RuBisCo. Wood-Ljungdahl: *acsB*, Acetyl CoA Synthetase; *cooS*, Anaerobic carbon monoxide dehydrogenase. reductive-TCA: *aclB*, ATP Citrate-lyase; *mcr*. 3HB/4HB: Malonyl CoA Reductase; *hbsT*, Thaumarchaeota 4-hydroxybutyryl-CoA Synthetase; *hbsC*, Crenarchaeotal 4-hydroxybutyryl-CoA Synthetase. **E) Sulfur oxidation/ reduction:** *sor*, sulfur oxygenase reductase; *dsrA*, dissimilatory sulfite reductase. **F) Energy conservation.** *atpA*, F-type ATP synthase. **G) Motility.** *motA*, flagella stator; *fliC*, flagella filament; *fliM*, flagella basal body. **H) Sociability.** Secretion Systems: *tolC*, type I secretion system outer membrane protein; *hlyD*, type I secretion system; *gspC*, type II general secretory pathway protein; *gspD*, type II general secretory pathway protein; *pscF*, type III secretion needle protein; *yscF*, type III secretion needle protein; *yscC*, type III secretion secretin; *virB1*, type IV secretion system protein; *virB5*, type IV secretion system protein; *vacA*, type V secretion system protein; *tssJ*, type VI secretion system protein; *clpV*, type VI secretion system ATPase. Quorum Sensing: *luxR*, Transcriptional activator protein; *luxI*, Acyl-homoserine-lactone synthase; *rpfC*, Sensory-regulatory protein. Gene abundance presented as violin plots (mean as estimate of center, box bounded by 95% Confidence interval (CI), density estimator to draw violin, observations as individual points). Dissimilar letters indicate $P < 0.05$ between contrasts, shared letters with an asterisk indicates $P < 0.10$ between contrasts (Mixed Effects Model, fixed: sample type; random: cucumber samples collected nearest from, cage/uncaged status; Tukey-Kramer HSD).

Table C1.: quantitative PCR (qPCR) and PCR master mix and thermocycler conditions

16S rRNA gene
sequencing

master mix (25 µL total)	PCR conditions
1 µL DNA	95°C 3:00 minutes
12.5 µL Promega Go Taq Master Mix (Madison, WI),	[95°C 0:45 minutes
10.25 µL H ₂ O	55°C 0:45 minutes
0.5 µL F primer	72°C 1:30 minutes]
0.5 µL R primer	25 cycles
0.25 µL Bovine Serum Albumin	

16S rRNA gene qPCR

master mix (10 µL total)	PCR conditions
1 µL DNA	95°C 3:00 minutes
5 µL Applied biosystems Power up SYBR Green Master Mix	[95°C 0:45 minutes
3 µL H ₂ O	55°C 0:45 minutes
0.5 µL F primer	72°C 1:30 minutes]
0.5 µL R primer	40 cycles

Table C3.: Common contaminant microbes filtered from amplicon sequence variant (ASV) table. All ASVs with taxonomy contained by the provided terms at the provided taxonomic rank and below were removed from the analysis. Taxonomic rank, the taxonomic level at which filtering was performed for the provided "taxon term"; taxon term, the name of the taxonomic group to be removed from the 16S gene count table

taxonomic rank	Taxon term
class	Thermoleophilia
family	Caldalkalibacillaceae
family	Corynebacteriaceae
family	Micrococcaceae
family	Orbaceae
family	Pirellulaceae
family	Sphingomonadaceae
genus	Bacillus
genus	Methylobacteriaceae
genus	Vulcaniibacterium

Table C4.: Marker Genes used for alignment with DIAMOND. Genes are sorted by metabolism (gene category), include the gene name (Gene Name), the source of the sequences (Gene Source), search terms used to find homologs in available databases, if applicable (Search Terms), the number of sequences used to create the DIAMOND database (num seqs), the length of the homologs used in the database (min length, average length, max length). RPKG GENES, genes used in the Reads per kilobase genome equivalent quantifications; cyanotoxin genes, genes used in the cyanobacterial metagenome assembled genome cyanotoxin analysis; Gene category, class of metabolism or physiology that the nested genes fall within; Gene name, containing gene name and/or abbreviation of the gene used in the analysis; gene source, the repository source for the genes used in the analysis [Greening Lab refers to Chris Greening's metabolic marker gene databases (<https://doi.org/10.26180/14431208.v1>), whilst UniProtKB refers to the UniProt Consortium UniProt Knowledgebase] preference was given to use gene databases from the Greening lab; search terms, these were search terms used to query the UniProt Knowledgebase if the gene was not available from the Greening Lab; number of seqs, refers to the number of sequences used to create the BLAST databases for sequence searches; min length (bp), the minimum length of a representative sequence - databases were amino acid files - so this value represents 3x the number of amino acids in the file; average length(bp), the average length of each protein database in base pairs - representing 3x the number of amino acids in the file - this value converted to kilobase - was used for normalization of count data in the reads per kilobase genome equivalent quantification metric; max length (bp), the maximum length in bp of any sequence in the protein database - with amino acids converted to bp with 3 x amino acid number. MAG geochemical genes, genes used to annotate cyanobacterial metagenome assembled genomes (MAG); Function, function was the geochemical function that was supported by the presence of the nested genes; gene/ pathway, the specific gene or pathway queried; Criteria for inclusion, the criteria used to classify a MAG as possessing the metabolic potential of interest; KO required, the KEGG orthologue that - if present - could serve as sufficient evidence for the metabolism's presence per the criteria for inclusion column; HMM/ Module, the exact hidden markov model or KEGG module of HMM's that were used to query gene presence

Gene Category	Gene Name	Gene Source	Search Terms (if applicable)	number of seqs	min length (bp)	average length (bp)	max length (bp)
Nitrogen Metabolism	Ammonia monooxygenase (<i>amoA</i>)	Greening Lab		66	561	762.6	849
	Copper containing nitrite reductase (<i>nirK</i>)	Greening Lab		338	903	1360.5	2997
	Cytochrome cd1 nitrite reductase (<i>nirS</i>)	Greening Lab		164	1566	1722	2103
	Dissimilatory nitrate reductase (<i>narG</i>)	Greening Lab		277	2853	3687	3882
	Hydrazine synthase (<i>hzsA</i>)	Greening Lab		14	954	2190.3	2463

Table C4 Continued.

Gene Category	Gene Name	Gene Source	Search Terms (if applicable)	number of seqs	min length (bp)	average length (bp)	max length (bp)
Nitrogen Metabolism	Nitric oxide reductase (<i>norB</i>)	Greening Lab		330	1338	2068.2	3327
	Nitrogenase (<i>nifH</i>)	Greening Lab		1271	684	858.9	1026
	Periplasmic nitrate reductase (<i>napA</i>)	Greening Lab		202	2214	2490.6	2844
Sulfur Metabolism	Anaerobic sulfite reductase (<i>asrA</i>)	Greening Lab		84	984	1024.8	1122
	adenosine 5'-phosphosulfate reductase // adenylylsulfate reductase (<i>aprA</i>)	UniProtKB	terms: EC:1.8.99.2 AND (gene_name:"aprA")	35	381	610.5	1998
	Dissimilatory sulfite reductase (<i>dsrA</i>)	Greening Lab		275	1020	1271.1	1458
	Dissimilatory sulfite reductase (oxidative <i>dsrA</i>)	Greening Lab		103	816	872.7	945
	Dissimilatory sulfite reductase (reductive <i>dsrA</i>)	Greening Lab		155	708	848.1	975
	Thiosulfohydrolase (<i>soxB</i>)	Greening Lab		235	1635	1719.3	1914
	Sulfur oxygenase reductase (<i>sor</i>)	Greening Lab		17	915	931.8	999
	Sat sulfate adenylyltransferase // EC 2.7.7.4 (<i>sat</i>)	UniProtKB	terms: "sulfate adenylyltransferase" AND (gene:sat)	6713	69	1161.6	10362
Carbon Fixation	Acetyl CoA synthase (<i>acsB</i>)	Greening Lab		371	981	1906.5	3582
	Anaerobic carbon monoxide dehydrogenase (<i>cooS</i>)	Greening Lab		300	1845	1941.9	2310
	ATP-citrate lyase (<i>acIB</i>)	Greening Lab		68	1107	1274.7	1479
	Crenarchaeotal 4-hydroxybutyryl-CoA synthetase (<i>hbsC</i>)	Greening Lab		12	1326	1415.4	1458
	Thaumarchaeota 4-hydroxybutyryl-CoA synthetase (<i>hbsT</i>)	Greening Lab		10	2094	2102.1	2145

Table C4 Continued.

Gene Category	Gene Name	Gene Source	Search Terms (if applicable)	number of seqs	min length (bp)	average length (bp)	max length (bp)
Carbon Fixation	Rubisco (<i>rbcL</i>)	Greening Lab		707	1092	1437	1911
	Malonyl-CoA reductase (<i>mcr</i>)	Greening Lab		6	3657	3668.4	3687
Methane Metabolism	Methyl CoM reductase (<i>mcrA</i>)	Greening Lab		195	1623	1688.1	2442
	Particulate methane monooxygenase (<i>pmoA</i>)	Greening Lab		85	507	751.2	885
	Soluble methane monooxygenase (<i>mmoA</i>)	Greening Lab		21	1575	1579.8	1584
CO Oxidation	Carbon monoxide dehydrogenase (<i>coxL</i>)	Greening Lab		709	1680	2402.1	3141
Oxygen Metabolism	Cytochrome aa3 oxidase (<i>coxA</i>)	Greening Lab		22832	1305	1712.4	1947
	Cytochrome bd oxidase (<i>cydA</i>)	Greening Lab		36188	1251	1451.4	1878
	Cytochrome bo3 oxidase (<i>cyoA</i>)	Greening Lab		8201	756	958.5	1134
	Cytochrome cbb3 oxidase (<i>ccoN</i>)	Greening Lab		11042	1137	1491.3	1704
Energy Conservation	F-type ATP synthase (<i>atpA</i>)	Greening Lab		26036	1230	1555.5	1845
Hydrogen Acquisition	[Fe]-hydrogenase	Greening Lab		25	1011	1041.3	1074
	[FeFe]-hydrogenase	Greening Lab		1221	441	609.9	828
	[NiFe]-hydrogenase	Greening Lab		2015	993	1514.1	1992
Motility	flagella filament (<i>fliC</i>)	SwissProt	with protein and gene names filtered to only include FliC	24	846	1434.3	1650
	flagella Basal body/hook (<i>fliM</i>)	SwissProt	with protein and gene names filtered to only include FliM	14	933	1004.4	1122
	flagella stator (<i>motA</i>)	SwissProt	with protein and gene names filtered to only include MotA	12	759	810.3	885

Table C4 Continued.

Gene Category	Gene Name	Gene Source	Search Terms (if applicable)	number of seqs	min length (bp)	average length (bp)	max length (bp)
Secretion system	T1SS Secretion protein (<i>hylD</i>)	UniProtKB	terms: HylD AND (protein_name:"Secretion protein HylD")	50	573	1203.3	1473
	T1SS Outer membrane protein (<i>tolC</i>)	UniProtKB	terms: TolC AND (gene:TolC) AND (protein_name:"Outer membrane protein TolC")	150	933	1389.6	1722
	T2SS protein (<i>gspC</i>)	UniProtKB	terms: (gene:GspC) AND (protein_name:"General secretory pathway protein C")	5	462	637.2	690
	T2SS protein (<i>gspD</i>)	UniProtKB	terms: (gene:GspD) AND (protein_name:"General secretory pathway protein D")	21	1587	2311.8	2502
	T3SS secretin (<i>yscC</i>)	UniProtKB	terms: (gene:YscC)	25	1464	2039.7	2763
	T3SS Needle protein (<i>yscF</i>)	UniProtKB	terms: (gene:YscF)	40	246	280.5	1068
	T3SS Needle protein (<i>pscF</i>)	UniProtKB	terms: (gene:PscF)	16	246	305.4	1068
	T4SS protein (<i>virB1</i>)	UniProtKB	terms: VirB1 AND (protein_name:"Type IV secretion system protein virB1")	25	582	725.4	1365
	T4SS protein (<i>virB5</i>)	UniProtKB	terms: VirB5 AND (protein_name:"Type IV secretion system protein virB5")	1789	174	714	3342
	T5SS protein (<i>vacA</i>)	UniProtKB	terms: VacA AND (protein_name:"Type IV secretion system protein virB5")	1789	174	714	3342
	T6SS protein (<i>clpV</i>)	UniProtKB	terms: VacA AND (gene_name:"ClpV")	26	588	2586	2883
	T6SS protein (<i>tssJ</i>)	UniProtKB	terms: TssJ	50	453	540.9	1347
Quorum sensing	acyl-homoserine-lactone synthase genes (<i>luxI</i>)	UniProtKB	terms: LuxI AND (gene:LuxI) AND (protein_name:"Acyl-homoserine-lactone synthase")	33	165	435.6	636
	<i>luxR</i>	UniProtKB	terms: LuxR AND (gene:LuxR) AND (protein_name:"Transcriptional activator protein LuxR")	153	198	701.4	873
	Sensory/regulatory protein (<i>rpfC</i>)	UniProtKB	terms: RpfC AND (gene:RpfC) AND (protein_name:"Sensory/regulatory protein RpfC[...]")	75	234	1243.5	6477

Table C4 Continued.

Gene Category	Gene Name	Gene Source	Search Terms (if applicable)	number of seqs	min length (bp)	average length (bp)	max length (bp)
Light Harvesting	Microbial rhodopsin RHO	Greening Lab		189	594	779.1	1062
	Photosystem I reaction centre (<i>psaA</i>)	Greening Lab		54	1818	2223.3	2595
	Photosystem II reaction centre (<i>psbA</i>)	Greening Lab		121	912	1011.9	1923
CYANOTOXIN GENES							
Gene Category	Gene Name	Gene Source	Search Terms (if applicable)	number of seqs			
Microcystin	microcystin synthetase gene cluster A (<i>mcyA</i>)	UniProtKB	(gene:mcyA) OR (protein_name:"Microcystin synthetase subunit A") AND (protein_name:mcyA)	525			
	microcystin synthetase gene cluster B (<i>mcyB</i>)	UniProtKB	: (gene:mcyB) OR (protein_name:"Microcystin synthetase subunit B") AND (protein_name:mcyB)	655			
	microcystin synthetase gene cluster C (<i>mcyC</i>)	UniProtKB	(gene:mcyC) OR (protein_name:"Microcystin synthetase subunit C") AND (protein_name:mcyC)	184			
	microcystin synthetase gene cluster D (<i>mcyD</i>)	UniProtKB	(gene:mcyD) OR (protein_name:"Microcystin synthetase subunit D") AND (protein_name:mcyD)	162			
	microcystin synthetase gene cluster E (<i>mcyE</i>)	UniProtKB	(gene:mcyE) OR (protein_name:"Microcystin synthetase subunit E") AND (protein_name:mcyE)	634			
	microcystin synthetase gene cluster F (<i>mcyF</i>)	UniProtKB	(gene:mcyF) OR (protein_name:"Microcystin synthetase subunit F") AND (protein_name:mcyF)	22			
	microcystin synthetase gene cluster G (<i>mcyG</i>)	UniProtKB	(gene:mcyG) AND (protein_name:"Microcystin synthetase")	58			
	microcystin synthetase gene cluster H (<i>mcyH</i>)	UniProtKB	(gene:mcyH) OR (protein_name:"Microcystin synthetase subunit H")	65			
	microcystin synthetase gene cluster I (<i>mcyI</i>)	UniProtKB	(gene:mcyI) OR (protein_name:"Microcystin synthetase subunit I")	18			

Table C4 Continued.

Gene Category	Gene Name	Gene Source	Search Terms (if applicable)	number of seqs
Microcystin	microcystin synthetase gene cluster J (<i>mcyJ</i>)	UniProtKB	(gene:mcyJ) AND (protein_name:"Microcystin synthetase")	59
Saxitoxin	saxitoxin gene A (<i>sxtA</i>)	UniProtKB	(gene:sxtA) AND (protein_name:saxitoxin)	149
	saxitoxin gene B (<i>sxtB</i>)	UniProtKB	(gene:sxtB)	12
	saxitoxin gene G (<i>sxtG</i>)	UniProtKB	(gene:sxtG)	45
	saxitoxin gene H (<i>sxtH</i>)	UniProtKB	(gene:sxtH)	20
	saxitoxin gene I (<i>sxtI</i>)	UniProtKB	(gene:sxtI)	15
	saxitoxin gene L (<i>sxtL</i>)	UniProtKB	(gene:sxtL)	7
	saxitoxin gene T (<i>sxtT</i>)	UniProtKB	(gene:sxtT)	8
Anatoxin-A	Anatoxin gene A (<i>anaA</i>)	UniProtKB	(gene:anaA)	8
	Anatoxin gene B (<i>anaB</i>)	UniProtKB	(gene:anaB)	8
	Anatoxin gene C (<i>anaC</i>)	UniProtKB	(gene:anaC)	60
	Anatoxin gene D (<i>anaD</i>)	UniProtKB	(gene:anaD)	8
	Anatoxin gene E (<i>anaE</i>)	UniProtKB	(gene:anaE)	7
	Anatoxin gene F (<i>anaF</i>)	UniProtKB	(gene:anaF)	24
	Anatoxin gene G (<i>anaG</i>)	UniProtKB	(gene:anaG)	5
Cylindrospermopsin	Cylindrospermopsin gene A (<i>cyrA</i>)	UniProtKB	(gene:cyrA) AND (protein_name:Cylindrospermopsin)	18
	Cylindrospermopsin gene B (<i>cyrB</i>)	UniProtKB	(gene:cyrB) AND (protein_name:"cyrB") AND (gene:cyrB) ++ (protein_name:"Cylindrospermopsin")	31
	Cylindrospermopsin gene C (<i>cyrC</i>)	UniProtKB	(gene:cyrC)	16
	Cylindrospermopsin gene D (<i>cyrD</i>)	UniProtKB	(gene:cyrD)	8
	Cylindrospermopsin gene E (<i>cyrE</i>)	UniProtKB	(gene:cyrE)	7
	Cylindrospermopsin gene F (<i>cyrF</i>)	UniProtKB	(gene:cyrF)	11
	Cylindrospermopsin gene G (<i>cyrG</i>)	UniProtKB	(gene:cyrG)	7

Table C4 Continued.

Gene Category	Gene Name	Gene Source	Search Terms (if applicable)	number of seqs
Cylindrospermopsin	Cylindrospermopsin gene H (<i>cyrH</i>)	UniProtKB	(gene:cyrH)	7
	Cylindrospermopsin gene I (<i>cyrI</i>)	UniProtKB	(gene:cyrI) AND (protein_name:Cylindrospermopsin)	22
	Cylindrospermopsin gene J (<i>cyrJ</i>)	UniProtKB	(gene:cyrJ)	79
	Cylindrospermopsin gene K (<i>cyrK</i>)	UniProtKB	(gene:cyrK)	12
	Cylindrospermopsin gene N (<i>cyrN</i>)	UniProtKB	(gene:cyrN) AND (protein_name:"Adenylyl-sulfate kinase")	6
	Cylindrospermopsin gene O (<i>cyrO</i>)	UniProtKB	(gene:cyrO)	11
MAG GEOCHEMICAL GENES				
Function	gene/ pathway	Criteria for inclusion	KO required	HMM/ Module
N ₂ fixation	nitrogenase molybdenum-iron protein alpha chain (<i>nifD</i>)	Presence of KO	K02586	TIGR01282.hmm
	nitrogenase molybdenum-iron protein beta chain (<i>nifK</i>)	Presence of KO	K02591	TIGR01286.hmm
	nitrogenase iron protein (<i>nifH</i>)	Presence of KO	K02588	TIGR01287.hmm
Carbon fixation	Reductive pentose phosphate cycle (Calvin-Benson-Bassham cycle; CBB)	60% of required CBB KO's present	K00855 // (K01601) // K00927 // (K05298 or K00150 or K00134) // (K01623 or K01624) // (K03841 or K02446 or K11532 or K01086) // K00615 // (K01623 or K01624) // (K01100 or K11532 or K01086) // K00615 //(K01807 or K01808)	M00165

Table C5.: Supplementary Table 5: cyanobacterial metagenome assembled genome (MAG) assembly statistics. cyanoID, Cyanobacterial metagenome assembled genome (MAG) unique identifier; binID, identifier used to disambiguate all MAGs generated in this study; Family, the taxonomic rank family assigned to the MAG; Genus, the taxonomic rank genus assigned to the MAG; Percent complete, percent completion of MAG as estimated by checkM, Percent contam, percent contamination of MAG as estimated by checkM; GC content, proportional abundance of guanine and cytosine nucleotides in each MAG; N50, weighted median statistic for MAG quality

cyanoID	binID	Family	Genus	genome size (Mb)	Percent complete	Percent contamination	GC content	N50
MAG1	05_out_wi_poop_bin.3	Xenococcaceae		7.05	99.27	1.1	0.38	36340
MAG2	08_in_wi_poop_bin.1	Xenococcaceae		7.55	98.03	2.22	0.456	19941
MAG3	09_out_wi_poop_bin.12	Xenococcaceae		4.7	84.27	2.33	0.376	12929
MAG4	06_in_wi_poop_bin.22	Xenococcaceae	Xenococcus	7.97	83.09	9.94	0.378	6115
MAG5	05_out_wi_sediment_bin.10	Spirulinaceae	Spirulina sp. (CCY15215)	6.53	91.24	1.7	0.459	11053
MAG6	05_out_wi_poop_bin.33	Prochloraceae		4.94	81.91	3.78	0.424	9465
MAG7	coassembly_bin.63	Phormidesmiaeae		5.04	87.4	3.26	0.509	12075
MAG8	09_out_wi_poop_bin.22	Synechococcaceae (PCC-7336)	Synechococcus	4.8	95.61	2.63	0.513	44029
MAG9	05_out_wi_poop_bin.19	Synechococcaceae (PCC-7336)	Synechococcus	4.9	100	3.95	0.512	57437
MAG10	03_out_wi_poop_bin.7	Microcystaceae_A	Crocospheera	5.37	82.38	7.75	0.377	8292
MAG11	05_out_wi_poop_bin.2	Microcystaceae_A	Crocospheera	4.78	94.39	3.47	0.388	31509
MAG12	02_in_wo_sediment_bin.3	Microcystaceae_A		4.41	83.39	1.94	0.39	8253
MAG13	coassembly_bin.85	Microcystaceae_A		4.62	89.47	7.03	0.386	36321
MAG14	06_in_wi_poop_bin.3	Microcystaceae_A	Crocospheera	4.89	85.37	1.79	0.374	32265

Table C5 Continued.

cyanoID	binID	Family	Genus	genome size (Mb)	Percent complete	Percent contamination	GC content	N50
MAG15	02_in_wi_poop_bin.16	Microcystaceae_A	Crocospaera	4.68	89.26	2.15	0.364	10047
MAG16	03_out_wi_sediment_bin.9	Microcystaceae_A	Crocospaera	4.57	83.58	2.15	0.369	12809
MAG17	02_in_wi_sediment_bin.5	Microcoleaceae	Trichodesmium	7.01	92.57	4.04	0.35	10137
MAG18	05_out_wi_sediment_bin.26	Desertifilaceae	Hormosilla	6.47	95.16	0.82	0.481	43680