UNIVERSITY OF CALIFORNIA SAN DIEGO

Phage and bacterial ecology in marine holobiont disease and competition

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Biology

by

Mark Little

Committee in charge:

San Diego State University

Professor Forest Rohwer, Chair
Professor David Lipson
Professor Nicholas Shikuma

University of California San Diego

Professor James Golden
Professor Jennifer Smith

2022

The Dissertation of Mark Little is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

San Diego State University

2022

DEDICATION

To my late mother, Lisa Little, and my late grandfathers, Lyle E. Belter and Louis J. Palmer.

EPIGRAPH

"But at the end of the day, the trees all get wheeled away"

Conor Oberst of Bright Eyes (from the "Noise Floor" LP)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AUC: area under the curve

DOC: dissolved organic carbon

PCR: polymerase chain reaction

KtW: Kill-the-winner model

MTE: metabolic theory of ecology

NOSC: nominal oxidation state of carbon

Phage: a virus that infects Bacteria or Archaea

Prophage: a phage that has inserted its genome into the bacterial genome (lysogen)

PtW: Piggyback-the-winner model

Stx: Shiga toxin

VLP: viral-like particle

VMR: virus-to-microbe ratio

ZOT: *Zonula occludens* Toxin

ACKNOWLEDGEMENTS

Thank you to my mentors Dr. Forest Rohwer and Dr. Linda Wegley Kelly. Thank you to the funding agencies and donor that supported me during my PhD: the Gordon and Betty Moore foundation and the National Science Foundation.

Chapter 1, in full, is published in Marine Disease Ecology (First Edit.), 2020. Oxford Press. Mark Little*, Maria I. Rojas*, Forest Rohwer. The dissertation author was the co-primary investigator and co-first author of this paper.

Chapter 2, in full, is published in the Proceedings of the National Academy of Sciences (PNAS), 2020. Ty N.F. Roach*, Mark Little, Milou G.I. Arts, Joel Huckeba, Andreas F. Haas, Emma E. George, Robert A. Quinn, Ana G. Cobian-Guemes, Douglas S. Naliboff, Cynthia B. Silveira, Mark J.A. Vermeij, Linda Wegley Kelly, Pieter C. Dorrestein, and Forest Rohwer. The dissertation author was the co-primary investigator and co-first author of this paper.

Chapter 3, in full, is published in Frontiers in Marine Science, 2021. Mark Little, Emma E. George, Milou G.I. Arts, Jade Shivak, Sean Benler, Joel Huckeba, Zachary A. Quinlan, Vittorio Boscaro, Benjamin Mueller, Ana G. Cobian-Guemes, Maria Isabel Rojas, Brandie White, Daniel Petras, Cynthia B. Silveira, Andreas F. Haas, Linda Wegley Kelly, Mark J.A. Vermeij, Robert A. Quinn, Patrick J. Keeling, Pieter C. Dorrestein, Forest Rohwer, and Ty N.F. Roach. The dissertation author was the co-primary investigator and co-first author of this paper.

VITA

2022 Doctor of Philosophy in Biology, University of California San Diego and San Diego State University

2013 Bachelors of Science in Biology with an emphasis in Marine Biology, San Diego State University

PUBLICATIONS

A. Gamba, D. Petras, **M. Little**, B. White, P.C. Dorrestein, F. Rohwer, R.A. Foster, A.C. Hartmann. Lipid saturation kinetics of host mitochondria and symbiotic algae in corals under high temperature stress. Frontiers in Marine Science. (2022)

M. Rojas, S. S. Giles, **M. Little**., R. Baron, I. Livingston, T. R. T. Dagenais, J. Baer., A. G. Cobián-Güemes, White, B., F. Rohwer. Swabbing the Urban Environment - A Pipeline for Sampling and Detection of SARS-CoV-2 From Environmental Reservoirs. J. Vis. Exp. doi:10.3791/62379 (2021).

M. Little*, E.E. George*, M.G.I. Arts, J. Shivak, S. Benler, J. Huckeba, Z.A. Quinlan, V. Boscaro, B. Mueller, A.G.C. Güemes, M.I. Rojas, B. White, D. Petras, C.B. Silveira, A.F. Haas, L.W. Kelly, M.J.A. Vermeij, R.A. Quinn, P.J. Keeling, P.C. Dorrestein, F. Rohwer and T.N.F. Roach. 2021. Three-Dimensional Molecular Cartography of the Caribbean Reef-Building Coral Orbicella faveolata. Front. Mar. Sci. 8:627724. doi: 10.3389/fmars.2021.627724

M. Little*, M.I. Rojas*, and F. Rohwer. 2020. Bacteriophage can drive virulence in marine pathogens. In Behringer (Ed.), Marine Disease Ecology (First Edit.). Oxford Press. ISBN: 9780198821649

T.N.F. Roach*, **M. Little***, M.G.I. Arts, J. Huckeba, A.F. Haas, E.E. George, R.A. Quinn, A.G. Cobián-Güemes, D.S. Naliboff, C.B Silveira, M.J.A. Vermeij, L. Wegley Kelly, P.C. Dorrestein, F. Rohwer, A multi-omic analysis of in situ coral-turf algal interactions. Proc. Natl. Acad. Sci. U. S. A. (2020).

C. Ghose, M. Ly, L.K. Schwanemann, J.H. Shin, K. Atab, J.J. Barr, **M. Little**, R.T. Schooley, J. Chopyk, D.T. Pride, The Virome of Cerebrospinal Fluid: Viruses Where We Once Thought There Were None. Front. Microbiol. 10, 2061 (2019).

I. Galtier d'Auriac, R.A. Quinn, H. Maughan, L.-F. Nothias, **M. Little**, C.A. Kapono, A. Cobian, B.T. Reyes, K. Green, S.D. Quistad, M. Leray, J.E. Smith, P.C. Dorrestein, F. Rohwer, D.D. Deheyn, A.C. Hartmann, Before platelets: the production of platelet-activating factor during growth and stress in a basal marine organism. Proc. Biol. Sci. 285 (2018).

J.E. Zlamal, T.K. Raab, **M. Little**, R.A. Edwards, D.A. Lipson, Biological chlorine cycling in the Arctic Coastal Plain. Biogeochemistry (2017).

R.A. Quinn, M.J.A. Vermeij, A.C. Hartmann, I. Galtier d'Auriac, S. Benler, A. Haas, S.D. Quistad, Y.W. Lim, **M. Little**, S. Sandin, J.E. Smith, P.C. Dorrestein, F. Rohwer, Metabolomics of reef benthic interactions reveals a bioactive lipid involved in coral defence. Proc. Biol. Sci. 283, doi:10.1098/rspb.2016.0469. (2016).

B. Knowles*, C.B. Silveira*, B. A. Bailey, K. Barott, V.A. Cantu, A.G. Cobián-Güemes, F.H. Coutinho, E.A. Dinsdale, B. Felts, K.A. Furby, E.E. George, K.T. Green, G.B. Gregoracci, A.F. Haas, J.M. Haggerty, E.R. Hester, N. Hisakawa, L.W. Kelly, Y.W. Lim, **M. Little**, A. Luque, T. McDole-Somera, K. McNair, L.S. de Oliveira, S.D. Quistad, N.L. Robinett, E. Sala, P. Salamon, S.E. Sanchez, S. Sandin, G.G.Z. Silva, J. Smith, C. Sullivan, C. Thompson, M.J.A. Vermeij, M. Youle, C. Young, B. Zgliczynski, R. Brainard, R.A. Edwards, J. Nulton, F. Thompson, F. Rohwer, Lytic to temperate switching of viral communities. Nature. 539, 123 (2016).

J.A. Grasis, T. Lachnit, F. Anton-Erxleben, Y.W. Lim, R. Schmieder, S. Fraune, S. Franzenburg, S. Insua, G. Machado, M. Haynes, **M. Little**, R. Kimble, P. Rosenstiel, F.L. Rohwer, T.C.G. Bosch, Species-specific viromes in the ancestral holobiont Hydra. PLoS One. 9, e109952 (2014).

ABSTRACT OF THE DISSERTATION

Phage and bacterial ecology in marine holobiont disease and competition

by

Mark Little

Doctor of Philosophy in Biology

University of California San Diego, 2022
San Diego State University, 2022

Professor Forest Rohwer, Chair

This dissertation work contains the following chapters: Chapter 1 revealed that bacterial pathogens of marine animals contain a higher abundance of prophages in their genomes than non-disease-causing host-associated bacteria. The work highlights the importance of horizontal gene transfer between bacteria and phages in the context of marine fauna, and reviews what is

known about prophages in the context of marine diseases. Additionally, a comparison of the prophage-encoding genes between pathogens and non-pathogens, found that predicted prophage-encoded genes in disease-causing bacteria are enriched for carbohydrate and nitrogen metabolism, virulence factors, and membrane transport. Chapter 2 consisted of utilizing multi-omics and microscopy to characterize coral-turf algal interactions *in situ* that led to a working model of the ecology of these benthic holobiont interactions. I applied metagenomic sequencing in parallel with metabolomics to uncover the underlying bacterial, viral, and biochemical processes associated with coral-turf algal competition and reef decline. This work emphasizes the important role of host-associated bacteria and viruses in the ecological outcome of competing coral-algal interactions. In this context, the study fits into what was coined the "Algal Feeding Hypothesis", which posits that changes in coral-algal interface communities are driven by bacteria that feed on algal-derived compounds. Chapter 3 focused on the molecular and microbial cartography of massive coral colonies using a multi-omics approach to understand the natural history of endangered massive coral colonies competing with benthic algae. In this work I combined 3D photogrammetry with metatranscriptomics, metabolomics, and metabarcoding to investigate all portions of the massive coral *Orbicella faveolata* visually and spatially *in situ.*

Chapter 1: Bacteriophage can drive virulence in marine pathogens

Introduction

Marine ecosystems around the globe are in dramatic decline caused by anthropogenic impacts such as pollution, overfishing, climate change, and increasingly prevalent diseases in ecologically important macro-organisms (Bateman et al., 2022; Bojko et al., 2022; Burge and Hershberger 2022). Many marine disease ecologists are reconsidering the strategies used to understand and investigate the etiology of these diseases. Recent studies incorporate the notion that an array of stressors can disrupt natural holobiont communities, leading to a variety of detrimental ecological outcomes often potentiated by microbial pathogenesis (Egan and Gardiner 2016; Morrow et al., 2022; Bateman et al., 2022; Morton et al., 2022). While microbial diseases can be caused by a variety of organisms, most bacteria involved in pathogenicity in marine environments contain horizontally acquired elements that are largely overlooked and play essential ecological and evolutionary roles. These elements are often carried by phage genomes integrated in the genome of the bacterial host. The expression of these prophage-encoded genes can confer pathogenicity and dysbiosis, the latter defined by an unbalanced composition of the host-associated microbial community.

Bacteriophages, or simply “phages,” are viruses that infect bacteria and have the unique ability to undergo one of two lifestyles, lytic or lysogenic. In the lytic cycle, upon infection, the phage uses the machinery of the bacterial host to replicate, synthesize new viral particles, and release its progeny, often killing the bacterial host by cell lysis (Echols 1972). During the lysogenic cycle, the phage genome is integrated into the bacterial genome and its replication occurs only as part of the normal cell cycle (reviewed in Young 1992). Interestingly, different

environmental and cellular cues can trigger the switch from one cycle to the other (Wommack and Colwell 2000).

Through infection, phages have the capability to move host genes between bacteria. When fragments of the host chromosome are packaged within the viral particles, bacterial DNA is shared through infection to a recipient bacterial cell. This horizontal transfer is termed transduction, and it can be the generalized transfer of a DNA sequence from a random position in the bacterial genome, or the specialized transfer of DNA from a specific location in the bacterial chromosome. In either case, the amount of packaged DNA is limited by the size of the viral capsid and hence of the original viral genome (Cui et al., 2014). In addition to phages, other mechanisms of horizontal gene transfer between bacteria occur via the transfer of transposons and plasmids. Transposons are DNA sequences that can jump from one location in the bacterial genome to another, generating gene duplications or truncating genes when their insertion interferes with the coding sequence of the gene. Plasmids, on the other hand, are small DNA molecules, often circular and double-stranded, that can replicate independently of the bacterial genome and can be transferred between bacterial hosts through conjugation.

Phages, generalized transducing agents, transposable elements, and plasmids are some of the major drivers of microbial evolutionary processes and therefore likely play a key role in microbial pathogenicity and dysbiosis. This chapter focuses on the role of phage-encoded elements in the context of the etiology of economically and ecologically relevant marine pathogenesis and dysbiosis. This section also provides a meta-analysis of all known, fully sequenced, marine bacterial host-associated pathogenic and non-pathogenic genomes that serve as a baseline for understanding how the ecology of horizontal gene-transfer carried out by phage contributes to the evolution of marine pathogens.

Transduction in the marine environment occurs at high rates and has been suggested to have the minimum capacity to move 10^{24} genes from viruses to host per year globally (Rohwer and Thurber 2009). Within the past 5 years, the influence of horizontally acquired genetic elements from viruses has gained traction in the field of marine disease ecology. Recent analyses of the genomes of multiple pathogenic Vibrio strains revealed prophage-encoded elements that contribute to the pathogenicity of the bacteria (Weynberg et al., 2015, Figure 1.4). These tripartite eukaryote–microbe–phage interactions likely determine many marine disease mechanisms. When a bacterium incorporates in its genome a viral genome, or acquires a prophage through infection, it is called a lysogen. Viral replication and survival occur through bacterial cell division, thus producing more lysogens as progeny. Phages that can initiate their incorporation into the chromosome of the host are known as temperate phages. It is well known that temperate phages have considerable gene repertoires that may enhance bacterial host fitness, and since phages are the most abundant biological entities on the planet, with an estimated 4.80×10^{31} phages on Earth, it is reasonable to predict that these viruses strongly influence the unfolding of marine pathogenesis across a variety of organisms and ecosystems (Cobián Güemes et al., 2016).

While prophages encompass around 25 percent of phages in the global phage gene pool, only forty-one prophage-mediated phenotypes have been observed or experimentally demonstrated (Bondy-Denomy and Davidson 2014; Casjens 2005). Prophages or temperate phages can enhance the fitness of their bacterial hosts in a variety of ways (Figure 1.1A-E), such as (1) conferring metabolic capacities through the acquisition of photosynthetic genes in Cyanobacteria (Rohwer and Thurber 2009), (2) encoding functional proteins such as anti-CRISPR systems in Pseudomonas aeruginosa which allow the bacteria to outcompete other

bacteria (Bondy-Denomy et al., 2014), and (3) exclusion factors like the Imm protein of the famous phage T4 in Escherichia coli that prevents other phages from infecting the lysogen (Lu and Henning 1994; Obeng et al., 2016). Horizontally acquired mutualistic viruses therefore allow lysogens to broaden their ecological niche space (Figure 1.1A-E). In many instances, the prophage can encode exotoxins that directly affect the host (Figure 1.1B). In addition, some prophage-encoded proteins have been shown to inhibit predation from bacterivorous protists (Figure 1.1D). We are only starting to shed light on the functional roles of integrated phage, but the fact that most bacterial genomes harbor about one to two prophages (Casjens 2003) indicates that these are significant players in a plethora of ecological dynamics. An assessment of temperateness, or the ability to initiate lysogeny or a lysogenic conversion of viruses in seawater, revealed that within phage communities, 80 percent of the members contain the potential for a temperate lifestyle (Breitbart et al., 2004). Clearly, there are more functions to be discovered considering the high prevalence of lysogeny, where most bacterial genomes harbor multiple prophage and at a maximum have been observed to comprise 20 percent of bacterial genomic sequence space (Canchaya et al., 2003; Casjens 2005).

The role of prophages in disease

In 1951, the first report of phage-mediated virulence was described in the bacterium *Corynebacterium diphtheriae*, the disease-causing agent of diphtheria. When non-virulent strains of *C. diphtheriae* were challenged with phages, the next generation of bacterial progeny presented a virulent phenotype (Freeman 1951). Since this first discovery of prophage-mediated bacterial fitness enhancements, it has been revealed that a large portion of strain-to-strain differences are due to phage-mediated horizontal gene transfer (Lawrence 2002).

Bacterial strains exhibiting pathogenicity have been shown to contain a higher proportion of phage genes compared to non-pathogenic strains, and currently twelve prophages encoding virulence genes have been discovered among seven relevant bacterial pathogens including *C. diptheriae*, *E. coli*, *S. enterica*, *P. aeruginosa*, *S. mitis*, *C. jejuni*, and *V. cholerae* (Busby et al., 2013; Davies et al., 2016). Temperate phages can produce a variety of exotoxins such as cholera, Shiga toxin, and botulism, and these types of prophage-mediated functions are extremely relevant in the case of marine disease pathogenesis. Since the first phage-mediated phenotype was observed in diphtheria, the *E. coli* prophage system has been studied extensively. This *E. coli* prophage encodes Shiga toxin (Stx) whose production is independent of phage lytic activity. Conversely, in the case of *C. diphtheriae*, the production and secretion of the toxin does not require lysis of the lysogen (Holmes 2000).

Some bacterial toxins, many of which contribute to pathogenicity, likely evolved to evade predation from other microorganisms, such as protists (Figure 1.1D). An example of this survival strategy is the aforementioned Shiga toxin, which confers *E. coli* anti-predatorial defense against the bacteriovore *Tetrahymena thermophila* (Lainhart et al., 2009). In the marine environment, a study on *Serratia marcescens* challenged a population of this bacterium against two bacterivorous predators with different feeding mechanism—*Acanthamoeba castellanii*, a surface feeder, and *Tetrahymena thermophila*, a particle feeder—and observed that the *S. marcescens* population became more resistant to the infection by lytic phages, presumably due to the acquisition of a prophage in their genome (Örmälä-Odegrip et al., 2015). These findings may suggest that predation pressure by bacterivores selects for bacteria carrying prophage in their genomes. These prophages potentially encode for proteins that either are directly toxic to bacteriovores or indirectly deter them from predating on the bacterial host. These phenomena are

relevant when considering the ability of pathogens to evade protist predation as well as infection by lytic phages in the marine environment.

Prophages in marine diseases

Understanding the distribution and role of temperate phages in marine bacterial pathogens is of high relevance considering their ability to exhibit superinfection exclusion to other phages targeting the same host range. This is important to consider because it allows for pathogenic lysogens to prevent lytic control (i.e., bacterial death via cell lysis) by other phages naturally present or introduced to the community as phage therapeutics. These dynamics are likely at large in instances of bacterially mediated marine disease and pathogenesis. To test this hypothesis, we searched for predicted temperate phages in publicly available marine bacterial pathogen and non-pathogen genomes from host-associated marine environments on a global scale. This meta-analysis utilized bioinformatic tools to demonstrate significantly higher proportion of prophages in pathogenic than in non-pathogenic marine host-associated bacteria (Figure 1.2).

The results mentioned above were not unexpected since virus-like particles have been identified as playing a role in ecologically relevant marine diseases. For example, putative phage hyperparasites have been described to be associated with the bacterium Candidatus *Xenohaliotis californiensis* (WS-RLO), which causes withering syndrome in abalone (Friedman and Crosson 2012). These virus-like particles have similar morphology to that of the Siphoviridae family of phages, although their genomes have not been sequenced (Cruz-Flores et al., 2016). While the function of these intracellular viruses in pathogenesis is not yet understood, due to their 50 nm size and the pleomorphic traits conferred to bacteria, it is possible that they constitute

generalized transducing agents. Furthermore, this system may be an example of how lytic viruses play a role in bacterial pathogenesis.

Evolutionary implications of prophages in marine diseases

Temperate phages are present in 40–50 percent of all known microbial genomes and in twenty-one of thirty known bacterial phyla (Canchaya et al., 2003; Touchon et al., 2016). In addition, lysogen abundance is more prevalent in pathogens and negatively correlated with spacer acquisition across CRISPR-Cas systems in respective host microbial genomes (Touchon et al., 2016). The contribution of phages in structuring microbial communities through predation and lysogeny is apparent, but the need for understanding functional roles in microbial disease ecology is often overlooked. Researchers have only begun to understand the roles of lysogens in nature in either pathogenic or ecological contexts. In this chapter we seek to shed light on what is known about phage-encoded function in bacterial pathogenesis and dysbiosis in marine disease.

While bacteriovore anti-predation conferred by horizontal gene transfer or phage infection has yet to be assessed in marine host-associated pathogenic bacteria, it is likely to explain the selection of systems that can lead to microbial disease in macro-organismal hosts. This process would require receptor-mediated endocytosis (RME), via cell-surface receptors of the eukaryotic host, where the exo-toxins against bacteriovore predation have a secondary effect on the macro-organismal host. These dynamics should be considered as marine disease ecologists investigate and attempt to address problems and develop or apply therapies in aquaculture and the environment. Here, prophages, the microbiological Trojan Horse, are clearly an important overlooked component of disease etiologies. Opportunely, genomics and

bioinformatics advance rapidly and allow for the evaluation of the contribution of prophage-encoded traits to the global pool of marine diseases (Frasca et al., 2022).

The events associated with the acquisition of a potentially mutualistic phage require lytic-to-temperate switching during an active infection event and this is defined as lysogeny. Lysogeny can introduce novel phenotypes to the bacterial host via gene expression (Brüssow et al., 2004). Recent metagenomic-based studies on lytic-to-temperate switching dynamics in marine ecosystems demonstrate that at high microbial abundances and within certain environments, there is a higher presence of integrase and excisionase genes, suggesting an ecological lytic-to-temperate switch in these viral communities (Knowles et al., 2016). Integrases are phage-encoded proteins that enable viruses to incorporate their genome into bacterial chromosomes, while excisionases allow prophages to exit the chromosome during induction to the lytic cycle. Therefore, the relative level of integrase and excisionase genes in viral communities suggests the functional potential for lysogeny (Knowles et al., 2016). Mathematical modeling further demonstrates that temperate phage lifestyles can be more prevalent under environmental conditions that favor bacterial growth (Maslov and Sneppen 2015). Although marine bacterial pathogeneses are currently increasing at unprecedented rates, these are relatively isolated phenomena when considering the spatial scale at which organisms exist in any given ecosystem.

The life history of temperate phages can shed light on the understanding of how prophages are distributed across bacterial phyla. The abundance of bacterial mechanisms that protect against phage infection, such as CRISPR-Cas systems, has been negatively correlated with the presence of prophages in a global dataset of bacterial genomes (Touchon et al., 2016). In this study, the bacteria with minimal doubling time, meaning those with genomes with the

fastest growing capability, were strongly correlated with the occurrence of lysogeny (Touchon et al., 2016). This finding adds further support to the observed increases of lysogeny in certain environments, as described in the Piggyback-the-Winner (PtW) model, where lysogen occurrence is correlated with higher microbial abundance (Knowles et al., 2016). In contrast, an alternative scenario can take place when virion numbers increase with bacterial abundance via lytic activity, and these dynamics are explained by the "Kill-the-Winner" (KtW) model (Thingstad 2000). In either case, bet-hedging strategies of allocating bacterial resources to viral lytic production and others to lysogeny greatly influence ecological dynamics where phage impact bacterial growth and function (Morton et al., 2022).

Meta-analysis of prophage-encoded functions

Prophage-encoded functions have yet to be assessed in pathogenic and dysbiotic host-associated system dynamics. This meta-analysis covers a total of eighty-nine publicly available complete marine genomes of host-associated bacteria, with sequences coming from thirty-two non-pathogenic and fifty-seven pathogenic isolates found in algae, invertebrates, and vertebrates (Klemetsen et al., 2018). The reference database used is the most extensive marine bacterial genome resource to date and includes comprehensive metadata on the source of the isolates, organismal pathology at the species level, bacterial genome length, data on encoded proteins per genome, and many other useful metadata (Klemetsen et al., 2018). The results of our analysis revealed a significantly higher abundance of prophage in marine pathogens when compared to non-pathogens (Figure 1.2).

A comparison of the number of hits to prophage-encoded genes between pathogenic and non-pathogenic host-associated marine bacteria revealed a significantly higher level of

prophages per genome across two different normalization methods (Figure 1.2). The log abundance of prophage hits was divided by either the respective bacterial genome length (Figure 1.2A) or the predicted number of bacterial-encoded proteins in the respective genomes (Figure 1.2B). Both normalization methods revealed a significantly higher number of prophage genetic elements in the pathogenic host-associated bacteria (Figure 1.2).

To understand the functional relevance of these predicted prophage-encoded elements, we analyzed the results with a computer program (PhiSpy) designed to find prophages in bacterial genomes. The predicted prophage-encoded gene hits were annotated against the SEED project subsystems database, which integrates the data generated by the Rapid Annotation of microbial genomes using Subsystems Technology (RAST), to gain insight into what known functions may be prophage-encoded in these ecologically relevant marine pathogens. The SEED database uses a system that organizes families of functional genes into categories. To determine which functions were positively enriched in prophage-encoded genes carried by pathogenic bacteria, the relative abundance of subsystems was averaged, and the averaged abundance of non-pathogenic bacteria was subtracted from that of pathogenic bacteria (Figure 1.3). The three highest positively enriched subsystems in pathogens were carbohydrate utilization, membrane transport, and virulence. The observation of carbohydrate utilization and membrane transport suggests that prophages within pathogenic marine bacteria encode other functions, beyond virulence factors, that are ecologically interesting (e.g., fitness enhancement), and that have a direct or indirect effect on host disease. Interestingly, ~ 50 percent of the prophage hits did not match to any known function in the SEED subsystems database, which confirms that there is still much more to learn about these ecologically relevant prophages and prophage-encoded elements. The results suggest PtW dynamics, where the functional genes acquired via lysogeny would

enable these pathogenic strains to outcompete the natural holobiont bacterial community members and ultimately lead to diseased states.

The tripartite eukaryote–microbe–phage dynamics discussed in this chapter are essential for marine biologists to consider when trying to understand the etiology and ecology of marine infectious diseases. While this chapter considers all fully sequenced host-associated marine bacterial pathogens at a more general level across global scales, it offers a baseline for future targeted studies on marine bacterial pathogenesis and prophage-encoded phenotypes. Future approaches include determining the specific protein or function that these genes encode to predict their role in fitness, pathogenicity, metabolism, competition, and microbial as well as ecological dynamics. This knowledge could inform researchers and clinicians when designing strategies to enrich desired bacterial traits, for instance to reverse dysbiosis in different ecosystems or directly beat pathogenesis.

Case study: The role of prophage in a coral pathogen

Vibrio coralliilyticus is a bacterial pathogen implicated in coral and oyster disease. All five documented strains of *V. coralliilyticus* harbor at least one or multiple prophages and are implicated in diseases of the corals *Pocillopora damicornis*, *Montipora aequituberculata*, and *Acropora cytherea*, and of the Pacific oyster *Crassotrea gigas*. These bacterial strains are associated with different pathogenic phenotypes across coral species, such as bleaching and white syndrome (WS) (Kimes et al., 2012; Ushijima et al., 2014). Traditionally, 16S rDNA is used as taxonomic marker gene for bacterial species-level identification; however, the field of metagenomics and whole-genome sequencing is expanding the ability to analyze these genomes and understand the functional capacity of horizontal gene transfer between prophages and

bacteria. Bioinformatic analyses show that *V. coralliilyticus* can carry integrated in its genome a complete prophage that codes for a *Zona occludens* toxin (ZOT) (Rohwer and Thurber 2009; Weynberg et al., 2015). This exotoxin gene has striking homology to that of *Vibrio cholerae* prophage-encoded cholera toxin (CTX) gene. *Vibrio cholerae* infection in humans causes cholera, an acute diarrheal illness that results from the incorporation and expression of the prophage CTXφ in the bacterial genome (Waldor and Makalanos 1996). This homology suggests that ZOT could play a role in coral and oyster disease through a mechanism similar to that of CTX (Figure 1.4).

Summary

1. Phages, the viruses that infect bacteria, can confer functions to the bacterial host that contribute to pathogenicity and dysbiosis through the lysogenic lifestyle.
2. Bacterial pathogens in the marine environment contain higher abundances of prophages in their genome than non-pathogenic bacteria.
3. A comparison of prophage genetic content between marine pathogens and non-pathogens revealed that pathogen associated prophages are enriched in genes encoding for carbohydrate metabolism, membrane transport, nitrogen metabolism, virulence, and others.
4. Horizontally acquired prophage-encoded DNA regions may play a large role in the ecology and evolution of marine diseases, due to the functions they confer.
5. Future studies on the bacteria that are associated with diseases should examine the regions of prophages in the genome for insights into the etiology and ecology of pathogenicity.

6. These approaches are relevant to non-genome associated datasets (e.g., metagenomic data), and much is to be discovered about the ecology and evolution of prophages in diseases as these types of analyses become better explored and implemented

Figures

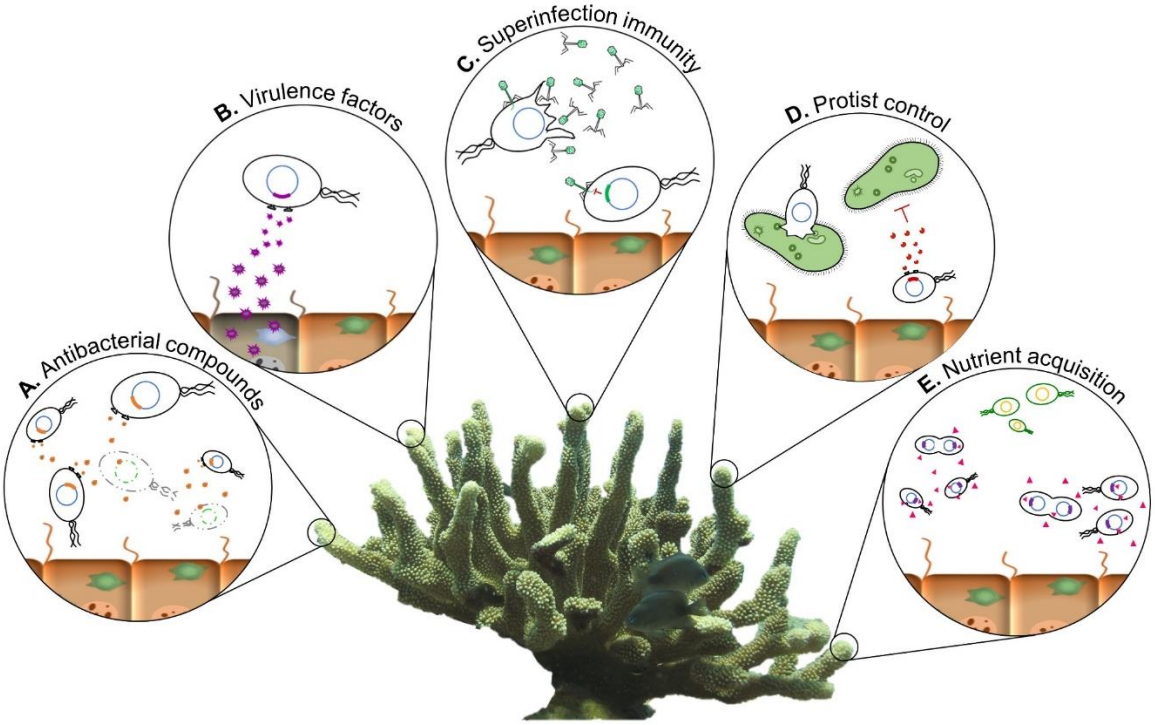


Figure 1.1 Conceptual overview of genes encoded by temperate phages that enhance lysogen fitness in ecologically relevant marine holobionts. (A) Prophage-encoded proteins that inhibit competing bacteria and enable expansion of the lysogen's niche space. (B) Prophage-encoded virulence factors that directly affect the host. (C) Superinfection immunity prevents lytic control of lysogens by lytic phages. On mucosal metazoan epithelium the progeny of spontaneous prophage inductions can cause lytic infection in competing commensal or pathogenic non-lysogens (Barr et al 2013; Silveira and Rohwer 2016). (D) Prophage-encoded proteins inhibit predation by unicellular protists that could have a negative secondary effect on the multicellular host. (E) Prophage-encoded genes that allow bacteria to expand their metabolic repertoire and niche.

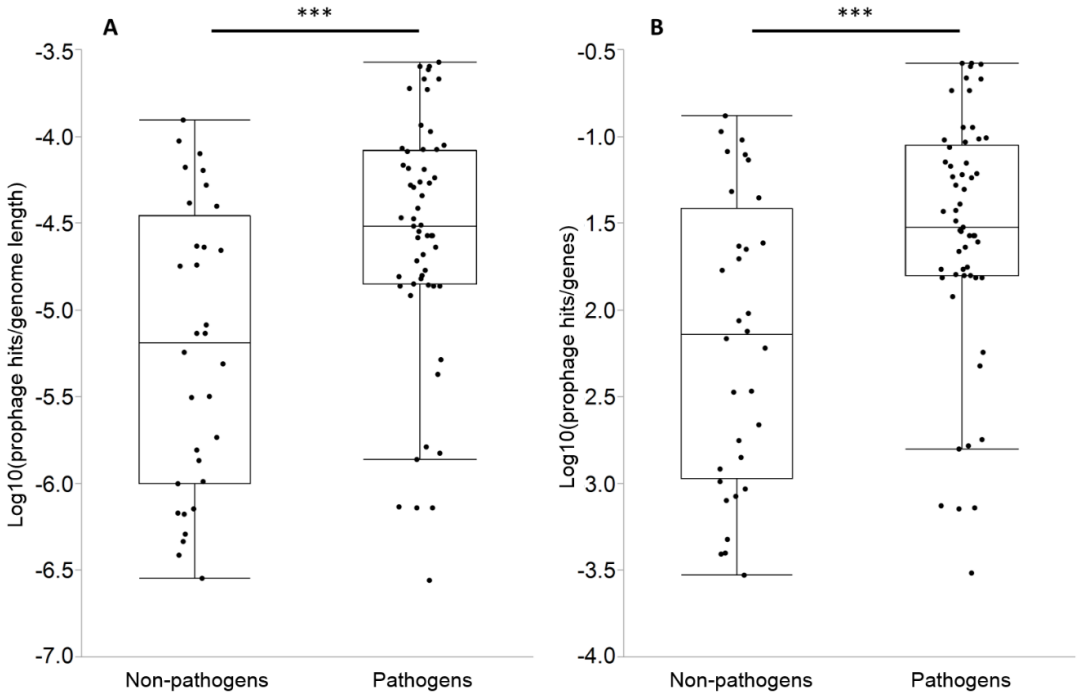


Figure 1.2 Prophage gene abundance in publicly available host-associated bacterial genomes in marine ecosystems. (A) Comparison of number of hits to predicted prophage-encoded genes in marine bacterial genomes normalized by the respective genome length for host-associated non-pathogens and pathogens. (B) Comparison of number of hits to predicted prophage-encoded genes in marine bacterial genomes normalized by total number of bacterial genes in the respective non-pathogen or pathogen classification. Data were accessed from public databases in March 2018. *** $p \leq 0.001$; two-sample Wilcoxon test.

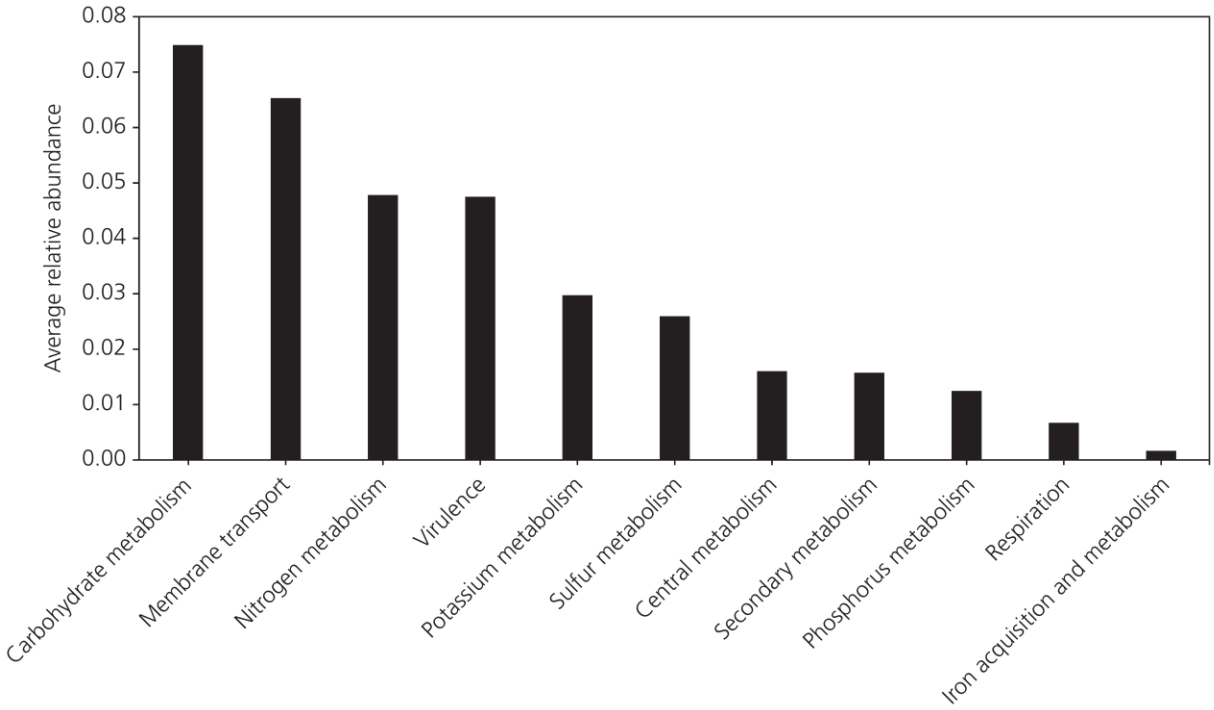


Figure 1.3 Subsystems level 1 functional profiles of prophage-encoded functions in publicly available host-associated bacterial genomes between non-pathogens and pathogens. Average abundance of subsystem level 1 genes of pathogens relative to the corresponding genes in non-pathogenic bacteria. Data were accessed from public databases in March 2018.

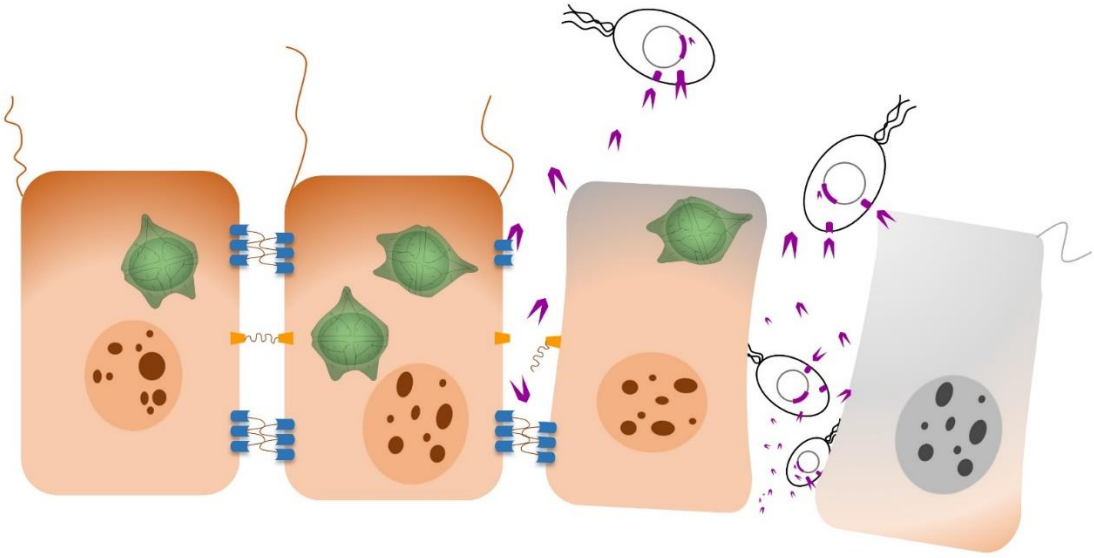


Figure 1.4 Prophage-encoded ZOT toxin of *Vibrio coralliilyticus* disrupting intercellular occluding junctions (tight junctions) that maintain integrity of epithelium during coral infection and disease.

Acknowledgements

Chapter 1, in full, is published in Marine Disease Ecology (First Edit.), 2020. Oxford Press. Mark Little*, Maria I. Rojas*, Forest Rohwer. The dissertation author was the co-primary investigator and co-first author of this paper.

I would like to thank Taylor O'Connell for developing the bioinformatic pipeline used to assess the genomes of interest for this chapter. This research was sponsored by the GBMF Investigator Award 3781 (to Forest Rohwer).

References

Barr, J. J., Auro, R., Furlan, M., Whiteson, K. L., Erb, M. L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A. S., Doran, K. S., Salamon, P., Youle, M., & Rohwer, F. (2013). Bacteriophage adhering to mucus provide a non–host-derived immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(26), 10771–10776.
Breitbart, M., Felts, B., Kelley, S., Mahaffy, J. M., Nulton, J., Salamon, P., and Rohwer, F. L. 2004. Diversity and population structure of a near-shore marine-sediment viral community. Proc Biol Sci 271, 565-574, doi:10.1098/rspb.2003.2628.

Brüssow, H., Canchaya, C., & Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews: MMBR*, *68*(3), 560–602.

Busby, B., Kristensen, D. M., & Koonin, E. V. (2013). Contribution of phage-derived genomic islands to the virulence of facultative bacterial pathogens. *Environmental Microbiology*, *15*(2), 307–312.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., & Brüssow, H. (2003). Prophage genomics. *Microbiology and Molecular Biology Reviews: MMBR*, *67*(2), 238–276.

Casjens, S. (2003). Prophages and bacterial genomics: what have we learned so far? *Molecular Microbiology*, *49*(2), 277–300.

Casjens, S. R. (2005). Comparative genomics and evolution of the tailed-bacteriophages. *Current Opinion in Microbiology*, *8*(4), 451–458.

Cobián Güemes, A. G., Youle, M., Cantú, V. A., Felts, B., Nulton, J., & Rohwer, F. (2016). Viruses as Winners in the Game of Life. *Annual Review of Virology*, *3*(1), 197–214.

Cruz-Flores, R., Cáceres-Martínez, J., Muñoz-Flores, M., Vásquez-Yeomans, R., Rodriguez, M. H., Del Río-Portilla, M. Á., Rocha-Olivares, A., & Castro-Longoria, E. (2016). Hyperparasitism by the bacteriophage (Caudovirales) infecting Candidatus Xenohaliotis californiensis (Rickettsiales-like prokaryote) parasite of wild abalone Haliotis fulgens and Haliotis corrugata from the Peninsula of Baja California, Mexico. *Journal of Invertebrate Pathology*, *140*, 58–67.
Cui, J., Schlub, T.E. and Holmes, E.C., 2014. An allometric relationship between the genome length and virion volume of viruses. Journal of virology, 88(11), pp.6403-6410.

Davies, E. V., James, C. E., Williams, D., O'Brien, S., Fothergill, J. L., Haldenby, S., Paterson, S., Winstanley, C., & Brockhurst, M. A. (2016). Temperate phages both mediate and drive adaptive evolution in pathogen biofilms. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8266–8271.

Egan, S., & Gardiner, M. (2016). Microbial Dysbiosis: Rethinking Disease in Marine Ecosystems. *Frontiers in Microbiology*, *7*, 991.

Friedman, C. S., & Crosson, L. M. (2012). Putative Phage Hyperparasite in the Rickettsial Pathogen of Abalone, "Candidatus Xenohaliotis californiensis." *Microbial Ecology*, *64*(4), 1064–1072.

Klemetsen, T., Raknes, I. A., Fu, J., Agafonov, A., Balasundaram, S. V., Tartari, G., Robertsen, E., & Willassen, N. P. (2018). The MAR databases: development and implementation of databases specific for marine metagenomics. *Nucleic Acids Research*, *46*(D1), D692–D699.

Knowles, B., Silveira, C. B., Bailey, B. A., Barott, K., Cantu, V. A., Cobián-Güemes, A. G., Coutinho, F. H., Dinsdale, E. A., Felts, B., Furby, K. A., George, E. E., Green, K. T., Gregoracci, G. B., Haas, A. F., Haggerty, J. M., Hester, E. R., Hisakawa, N., Kelly, L. W., Lim, Y. W., Little, M., Luque, A., McDole-Somera, T., McNair, K., de Oliveira, L. S., Quistad, S. D., Robinett, N. L., Sala, E., Salamon, P., Sanchez, S. E., Sandin, S., Silva, G. G. Z., Smith, J., Sullivan, C., Thompson, C., Vermeij, M. J. A., Youle, M., Young, C., Zgliczynski, B., Brainard, R., Edwards, R. A., Nulton, J., Thompson, F., & Rohwer, F. (2016). Corrigendum: Lytic to temperate switching of viral communities. *Nature*, *539*(7627), 123.

Lainhart William, Stolfa Gino, & Koudelka Gerald B. (2009). Shiga Toxin as a Bacterial Defense against a Eukaryotic Predator, Tetrahymena thermophila. *Journal of Bacteriology*, *191*(16), 5116–5122.

Lawrence, J. G. (2002). Gene transfer in bacteria: speciation without species? *Theoretical Population Biology*, *61*(4), 449–460.

Maslov, S., & Sneppen, K. (2015). Well-temperate phage: optimal bet-hedging against local environmental collapses. *Scientific Reports*, *5*, 10523.

Obeng, N., Pratama, A. A., & van Elsas, J. D. (2016). The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. *Trends in Microbiology*, *24*(6), 440–449.

Örmälä-Odegrip, A.-M., Ojala, V., Hiltunen, T., Zhang, J., Bamford, J. K. H., & Laakso, J. (2015). Protist predation can select for bacteria with lowered susceptibility to infection by lytic phages. *BMC Evolutionary Biology*, *15*, 81.

Rohwer, F., & Thurber, R. V. (2009). Viruses manipulate the marine environment. *Nature*, *459*(7244), 207–212.

Roux, S., Hallam, S. J., Woyke, T., & Sullivan, M. B. (2015). Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *eLife*, *4*, e08490.

Silveira, C. B., & Rohwer, F. L. (2016). Piggyback-the-Winner in host-associated microbial communities. *NPJ Biofilms and Microbiomes*, *2*, 16010.

Thingstad, T. F. (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography*, *45*(6), 1320–1328.

Touchon, M., Bernheim, A., & Rocha, E. P. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *The ISME Journal*, *10*(11), 2744–2754.

Ushijima, B., Videau, P., Burger, A. H., Shore-Maggio, A., Runyon, C. M., Sudek, M., Aeby, G. S., & Callahan, S. M. (2014). Vibrio coralliilyticus strain OCN008 is an etiological agent of acute Montipora white syndrome. *Applied and Environmental Microbiology*, *80*(7), 2102–2109.

Weynberg, K. D., Voolstra, C. R., Neave, M. J., Buerger, P., & van Oppen, M. J. H. (2015). From cholera to corals: Viruses as drivers of virulence in a major coral bacterial pathogen. *Scientific Reports*, *5*, 17889.

Wommack, K. E., & Colwell, R. R. (2000). Virioplankton: viruses in aquatic ecosystems. *Microbiology and Molecular Biology Reviews: MMBR*, *64*(1), 69–114.

Young, R. (1992). Bacteriophage lysis: mechanism and regulation. *Microbiological Reviews*, *56*(3), 430–481.

CHAPTER 2: A MULTIOMIC ANALYSIS OF IN SITU CORAL-TURF ALGAL INTERACTIONS

Abstract

Viruses, microbes, and host macroorganisms form ecological units called holobionts. Here, a combination of metagenomic sequencing, metabolomic profiling, and epifluorescence microscopy was used to investigate how the different components of the holobiont including bacteria, viruses, and their associated metabolites mediate ecological interactions between corals and turf algae. The data demonstrate that there was a microbial assemblage unique to the coral-turf algae interface displaying higher microbial abundances and larger microbial cells. This was consistent with previous studies showing that turf algae exudates feed interface and coral-associated microbial communities, often at the detriment of the coral. Further supporting this hypothesis, when the metabolites were assigned a nominal oxidation state of carbon (NOSC), we found that the turf algal metabolites were significantly more reduced (i.e., have higher potential energy) compared to the corals and interfaces. The algae feeding hypothesis was further supported when the ecological outcomes of interactions (e.g., whether coral was winning or losing) were considered. For example, coral holobionts losing the competition with turf algae had higher Bacteroidetes-to-Firmicutes ratios and an elevated abundance of genes involved in bacterial growth and division. These changes were similar to trends observed in the obese human gut microbiome, where overfeeding of the microbiome creates a dysbiosis detrimental to the long-term health of the metazoan host. Together these results show that there are specific biogeochemical changes at coral–turf algal interfaces that predict the competitive outcomes between holobionts and are consistent with algal exudates feeding coral-associated microbes.

Introduction

Coral and algae holobionts are assemblages of the macroorganisms and their associated viruses, bacteria, archaea, and protists (Knowlton and Rohwer 2003). The microbial portion of the holobiont, the microbiome, is often species-specific, temporally stable, and distinct between microhabitats within the host (Bourne et al., 2016; Frias-Lopez et al., 2002; Littman et al., 2009; Mouchka et al., 2010; Rohwer et al., 2002; Sunagawa et al., 2010; Sweet et al., 2011; Hester et al., 2016). The microbiome performs a wide array of functions that influence both host physiology (Rosenberg et al., 2007; Bourne et al., 2009; Lynch et al., 2016) and the biogeochemical cycling of matter and energy (Lesser et al., 2004; Wegley et al., 2007; Beman et al., 2008; Siboni et al., 2008; Raina et al., 2009; Fiore et al., 2010; Kimes et al., 2012; Haas et al., 2016; Ferrier-Pagès et al., 2016; Roach et al., 2017). The health of the holobiont is linked to the composition of its microbial constituents, which can be disrupted by various stressors and lead to dysbioses (Bourne et al., 2008; Mao-Jones et al., 2010). Host-associated microbes may also increase holobiont resistance and resilience to both local and global stressors (Ziegler et al., 2017). For instance, viral and bacterial symbionts may ward off potential pathogens through lysis, niche exclusion, and production of antibiotics (Ritchie 2006; Nissimov et al., 2009; Rypien et al., 2010). These interacting biological entities will influence the chemistry of holobionts, which will be reflected in the metabolites.

Understanding the roles of holobionts in ecosystem function has become increasingly important as many reefs that were formerly dominated by coral have been shifting to systems dominated by turf and fleshy macroalgae (McCook 1999; McCook et al., 2001; Hughes et al., 2003; Hughes et al., 2007; Smith et al., 2016). Turf algae are among the most abundant algal competitors that corals face (Barott et al., 2009; Barott et al., 2012; Haas and Wild 2010) and, as

such, play an important role in initiating algal phase shifts on coral reefs. Competition with turf algae is known to alter the microbial communities associated with corals (Barott and Rohwer 2012). However, the role of the microbiome and its associated metabolome in determining the outcome of these events (i.e., whether a coral wins or loses against its algal competitor) is still relatively unknown.

To examine how microbial diversity, metabolic capacity, and biochemistry affect ecological interactions between macroorganisms, both surface-associated water (n =18) and tissue (n = 36) samples were taken at a centimeter-scale spatial resolution across in situ coral–turf algal interactions from reefs on the island of Curaçao. All of these interactions involved turf algal holobionts interacting with either *Diploria strigosa* or *Orbicella faveolata* (formerly *Montastraea faveolata*) corals. Tissue samples were processed for metagenomic sequencing and metabolomic profiling, and the surface-associated samples were analyzed by epifluorescence microscopy (SI Appendix, Supplementary Methods). Microbial and viral abundance in the holobiont were directly quantified through microscopy. The microbiome was taxonomically and functionally profiled through metagenomic sequencing, and the molecular composition of the holobiont was assessed by untargeted metabolomic analysis. A machine learning approach was applied to identify which bacterial taxa, functional genes, and metabolites were most important for determining whether a coral was winning (i.e., had no algal overgrowth and no visible signs of paling or necrosis) or losing (i.e., had algal overgrowth and had visible signs of paling or necrosis) against its turf algal competitor. The results of this study show that there are specific functional genes, microbial taxa, and metabolites which distinguish coral, turf algae, and interface communities and that these functions, taxa, and metabolites are also linked to the competitive outcomes of these interactions. Specifically, the data demonstrate that there is an

emergent microbial community that forms at the interface between coral and turf algae, which is characterized by larger and more numerous bacterial cells, a greater proportion of Bacteroidetes, and a lower proportion of Firmicutes, as well as an enrichment in genes involved in bacterial cell growth and division. Furthermore, the same taxa and functional genes (i.e., Bacteroidetes, Firmicutes, and bacterial cell growth and division genes) are also significant predictors of whether the corals in these interactions were winning or losing against their turf algal competitor (i.e., whether the coral in the interaction was being overgrown by algae and showed visible signs of paling and/or tissue necrosis). In sum, the data presented here provide a comprehensive multiomic analysis of *in situ* coral–turf algal interactions, which illustrates the ecological role of the holobiont in organismal interactions in one of the most diverse and economically valuable ecosystems on Earth: coral reefs.

Results

Site Selection and Water Chemistry

All samples were collected from the Caribbean island of Curaçao in November 2015 (SI Appendix, Figure S2.1, Table S2.1, and Supplementary Methods). Site level water chemistry data including inorganic nitrogen, phosphate, and dissolved organic carbon (DOC) concentrations are shown in SI Appendix, Table S2.2. No sites were identified as outliers in the dataset for any of the measured water chemistry parameters, and these parameters were not significantly different between coral species ($n = 6$, all ANOVA P values ≥ 0.34) or competition outcome ($n = 6$, all ANOVA p-values ≥ 0.27) (SI Appendix, Table S2.2).

Concentrations of Surface-Associated Viruses and Microbes Across Coral–Algal Interactions

Samples for microscopic direct counts were taken by suctioning water and mucus directly off the surface of the coral, the turf algae, and the coral–turf algal interface with a syringe (Figure 2.1A). There was a general trend for the interfaces ($4.80 \times 10^6 \pm 2.46 \times 10^6$ cells·mL^{-1}; mean ± SEM) to exhibit higher microbial abundances than corals ($1.45 \times 10^6 \pm 2.12 \times 10^6$ cells·mL^{-1}; mean ± SEM) or turf algae ($2.50 \times 10^6 \pm 5.19 \times 10^6$ cells·mL^{-1}; mean ± SEM (Figure 2.1B). Coral-associated ($8.24 \times 10^6 \pm 1.56 \times 10^6$ VLPs·mL^{-1}; mean ± SEM) and interface-associated ($9.64 \times 10^6 \pm 2.44 \times 10^6$ virus-like particles [VLPs]·mL^{-1}; mean ± SEM) samples had a significantly higher concentration of viruses than turf algal-associated ($4.04 \times 10^6 \pm 2.14 \times 10^6$ VLPs·mL^{-1}; mean ± SEM) samples (Figure 2.1C). Furthermore, coral-associated (6.270 ± 2.222; mean ± SEM) samples had a significantly higher virus-to-microbe ratio (VMR) than the interface-associated (2.465 ± 1.156; mean ± SEM) or turf algae-associated (1.674 ± 0.470; mean ± SEM) samples (Figure 2.1D). Overall, VLPs and microbial cells exhibited a significantly

positive correlation ($R^2 = 0.116$, $p = 0.0008$) with a slope of less than one (m = 0.268), and the VMR showed a significantly negative correlation ($R^2 = 0.495$, $p = 0.0016$) with microbial cells (Figure 2.1E), indicating a reduced lytic production of viruses at higher microbial concentrations.

Epifluorescence micrographs were also used to determine microbial cell size (SI Appendix, Supplementary Methods), and the results showed that there were significantly larger microbes ($p < 0.001$) associated with the coral–algal interface (0.351 ± 0.0094 μm^3; mean ± SEM) when compared to the microbes associated with either the coral (0.212 ± 0.0134 μm3; mean ± SEM) or the turf algae (0.248 ± 0.0146 μm^3; mean ± SEM) alone (Figure 2.1F). Cell sizes were used to calculate a predicted community metabolic power output via metabolic theory of ecology (MTE). MTE calculations demonstrated that the combination of cell size and concentration yielded a significantly ($p < 0.001$) higher predicted community power output at the coral-algal interface ($280{,}389.167 \pm 140{,}518.682$ $W \cdot m^{-2}$; mean ± SEM) (Figure 2.1G).

Holobiont and Interface-Specific Bacterial Taxa, Functions, and Metabolites.

Biopsies for multiomic analysis were taken using an underwater power drill in a transect perpendicular to the coral–turf algal interface (Figure 2.1A) and processed for metagenomes (see SI Appendix, Table S2.3 for details on metagenomic libraries) and metabolomes (SI Appendix, Supplementary Methods). Ward's hierarchical clustering method was used to distinguish trends in the bacterial taxonomic and functional composition of metagenomes as well as for metabolomes. Overall, samples clustered by microorganism (Figure 2.2). Functional genes (Figure 2.2A), bacterial taxa (Figure 2.2B), and metabolites (Figure 2.2C) showed the strongest support for two groups, namely, coral and noncoral (i.e., turf algae plus interface). Functional genes and bacterial taxa also showed support for two subgroups within both the coral and

noncoral clusters. Within the coral cluster, the two different species of coral studied, *D. strigosa* and *O. faveolata*, were observed as distinct groups. Within the noncoral cluster, there was a subclustering of interaction interfaces versus turf algae (Figure 2.2A and B). These trends were recovered from the functional annotations at both level 1 (SI Appendix, Figure S2.2) and level 3 (Figure 2.2A and SI Appendix, Figure S.23) of the SEED database (Overbeek et al., 2005), and the taxonomic trends were observed at both the phylum (SI Appendix, Figure S2.4) and order (Figure 2.2B and SI Appendix, Figure S2.5) levels. No trends were observed in clustering by sample site.

Two-way cluster analysis showed that the interaction interface group was not merely a mix of the coral and algal groups, but rather had its own unique functional, taxonomic, and metabolomic profiles (SI Appendix, Figures S2.2–S2.6); however, the interface samples were more similar to the algal samples than to coral samples. The unique microbiome at the interface had a significantly higher relative abundance of Bacteroidetes, specifically those in the order Flavobacteriales ($p < 0.0001$), and a significantly lower relative abundance of Firmicutes ($p = 0.003$) when compared to the coral or turf algal microbiome (SI Appendix, Figure S2.7). Functional gene annotations demonstrated that there was a significant overrepresentation of genes involved in cell cycle and cell division ($p < 0.0001$) and a significant underrepresentation of genes involved in protein metabolism ($p = 0.006$) at the interface (SI Appendix, Figure S2.8). In the metabolomic dataset, 182 molecules were significantly more abundant at the interface relative to both coral and turf algae. All of these molecules were unknown (i.e., there were no spectral matches to the reference libraries) except for ceramide 18:1/16:0 ($p = 0.0039$; Figure 2.3A). This was a level 2 annotation according to the 2007 metabolomics standards initiative (39–41) (SI Appendix, Figure S2.9). Ceramide 18:1/16:0 and its less saturated form, ceramide

18:2/16:0 (level 2) were found in coral and interface samples but were completely absent in turf algal samples. The less saturated (18:2/16:0) form was found in equal abundance in both coral and interface, whereas the more saturated form (18:1/16:0) was found to be significantly higher at the interface (Figure 2.3A and B and SI Appendix, Figure S2.10).

To further investigate the metabolomic samples, each metabolite was assigned a nominal oxidation state of carbon (NOSC) and a Gibbs free energy of carbon oxidation under standard conditions ($\Delta G^{o}Cox$) based on their putative molecular formula (with mass accuracy of <1 ppm) (SI Appendix, Supplementary Methods). Overall, 7,751 of the total 8,427 features (i.e., ~92%) were assigned a putative molecular formula using SIRIUS4, with a mass accuracy cutoff of 0.0020 ppm. Metabolites in turf algal samples had significantly lower NOSCs and significantly higher $\Delta G°Cox$ values than did interface or coral samples (NOSC $p = 0.0486$; $\Delta G^{o}Cox$ $p = 0.0201$) (Figure 2.3C and D). This indicates that the biochemical compounds in turf algal samples were more reduced and were thus more energy rich. No difference in NOSC or ΔG was found between interactions where corals were winning or losing the competition with turf algae.

Machine Learning to Identify Bacterial Taxa, Functions, and Metabolites that Best Predict Competition Outcomes

Coral–turf algal interactions were classified as winning (i.e., coral winning) or losing (i.e., coral losing) based on the criteria proposed in Barott et al. (2012). Briefly, losing corals were classified as those corals that had significant algal overgrowth along with visible paling and/or tissue necrosis, whereas winning corals were corals that did not have algal overgrowth and did not show any signs of paling or necrosis. To distinguish which holobiont variables were linked to the competitive outcomes of the interactions, a random forests classification analysis was used. Random forests analysis showed that there were important functional genes (SI

Appendix, Figures S2.11–S2.13), bacterial taxa (SI Appendix, Figures S2.14–S2.16), and metabolites (SI Appendix, Figures S2.17–S2.19), which distinguished winning and losing interactions.

The top 10 most important variables in each analysis (SI Appendix, Figures S2.11-S2.19) were used to construct a two-way hierarchical dendrogram (Figure 2.4) to visualize distinguishable groups of winning and losing corals, interfaces, and algae. Using these important taxa (Figure 2.4D-F and SI Appendix, Table S2.4), functional genes (Figure 2.4A-C and SI Appendix, Table S2.5), and metabolites (Figure 2.4 G-I), samples clustered significantly by winning or losing interactions. Notably, all of the bacterial orders associated with losing interfaces were from a single bacterial phylum, Bacteroidetes.

To further elucidate the taxonomic and functional mechanisms involved in winning and losing coral–algal interactions, one-way ANOVAs were performed on all bacterial phyla and level 1 SEED metabolic categories followed by Tukey post hoc analysis on all significant variables (ANOVA p-values ≤ 0.05). This analysis revealed that there were only two phyla and two SEED subsystems that were significantly different at the interface relative to both coral and turf algal samples (i.e., ANOVA p-value ≤ 0.05 and Tukey p-value ≤ 0.05). The phylum Bacteroidetes was significantly enriched in interface samples, while the phylum Firmicutes was significantly depleted in the interface samples (Figure 2.5A and SI Appendix, Figure S2.7). Functionally, genes involved in cell division were significantly enriched in interface samples (Figure 2.5B and SI Appendix, Figure S2.8). The taxonomic shift toward Bacteroidetes and concomitant enrichment in cell division-related genes, as observed at the interface, was also observed on coral tissue of losing colonies (Figure 2.5A and SI Appendix, Figure S2.7). These results demonstrate a significantly higher Bacteroidetes-to-Firmicutes ratio at the interface and in

losing corals (Figure 2.5A). Regression analysis showed that the Bacteroidetes to Firmicutes ratio was significantly correlated with genes involved in cell division (Figure 2.5C) ($R^2 = 0.560$, $F(1,16) = 20.4$, $p = 0.0004$) as well as being significantly correlated with total microbial biomass (Figure 2.5D) ($R^2 = 0.402$, $F(1,16) = 10.76$, $p = 0.0047$). Taken together, these results suggest that an increase in Bacteroidetes is linked to an increase in cell division, greater bacterial cell size, and total biomass and that this shift toward faster growing, larger Bacteroidetes is linked to whether corals win or lose in their competitive interactions against benthic algae.

Discussion

The data presented herein illustrate that there are significant differences in the size, abundance, and community composition of microbes across in situ coral–turf algal interfaces (Figure 2.1). These differences show there is an emergent community that forms at the interface between coral and turf algae, which is characterized by larger and more numerous bacterial cells, a higher proportion of Bacteroidetes, a lower proportion of Firmicutes, an enrichment in genes involved in bacterial cell growth and division, and an increase in the potentially proapoptotic compound ceramide 18:1/16:0.

Microbial and Viral Abundances

The highest viral abundance was in the coral holobiont. The increased viral abundance in the coral-surface holobiont may be due to the bacteriophage adherence to mucus (BAM) dynamics described in Barr et al. (2013). BAM dynamics imply that bacterial viruses (bacteriophage or phage) adhere to mucus glycoproteins through noncovalent interactions with capsid proteins. Corals may use these mucus-attached phages to defend against invading bacterial pathogens. The combination of the differences in viral and bacterial abundances leads

to a significant difference in the VMR across the interface, where coral holobionts have the lowest microbial load but the highest VMR (Figure 2.1D). Furthermore, metagenomic analyses demonstrated higher levels of prophage in algal samples (SI Appendix, Figure S2.20), where VMR was the lowest (Figure 2.1D). Taken together, these data suggest a trend toward a decrease in lytic activity and an increase in viral lysogeny at higher microbial concentrations (Figure 2.1E and SI Appendix, Figure S2.20). A similar trend has been seen in other environments (Knowles et al., 2016; Coutinho et al., 2017), including the water column of tropical coral reefs, and has been described as Piggyback-the-Winner dynamics (Knowles et al., 2016). Piggyback-the-Winner posits that when VMRs are low, such as in interface and turf algal samples, there are more bacterial cells harboring lysogenic phage. This means that the bacterial assemblages at the interface and turf algae harbor a higher proportion of phage-encoded genes, which has been linked to increased pathogenicity of the overall community (Brüssow et al., 2004; Brüssow 2007; Munro et al., 2003; Aziz et al., 2005; Weynburg et al., 2015; Little et al., 2020). These microbe-phage dynamics may be another mechanism at play in the complex interactions of coral and algal holobionts.

Microbial Biomass and Energetics

The results also demonstrate that there are significantly larger microbial cells at the interface between coral and turf algae (Figure 2.1F). This change in cell size coupled to the cell concentrations leads to a higher predicted metabolic power output at the coral–algal interface (Figure 2.1G). This trend of higher power output has also been reported using calorimetry in controlled laboratory experiments (Roach et al., 2017). A higher microbial power output at the coral–algal interface means that the microbial populations here are using energy at a faster rate

and are dissipating more of that energy as heat (Roach et al., 2017; Roach et al., 2018). The increase in metabolic rate at the interface may be responsible for the reported decreased oxygen levels at the interface (Roach et al., 2017; Silveira et al., 2019; Haas et al., 2013; Gregg et al., 2013; Jorrisen et al., 2016). Understanding the direct links between bacterial taxa, cell size, power output, and biological oxygen demand may provide a more complete conceptual model of the way bacterial metabolism is involved in competitive interactions between benthic macroorganisms.

Emergent Microbiome and Metabolome at the Coral–Turf Algal Interface

Metabolomic samples showed clear differences between coral and noncoral holobionts. The interface exhibited a unique chemical signature, however, the metabolites driving differences at the interface were mostly unknown compounds. One known compound (level 2 according to the metabolomics standards initiative) was the potentially proapoptotic molecule, ceramide 18:1/16:0 (Figure 2.3A). Other bioactive lipids and proapoptotic inflammatory modulators have previously been shown to play a role in the coral holobiont (Quinn et al., 2016; Galtier d'Auriac et al., 2018; Roach et al., 2021), suggesting that non-self-recognition among some of the oldest extant holobionts (i.e., corals) involves bioactive lipids identical to those in highly derived taxa like humans. The data here further strengthen the hypothesis that major players of the immune response evolved during the pre-Cambrian era (Quistad et al., 2014). Furthermore, turf algal metabolites were found to have more negative nominal oxidation states of carbon and higher ΔG of carbon oxidation (Figure 2.3C and D), suggesting that the biochemicals in the turf algal holobiont are more reduced and, thus, more energy rich. It may be the combination of naive coral microbes being exposed to high-energy turf algal compounds at the

interface, which leads to the increase in size and power output of the bacterial cells at the interface (Figure 2.1F and G). The feeding of coral microbes on turf algal metabolites at the interface may also be in part responsible for the decrease in oxygen levels previously observed at the coral–algal interface (Roach et al., 2017; Haas et al., 2013). Thus, we propose the "algae feeding hypothesis" where reduced, high-energy turf algae exudates feed interface and coral-associated microbial communities, often to the detriment of the coral animal.

The metagenomic and metabolomic data show that there are specific bacterial taxa, functional genes, and metabolites that distinguish coral, turf algae, and interfaces (Figure 2.2), as well as winning and losing interactions (Figure 2.4 and SI Appendix, Tables S2.4 and S2.5). Furthermore, these data indicate that the interface is not merely a mix of coral and turf algal holobionts, but rather has its own emergent signature (SI Appendix, Figures S2.2–S2.6), which is more similar to the turf algal holobiont than the holobiont of coral (Figure 2.2). Specifically, members of the Bacteroidetes clade are overrepresented at the interaction interface, while the phylum Firmicutes is underrepresented at the interface (Figure 2.5 and SI Appendix, Figure S2.7). A similar trend is observed in coral samples, where Bacteroidetes are enriched in losing corals and Firmicutes are depleted in losing coral samples (SI Appendix, Figure S2.7). Thus, the data demonstrate that the Bacteroidetes-to-Firmicutes ratio is a significant predictor of whether a coral will win or lose in a competitive interaction with algae. The ratio of Bacteroidetes to Firmicutes is also a significant predictor of health status in other systems such as the human gut where this ratio has been linked to obesity (Ley et al., 2006) and in the human lung where it is linked to disease states in cystic fibrosis patients (Garg et al., 2017). The Bacteroidetes-to-Firmicutes ratio was a significant predictor of the amount of cell division genes and total microbial biomass in these holobionts (Figure 2.5D). Studies in mice and humans have shown

that the change in the Bacteroidetes-to-Firmicutes ratio can have significant impacts on energy output and biomass of microbial communities, with Bacteroidetes having an increased capacity to harvest energy from reduced compounds (Turnbaugh et al., 2006). Given that the abundance of Bacteroidetes and the Bacteroidetes-to-Firmicutes ratio was higher at the interface and in losing coral samples, we propose a working model (Figure 2.6) whereby the reduced metabolites released by turf algae select for an increase in Bacteroidetes relative to Firmicutes, which, in turn, leads to a faster growing bacterial community with larger cells and higher energy use rate. These fast-growing microbes can outcompete the corals for resources such as oxygen, which weaken the coral and lead to eventual algal overgrowth. This link between the energy content of algal metabolites, bacterial taxonomic composition, community metabolism, and coral health provides interesting insight into the ways that different components of the holobiont affect the outcome of ecological interactions and eventually shape entire community assemblages.

Despite the current progress in the field, it is worth noting that environmental multiomics still has a long way to go. Metabolomics databases are sparsely populated in regard to environmental metabolites making annotation difficult and leaving the majority of compounds unclassified. As this work and others (e.g., Robbins et al., 2019) have demonstrated the need to consider all components of the holobiont, it is clear that new methods and increased sequencing efforts will be needed to provide the amount of microbial coverage necessary to properly describe the roles of the less abundant components of the holobiont such as archaea and viruses. Thus, we highlight that future work is needed to provide more robust analyses of the coral and algal holobionts and their associated metabolites.

Conclusions

Overall, this study demonstrates that there are differences in both the surface-associated microbial community and the total holobionts of coral and turf algae, and that when these organisms interact, there is an emergent interface community. We hypothesize that this emergent community is driven by the coral microbiome feeding on the energy-rich exudates released by the adjacent turf algae, a phenomenon we term the algae feeding hypothesis. The data also show that specific bacterial groups such as Bacteroidetes and Firmicutes play a role in determining the competitive outcome of coral–turf algae interaction events. However, what this role is remains an open question and will require further investigation. In sum, we emphasize the role of host-associated microbial communities in ecological processes and highlight that the holobiont plays an important part in determining the outcome of coral–turf algal interactions and overall reef health.

Significance

All plants and animals are associated with communities of viruses and microbes that interact via a suite of metabolites. These components play critical roles in the success of these assemblages; however, the role of individual components (i.e., bacteria, viruses, metabolites) and how these govern ecological interactions between macroorganisms is not understood. This study investigates the role of holobiont components in coral–turf algal interactions. The data demonstrate that an emergent microbiome and metabolome form at the interface between coral and turf algae in competitive interactions. Machine learning analyses show that this emergent community predicts the outcome of these interactions. These results provide insight into rules of

community assembly in microbiomes and the roles of holobiont components in mediating ecological interactions.

Methods

Methods Overview

All samples were collected from in situ coral–turf algal interactions by divers on SCUBA around the Caribbean island of Curaçao in November of 2015. Surface-associated water samples and tissue biopsies were collected to assess the surface-associated epibiont and the tissue-associated microbial communities of corals and turf algae in direct competitive interactions with one another. Surface-associated water samples were processed for microscopy using both SYBR Gold and DAPI stains to quantify the abundance of bacteria and virus-like particles as well as the size of bacterial cells. Tissue biopsies were processed for metagenomics sequencing on a MiSeq platform and untargeted metabolomics via liquid chromatography tandem mass spectrometry. A machine learning approach was applied to metagenomic and metabolomic datasets to identify which bacterial taxa, functional genes, and metabolites were most important for determining whether a coral was winning or losing against its turf algal competitor. See SI Appendix, Supplementary Methods for a detailed explanation of all methods.

Data availability

Metagenomic sequence data from this study has been deposited into the Sequence Read Archive under the study accession code PRJNA597953. Metabolite library spectra files are available through the GNPS at the following link: (https://gnps.ucsd.edu/ProteoSAFe/result.jsp?task=7d1e1045428548bbad575ff12445a48c&view=advanced_view).

Supplemental Methods

Sample collection

All samples were collected by divers on SCUBA in November 2015 around the island of Curaçao (12.1696° N, 68.9900° W). Samples for microscopy were taken by suctioning water and mucus directly off the surface of the coral, the algae, and the coral-algal interface using a 3 mL blunt tip syringe. Coral and algal samples were taken 10 cm away from the interface (Figure 2.1) by suctioning 100 μL of surface-associated water and mucus every 0.5 cm over a 10 cm transect parallel to the interface, yielding a total sample volume of 2 mL. A similar process was conducted along a 10 cm transect of the interface.

Biopsies for metagenomic and metabolomic analysis were taken using an underwater power drill (Nemo Power Tools, Santa Clara, CA, USA). Biopsies, 1 cm in diameter and 1 cm in length, were collected in a transect perpendicular to the coral-algal interface (Figure 2.1). The coral and algal samples were taken 10 cm away from the interface on their respective sides. The Interface samples were taken directly at the interface of the coral and turf algae. Biopsies for metabolomic analysis were placed into 10 mL of LCMS grade 70% methanol and 30% water for metabolite extraction and later analyzed via liquid chromatography tandem mass spectrometry (LC-MS/MS). Samples for Metagenomic analysis were placed in RNA Later (Thermo Fisher Scientific, Waltham, MA, USA) and stored at -80° C after 30 min.

Epifluorescence microscopy

Microscopy samples were divided into two aliquots in order to analyze both viral abundance and microbial size via SYBR and DAPI staining respectively. SYBR aliquots were fixed with microscopy grade paraformaldehyde at 1% final concentration, vacuum filtered onto a

0.02 µm Anodisc filter (Whatman Inc., Florham Park, NJ, USA), stained with SYBR Gold (5 X final concentration; Invitrogen, Carlsbad, CA, USA), and mounted on microscope slides. DAPI aliquots were fixed with microscopy grade glutaraldehyde at 2% final concentration, vacuum filtered onto a 0.2 µm Anodisc filter (Whatman Inc., Florham Park, NJ, USA), stained with DAPI (5 $\mu g \cdot mL^{-1}$ final concentration; Invitrogen, Carlsbad, CA, USA), and mounted on microscope slides. The stained filters were imaged using an epifluorescence microscope (excitation/emission: 358/461 nm) at 600X magnification and were quantified using Image Pro software (Media Cybernetics).

Calculation of metabolic power output

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

$$I = i0M\alpha e^{-E/kT}$$

Where $i0$ is the mass-independent normalization constant, M is the wet weight of the organism in grams, and α is the scaling exponent. The effects of temperature on metabolic rate are accounted for by $e^{-E/kT}$ where E is the activation energy, k is Boltzmann's constant (8.62×10^{-5} eV· K^{-1}), and T is the water temperature (in Kelvin) at the site at the time of collection. Community-level metabolic rates were calculated by summing the individual metabolic rates (I) for all microbes in a sample.

Metagenome generation and analysis

Total DNA was extracted from coral punches using the AllPrep DNA/RNA kit (Qiagen). Metagenomic libraries were constructed using Illumina Nextera XT library preparation kits and sequenced at the SDSU sequencing facility via 96-plex sequencing with 600 cycles using the MiSeq platform with 2 x 300 pair-end read chemistry. Raw reads were quality-filtered by removing short reads (< 60 bp), reads with quality scores < 20, reads with >1% ambiguous bases, low complexity reads (entropy > 70), and duplicate reads, using the program PRINSEQ (Schmieder and Edwards 2011). Quality filtered metagenomic libraries were aligned using SUPERFOCUS (Silva et al., 2016) and the SEED hierarchical database of BLASTX-translated protein orthologs classified according to putative functional families (Overbeek et al., 2014). Relative abundances of taxa and functional gene classifications within each metagenome were used as input to the multivariate statistical and distance-based analyses described in the statistical analysis section.

Shotgun sequence metagenomic libraries generated a total of 19,388,513 raw reads with an average of 1,077,140 reads per sample (+ 148,402 raw reads per sample). This resulted in 18,819,251 quality-filtered reads with an average of 1,045,514 quality-filtered reads per sample (± 144,414 reads) of which an average of 12.3% (± 3%) were bacterial and < 0.5% (± 0.07%) were non-bacterial microbes. More information and statistics on metagenomic libraries can be found in Table S2.3.

Bioinformatics search for prophage in metagenomes

Unassembled metagenomic reads were queried using Fragment Recruit Assembly Purification (FRAP, https://github.com/yinacobian/frap) against a prophage protein database and

bacterial genome reference dataset. FRAP uses the SMALT pairwise sequence alignment program where we chose an ≥80% nucleotide identity for the complete query read against both databases (Ponstingl and Ning 2010). The prophage protein dataset consisted of 1.5 million contigs from PhiSpy predicted proteins in the NCBI RefSeq database and the bacterial genome database contained 66,000 complete bacterial genomes from NCBI RefSeq (Akhter et al., 2012; Pruitt et al., 2007). Our metric for percent prophage was determined by dividing the number of hits to the PhiSpy predicted protein database by the hits to the NCBI bacterial genome reference set.

Ultra-Performance Liquid Chromatography – Tandem Mass Spectrometry

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

A Maxis Q-TOF mass spectrometer (MS) (Bruker Daltonics) was coupled to the UPLC system, directly measuring the compounds coming off the LC-column. The spectrometer was

equipped with an electrospray ionization (ESI) source (200°C). Positive ion mode acquired MS spectra in the range of 50 – 2000 m/z. Prior to data collection, the spectrometer was externally calibrated with ESI-L Low Concentration Tuning Mix (Agilent technologies, Santa Clara, CA, USA). Throughout runs, Hexakis (1H,1H,3H-tetrafluorpropoxy) phosphazene (Synquest Laboratories, Alachua, FL, USA; m/z 922.0098) was used as the internal calibrant. Both MS1 and MS2 had a nebulizer gas pressure (nitrogen) of 2 bar, dry gas flow of 9 $L \cdot min^{-1}$ source temperature, and a capillary voltage of 4500 V. MS1 had a spectral rate of 3 Hz, and MS2 a spectral rate of 10 Hz. To obtain MS/MS fragmentation, the ten most intense ions per MS1 scan were introduced into the MS2, where they were fragmented using collision-induced dissociation. Automatic exclusion was used where an ion would be ignored in more than 3 scans, but when intensity was 2.5x the previous scan it would be re-fragmented.

Feature Table Generation

The raw datafiles from the MS machine were converted into .mzXML files with the Bruker Data Analysis software version 4.1. The .mzXML files are available on the MassIVE database (massive.ucsd.edu) under number MSV000080597 and MSV000080632 (same dataset). The .mzXML files were imported into MZmine (Pluskal et al., 2010). Mass detection threshold for MS1 was 3.00E+03, and 1.00E+02 for MS2. For building the chromatogram a minimum peak height of 6.00E+03, with a minimum peak duration of 1.00E-2 was set, together with a mass error of 2.50E+01 ppm, and 5.00E-02 m/z. For deconvolution, we used a baseline cutoff, with the minimum peak height of 4.00E+03, peak duration range of 0.01 to 3 min., and a baseline level of 1.00E+03. The m/z range for MS2 scan paring was set to 5.00E-02 Da, and the retention time range was set to 0.2 min. Isotope peaks were grouped with the m/z tolerance set to

5.00E-02 or 25 ppm, and a retention time tolerance of 0.1 and a maximum charge of 4. The representative isotope would be the most intense one. Features of different samples are aligned with an m/z tolerance of 5.00E0.2 m/z or 25 ppm with a weight for m/z of 75. Retention time tolerance for alignment was set to 0.1 with a weight of 25. The peak list is filtered for a minimum of 2 peaks in a row, and 2 peaks in an isotope pattern. Only peaks with an MS2 scan were kept. Duplicated peaks were filtered out using the m/z tolerance of 5.00E-02 m/z or 25 ppm, and a retention time tolerance of 0.1. Gap filling occurred with an intensity tolerance of 10%, m/z tolerance of 5.00E-02 m/z or 25 ppm, and a retention time tolerance of 0.15. The data was exported into a .CSV file (feature table) and a .mgf file for GNPS.

Molecular Network Generation

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

The molecular network used to identify the ceramide molecule is available here: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=15cfb993ddb24c4986f40c62510b9661.

Molecular Formula assignment and calculation of Nominal Oxidation State of Carbon

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

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

where a, b, c, d, e, and f are the numbers of C, H, N, O, P, and S atoms respectively in a given organic molecule and Z is net charge of the organic molecule. These NOSCs were then used to calculate the Gibbs Fee Energy of Carbon Oxidation (ΔG°Cox) of these compounds using the methods described in LaRowe and Van Cappellen (2011).

Statistical Analysis

All tests were conducted with an alpha of 0.05 (95% confidence level). A one-way analysis of variance (ANOVA) followed by a Tukey post hoc analysis were used to test for significant differences in viral abundance, bacterial abundance, bacterial size, bacterial metabolic power output, virus to microbe ratio, bacterial taxa, functional genes, NOSCs, $\Delta G^{o}Cox$, and ceramide abundance by treatments (i.e., an effect of coral, algae, interface). Data were further analyzed with linear regression comparing log (viral abundance) and virus to microbe ratio to log (bacterial abundance) and comparing cell division genes and total bacterial biomass to the Bacteroidetes to Firmicutes ratio. All the statistical analyses were performed using JMP 14

software (SAS Software). All dendrograms were produced using the Wards hierarchical clustering function in JMP 14. Two-way dendrogram heat maps, were produced by using the two-way clustering function in JMP 14.

All random forests analyses were performed in R using the 'rfPermute' in combination with the ‘randomForest’ package (Archer 2016). In supervised random forests, the competition outcome (i.e., winning or losing) of each sample was given and used for the learning process to identify winning or losing samples based on the metabolomic, taxonomic, and functional gene data.

Supervised random forests were done within a group of samples identified as either coral, algae, or interface. The variable importance plots (VIPs) from the random forests were used to identify the molecules, taxa, and functional genes that best distinguished winning and losing interactions within the three groups (i.e., coral, algae, interface). The top ten variables from each VIP were used to construct two-way dendrograms for distinguishing winning and losing interactions in each sample type in JMP 14.

Figures

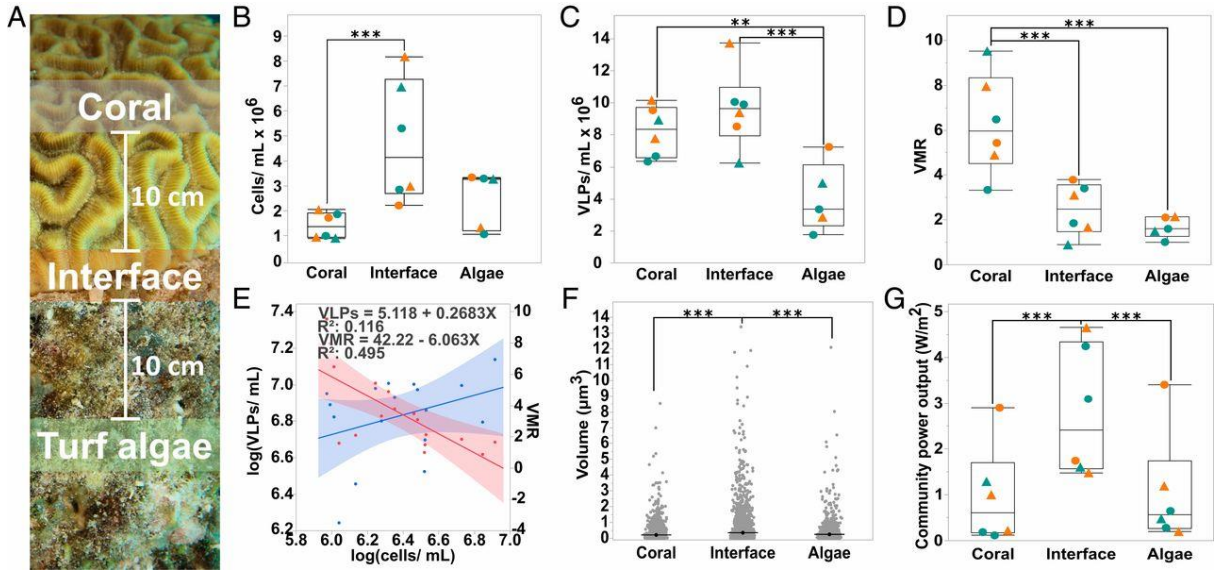


Figure 2.1 Analysis of coral-, interface-, and turf algal-associated virus and microbe communities by epifluorescence microscopy. (A) Sampling schematic. Separate surface-associated water samples for microbial and viral counts and tissue biopsies for multiomics were taken from the coral, the turf algae, and the interaction interface between the two macroorganisms. Interface samples were taken at the coral–turf algae interaction zone, and coral and turf algae water and tissue samples were taken 10 cm away from the interface. All samples were taken from coral–turf algal interactions at a 10–15 m depth. (B) Concentration of microbial cells per milliliter. (C) Concentration of VLPs per milliliter. (D) VMR. (E) Linear regression analysis of the concentration of VLPs (blue line) and VMR (red line) as a function of microbial cell concentration. (F) Microbial cell size. (G) The community power output in W/m2 as predicted by MTE. For B–D and G, triangles (▲) represent coral losing interactions, circles (•) represent coral winning interactions, orange symbols represent interactions with *D. strigosa*, and green symbols represent interactions with *O. faveolata* (n = 18, $^{**}p \leq 0.05$, $^{***}p \leq 0.01$)

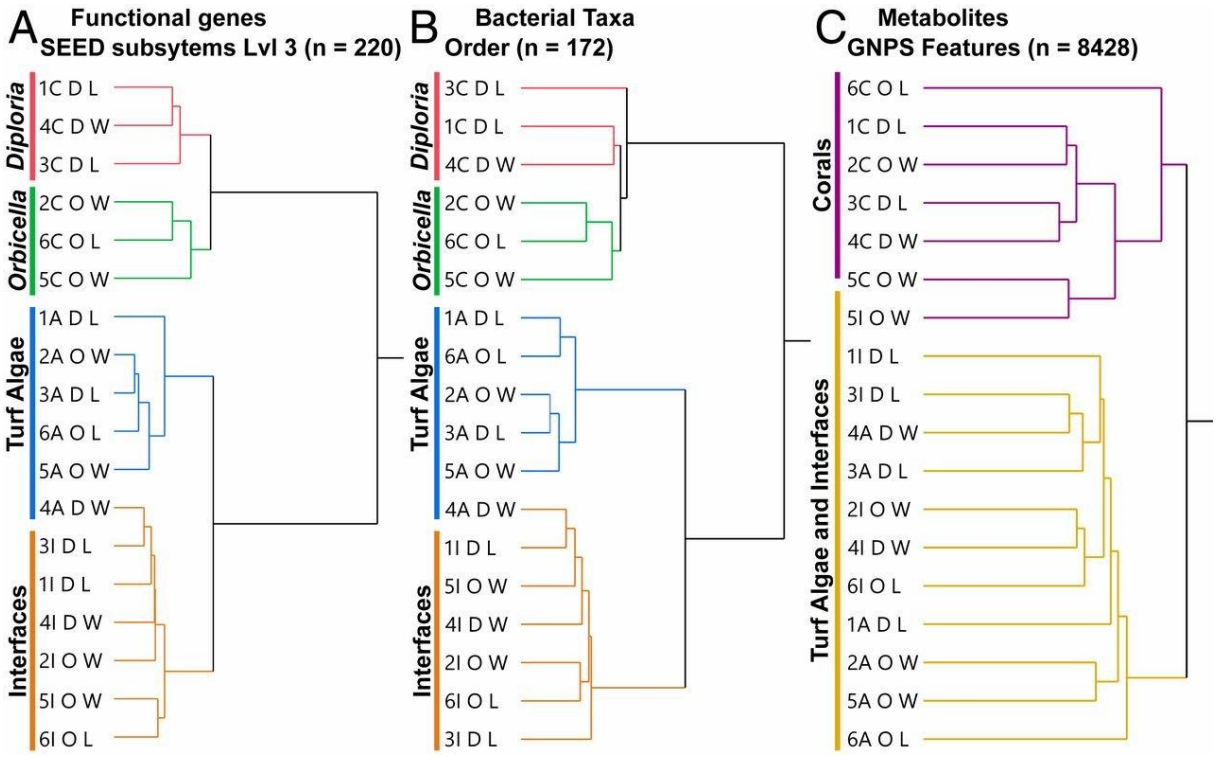


Figure 2.2 Hierarchical clustering of samples. Metagenomic and metabolomic data from coral, turf algae, and interface tissue samples were clustered using Ward's hierarchical clustering method based on the functional genes (A), bacterial taxa (B), and metabolites (C). The branch tips are labeled to describe the site number (1–6), the sample type (A, algae; C, coral; I, interface), the type of coral (D = *D. strigosa*, O = *O. faveolata*), and whether the coral in the interaction was winning (W) or losing (L).

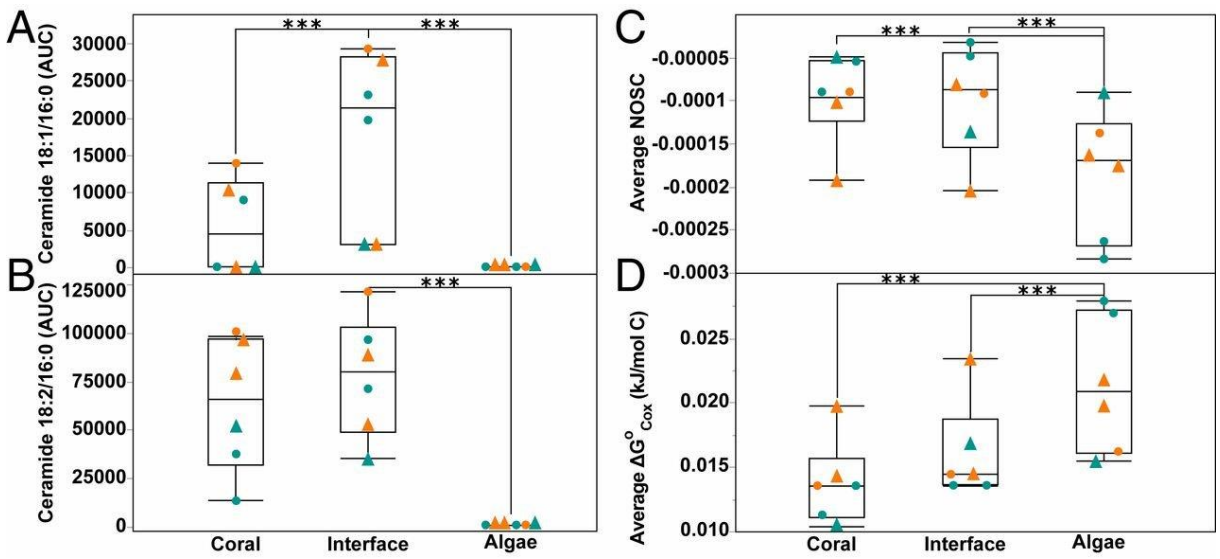


Figure 2.3 Metabolites across the coral-algal interface. The ceramide 18:1/16:0 abundance (A), the less saturated form, ceramide 18:2/16:0 (B) represented as the area under the curve (AUC). The average nominal oxidation state of carbon (NOSC) (C), and the average Gibbs free energy of carbon oxidation (ΔG°Cox) (D) (n = 18, ***p < 0.01). Triangles (▲) represent coral losing interactions and circles (•) represent coral winning interactions. Orange symbols represent interactions with *D. strigosa* and green symbols represent interactions with *O. faveolata*.

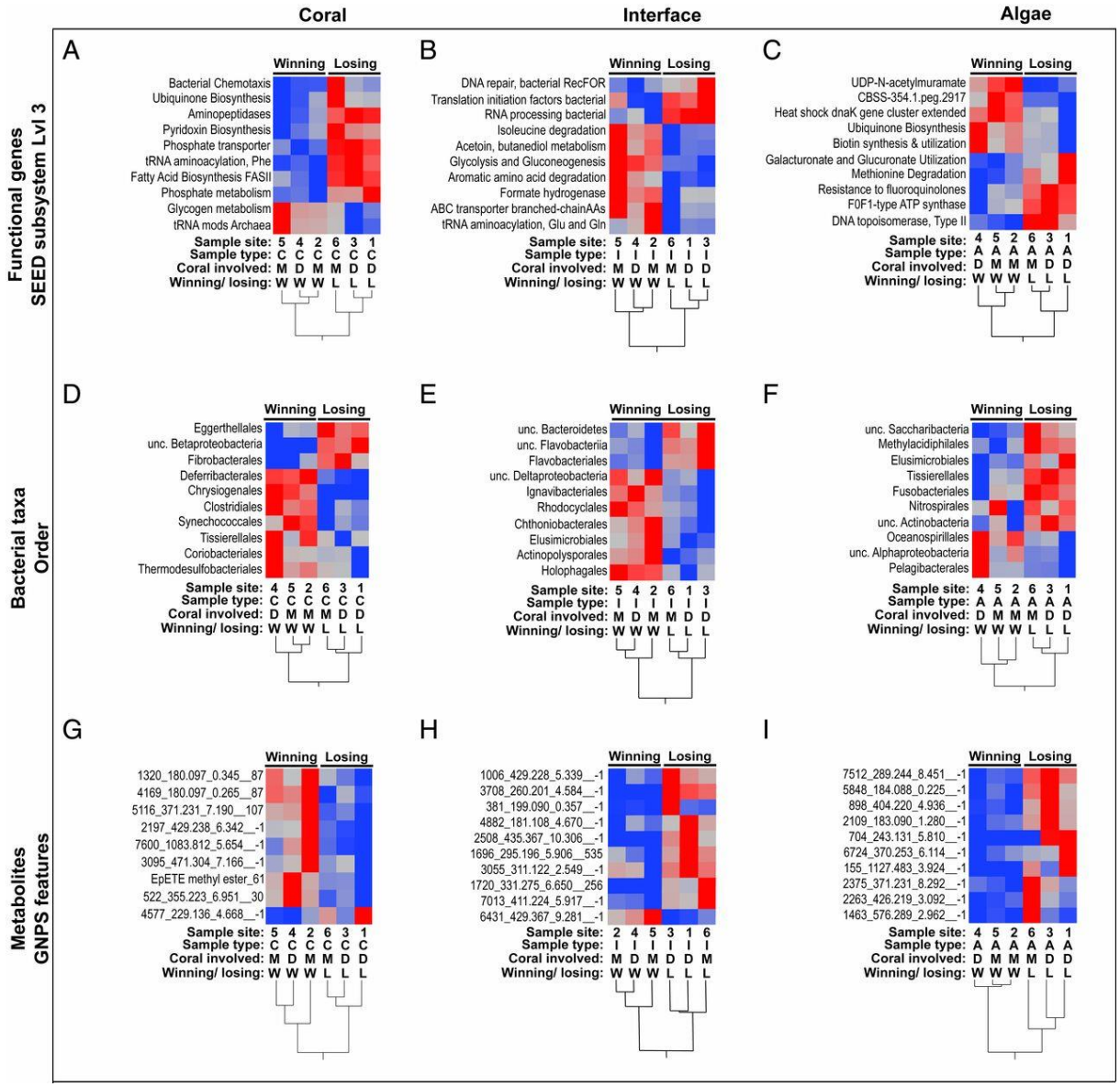


Figure 2.4 Two-way clustering of winning and losing interactions. Samples were clustered with Ward's hierarchical clustering method using the top random forest predictors for winning and losing corals (A, D, and G), interfaces (B, E, and H) and algae (C, F, and I) using functional genes (A–C), bacterial taxa (D–F), and metabolites (G–I) as inputs for the random forests. Abbreviations are as follows: A, algae; C, coral; D, *D. strigosa*; I, interface; L, losing; M, *O. faveolata*; W, winning. For metabolites, all the significant predictors of the coral–turf algal competition outcome were unannotated in the GNPS database. Thus, the first number is the GNPS cluster index, the following number is the mass-to-charge ratio, the third number is the retention time, and finally, the network subcluster ID where −1 indicates a single looped compound. For all heat maps, redder indicates relatively higher abundances and bluer indicates relatively lower abundances.

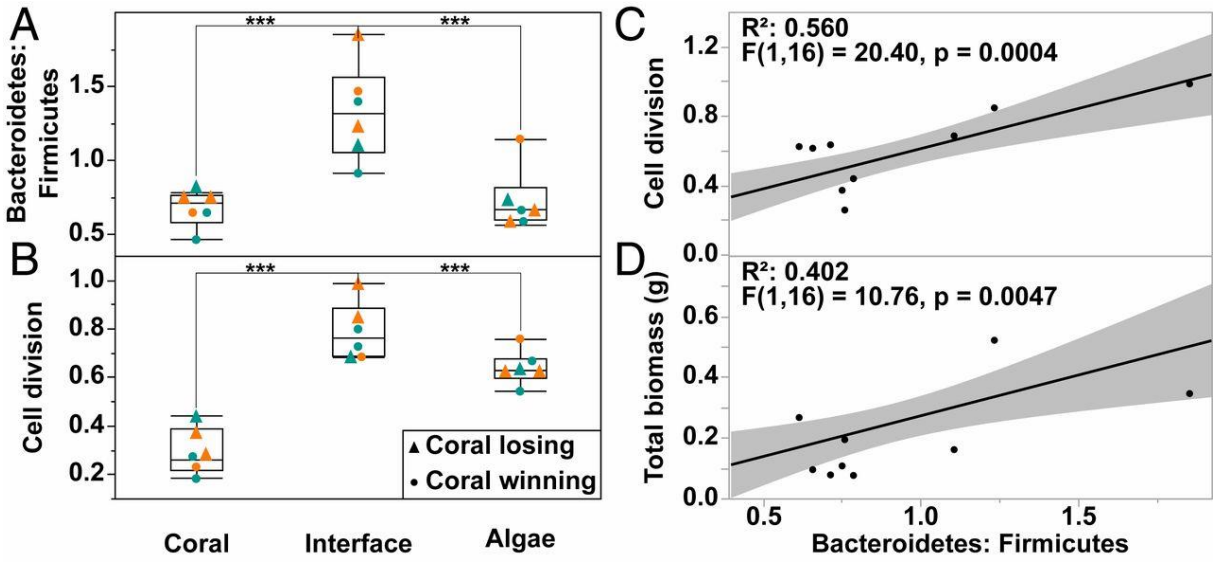


Figure 2.5 Results of ANOVA and subsequent regression analysis. Bacteroidetes-to-Firmicutes ratio (A), genes involved in cell division (B), regression analysis of the Bacteroidetes-to-Firmicutes ratio versus genes involved in cell division (C), and total microbial biomass in grams (D). $n = 18$. For A and B, triangles (▲) represent coral losing interactions and circles (●) represent coral winning interactions, and orange symbols represent interactions with *D. strigosa* and green symbols represent interactions with *O. faveolata* (***$p < 0.001$).

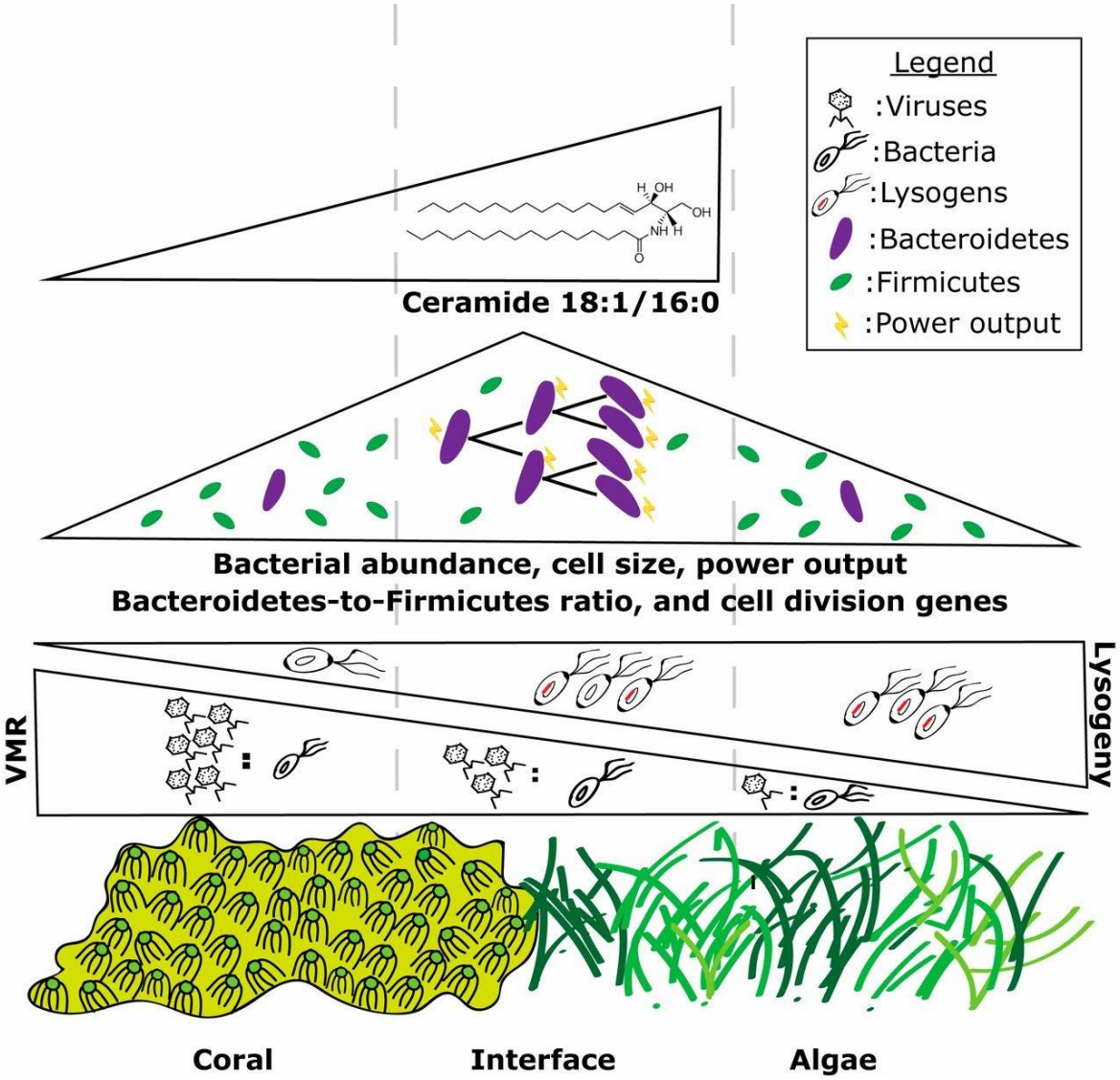
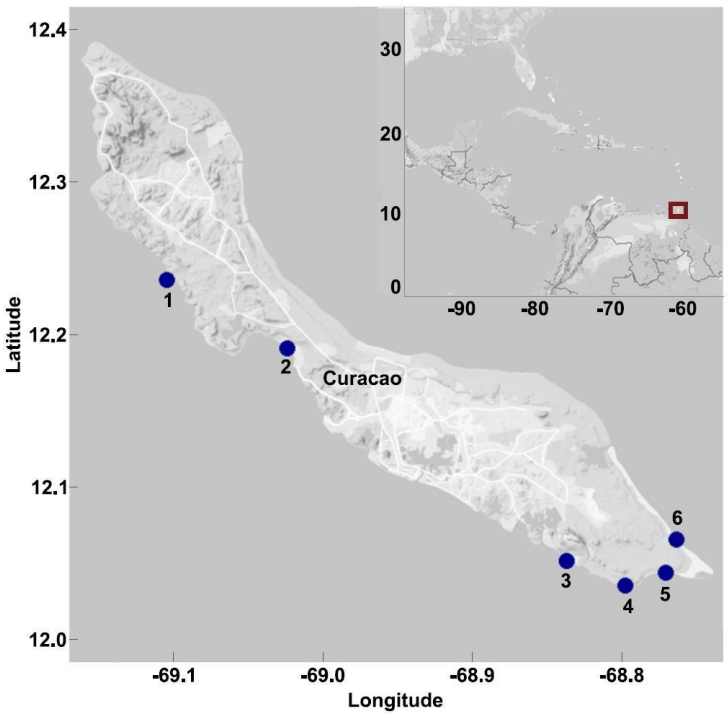
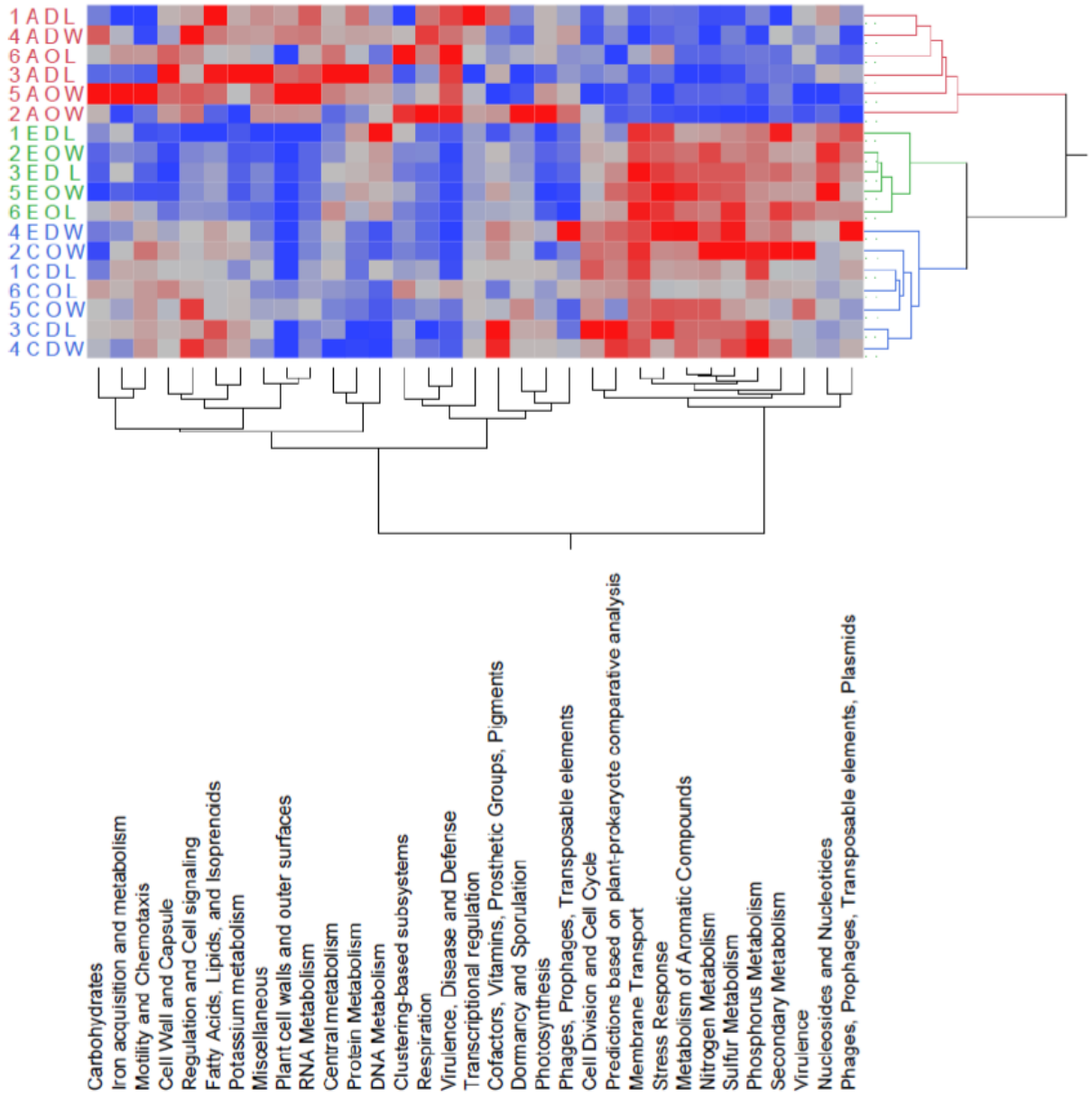


Figure 2.6 Working model of how different components of coral and turf algal holobionts mediate ecological interactions on the reef. Viruses are most abundant in the coral epibiont while bacteria are most abundant at the interface. This leads to the highest VMR in the coral holobiont and the lowest VMR in the turf algal holobiont. When coupled to the increased abundance of prophage found in the turf algal metagenomes this suggests increased lysogeny in the turf algal holobiont and increased lytic activity in the coral holobiont. Bacterial cell size, microbial metabolic power output, bacterial cell division, and the Bacteroidetes-to-Firmicutes ratio is highest at the interface. The potentially proapoptotic metabolite ceramide 18:1/16:0 is found only in the coral and interface samples but is most abundant at the interface.

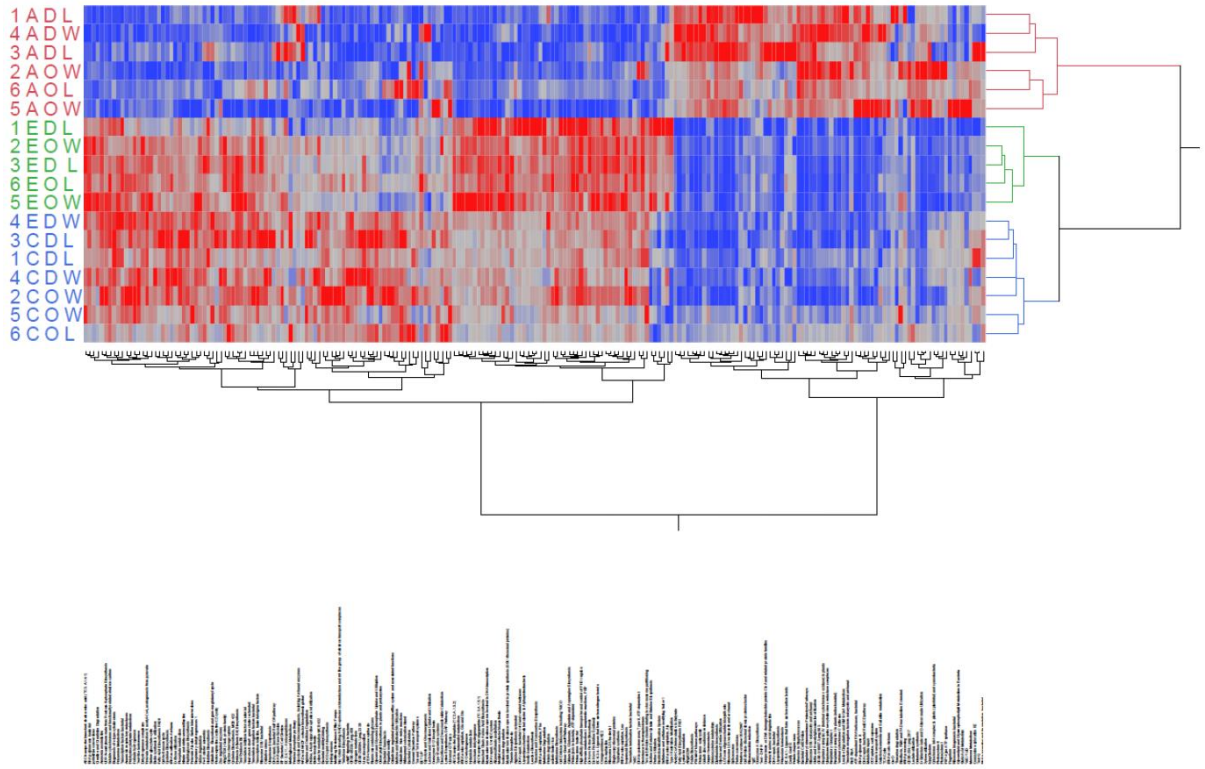
Supplementary Figures and Tables



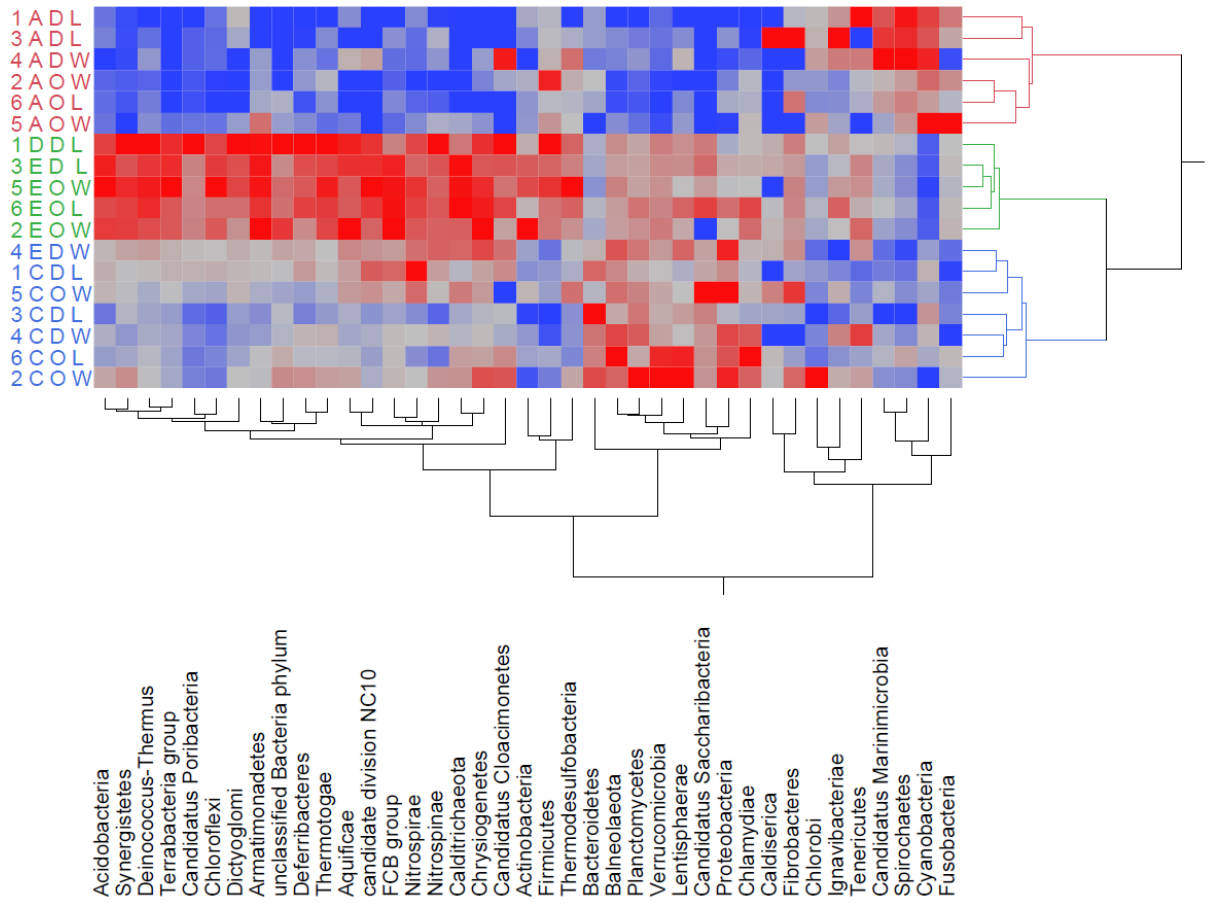
Supplemental Figure 2.1 Map of sampling sites around the island of Curaçao. Surface-associated samples were taken at each site by suction from the surface of the coral, the algae, and the interface over a 10 cm transect parallel with the interface. Tissue samples were taken at each site using an underwater power drill taking a biopsy 1 cm in diameter and 1 cm in depth. All samples were taken from coral-algal interactions at a 10-15 m depth.



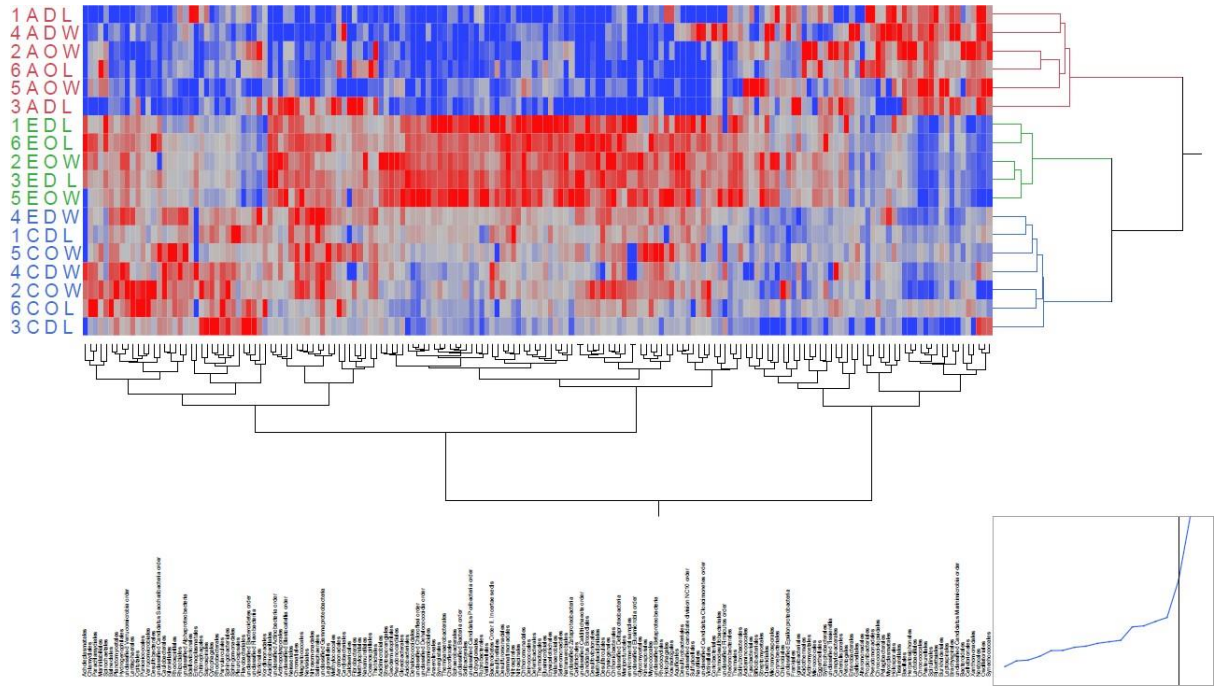
Supplemental Figure 2.2 Two-way heat map constructed using functional genes at level 1 of the SEED hierarchical database. Green, red, and blue branches represent the three significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.



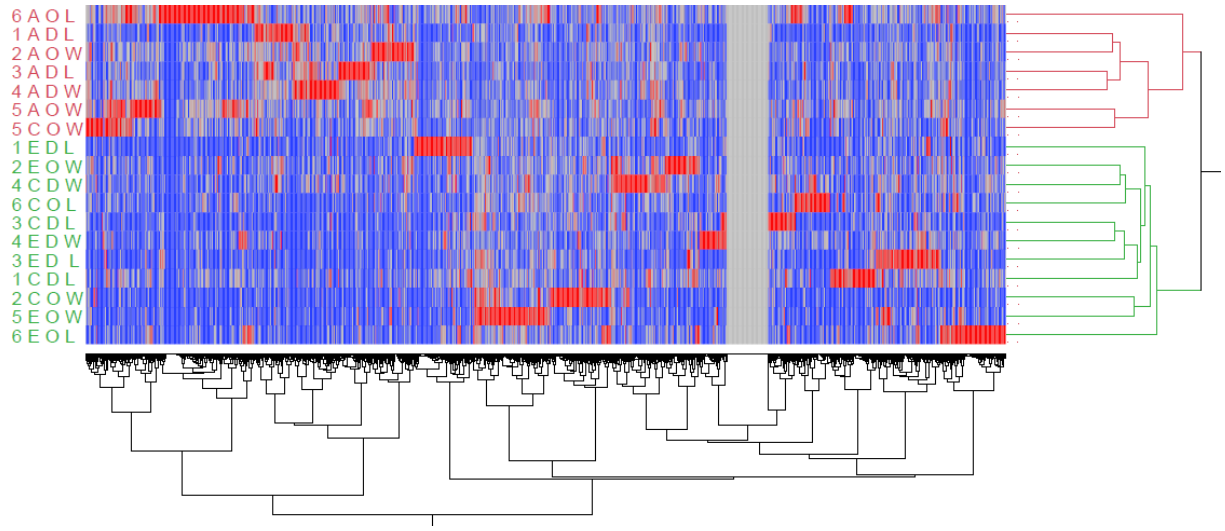
Supplemental Figure 2.3 Two-way heat map constructed using functional genes at level 3 of the SEED hierarchical database. Green, red, and blue branches represent the three significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.



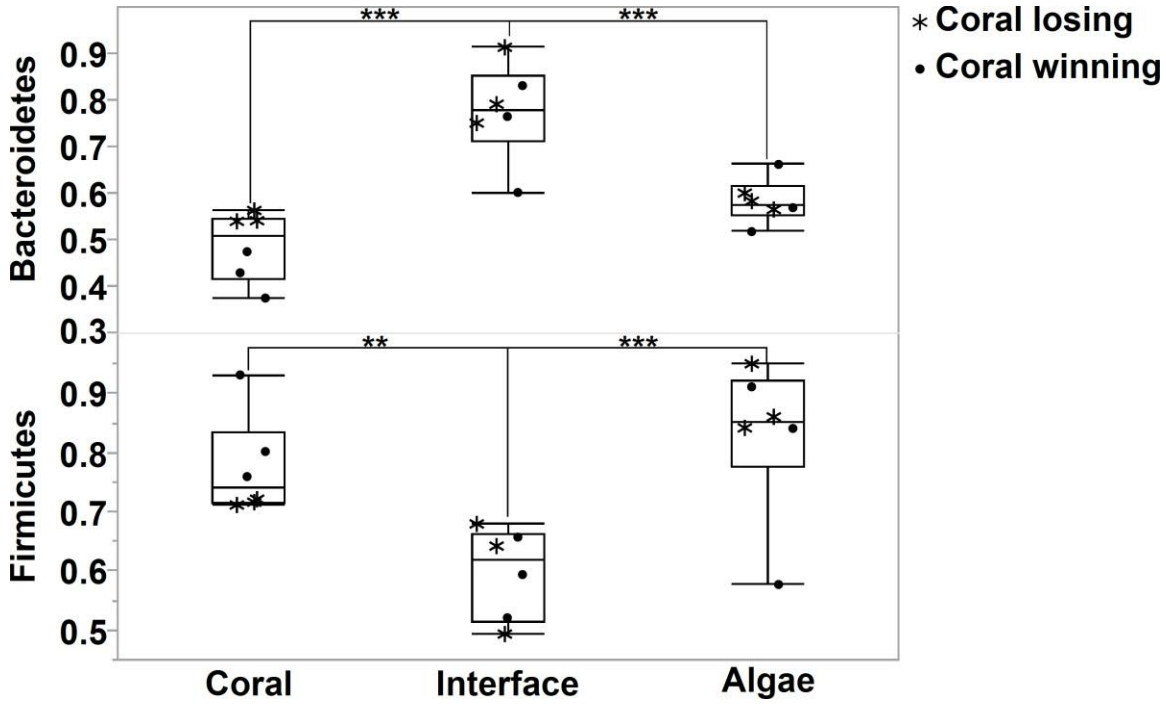
Supplemental Figure 2.4 Two-way heat map constructed using bacterial phyla. Green, red, and blue represent branches the three significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.



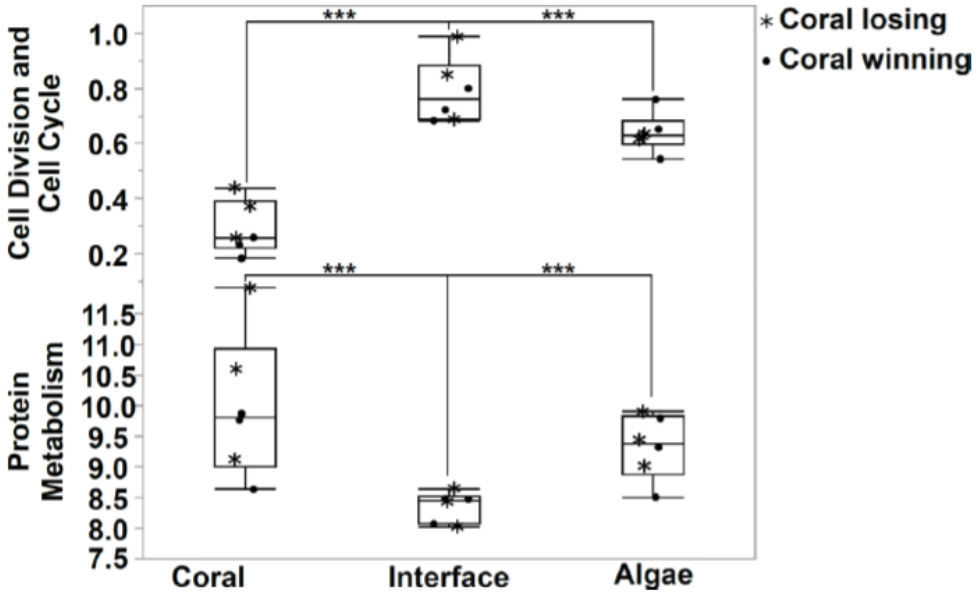
Supplemental Figure 2.5 Two-way heat map constructed using bacterial orders. Green, red, and blue branches represent the three significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.



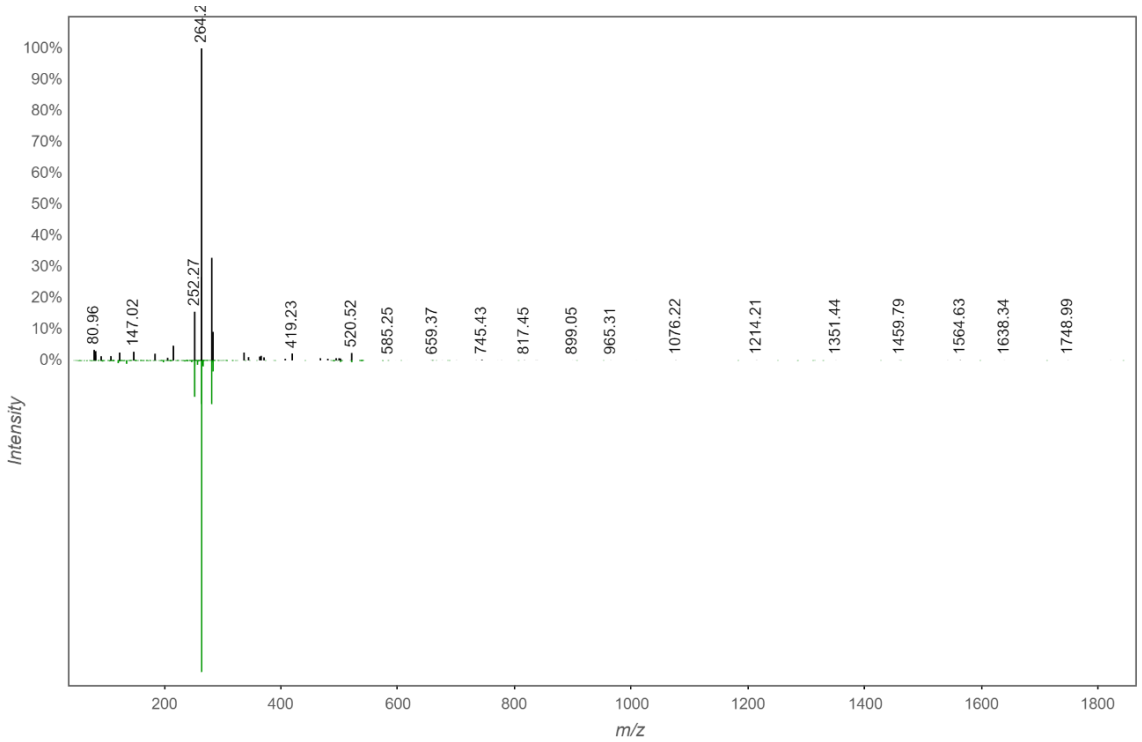
Supplemental Figure 2.6 Two-way heat map constructed using metabolites. Green and red branches represent the two significant clusters. The branch tips are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L). Redder indicates relatively higher abundances and bluer indicates relatively lower abundances.



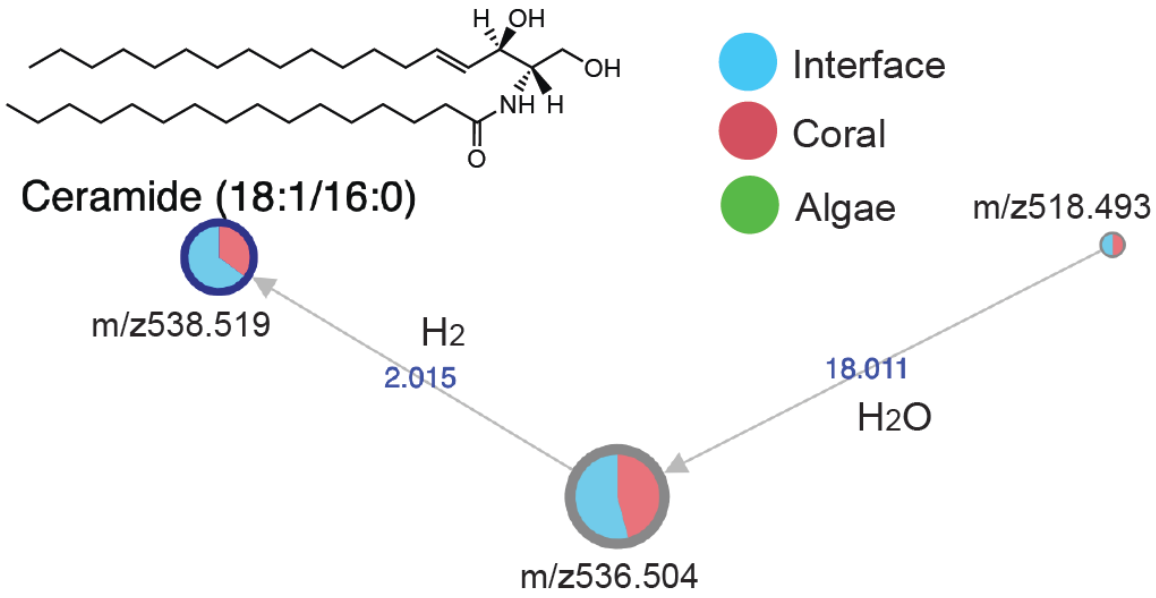
Supplemental Figure 2.7 Box plots of the percent relative abundance of the two phyla where the interface samples were significantly different than both the coral and the algal samples. (n = 18, Tukey post hoc **$p \leq 0.05$, ***$p \leq 0.01$)



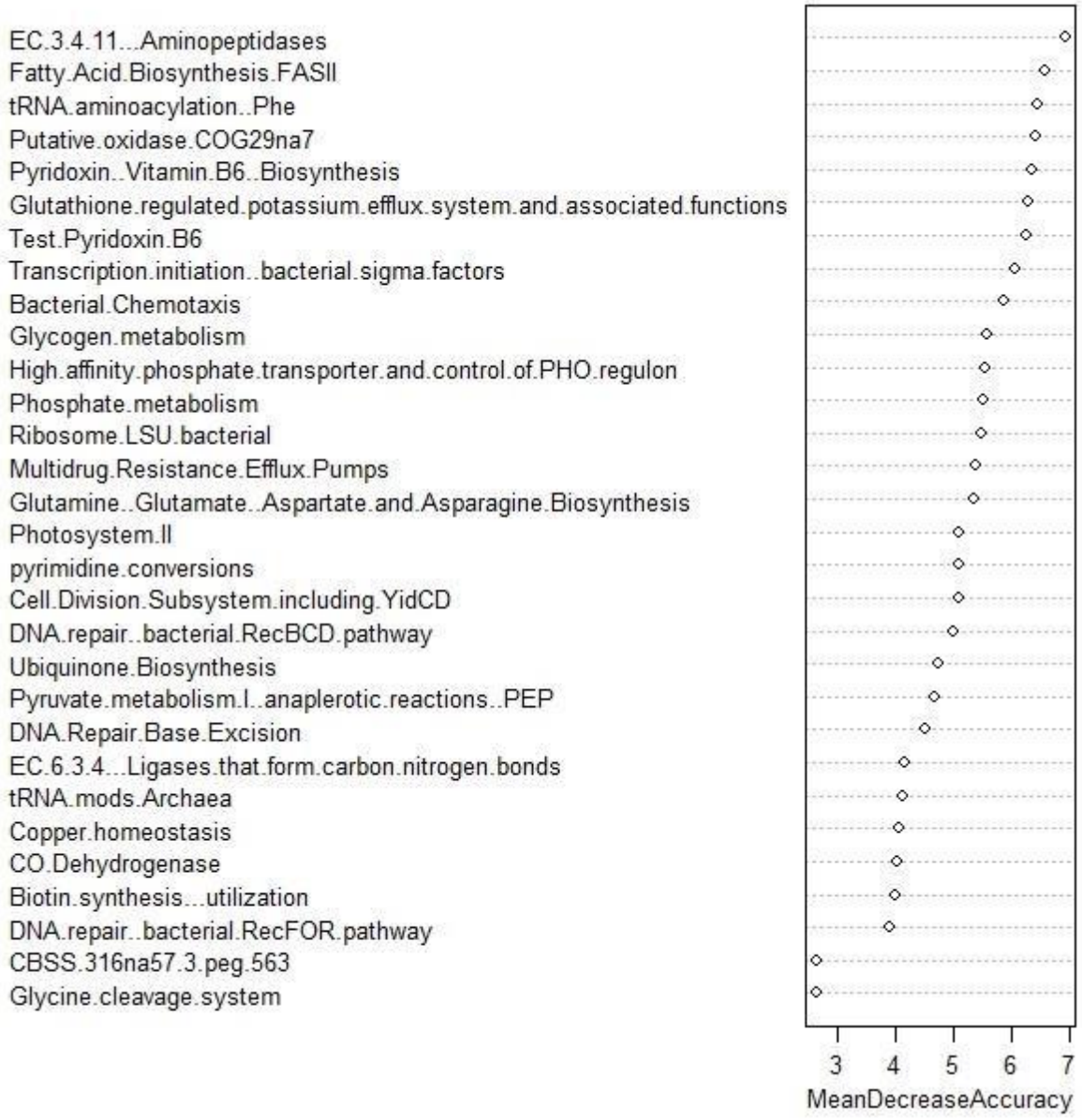
Supplemental Figure 2.8 Box plots of the percent relative abundance of the two level 1 SEED subsystems where the interface samples were significantly different than both the coral and the algal samples. (n = 18, Tukey post hoc *$p \leq 0.1$, **$p \leq 0.05$, ***$p \leq 0.01$)



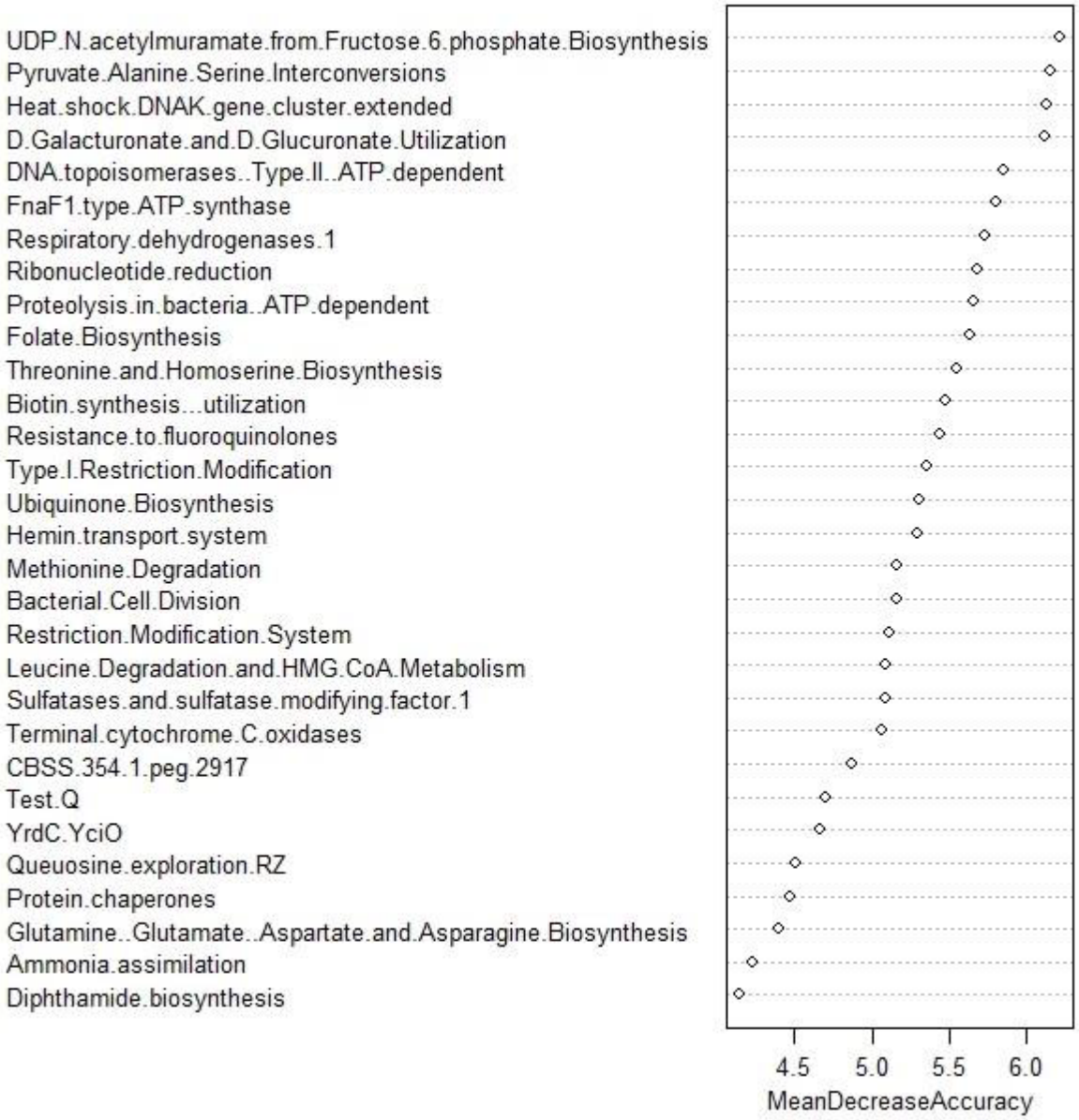
Supplemental Figure 2.9 Mirror plot of spectral match for ceramide 18:1/16:0. The GNPS library reference is on the bottom in green with the spectra for the compound in the coral and interface samples is shown above in black. Note, the high intensity peak at 264.2 is decisive for the 18:0 backbone.



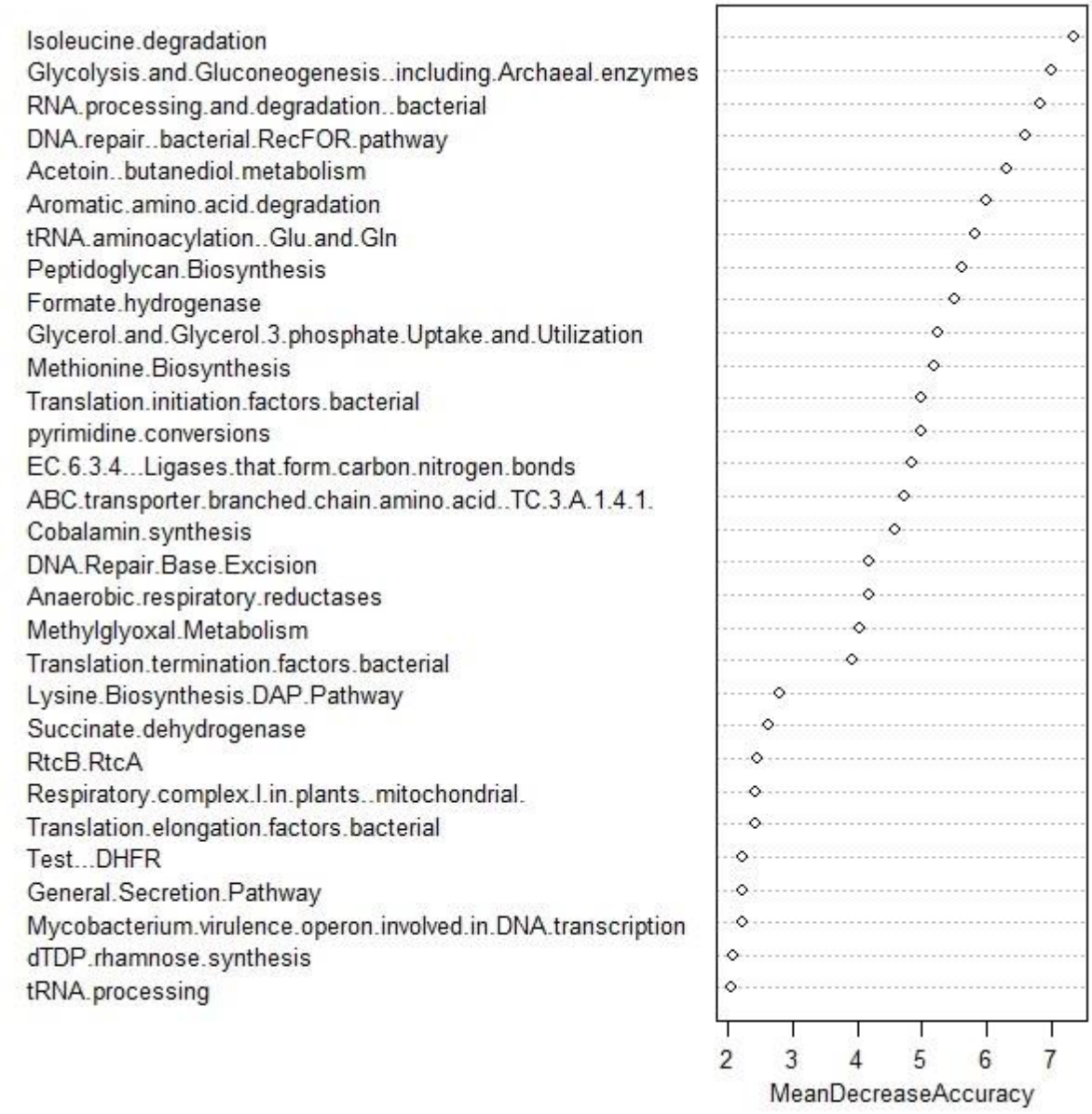
Supplemental Figure 2.10 Molecular network of ceramides in coral, algal, and interface metabolomics data from GNPS. The structure of known ceramide is shown along with its less saturated form. Edges are labeled by the mass difference between related nodes and known biochemical transformations are highlighted. The nodes are colored in a pie chart based on the total spectral intensity between coral, algae, and interface according to the color legend.



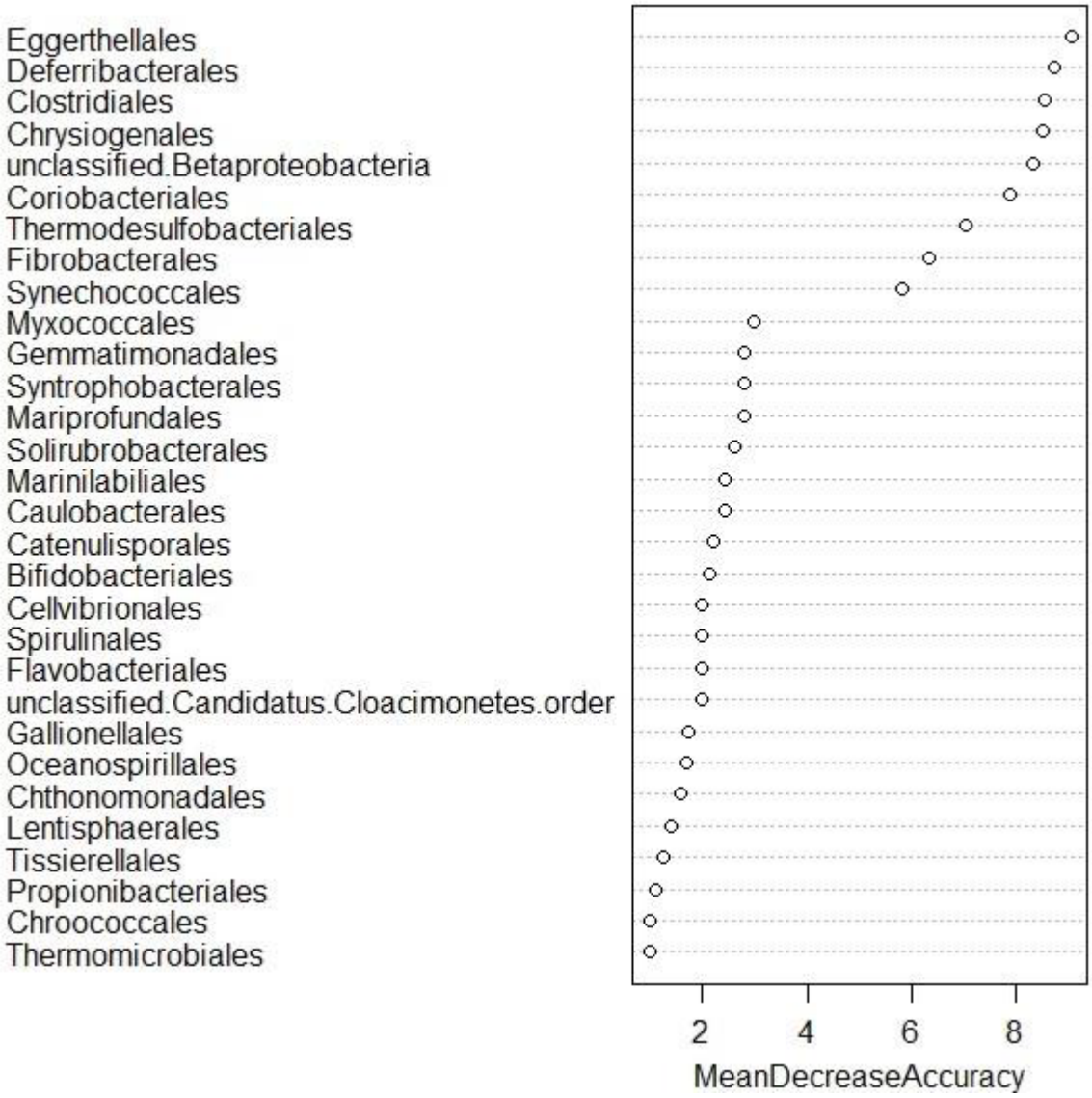
Supplemental Figure 2.11 Variable importance plot of functional genes (SEED level 3) from random forest classification analysis based on winning and losing corals. Variable are ranked from highest to lowest according to their mean decrease in accuracy.



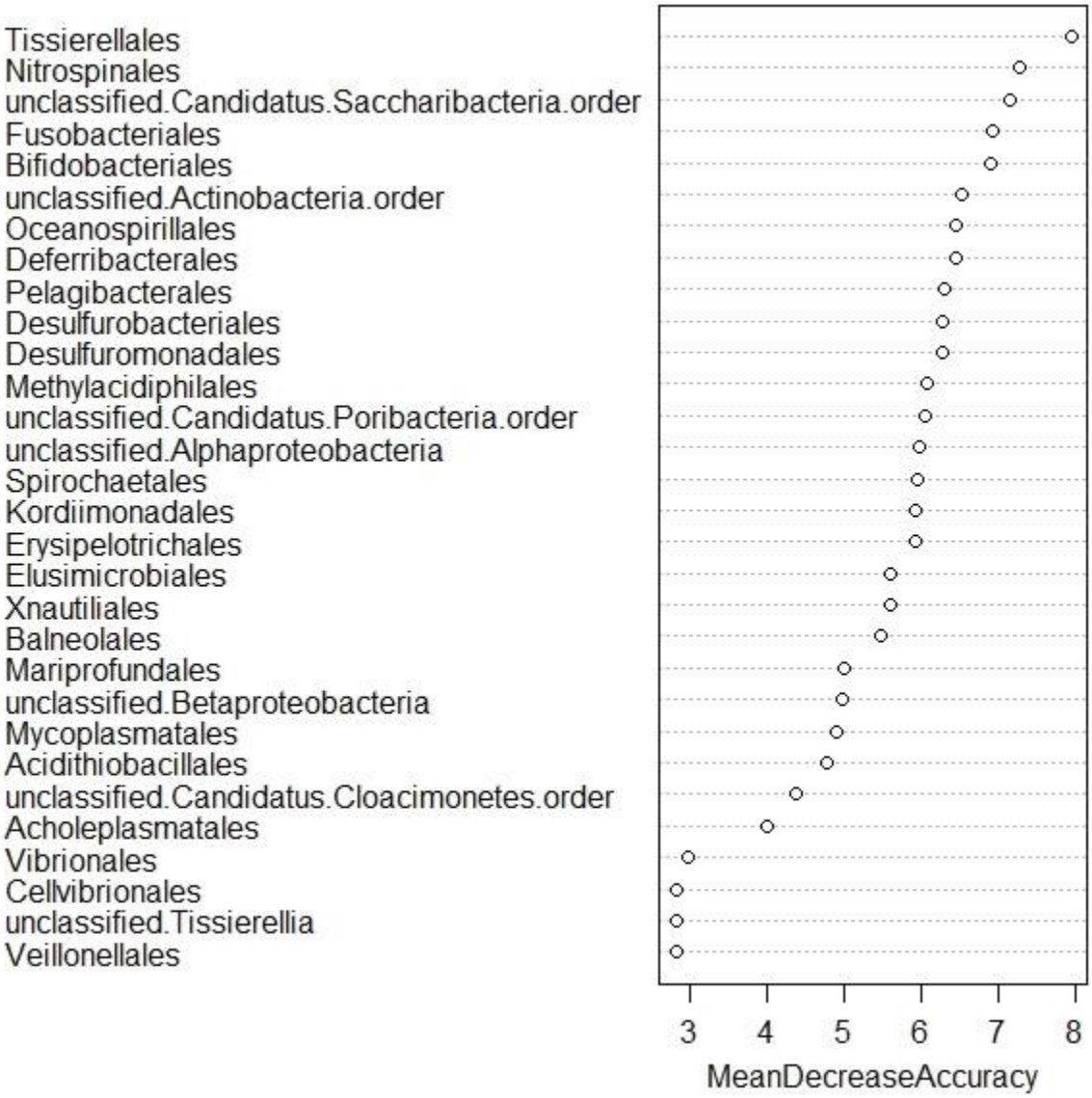
Supplemental Figure 2.12 Variable importance plot of functional genes (SEED level 3) from random forest classification analysis based on winning and losing algae. Variable are ranked from highest to lowest according to their mean decrease in accuracy.



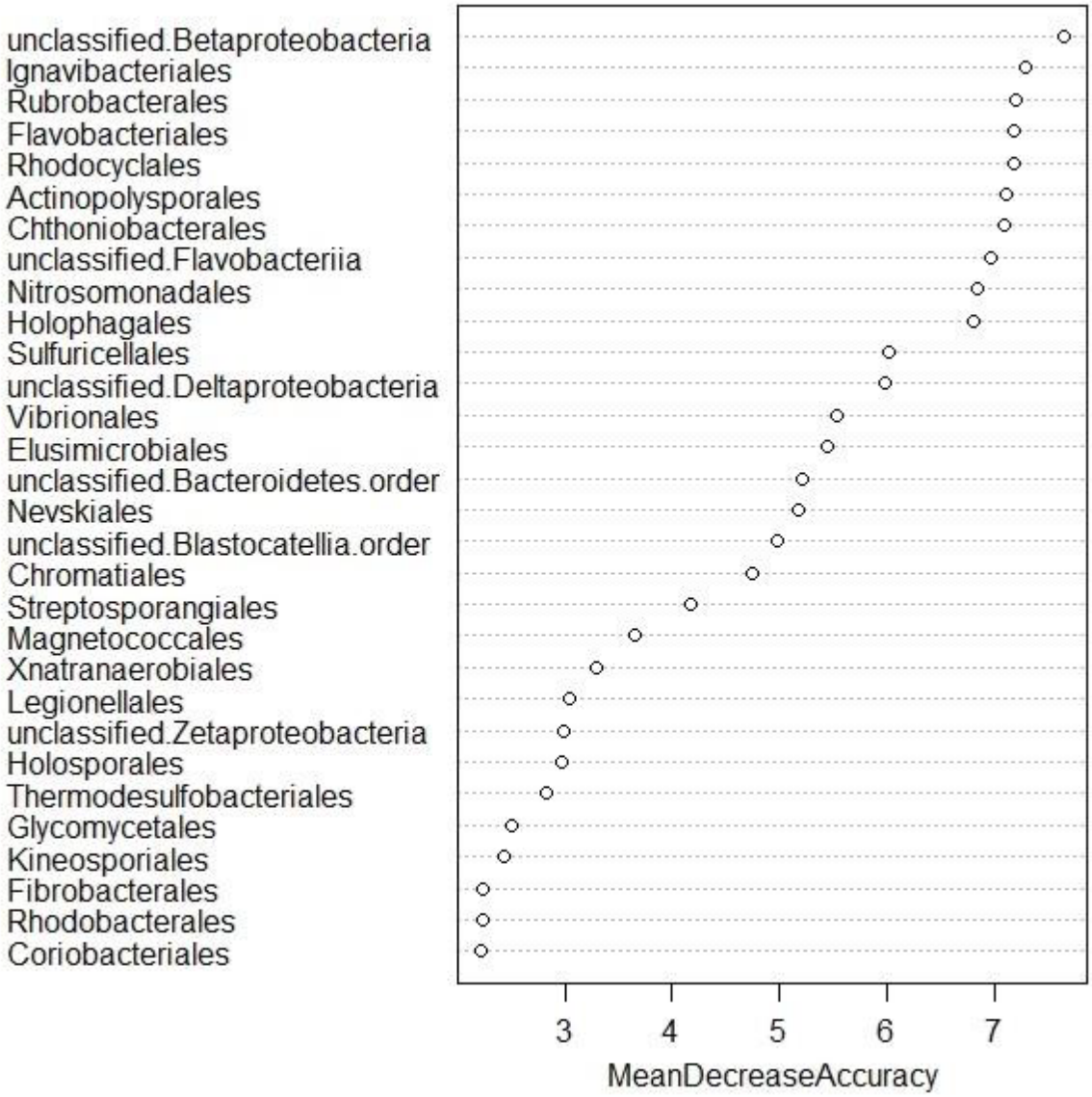
Supplemental Figure 2.13 Variable importance plot of functional genes (SEED level 3) from random forest classification analysis based on winning and losing interfaces. Variable are ranked from highest to lowest according to their mean decrease in accuracy.



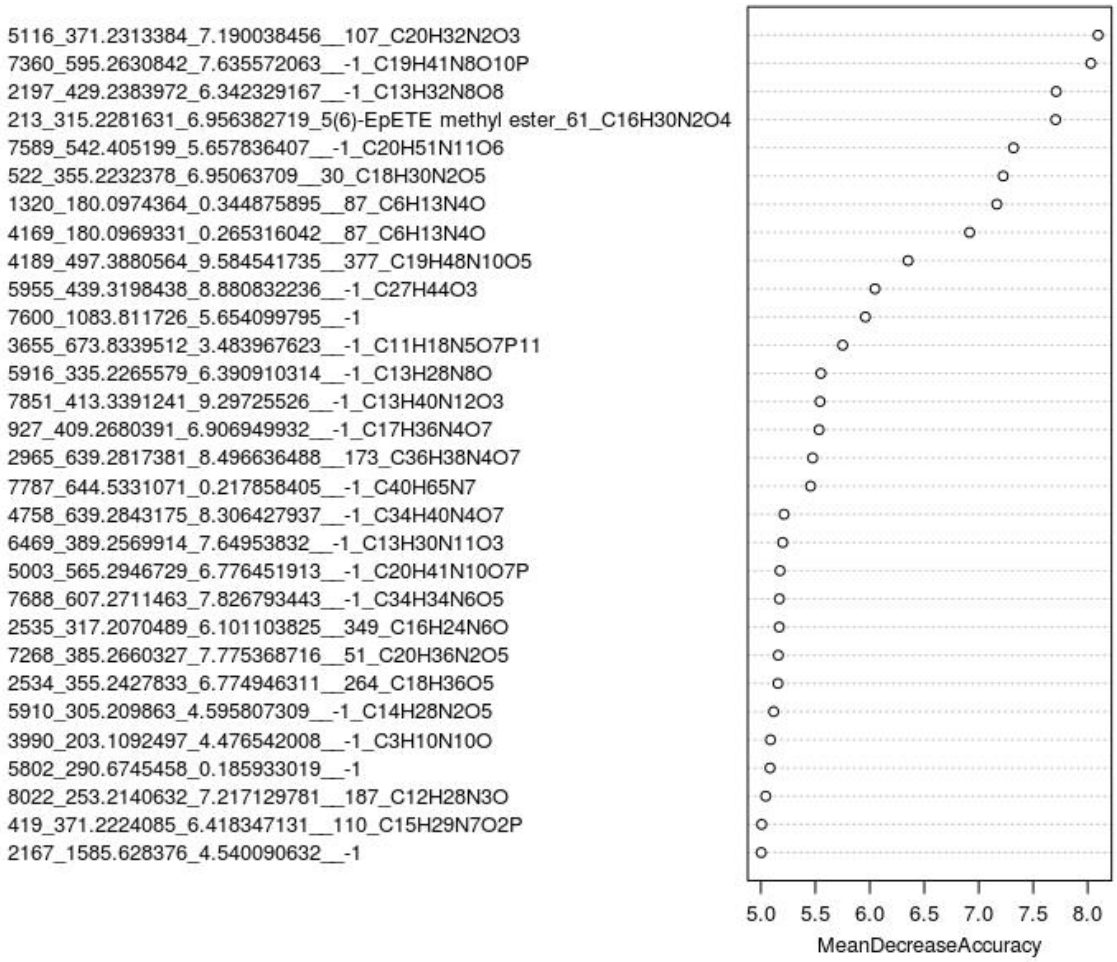
Supplemental Figure 2.14 Variable importance plot of bacterial orders from random forest classification analysis based on winning and losing corals. Variable are ranked from highest to lowest according to their mean decrease in accuracy.



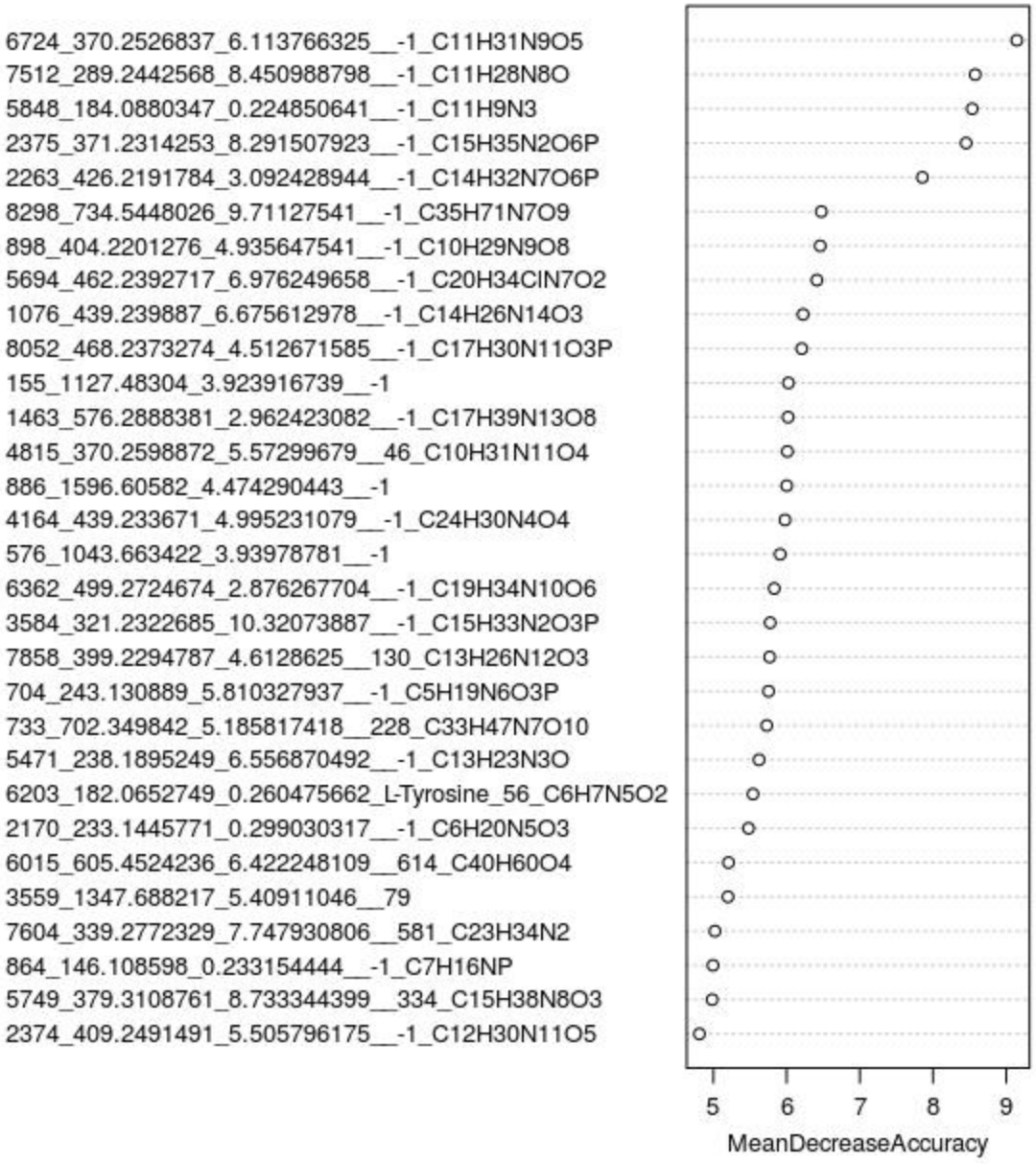
Supplemental Figure 2.15 Variable importance plot of bacterial orders from random forest classification analysis based on winning and losing algae. Variable are ranked from highest to lowest according to their mean decrease in accuracy.



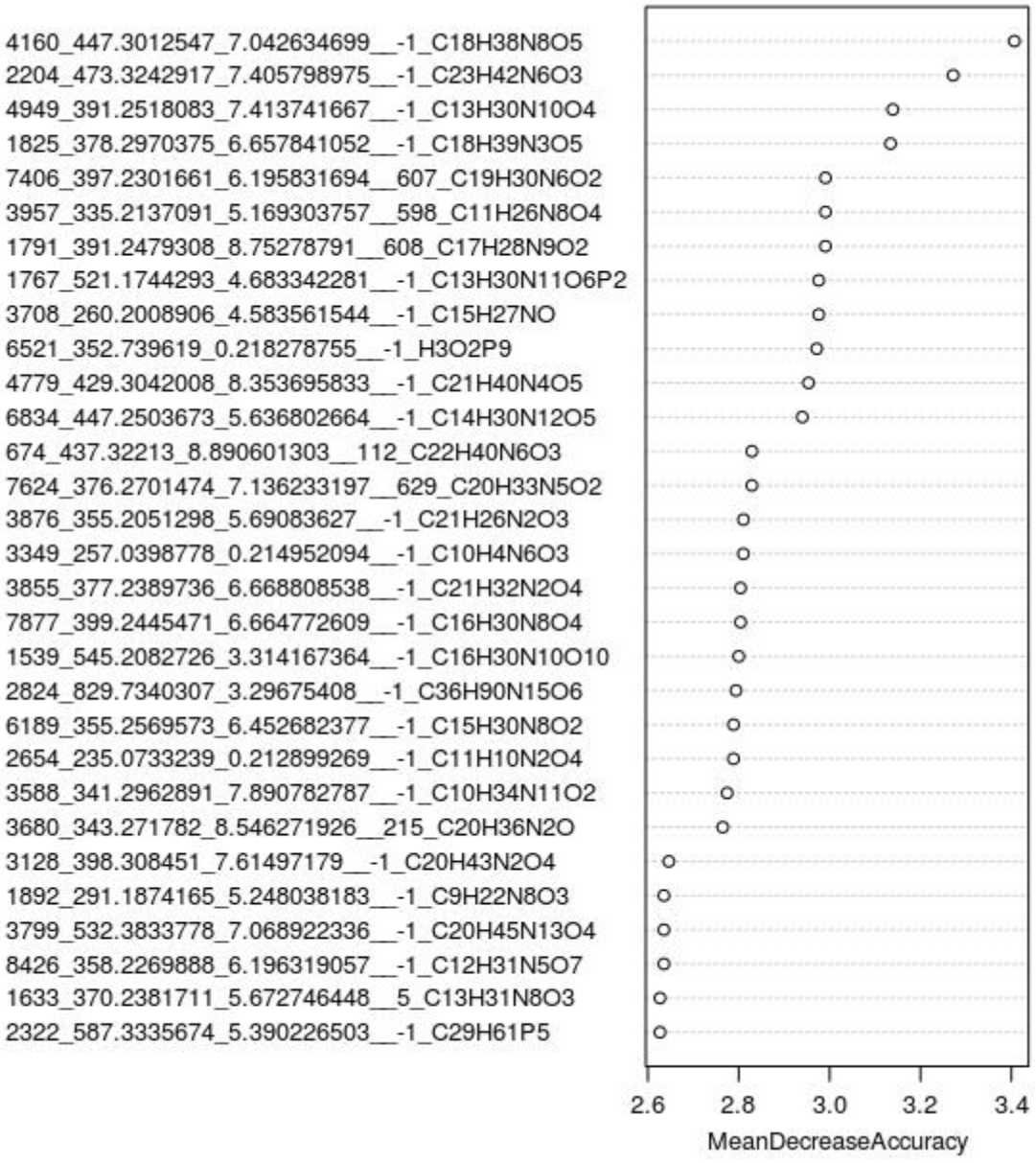
Supplemental Figure 2.16 Variable importance plot of bacterial orders from random forest classification analysis based on winning and losing interfaces. Variable are ranked from highest to lowest according to their mean decrease in accuracy.



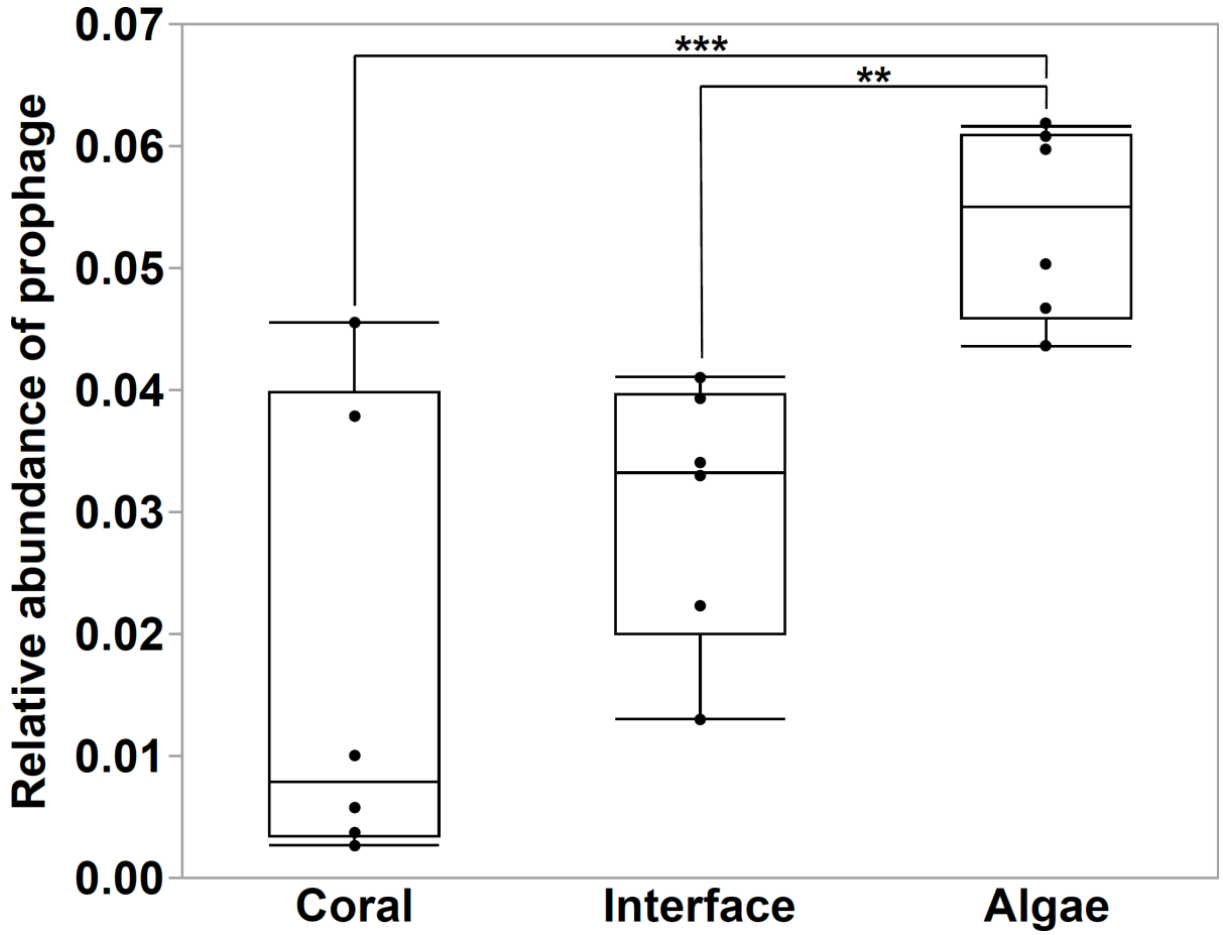
Supplemental Figure 2.17 Variable importance plot of metabolites from random forest classification analysis based on winning and losing corals. Variable are ranked from highest to lowest according to their mean decrease in accuracy. The first number is the GNPS annotation number, the following number is the mass to charge ratio, the third number is the retention time, the fourth number is the network subcluster ID where -1 indicates a single looped compound. This is followed by the putative molecular formula for all molecules where annotation was possible.



Supplemental Figure 2.18 Variable importance plot of metabolites from random forest classification analysis based on winning and losing algae. Variable are ranked from highest to lowest according to their mean decrease in accuracy. The first number is the GNPS annotation number, the following number is the mass to charge ratio, the third number is the retention time, the fourth number is the network subcluster ID where -1 indicates a single looped compound. This is followed by the putative molecular formula for all molecules where annotation was possible.



Supplemental Figure 2.19 Variable importance plot of metabolites from random forest classification analysis based on winning and losing interfaces. Variable are ranked from highest to lowest according to their mean decrease in accuracy. The first number is the GNPS annotation number, the following number is the mass to charge ratio, the third number is the retention time, the fourth number is the network subcluster ID where -1 indicates a single looped compound. This is followed by the putative molecular formula for all molecules where annotation was possible.



Supplemental Figure 2.20 Relative abundance of prophages in metagenomes. (n = 18, **$p \leq$ 0.05, ***$p \leq 0.01$)

Supplemental Table 2.1 Sample metadata.

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

Supplemental Table 2.2 Site-level water chemistry data. Site-level data including phosphate, nitrite, nitrite + nitrate, ammonium, and dissolved organic carbon (DOC) concentrations reported in micromolar (µM). Note, there were no significant differences between coral species, or between competition outcomes for any of the site level variables recorded.

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

Supplemental Table 2.3 Metagenomic library details. Just as in the main text, sample names are labeled to describe the site number (1-6), the sample type (C: coral, I: interface, A: algae), the type of coral (D = *D. strigosa*, O = *O. faveolata*, and whether the coral in the interaction was winning (W) or losing (L).

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

Supplemental Table 2.4 Microbial taxa that are significantly enriched in winning and losing interactions. Prokaryotic orders of bacteria that were significantly enriched as determined ANOVA are listed in the first column with the ANOVA *p*-value listed in the second column.

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

Supplemental Table 2.5 Functional genes that are significantly enriched in winning and losing interactions. Functional genes from Level 3 of the SEED database thawere significantly enriched as determined by ANOVA are listed in the first column with the ANOVA *p*-value listed in the second column.

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

Acknowledgements

Chapter 2, in full, is published in the Proceedings of the National Academy of Sciences (PNAS), 2020. Ty N.F. Roach*, Mark Little, Milou G.I. Arts, Joel Huckeba, Andreas F. Haas, Emma E. George, Robert A. Quinn, Ana G. Cobian-Guemes, Douglas S. Naliboff, Cynthia B. Silveira, Mark J.A. Vermeij, Linda Wegley Kelly, Pieter C. Dorrestein, and Forest Rohwer. The dissertation author was the co-primary investigator and co-first author of this paper.

I thank the Waitt Institute and the crew of the Plan B for their assistance in collecting samples. We thank the National Science Foundation for Grants G00009988 (to T.N.F.R.), OCE-1538567, IOS-1848671 (to L.W.K.), and OISE-1243541 (to F.L.R.). I thank the Undead Group—Sandi Calhoun, Kevin Green, Sean Benler, Nate Robinett, and Greg Peters—for their reading of the manuscript and useful feedback during the writing process.

References

Aziz, R. K., Edwards, R. A., Taylor, W. W., Low, D. E., McGeer, A., & Kotb, M. (2005). Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated M1T1 clone of Streptococcus pyogenes. *Journal of Bacteriology*, *187*(10), 3311–3318.

Barott, K. L., & Rohwer, F. L. (2012). Unseen players shape benthic competition on coral reefs. *Trends in Microbiology*, *20*(12), 621–628.

Barott, K. L., Williams, G. J., Vermeij, M. J. A., Harris, J., Smith, J. E., Rohwer, F. L., & Sandin, S. A. (2012). Natural history of coral- algae competition across a gradient of human activity in the Line Islands. *Marine Ecology Progress Series*, *460*, 1–12.

Barott, K., Smith, J., Dinsdale, E., Hatay, M., Sandin, S., & Rohwer, F. (2009). Hyperspectral and physiological analyses of coral-algal interactions. *PloS One*, *4*(11), e8043.

Barr, J. J., Auro, R., Furlan, M., Whiteson, K. L., Erb, M. L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A. S., Doran, K. S., Salamon, P., Youle, M., & Rohwer, F. (2013). Bacteriophage adhering to mucus provide a non–host-derived immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(26), 10771–10776.

Beman, J. M., Roberts, K. J., Wegley, L., Rohwer, F., & Francis, C. A. (2007). Distribution and diversity of archaeal ammonia monooxygenase genes associated with corals. *Applied and Environmental Microbiology*, *73*(17), 5642–5647.

Bourne, D. G., Garren, M., Work, T. M., Rosenberg, E., Smith, G. W., & Harvell, C. D. (2009). Microbial disease and the coral holobiont. *Trends in Microbiology*, *17*(12), 554–562.

Bourne, D. G., Morrow, K. M., & Webster, N. S. (2016). Insights into the Coral Microbiome: Underpinning the Health and Resilience of Reef Ecosystems. *Annual Review of Microbiology*, *70*, 317–340.

Bourne, D., Iida, Y., Uthicke, S., & Smith-Keune, C. (2008). Changes in coral-associated microbial communities during a bleaching event. *The ISME Journal*, *2*(4), 350–363.
Brüssow, H. (2007). Bacteria between protists and phages: from antipredation strategies to the evolution of pathogenicity. *Molecular Microbiology*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2958.2007.05826.x

Brüssow, H., Canchaya, C., & Hardt, W.-D. (2004). Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiology and Molecular Biology Reviews: MMBR*, *68*(3), 560–602, table of contents.

Coutinho, F. H., Silveira, C. B., Gregoracci, G. B., Thompson, C. C., Edwards, R. A., Brussaard, C. P. D., Dutilh, B. E., & Thompson, F. L. (2017). Marine viruses discovered via metagenomics shed light on viral strategies throughout the oceans. *Nature Communications*, *8*, 15955.

Ferrier-Pagès, C., Godinot, C., D'angelo, C., Wiedenmann, J., & Grover, R. (2016). Phosphorus metabolism of reef organisms with algal symbionts. *Ecological Monographs*, *86*(3), 262–277.

Fiore, C. L., Jarett, J. K., Olson, N. D., & Lesser, M. P. (2010). Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends in Microbiology*, *18*(10), 455–463.

Frias-Lopez, J., Zerkle, A. L., Bonheyo, G. T., & Fouke, B. W. (2002). Partitioning of bacterial communities between seawater and healthy, black band diseased, and dead coral surfaces. *Applied and Environmental Microbiology*, *68*(5), 2214–2228.

Galtier d'Auriac, I., Quinn, R. A., Maughan, H., Nothias, L.-F., Little, M., Kapono, C. A., Cobian, A., Reyes, B. T., Green, K., Quistad, S. D., Leray, M., Smith, J. E., Dorrestein, P. C., Rohwer, F., Deheyn, D. D., & Hartmann, A. C. (2018). Before platelets: the production of platelet-activating factor during growth and stress in a basal marine organism. *Proceedings. Biological Sciences / The Royal Society*, *285*(1884). https://doi.org/10.1098/rspb.2018.1307

Garg, N., Wang, M., Hyde, E., da Silva, R. R., Melnik, A. V., Protsyuk, I., Bouslimani, A., Lim, Y. W., Wong, R., Humphrey, G., Ackermann, G., Spivey, T., Brouha, S. S., Bandeira, N., Lin, G. Y., Rohwer, F., Conrad, D. J., Alexandrov, T., Knight, R., & Dorrestein, P. C. (2017). Three-Dimensional Microbiome and Metabolome Cartography of a Diseased Human Lung. *Cell Host & Microbe*, *22*(5), 705–716.e4.

Gregg, A., Hatay, M., Haas, A., Robinett, N., Barott, K., Vermeij, M., Marhaver, K., Meirelles, P., Thompson, F., & Rohwer, F. (2013). Biological oxygen demand optode analysis of coral reef-associated microbial communities exposed to algal exudates. *PeerJ*, *1*, e107.

H. Jorissen, C. Skinner, R. Osinga, D. de Beer, M. M. Nugues, Evidence for water mediated mechanisms in coral-algal interactions. Proc. Biol. Sci. 283, 20161137 (2016).

Haas, A. F., & Wild, C. (2010). Composition analysis of organic matter released by cosmopolitan coral reef-associated green algae. *Aquatic Biology*. https://www.int-res.com/abstracts/ab/v10/n2/p131-138/

Haas, A. F., Gregg, A. K., Smith, J. E., Abieri, M. L., Hatay, M., & Rohwer, F. (2013). Visualization of oxygen distribution patterns caused by coral and algae. *PeerJ*, *1*, e106.
Hester, E. R., Barott, K. L., Nulton, J., Vermeij, M. J., & Rohwer, F. L. (2016). Stable and sporadic symbiotic communities of coral and algal holobionts. *The ISME Journal*, *10*(5), 1157–1169.

Hughes, T. P., Baird, A. H., Bellwood, D. R., Card, M., Connolly, S. R., Folke, C., Grosberg, R., Hoegh-Guldberg, O., Jackson, J. B. C., Kleypas, J., Lough, J. M., Marshall, P., Nyström, M., Palumbi, S. R., Pandolfi, J. M., Rosen, B., & Roughgarden, J. (2003). Climate change, human impacts, and the resilience of coral reefs. *Science*, *301*(5635), 929–933.

Hughes, T. P., Rodrigues, M. J., Bellwood, D. R., Ceccarelli, D., Hoegh-Guldberg, O., McCook, L., Moltschaniwskyj, N., Pratchett, M. S., Steneck, R. S., & Willis, B. (2007). Phase shifts,

herbivory, and the resilience of coral reefs to climate change. *Current Biology: CB*, *17*(4), 360–365.

Jorissen, H., Skinner, C., & Osinga, R. (2016). Evidence for water-mediated mechanisms in coral–algal interactions. *Of the Royal …*. https://royalsocietypublishing.org/doi/abs/10.1098/rspb.2016.1137

Kauhanen, D., Sysi-Aho, M., Koistinen, K. M., Laaksonen, R., Sinisalo, J., & Ekroos, K. (2016). Development and validation of a high-throughput LC--MS/MS assay for routine measurement of molecular ceramides. *Analytical and Bioanalytical Chemistry*, *408*(13), 3475–3483.

Kimes, N. E., Grim, C. J., Johnson, W. R., Hasan, N. A., Tall, B. D., Kothary, M. H., Kiss, H., Munk, A. C., Tapia, R., Green, L., Detter, C., Bruce, D. C., Brettin, T. S., Colwell, R. R., & Morris, P. J. (2012). Temperature regulation of virulence factors in the pathogen *Vibrio coralliilyticus*. *The ISME Journal*, *6*(4), 835–846.

Knowles, B., Silveira, C. B., Bailey, B. A., Barott, K., Cantu, V. A., Cobián-Güemes, A. G., Coutinho, F. H., Dinsdale, E. A., Felts, B., Furby, K. A., George, E. E., Green, K. T., Gregoracci, G. B., Haas, A. F., Haggerty, J. M., Hester, E. R., Hisakawa, N., Kelly, L. W., Lim, Y. W., Little, M., Luque, A., McDole-Somera, T., McNair, K., de Oliveira, L. S., Quistad, S. D., Robinett, N. L., Sala, E., Salamon, P., Sanchez, S. E., Sandin, S., Silva, G. G. Z., Smith, J., Sullivan, C., Thompson, C., Vermeij, M. J. A., Youle, M., Young, C., Zgliczynski, B., Brainard, R., Edwards, R. A., Nulton, J., Thompson, F., & Rohwer, F. (2016). Corrigendum: Lytic to temperate switching of viral communities. *Nature*, *539*(7627), 123.

Knowlton, N., & Rohwer, F. (2003). Multispecies microbial mutualisms on coral reefs: the host as a habitat. *The American Naturalist*, *162*(4 Suppl), S51–S62.

Lesser, M. P., Mazel, C. H., Gorbunov, M. Y., & Falkowski, P. G. (2004). Discovery of symbiotic nitrogen-fixing cyanobacteria in corals. *Science*, *305*(5686), 997–1000.

Ley, R. E., Turnbaugh, P. J., Klein, S., & Gordon, J. I. (2006). *Microbial ecology: human gut*

Little, M., Rojas, M. I., & Rohwer, F. (2020). Bacteriophage can drive virulence in marine pathogens. *Marine Disease Ecology*. https://books.google.com/books?hl=en&lr=&id=fcXLDwAAQBAJ&oi=fnd&pg=PA73&dq=Bacteriophage+can+drive+virulence+marine+pathogens&ots=P4xFPqMm0P&sig=E-uwrTFj-P7NeSK9qp7pVujtjJU

Littman, R. A., Willis, B. L., Pfeffer, C., & Bourne, D. G. (2009). Diversities of coral-associated bacteria differ with location, but not species, for three acroporid corals on the Great Barrier Reef. *FEMS Microbiology Ecology*, *68*(2), 152–163.

Lynch, S. V., & Pedersen, O. (2016). The Human Intestinal Microbiome in Health and Disease. *The New England Journal of Medicine*, *375*(24), 2369–2379.

Mao-Jones, J., Ritchie, K. B., Jones, L. E., & Ellner, S. P. (2010). How microbial community composition regulates coral disease development. *PLoS Biology*. https://journals.plos.org/plosbiology/article/file?type=printable&id=10.1371/journal.pbio.1000345

McCook, L. J. (1999). Macroalgae, nutrients and phase shifts on coral reefs: scientific issues and management consequences for the Great Barrier Reef. *Coral Reefs* , *18*(4), 357–367.

McCook, L., Jompa, J., & Diaz-Pulido, G. (2001). Competition between corals and algae on coral reefs: a review of evidence and mechanisms. *Coral Reefs* , *19*(4), 400–417.

Mouchka, M. E., Hewson, I., & Harvell, C. D. (2010). Coral-associated bacterial assemblages: current knowledge and the potential for climate-driven impacts. *Integrative and Comparative Biology*, *50*(4), 662–674.

Munro, J., Oakey, J., Bromage, E., & Owens, L. (2003). Experimental bacteriophage-mediated virulence in strains of Vibrio harveyi. *Diseases of Aquatic Organisms*, *54*(3), 187–194.

Nissimov, J., Rosenberg, E., & Munn, C. B. (2009). Antimicrobial properties of resident coral mucus bacteria of Oculina patagonica. *FEMS Microbiology Letters*, *292*(2), 210–215.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A. C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G. D., Rodionov, D. A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., & Vonstein, V. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, *33*(17), 5691–5702.

Qu, F., Zhang, H., Zhang, M., & Hu, P. (2018). Sphingolipidomic Profiling of Rat Serum by UPLC-Q-TOF-MS: Application to Rheumatoid Arthritis Study. *Molecules* , *23*(6). https://doi.org/10.3390/molecules23061324

Quistad, S. D., Stotland, A., Barott, K. L., Smurthwaite, C. A., Hilton, B. J., Grasis, J. A., Wolkowicz, R., & Rohwer, F. L. (2014). Evolution of TNF-induced apoptosis reveals 550 My of functional conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), 9567–9572.

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*, *75*(11), 3492–3501.

Ritchie, K. B. (2006). Regulation of microbial populations by coral surface mucus and mucus-associated bacteria. *Marine Ecology Progress Series*, *322*, 1–14.

Roach, T. N. F., Abieri, M. L., George, E. E., Knowles, B., Naliboff, D. S., Smurthwaite, C. A., Kelly, L. W., Haas, A. F., & Rohwer, F. L. (2017). Microbial bioenergetics of coral-algal interactions. *PeerJ*, *5*, e3423.

Roach, T. N. F., Dilworth, J., Jones, A. D., Quinn, R. A., & Drury, C. (2020). Metabolomic signatures of coral bleaching history. *bioRxiv*. https://www.biorxiv.org/content/10.1101/2020.05.10.087072v1.abstract

Roach, T. N. F., Salamon, P., & Nulton, J. (2018). Application of finite-time and control thermodynamics to biological processes at multiple scales. *Journal of Non*. https://www.degruyter.com/view/j/jnet.2018.43.issue-3/jnet-2018-0008/jnet-2018-0008.xml

Robbins, S. J., Singleton, C. M., Chan, C. X., Messer, L. F., Geers, A. U., Ying, H., Baker, A., Bell, S. C., Morrow, K. M., Ragan, M. A., Miller, D. J., Forêt, S., ReFuGe2020 Consortium, Voolstra, C. R., Tyson, G. W., & Bourne, D. G. (2019). A genomic view of the reef-building coral Porites lutea and its microbial symbionts. *Nature Microbiology*, *4*(12), 2090–2100.

Rohwer, F., Seguritan, V., Azam, F., & Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series*, *243*, 1–10.

Rosenberg, E., Koren, O., Reshef, L., Efrony, R., & Zilber-Rosenberg, I. (2007). The role of microorganisms in coral health, disease and evolution. *Nature Reviews. Microbiology*, *5*(5), 355–362.

Rypien, K. L., Ward, J. R., & Azam, F. (2010). Antagonistic interactions among coral-associated bacteria. *Environmental Microbiology*, *12*(1), 28–39.

Siboni, N., Ben-Dov, E., Sivan, A., & Kushmaro, A. (2008). Global distribution and diversity of coral-associated Archaea and their possible role in the coral holobiont nitrogen cycle. *Environmental Microbiology*, *10*(11), 2979–2990.

Silveira, C. B., Luque, A., Roach, T. N., Villela, H., Barno, A., Green, K., Reyes, B., Rubio-Portillo, E., Le, T., Mead, S., Hatay, M., Vermeij, M. J., Takeshita, Y., Haas, A., Bailey, B., & Rohwer, F. (2019). Biophysical and physiological processes causing oxygen loss from coral reefs. *eLife*, *8*. https://doi.org/10.7554/eLife.49114

Smith, J. E., Brainard, R., Carter, A., Grillo, S., Edwards, C., Harris, J., Lewis, L., Obura, D., Rohwer, F., Sala, E., Vroom, P. S., & Sandin, S. (2016). Re-evaluating the health of coral reef communities: baselines and evidence for human impacts across the central Pacific. *Proceedings. Biological Sciences / The Royal Society*, *283*(1822). https://doi.org/10.1098/rspb.2015.1985

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W.-M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., Reily, M. D., Thaden, J. J., & Viant, M. R. (2007). Proposed minimum reporting standards for chemical analysis. *Metabolomics: Official Journal of the Metabolomic Society*, *3*(3), 211–221.

Sunagawa, S., Woodley, C. M., & Medina, M. (2010). Threatened corals provide underexplored microbial habitats. *PloS One*, *5*(3), e9554.

Sweet, M. J., Croquer, A., & Bythell, J. C. (2011). Bacterial assemblages differ between compartments within the coral holobiont. *Coral Reefs* , *30*(1), 39–52.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., & Magrini, V. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*.

Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., & Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral Porites astreoides. *Environmental Microbiology*, *9*(11), 2707–2719.

Weynberg, K. D., Voolstra, C. R., Neave, M. J., Buerger, P., & van Oppen, M. J. H. (2015). From cholera to corals: Viruses as drivers of virulence in a major coral bacterial pathogen. *Scientific Reports*, *5*, 17889.

Ziegler, M., Seneca, F. O., Yum, L. K., Palumbi, S. R., & Voolstra, C. R. (2017). Bacterial community dynamics are linked to patterns of coral heat tolerance. *Nature Communications*, *8*, 14213.

Supplemental References

Akhter, S., Aziz, R. K., & Edwards, R. A. (2012). PhiSpy: a novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Research*, *40*(16), e126.

Archer, E. (2016). rfPermute: Estimate permutation p-values for Random Forest importance metrics. *R Package Version*, *1*(2).

Graham, E. B., Tfaily, M. M., Crump, A. R., Goldman, A. E., Bramer, L. M., Arntzen, E., Romero, E., Resch, C. T., Kennedy, D. W., & Stegen, J. C. (2017). Carbon inputs from riparian vegetation limit oxidation of physically bound organic carbon via biochemical and thermodynamic processes. *Journal of Geophysical Research: Biogeosciences*, *122*(12), 3188–3205.

LaRowe, D. E., & Van Cappellen, P. (2011). Degradation of natural organic matter: A thermodynamic analysis. *Geochimica et Cosmochimica Acta*, *75*(8), 2030–2042.

McDole, T., Nulton, J., Barott, K. L., Felts, B., Hand, C., Hatay, M., Lee, H., Nadon, M. O., Nosrat, B., Salamon, P., Bailey, B., Sandin, S. A., Vargas-Angel, B., Youle, M., Zgliczynski, B. J., Brainard, R. E., & Rohwer, F. (2012). Assessing coral reefs on a Pacific-wide scale using the microbialization score. *PloS One*, *7*(9), e43233.

Olivon, F., Grelier, G., Roussi, F., Litaudon, M., & Touboul, D. (2017). MZmine 2 Data-Preprocessing To Enhance Molecular Networking Reliability. *Analytical Chemistry*, *89*(15), 7836–7840.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(Database issue), D206–D214.

Pluskal, T., Castillo, S., Villar-Briones, A., & Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, *11*, 395.

Ponstingl, H., & Ning, Z. (2010). SMALT-a new mapper for DNA sequencing reads. *F1000 Posters*, *1*(L313). https://f1000research.com/assets/download/327

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, *35*(Database issue), D61–D65.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* , *27*(6), 863–864.

Silva, G. G. Z., Green, K. T., Dutilh, B. E., & Edwards, R. A. (2016). SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* , *32*(3), 354–361.

CHAPTER 3: Three-dimensional molecular cartography of the caribbean reef-building coral *Orbicella faveolata*

Abstract

All organisms host a diversity of associated viruses, bacteria, and protists, collectively defined as the holobiont. While scientific advancements have enhanced the understanding of the functional roles played by various components of the holobiont, there is a growing need to integrate multiple types of molecular data into spatially and temporally resolved frameworks. To that end, we mapped 16S and 18S rDNA metabarcoding, metatranscriptomics, and metabolomic data onto three-dimensional reconstructions of coral colonies to examine microbial diversity, microbial gene expression, and biochemistry on two colonies of the ecologically important, reef-building coral, *Orbicella faveolata* and their competitors (i.e., adjacent organisms interacting with the corals: fleshy algae, turf algae, hydrozoans, and other corals). Overall, no statistically significant spatial patterns were observed among the samples for any of the data types; instead, strong signatures of the macroorganismal hosts (e.g., coral, algae, hydrozoa) were detected, in the microbiome, the transcriptome, and the metabolome. The 16S rDNA analysis demonstrated higher abundance of Firmicutes in the coral microbiome than in its competitors. A single bacterial amplicon sequence variant from the genus Clostridium was found exclusively in all *O. faveolata* samples. In contrast to microbial taxa, a portion of the functionally annotated bacterial RNA transcripts (6.86%) and metabolites (1.95%) were ubiquitous in all coral and competitor samples. Machine learning analysis of microbial transcripts revealed elevated T7-like cyanophage- encoded photosystem II transcripts in *O. faveolata* samples, while sequences involved in bacterial cell division were elevated in turf algal and interface samples. Similar analysis of metabolites revealed that bacterial-produced antimicrobial and antifungal compounds

were highly enriched in coral samples. This study provides insight into the spatial and biological patterning of the coral microbiome, transcriptome, and metabolome.

Introduction

The holobiont is the sum of an organism and all of its associated microbes and viruses, which interact through a complex suit of biochemicals. Technologies recently adopted for marine molecular ecology have widened the lens with which ecologists can investigate a holobiont, including standard marker amplicon sequencing, metabarcoding, metagenomics, metatranscriptomics, metaviromics, metaproteomics, metabolomics, and others. Each of these molecular approaches on their own offer a unique view into facets of macroorganismal, microbial, viral, or biochemical dynamics (Hasin et al., 2017; Pinu et al., 2019; Roach et al., 2020, 2021a). For example, 16S and 18S rDNA amplicon sequencing have allowed for targeted taxonomic profiling of prokaryotic and eukaryotic microbiomes in many systems from corals to humans (e.g., Rohwer et al., 2002; Turnbaugh et al., 2007). Metagenomics, or the shotgun sequencing of DNA, has allowed for the identification and quantification of microbial and viral taxa, and functional genes (Handelsman et al., 1998; Breitbart et al., 2002; Edwards and Rohwer, 2005; Franzosa et al., 2018). Metagenomics also provides the ability to assemble genomes de novo from community-level sequencing, and identify and describe new gene families (Parks et al., 2017; Ovchinnikov et al., 2017). Furthermore, purification-based metagenomic methods, such as metaviromics, utilize physical and chemical properties to isolate and sequence specific gene pools (e.g., viral pool, archaeal pool) that would otherwise comprise small percentages of total DNA present in the holobiont (Thurber et al., 2009; Poulos et al., 2018). Together, these emerging approaches have improved the capacity for understanding the complex nature of

holobiont biology and ecology, by providing the means to investigate all of the holobiont constituents in situ.

Metatranscriptomics, or the shotgun style sequencing of RNA, allows for the quantification of total gene expression in a community (Wang et al., 2009). In addition, amplicon sequencing of conserved coding regions of the genome (of both macroorganisms and microbes) has provided species-level and higher taxonomic identifications of operational taxonomic units (OTUs) or amplicon sequence variants (ASVs) to single nucleotide differences (Callahan et al., 2016; Thompson et al., 2017). Metabolomics provides insight into the physiological and metabolic activity of a holobiont by identifying and quantifying small molecules present in the holobiont (Quinn et al., 2016; Sogin et al., 2016; Hartmann et al., 2017; Matthews et al., 2020). Many of these methods complement each other, for example, open-reading frame protein predictions from metagenomic data can provide information for new proteins or molecules in metabolomic datasets (Wang et al., 2014). High throughput sequencing of amplified DNA or environmental DNA (eDNA) using standard gene markers, known as metabarcoding, has allowed for a massive number of censuses of single and multicellular marine eukaryotes such as algae, invertebrates, and vertebrates (Blaxter et al., 2005; Taberlet et al., 2012; Leray and Knowlton, 2015). Integrating these tools is essential for understanding the roles of microbes, viruses, and biochemicals in large-scale ecological processes. Combining these approaches can improve the capacity for scientists to observe and link states and processes from all aspects of the central dogma of biology in an ecosystem.

In addition to molecular approaches for ecology, developments in the literal ‘lens’- i.e., photographic imaging technology- have also increased the amounts and types of information that can be obtained from natural environments (Burns et al., 2015; Roach et al., 2021b).

Photogrammetric three- dimensional modeling provides a more realistic visualization and representation of environments than traditional two- dimensional image collections and has allowed for new types of spatial analysis (Burns et al., 2016; George et al., 2021). For example, analyses of 3D reconstructions of corals and their competitors have demonstrated that the geometric properties of corals 'in part' determine the outcome of competitive coral- algal interactions in several coral genera (George et al., 2021). Similar techniques for 3D reconstruction have also been used in terrestrial landscapes (Schneider et al., 2014) and freshwater ecosystems (Marazuela et al., 2018).

The combination of molecular data with spatially explicit 3D cartography tools are now allowing for unprecedented insight into pattern and process in many systems. For example, metabolite mapping on the human body demonstrated that the composition of molecules on the human skin varied depending on people's daily routine (Bouslimani et al., 2015). These technologies have yet to be fully utilized in marine environments due to the technological difficulties of in situ sampling and data collection. Here, we combine 16S rDNA amplicon sequencing, 18S rDNA amplicon sequencing, metatranscriptomics, and metabolomics with three-dimensional molecular cartography tools to provide 3D molecular maps of several components of coral and algae holobionts, allowing us to investigate the spatial distribution and ecological drivers of microbial taxa, transcripts, and small molecules.

The goal of this study was to use cutting edge -omics and imaging techniques to explore the processes occuring where different benthic holobionts meet (i.e. interfaces), providing insight into the benthic interaction ecology of the ecologically important, reef-building coral *Orbicella faveolata. O. faveolata* (previously *Montastraea faveolata*), or the mountainous star coral, is native to the Caribbean Sea and Gulf of Mexico.

O. faveolata is a massive reef-building coral currently listed as endangered by the International Union for Conservation of Nature (IUCN) (Aronson et al., 2008), with an estimated 50% abundance loss in the last 30 years. On healthy reefs, *O. faveolata* often dominate the reefs at depths of 10–20 m, and these corals face several threats including coral bleaching, algal overgrowth, and diseases such as yellow band disease, black band disease, and white plague (Kimes et al., 2013).

Here, several different molecular techniques were used to investigate specific components of *O. faveolata* holobionts. 16S and 18S rDNA sequencing data served as a metric to assess total bacterial and protistan taxonomic diversity, and to provide a general overview of the three-dimensional composition of *O. faveolata*'s microbiome. A total RNA sequencing approach was used to establish holobiont gene expression profiles, and metabolomic profiling allowed for the detection of small molecules, yielding a snapshot of the holobiont's biochemistry. We combined these approaches to test the hypothesis that holobiont taxa and function may be driven by spatial patterns (e.g., physical distance), but found no evidence supporting this. Rather, biological variables (e.g., the macroorganismal hosts) proved to be more important for taxonomic, functional, and molecular composition. Integration of multiple data types in a spatial framework provides a better understanding of the distinct components of the holobiont and how they mediate ecological interactions between macroorganisms. Future work should focus on scaling these methods both in terms of higher sampling resolution and sampling across larger spatial scales to further explore unique molecular profiles for specific organisms or physiological states (e.g., stress, disease, competition) and to examine if there are breakpoints where spatial patterns do exist.

Results

Three-Dimensional Photomosaics and Spatial Analysis

Analysis of the 3D photomosaic models allowed for the calculation of geometric properties of the corals and revealed that the East Point *O. faveolata* colony had a maximum height of 1.80 m and a maximum diameter of 1.70 m with a total surface area of 196,388 cm2, whereas the Water Factory colony was smaller with a maximum height of 1.06 m and a maximum diameter of 0.70 m with a total surface area of 8,973 cm^2. The East Point *O. faveolata* colony had an interface length of 969 cm (colony perimeter length), and the main competing organism interacting with the *O. faveolata* colony included turf algae (46.81% of the interface), along with a closely related species of coral, *Orbicella annularis* (17.80% of the interface). Coral overhangs where competitors were absent made up 24.12% of the interface, and interactions with crustose coralline algae (CCA) were limited to 11.27%. The Water Factory *O. faveolata* colony had an interface length of 545 cm, and the main interactions involved turf algae (27.36%), along with a direct interaction between the coral colony and *Millepora complanata* (fire coral/hydrozoan, 32.24%). The remainder of the interface consisted of coral overhangs (32.36%) and interactions with CCA (8.04%). The coral colonies and interacting organisms were within a 2.5 m 2.5 m (6.25 m^2) reef plot at East Point and a 1 m x 1 m (1 m^2) reef plot at Water Factory. Molecular cartography and subsequent linear regression analysis of spatial patterning demonstrated no evidence for spatial autocorrelation or isolation by distance, as there was no significant correlation between pairwise physical distance and pairwise sample similarity for any data type within the coral colonies (16S rDNA, 18S rDNA, metatranscriptomes, or metabolomes) (all R^2s < 0.04 and all p values > 0.38, Figure 3.2). Additional sampling may have improved these correlations; however, a power analysis revealed that 28 samples would be

required to see a significant trend ($p < 0.05$) using a strong effect size (0.5) and a power of 80% (Supplementary Table 3.1). Since 21-29 samples for each data type were used in this study, additional sampling would not likely improve the weak correlations observed between physical distance and sample similarity (Supplementary Table 3.2).

Diversity of Bacterial Communities: 16S rDNA

A total of 656,698 16S rDNA sequences were generated with an average of 24,322 sequences per sample ($n = 27$). A total of 3,034 amplicon sequence variants (ASVs) were identified, and ASVs identified as chloroplasts, mitochondria, and long branching chimeras were removed. Overall, 16S rDNA sequencing demonstrated that samples were not spatially structured ($p = 0.42$, Figure 3.2A) but were rather found to cluster by community type (Figures 3.3A,B; Supplementary Figure 3.1). At the ASV level, there was significant clustering of coral samples separate from all other samples (Figures 3.3A,B PERMANOVA $p < 0.003$) as well as significant differences between corals and interface samples (Figures 3.3A,B, PERMANOVA $p = 0.003$). This was largely driven by the differences in the *O. faveolata* microbiome compared to the microbiome of turf algae and interface samples (Figure 3.3B, PERMANOVA $p = 0.02$ and 0.02, respectively). Similar trends were observed at the phylum level, though PERMANOVA *p*-values were generally less significant than at the ASV level (Supplementary Figure 3.2). Furthermore, when comparing between colony versus within colony variance, there was a significant difference in 16S rDNA annotations between colonies at the ASV level (Figure 3.3B PERMANOVA $p = 0.04$), but not at the phylum level (Supplementary Figure 3.2 PERMANOVA $p = 0.73$). Interestingly, the bacterial microbiome of the two hydrozoan

competitor samples did not cluster separately as an outgroup, but were rather more similar to turf algae than to *O. faveolata* (Figures 3.3A,B).

At the ASV level, there was an overall trend for the coral microbiome to be less diverse than interfaces or competitor (turf algae and *M. complanata*) microbiomes (Figure 3.3D), with coral samples having the lowest Shannon entropy (4.43 ± 0.92), followed by the interface (5.11 ± 2.34), and finally the competitor samples (5.31 ± 1.55). It is also notable that turf algal samples had the highest entropy (6.24 ± 0.65) of all sample types. An even stronger trend of diversity was observed for rarefaction species richness (referred to as richness), with coral samples being significantly less rich than interface ($p = 0.002$) or competitor samples ($p = 0.01$). This pattern was also observed at the phylum level (Supplementary Figure 3.2).

Random forests analysis demonstrated that the bacteria in the phyla Firmicutes (specifically Clostridium), Bacteroidetes (specifically Flavobacterium) and Proteobacteria (specifically Alphaproteobacteria) were among the most important bacterial ASVs in predicting the sampling area (i.e., coral, interface, competitor). This was further supported by 9 of the top 10 most important random forests predictors at the ASV level coming from these same three bacterial phyla (i.e., Firmicutes, Bacteroidetes, and Proteobacteria) (Supplementary Figure 3.3). Interestingly, many of the top BLAST hits were to previously sampled *O. faveolata* (Supplementary Figure 3.3).

The ubiquity-abundance analysis revealed that a single ASV, in the genus Clostridium was uniquely ubiquitous in *O. faveolata* (i.e., a Clostridium sp. was found in all *O. faveolata* samples and only in *O. faveolata* samples, Figure 3.3C). The relative abundance of the Clostridium ASV (d6d6d57d15072e10d222a16f3e6d97e9) in *O. faveolata* samples ranged from 1.31% to 0.03% with an average of 0.47%. The Clostridium ASV was also one of the top ten

predictors in the random forest analysis, indicating the potential importance for this ASV in the core microbiome of *O. faveolata* (Supplementary Figure 3.2). BLAST results showed that the Clostridium ASV shared 100% sequence identity with an uncultured bacterium clone collected from *O. faveolata* in Panamá in 2008 (Sunagawa et al., 2010). The Clostridium ASV was highly abundant and ubiquitously distributed across all *O. faveolata* samples, but absent from any and all interface or competitor samples (Video S3.1). The phylum Firmicutes, which includes Clostridium, also showed a similar distribution across the coral colonies and their competitors, however Firmicutes was highly abundant in all corals including *O. annularis* (Figure 3.4). Other phyla, such as Cyanobacteria, displayed the opposite pattern, with low relative abundances in *O. faveolat*a and higher relative abundances in turf algal competitors and interfaces (Figure 3.4C). Several rare bacterial taxa were also present at both sites. Four unique ASVs belonging to Endozoicomonas, a symbiont of marine animals, were found at low relative abundances (0.02–0.99%) in the *O. faveolata* colonies from both sites, and the bacterial symbiont was present in 60% of the coral samples. Five different Endozoicomonas ASVs were present in the hydrozoan competitor, *M. complanata*, at higher relative abundances (0.65–7.49%), suggesting *O. faveolata* and *M. complanata* specific symbionts. ASVs belonging to the understudied TM6 phylum were also identified at low relative abundances (0.01– 0.26%), specifically from two interface samples and one turf algae sample at Water Factory along with one East Point *O. faveolata* sample.

Diversity of Microbial Eukaryotes: Non-metazoan 18S rDNA

A total of 2,972,176 18S rDNA sequences were generated with an average of 102,488 sequences per sample (n = 29). A total of 541 ASVs were identified and ASVs identified as bacteria and long branching chimeras were removed. Symbiodiniaceae, specifically

Cladocopium spp., was the most abundant taxon in coral samples, whereas red (Florideophyceae) and green (Ulvophyceae) algae were the dominant taxa in turf algae samples (Supplementary Figure 3.4). ASVs belonging to these taxa were also the main drivers of the clustering observed in the PCoA and NMDS (Supplementary Figures 3.1, 3.4). The majority of the ubiquitous ASVs unique to *O. faveolata* were also Symbiodiniaceae (Supplementary Figure 3.5), and a phylogenetic analysis showed no clustering of ASVs based on coral colony or coral species (Supplementary Figure 3.6). However, Symbiodiniaceae ASVs from *M. complanata* samples formed a well-supported clade that was distinct from the coral Symbiodiniaceae clade, suggesting hydrozoan-specific and coral-specific eukaryotic symbionts (Supplementary Figure 3.6). BLASTn results showed that the coral Symbiodiniaceae ASVs shared a 99.01% identity with several Clade C symbionts (Cladocopium spp.) whereas *M. complanata* ASVs had a 98.77% identity to Clade A symbionts (*Symbiodinium* spp.).

Corallicolid, an intracellular apicomplexan symbiont of coral, was present at low abundances in three coral samples including the competing *O. faveolata* colony at East Point (northern transect, sample V, relative abundance = 0.075%), the competing *O. annularis* colony at East Point (southern transect, sample V, relative abundance = 0.29%) and the interface between *O. faveolata* and *M. complanata* at Water Factory (western transect, sample III, relative abundance = 0.36%). This symbiont has been reported in *O. faveolata*, *O. annularis*, and several other cnidarians (Kwong et al., 2019), but its presence in some coral colonies and absence from nearby conspecific colonies suggests a sporadic association with specific coral species.

Function of Bacterial Communities: Metatranscriptomes

Similar to 16S and 18S rDNA, the functional annotation of bacterial metatranscriptomes also demonstrated significant clustering of samples by host organism, with *O. faveolata* samples being significantly different than interface (PERMANOVA $p < 0.01$) and competitor samples (PERMANOVA $p < 0.001$) (Figures 3.5A,B and Supplementary Figure 3.1). In stark contrast to 16S rDNA data where the hydrozoan outgroup (pink triangles in Figure 3.5A) grouped with turf algae, the bacterial transcriptomes group hydrozoan samples with coral, suggesting a potential taxonomic, functional decoupling in these microbiomes. Of the 5,206 functionally annotated bacterial RNA transcripts from all host organisms, 357 (6.86%) were ubiquitous to all samples, and no functions were unique to *O. faveolata*; a strong contrast to the 16S rDNA results (Figure 3.3C). Of the ubiquitous transcripts, the NAD(P)H-quinone oxidoreductase function was the most abundant, with a mean relative abundance of 12.4% in all samples; furthermore, NAD(P)H-quinone oxidoreductase was also shown to be the most important predictor for determining sample type in the random forests classification analysis (Supplementary Figure 3.7).

In agreement with the 16S and 18S rDNA results, the distribution of the functionally annotated bacterial transcripts across samples resulted in significantly lower diversity in coral samples (Shannon, 5.08 ± 0.74; Richness, 939.55 ± 356.53) and significantly higher diversity in other sample types (Shannon, 5.63 ± 0.77; Richness, 1289.83 ± 359.25), in particular turf algal competitors (Shannon, 5.88 ± 0.56; Richness, 1404.71 ± 262.29) (Figure 3.5D).

Random forests analysis revealed that microbial transcripts were enriched in phage-encoded photosystems in coral samples and enriched for transcripts involved in bacterial cell division, motility, and chemotaxis in turf algal and interface samples (Figure 3.6). At all levels of the SEED hierarchy (i.e., Levels 1,2,3, and function) phage related genes were found to be

significantly enriched in coral samples and a driver of the clustering seen in the data (Figure 3.6). Specifically, we found the photosystem II protein D1 (PsbA) transcript to be driving the T7-like cyanophage pattern at higher levels (Supplementary Figures 3.8, 3.9). This has been visualized in a two-way dendrogram in Figure 3.6A and via a 3D molecular heatmap in Figure 3.6B. Other random forests top predictors included bacterial respiration along with cell regulation and signaling, both of which were more abundant in coral than the interface or competitors. The opposite pattern was observed for functions involving membrane transport, cell wall and capsule, iron acquisition and metabolism, virulence, dormancy and sporulation, and phages (prophages, transposable elements, plasmids) (Figure 3.6A).

Associated Biochemicals: Metabolomes

The metabolomic data set consisted of 2,315 unique features of which 53 (2.28%) had a spectral match to a known compound in the GNPS database with another 230 analog hits (i.e., contains a known mass shift and has a similar spectrum). Similar to 16S and 18S rDNA, and transcriptomic data, metabolites demonstrated significant clustering of *O. faveolata* samples versus their competitors (Figures 3.7A,B and Supplementary Figure 3.1, PERMANOVA p = 0.03). However, unlike other data types, metabolomic analysis was not able to distinguish interfaces as separate groups (Figure 3.7B, PERMANOVA p = 0.102). The main drivers of the observed clustering included several unknown metabolites along with a glycerophosphocholine (m/z 963.71, ClassyFire subclass) and phenylalanine (m/z 166.09, library hit) (Figure 3.7A). The unknown metabolites had mass-charge ratios of 146.10, 229.15, 482.36, and 510.39, and the unknown metabolites with mass-charge ratios of 482.36 and 510.39 had similar ratios to C16 Lyso-PAF (m/z 482.36) and C18 Lyso-PAF (m/z 510.42), respectively (Quinn et al., 2016). The

C18 Lyso-PAF metabolite (m/z 510.39) was also the second top predictor of coral, turf algae, and interface groups in the random forests analysis on all metabolites (Figure 3.8 and Supplementary Figure 3.10).

The ubiquity-abundance analysis revealed that out of 2,315 metabolites, 45 were found in all samples (1.94%) while 219 and 39 metabolites were unique to coral (*O. faveolata* and *O. annularis*) and turf algae, respectively. Of the coral specific metabolites, 136 were only present in *O. faveolata*. (Figure 3.7C). Of those compounds found only in *O. faveolata*, the cyanobacterial-produced antifungal compound lobocyclamide (dereplicator hit) and the Streptomyces-produced antimicrobial coleimycin (analog hit) were among the most ubiquitous compounds (Supplementary Figure 3.10). Five unknown metabolites with mass to charge ratios of 467.31, 240.10, 432.34, 1149.74, and 397.21 were also uniquely ubiquitous in *O. faveolata* (ubiquity = 0.92). Unlike 16S, 18S, and transcriptome samples, the random forests analysis of metabolites did not find the ubiquitous molecules among the top compounds for predicting sample type. Instead, the random forests analysis identified two fatty acid esters (ClassyFire annotation) connected in the network, a tyrosine-proline dipeptide (analog hit), and a ceramide (analog hit) to be the top predictors of sample type (Figure 3.8 and Supplementary Figure 3.9).

Richness and entropy in metabolomic samples behaved opposite to that of the 16S rDNA, 18S rDNA, and RNA datasets (Figure 3.7D), with corals generally possessing a more diverse metabolome (Shannon = 6.03 ± 0.41, Richness = 1607 ± 364) compared to interfaces (Shannon = 5.43 ± 0.74, $p = 0.03$, Richness = 1067 ± 578, $p = 0.11$) or competitors (Shannon = 5.32 ± 0.55, $p = 0.005$, Richness = 727 ± 389, $p < 0.001$).

Relationship Between Diversity of Molecular Data Types

There were no significant correlations between the entropy (H') (all R^2s 0.07 and *p*-values 0.33, Supplementary Figure 3.11) or richness (all R^2s 0.02 and all *p*-values 0.52, Supplementary Figure 3.11) among different data types, indicating potential decouplings in taxonomic, functional, and biochemical profiles. An analysis of all data types sampled within the 2.5 m × 2.5 m reef site at East Point (6.25 m^2) and a 1 m x 1 m reef area at Water Factory (1 m^2) also showed differences in entropy between sites. The Water Factory site had the most diverse 18S rDNA (entropy = 3.71) and bacterial function (transcriptomes, entropy = 6.22), whereas East Point had the highest diversity of 16S rDNA (entropy = 4.33) and metabolites (entropy = 6.75) (Supplementary Figure 3.12). Furthermore, when all samples were considered, metabolomes were seen to be the most diverse of all data types (entropy = 6.76), while 18S rDNA was the least diverse (entropy = 3.83) (Supplementary Figure 3.12).

Integration of Molecular Data Types

A mmvec neural net analysis was used to generate conditional probabilities of features in the different data types (i.e., 16S rDNA, transcripts, and metabolites) (Morton et al., 2019). This analysis revealed broad patterns amongst all data types with conditional probabilities ranging from 1.05×10^{-06} - 0.720 for 16S rDNA vs transcripts, from 6.7×10^{-10} - 0.297 for 16S rDNA vs metabolites, and from 1.93×10^{-08} - 0.301 for transcripts vs metabolites (Figure 3.9 and Supplementary Figure 3.13). To further analyze these patterns, we selected the top predictors from the previous random forests analyses to construct two-way heatmaps (Figure 3.9). The highest of all conditional probabilities from this analysis was between Chlamydiae and transcripts for dormancy and sporulation (0.720), which was an order of magnitude higher than

the average conditional probability for this data set (0.030 - 0.065). Other notably high conditional probabilities for 16S genes vs transcripts include the following: Nitrogen metabolism and Planctomycetes (0.525), phosphorus metabolism and Proteobacteria (0.210), photosynthesis with Cyanobacteria (0.122), and potassium metabolism and Planctomycetes (0.118). Notably low conditional probabilities included the relationships between Plantomycetes and stress response (9.93×10^{-6}), secondary metabolism (1.05×10^{-6}), phosphorus metabolism (1.33×10^{-5}), motility and chemotaxis (2.71×10^{-5}), dormancy and sporulation (6.36×10^{-5}), and virulence (9.13×10^{-5}).

When relating metabolomic profiles to 16S genes, the average conditional probability was much lower (1.35×10^{-3} - 0.00217). Notably high conditional probabilities between 16S and metabolites included glycerophosphoethanolamine and Firmicutes (0.019), terpene lactones and Chloroflexi (0.012), quinuclidines and Firmicutes (0.007). Notably low conditional probabilities included the following relationships with Spirochaetes and fatty acyls (1.35×10^{-07}), peptides (4.51×10^{-07}), and glycerophosphoethanolamine (4.46×10^{-07}).

The conditional probabilities of several microbial transcripts and metabolites varied between Firmicutes and Bacteroidetes (Figure 3.9), two of the main phyla identified in the previous analyses. Dormancy and sporulation along with cell division and cell cycle had a greater conditional probability with Bacteroidetes (0.270 and 0.039) relative to Firmicutes (0.010 and 0.007), whereas the conditional probability of phosphorus metabolism, RNA metabolism, and virulence was higher with Firmicutes (0.134, 0.060 and 0.097) than with Bacteroidetes (0.027, 0.004 and 0.002). The conditional probability of Phage group 1 (phages, prophages and transposable elements) with Firmicutes (0.040) was also higher than with Bacteroidetes (0.009). For metabolites, glycerophosphoethanolamine and cis-9- hexadecenoic acid (8S-HETrE) had a

higher probability with Firmicutes (0.019 and 0.001) than with Bacteroidetes (0.002 and 7.04 x 10^{-4}). Firmicutes and these metabolites were also more abundant in corals than in other organisms (Figures 3.4, 3.8), highlighting a potential association between the coral microbiome and metabolome.

Discussion

Coral and algal competition has been studied extensively utilizing molecular techniques, such as amplicon sequencing, metagenomics, and metabolomics (Barott et al., 2012; Haas et al., 2016; Quinn et al., 2016; Pratte et al., 2018; Silveira et al., 2019; Clements et al., 2020; Roach et al., 2020). Here, we integrate bacterial and protistan metabarcoding, metatranscriptomics, and metabolomics across 3D models of the endangered reef-building coral *Orbicella faveolata* and its competitors. This approach allowed us to integrate and visualize molecular data in a spatial framework to investigate spatial and biological patterns within and between *O. faveolata* colonies and their competitors (Figure 3.1).

All molecular profiles (DNA, RNA, and metabolites) revealed ecologically relevant differences between the *O. faveolata* colonies, the interface, and the corresponding benthic competitors (Figures 3.2–3.8). Holobionts clustered by microbial composition (16S and 18S), bacterial gene function (transcriptomes), and metabolites (Figures 3.3, 3.5, 3.7). Many uniquely ubiquitous microbes and metabolites, and sporadic microbial symbionts were found to be associated with *O. faveolata*. No spatial patterns were observed, suggesting that a single coral colony is a diverse and dynamic consortium of host, microbes, and associated functions/metabolites (Figure 3.2). However, larger sampling efforts may be warranted to investigate this further and the scale of sampling should be taken into consideration. We also utilized machine-learning algorithms to compare thousands of variables from different data types (Figures 3.4, 3.6, 3.8) and found potential associations between molecular functions and various members of the coral holobiont (Figures 3.9, 3.10).

Uniquely Ubiquitous and Sporadic Coral Symbionts

A putative Clostridium sp. ASV was uniquely ubiquitous to the *O. faveolata* colonies, pointing to it as a core member of the *O. faveolata* holobiont (Figure 3.3C and Supplementary Video 3.1). Our findings are corroborated by a previous study that reported this same Clostridium sp. ASV to be present in *O. faveolata* from Panamá in 2008 (Sunagawa et al., 2010), which suggests stability of the ASV as an *O. faveolata* symbiont in the Caribbean Sea. Clostridium are commonly found in corals, likely inhabiting hypoxic portions of the *O. faveolata* holobiont. They have been found in a wide variety of coral species across large geographic spatial scales, where they may occupy low-oxygen mucosal microniches generated by the breakdown of complex carbon (McKew et al., 2012). Our findings highlight the need for future studies on compartmentalization of the coral holobiont, for example, using multi-omics with respect to different anatomical regions of the coral (Sweet et al., 2011).

Other microbial symbionts showed differential associations with the coral holobionts and their respective competitors. Endozoicomonas, a bacterial symbiont of marine animals (Neave et al., 2016), was found at low abundances in the majority of *O. faveolata* samples at both sites. Conversely, Endozoicomonas were highly present in the hydrozoan competitor, *M. complanata*. Endozoicomonas ASVs were unique to each cnidarian host, suggesting at least two species of Endozoicomonas, each specific to *O. faveolata* and *M. complanata*. Endozoicomonas is likely involved in the cnidarian sulfur cycle (Tandon et al., 2020), and may play a relevant functional role in the holobionts. Corallicolid, an apicomplexan coral symbiont, was also present in some coral colonies and absent from nearby conspecific colonies, showing a potential sporadic association with specific coral species. At East Point, corallicolids were found in the *O. faveolata* and *O. annularis* colonies interacting with the coral of interest, while it was present in only one

O. faveolata interaction sample at the Water Factory site. Corallicolids are known to infect a diverse range of cnidarian hosts, however, their ecological role remains unknown (Kwong et al., 2019), and additional sampling is required to determine the apicomplexan's association with various coral species.

Several bacterial patterns at the phylum level were observed to be specific to the coral holobionts. Broadly, abundances of the phylum Firmicutes, which contains the genus Clostridium, were higher in the coral samples, and abundances of Bacteroidetes were higher in turf algae samples resulting in significantly higher Bacteroidetes-to-Firmicutes ratios at the interface between coral and turf algae (Figure 3.4). Interestingly, Bacteroidetes-to- Firmicutes ratios have been linked to coral health; in specific, this ratio has been shown to be elevated in coral holobionts that are losing against turf algal competitors and also in disease states (Closek et al., 2014; Roach et al., 2017, 2020). The differential abundances of these phyla have been studied more in depth with regards to human health, where the gut microbiome profile of obese individuals is linked to a decrease in Bacteroidetes-to- Firmicutes ratios (Turnbaugh et al., 2006). This suggests that the correlation between the relative abundances of Bacteroidetes and Firmicutes, and host physiology is potentially specific to the animal host, and likely has to do with the delicate balance of anaerobic and aerobic respiratory processes.

Relevant Transcripts and Metabolites

Other bacterial phyla, such as Cyanobacteria, displayed relatively low abundances in coral samples compared to the interface and competitor samples. Conversely, transcripts from a T7-like cyanophage (family Podoviridae, Cyanopodovirus), specifically, photosystem II psbA transcripts, were found in high abundance in coral samples compared to interface and competitor

samples (Figures 3.4, 3.5). Other coral studies have found abundant cyanobacterial photosystem genes, including psbA and psbD, that were potentially acquired through phage infections (Weynburg et al., 2017). Viral-mediated horizontal gene-transfer of photosystems genes, and other genes, from and to the bacterial host has been observed (Chénard and Suttle, 2008; Thompson et al., 2011; Little et al., 2020); furthermore, a bacteriophage- adherence-to-mucus model suggests that phage interact with mucosal surfaces, such as the surface of a coral, through outer capsid protein domains (Hoc) previously identified in Cyanophage (Sullivan et al., 2005; Barr et al., 2013; Silveira and Rohwer, 2016). Our observation of the expression of T7- like cyanophage photosystem genes may suggest these viruses are active in coral mucus and possibly a widespread feature of coral holobionts.

Metabolomic analysis highlighted 136 known molecules as being unique to the *O. faveolata* holobiont (Figure 3.7C). Among the metabolites unique to *O. faveolata*, the bacterial produced metabolites, lobocyclamide, a known antifungal compound produced by Cyanobacteria (MacMillan et al., 2002), was one of the most ubiquitous (Supplementary Figure 3.10). Bacterial-produced antifungals have been found previously in Gorgonian octocorals, where *Pseudoalteromonas* sp. have been shown to produce them in a light-dependent manner (Moree et al., 2014). Another uniquely ubiquitous *O. faveolata* metabolite was the Streptomyces-produced antimicrobial coleimycin (Supplementary Figure 3.10). The significant enrichment of these bacterial-produced antimicrobials highlights the potential for bacterial modulation of the host microbiome.

Other compounds found to be elevated in the *O. faveolata* holobiont included several putatively bioactive lipids such as lyso-platelet activating factor and a ceramide analog (Figures 3.7, 3.8, 3.9). These lipids can serve as modulators of the coral immune system, and have been

shown to play a role in corals' response to various stressors including turf algae overgrowth and temperature stress (Quinn et al., 2016; Galtier d'Auriac et al., 2018; Roach et al., 2020, 2021a). Our findings highlight the importance of these and other bioactive lipids and bring into question the role of the coral immune system in ecological interactions (Quistad et al., 2014).

Diversity and Spatial Patterns

Overall, corals had lower diversity of microbial composition and function, but displayed higher metabolite diversity compared to interfaces and turf algae. This is consistent with previous studies on coral and algae interactions where the bacterial consortium within corals was found to be less taxonomically rich and diverse than algal holobionts (Figure 3.3D; Barott et al., 2011). In terms of spatial patterns of the taxonomic, functional, and metabolite data, we observed no evidence for significant isolation by distance or spatial autocorrelation (Figure 3.2). Several other environmental factors may impact spatial patterns within a coral colony, such as water movement over the coral surface (e.g., boundary layers) or incident light; molecular factors such as gene regulation and post-translational modifications could also affect this spatial patterning.

Co-occurrence Networks Linking Omics Datasets

Understanding the distribution of various components of the holobiont in context to each other remains a largely under investigated aspect of molecular ecology in host-associated systems. To this end, an artificial neural network approach (mmvec) was used to calculate the conditional probabilities of the multi-omics annotations produced in this study (Morton et al., 2019). The network analysis revealed strong co- occurrences between Cyanobacteria and photosynthesis along with Planctomycetes and nitrogen metabolism transcripts (Figure 3.9).

Planctomycetes is one of the only known bacterial groups to perform anaerobic ammonium oxidation (anammox), a process that significantly contributes to the nitrogen cycle (Jetten et al., 2009). Interestingly, increased levels of bacterial nitrogen fixation in the coral holobiont have been hypothesized to contribute to coral disease states and disrupture in the stability between corals and their endosymbiotic partners (Shashar et al., 1994; Santos et al., 2014; Rädecker et al., 2015). Because most Planctomycetes have strong associations with macroalgae, lack peptidoglycan walls (e.g., exhibit resistance to antimicrobials produced by other bacteria), and possess holdfasts for attachment, it is probable that members of this phylum contribute to processes associated with algal competition and coral bleaching. Additionally, the highest co-occurrence found was between transcripts for dormancy and sporulation and the bacterial phylum, Chlamydiae (Figure 3.9). All members of Chlamydiae are obligate intracellular symbionts of eukaryotes and infect a wide range of hosts including protists and animals (Horn et al., 2000; Roulis et al., 2013). The relationship between these symbionts and dormancy and sporulation functions provides insight into the lifecycle of these intracellular bacteria. The human pathogen, Chlamydia, can exist in a metabolically reduced state known as an elementary body that allows for long- term survival in nutrient-poor conditions (Hogan et al., 2004). This 'spore-like state' would likely be common on coral reefs where host conditions vary due to the fluctuating environment. However, the potential hosts of these specific Chlamydiae symbionts remain unknown, as is the case for the majority of Chlamydiae (Dharamshi et al., 2020).

The neural network analysis also linked the presence of specific transcripts and metabolites with Firmicutes and Bacteroidetes. As previously mentioned, Firmicutes were highly abundant in coral samples (Figure 3.4), and this group co-occurred with phage, prophage, and transposable element transcripts (Figure 3.9) which were also highly abundant in coral samples

(Figure 3.6). On the other hand, cell division and cell cycle transcripts were more abundant in interface and algal samples, and the transcripts co-occurred with Bacteroidetes (Figure 3.9). These patterns where Bacteroidetes and cell division/growth associate with the interface and algae, whereas Firmicutes and phage related functions associated with coral provide further evidence of the functional roles of these taxonomic groups. The co-occurrence of glycerophosphoethanolamine and cis-9-hexadecenoic acid (8S-HETrE), two highly abundant coral metabolites, with Firmicutes (Figure 3.9) also suggests a potential association between the coral microbiome and metabolome (Figure 3.10).

Current Limitations and Future Directions

Here, we highlight the novelty and robust power of incorporating multi-omics into natural history and ecological frameworks (Figure 3.10). However, among the limitations of our findings is the use of database-dependent analyses. In the future, tools such as artificial neural networks (Mendez et al., 2019; Cantu et al., 2020) can be used in bioinformatic and chemoinformatic approaches to shed light on sequence and mass-spectrometry data that is absent in repositories. Furthermore, less invasive sampling techniques could be used to capture molecular dynamics over relevant temporal scales (e.g., before and after stress events). Our data reveals important in situ molecular ecology of the coral holobiont, and follow-up studies should extend these sampling schemes to entire reefs to investigate holobionts over multiple spatial and biogeographic scales.

This study focused on investigating the spatial and ecological drivers of two coral colonies in unprecedented detail using four different omics techniques combined with photogrammetric spatial reconstruction. The lack of spatial patterns observed is potentially a

result of other uninvestigated variables such as environmental factors or boundary layer processes. The lack of spatial structuring at this scale may also be due to the resolution and scale of sampling. For example, samples could be collected more densely or distributed over larger areas to investigate these trends at other scales such as host-associated microenvironments or across whole reef systems. Because these methodologies are becoming more affordable (i.e., the cost of DNA/RNA sequencing and mass spectrometry), scaling these approaches to address holobiont spatial aspects at the whole reef- level or even the level of island chains or ocean basins is now feasible. In addition, 3D image reconstruction has been applied at the reef-scale and is relatively straightforward to implement (e.g., Edwards et al., 2017; Roach et al., 2021b). Future molecular- mapping methods combined with environmental and temporal data will elucidate additional functions performed by the diverse members of the coral holobiont that, in turn, influence molecular to reef-scale processes.

Conclusions

The data herein revealed that microbial composition, function, and molecular profiles are strong drivers of differences between coral and algae holobionts, and at the spatial scales examined, no evidence of spatial patterns within the *O. faveolata* colonies were observed for the various data types. This dataset revealed uniquely ubiquitous microbes and metabolites, along with potentially sporadic microbial symbionts associated with this endangered reef-building coral. The 16S rDNA data from *O. faveolata* demonstrated that a single ubiquitous Clostridium ASV was present in all colony samples, and abundant T7- like cyanophage transcriptional functions were also found in *O. faveolata*, adding new insight into the viral fraction of the coral holobiont. Metabolomics profiling displayed high levels of bioactive lipids in the *O. faveolata* samples, highlighting their role in coral immune responses. In addition, the co-occurrence of bacterial taxa, microbial transcripts, and small molecules revealed specific bacterial phyla with strong associations to various metabolic functions, such as Planctomycetes and nitrogen metabolism along with Chlamydiae and dormancy functions. These methodologies provide a new framework to examine holobiont structuring and the molecular ecology of the coral holobiont.

Methods

Sample Collection

Two individual colonies of *Orbicella faveolata* were selected as "corals of interest" at two separate dive sites (Water Factory, 12◦06 31.7 N 68◦57 13.4 W and East Point, 12°02 35.1 N 68°45 43.5 W) on the island of Curaçao in the southern Caribbean (Figure 3.1). Samples from the *O. faveolata* colonies and their interacting organisms (turf algae, *Millepora complanata*, *Orbicella annularis*, and one conspecific *O. faveolata*) were collected along transects using the cardinal directions (Figure 3.1). For each North, East, South and West transect, two coral samples (I and II), one interaction/interface (III) and two competing organism (IV and V) samples were collected. The interaction/interface sample was taken from the border where the two organisms touch, and includes tissue from both of the organisms. Three tissue biopsies (1 cm diameter × 1 cm depth) from each location (I, II, III, IV, and V) were sampled for metabolomics, metatranscriptomics and microbial diversity (i.e., 16S and 18S rDNA sequencing). The tissue biopsies were collected by divers on SCUBA using a diamond encrusted core drill bit (Lasco Diamond Products). Samples were stored in ziploc bags until their return to a moored vessel with indoor laboratory space where samples for metatranscriptomics (RNA) were fixed with RNAlater (ThermoFisher Scientific cat# AM7020) in cryovials and flash frozen in liquid nitrogen. Samples for mass spectrometry (metabolomics) were stored in 70% LC/MS grade methanol/water in glass vials and stored at 20°C. Samples for metabarcoding (DNA) were also flash frozen in liquid nitrogen. All molecular samples were later moved to −80°C prior to processing and analysis.

Photogrammetry, 3D Model Construction and Molecular Cartography

Over 200 photographs were taken of each *O. faveolata* colony and their interacting organisms at the two separate sites (Water Factory and East Point) using a Canon Rebel T4i with a 35-mm lens and two Keldan 800 lumen video lights to illuminate the corals uniformly (George et al., 2021). A ruler, placed along the interface of the coral colonies, was photographed to set the scale for the resulting 3D digital models. Pre-sampling images were used for the 3D coral reconstruction, and post-sampling images were taken to map the sample location on the 3D models. Photos from each coral colony were color corrected and subsequently imported into AutodeskQR ReCapTM Photo 21.0 to create spatially accurate 3D coral reconstructions (Burns et al., 2015; Leon et al., 2015). The models were scaled using the in-reef ruler incorporated into the 3D-model and the ReCapQR Photo 'set scale and units' tool. The models were exported as a STL file as well as an OBJ file that included corresponding MTL and JPG files.

The STL models were then imported into MeshLab 2020.03, and the spatial coordinates for each sampling point in the 3D models were determined using methods from Protsyuk et al. (2018). The location of each sample was selected on the 3D model surface using the 'reference scene' tool in MeshLab, and the x, y, z coordinates for each sample location were recorded. The corresponding sample data was then added to the spatial coordinates in a CSV file. Three-dimensional molecular heatmaps were generated using the online software platform ili (Protsyuk et al., 2018). These maps were constructed with the OBJ files with corresponding MTL and JPG files as the input for the 3D surfaces and CSV files for the -omics data and spatial coordinates. Colored mesh was overlaid using the JPG file, and a linear 'hot' color scale was used for all 3D molecular heatmaps. The data was displayed using a spot border capacity and radius of 1.

DNA/RNA Extraction

DNA was extracted from tissue biopsies using DNeasy PowerBiofilm kit (Qiagen Cat No./ID: 2400-50) and quantified using Qubit DNA HS assay kit. Total RNA was extracted from the samples using the AllPrep DNA/RNA kit (Qiagen Cat No./ID: 80204). DNA was used for amplicon sequencing and the RNA portion was dedicated for metatranscriptomes. RNA was quantified using Qubit RNA HS assay kit and checked on an Agilent Bioanalyzer 2100 (Cat. 5067-1513) before proceeding to library preparation.

Amplicon Sequencing

To determine the bacterial diversity of *O. faveolata* and its competitors, extracted DNA was sent to the Integrated Microbiome Resource (IMR) at Dalhousie University where the V6-V8 region of the 16S rDNA was amplified, and the amplicon samples were sequenced with Illumina MiSeq using 300 x 300 bp paired end V3 chemistry, resulting in 656,698 reads. To study the protist diversity, PCR amplification of 18S rDNA was performed using non-metazoan 18S rDNA primers (UNonMet). The cleaned and purified PCR products were also sent to the IMR at Dalhousie University where the V4 region of the 18S rDNA was amplified, and the amplicon samples were sequenced with Illumina MiSeq using 300 300 bp paired end V3 chemistry, resulting in 2,972,176 reads.

The paired-end reads were demultiplexed and denoised using DADA2 as part of the Qiime 2TM pipeline (Bolyen et al., 2019), and sequences were trimmed using the parameters "-p-trim- left-f 19 -p-trim-left-r 20 -p-trunc-len-f 290 -p-trunc-len-r 290" for 16S amplicon sequences and "-p-trim-left-f 15 -p-trim-leftr19 -p-trunc-len-f 290 -p-trunc-len-r 290" for 18S amplicon sequences. Taxonomic classification was performed using the q2-feature-classifier

(Bokulich et al., 2018) trained against the SILVA database version 132, and phylogenetic trees were built with FastTree as part of the Qiime 2TM pipeline. Long branching amplicon sequence variants (ASVs) were removed from 16S and 18S rDNA datasets, along with ASVs classified as plastids and mitochondria in the 16S rDNA dataset. For 16S rDNA, ASVs were collapsed into a Qiime 2 taxonomy table at the phylum level, and analyses were conducted on both bacterial ASVs and phyla.

Metatranscriptomics

RNA libraries were generated using Illumina TruSeq RNA Library Prep Kit v2 (Cat: RS-122-2001) and spiked into 3 lanes of an Illumina Hiseq 2500 run using 150 x 150 bp paired end (PE) chemistry, yielding a total of 282,578,680 quality- filtered PE reads with an average of 7,436,281 PE reads per sample (Supplementary Table 3.1). Samples were demultiplexed using the manufacturer's software (basespace.illumina.com). Reads were quality filtered using PRINSEQ with parameters "-ns_max 0 -derep -lc_entropy = 0.5 -trim_qual_right = 15-trim_qual_left = 15 -trim_qual_type mean -trim_qual_rule lt -trim_qual_window 2 -min_len 30 -min_qual_mean 20 - rm_header" (Schmieder and Edwards, 2011). Only read 1 sequences were used in the microbial functional annotation.

Bacterial transcripts were identified in our metatranscriptomic sequence dataset using the mapping algorithm SUPER-FOCUS against the RAST-SEED database for microbial functions with default parameters (Overbeek et al., 2014; Silva et al., 2016). Annotations for specific eukaryote coding regions were conducted using FRAP, which utilizes SMALT mapping, with a 96% identity over 100% of the length of the quality- filtered and trimmed sequence read (Ponstingl and Ning, 2010). Sequences mapping to coral and Symbiodinium reference genomes

were removed prior to the microbial subsystem annotation (Supplementary Table 3.2). A compilation of publicly available coral genomes and Symbiodinium genomes from reefgenomics.org were used to identify coral and Symbiodinium- associated reads (Liew et al., 2016).

Ultra-Performance Liquid Chromatography – Tandem Mass Spectrometry

Methanolic extracts were analyzed as described in Roach et al. (2020). Analysis was done using Ultra Performance Liquid Chromatography (UPLC) with a Kinetex C18 reverse phase column (Phenomenex Inc.) connected to a Maxis Q-TOF Mass Spectrometer (Bruker Daltonics). The raw datafiles from the MS machine were converted into .mzXML files with the Bruker Data Analysis software version 4.1. The .mzXML files are available on the MassIVE database (massive.ucsd.edu) under numbers MSV000080597 and MSV000080632 (same dataset). The .mzXML files were imported into MZmine2 (Pluskal et al., 2010) of which a beta version was used (2.37.1.corr17.7, see Supplementary Methods). Thresholds used were the same as in Roach et al. (2020) only the ADAP chromatogram builder function was used with the following settings: a minimum height intensity was 3.00×10^3, the group intensity threshold 3.00×10^3, minimum group size in number of scans was set to 3, together with a mass tolerance of 25 ppm or 0.05 m/z.

In addition to the MZmine workflow, the MZmine metacorrelate function was used which performed a Pearson correlation analysis with the following settings: Retention time tolerance: 0.1 min, the minimum height: $4.00 \times 10^3$and noise level: 3.00×10^3. The minimum number of samples in all was set to an absolute value of 2, and there was no minimal number of samples per group. 60% of the intensity overlap was set as a minimum and gap filled features were excluded. Minimum number of data points was set to 5, and two points on the edge. The

shape correlation was set to 85% minimum. Ion Identity networking function searched for all modifications and all adducts from the positive ion mode within a 0.05 – 25 ppm, with a minimum height of 4.00 x 10^3. The maximum charge was set to 4, and there was a maximum for 3 molecules per cluster. For more details on the MZmine workflow and specific settings see Supplementary Materials.

Metabolomics: Molecular Networking and Spectral Library Searching

The Feature-Based Molecular Networking (FBMN) workflow (Nothias et al., 2020) in the GNPS environment (Wang et al., 2016) was used to create the Ion Identity Network (IIN) (job ID: 415d51c0fd25492187b2d822eaa6f217). All fragment ions within 17 Da of the precursor m/z were removed. MS/MS spectra were filtered by selecting the top 6 fragment ions within a 50 Da window throughout the spectrum. Mass tolerances were set for both the precursor ion and the MS/MS fragment ion to 0.05 Da. A molecular network was created with edges having a cosine score above 0.7 and more than 4 matched peaks. Nodes had to appear in each other's top 20 of most similar nodes, and the maximum size of a molecular family was set to 500. If a molecular family was larger than 500, the lowest scoring edges were removed.

Library spectra were filtered in the same way as the input data. For a positive library match the spectra needed a minimum of 4 matching peaks and a cosine score above 0.7. Analog search mode was used by searching against MS/MS spectra within a difference of 100.0 Da of the precursor ion. The edges from the IIN were added separately by the output file of MZmine. Information on the chemical structure was enhanced using different workflows within the GNPS environment. Network Annotation Propagation (job ID: 729d9f96b8614ba9a18d751 7926d2462) (da Silva et al., 2018; Kyo Bin Kang et al., 2018), Dereplicator (job ID:

ada96df6153e4b5e9be24423cc4ff760) (Mohimani et al., 2017) and MS2LDA_MotifDB (job ID: 16f297cc335644349f85440fc9af89d6) (van der Hooft et al., 2016; van der Hooft et al., 2017; Wandy et al., 2018; Rogers et al., 2019) were combined with the GNPS Library search and incorporated to the network using GNPS MolNetEnhancer (Ernst et al., 2019), assigning Chemical Class annotations using the ClassyFire (Djoumbou Feunang et al., 2016) chemical ontology (job ID: beee6b518aa54bc1aaaa7ba9d1f9ec1c). The network was downloaded into Cytoscape (Shannon et al., 2003) for further network visualization.

GNPS annotated 53 features based on MS2 spectra as known compounds by matching MS2 spectra to library spectra. Another 238 features were highly similar to MS2 spectra and marked as analog hits. Information on feature structures from the different workflows were integrated with MolNetEnhancer and resulted in 564 Classyfire annotations on the level of Kingdom and Superclass, 562 on class level, and 420 on subclass level. A total of 265 runs were used as input for MZmine, which detected 3355 features. Analysis on MS1 level was done on 31 samples and 24 blanks. Contaminant features were flagged and removed if the maximum peak area in one of the blanks was larger than half of the mean peak area of all samples. Peak area background noise level was set on 4929 as was the smallest peak area before the gap filling step in MZmine. Features needed to pass the peak area background noise level threshold in at least 2 samples. In addition, feature peak areas smaller than two-fold background noise (peak area of 10000) were set to 0, but features were kept. This resulted in 2315 features.

Statistical Analyses

For the 16S and 18S rDNA amplicon sequencing, the abundance tables were rarefied by subsampling the reads to account for library size and imported using QIIME 2R1. The transcriptome abundance data tables were taken using read hits counts normalized by the total number of hits per library, the metabolomic data table consists of areas under the curve. Maximum likelihood trees of 18S rDNA were built in IQ-TREE v1.5.4. (Nguyen et al., 2015). The spatial analysis was conducted using physical distances measured over the 3D coral's surface. The shortest 3D topological distance between two sample points was calculated using the ReCap® Photo 'measure distance' tool, and a pairwise physical distance matrix was generated for all samples within the coral colonies. The spatial patterns of the different data types were then analyzed using linear regression analyses in R version 3.4.4 (R Core Team, 2017) with ggplotRegression (Wickham and Chang, 2016), and power analyses were conducted using the R pwr package 1.3-0 (Champely et al., 2017).

All statistical analyses were conducted in R version 3.4.4 (R Core Team, 2017). Abundance data tables of each data type were manipulated with dplyr and tidyverse and used to create Bray-Curtis dissimilarity matrices with the Vegan package 2.5-2 (Oksanen et al., 2016). Additionally, the Vegan and pairwise Adonis packages were used to generate PERMANOVA test statistics on the Bray-Curtis dissimilarity matrices and provided pairwise comparisons between variables. The Principle Coordinates and eigenvectors were generated by the Ape package 5.4-1 with a Cailliez correction for negative eigenvalues (Paradis and Schliep, 2019) and the Non-metric Multi-dimensional Scaling (NMDS) analysis was conducted using the metaMDS function of the Vegan package 2.5-2 (Oksanen et al., 2016). Principal Coordinate Analysis coordinates with eigenvectors and NMDS were plotted with ggplot2 (Wickham and

Chang, 2016), and heatmaps of relative abundances were plotted using the heatmap function in R. Supervised and unsupervised random forests analyses were run using rfPermute 2.1.81 (Archer, 2016), and 3,000 trees were built using a set seed of 25. Shannon entropy and rarefaction species richness was calculated using the Vegan package 2.5-2. F-Tests (two-sample for variances) and t-Tests (two-sample assuming equal variance or assuming unequal variance) were used for diversity metrics.

Co-occurrence Analysis

To integrate the different datasets, the mmvec tool3 was used to calculate conditional probabilities of the occurrence of transcriptomes or metabolites with the occurrence of specific bacteria on a phylum level (Morton et al., 2019). The standalone version of mmvec was used and co-occurrence probabilities were extracted by applying a softmax transformation. The conditional probabilities of the top predictors were visualized in a two way heatmap using JMP (JMP, Version 14. SAS Institute Inc., Cary, NC, United States, 1989-2021).

Ubiquity Calculations

The methods from Hester et al., 2016 were used to calculate the ubiquity and relative abundance of the 16S and 18S rDNA ASVs, bacterial transcripts, and metabolites. Briefly, the ubiquity of an individual ASV/transcript/metabolite was determined by the number of samples the ASV/transcript/metabolite was present in divided by the total number of samples. For example, a ubiquity of 1 resulted from the ASV/transcript/metabolite appearing in 100% of the samples. The relative abundance was calculated as a proportion of the entire community (e.g.,

proportion a specific ASV/transcript/metabolite to all ASVs/transcripts/metabolites from all samples). Ubiquity-abundance was plotted using ggplot2 (Wickham and Chang, 2016).

Figure and Tables

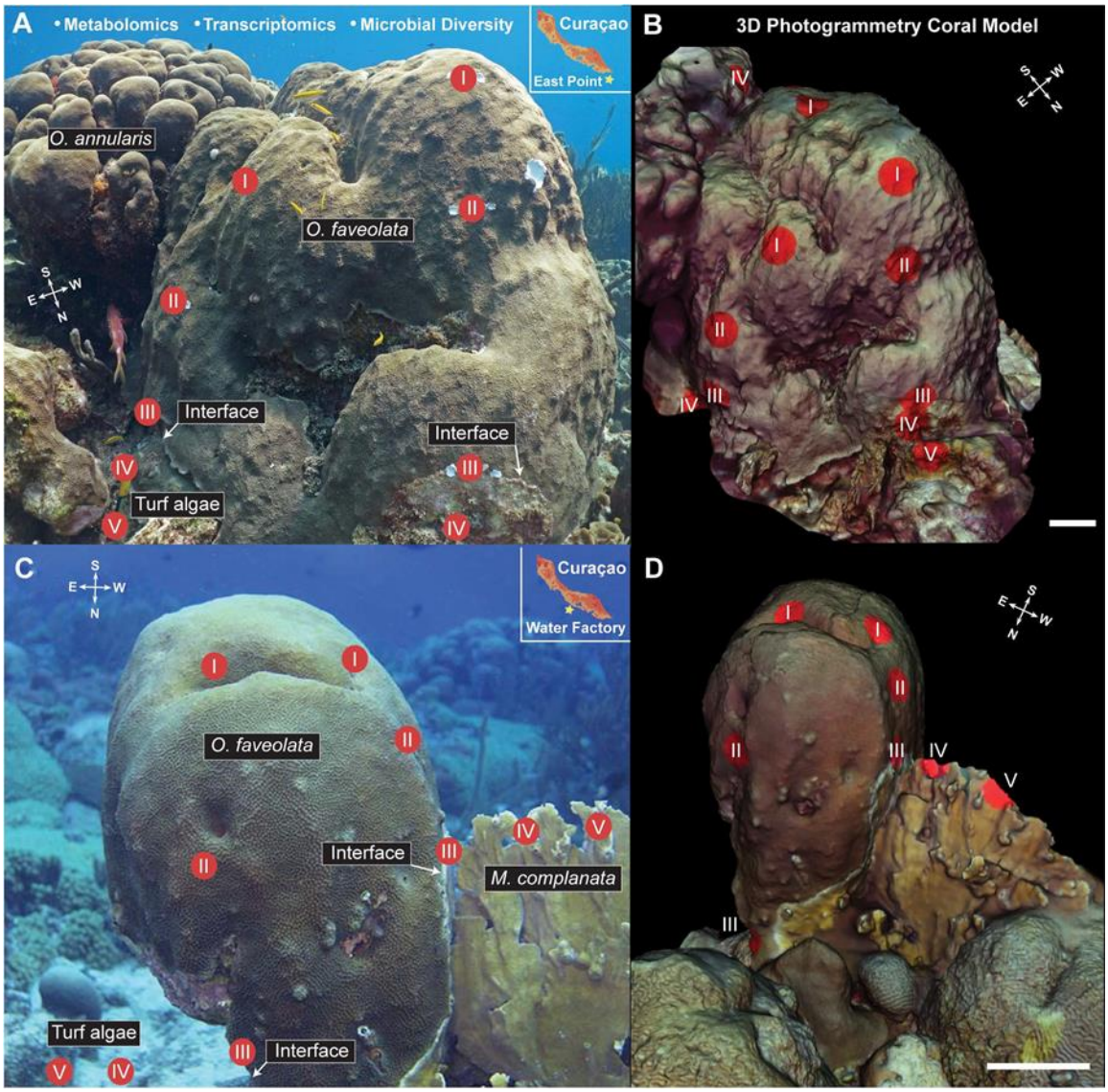


Figure 3.1 Images (A,C) and 3D models (B,D) of the *Orbicella faveolata* corals with sampling points overlaid for visualization. Inset in panels (A,C) show the geographic location of the sampling efforts. Compass insets indicate the orientation of the coral colonies for cardinal direction. Three samples were taken from each sampling point (I-V), one for 16S and 18S rDNA amplicon sequencing, one for metatranscriptomic sequencing, and one for LC MS/MS metabolomic analysis for a total of 60 samples (24 coral, 12 interface, and 24 competitor) taken at each site (Water factory and East Point). Models were created using Autodesk ReCap Photo and 3D molecular mapping was performed using ili (scale bar = 20 cm). White arrows indicate the interface between *O. faveolata* and the interacting competitors (turf algae, *Orbicella annularis*, *Millepora complanata*). White arrows indicate the interface between *O. faveolata* and the interacting competitors (turf algae, *Orbicella annularis*, *Millepora complanata*).

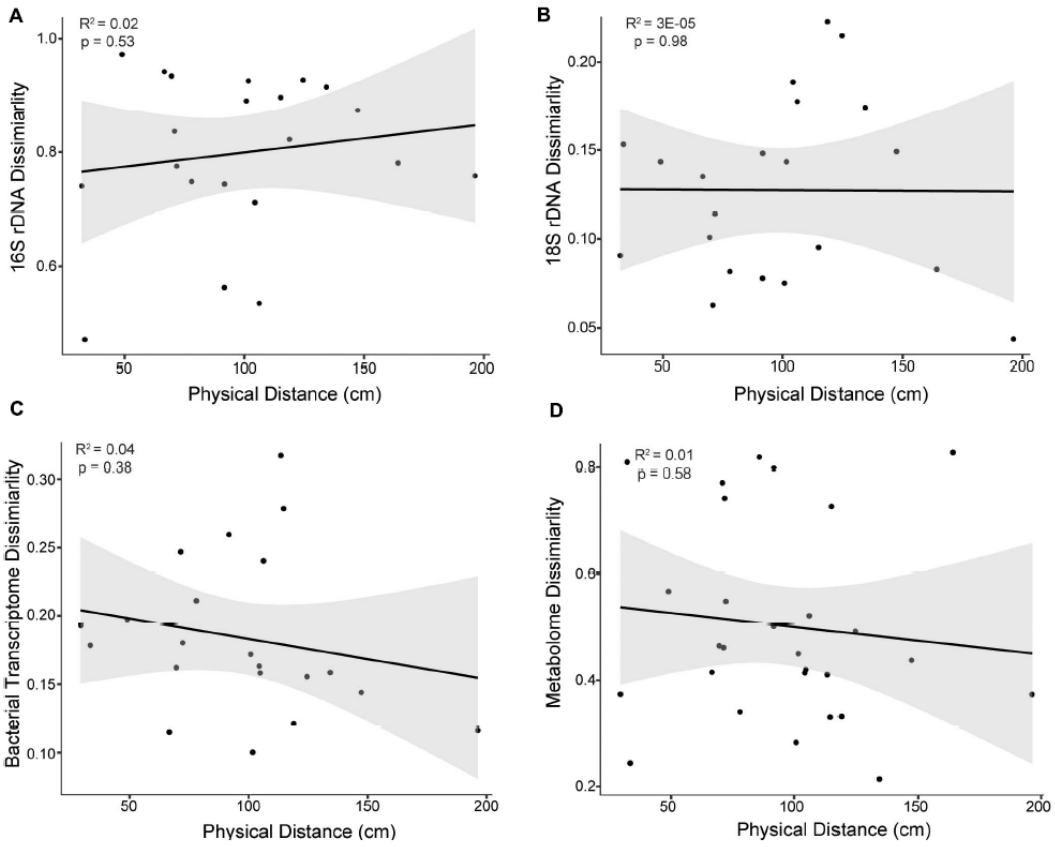


Figure 3.2 Bray-Curtis dissimilarity of (A) 16S rDNA, (B) 18S rDNA, (C) bacterial transcriptomes, and (D) metabolomes as a function of the physical distance between samples of the East Point *Orbicella faveolata* colony. The physical distance represents the shortest distance between two sample points along the 3D coral model surface. No significant correlations were observed ($p > 0.05$) and similar results were found for the Water Factory *O. faveolata* colony. Black lines represent the fit line and gray areas represent the fit confidence (95%).

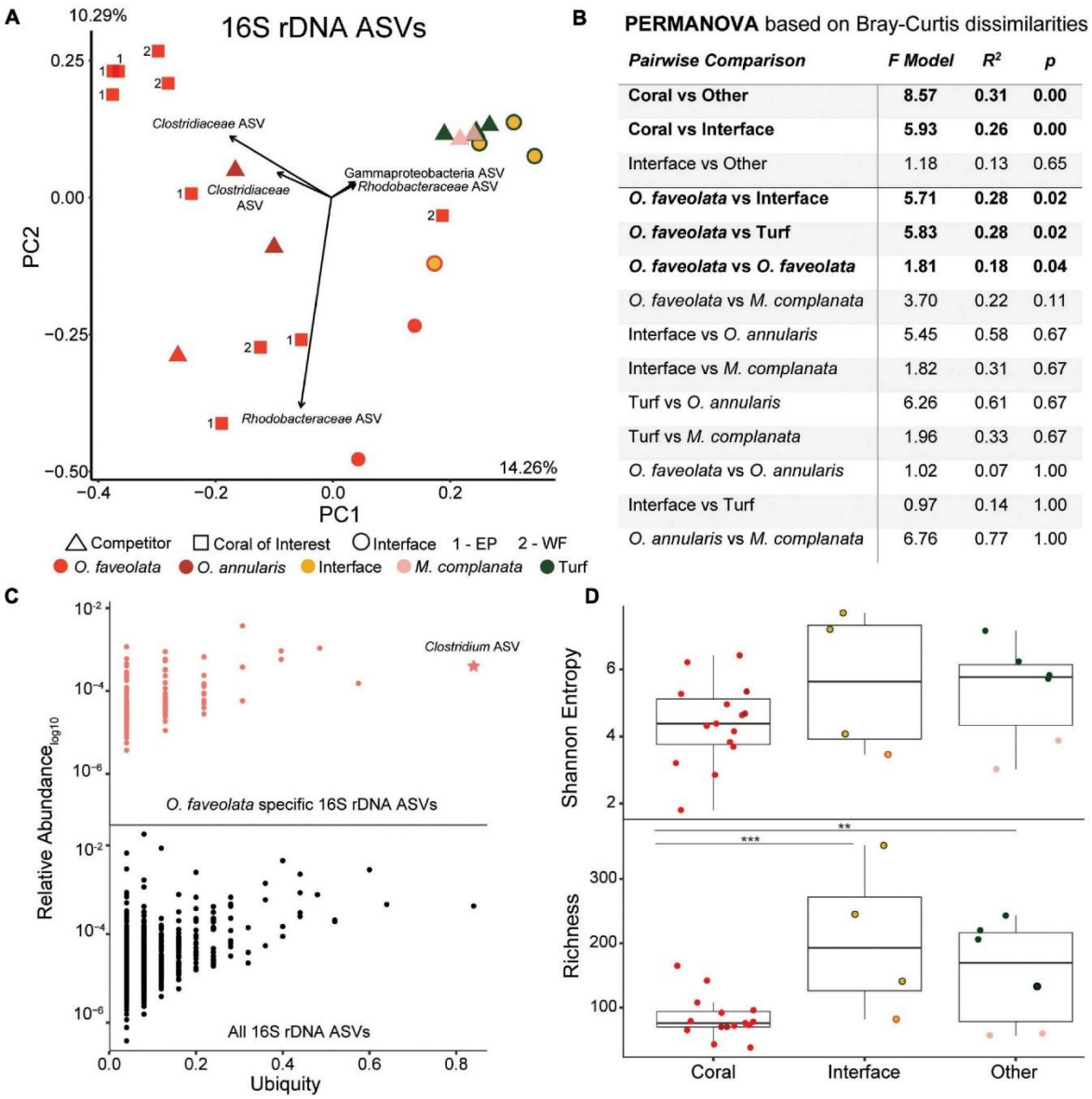


Figure 3.3 Overview of the 16S rDNA amplicon data. (A) Principal components plot of the 16S rDNA ASVs from the coral colonies of interest along with interfaces and their respective benthic competitors. EP, East Point coral of interest; WF, Water Factory coral of interest. (B) PERMANOVA results based on Bray-Curtis dissimilarities. (C) Relative abundance (proportion of the entire community) versus ubiquity of each ASV from all samples (bottom plot) and those unique to *O. faveolata* samples (top plot). Ubiquity for unique *O. faveolata* ASVs is normalized to total *O. faveolata* samples. The star represents the most ubiquitous ASV found only in *O. faveolata* (i.e., a taxa that is uniquely ubiquitous in *O. faveolata*). (D) Diversity metrics of coral, interface and other (turf algae and *M. complanata*) samples. Interface and other samples had significantly higher bacterial richness than coral samples (** $p \leq 0.01$, *** $p \leq 0.001$).

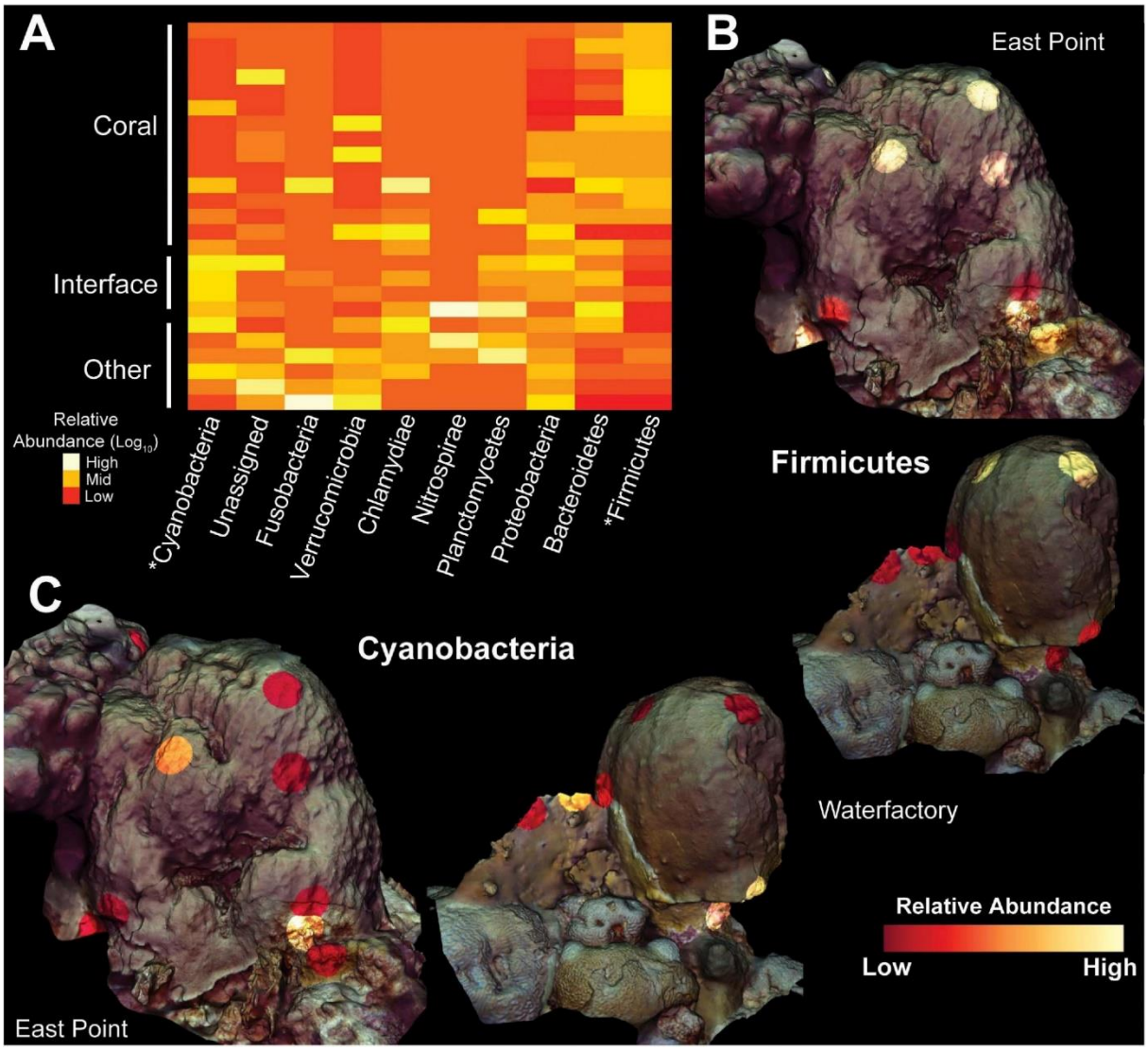
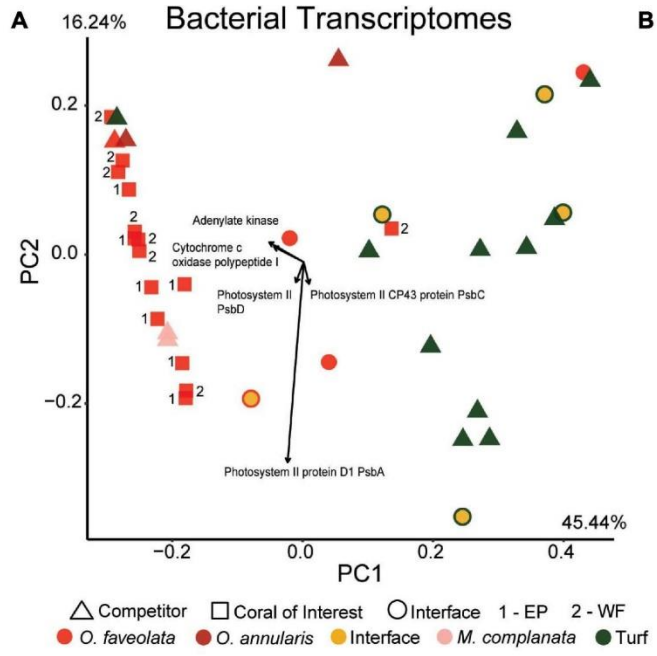


Figure 3.4 Heatmap and 3D cartography of bacterial phyla from *Orbicella faveolata* colonies, competitors and interfaces. (A) Heatmap of phyla selected from the supervised random forests analysis. Relative abundances are log-transformed and colors are normalized to columns with red representing the lowest relative abundance and white representing the highest. Asterisks highlight phyla shown in 3D models. 3D coral models mapped with the relative abundances of (B) Firmicutes and (C) Cyanobacteria from both sampling sites. The relative abundance of Firmicutes ranges from 0.21–93.5% at East Point and 0.0–89.7% at Water Factory. For Cyanobacteria, the relative abundance ranges from 0.0–14.4% at East Point and 0.0–0.49% at Water Factory.



B

PERMANOVA based on Bray-Curtis dissimilarities

Pairwise Comparison	*F Model*	R^2	*p*
Coral vs Other	**14.28**	**0.31**	**0.00**
Coral vs Interface	**8.80**	**0.27**	**0.01**
Other vs Interface	0.10	0.01	1.00
***O. faveolata* vs Turf**	**23.62**	**0.46**	**0.01**
***O. faveolata* vs Interface**	**10.50**	**0.32**	**0.03**
Turf vs *M. complanata*	7.01	0.39	0.19
O. faveolata vs *O. faveolata*	1.07	0.03	0.29
Turf vs *O. annularis*	3.89	0.26	0.39
O. faveolata vs *O. annularis*	2.75	0.13	0.72
Interface vs *M. complanata*	4.17	0.45	0.92
Interface vs *O. annularis*	2.15	0.30	1.00
O. annularis vs *M. complanata*	1.42	0.41	1.00
O. faveolata vs *M. complanata*	0.40	0.02	1.00
Turf vs Interface	0.22	0.02	1.00

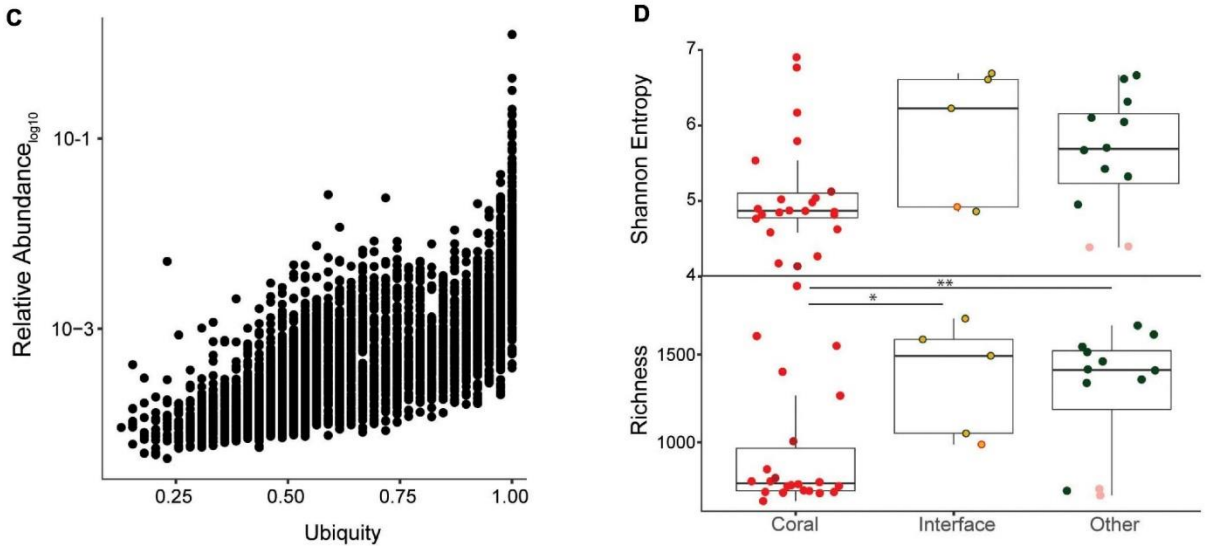


Figure 3.5 Overview of the bacterial metatranscriptomes at the SEED functional level. (A) Principal components plot of the coral colonies of interest along with interfaces and their respective benthic competitors. EP, East Point coral of interest; WF, Water Factory coral of interest. (B) PERMANOVA results based on Bray-Curtis dissimilarities. (C) Relative abundance as a function of ubiquity of each bacterial transcript from all samples. The majority of bacterial transcripts were found in most samples (ubiquitous), and no *Orbicella faveolata* specific bacterial transcripts were observed. (D) Diversity metrics of coral, interface and other (turf algae and *Millepora complanata*) samples at the SEED functional level. Richness of bacterial transcripts was significantly higher in interface and other samples than in corals (* $p \leq 0.05$, ** $p \leq 0.01$).

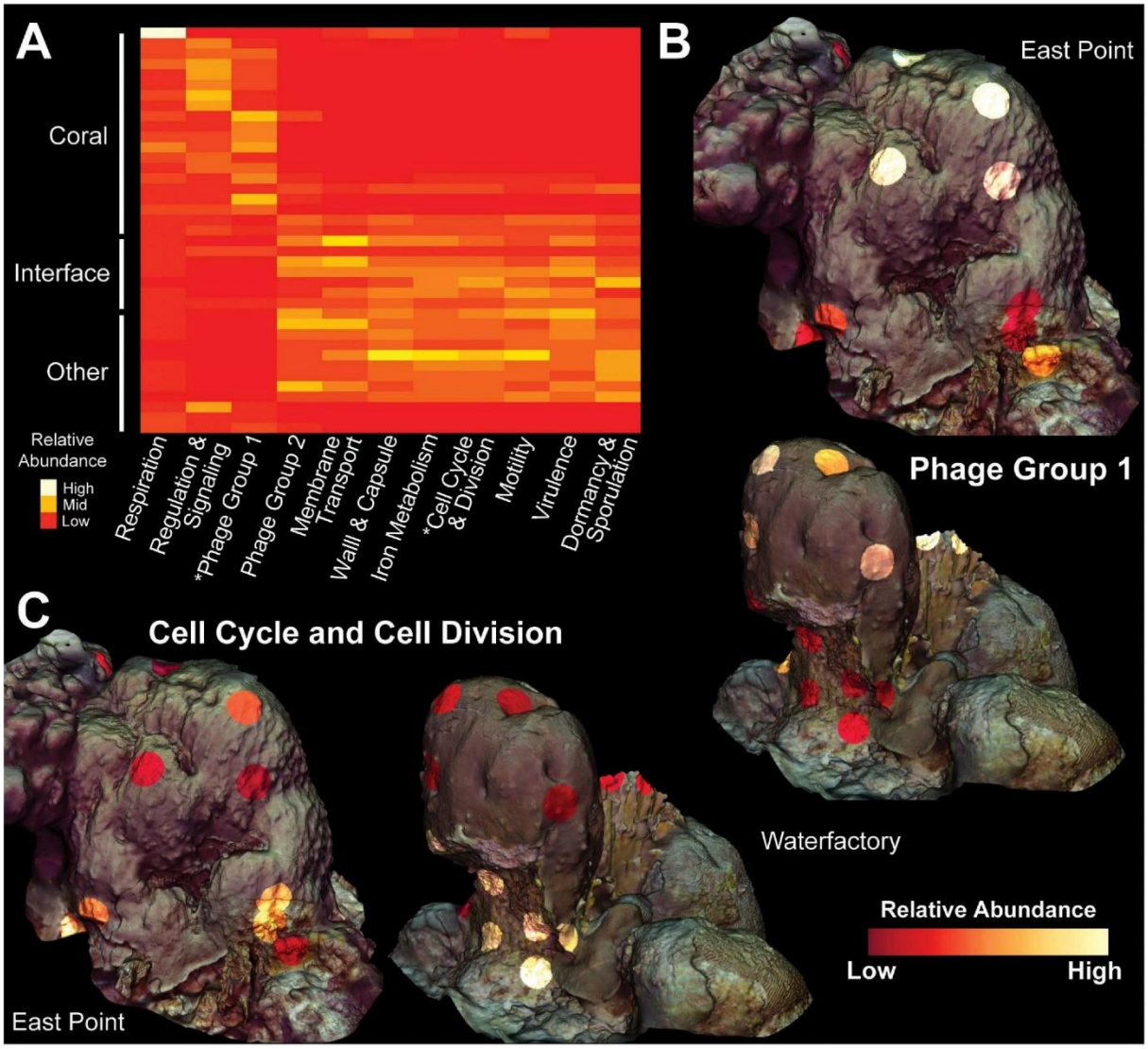
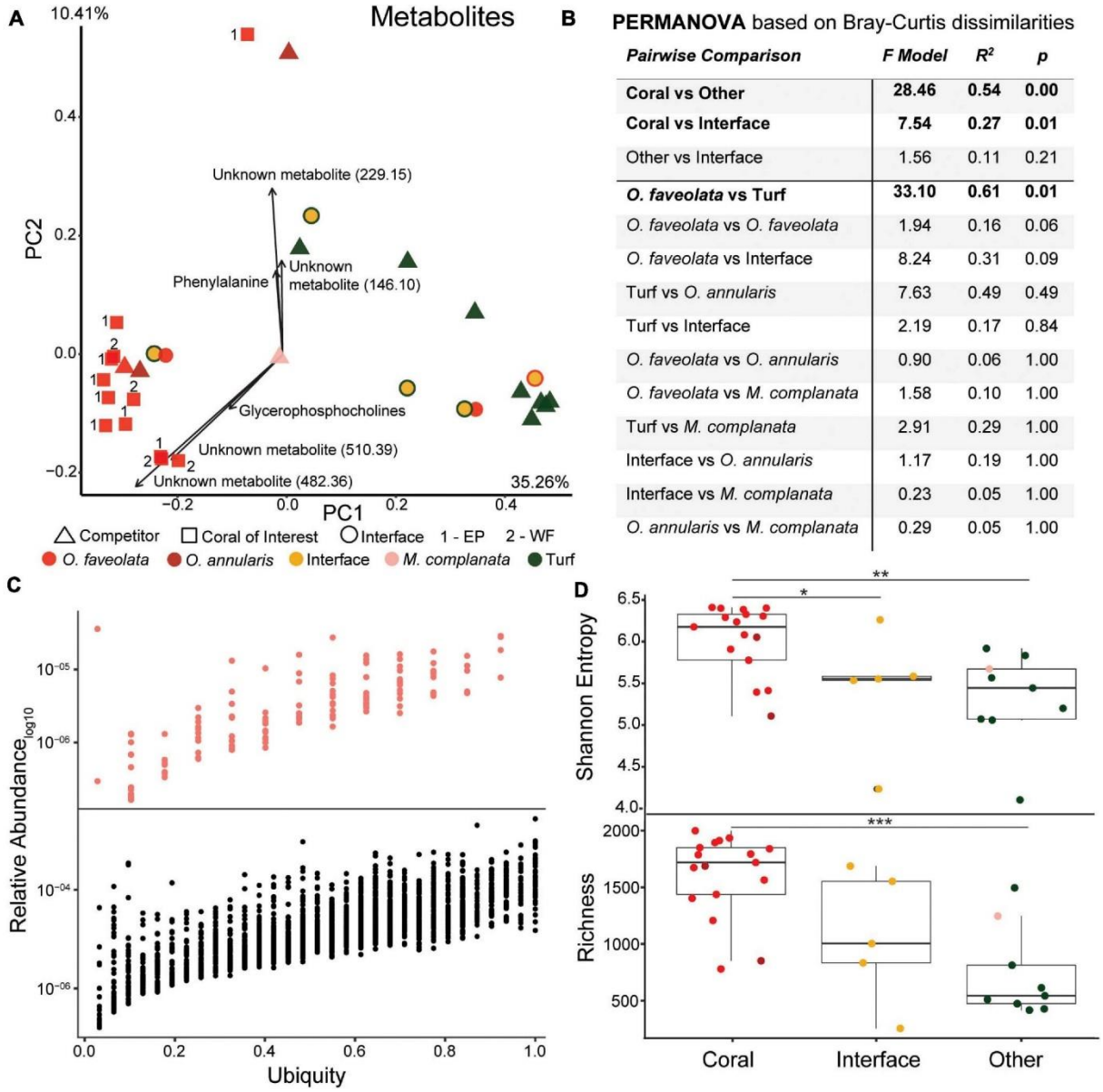


Figure 3.6 Molecular heatmap of bacterial metatranscriptomes from *Orbicella faveolata* colonies, competitors and interfaces. (A) Heatmap of SEED Level 1 categories selected from the supervised random forests analysis. Phage Group 1 includes phages, prophages and transposable elements while Phage Group 2 also includes plasmids. Colors are normalized to columns with red representing the lowest relative abundance and white representing the highest. Asterisks highlight Level 1 categories shown in 3D models. 3D coral models mapped with the relative abundance of bacterial transcripts belonging to (B) Phage Group 1 and (C) Cell Cycle and Cell Division at both sampling sites. The relative abundance of Phage Group 1 transcripts range from 0.06–0.74% at East Point and 0.002–0.42% at Water Factory. For Cell Cycle and Cell Division, the relative abundances range from 0.03–0.53% at East Point and 0.02–0.28% at Water Factory.



Pairwise Comparison	*F Model*	R^2	*p*
Coral vs Other	**28.46**	**0.54**	**0.00**
Coral vs Interface	**7.54**	**0.27**	**0.01**
Other vs Interface	1.56	0.11	0.21
***O. faveolata* vs Turf**	**33.10**	**0.61**	**0.01**
O. faveolata vs *O. faveolata*	1.94	0.16	0.06
O. faveolata vs Interface	8.24	0.31	0.09
Turf vs *O. annularis*	7.63	0.49	0.49
Turf vs Interface	2.19	0.17	0.84
O. faveolata vs *O. annularis*	0.90	0.06	1.00
O. faveolata vs *M. complanata*	1.58	0.10	1.00
Turf vs *M. complanata*	2.91	0.29	1.00
Interface vs *O. annularis*	1.17	0.19	1.00
Interface vs *M. complanata*	0.23	0.05	1.00
O. annularis vs *M. complanata*	0.29	0.05	1.00

Figure 3.7 Overview of metabolome data. (A) Principal components plot of the coral colonies of interest along with interfaces and their respective benthic competitors. Main driving metabolites shown along with mass to charge ratios for unknown metabolites. EP, East Point coral of interest; WF, Water Factory coral of interest. (B) PERMANOVA results based on Bray-Curtis dissimilarities. (C) Relative abundance versus ubiquity of each metabolite from all samples (bottom plot) and those unique to *Orbicella faveolata* samples (top plot). Ubiquity for unique *O. faveolata* metabolites is normalized to total *O. faveolata* samples. (D) Diversity metrics of coral, interface and other (turf algae and *M. complanata*) samples (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$).

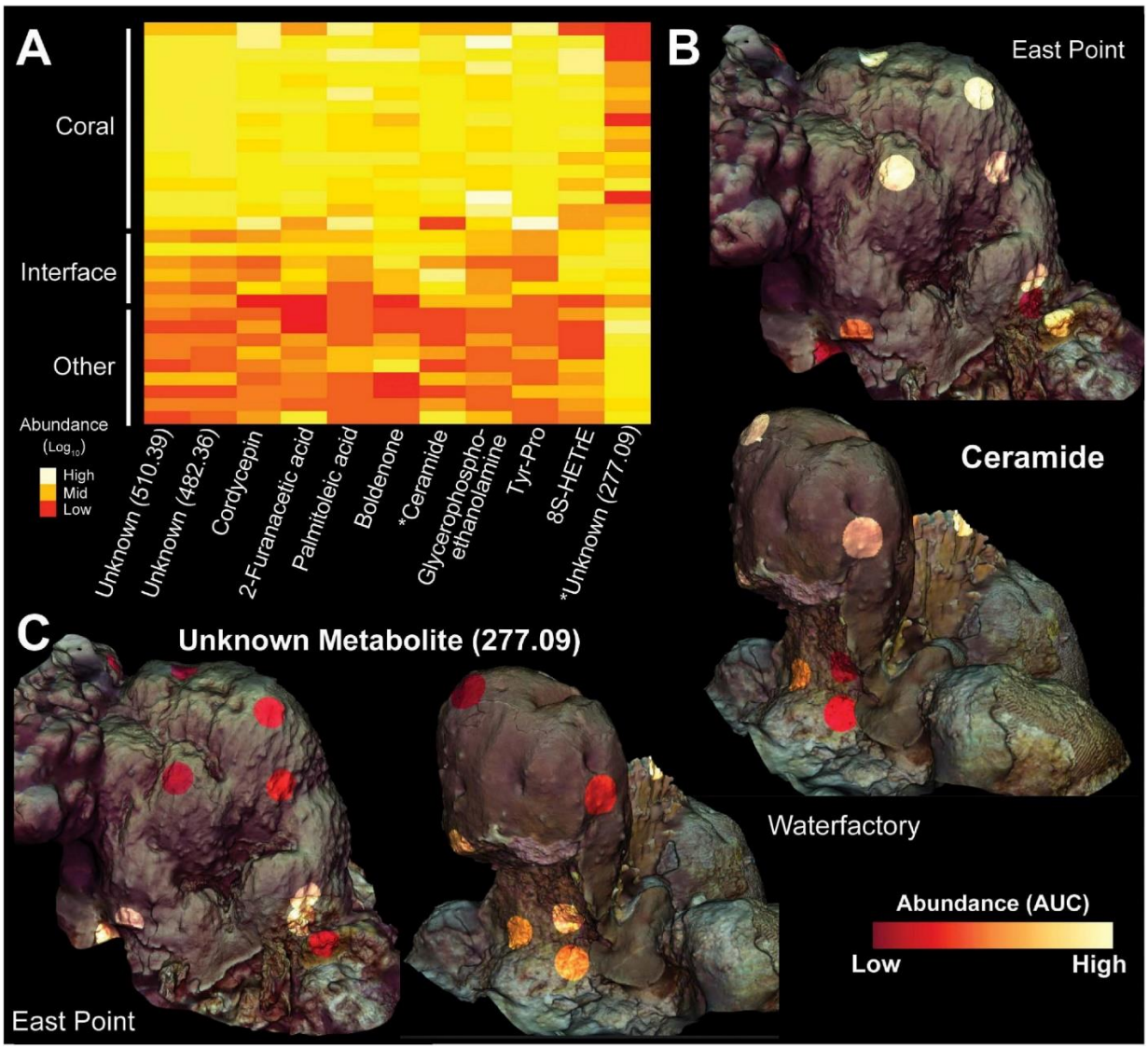


Figure 3.8 Heatmap and 3D cartography of metabolites from *Orbicella faveolata* colonies, competitors and interfaces. (A) Heatmap of metabolites selected from the supervised random forests analysis. Mass to charge ratios are shown for unknown metabolites. Abundances (Area Under the Curve/AUC) are log-transformed and colors are normalized to columns with red representing the lowest abundance and white representing the highest. Asterisks highlight metabolites shown in the 3D models. 3D coral models mapped with the abundances of (B) ceramide and (C) an unknown metabolite (277.09) from both sampling sites. Ceramide abundance ranges from 0-124,247 AUC at East Point and 0-69,148 AUC at Water Factory. For the unknown metabolite (277.09), the abundance ranges from 0-338,137 AUC at East Point and 0-137,151 AUC at Water Factory.

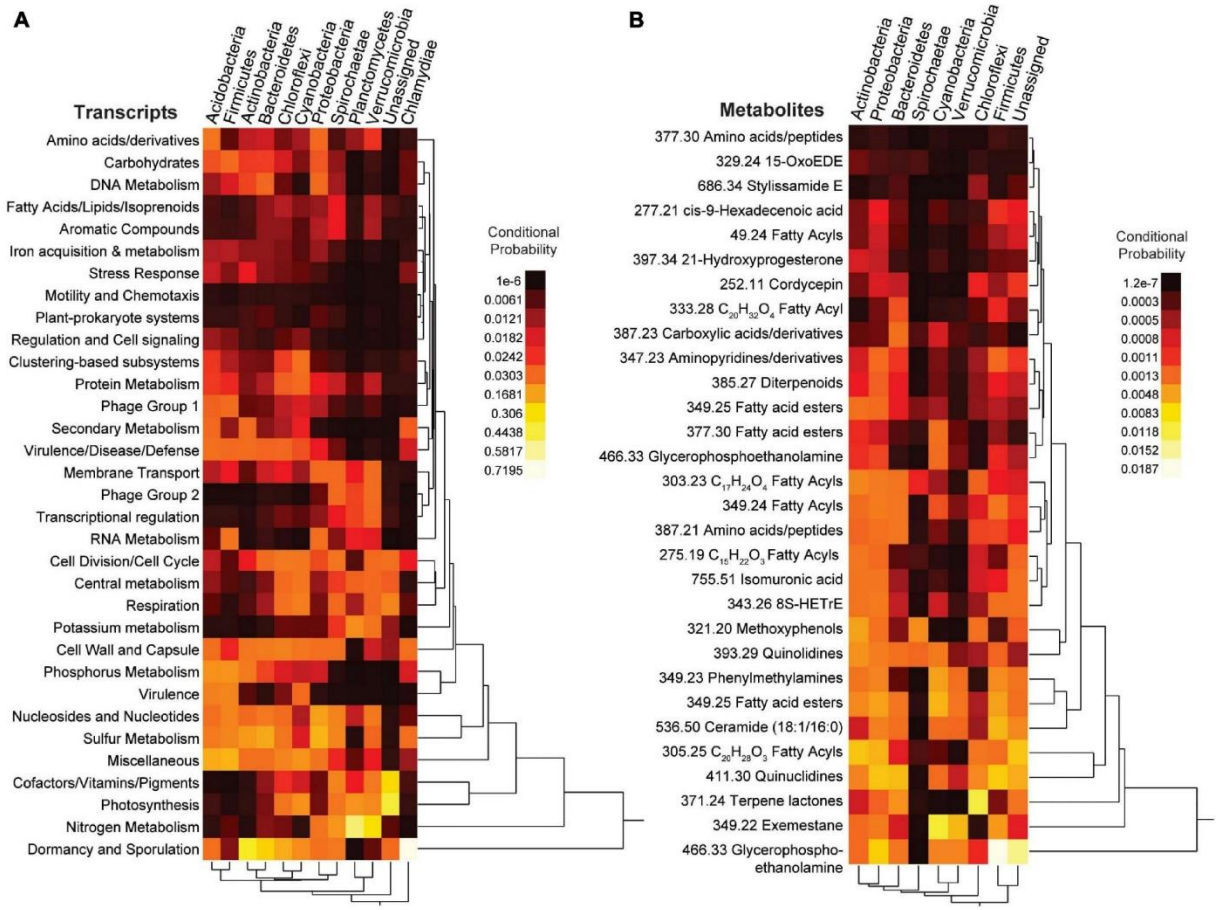


Figure 3.9 Heatmaps from the neural network (mmvec) analysis. The conditional probability (dark red/black for low and yellow/white for high) that the presence of a transcript or metabolite corresponds to the presence of a specific bacterial phylum is shown. (A) Top SEED Level 1 transcript predictors from supervised random forest analysis; (B) Top known metabolite predictors from supervised random forest analysis. For metabolites the number in front of the annotation is the mass to charge ratio.

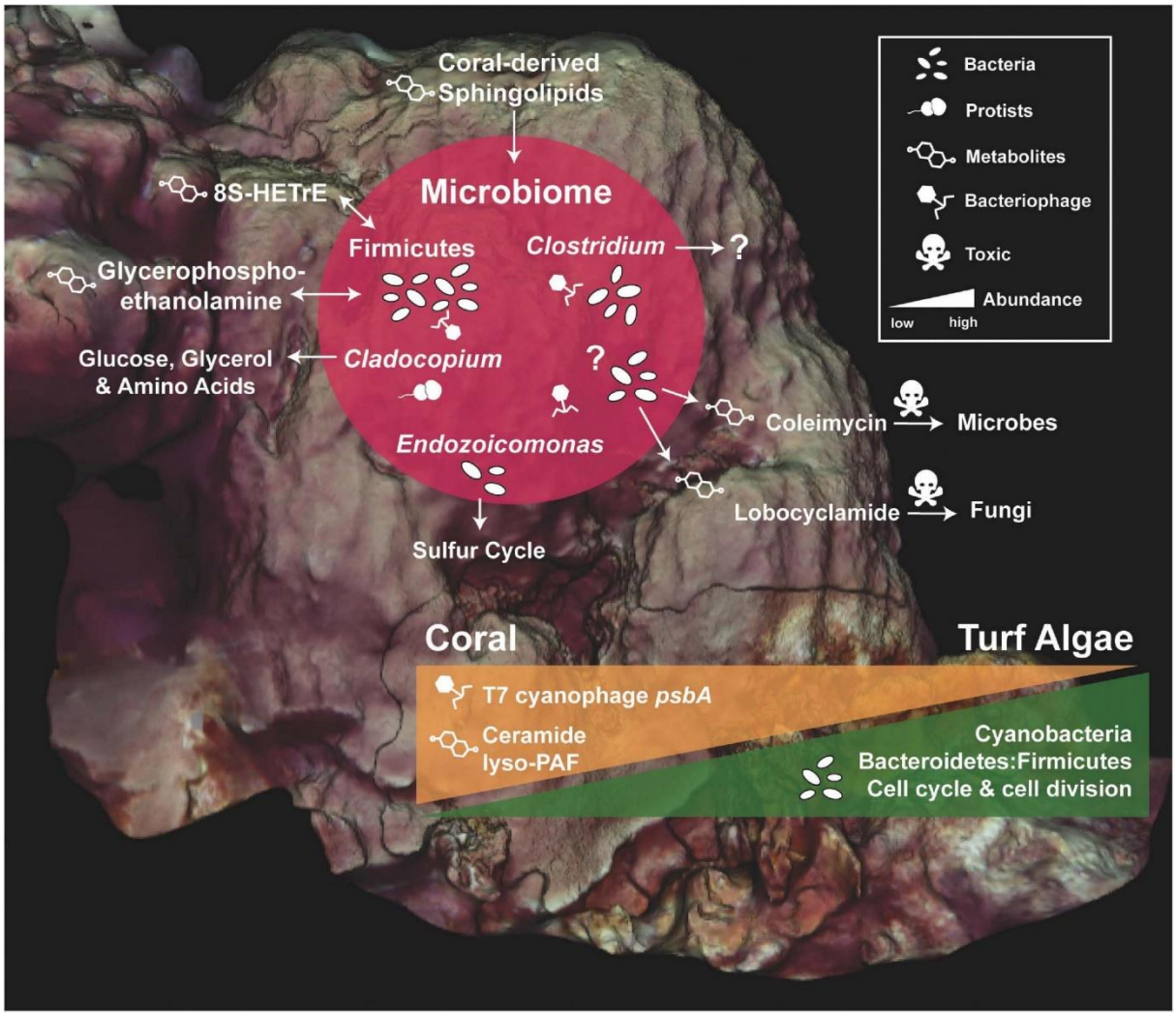
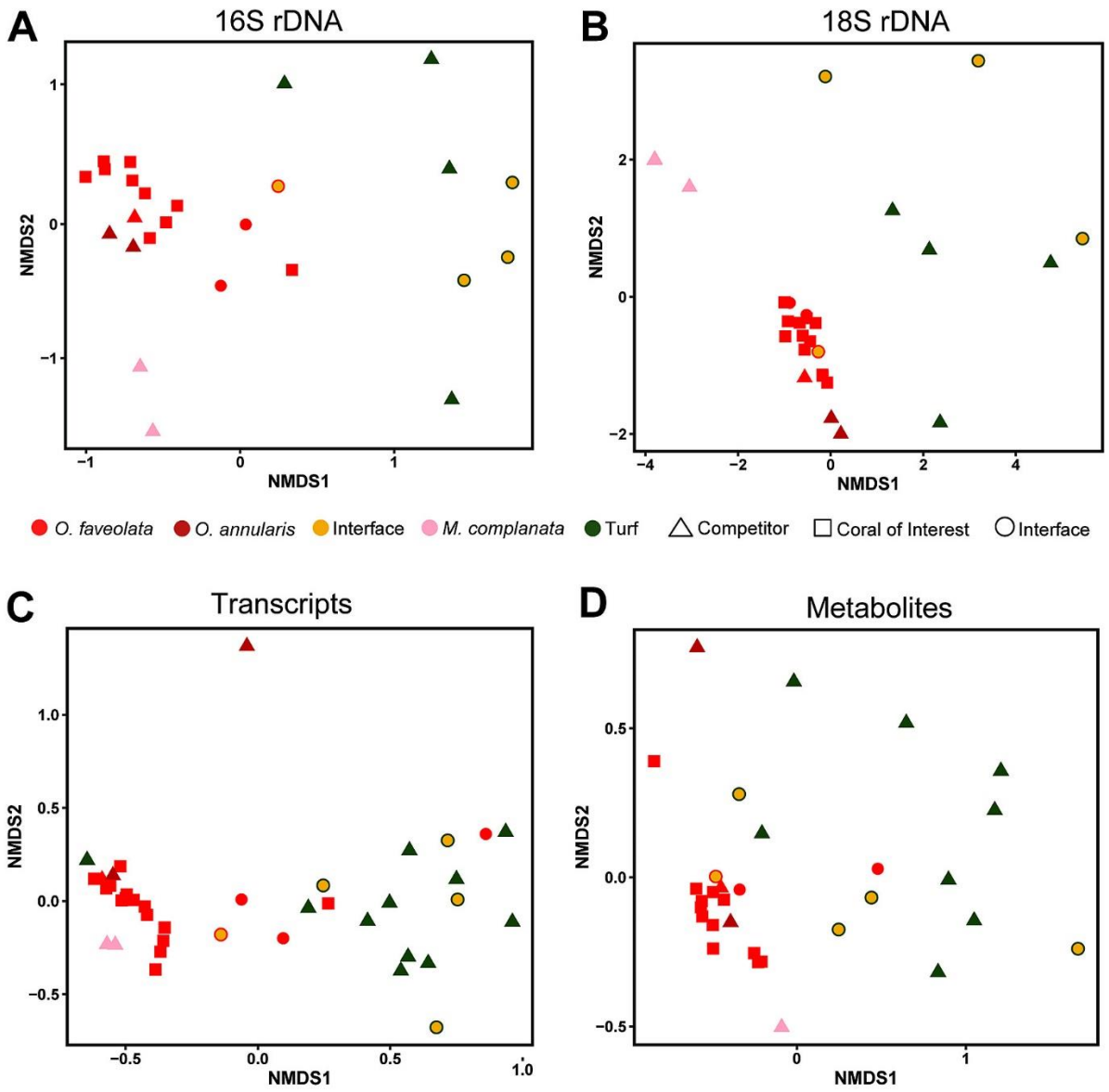
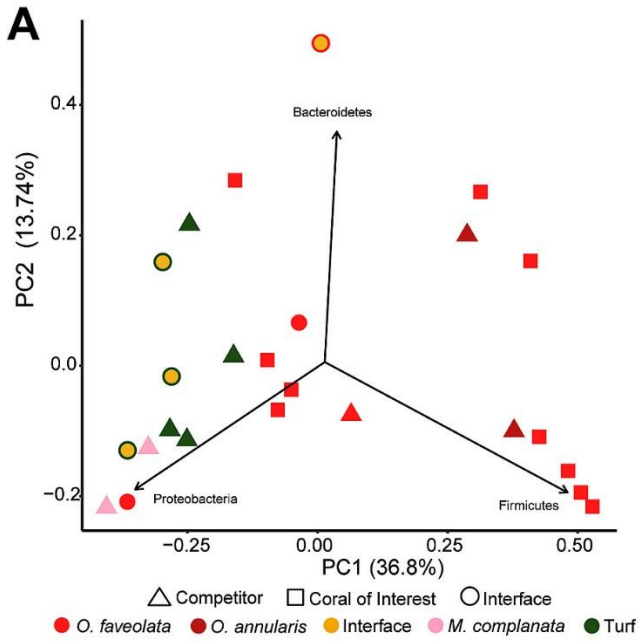


Figure 3.10 Summary of the *Orbicella faveolata* holobiont and its main competitor, turf algae. Core microbial members of *O. faveolata* are shown along with their potential functions and interactions with the coral host (red circle). Question marks represent functions that are unknown. The change in abundance of specific bacterial groups, bacterial and phage functions, and metabolites across coral and turf algae competitors are illustrated by orange and green triangles.

Supplemental Figures and Tables

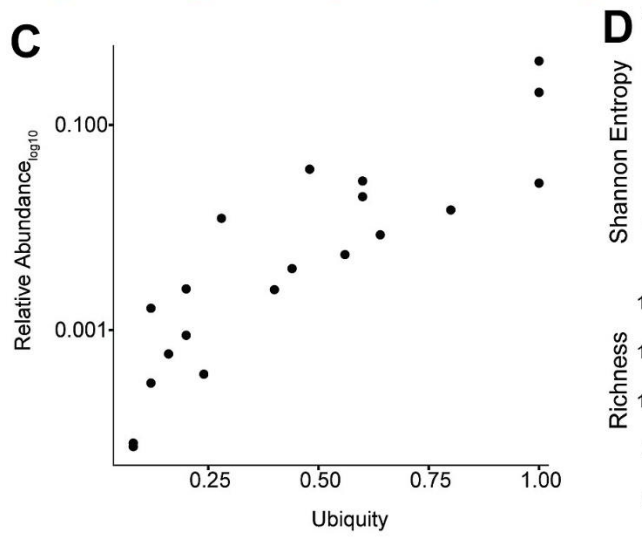


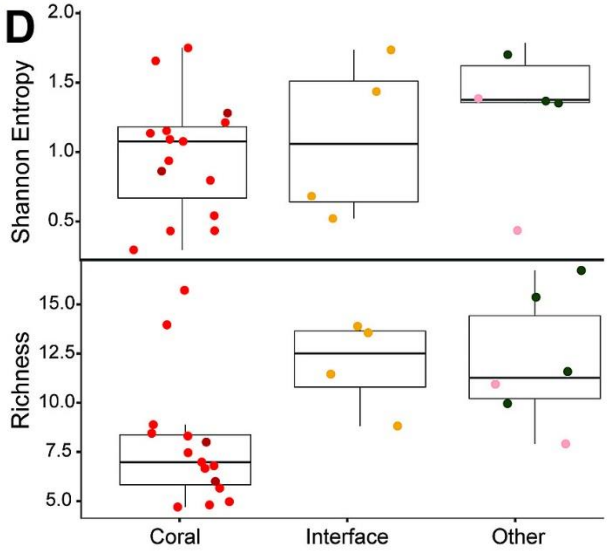
Supplemental Figure 3.1 Non-metric Multi-dimensional Scaling (NMDS) plot of coral colonies of interest along with interfaces and their respective benthic competitors. Stress values for all datasets are less than 2. A/ 16S rDNA. B) 18S rDNA. C) Transcripts at the SEED functional level. D) Metabolites.



B **PERMANOVA** based on Bray-Curtis dissimilarities

Pairwise Comparison	*F Model*	R^2	*p value*
Coral vs Other	**5.62**	**0.23**	**0.04**
Coral vs Interface	3.63	0.18	0.15
Other vs Interface	0.32	0.04	1.00
Interface vs *O. annularis*	7.65	0.66	0.67
Turf vs *O. annularis*	14.06	0.78	0.67
Turf vs *M. complanata*	2.13	0.35	0.67
O. faveolata vs Turf	2.95	0.16	0.83
O. faveolata vs Interface	2.90	0.16	0.85
O. faveolata vs *M. complanata*	2.06	0.14	1.00
O. annularis vs *M. complanata*	15.57	0.89	1.00
O. faveolata vs *O. annularis*	0.89	0.06	1.00
Interface vs *M. complanata*	0.81	0.17	1.00
Interface vs Turf	0.29	0.05	1.00



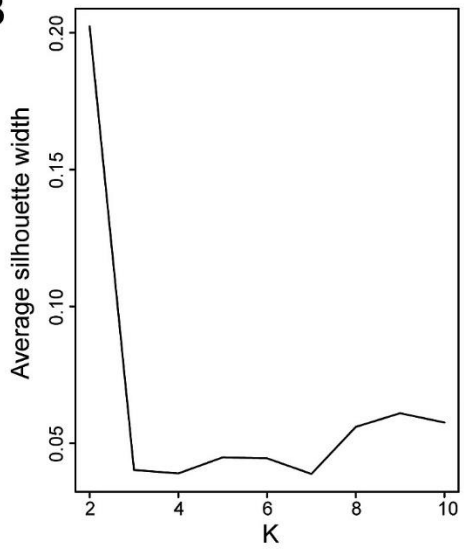


Supplemental Figure 3.2 Overview of the 16S rDNA amplicon data at the phylum level. A) Principal components plot of the coral colonies of interest along with interfaces and their respective benthic competitors. B) PERMANOVA results based on Bray-Curtis dissimilarities. C) Relative abundance (proportion of the entire community) versus ubiquity of each phylum from all samples. D) Diversity metrics of coral, interface and other (turf algae and *M. complanata*) samples.

A

16S rDNA ASV	Taxa	Mean Decrease in Accuracy	*p*
310c227273d6fa9190bc6d8fde2b3765	Firmicutes; *Clostridium* (*sensu stricto*)	7.48	0.01
6885939305af4557209bfa56038018c0	Firmicutes; Clostridia	7.48	0.02
9fcfa7e555c6dabbe363302dfdc4c56c	Bacteroidetes; *Bacteroides*	7.37	0.01
538a8597c801e549ee3ab703923a9b46	Proteobacteria; Marine Methylotrophic Group 3	6.57	0.02
16c3ebddb379253a6c3b6dec50884083	Proteobacteria; *Thalassospira*	6.43	0.02
d6d6d57d15072e10d222a16f3e6d97e9	Firmicutes, uncultured Clostridia bacterium	6.14	0.03
95b071fbe5e82af9043aa871f47f2366	Firmicutes, *Ruminococcaceae*	6.11	0.04
e84269729590c7dd67bb58ffb1305533	Firmicutes; *Clostridium* (*sensu stricto*)	5.85	0.02
372ec76d916ddd3cfb4aa3488ea70471	Proteobacteria; *Ruegeria*	5.68	0.01
a308ae306a4bff2c3e9323544c69a9f9	Actinobacteria; *Acidimicrobiales*	5.64	0.01

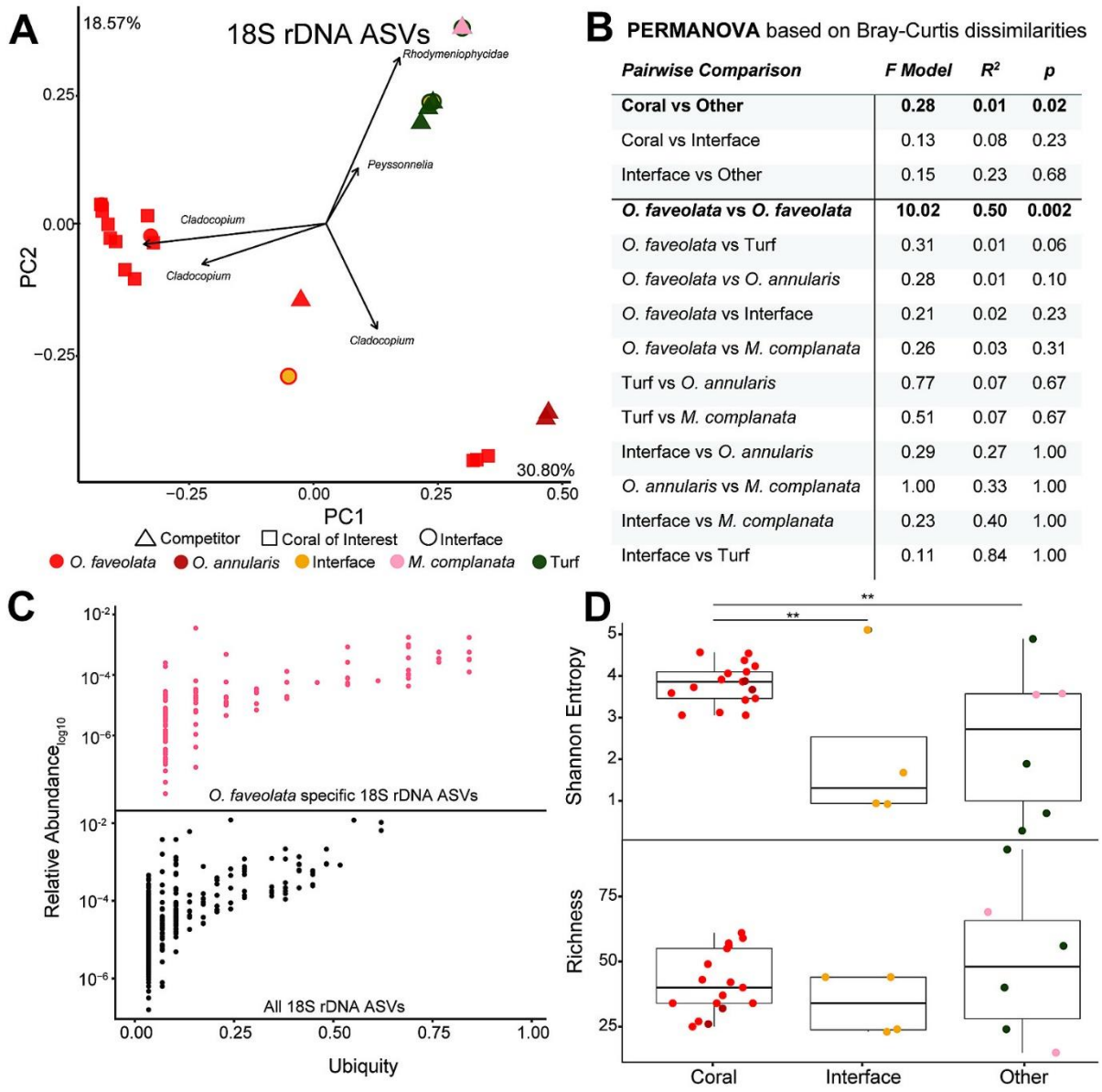
B



C

16S rDNA ASV/Taxa	Ubiquity	Relative Abundance (%)
d6d6d57d15072e10d222a16f3e6d97e9 Firmicutes, *Clostridium*	0.91	0.42
b79f35ceecada3be3bfbd68ccd383ab3 Bacteroidetes	0.64	0.15
d2fa4913f74bf3c73d266f1f180efce3 CFB group	0.55	1.08
193b8a5be3a5fb21ef6ec98483feef31 CFB group	0.45	0.58
c26c7b807c4f7d21f4ee4ad42873d37f Firmicutes, *Halanaerobiaceae*	0.45	0.92

Supplemental Figure 3.3 Random Forest and ubiquity metrics for 16S rRNA amplicon data. A) Top ten ASV predictors of coral, interface and other categories from the supervised random forest. Taxonomy shown at the phylum level and at the lowest taxonomic rank possible based on Qiime2 taxonomy and BLAST analysis. B) silhouette plot from the unsupervised random forest. C) The most abundant, uniquely ubiquitous ASVs found in and only *O. Faveolata* samples. Ubiquity ranges from 0-1.



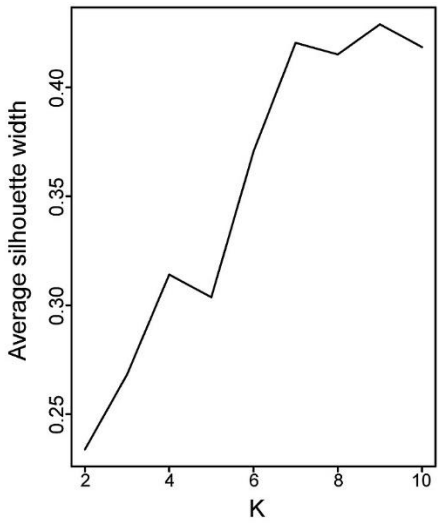
Pairwise Comparison	*F Model*	R^2	*p*
Coral vs Other	**0.28**	**0.01**	**0.02**
Coral vs Interface	0.13	0.08	0.23
Interface vs Other	0.15	0.23	0.68
O. faveolata* vs *O. faveolata	**10.02**	**0.50**	**0.002**
O. faveolata vs Turf	0.31	0.01	0.06
O. faveolata vs O. annularis	0.28	0.01	0.10
O. faveolata vs Interface	0.21	0.02	0.23
O. faveolata vs *M. complanata*	0.26	0.03	0.31
Turf vs *O. annularis*	0.77	0.07	0.67
Turf vs *M. complanata*	0.51	0.07	0.67
Interface vs *O. annularis*	0.29	0.27	1.00
O. annularis vs *M. complanata*	1.00	0.33	1.00
Interface vs *M. complanata*	0.23	0.40	1.00
Interface vs Turf	0.11	0.84	1.00

Supplemental Figure 3.4 Overview of the 18S rDNA amplicon data. A) Principal components plot of the 18S rRNA gene ASVs from the coral colonies of interest along with interfaces and their respective benthic competitors. B) PERMANOVA results based on Bray-Curtis dissimilarities. C) Relative abundance (proportion of the entire community versus ubiquity of each ASV from all samples (bottom plot) and unique to *O. faveolata* samples (top plot). Ubiquity for the unique *O. faveolata* ASVs is normalized to total *O. faveolata* samples. D) Diversity metrics of coral, interface and other (turf algae and *M. complanata*) samples (**p ≤ 0.01, ***p ≤ 0.001)

A

18S rDNA ASV	Taxa	Mean Decrease in Accuracy	*p*
4ce3d13cde6a652c48857d29e44de316	*Symbiodiniaceae*, *Cladocopium*	9.95	0.01
e3442b2d10018cd51a2dc2699d2e1914	*Symbiodiniaceae*, *Cladocopium*	9.93	0.01
994e0d0b44df33b0218f107f20d2e1e2	*Symbiodiniaceae*, *Cladocopium*	9.78	0.01
e04926661deeab113372cb2744adf05c	*Symbiodiniaceae*, *Cladocopium*	8.33	0.02
50101780dc5cf690ad2c4977bb912e4a	*Symbiodiniaceae*, *Cladocopium*	8.01	0.01
ad1633a62e7ac0711e123377ebba520b	*Symbiodiniaceae*, *Cladocopium*	7.93	0.03
963ffbcd475118907839071c5524b0e9	*Symbiodiniaceae*, *Cladocopium*	7.91	0.01
519106a9462ca47b5155c8bb60ab84e0	*Symbiodiniaceae*, *Cladocopium*	7.85	0.01
3461e61802d934135e4d41deaa152a57	*Symbiodiniaceae*, *Cladocopium*	7.80	0.01
979ea08afa1d54ded7cc3feece5ca8c5	*Symbiodiniaceae*, *Cladocopium*	7.66	0.02

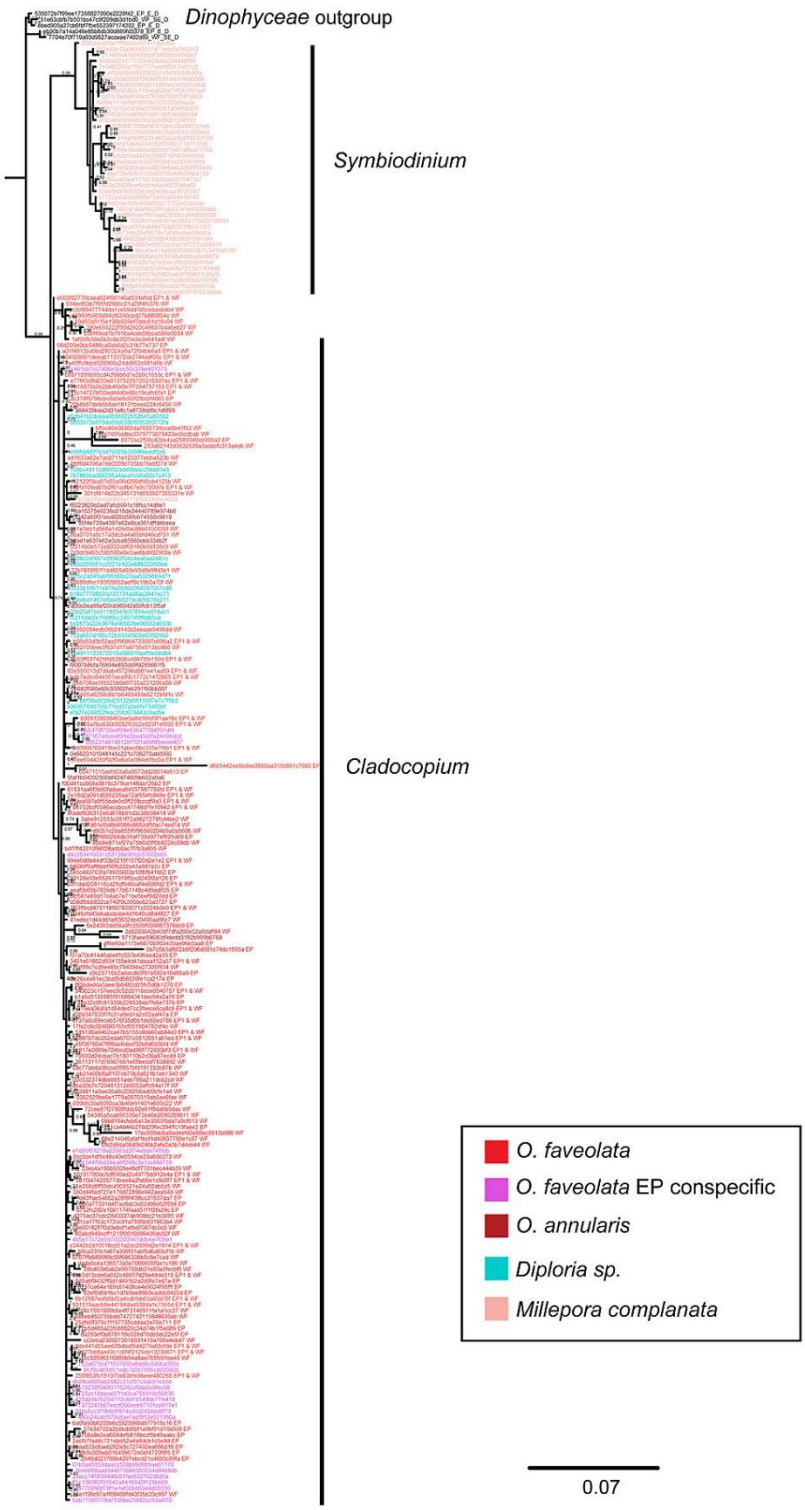
B



C

18S rDNA ASV/Taxa	Ubiquity	Relative Abundance (%)
4ce3d13cde6a652c48857d29e44de316 *Symbiodiniaceae*, *Cladocopium*	0.85	2.13
3461e61802d934135e4d41deaa152a57 *Symbiodiniaceae*, *Cladocopium*	0.85	0.87
250f853fc191070e83bfe36eee480265 *Symbiodiniaceae*, *Cladocopium*	0.85	0.57
f1fd16870c0e2bb46b0e7f7054757159 *Symbiodiniaceae*, *Cladocopium*	0.85	0.55
a01f4912bd0bd282324a9a72f94bb6a5 *Symbiodiniaceae*, *Cladocopium*	0.85	0.26

Supplemental Figure 3.5 Random Forest and ubiquity metrics for 18S rRNA amplicon data. A) Top ten ASV predictors of coral, interface and other categories from the supervised random forest. Taxonomy shown at the lowest taxonomic rank possible based on QIIME2 taxonomy and BLAST analysis. B) Silhouette plot from unsupervised random forest. C) The most abundant, uniquely ubiquitous ASVs found in and only *O. faveolata* samples. Ubiquity ranges from 0-1.

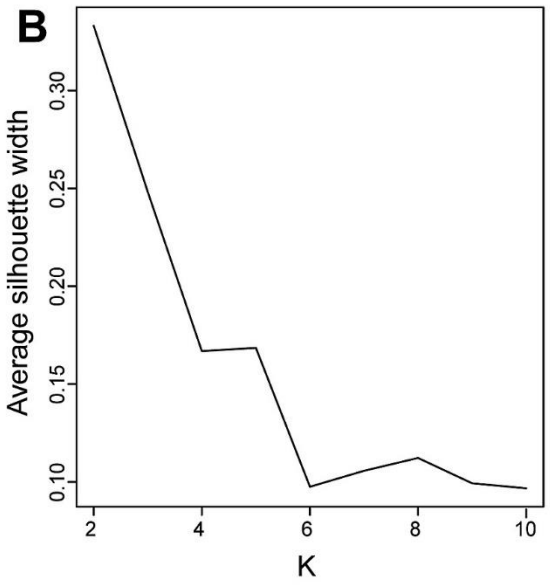


Supplemental Figure 3.6 Maximum likelihood tree inferred under the TIM2+I+G4 of the 18S rRNA gene sequences of symbiodiniaceae and Dinophyceae.

A

Transcript (SEED functional level)	UniProt/EC Identifier	Mean Decrease in Accuracy	p value
NAD(P)H-plastoquinone oxidoreductase	EC 1.6.5.	3.23	0.01
Chaperone protein HscA	P0A6Z1	3.02	0.01
glutaminyl-tRNA synthase (glutamine-hydrolysing)	EC_6.3.5.7	2.89	0.01
Ubiquinol cytochrome c reductase cytochrome B subunit	EC 7.1.1.8	2.86	0.01
DinG family ATP dependent helicase YoaA	A0A385AQ58	2.57	0.01
Plasmin protease S01.233		2.54	0.01
MAP kinase ERK8	Q8TD08	2.52	0.01
FIG005069 hypothetical protein	A0A0H5CFE9	2.51	0.01
Protein translocase subunit SecF		2.46	0.01
UDP-N-acetylmuramoylalanyl-D-glutamate-2,6-diamino-pimelate ligase	EC_6.3.2.13.	2.44	0.01

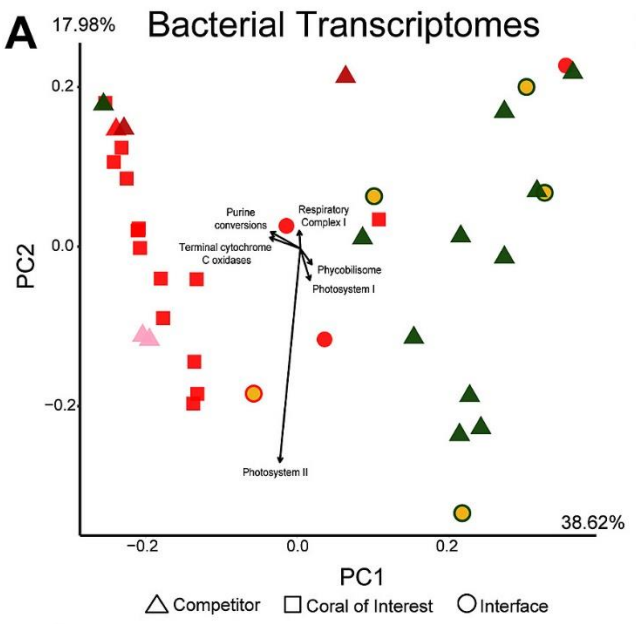
B



C

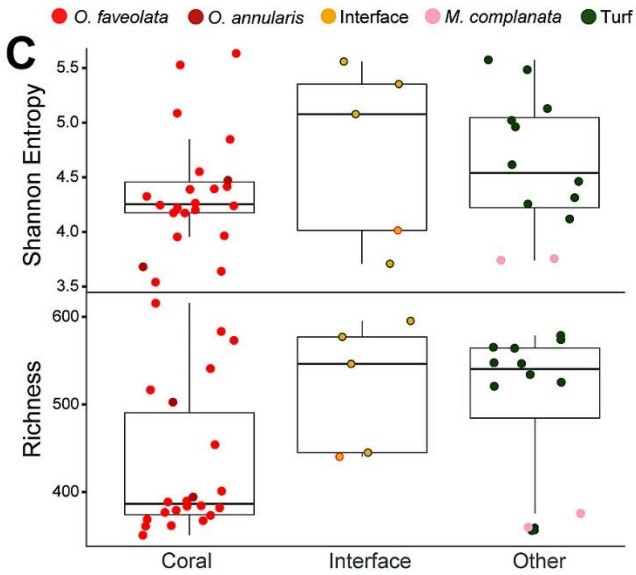
Transcript (SEED Functional Level)	Relative abundance (%)
NAD(P)H-quinone oxidoreductase subunit 5, chloroplastic	12.41
Chaperone protein HscA	4.31
Glutamyl tRNA Gln amidotransferase subunit A	3.20
Ubiquinol-cytochrome c reductase	2.03
DinG family ATP dependent helicase YoaA	1.81
Plasmin protease S01.233	1.76
MAP kinase ERK8	1.47
FIG005069 hypothetical_protein	1.44
Protein translocase subunit SecF	1.37
UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2,6-diaminopimelate ligase	1.15

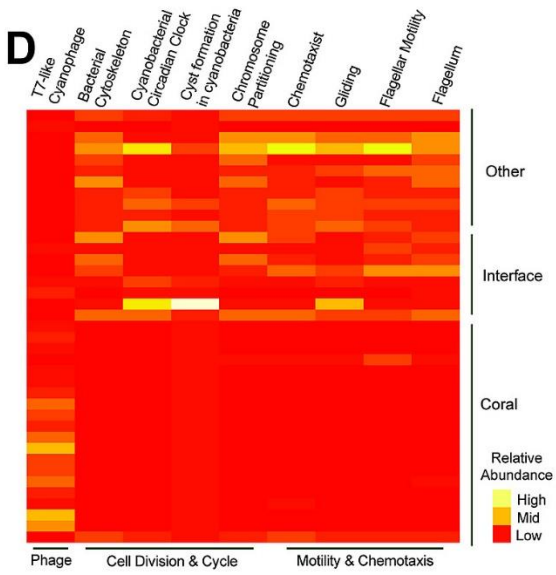
Supplemental Figure 3.7 Random Forest and ubiquity metrics for bacterial transcripts at the SEED functional level. A) top ten transcript predictors of coral, interface and other categories from the supervised random forest. B) Silhouette plot from unsupervised random forest. C) The top 10 most relatively abundant transcripts present in all samples (ubiquity of 1). No transcripts were unique to *O. faveolata*.



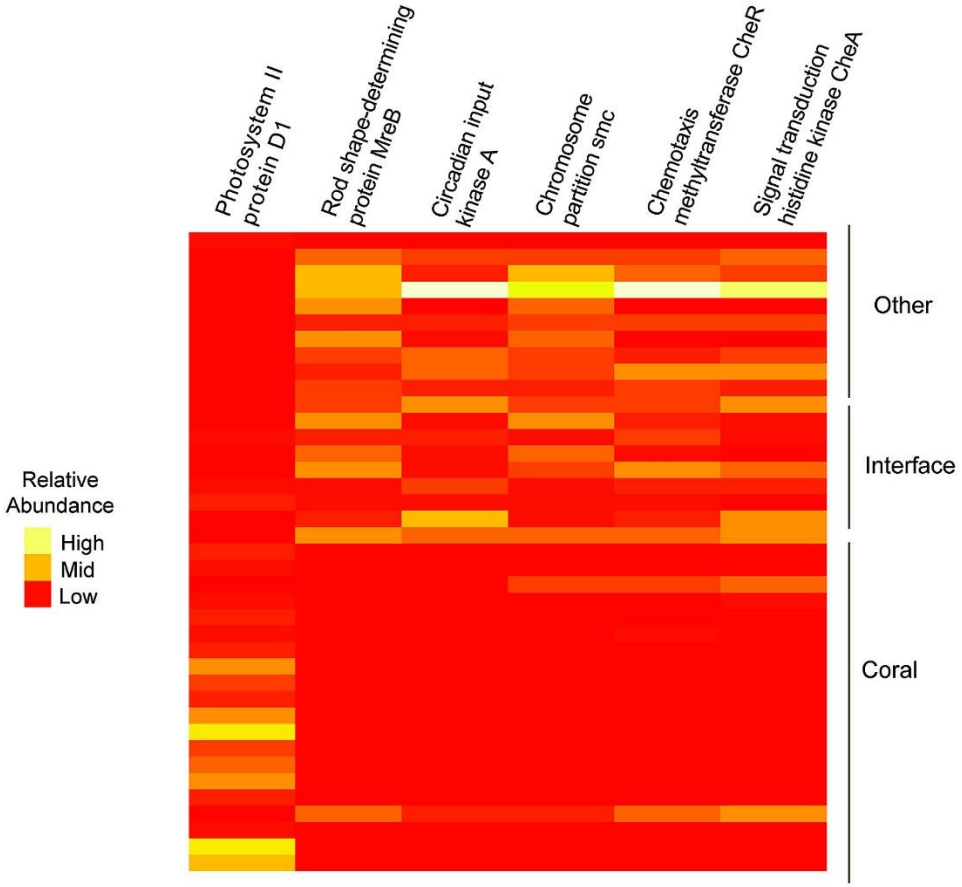
B **PERMANOVA** based on Bray-Curtis dissimilarities

Pairwise Comparison	*F Model*	R^2	*p value*
Coral vs Other	**11.91**	**0.27**	**0.00**
Coral vs Interface	**7.50**	**0.24**	**0.02**
Other vs Interface	0.11	0.01	1.00
***O. faveolata* vs Turf**	**20.65**	**0.42**	**0.01**
***O. faveolata* vs Interface**	**9.37**	**0.30**	**0.01**
Turf vs *M. complanata*	6.92	0.39	0.14
O. faveolata vs *O. faveolata*	1.56	0.04	0.19
O. faveolata vs *O. annularis*	3.31	0.15	0.42
Turf vs *O. annularis*	3.89	0.26	0.42
Interface vs *M. complanata*	4.32	0.46	0.97
Interface vs *O. annularis*	2.12	0.30	1.00
O. annularis vs *M. complanata*	1.44	0.42	1.00
O. faveolata vs *M. complanata*	0.63	0.03	1.00
Turf vs Interface	0.20	0.01	1.00





Supplemental Figure 3.8 Overview of the bacterial transcriptome data at the third SEED level (Level 3). A) Principal components plot of the coral colonies of interest along with interfaces and their respective benthic competitors. B) PERMANOVA results based on Bray-Curtis dissimilarities. C) Diversity metrics of coral, interface and other (turf algae and *M. complanata*) bacterial transcripts. D) Heatmap of selected SEED level 3 categories. These subcategories. These subcategories belong to the following SEED level 1 groups: phage-related, cell division and cell cycle, motility and chemotaxis.

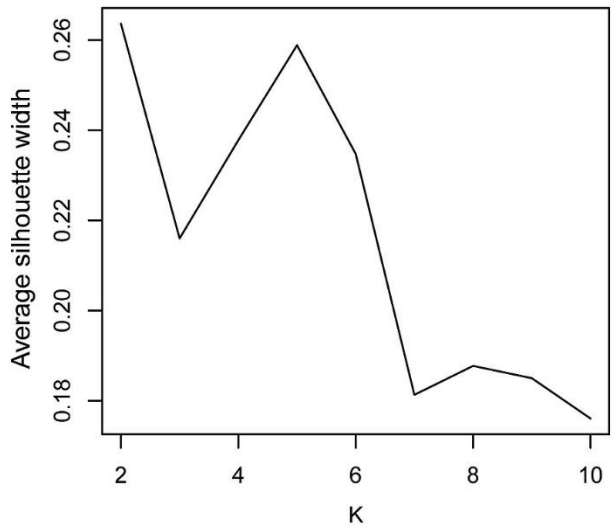


Supplemental Figure 3.9 Heatmap of selected SEED functional categories. The photosystem II protein D1 (PsbA) drove the T7-like cyanophage pattern seen at level 3 (previous figure). Several functional SEED categories contributed to the increased relative abundance of cell division and cell cycle along with motility and chemotaxis in turf and interface. However, only a few of those functional categories are included here.

A

Metabolite	ID	Mean Decrease in Accuracy	*p*
Fatty acid ester	3955_349.2518_5.9965_177	7.56	0.01
Tyr-Pro	2657_263.1324_2.8507_.1	6.63	0.01
Ceramide	6287_536.5013_10.5813_121	6.51	0.01
Psoriasis feature/Glycerophosphoethanolamines	2872_466.3311_7.012_49	6.49	0.01
Amino acids, peptides and analogue	275_377.3041_7.548_22	6.30	0.01
8-Hydroxy-5Z,9E,11Z,14Z-eicosatetraenoic acid (8-Hete)	4361_343.2642_8.5931_8	6.19	0.01
Isomuronic acid/Glycerophosphocholines	18750_755.5144_5.9912_7	5.97	0.01
Fatty acid ester	5862_377.2946_6.6653_177	5.67	0.01
cis-9-Hexadecenoic acid, heptyl ester	2373_277.1994_6.1799_35	5.57	0.01
15-Oxo-11Z,13E-eicosadienoic acid (15-OxoEDE)	2356_329.2407_7.6721_8	5.54	0.01

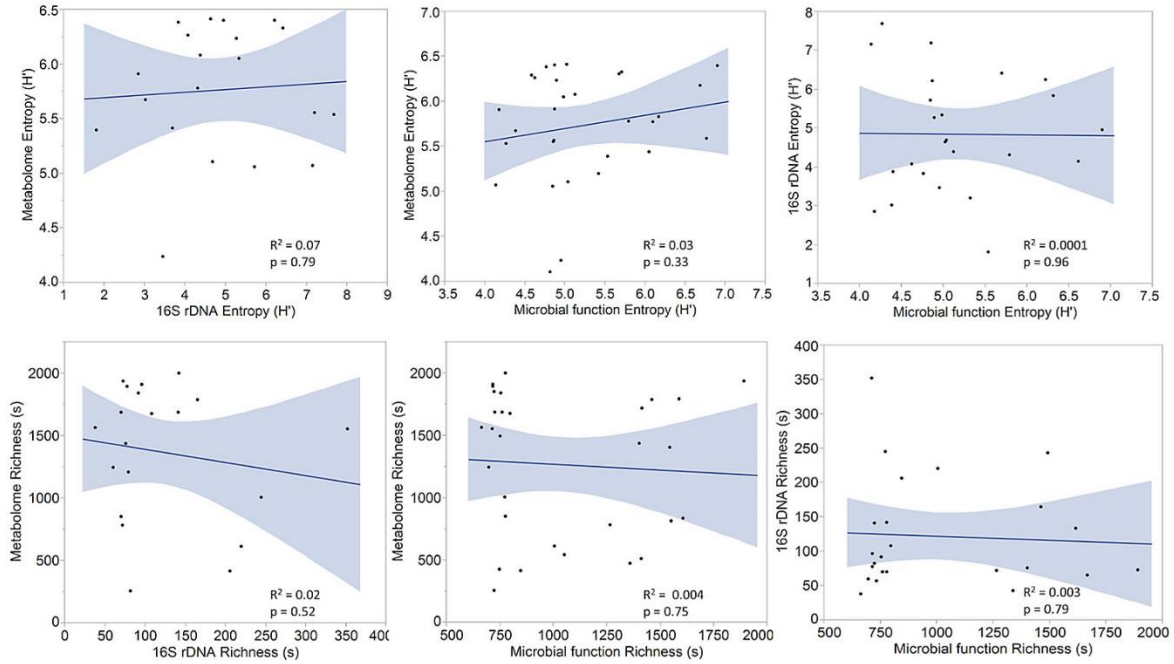
B



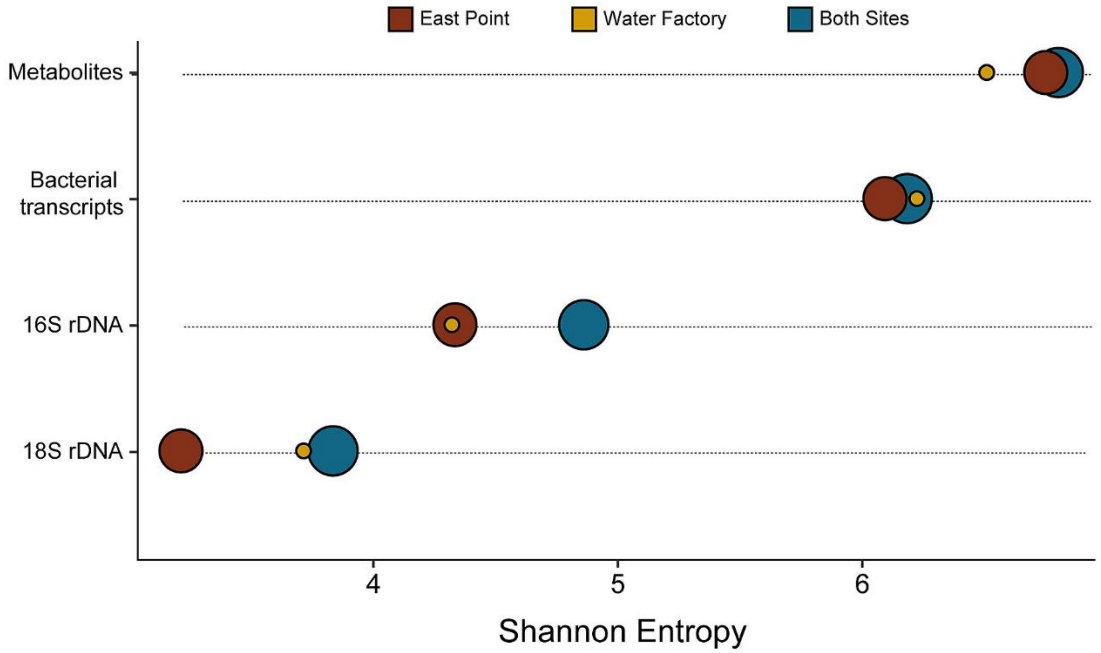
C

Metabolite	Ubiquity
Stylissamide B	0.92
Phenylethylamide	0.85
Piperidines (352.23)	0.69
Fenobucarb	0.69
Coumarins and derivatives (443.25)	0.62
Lobocyclamide	0.62
(4R)-4-[(3S,5S,7S,9S,10S,13R,14S,17R)-3,7-Dihydroxy-10,13-dimethyl-2,3,4,5,6,7,8,9,11,12,14,15,16,17-tetradecahydro-1H-cyclopenta[a]phenanthren-17-yl]pentanoic acid	0.54
Coelimycin P1	0.39
[2-Hydroxy-3-[3,4,5-trihydroxy-6-(hydroxymethyl)oxan-2-yl]oxypropyl] (9Z,12Z,15Z)-octadeca-9,12,15-trienoate	0.39
S-4-benzyl-3-((S)-3-hydroxy-2,2-dimethyloctanoyl)oxazolidin-2-one	0.39

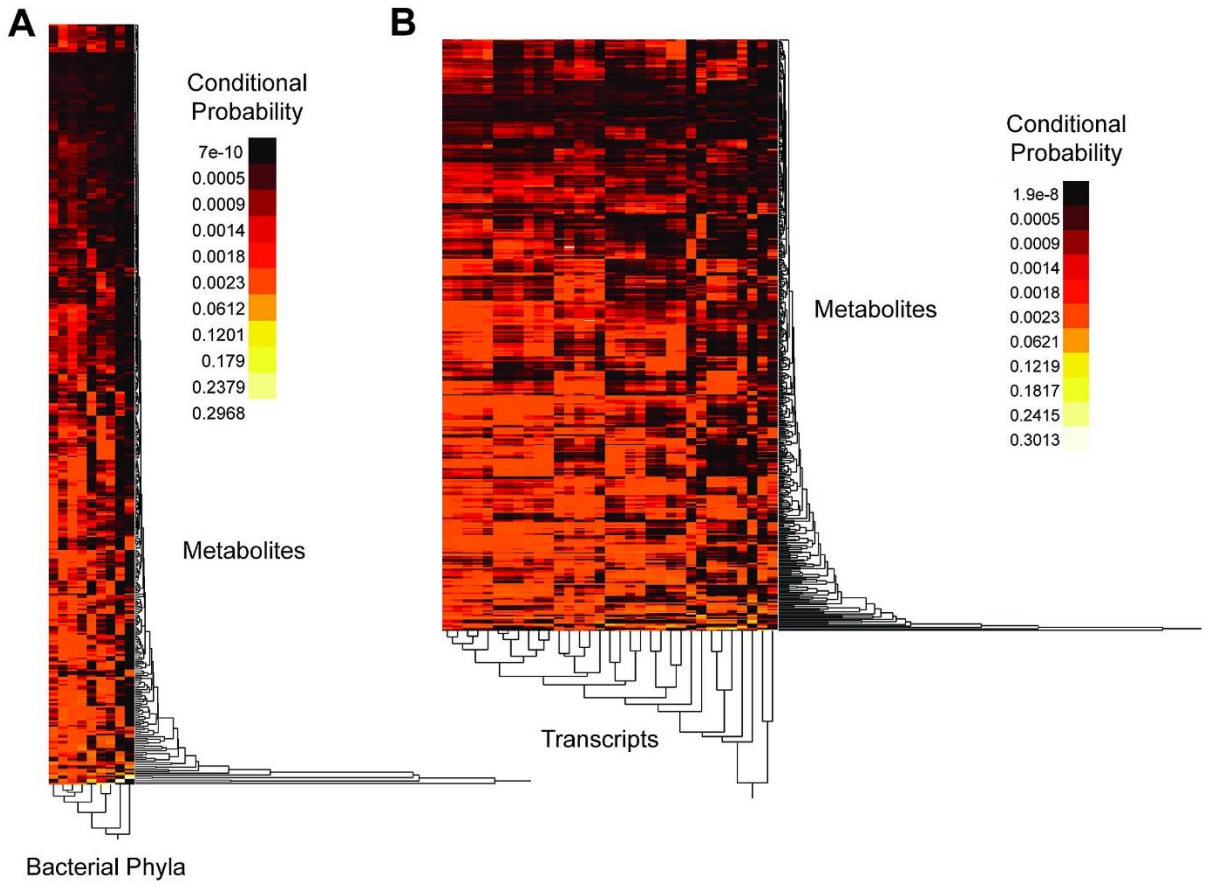
Supplemental Figure 3.10 Random Forest results for annotated metabolites. A) Top ten annotated metabolite predictors of coral, interface and other categories from the supervised random forest. Supervised random forest 483.38 and 510.39 as the top predictors (mean decrease in accuracy, 4.83 and 4.76, respectively). B) Silhouette plot from unsupervised random forest. C) The most abundant, uniquely ubiquitous annotated metabolites found in and only in *O. faveolata* samples. Five unknown metabolites with a mass to charge ratio of 467.31, 240.10, 432.34, 1149.74, and 397.21 were also uniquely ubiquitous (0.92) in *O. faveolata*. Ubiquity ranges from 0-1.



Supplemental Figure 3.11 Regression analyses of metabolomic, 16S rDNA amplicons, and microbial functional Shannon Entropy (H') and Richness (s). 16S amplicon relative abundances for these analyses were conducted from ASV relative abundances. Microbial functions for these analyses were from SEED subsystems functional level relative abundances. No significant correlations were found across the three data types. For metabolomes, relative abundance of all hits were used. Blue lines represent the fit line and blue areas represent the fit confidence (95%).



Supplemental Figure 3.12 The range of Shannon Entropy of all sample types from East Point, Water Factory and both sites combined. The size of the circles represent the reef area sampled at East Point (6.5 m^2), Water Factory (1 m^2), and total area (7.5 m^2).



Supplemental Figure 3.13 Heatmaps from the neural network (MMVEC) analysis. The conditional probability (dark red for low and yellow/white for high) that the presence of a metabolite co-occurs with a bacterial phylum or transcript is shown. A) All metabolites related to all bacterial phyla. B) All metabolites related to all SEED level 1 transcripts.

Supplemental Table 3.1 Power analysis of the sampling size (*n*) required to be statistically significant ($p \leq 0.5$) with variable effect sizes (r) and a power of 80%.

Data	n	r	p	power
Sample set 1	**903**	0.1	0.05	0.8
Sample set 2	**84**	0.3	0.05	0.8
Sample set 3	**28**	0.5	0.05	0.8

Supplemental Table 3.2 Power analysis using sample sizes (*n*) from this study to determine the power at variable effect sizes (*r*) that are statistically significant ($p \leq 0.05$).

Data	n	r	p	power
18S rDNA	21	0.5	0.05	**0.66**
18S rDNA	21	0.3	0.05	**0.27**
18S rDNA	21	0.1	0.05	**0.07**
16S rDNA	21	0.5	0.05	**0.66**
16S rDNA	21	0.3	0.05	**0.27**
16S rDNA	21	0.1	0.05	**0.07**
Metabolites	28	0.5	0.05	**0.79**
Metabolites	28	0.3	0.05	**0.35**
Metabolites	28	0.1	0.05	**0.08**
Transcripts (Lvl 3)	21	0.5	0.05	**0.66**
Transcripts (Lvl 3)	21	0.3	0.05	**0.27**
Transcripts (Lvl 3)	21	0.1	0.05	**0.07**

Supplemental Table 3.3 Sequence mapping identification results including the percentage of reads assigned to microbial subsystems, coral reference genomes, and Symbiodinium reference genomes.

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

Acknowledgements

Chapter 3, in full, is published in Frontiers in Marine Science, 2021. Mark Little, Emma E. George, Milou G.I. Arts, Jade Shivak, Sean Benler, Joel Huckeba, Zachary A. Quinlan, Vittorio Boscaro, Benjamin Mueller, Ana G. Cobian-Guemes, Maria Isabel Rojas, Brandie White, Daniel Petras, Cynthia B. Silveira, Andreas F. Haas, Linda Wegley Kelly, Mark J.A. Vermeij, Robert A. Quinn, Patrick J. Keeling, Pieter C. Dorrestein, Forest Rohwer, and Ty N.F. Roach. The dissertation author was the co-primary investigator and co-first author of this paper.

I would like to thank the following organizations for funding this research: the National Science Foundation for Grants G00009988 (to TR), the Gordon and Betty Moore Foundation for Grants #3781 (to FR), #9702 (to FR), and #9201 (to PK), and the University of British Columbia (International Doctoral Fellowship to EG). I would like to thank the staff of CARMABI research station.

References

Archer, E. (2016). rfPermute: Estimate permutation p-values for Random Forest importance metrics. *R Package Version*, *1*(2).

Aronson, R., Bruckner, A., Moore, J., Precht, B., & Weil, E. (2008). Montastraea faveolata. *IUCN Red List of Threatened Species*.

Barott, K. L., Rodriguez-Brito, B., Janouškovec, J., Marhaver, K. L., Smith, J. E., Keeling, P., & Rohwer, F. L. (2011). Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral Montastraea annularis. *Environmental Microbiology*, *13*(5), 1192–1204.

Barott, K. L., Williams, G. J., Vermeij, M. J. A., Harris, J., Smith, J. E., Rohwer, F. L., & Sandin, S. A. (2012). Natural history of coral- algae competition across a gradient of human activity in the Line Islands. *Marine Ecology Progress Series*, *460*, 1–12.

Barr, J. J., Auro, R., Furlan, M., Whiteson, K. L., Erb, M. L., Pogliano, J., Stotland, A., Wolkowicz, R., Cutting, A. S., Doran, K. S., Salamon, P., Youle, M., & Rohwer, F. (2013). Bacteriophage adhering to mucus provide a non–host-derived immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(26), 10771–10776.

Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., & Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1462), 1935–1943.

Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., Huttley, G. A., & Gregory Caporaso, J. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*, *6*(1), 90.

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., Cope, E. K., Da Silva, R., Diener, C., Dorrestein, P. C., Douglas, G. M., Durall, D. M., Duvallet, C., Edwardson, C. F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J. M., Gibbons, S. M., Gibson, D. L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G. A., Janssen, S., Jarmusch, A. K., Jiang, L., Kaehler, B. D., Kang, K. B., Keefe, C. R., Keim, P., Kelley, S. T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M. G. I., Lee, J., Ley, R., Liu, Y.-X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B. D., McDonald, D., McIver, L. J., Melnik, A. V., Metcalf, J. L., Morgan, S. C., Morton, J. T., Naimey, A. T., Navas-Molina, J. A., Nothias, L. F., Orchanian, S. B., Pearson, T., Peoples, S. L., Petras, D., Preuss, M. L., Pruesse, E., Rasmussen, L. B., Rivers, A., Robeson, M. S., 2nd, Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S. J., Spear, J. R., Swafford, A. D., Thompson, L. R., Torres, P. J., Trinh, P., Tripathi, A., Turnbaugh, P. J., Ul-Hasan, S., van der Hooft, J. J. J., Vargas, F., Vázquez-Baeza, Y., … Caporaso, J. G. (2019).

Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857.

Bouslimani, A., Porto, C., Rath, C. M., Wang, M., Guo, Y., Gonzalez, A., Berg-Lyon, D., Ackermann, G., Moeller Christensen, G. J., Nakatsuji, T., Zhang, L., Borkowski, A. W., Meehan, M. J., Dorrestein, K., Gallo, R. L., Bandeira, N., Knight, R., Alexandrov, T., & Dorrestein, P. C. (2015). Molecular cartography of the human skin surface in 3D. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(17), E2120–E2129.

Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F., & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(22), 14250–14255.

Burns, J., Delparte, D., Gates, R. D., & Takabayashi, M. (2015). Integrating structure-from-motion photogrammetry with geospatial software as a novel technique for quantifying 3D ecological characteristics of coral reefs. *PeerJ*, *3*, e1077.

Burns, J. H. R., Delparte, D., Kapono, L., & Belt, M. (2016). Assessing the impact of acute disturbances on the structure and composition of a coral community using innovative 3D reconstruction techniques. *Methods in*. https://www.sciencedirect.com/science/article/pii/S2211122015300116

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.

Cantu, V. A., Salamon, P., Seguritan, V., Redfield, J., Salamon, D., Edwards, R. A., & Segall, A. M. (2020). PhANNs, a fast and accurate tool and web server to classify phage structural proteins. In *Cold Spring Harbor Laboratory* (p. 2020.04.03.023523). https://doi.org/10.1101/2020.04.03.023523

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & De Rosario, H. (2017). *pwr: Basic functions for power analysis*. https://nyuscholars.nyu.edu/en/publications/pwr-basic-functions-for-power-analysis

Chénard, C., & Suttle, C. A. (2008). Phylogenetic diversity of sequences of cyanophage photosynthetic gene psbA in marine and freshwaters. *Applied and Environmental Microbiology*, *74*(17), 5317–5324.

Clements, C. S., Burns, A. S., Stewart, F. J., & Hay, M. E. (2020). Seaweed-coral competition in the field: effects on coral growth, photosynthesis and microbiomes require direct contact. *Proceedings. Biological Sciences / The Royal Society*, *287*(1927), 20200366.

Closek, C. J., Sunagawa, S., DeSalvo, M. K., Piceno, Y. M., DeSantis, T. Z., Brodie, E. L., Weber, M. X., Voolstra, C. R., Andersen, G. L., & Medina, M. (2014). Coral transcriptome and

bacterial community profiles reveal distinct Yellow Band Disease states in Orbicella faveolata. *The ISME Journal*, *8*(12), 2411–2422.

da Silva, R. R., Wang, M., Nothias, L.-F., van der Hooft, J. J. J., Caraballo-Rodríguez, A. M., Fox, E., Balunas, M. J., Klassen, J. L., Lopes, N. P., & Dorrestein, P. C. (2018). Propagating annotations of molecular networks using in silico fragmentation. *PLoS Computational Biology*, *14*(4), e1006089.

Dharamshi, J. E., Tamarit, D., Eme, L., Stairs, C. W., Martijn, J., Homa, F., Jørgensen, S. L., Spang, A., & Ettema, T. J. G. (2020). Marine Sediments Illuminate Chlamydiae Diversity and Evolution. *Current Biology: CB*, *30*(6), 1032–1048.e7.

Djoumbou Feunang, Y., Eisner, R., Knox, C., Chepelev, L., Hastings, J., Owen, G., Fahy, E., Steinbeck, C., Subramanian, S., Bolton, E., Greiner, R., & Wishart, D. S. (2016). ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *Journal of Cheminformatics*, *8*(1), 61.

Edwards, C. B., Eynaud, Y., Williams, G. J., & Pedersen, N. E. (2017). Large-area imaging reveals biologically driven non-random spatial patterns of corals at a remote reef. *Coral Reefs*. https://link.springer.com/article/10.1007/S00338-017-1624-3

Ernst, M., Kang, K. B., Caraballo-Rodríguez, A. M., Nothias, L.-F., Wandy, J., Chen, C., Wang, M., Rogers, S., Medema, M. H., Dorrestein, P. C., & van der Hooft, J. J. J. (2019). MolNetEnhancer: Enhanced Molecular Networks by Integrating Metabolome Mining and Annotation Tools. *Metabolites*, *9*(7). https://doi.org/10.3390/metabo9070144

Franzosa, E. A., McIver, L. J., Rahnavard, G., Thompson, L. R., Schirmer, M., Weingart, G., Lipson, K. S., Knight, R., Caporaso, J. G., Segata, N., & Huttenhower, C. (2018). Species-level functional profiling of metagenomes and metatranscriptomes. *Nature Methods*, *15*(11), 962–968.

Galtier d'Auriac, I., Quinn, R. A., Maughan, H., Nothias, L.-F., Little, M., Kapono, C. A., Cobian, A., Reyes, B. T., Green, K., Quistad, S. D., Leray, M., Smith, J. E., Dorrestein, P. C., Rohwer, F., Deheyn, D. D., & Hartmann, A. C. (2018). Before platelets: the production of platelet-activating factor during growth and stress in a basal marine organism. *Proceedings. Biological Sciences / The Royal Society*, *285*(1884). https://doi.org/10.1098/rspb.2018.1307

George, E. E., Mullinix, J. A., Meng, F., Bailey, B. A., Edwards, C., Felts, B., Haas, A. F., Hartmann, A. C., Mueller, B., Roach, T. N. F., Salamon, P., Silveira, C., Vermeij, M. J. A., Rohwer, F., & Luque, A. (2021). Space-filling and benthic competition on coral reefs. *PeerJ*, *9*, e11213.

Haas, A. F., Fairoz, M. F. M., Kelly, L. W., Nelson, C. E., Dinsdale, E. A., Edwards, R. A., Giles, S., Hatay, M., Hisakawa, N., Knowles, B., Lim, Y. W., Maughan, H., Pantos, O., Roach, T. N. F., Sanchez, S. E., Silveira, C. B., Sandin, S., Smith, J. E., & Rohwer, F. (2016). Global microbialization of coral reefs. *Nature Microbiology*, *1*(6), 16042.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, *5*(10), R245–R249.

Hartmann, A. C., Petras, D., Quinn, R. A., Protsyuk, I., Archer, F. I., Ransome, E., Williams, G. J., Bailey, B. A., Vermeij, M. J. A., Alexandrov, T., Dorrestein, P. C., & Rohwer, F. L. (2017). Meta-mass shift chemical profiling of metabolomes from coral reefs. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(44), 11685–11690.

Hasin, Y., Seldin, M., & Lusis, A. (2017). Multi-omics approaches to disease. *Genome Biology*, *18*(1), 83.

Hester, E. R., Barott, K. L., Nulton, J., Vermeij, M. J., & Rohwer, F. L. (2016). Stable and sporadic symbiotic communities of coral and algal holobionts. *The ISME Journal*, *10*(5), 1157–1169.

Hogan, R. J., Mathews, S. A., Mukhopadhyay, S., Summersgill, J. T., & Timms, P. (2004). Chlamydial persistence: beyond the biphasic paradigm. *Infection and Immunity*, *72*(4), 1843–1855.

Horn, M., Wagner, M., Müller, K.-D., Schmid, E. N., Fritsche, T. R., Schleifer, K.-H., & Michel, R. (2000). Neochlamydia hartmannellae gen. nov., sp. nov.(Parachlamydiaceae), an endoparasite of the amoeba Hartmannella vermiformisThe GenBank accession number for the sequence reported in this paper is AF177275. *Microbiology*, *146*(5), 1231–1239.

Jetten, M. S. M., van Niftrik, L., Strous, M., Kartal, B., Keltjens, J. T., & Op den Camp, H. J. M. (2009). Biochemistry and molecular biology of anammox bacteria. *Critical Reviews in Biochemistry and Molecular Biology*, *44*(2-3), 65–84.

Kang, K. B., Park, E. J., da Silva, R. R., Kim, H. W., Dorrestein, P. C., & Sung, S. H. (2018). Targeted isolation of neuroprotective dicoumaroyl neolignans and lignans from Sageretia theezans using in silico molecular network annotation propagation-based dereplication. *Journal of Natural Products*, *81*(8), 1819–1828.

Kimes, N. E., Johnson, W. R., Torralba, M., Nelson, K. E., Weil, E., & Morris, P. J. (2013). The M ontastraea faveolata microbiome: ecological and temporal influences on a C aribbean reef-building coral in decline. *Environmental Microbiology*, *15*(7), 2082–2094.

Kwong, W. K., Del Campo, J., Mathur, V., Vermeij, M. J. A., & Keeling, P. J. (2019). A widespread coral-infecting apicomplexan with chlorophyll biosynthesis genes. *Nature*, *568*(7750), 103–107.

Leon, J. X., Roelfsema, C. M., Saunders, M. I., & Phinn, S. R. (2015). Measuring coral reef terrain roughness using “Structure-from-Motion”close-range photogrammetry. *Geomorphology* , *242*, 21–28.

Leray, M., & Knowlton, N. (2015). DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(7), 2076–2081.

Liew, Y. J., Aranda, M., & Voolstra, C. R. (2016). Reefgenomics.Org - a repository for marine genomics data. *Database: The Journal of Biological Databases and Curation*, *2016*. https://doi.org/10.1093/database/baw152

Little, M., Rojas, M. I., & Rohwer, F. (2020). Bacteriophage can drive virulence in marine pathogens. *Marine Disease Ecology*. https://books.google.com/books?hl=en&lr=&id=fcXLDwAAQBAJ&oi=fnd&pg=PA73&dq=Bacteriophage+can+drive+virulence+marine+pathogens&ots=P4xFPqMm0P&sig=E-uwrTFj-P7NeSK9qp7pVujtjJU

MacMillan, J. B., Ernst-Russell, M. A., de Ropp, J. S., & Molinski, T. F. (2002). Lobocyclamides A−C, Lipopeptides from a Cryptic Cyanobacterial Mat Containing Lyngbya confervoides. *The Journal of Organic Chemistry*, *67*(23), 8210–8215.

Marazuela, M. A., Vázquez-Suñé, E., Custodio, E., Palma, T., García-Gil, A., & Ayora, C. (2018). 3D mapping, hydrodynamics and modelling of the freshwater-brine mixing zone in salt flats similar to the Salar de Atacama (Chile). *Journal of Hydrology*, *561*, 223–235.

Matthews, J. L., Cunning, R., Ritson-Williams, R., Oakley, C. A., Lutz, A., Roessner, U., Grossman, A. R., Weis, V. M., Gates, R. D., & Davy, S. K. (2020). Metabolite pools of the reef building coral Montipora capitata are unaffected by Symbiodiniaceae community composition. *Coral Reefs* . https://doi.org/10.1007/s00338-020-01999-3

McKew, B. A., Dumbrell, A. J., Daud, S. D., Hepburn, L., Thorpe, E., Mogensen, L., & Whitby, C. (2012). Characterization of Geographically Distinct Bacterial Communities Associated with Coral Mucus Produced by Acropora spp. and Porites spp. *Applied and Environmental Microbiology*, *78*(15), 5229–5237.

Mendez, K. M., Broadhurst, D. I., & Reinke, S. N. (2019). The application of artificial neural networks in metabolomics: a historical perspective. *Metabolomics: Official Journal of the Metabolomic Society*, *15*(11), 142.

Moree, W. J., McConnell, O. J., Nguyen, D. D., Sanchez, L. M., Yang, Y.-L., Zhao, X., Liu, W.-T., Boudreau, P. D., Srinivasan, J., Atencio, L., Ballesteros, J., Gavilán, R. G., Torres-Mendoza, D., Guzmán, H. M., Gerwick, W. H., Gutiérrez, M., & Dorrestein, P. C. (2014). Microbiota of healthy corals are active against fungi in a light-dependent manner. *ACS Chemical Biology*, *9*(10), 2300–2308.

Morton, J. T., Aksenov, A. A., Nothias, L. F., Foulds, J. R., Quinn, R. A., Badri, M. H., Swenson, T. L., Van Goethem, M. W., Northen, T. R., Vazquez-Baeza, Y., Wang, M., Bokulich, N. A., Watters, A., Song, S. J., Bonneau, R., Dorrestein, P. C., & Knight, R. (2019). Learning representations of microbe–metabolite interactions. *Nature Methods*, *16*(12), 1306–1314.

Neave, M. J., Apprill, A., Ferrier-Pagès, C., & Voolstra, C. R. (2016). Diversity and function of prevalent symbiotic marine bacteria in the genus Endozoicomonas. *Applied Microbiology and Biotechnology*, *100*(19), 8315–8324.

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274.

Nothias, L.-F., Petras, D., Schmid, R., Dührkop, K., Rainer, J., Sarvepalli, A., Protsyuk, I., Ernst, M., Tsugawa, H., Fleischauer, M., Aicheler, F., Aksenov, A. A., Alka, O., Allard, P.-M., Barsch, A., Cachet, X., Caraballo-Rodriguez, A. M., Da Silva, R. R., Dang, T., Garg, N., Gauglitz, J. M., Gurevich, A., Isaac, G., Jarmusch, A. K., Kameník, Z., Kang, K. B., Kessler, N., Koester, I., Korf, A., Le Gouellec, A., Ludwig, M., Martin H, C., McCall, L.-I., McSayles, J., Meyer, S. W., Mohimani, H., Morsy, M., Moyne, O., Neumann, S., Neuweger, H., Nguyen, N. H., Nothias-Esposito, M., Paolini, J., Phelan, V. V., Pluskal, T., Quinn, R. A., Rogers, S., Shrestha, B., Tripathi, A., van der Hooft, J. J. J., Vargas, F., Weldon, K. C., Witting, M., Yang, H., Zhang, Z., Zubeil, F., Kohlbacher, O., Böcker, S., Alexandrov, T., Bandeira, N., Wang, M., & Dorrestein, P. C. (2020). Feature-based molecular networking in the GNPS analysis environment. *Nature Methods*, *17*(9), 905–908.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'hara, R. B., Simpson, G. L., Solymos, P., & Others. (2016). vegan: Community Ecology Package. R package version 2.4-3. *Vienna: R Foundation for Statistical Computing.*

Ovchinnikov, S., Park, H., Varghese, N., Huang, P.-S., Pavlopoulos, G. A., Kim, D. E., Kamisetty, H., Kyrpides, N. C., & Baker, D. (2017). Protein structure determination using metagenome sequence data. *Science*, *355*(6322), 294–298.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., & Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, *42*(Database issue), D206–D214.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., Hugenholtz, P., & Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, *2*(11), 1533–1542.
Pinu, F. R., Beale, D. J., Paten, A. M., Kouremenos, K., Swarup, S., Schirra, H. J., et al. (2019). Systems biology and multi-omics integration: viewpoints from the metabolomics research community. Metabolites 9:76. doi: 10.3390/ metabo9040076

Pluskal, T., Castillo, S., Villar-Briones, A., & Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics*, *11*, 395.

Ponstingl, H., & Ning, Z. (2010). SMALT-a new mapper for DNA sequencing reads. *F1000 Posters*, *1*(L313). https://f1000research.com/assets/download/327

Poulos, B. T., John, S. G., & Sullivan, M. B. (2018). Iron Chloride Flocculation of Bacteriophages from Seawater. *Methods in Molecular Biology* , *1681*, 49–57.

Pratte, Z. A., Longo, G. O., Burns, A. S., Hay, M. E., & Stewart, F. J. (2018). Contact with turf algae alters the coral microbiome: contact versus systemic impacts. *Coral Reefs* , *37*(1), 1–13.

Protsyuk, I., Melnik, A. V., Nothias, L.-F., Rappez, L., Phapale, P., Aksenov, A. A., Bouslimani, A., Ryazanov, S., Dorrestein, P. C., & Alexandrov, T. (2018). 3D molecular cartography using LC--MS facilitated by Optimus and'ili software. *Nature Protocols*, *13*(1), 134.

Quinn, R. A., Vermeij, M. J. A., Hartmann, A. C., Galtier d'Auriac, I., Benler, S., Haas, A., Quistad, S. D., Lim, Y. W., Little, M., Sandin, S., Smith, J. E., Dorrestein, P. C., & Rohwer, F. (2016). Metabolomics of reef benthic interactions reveals a bioactive lipid involved in coral defence. *Proceedings. Biological Sciences / The Royal Society*, *283*(1829). https://doi.org/10.1098/rspb.2016.0469

Quistad, S. D., Stotland, A., Barott, K. L., Smurthwaite, C. A., Hilton, B. J., Grasis, J. A., Wolkowicz, R., & Rohwer, F. L. (2014). Evolution of TNF-induced apoptosis reveals 550 My of functional conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), 9567–9572.

Rädecker, N., Pogoreutz, C., Voolstra, C. R., Wiedenmann, J., & Wild, C. (2015). Nitrogen cycling in corals: the key to understanding holobiont functioning? *Trends in Microbiology*, *23*(8), 490–497.

RCore Team. (2013). *R: A language and environment for statistical computing*. https://cran.microsoft.com/snapshot/2014-09-08/web/packages/dplR/vignettes/xdate-dplR.pdf

Roach, T. N. F., Abieri, M. L., George, E. E., Knowles, B., Naliboff, D. S., Smurthwaite, C. A., Kelly, L. W., Haas, A. F., & Rohwer, F. L. (2017). Microbial bioenergetics of coral-algal interactions. *PeerJ*, *5*, e3423.

Roach, T. N. F., Little, M., Arts, M. G. I., Huckeba, J., Haas, A. F., George, E. E., Quinn, R. A., Cobián-Güemes, A. G., Naliboff, D. S., Silveira, C. B., Vermeij, M. J. A., Kelly, L. W., Dorrestein, P. C., & Rohwer, F. (2020). A multiomic analysis of in situ coral–turf algal interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(24), 13588–13595.

Roach, T. N. F., Dilworth, J., H, C. M., Jones, A. D., Quinn, R. A., & Drury, C. (2021a). Metabolomic signatures of coral bleaching history. *Nature Ecology & Evolution*, *5*(4), 495–503.

Roach, T. N. F., Yadav, S., Caruso, C., Dilworth, J., Foley, C. M., Hancock, J. R., Huckeba, J., Huffmyer, A. S., Hughes, K., Kahkejian, V. A., Madin, E. M. P., Matsuda, S. B., McWilliam,

M., Miller, S., Santoro, E. P., Rocha de Souza, M., Torres-Pullizaa, D., Drury, C., & Madin, J. S. (2021b). A Field Primer for Monitoring Benthic Ecosystems using Structure-from-Motion Photogrammetry. *Journal of Visualized Experiments: JoVE*, *170*. https://doi.org/10.3791/61815
Rogers, S., Ong, C. W., Wandy, J., Ernst, M., Ridder, L., and van der Hooft,
J. J. J. (2019). Deciphering complex metabolite mixtures by unsupervised and supervised substructure discovery and semi-automated annotation from MS/MS spectra. Faraday Discuss. 218, 284–302. doi: 10.1039/c8fd00235e

Rohwer, F., Seguritan, V., Azam, F., & Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series*, *243*, 1–10.

Roulis, E., Polkinghorne, A., & Timms, P. (2013). Chlamydia pneumoniae: modern insights into an ancient pathogen. *Trends in Microbiology*, *21*(3), 120–128.

Santos, H. F., Carmo, F. L., Duarte, G., Dini-Andreote, F., Castro, C. B., Rosado, A. S., van Elsas, J. D., & Peixoto, R. S. (2014). Climate change affects key nitrogen-fixing bacterial populations on coral reefs. *The ISME Journal*, *8*(11), 2272–2279.

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* , *27*(6), 863–864.

Schneider, F. D., Leiterer, R., Morsdorf, F., Gastellu-Etchegorry, J.-P., Lauret, N., Pfeifer, N., & Schaepman, M. E. (2014). Simulating imaging spectrometer data: 3D forest modeling based on LiDAR and in situ data. *Remote Sensing of Environment*, *152*, 235–250.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*(11), 2498–2504.

Shashar, N., Cohen, Y., Loya, Y., & Sar, N. (1994). Nitrogen fixation (acetylene reduction) in stony corals: evidence for coral-bacteria interactions. *Marine Ecology Progress Series*, *111*(3), 259–264.

Silva, G. G. Z., Green, K. T., Dutilh, B. E., & Edwards, R. A. (2016). SUPER-FOCUS: a tool for agile functional analysis of shotgun metagenomic data. *Bioinformatics* , *32*(3), 354–361.

Silveira, C. B., & Rohwer, F. L. (2016). Piggyback-the-Winner in host-associated microbial communities. *NPJ Biofilms and Microbiomes*, *2*, 16010.

Silveira, C. B., Luque, A., Roach, T. N., Villela, H., Barno, A., Green, K., Reyes, B., Rubio-Portillo, E., Le, T., Mead, S., Hatay, M., Vermeij, M. J., Takeshita, Y., Haas, A., Bailey, B., & Rohwer, F. (2019). Biophysical and physiological processes causing oxygen loss from coral reefs. *eLife*, *8*. https://doi.org/10.7554/eLife.49114

Sogin, E. M., Putnam, H. M., Anderson, P. E., & Gates, R. D. (2016). Metabolomic signatures of increases in temperature and ocean acidification from the reef-building coral, Pocillopora damicornis. *Metabolomics: Official Journal of the Metabolomic Society*, *12*(4), 71.

Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F., & Chisholm, S. W. (2005). Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. *PLoS Biology*, *3*(5), e144.

Sweet, M. J., Croquer, A., & Bythell, J. C. (2011). Bacterial assemblages differ between compartments within the coral holobiont. *Coral Reefs* , *30*(1), 39–52.

Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, *21*(8), 1789–1793.

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.

Thompson, L. R., Zeng, Q., Kelly, L., Huang, K. H., Singer, A. U., Stubbe, J., & Chisholm, S. W. (2011). Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(39), E757–E764.

Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The human microbiome project. *Nature*, *449*(7164), 804–810.

Turnbaugh, P. J., Ley, R. E., Mahowald, M. A., & Magrini, V. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*.

van Der Hooft, J. J. J., Wandy, J., Barrett, M. P., Burgess, K. E. V., & Rogers, S. (2016). Topic modeling for untargeted substructure exploration in metabolomics. *Proceedings of the National Academy of Sciences*, *113*(48), 13738–13743.

van der Hooft, J. J. J., Wandy, J., Young, F., Padmanabhan, S., Gerasimidis, K., Burgess, K. E. V., Barrett, M. P., & Rogers, S. (2017). Unsupervised Discovery and Comparison of Structural Families Across Multiple Samples in Untargeted Metabolomics. *Analytical Chemistry*, *89*(14), 7569–7577.

Wandy, J., Zhu, Y., van der Hooft, J. J. J., Daly, R., Barrett, M. P., & Rogers, S. (2018). Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry. *Bioinformatics*, *34*(2), 317–318.

Wang, D.-Z., Xie, Z.-X., & Zhang, S.-F. (2014). Marine metaproteomics: current status and future directions. *Journal of Proteomics*, *97*, 27–35.

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapono, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. V., Meehan, M. J., Liu, W.-T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., Kersten, R. D., Pace, L. A., Quinn, R. A., Duncan, K. R., Hsu, C.-C., Floros, D. J., Gavilan, R. G., Kleigrewe, K., Northen, T., Dutton, R. J., Parrot, D., Carlson, E. E., Aigle, B., Michelsen, C. F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B. T., Gerwick, L., Liaw, C.-C., Yang, Y.-L., Humpf, H.-U., Maansson, M., Keyzers, R. A., Sims, A. C., Johnson, A. R., Sidebottom, A. M., Sedio, B. E., Klitgaard, A., Larson, C. B., Boya P, C. A., Torres-Mendoza, D., Gonzalez, D. J., Silva, D. B., Marques, L. M., Demarque, D. P., Pociute, E., O'Neill, E. C., Briand, E., Helfrich, E. J. N., Granatosky, E. A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J. J., Zeng, Y., Vorholt, J. A., Kurita, K. L., Charusanti, P., McPhail, K. L., Nielsen, K. F., Vuong, L., Elfeki, M., Traxler, M. F., Engene, N., Koyama, N., Vining, O. B., Baric, R., Silva, R. R., Mascuch, S. J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P. G., Dai, J., Neupane, R., Gurr, J., Rodríguez, A. M. C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B. M., … Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, *34*(8), 828–837.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*(1), 57–63.

Weynberg, K. D., Laffy, P. W., Wood-Charlson, E. M., Turaev, D., Rattei, T., Webster, N. S., & van Oppen, M. J. H. (2017). Coral-associated viral communities show high levels of diversity and host auxiliary functions. *PeerJ*, *5*, e4054.

Wickham, H., & Chang, W. (2016). An implementation of the grammar of graphics. *WWW: http://ggplot2. Org*. https://www.rmanchester.org/wp-content/uploads/sites/3/presentations/ManchesterR_-_Rmanchester_TIKZ_-_Graeme_Hutcheson_-_20141030.pdf