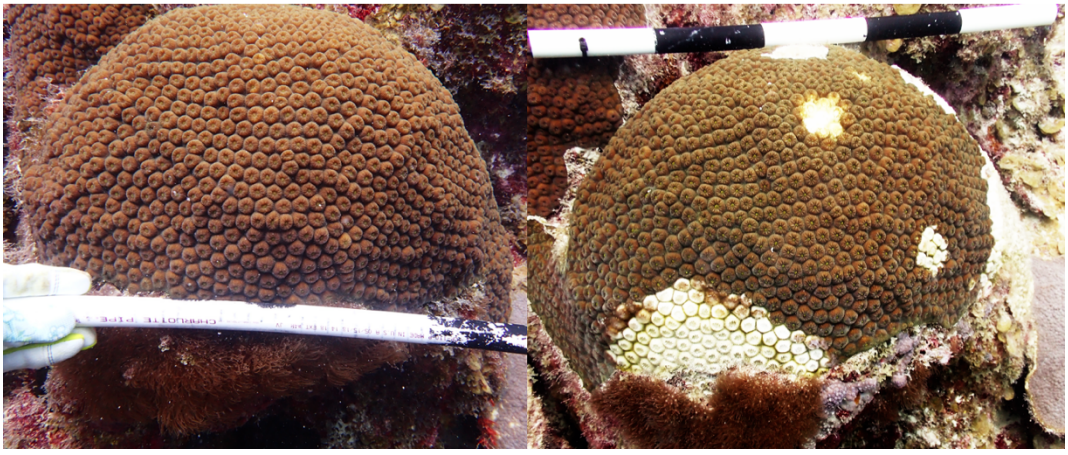
Tracking Bacterial, Viral, and Gene Expression Shifts throughout SCTLD Development in the Dry Tortugas



Tracking Bacterial, Viral, and Gene Expression Shifts throughout SCTLD Development in the Dry Tortugas

Final Report

Prepared By:

Grace Klinges[1]
Lauren Fuess[2]
Alex Veglia[3]
Ashley Rossin[4]
Daniel Holstein[4]

[1]Arizona State University, [2]Texas State University, [3]University of Puerto Rico, Mayaguez, [4]Louisiana State University

June 2024

Completed in Fulfillment of C1E0A5 for

Florida Department of Environmental Protection
Coral Protection and Restoration Program
8000 N Ocean Dr.
Dania Beach, FL 33004

This report should be cited as follows:
{Klinges G, Fuess L, Veglia A, Rossin A, Holstein D. 2024. Tracking Bacterial, Viral, and Gene Expression Shifts throughout SCTLD Development in the Dry Tortugas. Florida DEP. Miami, FL. 37.}

This report was funded through a contract agreement from the Florida Department of Environmental Protection's (DEP) Coral Protection and Restoration Program. The views, statements, findings, conclusions, and recommendations expressed herein are those of the author(s) and do not necessarily reflect the views of the State of Florida or any of its subagencies.

Acknowledgements
We thank Marina Villoch for her considerable support leading both laboratory benchwork and fieldwork for this project, and Erin Shilling (TXST) for assistance with computational analyses. We thank Mote scientists Zachary Craig and Ian Combs as well as NPS staff members Karli Hollister, Jordan Holder, Rachel Johns, Evan Hovey, and Melissa Heres for their support in the field. We also thank Clayton Pollock and Erinn Muller for assisting with study design and project coordination. The collection of samples associated with this project was supported by EPA South Florida Program grant number 02D20722. These samples were collected under NPS permit DRTO-2022-SCI-0006 and are associated with NPS Study No. DRTO-00131.

Management Summary

The rapid spread of SCTLD throughout Florida's Coral Reef has hampered efforts to identify the causative pathogen, due in part to the difficulty of finding disease-susceptible corals with no prior disease exposure. This project leveraged integrative multi-omic and histological analyses to characterize SCTLD pathogenesis and coral immunity in four species of coral collected from Dry Tortugas National Park before and after the arrival of SCTLD. Histopathology of samples from this study showed consistent patterns to other histological analyses of SCTLD, although healthy samples from the Dry Tortugas exhibited an unusual amount of exocytosis. Initial analyses of coral microbial function and taxonomy based on shotgun sequencing data suggest that while microbial communities of corals naïve to SCTLD in Dry Tortugas National Park may differ from other corals sampled in the Florida Keys, microbial community function and composition of diseased samples from this study were similar to samples that have been previously sequenced. We found that diseased coral samples across coral species harbored viruses hypothesized to interact with the algal symbiont and viruses (including herpesvirus) that may influence host immunity capacity against other infections. Immune function in samples of healthy coral varied by species and patterns of change in function followed observed patterns of disease prevalence by month. Continued analyses supported by Florida DEP will focus on the development of robust, integrative analytical approaches to leverage the true power of this data set.

Executive Summary

The rapid spread of SCTLD throughout Florida's Coral Reef has had devastating impacts on these essential coastal ecosystems. This rapid spread has hampered efforts to study many aspects of disease biology, including investigation of causative agents and factors which contribute to host resilience. Characterization of these traits is essential to creation of improve management strategies for Florida's Coral Reefs, but depends on availability of samples from disease-susceptible corals with no prior disease exposure. This project leveraged unique samples from Dry Tortugas National Park (DRTO) samples before and during SCTLD arrival to investigate numerous aspects of SCTLD biology. By combining integrative 'omic and histological analyses of corals sampled through time we provide novel insight regarding the patterns underlying SCTLD outbreaks across multiple species of corals. Histological analyses of samples confirmed signatures of disease consistent with previous analysis of SCTLD corals, including vacuolization of symbionts. Similarly, while microbial communities of DRTO corals pre-disease were unique compared to other Florida

Keys samples, signatures of disease progression in the microbial communities were highly similar to those previously sequenced. Notably, while corals from DRTO had high levels of beneficial microbes, including *Endozoicomonas*, presence of these bacteria failed to protect against disease onset, and populations of these beneficial microbes declined through disease onset. This suggests intervention to maintain beneficial microbial communities, through methods such as probiotic treatment, may be necessary to maintain these communities and prevent disease. Viral community analyses similarly identified patterns consistent with previous studies of SCTLD. Viruses associated with Symbiodiniaceae were present in both healthy and diseased samples, pointing to a role of opportunistic pathogens in SCTLD. Viruses with potential impacts on coral immunity were also documented, further implicating roles of opportunistic associations in SCTLD. Finally, transcriptomic analyses revealed notable patterns of changes in host coral immunity across disease progression in three out of the four studied species. Notably, all three of these species were characterized by decreases in expression of key immune genes over time in healthy and apparently healthy coral tissue. These patterns of immune decline corresponded with peak disease onset in each species, suggesting a biological rather than environmental driver. Combined, our analyses to date suggest that even remote, previously healthy corals can succumb to SCTLD, and emphasize the need for proactive and comprehensive management approaches. Ongoing analyses of these data, supported by a 2024-2025 FDEP contract, will incorporate integrative approaches to generate important biological insight which will guide the development of improved management approaches necessary for the protection of theses reefs.

Table of Contents

1. BACKGROUND

1.1. Introduction

Florida's Coral Reef is currently experiencing a multi-year disease-related mortality event that has resulted in massive die-offs in multiple coral species. This die off event has been attributed to the spread of a novel coral disease, stony coral tissue loss disease (SCTLD). Approximately 21 species of coral, including both Endangered Species Act-listed and primary reef-building species, have displayed tissue loss lesions which often result in whole colony mortality. First observed near Virginia Key in late 2014, the disease has since spread to the northernmost extent of Florida's Coral Reef, and southwest past the Marquesas in the Lower Florida Keys. The best available information indicates that the disease outbreak is continuing to spread west and throughout the Caribbean.

The rapid spread of SCTLD throughout Florida's Coral Reef has hampered efforts to identify the causative pathogen, due in part to the difficulty of finding disease-susceptible corals with no prior disease exposure. Comparative assessments of pathogen abundances in "healthy" vs diseased corals may be confounded by the presence of latent, asymptomatic infection in healthy controls. Many bacterial taxa found in diseased corals (incl. Rhizobiales and Rhodobacterales) have also been identified in considerable abundance in apparently healthy conspecifics. Temporal analyses of colonies from pre-exposure to necrosis may help reduce background variation of coral-associated bacterial and viral communities and in coral immune function. Repeated sampling of colonies previously naïve to SCTLD not only provides a better baseline to assess microbiome composition of healthy corals, but also eliminates the potential for inter-colony variation in microbial composition that may have inhibited prior discovery of the causative agent. Furthermore, temporal approaches allow for improved investigation of mechanisms of coral response to SCTLD, including those which may confer disease resistance.

1.2. Project Goals and Tasks

This project leveraged integrative multi-omic and histological analyses to characterize samples of four species of coral collected from Dry Tortugas National Park before and after the arrival of SCTLD. This included samples of SCTLD-naïve, SCTLD-exposed but apparently healthy, and diseased coral health states. This integrative approach will provide insight regarding the etiological agent of SCTLD, and help identify important mechanisms of coral response and resilience to SCTLD. The outcomes of this project will be incorporated into an on-going coral disease response effort for Florida's Coral Reef.

2. METHODS

2.1. Metagenomic Analyses

Coral samples stored in DNA/RNA shield were homogenized using MP Biomedicals Lysis Matrix A, and a portion of this homogenate was then used for DNA extraction using the Omega E.Z.N.A.® DNA/RNA Isolation Kit, following the manufacturer's protocol. Quality & quantity were assessed using a NanoDrop spectrophotometer. Extracted DNA from the samples was then submitted to Genohub in November of 2023. At Genohub, final quality control was performed using qPCR and via Agilent Tapestation. Whole-genome sequencing was then performed to a

minimum depth of 50 million 150bp PE reads per sample on one lane of a NovaSeq X Plus (10B). Raw sequencing products were submitted to the NCBI Sequence Read Archive (SRA) as project numbers SUB14296846 and SUB14297148.

Quality control and adapter removal was performed using *fastp* (v. 0.22.0, (Chen et al., 2018)) with an average allowable phred score of 25, minimum length of 50bp, a complexity threshold of 30%, and with reads with greater than two Ns (ambiguous base calls) discarded. Base correction was enabled in overlap regions and any tails containining polyG or polyX sequences were trimmed. On average, 13.13% of reads were lost with quality control per sample. The maximum number of reads per sample after filtration was 241,998,208 reads and the minimum was 47,370,064. Error correction was performed on all reads using the SPAdes error correction module (Nurk et al., 2017) using a k-mer size of 21, with coverage threshold calculation enabled and with mismatch correction enabled.

Host and symbiont reads were removed using *bowtie2* (v. 2.4.2). *Bowtie2* alignment was performed to the available genomes for *Montastraea cavernosa* (https://www.bco-dmo.org/dataset/875253), *Orbicella faveolata* (GCA_002042975.1), *Durusdinium trenchii* (GCA_963969995.1), *Breviolum minutum* (GCA_000507305.1), *Cladocopium goreaui* (GCA_947184155.1), and *Symbiodinium fitti* (GCA_003297005.1). Local alignments were performed and reads that failed to align concordantly to reference genomes (i.e., reads that did not match to symbiont or host genomes) were retained. We also utilized the draft transcriptome of *Colpophyllia natans* produced by RNAseq analysis (see "Transcriptomic Analyses" below) to produce a rough DNA reference library for this species and aligned samples of *Colpophyllia natans* to this draft.

After host/symbiont filtration, functional characterization of samples was performed using HUMAnN3 (v3.8, (Beghini et al., 2021) to profile genes, pathways, and modules from initial assemblies. Reads were annotated via MetaPhlAn4 (v 4.1.1, Blanco-Miguez et al., 2023) by aligning to both the ChocoPhlAn DNA database and the UniRef50 protein database. Outputs of this analysis included gene family abundance files (the presence of groups of evolutionarily-related protein coding sequences) and pathway abundance files (a measure of pathway completion of a function of abundance of the pathway's component reactions). Pathway abundance tables were then joined into a single file using *humann_join_tables* and renormalized to copies per million (CPM) using *humann_renorm_table*. Pathways from HUMAnN3, counts, and sample metadata were normalized as relative abundance, imported into RStudio (v.2022.12.0+353) with R (v.4.2.2), and plotted using ggplot2. Taxonomic characterization of each sample (unbinned) was performed using MetaPhlAn4 (v 4.1.1,(Blanco-Miguez et al., 2023) using the ChocoPhlAn DNA database. To determine taxonomic groups that were statistically differentially abundant between disease status types, the R package *corncob* (Martin et al., 2020) as used.

Reads were also used to produce metagenome-assembled genomes (MAGs) by assembling *de novo* using three parallel methods: MEGAHIT (v.1.2.9, Li et al., 2016), metaSPAdes (v.3.13.1, (Nurk et al., 2017), and MetaVelvet (v1.1.01, Namiki et al., 2012). MEGAHIT and MetaVelvet were performed using presets for large metagenomes. MetaSPAdes assembly was performed within the *metawrap* wrapper

using default parameters. Initial assemblies were concatenated by sample and then binned using *MaxBin* (v. 2.2.7, Wu et al., 2016) to cluster metagenomic contigs into different bins, each consisting of contigs from one microbial species. Quality of these initial bins was then assessed using *CheckM* (v. 1.2.2, Parks et al., 2015), which uses single-copy marker gene sequences to evaluate the completeness and contamination of a genome. CheckM was also used to assess taxonomy using the marker lineage workflow. At this stage, completion is low, which is to be expected due the large amount of bacterial diversity observed in these samples from initial analysis via the 16S rRNA marker gene. Assemblies will be refined by combining taxonomically congruent bins across samples.

2.2. Viral Community Analyses

Virus community analyses were performed on previously quality-controlled DNA and RNA sequencing reads (see sections 2.1 and 2.3). To reduce the computational burden of downstream analyses, all cleaned libraries were normalized using the program bbnorm.sh from the BBMap toolkit (https://sourceforge.net/projects/bbmap/). Because the field of coral reef virology is still in its infancy (Thurber et al., 2017), there is no available coral reef virus databases adequate for virus read alignment. Thus, the first step in the virus community analysis was "raw" assembly of each cleaned DNA and RNA sequencing libraries with SPAdes (v.3.15.5, Nurk et al., 2017) using the "--meta" and "--rnaviral" options, respectively.

Virus-like sequences were detected in the DNA assembly files using the program geNomad (v1.8.0, Camargo et al., 2023). Each sequence classified as virus-like by geNomad was then extracted from their original assembly file and concatenated into a single fasta file. The single fasta file containing all virus-like sequences from the metagenome assemblies was then used as input for binning collapsing with vRhyme (v1.1.0, Kieft et al., 2022) to construct virus metagenome-assembled genomes (vMAGS). Sequences that were not included in bins were then extracted from the vRhyme input fasta file and combined with the final vMAGs producing a final fasta file containing all likely virus sequences from DNA sequencing libraries. Quality and completeness of all virus vMAGs and unbinned sequences (both of which representing either genome-fragments or whole genomes) using the program CheckV (v1.0.1, Nayfach et al., 2021). Taxonomic classification for the final set of virus-like sequences was generated using the program geNomad (v1.8.0). Metadata information was for each sequence was extracted from CheckV and geNomad output files to produce a final sequence metadata file containing information such as sequence length, quality, completeness and predicted taxonomic classification. A read counts matrix with all samples was produced by mapping non-normalized, cleaned reads from each DNA sequencing library to the final fasta file containing virus-like sequences with the programs minimap2 (v2.28, Li, 2021) and SAMtools (v1.20, Danecek et al., 2021)).

Virus-like contigs were first identified within each individual RNA sequencing library using the program Deep6 (Finke et al., 2023). Any sequence predicted to be viral with a group score greater than or equal to 0.7 was then extracted and sequences were pooled into a single fasta file. Next, short (<500 bp), misassembled, and redundant contigs were removed from the fasta file containing all Deep6-predicted virus sequences and a non-redundant assembly file was produced using the program

EvidentialGene tr2aacds pipeline v2019.05.14 Gilbert, 2019). A second round of Deep6 virus sequence prediction was then performed on the non-redundant virus-like contig fasta file to retain only the most probable virus-like sequences. The final non-redundant virus-like contig fasta file was then assessed with geNomad (v1.8.0) to obtain putative taxonomy of sequences. Next, quality and completeness of the virus-like sequences were checked using the program CheckV (v1.0.1). Metadata for each sequence was then extracted from Deep6, geNomad, and CheckV output files to produce a final sequence metadata file containing identical types of information within the DNA sequence metadata file. A read counts matrix with all samples was generated by mapping cleaned, normalized reads against the final non-redundant virus-like contig file using the program kallisto (v0.50.1, (Bray et al., 2016).

The sequence classification results, expression counts, and sample metadata were imported into RStudio (v.2022.12.0+353) with R (v.4.2.2) and combined into a single phyloseq object (one for DNA sequencing results files and one for RNA sequencing results files) using the phyloseq package McMurdie & Holmes, 2013). To determine the virus sequences/taxonomic groups (when available) that had statistically significant increase in abundance within diseased coral tissue samples, the R package *DESeq2* was used with the Wald test, "fittype = local". Differential abundance of these upregulate virus taxa in the RNA sequencing data was visualized in a heatmap using the abundance_heatmap() function from the *phylosmith* R package and the trimmed mean of M-values (TMM) normalized counts table.

2.3. Transcriptomic Analyses

Coral samples stored in DNA/RNA shield were homogenized using MP Biomedicals Lysis Matrix A, and a portion of this homogenate was then used for RNA extraction using the Omega E.Z.N.A.® DNA/RNA Isolation Kit, following the manufacturer's protocol. Quality & quantity were assessed using a Nanodrop spectrophotometer. Extracted RNA from the samples was then submitted to Genohub in November of 2023. At Genohub, mRNA library prep with polyA tail enrichment was conducted and the library was then sequenced on the NovaSeq PE150 platform.

Resulting read files were first processed using the program fastp to remove adapters and filter out poor-quality reads (Chen et al., 2018). The following parameters were used for fastp: minimum quality score of 25, minimum phred quality score of 20, minimum read length of 50 bp, minimum complexity threshold of 30%, removal of any polyG or polyX tails, n base limit of 2, and base correction enabled in overlap regions.

De novo transcriptomes were generated from a random set of five samples for all coral species following protocols described in Beavers et al. 2023. "Raw" transcriptome assemblies were generated using Trinity v2.9.1 (Grabherr et al., 2011). The resulting transcriptomes were then filtered to retain only the longest isoform per gene and blasted against a master coral transcriptome to retain only coral-derived transcripts (https://zenodo.org/records/7838980). Next, the program TransDecoder was used to identify transcripts with open reading frames and transcripts without an identified coding region were removed (Haas & Papanicolaou, 2016). The program cd-hit was used to remove redundant transcripts (Li & Godzik, 2006). Resultant transcriptome completeness was assessed using BUSCO (Simao et al., 2015). Finally,

the transcriptomes were annotated via comparison to the UniProt database using blast v2.15.0.

Prior to alignment, read files were processed using BBSplit (Bushnell) to separate host and Symbiodiniaceae reads, as returned sequencing data includes transcripts from both. Raw reads were processed in BBSplit against the generated de novo transcriptomes and references transcriptomes of species representative of each of the four dominant genera of Symbiodiniaceae: Symbiodinium microadriaticum (Aranda et al., 2016), Breviolum minutum (Parkinson et al., 2016), Cladocopium goreaui (Davies et al., 2016), and Durusdinium trenchii (Bellantuono et al., 2018). The resulting host reads were then aligned to the generated de novo transcriptomes using Salmon (Patro et al., 2017).

Symbiont read count matrices were also generated using BBSplit output. In addition to splitting out host and Symbiodiniaceae reads, BBSplit also reports proportions of reads aligned to each reference transcriptome. Based on this output each sample was assigned a dominant symbiont type for symbiont gene expression analyses. Samples were then split by dominant species type and BBSplit symbiont output reads were aligned to the respective reference Symbiodineaceae transcriptome (listed above) using Salmon (Patro et al., 2017). The symbiont read count matrices will be analyzed at a later date.

Generated host read matrixes were analyzed to identify differences in gene expression patterns across time points and tissue status using the R package limma (Ritchie et al., 2015). Limma analysis was chosen as the program allows for a repeated measures approach. Gene expression analyses were conducted independently for each coral species. Reads were first normalized using the calcNormFactors function, and then analyzed using the voomLmFit function with blocking based on sample genotype (repeated measures approach). For each species all pairwise contrasts between sample groups with a minimum *n* of 3 were run. **Table 1** lists pairwise comparisons run for each species. Differentially expressed genes were identified based on 10% FDR (adjust $p < 0.10$). Immune genes were identified based on gene ontology terms using the keywords: “immun”, “bacteria”, ”defense”, ”inflamm”, “wound heal”, and “melanin bio”. All graphs were produced using raw count values with the R package ggplot2 (Wickham, 2016).

Table 1. *List of pairwise comparisons used for transcriptomic analyses; H= healthy, AH= apparently healthy, DM= disease margin*

Species	Contrasts
C. natans	Sept_H vs June_AH, Sept_H vs. Aug_AH, June_AH vs. Aug_AH June_AH vs. June_DM
O. faveolata	Sept_H vs June_H, Sept_H vs. Aug_H, June_H vs. Aug_H Sept_H vs June_AH, June_H vs. June_AH, Aug_H vs. June_AH June_AH vs. June_DM
O. franksi	Sept_H vs June_H, Sept_H vs. Aug_H, June_H vs. Aug_H Sept_H vs. Aug_AH, June_H vs. Aug_AH
M. cavernosa	Sept_H vs June_H, Sept_H vs. Aug_H, June_H vs. Aug_H

2.4. Histological Analyses

Samples were fixed in zinc-buffered formalin (Z-fix, Anatech), then seawater 24 hours, then stored in 70% ethanol and shipped to Louisiana State University. Corals were decalcified with a 1% HCl EDTA solution and stored in 70% ethanol until processed. Corals were processed using a Leica ASP6025, embedded in paraffin wax blocks on a Leica EG1150H embedding machine, and sectioned at five mm thickness on a Leica RM2125RTS microtome. All samples were sectioned in both cross and longitudinal orientation with three to five polyps in each orientation. Seven sections were made 500 µm apart. Histological slides were stained with hematoxylin and eosin stain on a Leica ST5020, viewed on an Olympus BX41 microscope with an Olympus SC180 camera attachment, and analyzed using ImageJ software.

Slides were analyzed across two methodologies, tissue quantification and measurements following Rossin et al. (*in prep*). Tissue quantification was split between consistency and intensity of disease signs ranked as absent, low, medium, and high. The disease signs noted were necrosis, vacuolization, exocytosis, gastrodermal separation, and degraded symbionts. Consistency referred to the signs occurring over the five slides analyzed – regardless of intensity of sign. Intensity referred to the degree of the disease sign when it was seen. Additionally, certain tissue parameters were noted for presence/absence: eroded gastrodermis, amoebocytes, loss of eosin from the mesoglea, loss of structural integrity, and fungus or sponge. This quantification was then compared between species and time points.

Disease measurements were performed using five 60,000 mm^2 images per tissue sample. Each micrograph was split into twelve 5000 mm^2 grid-cells. A random number generator determined which section the cells were measured within. The areas of 15 symbionts within their vacuoles were measured per sample. Additionally, presence of gastrodermal separation and degraded symbionts was noted, as well as the proportion of symbionts undergoing exocytosis within the grid-cell of interest.

A Bayesian hierarchical linear model was used to detect differences between apparently healthy and diseased samples according to four histological measurements:

vacuolization, *symbiont size*, *degraded symbiont presence*, *proportion exocytosis*, and *gastrodermal separation*. This model used non-informative priors. A second Bayesian hierarchical model was employed in which the first level of the model was a binomial generalized linear model with an intercept and five predictors, which were each modeled as random effects with $j = 6$ levels for species.

Where y_i is the binomially-distributed response variable of disease state (0 = putatively healthy and 1 = diseased), a is the intercept, b_{1-4} are the slopes, and x_{1-5} are the predictor values, exocytosis, gastrodermal separation, symbiont size, degraded symbiont presence, and vacuolization. The subscript j indexes the random effect of species ($j = 6$) and the subscript i indexes the observed data ($i = n$). The second level of the model was unconditional; i.e., no model was applied to the random effect estimates and only grand mean estimates of the five level one parameters were estimated. Model variance-covariance was estimated using a scaled-inverse Wishart distribution. The model was run in JAGS (version 1.5.2) using the package JAGSUI (Su & Yajima, 2015) in R. All parameters were given diffuse normal priors. Models were initialized with a randomly selected value for all five parameters from a normal distribution with a mean of zero and standard deviation of one. We ran three Markov chain Monte Carlo (MCMC) chains each for 40,000 iterations but removed 5,000 for burn in and thinned by two, for a total of 105,000 iterations used for posterior analysis.

Model convergence was evaluated from the values, where < 1.1 indicated convergence. Additionally, we plotted all posteriors and visually confirmed convergence. We interpreted predictor effects based on where 0 was in relation to their posterior distributions. We did not take full advantage of the Bayesian hierarchical model potential as we did not have informed priors or utilize a second level model in this study. All Bayesian hierarchical models converged below 1.1 and were accepted.

3. RESULTS

3.1. Metagenomic Analyses

3.1.1.Filtration

We sequenced samples of four species of coral, resulting in an average of 33,781,130 reads per sample for *Orbicella faveolata*, an average of 42,083,762 reads per sample for *Orbicella franksi*, an average of 29,069,371 reads per sample for *Montastraea cavernosa*, and an average of 24,240,705 reads per sample for *Colpophyllia natans.* After quality control, these 133 metagenomic samples were aligned to the available genomes for *Montastraea cavernosa* (https://www.bco-dmo.org/dataset/875253), *Orbicella faveolata* (GCA_002042975.1), *Durusdinium trenchii* (GCA_963969995.1), *Breviolum minutum* (GCA_000507305.1), *Cladocopium goreaui* (GCA_947184155.1), and *Symbiodinium fitti* (GCA_003297005.1). Reads from *Colpophyllia natans* were also aligned to the draft transcriptome of *Colpophyllia natans* produced by RNAseq analysis. Only reads that did not align to coral host or algal symbiont reference genomes were retained for downstream analyses. Samples were sequentially aligned to reference genomes in the order listed in **Table 2**, and as such, the percent aligning to each genome refers to the percent of remaining reads that aligned, not the percent of initial reads. For example, for samples of *Montastraea cavernosa* 10.89% of initial reads aligned to the genome of *Orbicella faveolata*. Then an average of 83.38% of remaining reads aligned to the

genome of *M. cavernosa*. Although samples of both *Orbicella* species aligned well to the *O. faveolata* genome and samples of *M. cavernosa* aligned well to the *M. cavernosa* available genome, *C. natans* samples did not align well to the *C. natans* transcriptome produced from RNAseq reads. This is likely a result of large non-coding regions of *C. natans* that are absent from the transcriptome and which are likely present in DNA sequences from these samples. After quality control and filtration, we retained an average of 9,802,537 reads per sample for *Orbicella faveolata*, an average of 13,011,951 reads per sample for *Orbicella franksi*, an average of 4,748,355 reads per sample for *Montastraea cavernosa*, and an average of 14,599,950 reads per sample for *Colpophyllia natans.*

After aligning the metagenomic reads to these reference genomes, a substantial amount of host and algal symbiont reads were successfully removed. This step was critical to reduce the background noise and enhance the detection of other microbial sequences within the samples. However, subsequent BLAST queries indicated that a significant level of coral host contamination remains in the metagenomic samples, with sequence BLAST homology matching other species of coral not included in this dataset. This has hampered the assembly of microbial genomes from metagenomic dataset. We will continue to refine this dataset by aligning reads to available reference genomes from other coral species.

Table 2. *A. Percent alignment of samples from this dataset to available reference genomes and transcriptomes from species included in this study. B. Percent alignment of samples from this dataset to available reference genomes for dominant symbiont profiles identified from ITS2 data from these samples.*

A. Host alignment	1. Average % aln to *Orbicella faveolata*	2. Average % aln to *Montastraea cavernosa*	3. Average % aln to *Colpophyllia natans* draft transcriptome
CNAT	4.74	8.54	15.79
MCAV	10.89	83.38	NA
OFAV	72.17	9.48	NA
OFRA	67.96	9.41	NA

B. Symbiont alignment	4. Average % aln to *Durusdinium trenchii*	5. Average % aln to *Symbiodinium fitti*	6. Average % aln to *Cladocopium goreaui*	7. Average % aln to *Breviolum minutum*
CNAT	0.14	0.49	3.04	38.35
MCAV	0.14	0.48	39.15	4.08
OFAV	0.15	0.48	35.12	5.05
OFRA	0.15	0.52	34.93	3.01

3.1.2. *Functional and Taxonomic Annotations*

In order to best characterize the microbial community composition and functional capacity in these samples, we elected to analyze the dataset looking both at unassembled reads, reflecting the entire microbial community in each sample, as well as at metagenome-assembled genomes, which provide greater resolution into the dominant taxa within our dataset. From complete metagenomic samples (unassembled

reads) a total of 464 functional pathways were preliminarily identified by HUMAnN3. The output abundance tables were exported in Reads Per Kilobase of transcript and normalized using the HUMAnN3 relative abundance function. The relative abundance of these pathways was assessed by disease status (e.g., Naïve, Diseased Lesion, Diseased Apparently Healthy, and Apparently Healthy) within species as well as averaged by disease status across species. The most abundant classified pathway across all samples was the prokaryotic TCA cycle, which averaged 0.252% ± 0.088% of reads in diseased samples, 0.222% ± 0.061% of reads in diseased- apparently healthy samples, 0.243% ± 0.086% of reads in healthy samples, and 0.316% ± 0.186% of reads in naïve samples across all species. Also found in high relative abundance were the pathways for ATP biosynthesis, guanosine deoxyribonucleotides *de novo* biosynthesis, aerobic respiration via cytochrome c, and the ammonia assimilation cycle (**Table 3, Fig 1**). No statistically significant differences were found in abundance of pathways across either coral host species or disease status. Nonetheless, abundance of the pathways TCA Cycle I, ATP biosynthesis, and guanosine ribonucleotide *de novo* biosynthesis trended higher in apparently healthy and naïve samples than in diseased and diseased-apparently healthy samples. In contrast. The abundance of pathways associated with assimilatory nitrate reduction, assimilatory sulfate reduction IV, and assimilatory sulfate reduction I were elevated in samples of diseased and diseased-apparently healthy tissue in comparison to apparently healthy (no disease) and naïve samples. Pathways assigned to photosynthesis light reactions were elevated in diseased apparently healthy samples compared to other treatments. These patterns were consistent when examined within species (**Fig. 2**).

Table 3: *Average and standard deviation of relative abundance of 15 most highly abundant functional pathways, averaged across species within tissue status types.*

		APPARENTLY HEALTHY	HEALTHY	
TCA: TCA CYCLE I (PROKARYOTIC)	0.252 ± 0.175	0.222 ± 0.061	0.243 ± 0.86	0.316 ± 0.186
PWY-7980: ATP BIOSYNTHESIS	0.171 ± 0.196	0.253 ± 0.066	0.251 ± 0.115	0.251 ± 0.528
PWY-7226: GUANOSINE DEOXYRIBONUCLEOTIDES DE NOVO BIOSYNTHESIS I	0.228 ± 0.081	0.216 ± 0.092	0.248 ± 0.057	0.216 ± 0.053
PWY-3781: AEROBIC RESPIRATION I (CYTOCHROME C)	0.235 ± 0.164	0.153 ± 0.084	0.257 ± 0.092	0.212 ± 0.085
PWY-7221: GUANOSINE RIBONUCLEOTIDES DE NOVO BIOSYNTHESIS	0.144 ± 0.146	0.174 ± 0.259	0.271 ± 0.035	0.256 ± 0.060
AMMASSIM-PWY: AMMONIA ASSIMILATION CYCLE III	0.157 ± 0.083	0.247 ± 0.234	0.168 ± 0.011	0.209 ± 0.284
PWY-5690: TCA CYCLE II (PLANTS AND FUNGI)	0.216 ± 0.261	0.130 ± 0.087	0.245 ± 0.076	0.182 ± 0.040
PWY-6608: GUANOSINE NUCLEOTIDES DEGRADATION III	0.202 ± 0.139	0.196 ± 0.067	0.166 ± 0.046	0.204 ± 0.480
PWY-7663: GONDOATE BIOSYNTHESIS (ANAEROBIC)	0.139 ± 0.203	0.244 ± 0.074	0.202 ± 0.028	0.166 ± 0.061
PWY490-3: NITRATE REDUCTION VI (ASSIMILATORY)	0.221 ± 0.089	0.146 ± 0.056	0.172 ± 0.040	0.181 ± 0.155
PWY-6599: GUANINE AND GUANOSINE SALVAGE II	0.162 ± 0.093	0.151 ± 0.221	0.191 ± 0.399	0.151 ± 0.114
PWY-101: PHOTOSYNTHESIS LIGHT REACTIONS	0.057 ± 0.785	0.221 ± 0.111	0.090 ± 0.182	0.021 ± 0.177
PWY1ZNC-1: ASSIMILATORY SULFATE REDUCTION IV	0.121 ± 0.021	0.069 ± 0.041	0.054 ± 0.110	0.055 ± 0.051
SO4ASSIM-PWY: ASSIMILATORY SULFATE REDUCTION I	0.038 ± 0.742	0.095 ± 0.032	0.022 ± 0.097	0.046 ± 0.090
PWY-7111: PYRUVATE FERMENTATION TO ISOBUTANOL (ENGINEERED)	0.069 ± 0.210	0.058 ± 0.060	0.020 ± 0.113	0.018 ± 0.049

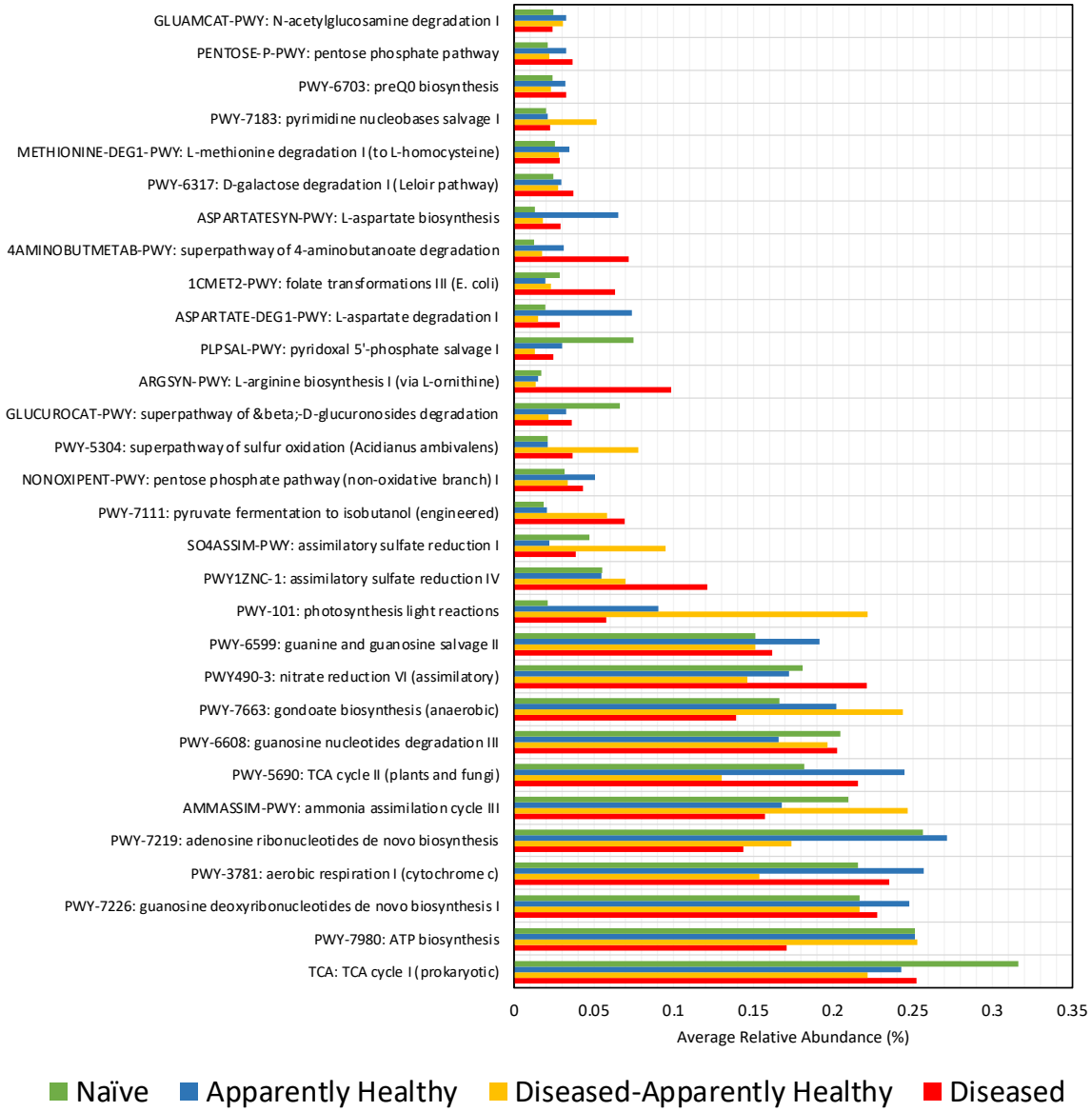


***Fig 1**: Relative abundance of the 30 most highly abundant pathways across all samples, as averaged by disease status.*

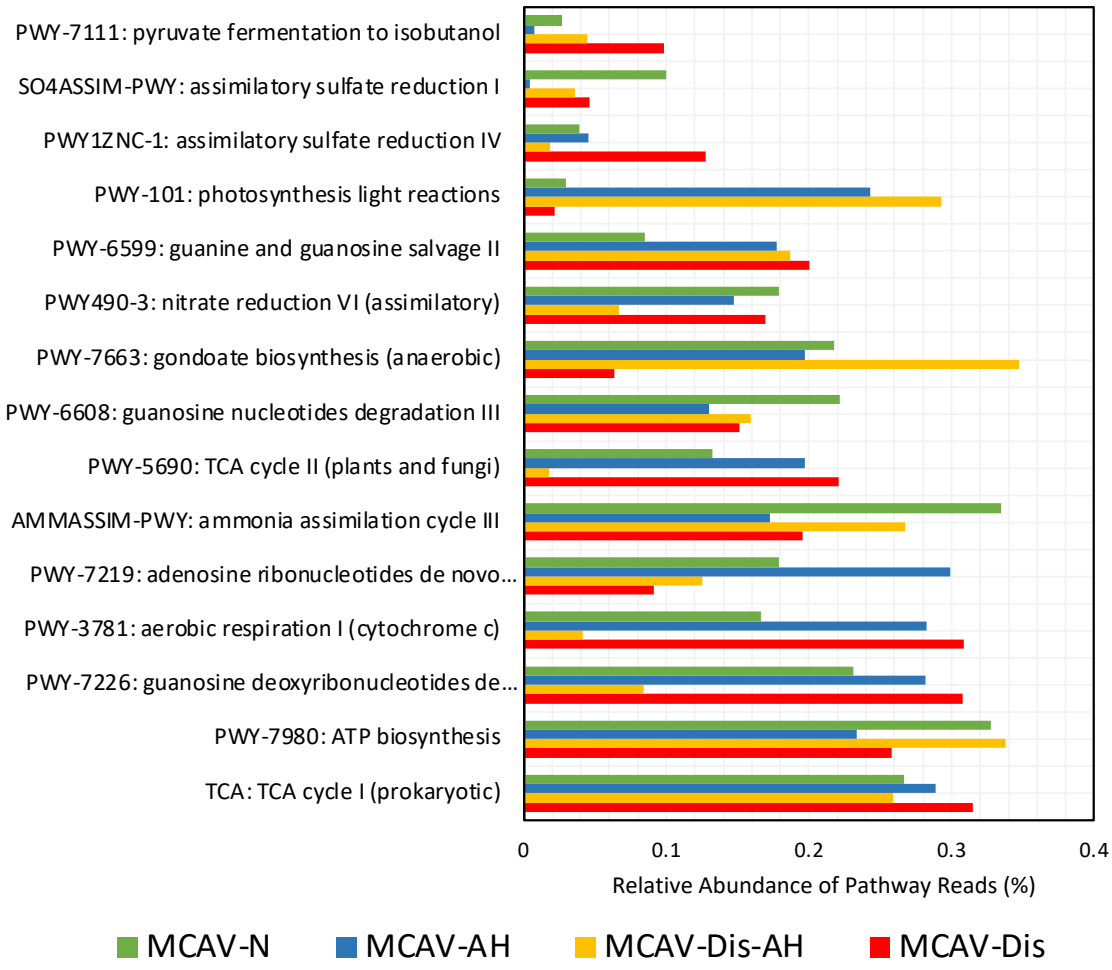


***Fig 2**. Relative abundance of the most highly abundant pathways across samples of Montastraea cavernosa (N = 32), as averaged by disease status.*

Taxonomic characterization of each sample was performed using MetaPhlAn4 (v 4.1.1, (Blanco-Miguez et al., 2023) using the ChocoPhlAn DNA database and the abundance of these taxonomic groups were compared across samples using the R package *corncob* (Martin et al., 2020). Corncob utilizes beta-binomial models to test hypotheses about the effects of covariates on relative abundance of microbial taxa from high throughput sequencing data. As samples of *M. cavernosa* had the most taxa successfully annotated (N = 521), we here present differential abundance data from this species (**Fig. 3**).

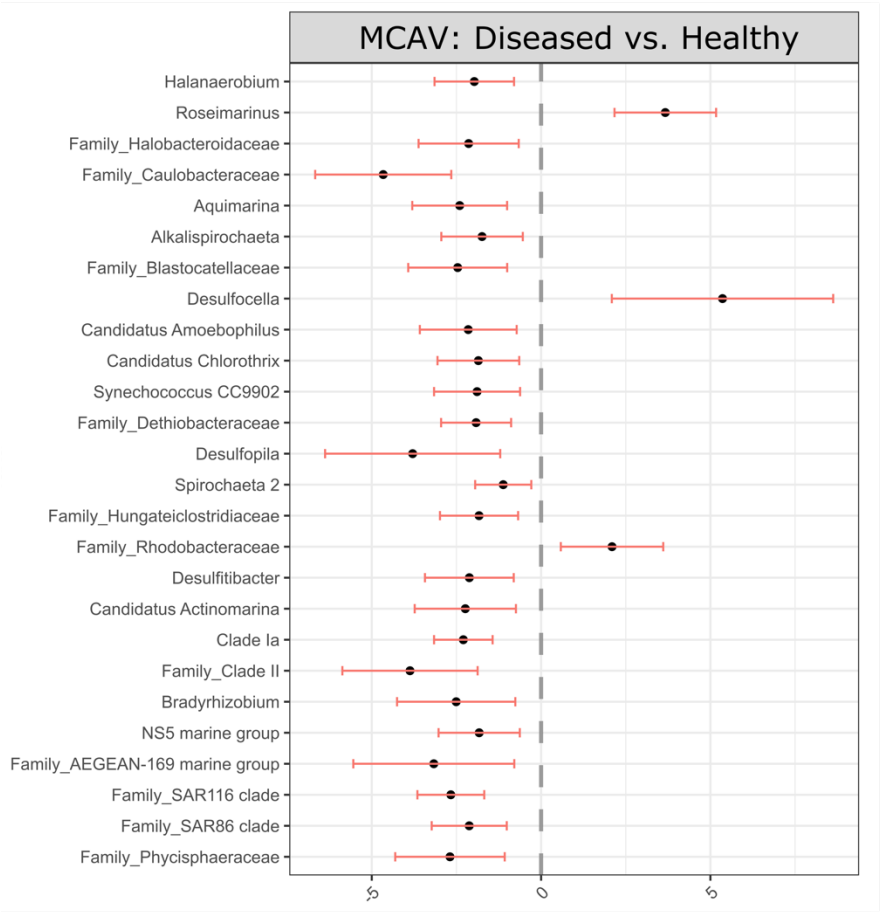


Fig 3. *Differential abundance of taxa found in samples of M. cavernosa as a function of health status, assessed using the R package 'corncob' (v 0.1.0). The relative abundance of each taxonomic group as identified by MetaPhlAn4 was modeled as a function of health status, and the coefficient indicates differences in ASV relative abundance with positive coefficients (x-axis) indicating taxa that are more abundant in diseased samples, and negative coefficients indicating taxa are less abundant. Bars represent 95% confidence intervals of abundance as modeled across samples within groups. A false discovery rate (FDR) cutoff of 0.05 was used.*

3.2. Viral Community Analyses

3.2.1. DNA sequencing virus contig assembly, prediction, and differential abundance analysis

There were 130,862,039 contigs assembled across species with 20,394,501 from *C. natans* libraries, 27,413,323 from *M. cavernosa* libraries, 37,486,939 from *O. faveolata* libraries, and 45,567,276 from *O. franksi* libraries. From the 130,862,039 total contigs, geNomad identified 54,642 as virus-like having homology to known virus sequences. vRhyme binned 12,177 virus-like contigs into 2,355 bins representing vMAGS. The vMAGs and the non-binned sequences (which may still represent individual viruses) were combined into one final virus

reference file containing a total of 44,820 sequences. Sequence lengths ranged from 300-359,601 base pairs long indicated genome/gene fragments as well as putative whole genomes. CheckV determined that 97% (43,509) of these sequences represent genome-fragments, while 1,311 sequences were of high-quality having high completeness (>90%). The program geNomad was able to provide taxonomic classifications to the realm level for 27,006 (40%) sequences, 21,727 belonged to Duplodnaviria (dsDNA viruses), 4,787 to Varidnaviria (dsDNA viruses), 527 to Riboviria (RNA viruses), and 20 to Monodnaviria (ssDNA viruses). All sequences that were assigned to the realm Duplodnaviria (the major fraction of geNomad-classified sequences) were determined to belong to class Caudoviricetes, a group of tailed prokaryotic viruses. As for eukaryotic viruses, sequences from classes of known algae-infecting viruses like Algavirales (n=51) and hypothesized coral-infecting viruses such as Herpesvirales (n=497) were also present.

Differential abundance analysis with DESeq2 revealed 31 sequences that were significantly more abundant in diseased tissue samples across all species. There were 26 sequences from the realm Duplodnaviria with the remaining five sequences lacking a realm assignment. The significantly upregulated Duplodnaviria sequences were assigned to class Caudoviricetes and only one of these sequences was able to be classified to a taxonomic level beyond class. The one sequence was classified as family Autographiviridae, a group of bacteria-infecting viruses (phages) that are thought to cause lytic infections of gram-negative bacteria.

3.2.2. RNA sequencing virus contig assembly, prediction, and differential abundance analysis

Assembly of RNA sequencing reads from all species yielded 8,978,237 assembled contigs. There were 1,740,640, 1,702,441, 2,976,155 and 2,559,001 from *C. natans, M. cavernosa, O. faveolata*, and *O. franksi* libraries, respectively. The initial Deep6 screening identified 1,837,778 virus-like transcripts that were then reduced to 990,976 by removing error-derived sequences and sequence redundancy with EvidentialGenes. An additional screening with Deep6 lead to the removal of an additional 17,331 producing a final, non-redundant virus-like reference file containing 973,645 sequences. The lengths of virus-like sequences ranged from 500-25,141 nucleotides. Many reference sequences (973,277) were determined to be genome-fragments (whole or partial transcripts) by CheckV, while 368 were of high-quality with high completeness (>90%). The taxonomy results sourced from Deep6 or geNomad indicated that 865,612 sequences were from Riboviria (RNA viruses), 48,426 from Varidnaviria (dsDNA viruses), 45,982 from Duplodnaviria (dsDNA viruses), and 13,713 from Monodnaviria (ssDNA viruses). Notably, sequences from suspected Symbiodiniaceae-infecting filamentous +ssRNA virus orders order Tymovirales (n=67) and Patavirales (n=8) were identified in the data.

DESeq2 analyses identified 128,791 sequences that were upregulated within diseased tissue sample type across species. The bulk of upregulated sequences were classified as RNA virus (n=117,838; realm Riboviria), although most (n=117,838) were unclassified past the taxonomic rank of realm. Of those

sequences that were able to be assigned to an order (n=42), 17 were retroviral (Ortervirales) and 15 were assigned to Picornavirales. Additionally, sequences assigned to Tymovirales (n=4) and Patatavirales (n=1) were upregulated in diseased tissue types (**Fig 4**)

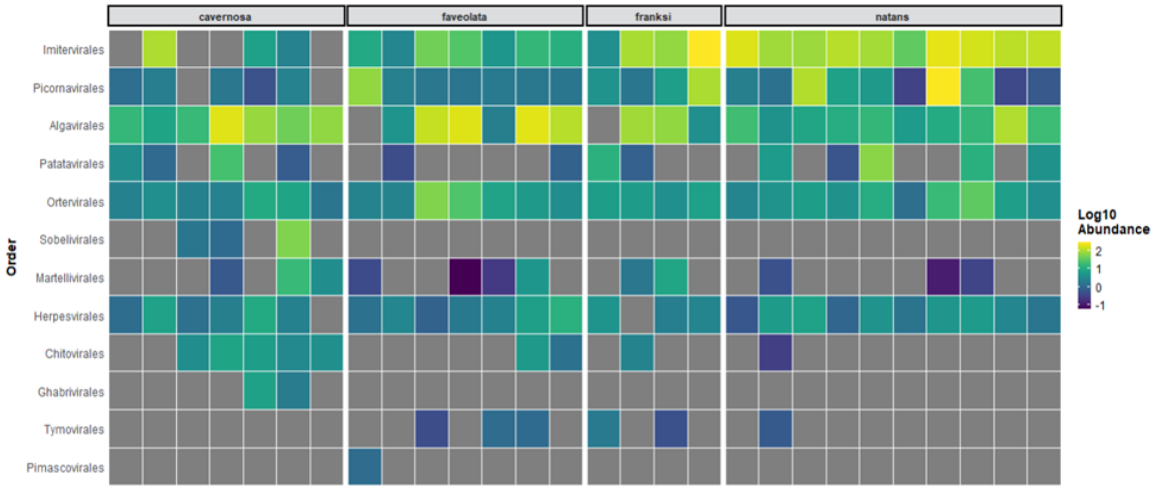


Fig 4: *Abundance heatmap visualizing increased abundances of sequences from RNA sequence data analysis determined to be upregulated in diseased tissue types across coral species by DESeq2 analysis. Each column represents a sample, and each row represents a virus order. Color indicates log10 relative abundance with yellow indicating high abundance and purple indicating lower relative abundance (to the samples included in the figure).*

3.3. Transcriptomic Analyses

3.3.1.Colpophyllia natans gene expression patterns

For *C. natans*, a total of 252,705,470 paired reads were assembled into a raw assembly of 865,329 transcripts. Following the cleaning steps, a final reference transcriptome with 44,208 transcripts and an N50 of 2,357 was obtained. The BUSCO scores for the final assembly were 89.2% complete (834 single-copy, 17 duplicate-copy, total = 851), 29 fragmented (3%), and 74 missing (7.8%). Comparison to the UniProtKB database resulted in a total of 42,049 annotated sequences (95.1%). Individual sample mapping rates to the resultant de novo transcriptome ranged from 59.78-87.2%.

Differential expression across all pairwise comparisons of *C. natans* samples anged from 141-293 differentially expressed genes (**Table 4**). The greatest differences in gene expression were observed between apparently healthy and disease margin tissue during June sampling points (sites at 25-50% disease prevalence). Few identifiable immune genes were differentially expressed between groups (~2-8% of total differentially expressed genes). However, several of these immune genes were consistently differentially expressed across healthy/apparently healthy samples (Sept healthy vs. June apparently healthy vs. August apparently healthy; **Fig 5, Table 5**)

***Table 4**: gene expression results of transcriptomic analysis of C. natans samples; H=healthy, AH= apparently healthy, DM= disease margin*

Comparison	Total Differentially Expressed Genes	Immune Differentially Expressed Genes
Sept_H vs. June_AH	141	3
Sept_H vs. Aug_AH	194	15
June_AH vs. Aug_AH	293	19
June_AH vs. June_DM	684	31

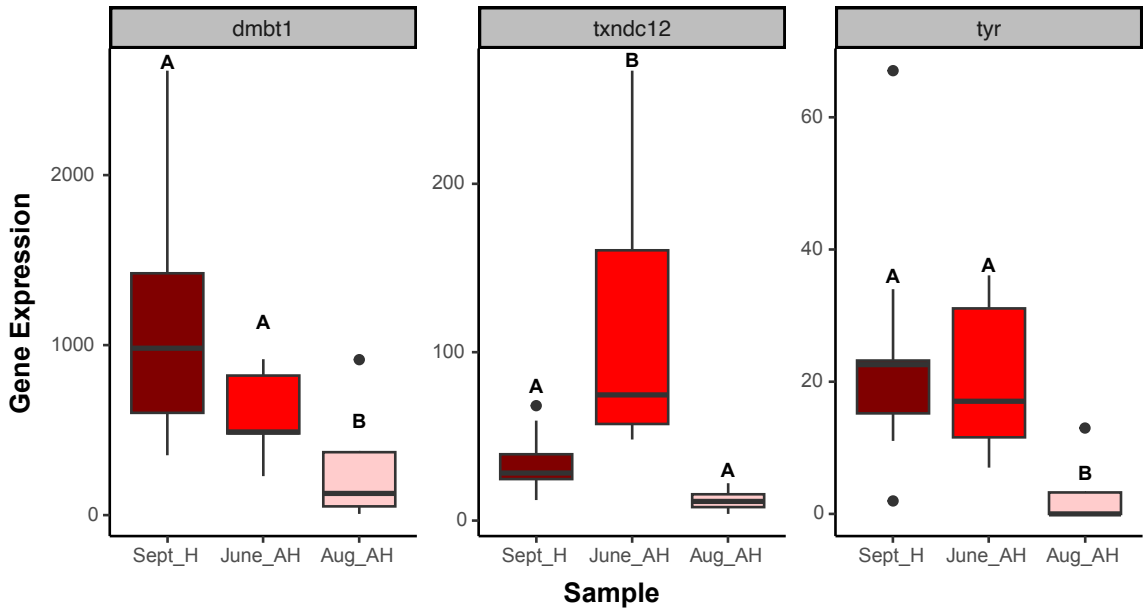


***Fig 4**: expression of select C. natanas immune genes which vary in expression across healthy (H) and apparently healthy (AH) tissue over time. Letters indicate statistical groups. p values for each comparison shown in **table 5**. dmbt1= deleted in malignant brain tumor protein 1; txndc12= Thioredoxin domain-containing protein 1; tyr= tyrosinase*

***Table 5**: differential expression results for C. natans comparing expression of key immune genes across healthy (H) and apparently heathy (AH) tissue over time; coef is a metric of the degree of change in expression between the compared sample types.*

Gene	Sept_H vs. June_AH coef	Sept_H vs. June_AH padj	Sept_H vs. Aug_AH coef	Sept_H vs. Aug_AH padj	June_AH vs. Aug_AH coef	June_AH vs. Aug_AH padj
dmbt1	0.766	0.546	3.024	**0.0137***	2.26	**0.0870***
txndc12	-1.66	**0.0624***	1.16	0.497	2.82	**0.0207***
tyr	-0.0386	0.996	3.69	**0.0671***	3.73	**0.0778***

3.3.2. Orbicella faveolata gene expression patterns

For *O. faveolata*, total of 256,119,936 paired reads were assembled into a raw assembly of 1,187,062 transcripts. Following the cleaning steps, a final reference transcriptome with 81,810 transcripts and an N50 of 1,823 was obtained. The BUSCO scores for the final assembly were 93.5% complete (848 single-copy, 44 duplicate-copy, total = 892), 23 fragmented (2.4%), and 39 missing (4.1%). Comparison to the UniProtKB database resulted in a total of 78,610 annotated sequences (96.1%). Individual sample mapping rates to the resultant de novo transcriptome ranged from 68.57-88.75%.

Differential expression across all pairwise comparisons of *O. faveolata* samples ranged from 11-1031 differentially expressed genes (**Table 6**). The greatest differences in gene expression were observed between healthy samples collected in September and August. Few identifiable immune genes were differentially expressed between groups (~5% of total differentially expressed genes for most comparisons). However, several of these immune genes were consistently differentially expressed across healthy/apparently healthy samples (Sept s. June vs. August healthy; **Fig 5, Table 7**)

***Table 6**: gene expression results of transcriptomic analysis of O. faveolata samples; H=healthy, AH= apparently healthy, DM= disease margin*

Comparison	Total Differentially Expressed Genes	Immune Differentially Expressed Genes
Sept_H vs. June_H	274	14
Sept_H vs. Aug_H	1031	45
June_H vs. Aug_H	97	6
Sept_H vs. June_AH	63	3
June_H vs. June_AH	11	0
June_H vs. June_DM	83	4
June_AH vs. June_DM	67	3

***Table 7**: differential expression results for O. faveolata comparing expression of key immune genes across healthy (H) tissue over time; coef is a metric of the degree of change in expression between the compared sample types.*

Gene	Sept_H vs. June_H coef	Sept_H vs. June_H padj	Sept_H vs. Aug_H coef	Sept_H vs. Aug_H padj	June_H vs. Aug_H coef	June_H vs. Aug_H padj
dmbt1	2.39	**0.0621***	2.59	**0.0866***	0.198	0.995
endod1	3.60	**0.0587***	3.57	**0.09998***	-0.0248	0.9997
gp2	1.88	**0.0762***	2.14	**0.0973***	0.263	0.991
mmp25	2.12	**0.0492***	2.42	**0.0673***	0.296	0.990
rnf213a	-2.18	**0.0484***	-2.796	**0.0133***	-0.613	0.863

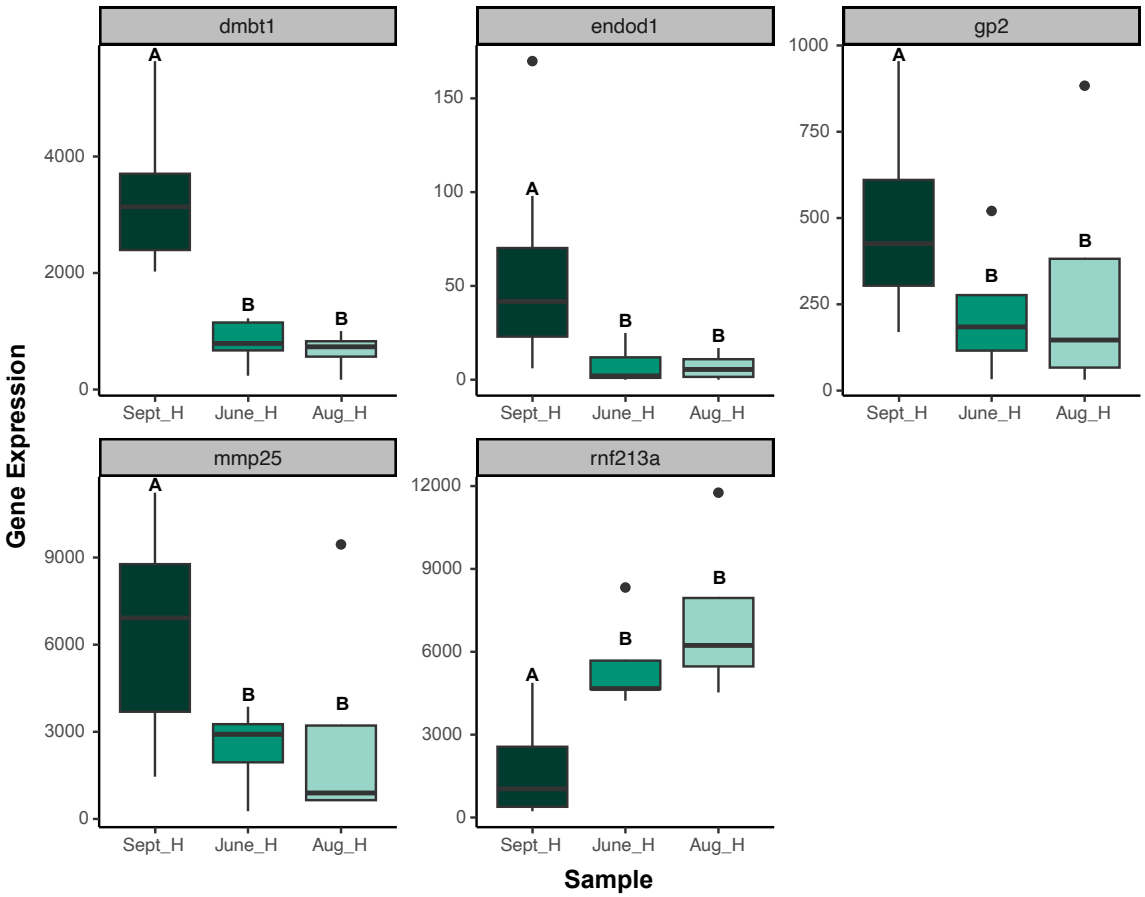


Fig 5*: expression of select O. faveolata immune genes which vary in expression across healthy (H) tissue over time. Letters indicate statistical groups. p values for each comparison shown in* ***table*** *7. dmbt1= deleted in malignant brain tumor protein 1; endod1= endonuclease domain-containing protein 1; gp2= pancreatic secretory granule membrane major glycoprotein GP2; mmp25= matrix metalloproteinase-25; rnf213a= E3 ubiquitin-protein ligase rnf213-alpha*

3.3.3. Orbicella franksi gene expression patterns

For *O. franksi*, a total of 273,084,162 paired reads were assembled into a raw assembly of 1,326,040 transcripts. Following the cleaning steps, a final reference transcriptome with 77,416 transcripts and an N50 of 1,951 was obtained. The BUSCO scores for the final assembly were 91.7% complete (818 single-copy, 57 duplicate-copy, total = 875), 28 fragmented (2.9%), and 51 missing (5.4%). Comparison to the UniProtKB database resulted in a total of 74,404 annotated sequences (96.1%). Individual sample mapping rates to the resultant de novo transcriptome ranged from 71.03-87.9%.

Differential expression across all pairwise comparisons of *O. franski* samples ranged from 219-10644 differentially expressed genes (**Table 8**). The greatest differences in gene expression were observed between healthy samples collected in September and August. Few identifiable immune genes were differentially expressed between groups (~4-9% of total differentially expressed genes). However, several of these immune genes were consistently differentially expressed across healthy/apparently healthy samples (Sept s. June vs. August healthy; **Fig 6, Table 9**)

Table 8*: gene expression results of transcriptomic analysis of O. franksi samples; H=healthy, AH= apparently healthy, DM= disease margin*

Comparison	Total Differentially Expressed Genes	Immune Differentially Expressed Genes
Sept_H vs. June_H	4876	192
Sept_H vs. Aug_H	10644	561
June_H vs. Aug_H	2315	202
Sept_H vs. Aug_AH	1478	70
June_H vs. Aug_AH	219	15

***Table 9**: differential expression results for O. franksi comparing expression of key immune genes across healthy (H) tissue over time; coef is a metric of the degree of change in expression between the compared sample types.*

Gene	Sept_H vs. June_H coef	Sept_H vs. June_H padj	Sept_H vs. Aug_H coef	Sept_H vs. Aug_H padj	June_H vs. Aug_H coef	June_H vs. Aug_H padj
dapk	0.941	**0.0823***	2.50	**0.00383****	1.56	**0.0430***
dmbt1.a	1.29	**0.0382***	2.19	**0.01303***	0.901	0.287
dmbt1.b	1.46	**0.0749***	3.34	**0.0106***	1.878	0.124
glyp	0.777	**0.0568***	1.70	**0.00547****	0.920	**0.0962***
otulin	0.952	**0.0923**	2.50	**0.00514****	1.55	**0.0584***
pde2a	0.919	**0.0786***	2.33	**0.00429****	1.41	**0.0560***
rnf213b	1.69	**0.0447***	4.89	**0.00749****	3.20	**0.0554***
znfx1.a	1.13	**0.0462***	2.50	**0.00393****	1.36	**0.0767***
znfx1.b	1.81	**0.0374***	4.426	**0.00324****	2.62	**0.0420***

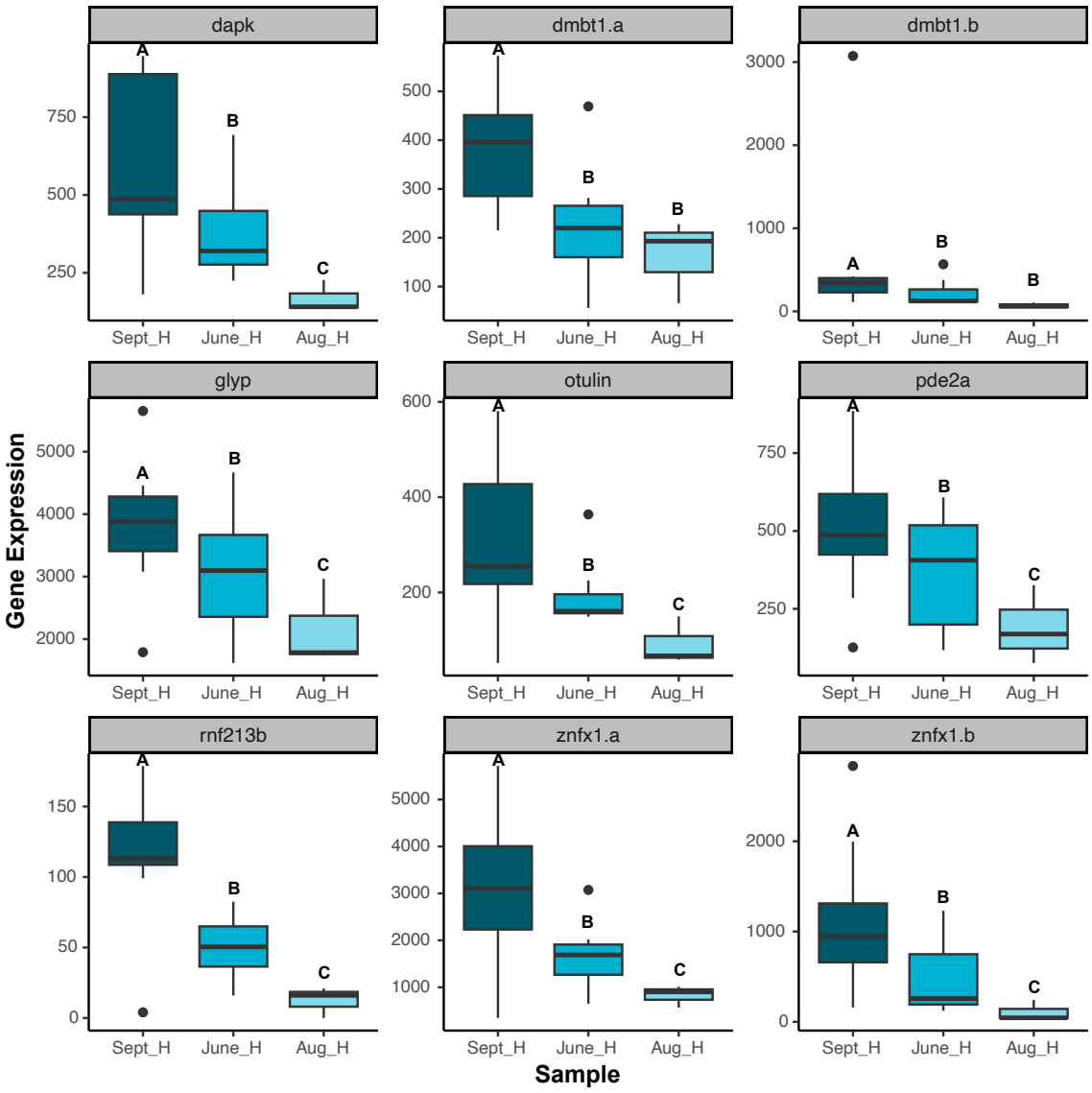


***Fig 6**: expression of select O. franksi immune genes which vary in expression across healthy (H) tissue over time. Letters indicate statistical groups. p values for each comparison shown in **table 9**. dapk= death-associated protein kinase 1; dmbt1= deleted in malignant brain tumor protein 1; glyp= glycogen phosphorylase; otulin= Ubiquitin thioesterase otulin; pde2a = cGMP-dependent 3',5'-cyclic phosphodiesterase;rnf213b= E3 ubiquitin-protein ligase rnf213-beta; znfx1= NFX1-type zinc finger-containing protein 1*

3.3.4. Montastraea cavernosa gene expression patterns

For *M. cavernosa,* a total of 286,186,657 paired reads were assembled into a raw assembly of 993,661 transcripts. Following the cleaning steps, a final reference transcriptome with 86,030 transcripts and an N50 of 1,630 was obtained. The BUSCO scores for the final assembly were 93% complete (857 single-copy, 31 duplicate-copy, total = 888), 28 fragmented (2.9%), and 38 missing (4.1%). Comparison to the UniProtKB database resulted in a total of 82,244 annotated sequences (95.6%). Individual sample mapping rates to the resultant de novo transcriptome ranged from 40.15-85.97%.

No differences in gene expression were observed between healthy colonies sampled across time points (Sept, June, August)

3.4. Histological Analyses

Preliminary histological analysis revealed *O. faveolata* and *M. cavernosa* to show traditional SCTLD signs of higher vacuolization in diseased samples than in healthy (**Fig. 7**). There is high exocytosis in healthy samples, which is different from previous SCTLD analyses. There is a larger minimum size of symbionts in diseased samples as well.

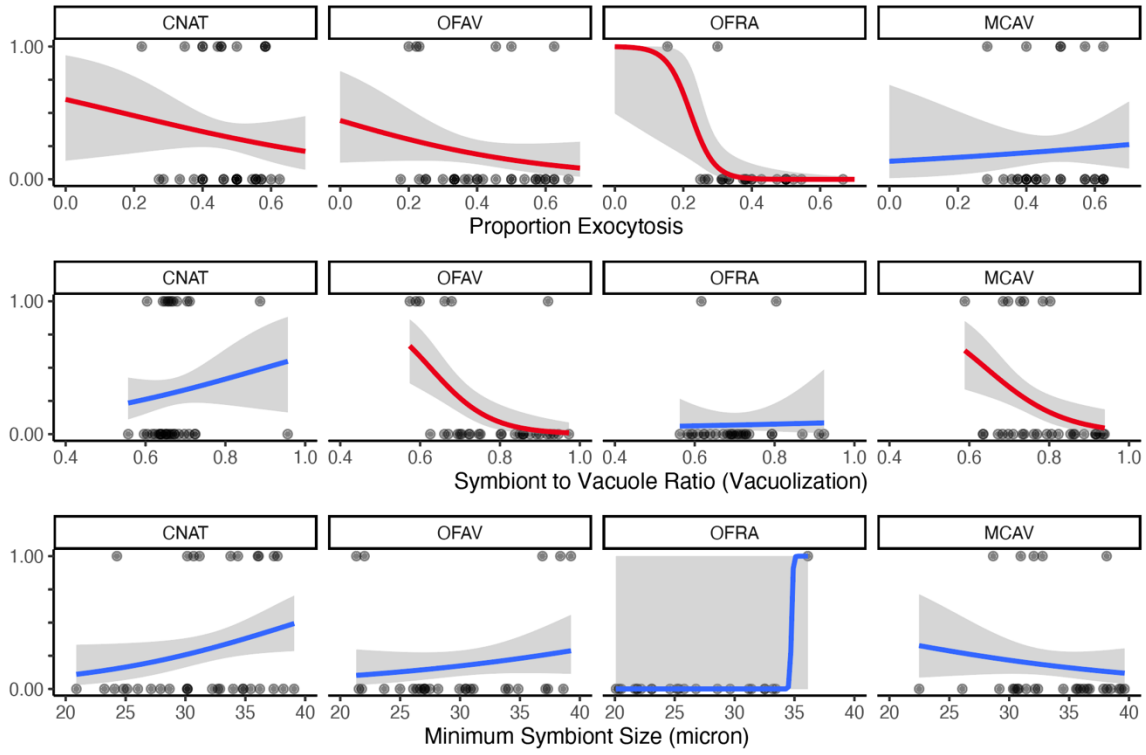


***Fig** 7: Logistic regression depicting Bayesian analyses, on Y-axis, one indicates diseased and zero indicates healthy.*

4. DISCUSSION AND MANAGEMENT RECOMMENDATIONS

4.1. Discussion

4.1.1. Metagenomics

Initial analyses of coral microbial function and taxonomy based on shotgun sequencing data suggest that while microbial communities of corals naïve to SCTLD in Dry Tortugas National Park may differ from other corals sampled in the Florida Keys, microbial community function and composition of diseased samples from this study were similar to samples that have been previously sequenced.

Previously, the taxa *Vibrio, Arcobacter, Algicola,* and *Planktotalea* were found to be consistently enriched in lesion tissue in the coral species *Montastraea cavernosa*, *Diploria labyrinthiformis* and *Dichocoenia stokesii* (Meyer et al., 2019). In contrast, a separate study found members of Rhodobacterales and Rhizobiales to be enriched in lesions from *Stephanocoenia intersepta*, *D. labyrinthiformis*, *D. stokesii* and *Meandrina meandrites* (Rosales et al., 2020). In this study, we have so far identified a significant enrichment in the family Rhodobacteraceae and the genus *Desulfocella*, though examination of the 16S rRNA from shotgun sequencing data in the next year of analysis will continue to elucidate the taxonomic profiles of our samples. Although not found to be significant from metagenomic data, in a separate analysis of 16S rRNA amplicon sequence data from these same samples (with significantly lower sequencing depth than the present study) we determined that the bacterial genus *Endozoicomonas* was found in considerable abundance across naïve samples of the species *Orbicella faveolata* and *O. franksi* (Klinges, unpublished). Interestingly, this genus, which is a proposed commensal or beneficial taxon in coral hosts, was not found at >2% abundance in any samples of *Colpophyllia natans* or *Montastraea cavernosa*. The absence of this proposed symbiont may contribute to the observed elevated disease susceptibility we saw in these species.

Analysis of community functional capacity based on shotgun sequencing data indicated that aerobic microbial activity was high across all species and health status types. We found elevated abundance of reads mapping to PWY-3781 associated with aerobic respiration via cytochrome *c* and the TCA cycle (involved in energy production by aerobic microbes). These pathways may be associated with bacteria identified in 16S rRNA analyses including the genera *Endozoicomonas*, *Tistlia*, and *Terasakiellaceae*, which are all aerobic. Pathways associated with aerobic respiration via cytochrome *c* were found to be elevated across all SCTLD disease status types in previous studies (Rosales et al., 2022). However, we also found high abundance of pathways associated with sulfate reduction and oxidation as well as nutrient utilization (ammonia assimilation and nitrate reduction). These pathways may be associated with bacteria involved in sulfur cycling, which have been previously identified as a biomarker for SCTLD in *Montastraea cavernosa* (Becker et al., 2022)\ In 16S rRNA analyses, we uncovered the high abundance of the sulfur reducing taxa *Desulfocella*, *Desulfovibrio*, and *Halodesulfovibrio*, as well as numerous anaerobic taxa including *Roseimarinus, Fusibacter,* and *Halanaerobium* that may be utilizing nitrate during organic matter oxidation. Pathways associated with sulfate reduction were elevated in diseased corals in comparison to healthy conspecifics, though this was not found to be statistically significant, possibly due to unbalanced sample sizes. The production and degradation of the nucleosides adenosine (a component of ATP) and guanosine were elevated across all species and health status types.

We also found a high relative abundance of reads mapping to pathways traditionally associated with plants, such as the plant/fungal TCA cycle and photosynthesis light reactions. The elevated abundance of these pathways may be reflective of remaining algal symbiont reads in our dataset, but it is also possible that these pathways are associated with cyanobacteria. We will continue to explore this possibility through the characterization of remaining prokaryotic and eukaryotic reads in the dataset.

4.1.2. Viral Community

Several lines of evidence suggest a contribution of coral-associated virus communities to stony coral tissue loss disease, specifically in the observed disruption of the coral-dinoflagellate symbiosis (Landsberg et al., 2020). This includes: 1. i) impeded or ceased lesion progression in dinoflagellate-free (or "bleached") coral tissues (Meiling et al., 2020), ii) degredation of symbiosomes within coral gastrodermal cells (Landsberg et al., 2020), iii) transmission electron microscopy detection of putative filamentous virus-like particles associated with Symbiodiniaceae cells in SCTLD-affected and SCTLD-exposed corals (Work et al., 2021), iv) *in silico* chracterization of filamentous virus-like genomes in SCTLD-affected/exposed coral metatranscriptomes (Veglia et al., 2022), v) upregulation of antiviral pathways in SCTLD-affected coral metatranscriptomes (Beavers et al., 2023), vi) evidence of upregulation of multiple virus taxa within diseased and disease-exposed coral metatranscriptomes (Vega Thurber & Correa, 2023; Veglia, 2023). In the first year of this project, the goal was to identify all putative virus sequences and determine the virus taxa determined to be upregulated in diseased tissue types and compare to previous findings from US Virgin Islands and Florida.

The running hypothesis for the role/contribution of viruses to SCTLD etiology is that of a community contribution in the form of opportunistic infections (Vega Thurber & Correa, 2023; Veglia, 2023). In this scenario, an initial etiological agent initiates dysbiosis within the coral holobiont and this is then further exacerbated by increased infection activity from several virus taxa capable of infecting the coral or its symbionts. The preliminary results from DNA and RNA sequencing data analyses in this project indicate similar findings where several prokaryotic and eukaryotic virus taxa are upregulated in the diseased coral tissue samples (**Fig 4**). Similar to the previous studies, notable virus taxa that are hypothesized to interact with Symbiodiniaceae like Algavirales, Imitervirales, Tymovirales, and Patatavirales were seen with increased abundance in diseased tissue samples. Algavirales and Imitervirales are both giant dsDNA viruses and their detection across coral genera containing distinct Symbiodiniaceae communities suggests that these virus groups are common in healthy and diseased corals. However, their ecology and the impact of their interactions with the putative Symbiodiniaceae host within coral holobionts is completely unknown emphasizing the need for future research in this area. Furthermore, the +ssRNA filamentous RNA viruses within Tymovirales and Patatavirales have been of great interest in the context of SCTLD (Veglia et al., 2022; Work et al., 2021). These viruses are also present within healthy and diseased coral tissues indicating

a potential for opportunistic infections and a contribution to SCTLD. That said, this would not suggest these viruses are the sole etiological agent of SCTLD, agreeing with previous observations (see (Vega Thurber & Correa, 2023; Veglia, 2023; Veglia et al., 2022; Work et al., 2021).

Viruses are known to influence host immune properties leaving them susceptible or more prepared to fight incoming pathogens (Neil & Cadwell, 2018). One such group of viruses shown to be capable of this is herpes viruses, a group hypothesized to infect the coral host and is seemingly ubiquitous in coral holobionts (Thurber et al., 2017). Our results suggest that herpesviruses are prevalent across coral taxa on Dry Tortuga reefs and their community diversity and activity may contribute to differences in SCTLD susceptibility and severity. There is a significant amount of work needed to be done to obtain foundational information for several of the common virus groups commonly reported in coral metagenomes/metatranscriptomes. Unfortunately, without this foundational information, it is difficult to interpret results like those shown here. Future work should look to investigate these groups (i.e., Algavirales, Imitervirales, and Herpesvirales) and produce empirical data related to their prevalence and diversity across individuals within and across coral species.

4.1.3. Transcriptomics

Preliminary analyses of gene expression patterns across time and tissue types revealed strong, species-specific patterns of gene expression. Herein we will focus our discussion on patterns of differential expression associated with comparisons of healthy and apparently healthy tissue as these were our most robust comparisons to date. Changes in gene expression in healthy and apparently healthy tissue over time were highly species-specific. While healthy colonies of *Montastrea cavernosa* displayed no significant differences in gene expression over the nearly one-year sampling period (and disease progression to 50%), *Orbicella franksi* colonies showed strong changes in gene expression. Over 10,000 transcripts were differentially expressed between colonies sampled in September, and those sampled the following August. *O. faveolata* and *C. natans* showed intermediate patterns of changes.

In all three species with significant differences, changes in expression of immune transcripts were more subtle, with identifiable immune transcripts comprising anywhere from ~2-8% of differentially expressed transcripts. However, these subsets of transcripts did display notable patterns of rate and directionality of changes, especially when comparing closely related *O. faveolata* and *O. franksi*. Expression of most immune transcripts in both species declined through the sampling periods, with the rate of changes in theses transcripts in healthy colonies coinciding with peak disease incidence. The most dramatic changes in gene expression in *O. faveolata* were observed between September and June (peak disease abundance), stabilizing in August. In contrast, *O. franksi* gene expression steadily declined from September thru June to August (peak disease abundance). The discordance in gene expression patterns between the two species suggests that the patterns were indeed driven by increases in disease prevalence rather than responses to environmental conditions

A few notable conclusions can also be drawn when considering patterns of response of specific immune genes. First, patterns of changes in immune genes did not show any noticeable trends towards antiviral or antibacterial pathways. Both viruses (Beavers et al., 2023; Work et al., 2021) and bacterial (Rosales et al., 2020; Rosales et al., 2023) agents have been suggested as the putative cause of SCTLD, with viral agents being the most widely recognized putative pathogen. Current analysis of gene expression patterns fails to support one putative agent over another. However, upcoming analyses using more integrative and comprehensive approaches should prove more fruitful in teasing apart these signatures and aiding in pathogen identification.

Several key immune transcript and processes were commonly differentially expressed across comparisons and species. The most notable of these is changes in the putative antimicrobial protein, deleted in malignant brain tumor protein 1 (DMBT1). DMBT1 is a component of TLR and NFKB signaling, (Bikker et al., 2002), which lead to the production of antimicrobial compounds. This gene has frequently been cited as an important component of the immune response of numerous tropical corals (Fuess et al., 2016; Wright et al., 2017). In all three species where patterns of differential expression were observed (*C. natans, O. faveolata, O. franksi*), DMBT1 expression decreased through disease progression on reefs, both in apparently healthy (*C. natans*) and healthy (*Oribcellas*) tissue. This downregulation of an important immune gene as disease reaches reefs suggests potential immune suppression of even healthy corals. A second important transcript involved in bacterial responses, E3 ubiquitin-protein ligase rnf213 (RNF213), was also differentially expressed across healthy tissues in the *Orbicella sp.* RNF213 is key for detection of bacterial LPS and activation of immune responses (Zhang et al., 2023). Notably, RNF213 homologs increased in expression through disease onset in *O. faveolata*, but decreased over the same period in *O. franksi*. Further investigation regarding changes in expression of key antimicrobial genes is essential to understand the roles of these changes in either facilitating or preventing disease.

4.1.4. Histology

Preliminary analyses indicate the disease seen at Dry Tortugas is consistent with SCTLD analyses previously done. Vacuolization of the symbiont is a consistent disease sign for SCTLD-infected colonies. The healthy samples having higher exocytosis than the diseased is odd, but could be indicative of a higher turnover of symbionts for shuffling.

4.2. Future Steps

We have generated an unprecedented sequencing data set comprising both sequencing (DNA/RNA) and phenotypic (histology/TEM) data. This sample set is notable both for its thoroughness (high degree of within-sample sequencing) and breadth (large number of samples per species and repeated sampling across time). In addition, samples from this project were sequenced to a very high depth (10's of millions of reads per sample). Consequently, this dataset presents a unique and unparalleled wealth of biological data pertinent to SCTLD; specifically, our sample set can be mined to gain new insight regarding putative etiological agents, biological progression and stages of disease, and markers of disease resistance. Currently we have

only just begun to scratch the surface of what can be done with this data, providing putative markers of disease resistance and etiological agents using species- and data-type-independent analyses. Moving forward we will focus on the development of robust, integrative analytical approaches which will leverage the true power of this data set. Specifically, supported by FY25 DEP funding, we will use advanced 'omic analyses to: 1) identify with confidence putative etiological agents and their effects on host and symbiont biology; 2) identify host and symbiont traits of interest which confer disease resistance, or lack thereof (both species-specific and species-independent); 3) compare biological progression of disease across species to improve understanding of disease mechanisms and potentially develop new prophylactic techniques.

4.3. Management Recommendations

As one of the first SCTLD time-series analyses performed on samples of corals from pre-exposure to exposure to disease development, this study has considerable implications for our understanding of disease development, especially in remote coral reefs.

- Corals in this study developed the hallmark histological signs of SCTLD despite their location in a remote region of the lower Florida Keys far from significant local stressors. Indeed, it is possible that their stress-naïve state pre-exposure made them more susceptible to SCLTD, as corals appeared to exhibit signs of immunosuppression during disease development. Practitioners and managers should ensure that corals in the remote reaches of the Caribbean are regularly monitored for disease development, and further studies should be performed in these remote areas to validate these results.
- Despite the initial presence of beneficial and commensal bacterial taxa including *Endozoicomonas*, corals in this study nonetheless developed disease and beneficial communities were lost during disease development. Further research into probiotics could include methods to enhance and maintain these protective microbial communities effectively even under disease pressure.
- As coral immune function during SCTLD development varied by species, even within the same coral genus, restoration projects should include a genetically diverse array of corals to enhance the overall resilience of reefs to SCTLD and other stressors. Genetic diversity can offer a range of immune responses and adaptive capacities.
- Further study is needed to investigate the mechanisms behind the differences in immune responses between coral species. Understanding these mechanisms can lead to the development of targeted therapies and better inform species selection for restoration.
- Integrative analysis is needed between datasets produced in this study to evaluate the interaction of coral gene expression, the abundance of viral and bacterial groups, bacterial function, and disease histopathology. This integrative analysis will better inform our understanding of the relative contributions of bacterial and viral players to disease development and the relationship to coral immunity and pathogenesis.

- Continued research is needed on the role of environmental factors (e.g., water temperature, pollution) in SCTLD development and spread, in particular as bacterial genes involved in nutrient and sulfur cycling were highly abundant in diseased samples. Understanding these factors can inform broader conservation strategies that address multiple stressors impacting coral health.

By integrating these recommendations into management strategies, we can enhance our ability to mitigate SCTLD impacts on coral reefs, promote coral resilience, and ensure the long-term health of these vital ecosystems. The finding that even remote, previously healthy corals can succumb to SCTLD emphasizes the need for proactive and comprehensive management approaches.

5. REFERENCES CITED

Aranda, M., Li, Y., Liew, Y. J., Baumgarten, S., Simakov, O., Wilson, M. C., Piel, J., Ashoor, H., Bougouffa, S., Bajic, V. B., Ryu, T., Ravasi, T., Bayer, T., Micklem, G., Kim, H., Bhak, J., LaJeunesse, T. C., & Voolstra, C. R. (2016, Dec 22). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci Rep, 6*, 39734. https://doi.org/10.1038/srep39734

Beavers, K. M., Van Buren, E. W., Rossin, A. M., Emery, M. A., Veglia, A. J., Karrick, C. E., MacKnight, N. J., Dimos, B. A., Meiling, S. S., Smith, T. B., Apprill, A., Muller, E. M., Holstein, D. M., Correa, A. M. S., Brandt, M. E., & Mydlarz, L. D. (2023, May 22). Stony coral tissue loss disease induces transcriptional signatures of in situ degradation of dysfunctional Symbiodiniaceae. *Nat Commun, 14*(1), 2915. https://doi.org/10.1038/s41467-023-38612-4

Becker, C. C., Brandt, M., Miller, C. A., & Apprill, A. (2022, Mar). Microbial bioindicators of Stony Coral Tissue Loss Disease identified in corals and overlying waters using a rapid field-based sequencing approach. *Environ Microbiol, 24*(3), 1166-1182. https://doi.org/10.1111/1462-2920.15718

Beghini, F., McIver, L. J., Blanco-Miguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A., Manghi, P., Scholz, M., Thomas, A. M., Valles-Colomer, M., Weingart, G., Zhang, Y., Zolfo, M., Huttenhower, C., Franzosa, E. A., & Segata, N. (2021, May 4). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife, 10*. https://doi.org/10.7554/eLife.65088

Bellantuono, A. J., Dougan, K. E., Granados-Cifuentes, C., & Rodriguez-Lanetty, M. (2018). https://doi.org/10.1101/508184

Bikker, F. J., Ligtenberg, A. J., Nazmi, K., Veerman, E. C., van't Hof, W., Bolscher, J. G., Poustka, A., Nieuw Amerongen, A. V., & Mollenhauer, J. (2002, Aug 30). Identification of the bacteria-binding peptide domain on salivary agglutinin (gp-

340/DMBT1), a member of the scavenger receptor cysteine-rich superfamily. *J Biol Chem, 277*(35), 32109-32115. https://doi.org/10.1074/jbc.M203788200

Blanco-Miguez, A., Beghini, F., Cumbo, F., McIver, L. J., Thompson, K. N., Zolfo, M., Manghi, P., Dubois, L., Huang, K. D., Thomas, A. M., Nickols, W. A., Piccinno, G., Piperni, E., Puncochar, M., Valles-Colomer, M., Tett, A., Giordano, F., Davies, R., Wolf, J., Berry, S. E., Spector, T. D., Franzosa, E. A., Pasolli, E., Asnicar, F., Huttenhower, C., & Segata, N. (2023, Nov). Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol, 41*(11), 1633-1644. https://doi.org/10.1038/s41587-023-01688-w

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016, May). Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol, 34*(5), 525-527. https://doi.org/10.1038/nbt.3519

Bushnell, B. *BBMap*. sourceforge.net/projects/bbmap/

Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S. G., Nayfach, S., & Kyrpides, N. C. (2023, Sep 21). Identification of mobile genetic elements with geNomad. *Nat Biotechnol*. https://doi.org/10.1038/s41587-023-01953-y

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018, Sep 1). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), i884-i890. https://doi.org/10.1093/bioinformatics/bty560

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021, Feb 16). Twelve years of SAMtools and BCFtools. *Gigascience, 10*(2). https://doi.org/10.1093/gigascience/giab008

Davies, S. W., Marchetti, A., Ries, J. B., & Castillo, K. D. (2016). Thermal and pCO2 Stress Elicit Divergent Transcriptomic Responses in a Resilient Coral. *Frontiers in Marine Science, 3*. https://doi.org/10.3389/fmars.2016.00112

Finke, J. F., Kellogg, C. T. E., & Suttle, C. A. (2023, Feb 16). Deep6: Classification of Metatranscriptomic Sequences into Cellular Empires and Viral Realms Using Deep Learning Models. *Microbiol Resour Announc, 12*(2), e0107922. https://doi.org/10.1128/mra.01079-22

Fuess, L. E., Pinzon, C. J., Weil, E., & Mydlarz, L. D. (2016, Sep). Associations between transcriptional changes and protein phenotypes provide insights into immune regulation in corals. *Dev Comp Immunol, 62*, 17-28. https://doi.org/10.1016/j.dci.2016.04.017

Gilbert, D. G. (2019). Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? https://doi.org/10.1101/829184

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., & Regev, A. (2011, May 15). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol, 29*(7), 644-652. https://doi.org/10.1038/nbt.1883

Haas, B., & Papanicolaou, A. (2016). TransDecoder (find coding regions within transcripts). *Google Scholar*.

Kieft, K., Adams, A., Salamzade, R., Kalan, L., & Anantharaman, K. (2022, Aug 12). vRhyme enables binning of viral genomes from metagenomes. *Nucleic Acids Res, 50*(14), e83. https://doi.org/10.1093/nar/gkac341

Landsberg, J. H., Kiryu, Y., Peters, E. C., Wilson, P. W., Perry, N., Waters, Y., Maxwell, K. E., Huebner, L. K., & Work, T. M. (2020). Stony Coral Tissue Loss Disease in Florida Is Associated With Disruption of Host–Zooxanthellae Physiology. *Frontiers in Marine Science, 7*. https://doi.org/10.3389/fmars.2020.576013

Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., Yamashita, H., & Lam, T. W. (2016, Jun 1). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods, 102*, 3-11. https://doi.org/10.1016/j.ymeth.2016.02.020

Li, H. (2021, Dec 7). New strategies to improve minimap2 alignment accuracy. *Bioinformatics, 37*(23), 4572-4574. https://doi.org/10.1093/bioinformatics/btab705

Li, W., & Godzik, A. (2006, Jul 1). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics, 22*(13), 1658-1659. https://doi.org/10.1093/bioinformatics/btl158

Martin, B. D., Witten, D., & Willis, A. D. (2020, Mar). Modeling Microbial Abundances and Dysbiosis with Beta-Binomial Regression. *Ann Appl Stat, 14*(1), 94-115. https://doi.org/10.1214/19-aoas1283

McMurdie, P. J., & Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One, 8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

Meiling, S., Muller, E. M., Smith, T. B., & Brandt, M. E. (2020). 3D Photogrammetry Reveals Dynamics of Stony Coral Tissue Loss Disease (SCTLD) Lesion

Progression Across a Thermal Stress Event. *Frontiers in Marine Science, 7*. https://doi.org/10.3389/fmars.2020.597643

Meyer, J. L., Castellanos-Gell, J., Aeby, G. S., Hase, C. C., Ushijima, B., & Paul, V. J. (2019). Microbial Community Shifts Associated With the Ongoing Stony Coral Tissue Loss Disease Outbreak on the Florida Reef Tract. *Front Microbiol, 10*, 2244. https://doi.org/10.3389/fmicb.2019.02244

Namiki, T., Hachiya, T., Tanaka, H., & Sakakibara, Y. (2012, Nov 1). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res, 40*(20), e155. https://doi.org/10.1093/nar/gks678

Nayfach, S., Camargo, A. P., Schulz, F., Eloe-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021, May). CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat Biotechnol, 39*(5), 578-585. https://doi.org/10.1038/s41587-020-00774-7

Neil, J. A., & Cadwell, K. (2018, Sep 15). The Intestinal Virome and Immunity. *J Immunol, 201*(6), 1615-1624. https://doi.org/10.4049/jimmunol.1800631

Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017, May). metaSPAdes: a new versatile metagenomic assembler. *Genome Res, 27*(5), 824-834. https://doi.org/10.1101/gr.213959.116

Parkinson, J. E., Baumgarten, S., Michell, C. T., Baums, I. B., LaJeunesse, T. C., & Voolstra, C. R. (2016, Feb 11). Gene Expression Variation Resolves Species and Individual Strains among Coral-Associated Dinoflagellates within the Genus Symbiodinium. *Genome Biol Evol, 8*(3), 665-680. https://doi.org/10.1093/gbe/evw019

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015, Jul). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res, 25*(7), 1043-1055. https://doi.org/10.1101/gr.186072.114

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017, Apr). Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods, 14*(4), 417-419. https://doi.org/10.1038/nmeth.4197

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015, Apr 20). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res, 43*(7), e47. https://doi.org/10.1093/nar/gkv007

Rosales, S. M., Clark, A. S., Huebner, L. K., Ruzicka, R. R., & Muller, E. M. (2020). Rhodobacterales and Rhizobiales Are Associated With Stony Coral Tissue Loss Disease and Its Suspected Sources of Transmission. *Front Microbiol, 11*, 681. https://doi.org/10.3389/fmicb.2020.00681

Rosales, S. M., Huebner, L. K., Clark, A. S., McMinds, R., Ruzicka, R. R., & Muller, E. M. (2022). Bacterial Metabolic Potential and Micro-Eukaryotes Enriched in Stony Coral Tissue Loss Disease Lesions. *Frontiers in Marine Science, 8*. https://doi.org/10.3389/fmars.2021.776859

Rosales, S. M., Huebner, L. K., Evans, J. S., Apprill, A., Baker, A. C., Becker, C. C., Bellantuono, A. J., Brandt, M. E., Clark, A. S., Del Campo, J., Dennison, C. E., Eaton, K. R., Huntley, N. E., Kellogg, C. A., Medina, M., Meyer, J. L., Muller, E. M., Rodriguez-Lanetty, M., Salerno, J. L., Schill, W. B., Shilling, E. N., Stewart, J. M., & Voss, J. D. (2023, Mar 9). A meta-analysis of the stony coral tissue loss disease microbiome finds key bacteria in unaffected and lesion tissue in diseased colonies. *ISME Commun, 3*(1), 19. https://doi.org/10.1038/s43705-023-00220-0

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015, Oct 1). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics, 31*(19), 3210-3212. https://doi.org/10.1093/bioinformatics/btv351

[Record #390 is using a reference type undefined in this output style.]

Thurber, R. V., Payet, J. P., Thurber, A. R., & Correa, A. M. (2017, Apr). Virus-host interactions and their roles in coral reef health and disease. *Nat Rev Microbiol, 15*(4), 205-216. https://doi.org/10.1038/nrmicro.2016.176

Vega Thurber, R. L., & Correa, A. M. S. (2023). *Meta-transcriptomics to determine if and how viruses are involved in SCTLD infection status and/or disease susceptibility*.

Veglia, A. J. (2023). *Detecting and interpreting viral dynamics in marine invertebrate holobionts* Rice University].

Veglia, A. J., Beavers, K., Van Buren, E. W., Meiling, S. S., Muller, E. M., Smith, T. B., Holstein, D. M., Apprill, A., Brandt, M. E., Mydlarz, L. D., & Correa, A. M. S. (2022, Feb 17). Alphaflexivirus Genomes in Stony Coral Tissue Loss Disease-Affected, Disease-Exposed, and Disease-Unexposed Coral Colonies in the U.S. Virgin Islands. *Microbiol Resour Announc, 11*(2), e0119921. https://doi.org/10.1128/mra.01199-21

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Work, T. M., Weatherby, T. M., Landsberg, J. H., Kiryu, Y., Cook, S. M., & Peters, E. C. (2021). Viral-Like Particles Are Associated With Endosymbiont Pathology in Florida Corals Affected by Stony Coral Tissue Loss Disease. *Frontiers in Marine Science, 8*. https://doi.org/10.3389/fmars.2021.750658

Wright, R. M., Kenkel, C. D., Dunn, C. E., Shilling, E. N., Bay, L. K., & Matz, M. V. (2017, Jun 1). Intraspecific differences in molecular stress responses and coral pathobiome contribute to mortality under bacterial challenge in Acropora millepora. *Sci Rep, 7*(1), 2609. https://doi.org/10.1038/s41598-017-02685-1

Wu, N., Zhong, M. C., Roncagalli, R., Perez-Quintero, L. A., Guo, H., Zhang, Z., Lenoir, C., Dong, Z., Latour, S., & Veillette, A. (2016, Apr). A hematopoietic cell-driven mechanism involving SLAMF6 receptor, SAP adaptors and SHP-1 phosphatase regulates NK cell education. *Nat Immunol, 17*(4), 387-396. https://doi.org/10.1038/ni.3369

Zhang, Y., Yuan, Y., Jiang, L., Liu, Y., & Zhang, L. (2023). The emerging role of E3 ubiquitin ligase RNF213 as an antimicrobial host determinant. *Front Cell Infect Microbiol, 13*, 1205355. https://doi.org/10.3389/fcimb.2023.1205355