

An Integrative 'Omic Approach Leveraging Historical SCTL D Data to Identify Predictive Markers of SCTL D Resistance



An Integrative 'Omic Approach Leveraging Historical SCTL D Data to Identify Predictive Markers of SCTL D Resistance

Final Report

Prepared By:

Lauren E. Fuess¹
Grace Klinges²
Alex Veglia³
Ashley Rossin⁴
Michael Studivan⁴
Blake Ushijima⁵

¹Texas State University, ²Arizona State University, ³University of Puerto Rico, Mayaguez, ⁴NOAA AOML, ⁵ University of North Carolina Wilmington

June 2025

Completed in Fulfillment of PO# C3D3C6 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:

(Fuess et al., 2025 Rossin A, Studivan M, Ushijima B. 2025. An Integrative 'Omic Approach Leveraging Historical SCTL D Data to Identify Predictive Markers of SCTL D Resistance. 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. Analyses of bacterial communities uncovered taxa previously believed to be indicators of SCTLD in apparently healthy samples collected from nearby diseased colonies, suggesting that site-wide coral communities are likely impacted by disease presence and affirming our previous conclusion that many apparently healthy samples in existing datasets do not represent truly naive colonies. Our investigation of coral virome dynamics continues to support a community-wide contribution to SCTLD etiology, marked by increased activity of core viral groups (those consistently present in healthy corals) in diseased tissues. While sequencing approaches revealed these shifts, preliminary transmission electron microscopy analysis did not show consistent patterns in virus-like particle abundance that could support pathogen identification or SCTLD diagnosis. Continued research on coral viruses on Florida reefs is critical to establish baseline knowledge, improve diagnostic power, and strengthen response efforts for SCTLD and future coral disease outbreaks. Our histological data integration is an important first step towards improved methods for coral diagnostics and disease identification. When considering transcriptomic data, our inability to identify species independent markers of disease resistance highlights the need for improved genomic resources for species of interest. Still, we demonstrate that early response to SCTLD is highly consistent across species and involves several arms of immunity, suggesting that the lack of ability to sustain this response may be a more direct cause of mortality. Intervention to prolong coral immune responses may benefit reefs.

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 improved 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.

Assessment of bacterial community function has revealed clear patterns of host and microbial responses associated with transitions from a healthy to diseased state. Key bacterial taxa associated with SCTLD, such as *Halarcobacter* and *Desulfocella*, were predictive of disease state and led to observed enrichment in genes associated with anaerobic and sulfur-related metabolism. Analyses of raw metagenomic reads, more so than complete genome assembly, have allowed for more complete functional profiling of the bacterial community while expanding our analyses past dominant microbial members. Bacterial functional groups were correlated with conserved host genes, which identified highly connected modules, particularly linking host genes for DNA repair and apoptosis with microbial anaerobic respiration and toxic metabolite production. These findings support a model in which disease progression is driven by a host-microbe feedback loop that impairs healing and accelerates tissue degradation.

Virus community analyses provided further evidence of increased activity among diverse core viral lineages (those consistently present in healthy corals) within diseased tissues, suggesting a cumulative viral contribution to SCTLD etiology rather than the involvement of a single novel viral pathogen. These patterns are consistent with upregulated antiviral responses observed in host gene expression data. However, no consistent patterns in virus-like particle abundance were observed in transmission electron microscopy (TEM) images from either *in situ* (this project) or *ex situ* (C21169; PI Ushijima) diseased tissue samples. Establishing core virome members for coral species inhabiting Florida's reefs, as done here, improves our ability to identify potential pathogens by helping distinguish novel invaders from active members of the resident viral community. Despite this progress, our ability to investigate coral virus communities remains limited by the lack of baseline data from non-diseased conditions.

Histological data generated herein has been combined with a larger set of disease histology in an important first step towards generation of improved methods for coral diagnostics from histological samples. This integration will allow for improved methods for diagnosing coral diseases and assessing health via histology in a variety of settings, including nurseries and protected areas.

Analysis of transcriptomic data failed to identify species-independent markers of disease resistance, but did find several species-independent markers of response, particularly at early stage of disease. These included orthologs related to immune recognition, reactive oxygen processing, inflammation, and antiviral processes. Additionally, integration of ortholog data with microbial and viral community data demonstrated strong impacts of proposed host and Symbiodinaceae symbiotic microbes on host gene expression and immunity, regardless of host species.

This project also included tasks associated with other FDEP projects, including a related laboratory disease challenge experiment with a time series aspect (C21169; PI- Ushijima). Immunological analysis of samples collected from that project did not find any differences in immune response between control and exposed colonies of *Montastraea cavernosa*. Immune activity did significantly change over time, likely as a result of experimental stress (lack of feeding, water changes, etc.), which may have dampened our ability to detect experimental effects.

Table of Contents

1. BACKGROUND	7
1.1. Introduction	7
1.2. Project Goals and Tasks.....	7
2. METHODS	7
2.1. Metagenomic Analyses	7
2.1.1. Taxonomic identification and diversity measures of additional bacterial and archaeal taxa	7
2.1.2. Assessment of microbial functional activity	8
2.1.3. Integrative analyses of metagenomic data with transcriptomic data	8
2.2. Viral Community Analyses	9
2.2.1. Processing of sequencing data for virus sequence detection	9
2.2.2. Virus sequence classification	10
2.2.3. Virus sequence quantification and functional annotation	10
2.2.4. Virus community & gene landscape analyses	10
2.3. Histological Analyses	11
2.4. Transcriptomic Analyses	12
2.4.1. De novo transcriptome refinement	13
2.4.2. Ortholog Analyses- Identification of species specific markers of SCTLD resistance	22
2.4.3. Integrative Analyses: Linking Transcriptomic, Microbial, Viral, and Histological Data	25
2.5. Immunological Analysis of samples from project C21169.....	26
2.6. Analysis of TEM images from coral sampled under projects C1E0A5 and C21169.....	27
3. RESULTS	28
3.1. Metagenomic Analyses	28
3.1.1. Analyses identifying bacteria and microeukaryotes from metagenomic data	28
3.1.2. Assessments of alpha and beta diversity from metagenomic data	28
3.1.3. Species-independent markers of disease status	30
3.1.4. Integrative analyses of metagenomic data with transcriptomic data	32
3.2. Viral Community Analyses	33
3.2.1. Virus sequence detection and classification success	33
3.2.2. Core virome characterization and evidence for phyllosymbiosis.....	34
3.3. Histological Analyses	40
3.4. Transcriptomic & Associated Integrative Analyses	40
3.4.1. Species independent markers of disease resistance.....	40
3.4.2. Species independent correlations between microbiome/viral community composition and histological markers and host gene expression	43
3.5. Immunological Analysis of samples from project C21169.....	47
3.6. Analysis of TEM images from coral sampled under projects C1E0A5 and C21169.....	50

3.6.1. TEM image analysis from Dry Tortugas corals sampled under C1E0A5	50
3.6.2. TEM image analysis from time series experiment under project C21169	53
4. DISCUSSION AND MANAGEMENT RECOMMENDATIONS	54
4.1. Discussion	54
4.1.1. Metagenomics	54
4.1.2. Virus sequence analyses	55
4.1.3. Histology	57
4.1.4. Transcriptomics	58
4.1.5. Immunological Analysis of samples from project C21169	60
4.1.6. Analysis of TEM images from coral sampled under projects C1E0A5 and C21169	60
4.2. Future Steps	60
4.3. Management Recommendations	61
5. REFERENCES CITED	62

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

2.1.1. Taxonomic identification and diversity measures of additional bacterial and archaeal taxa

Analyses in this funding year focused on the characterization of additional micro-eukaryotes and bacteria from previously sequenced samples and synergistic analyses between metagenomic and metatranscriptomic data. PhyloFlash 3.4 (Gruber-Vodicka et al., 2020) was used to map SSU rRNA sequences from metagenomic data against the SILVA SSU database. Alpha and beta diversity analyses were performed on count tables and taxa assignments generated by phyloFlash. Data were aggregated to the lowest possible taxonomic assignments and were normalized by read count across samples (rarefied). Differences observed in Shannon–Wiener diversity index between disease classes and species were tested with the Kruskal–Wallis rank sum test and the Pairwise Wilcoxon Rank Sum Test with FDR multiple test correction was used for pairwise analyses. Due to the presence of numerous high diversity samples, centered log-ratio (CLR) transformation performed to capture ratio relationships between taxa using the tool *clr* from the microbiome package. Beta diversity analyses were performed on the unrarefied raw counts table produced by phyloFlash. Principal components analysis (PCA) was performed using Euclidean distances calculated from the CLR-transformed dataset (producing Aitchison distance) with the *phyloseq* command ordinate (as RDA without constraints). Permutational Multivariate Analysis of Variance (PERMANOVA) was performed on Euclidean

distances to test for differences in beta diversity of bacterial community compositions among groups. PERMANOVA was performed using the function `adonis` from the package `vegan` (v2.5.579; Oksanen et al., 2025) and was followed by pairwise analysis of variance with `pairwiseAdonis` (v0.0180; Martinez Arbizu, 2020) using Euclidean distance and 999 permutations.

To identify microbial taxa predictive of coral disease states, we conducted a Random Forest classification analysis using taxonomic and abundance output from `phyloFlash`. Data were aggregated to the genus level unless taxonomic annotation was not available, in which case a higher taxonomic level was used. Samples were grouped into four disease classes: Disease Lesion, Diseased Unaffected (healthy tissue from diseased colonies), Apparently Healthy, and Naïve. The Random Forest model was trained to classify samples into these categories based on microbial community composition. We used the `randomForest` package in R, setting the number of trees (`ntree`) to 1000 and allowing the model to determine the optimal number of variables tried at each split (`mtry`). Out-of-bag (OOB) error rates were calculated to estimate model accuracy, both overall and class-specific. Feature importance was assessed using two metrics: Mean Decrease in Accuracy (MDA), which reflects the loss in classification accuracy when a feature is removed, and Mean Decrease in Gini (MDG), which measures how much a feature contributes to reducing impurity at decision nodes. To visualize important predictors, we plotted taxa with the highest MDA and MDG values, both across all disease classes and within individual class comparisons. Model reproducibility was monitored by examining variability in results across multiple model runs. All data preprocessing and visualization were performed in R using `ggplot2` and related tidyverse packages.

2.1.2. Assessment of microbial functional activity

Characterization of microbial functional activity 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-Míguez 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`.

2.1.3. Integrative analyses of metagenomic data with transcriptomic data

To explore relationships between host functional activity and microbial metabolic potential, we began with normalized read counts of Single Copy Orthologs (SCOs, as identified by OrthoFinder, see methods in section 2.4 below) per sample and relative abundances of microbial metabolic pathways (as identified by HUMAnN3) per sample. 7,462 SCOs and 468 microbial pathways conserved across all species and sample types were identified. To investigate the functional relationships between host Single Copy Orthologs (SCOs) and microbial metabolic pathways, correlation analysis and network analysis were performed separately on each disease class. Apparently Healthy and Diseased Unaffected samples were grouped together based on their similarity in other analyses. Before splitting data by class, SCOs that tended to occur together ($r > 0.9$) and pathways that occurred together ($r > 0.9$) were clustered to reduce dimensionality using a Spearman correlation analysis on normalized SCO counts and relative abundances of microbial pathways. Hierarchical clustering (average linkage) was then applied to the absolute correlation distance matrix ($1 - r$) to group highly positively co-occurring terms into clusters, using a fixed height threshold ($h = 0.1$, corresponding to $r > 0.9$) to define clusters. A second Spearman correlation analysis was then performed on the clusters of SCOs/pathways on data divided by disease class, including correlations between pathways and SCOs that were both positive (when abundance of X pathway increased, abundance of Y SCO increased) and negative (when abundance of X pathway

increased, abundance of Y SCO decreased, or vice versa). Resulting p-values were adjusted for multiple testing using the False Discovery Rate (FDR) method to control for type I error due to the large number of comparisons. Correlations exceeding the threshold $|r| > 0.9$ were visualized as a bipartite network using the *igraph* and *ggraph* packages in R. The network layout was generated using the Fruchterman-Reingold ("fr") force-directed algorithm, which simulates attractive forces between connected nodes and repulsive forces between all nodes to produce an interpretable spatial arrangement. This visualization approach emphasizes network topology, highlighting hubs (nodes with high degree) and bottlenecks (nodes with high betweenness centrality) that may play key roles in mediating functional interactions within the disease-associated microbiome.

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 prior to full pipeline processing using the program `bbnorm.sh` from the BBMap toolkit (Bushnell). The assembly, detection, and taxonomic classification of viral sequences are inherently challenging, particularly in non-model systems such as coral reefs and coral tissues, where the diversity of viruses across most coral species and regions remains largely unknown. In phase II, we built on phase I efforts to address these challenges by leveraging multiple programs at each pipeline stage (i.e., assembly, detection, classification) and by developing new bioinformatic tools to support future coral reef virus studies in Florida and beyond.

2.2.1. Processing of sequencing data for virus sequence detection

Accurate and meaningful results that allow us to understand the roles of viruses in SCTLD rely on high-quality sequence assembly. Starting with RNA sequencing data, the objective was to generate high-quality RNA/DNA viral transcript sequences and RNA virus genomes. To achieve this, we applied several sequence assembly algorithms, leveraging their respective strengths to produce multiple assemblies per library, thereby increasing the likelihood and quality of virus sequence assembly and subsequent recovery and classification. All normalized individual sequencing libraries were first assembled with three different SPAdes (v3.15.5; Nurk et al., 2017) algorithms which included metaSPAdes (`--meta`), rnaSPAdes (`--rna`), and rnaviralSPAdes (`--rnaviral`). An additional assembly using Trinity (v2.15.2; Grabherr et al., 2011) was generated for each species by assembling pooled read libraries per species. Each assembly produced was then screened for virus-like sequences using the program deep6 (Finke et al., 2023). All virus-like sequences identified by deep6 were then pooled into a single fasta file. From the file containing all deep6 predicted virus sequences, a non-redundant, error-free first pass reference set of putative virus sequences was generated using the program EvidentialGenes (sourceforge.net/projects/evidentialgene/).

Virus sequences can be difficult to assemble due to their high variability and repetitive regions. To address this, we performed an additional "mega-assembly" by assembling all virus-like reads together, as this has been shown to improve sequence contiguity (Hofmeyr et al., 2020; Vosloo et al., 2021). Cleaned reads were mapped to the initial reference set of virus-like sequences using Bowtie2 (v2.5.4; Langmead & Salzberg, 2012). All reads that aligned to the reference sequences were then extracted from their source libraries and pooled into a single set of forward and reverse reads. The pooled reads were then re-assembled with SPAdes using the metaviralSPAdes, rnaSPAdes, and rnaviralSPAdes algorithms, as well as with Trinity. This new assembly was again screened for virus-like sequences using the program deep6. All deep6 predicted virus sequences were then pooled with the first pass reference set of sequences. Finally, EvidentialGenes was then used to produce a non-redundant, error-free final reference set of putative virus sequences recovered from all the RNA sequencing data produced.

For recovery of virus sequences from metagenomic data, first all normalized metagenomic libraries (n = 133) were assembled using SPAdes (v3.15.5; Nurk et al., 2017) with the "`--meta`" option. In

addition, reads not assigned to a sample during demultiplexing (~100–120 Gbp of non-normalized data) were also normalized and then assembled using MEGAHIT (v1.2.9 due to the higher computational requirements of SPAdes. All assembled contigs/scaffolds across all libraries were then combined into a single fasta file and deduplicated using Seqkit (v2.10.0; rmup -s; Shen et al., 2024) to remove exact sequence redundancies. To remove cellular sequences from the non-redundant sequence set, sequences were aligned using minimap2 (v2.29; -x asm5; Li, 2021) to a custom database containing all publicly available coral (n = 142) and dinoflagellate (n = 32) genomes from NCBI's GenBank, downloaded using the command-line tool ncbi-genome-download (v0.3.3; <https://github.com/kblin/ncbi-genome-download>). All sequences that did not align to coral or dinoflagellate sequences were then extracted and an additional round of cellular sequence screening (this time focused on prokaryotic and other micro-eukaryotic groups) was performed with the program CAT (v6.0.1; Hauptfeld et al., 2024) using NCBI NR database. All sequences not classified as cellular were extracted and collapsed by 100% nucleotide identity using CD-HIT-EST v4.8.1 (Fu et al., 2012) with the parameters -c 1.0 and -aS 1.0, thereby removing exact duplicates regardless of length. The deduplicated putative virus-like sequences from the metagenome assemblies were then used as input for binning with vRhyme (v1.1.0; Kieft et al., 2022) to construct virus metagenome-assembled genomes (vMAGs). Sequences not included in bins were extracted from the vRhyme input FASTA file and combined with the final vMAGs, producing a final FASTA file containing all likely viral sequences from DNA sequencing libraries, which was then processed through the classification pipeline.

2.2.2. *Virus sequence classification*

Given the unexplored nature of virus diversity associated with coral reefs, a multi-tool analysis was conducted to classify putative virus sequences. Putative virus sequences from both the metagenome and metatranscriptomes processing were analyzed using the following programs: i. geNomad (v1.11.0; Camargo et al., 2024), ii. Cenote-Taker 3 (v3.4.0; <https://github.com/mtisza1/Cenote-Taker3>); iii. CAT (v6.0.1; Hauptfeld et al., 2024); iv. VirSorter2 (v2.2.4; Guo et al., 2021); v. CheckV (v1.0.1; Nayfach et al., 2021); vi. VITAP (v1.7.1; Zheng et al., 2025); and viCAT (unpublished; developed by Alex Veglia, 2025, currently in preparation for release). Results from each of the seven programs were then compared and combined using the program viSUM (unpublished; developed by Alex Veglia, 2025, currently in preparation for release). The program viSUM aggregates the outputs of multiple viral classification tools, compares their assignments, and generates a unified consensus for each sequence with associated support scores. All downstream analyses were performed on sequences that viSUM identified as having evidence of being viral based on one or more of the classification/detection programs used.

2.2.3. *Virus sequence quantification and functional annotation*

Read count matrices were generated for DNA (metagenome) and RNA (metatranscriptomes) virus sequence sets by aligning cleaned, normalized RNA sequencing reads using the program kallisto (v0.50.1; Bray et al., 2016). Read count matrices were then processed with RSEM (v1.3.3; Li & Dewey, 2011) to produce trimmed mean of M-values (TMM) normalized counts table to allow for cross sample comparisons. Functional annotation of virus genes was done using the program MetaCerberus (v1.4.0; Figueroa III et al., 2024) with the following databases: VOG (<https://vogdb.org/>; Oct 2024), pVOG (2016; Grazziotin et al., 2017), PHROG (Jun 2022; Terzian et al., 2021), AMRFinder-fams (Feb 2024; Feldgarden et al., 2021), NFixDB (Jan 2024; Bellanger et al., 2024), and GVDB (2021; Aylward et al., 2021)

2.2.4. *Virus community & gene landscape analyses*

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 (analyses described are focused on RNA sequencing results files) using the *phyloseq* R package (McMurdie &

Holmes, 2013). To identify core viromes (at the class tank level) for each coral species, we first subset the TMM-normalized phyloseq object by host species. For each subset, counts for virus transcripts with confidence scores above 0.28 (indicating minimum three programs identified the sequence as viral) were aggregated at the viral class level using the `rowsum()` function. We then calculated the prevalence of each class as the proportion of samples in which it was present. Core classes were defined as those detected in $\geq 95\%$ of samples for a given species. The union of all species-specific core classes was used to identify corresponding viral transcripts across the full dataset. These transcripts and their counts data were extracted to generate a new phyloseq object representing the combined core virome. Bray–Curtis dissimilarity was calculated using the new phyloseq object with core virome transcripts with the `vegdist()` function from the `vegan` package. Species were hierarchically clustered based on this distance matrix using average linkage (UPGMA), and the resulting dendrogram was visualized to assess inter-species similarity in core viral transcript profiles.

Virus community alpha diversity was calculated per sample with the raw counts data using the Shannon index (H') as implemented in *phyloseq* (i.e. via `estimate_richness(phyobj, measures="Shannon")`, which calls `vegan::diversity` with `index="shannon"`). We then tested for overall differences in Shannon diversity across the three sampling time points (T0, T1, T2) using a Kruskal–Wallis rank-sum test (`kruskal.test()` in R's *stats* package), because residuals from a linear model of Shannon failed normality (Shapiro–Wilk via `shapiro.test()`, also in *stats*). For any significant Kruskal–Wallis result, we performed pairwise Wilcoxon rank-sum tests (`wilcox.test()` in *stats*) with Benjamini–Hochberg adjustment (`p.adjust(method="BH")`) to identify which time-point pairs differed. The same procedure (Kruskal–Wallis followed by pairwise Wilcoxon) was applied separately within each coral species to assess time-point effects on α -diversity. Finally, to evaluate whether health status (healthy vs diseased vs quiesced) influenced Shannon diversity, we ran a Kruskal–Wallis test on Shannon by health category (again with Wilcoxon post hoc as above) across all species.

Beta diversity was calculated on TMM-normalized counts by first transforming each library to relative transcript abundances. Bray–Curtis dissimilarities were then computed using the `distance()` function in *phyloseq* with `method = "bray"`. We confirmed that multivariate dispersion did not differ by time point, sample site, or health status (all `betadisper()` + `permutest()` $p \gg 0.05$), indicating no significant heterogeneity of spread. We then ran a PERMANOVA using `adonis2(bc_dist ~ host_species + time_point + health_status, data = meta_df, permutations = 999, by = "margin")`. In this model, “host_species” captures differences among coral species, “time_point” tests for changes across the outbreak timeline, and “health_status” compares healthy versus diseased/quiesced samples. To examine within-species effects, we subset the data by coral species and reran `adonis2(bc_sub ~ time_point + health_status, data = meta_sub, permutations = 999, by = "margin")` on the corresponding Bray–Curtis distance matrix. This two-factor, within-species PERMANOVA evaluated whether viral β -diversity differed by time point or health status within each coral species independently.

Differential abundance of viral transcripts was assessed with DESeq2 (Love et al., 2014) using a negative-binomial GLM with the formula `~ host_species + tissue_type`, where `host_species` (cavernosa, faveolata, franksi, natans) was included as a covariate to account for baseline differences in each coral's virome, and `tissue_type` (healthy as the reference) was used to estimate the log₂-fold-change in disease_margin versus healthy samples across all species. Similar to core virome analyses, only virus sequences with confidence scores above 0.28 (indicating minimum three programs identified the sequence as viral) were used for differential expression analyses. Differential abundance of these upregulated 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. Gene information for differentially expressed (upregulated) transcripts were then extracted from functional annotation results and grouped by potential function.

2.3. Histological Analyses

Samples were fixed in zinc-buffered formalin (Z-fix, Anatech), then seawater for 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 μ m 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 review*). 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 mm² images per tissue sample. Each micrograph was split into twelve 5000 mm² 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.

2.4. Transcriptomic Analyses

The analyses conducted in phase II are directly complementary to data produced during phase I of this project. Herein we performed the following tasks: 1) refined *de novo* transcriptomes for each of our sampled species and generated improved read count matrices, 2) identified single copy orthologs

consistent across all species for identification of species-independent markers of resistance, 3) conducted differential expression modeling to identify species-independent markers of resistance, and 4) conducted integrative, correlative analyses to identify species-independent associations between ortholog expression and microbiota, viral community members, and histological traits of interest. Full details on sample collection and initial processing can be found in the report for year 1 (Klinges et al., 2024).

2.4.1. De novo transcriptome refinement

De novo transcriptomes were generated from five random samples for each coral species using a modification of previously established protocols (Beavers et al., 2023). Before assembly, two pre-filtering steps were implemented to remove as many non-coral reads as possible. First, any reads aligning to potential eukaryotic contamination were filtered out from the available viral, human, fungi, and protozoa genomic databases from kraken2 (Wood et al., 2019). Next the reads were filtered using BBSplit (Bushnell) against all publicly available Symbiodiniaceae transcriptomes that could be located ($n = 9$, **Table 1**).

Table 1. List of transcriptomes used for BBSplit pre-filtering step in transcriptome assembly and read filtering. Asterisks denote those which were also used for BBSplit filtering/host and symbiont separation of reads.

Species	Associated Publication	Download Link
<i>Symbiodinium microadriaticum</i> *	(Aranda et al., 2016)	http://smic.reefgenomics.org/download/Smic.transcriptomeRef.cov5.fa.gz
<i>Symbiodinium</i> spp., "Y106"	(Shoguchi et al., 2018)	https://marinegenomics.oist.jp/symb/download/syma_transcriptome_37.fasta.gz
<i>Breviolum aenigmaticum</i>	(Parkinson et al., 2016)	http://zoox.reefgenomics.org/download/Symbiodinium_aenigmaticum.tar.gz
<i>Breviolum minutum</i> *	(Parkinson et al., 2016)	http://zoox.reefgenomics.org/download/Symbiodinium_minutum.tar.gz
<i>Breviolum pseudominutum</i>	(Parkinson et al., 2016)	http://zoox.reefgenomics.org/download/Symbiodinium_pseudominutum.tar.gz
<i>Breviolum psygmophilum</i>	(Parkinson et al., 2016)	http://zoox.reefgenomics.org/download/Symbiodinium_psygmophilum.tar.gz
<i>Cladocopium proliferum</i> (formerly thought to be <i>C. goreau</i>)	(Davies et al., 2016)	http://ssid.reefgenomics.org/
<i>Cladocopium</i> spp., "Y103"*	(Shoguchi et al., 2018)	https://marinegenomics.oist.jp/symb/download/symC_transcriptome_40.fasta.gz
<i>Durusdinium trenchii</i> *	(Bellantuono et al., 2019)	https://datadryad.org/stash/downloads/file_stream/258199

“Raw” transcriptome assemblies were then generated using these filtered reads with Trinity v2.9.1 (Grabherr et al., 2011). The resulting transcriptome was then filtered to retain only the longest isoform per gene. Next, the program TransDecoder was used to identify transcripts with open reading frames and transcripts without an identified coding region were removed (Haas). The program CD-HIT was used to remove redundant transcripts (Fu et al., 2012). As a final filtration step to ensure as much symbiont contamination as possible was removed, the resultant transcriptome was blasted against a custom database assembled from all publicly available Symbiodiniaceae proteomes ($n = 17$, **Table 2**)

following established protocols (Stankiewicz et al., 2025). Resultant transcriptome completeness was assessed using BUSCO (Manni et al., 2021). Finally, the transcriptome was annotated via comparison to the UniProt database using blast v2.15.0 (Camacho et al., 2009).

Table 2. List of proteomes used for blast filtering step in transcriptome assembly.

Species	Associated Publication	Download Link
<i>Symbiodinium fitti</i>	(Reich et al., 2021)	https://datadryad.org/stash/dataset/doi:10.5061/dryad.xgxd254g8
<i>Symbiodinium linucheae</i>	(Gonzalez-Pech et al., 2021)	https://espace.library.uq.edu.au/view/UQ:f1b3a11
<i>Symbiodinium microadriaticum</i>	(Gonzalez-Pech et al., 2021)	https://espace.library.uq.edu.au/view/UQ:f1b3a11
<i>Symbiodinium natans</i>	(Gonzalez-Pech et al., 2021)	https://espace.library.uq.edu.au/view/UQ:f1b3a11
<i>Symbiodinium necroappatans</i>	(Gonzalez-Pech et al., 2021)	https://espace.library.uq.edu.au/view/UQ:f1b3a11
<i>Symbiodinium tridacnidorum</i>	(Gonzalez-Pech et al., 2021)	https://espace.library.uq.edu.au/view/UQ:f1b3a11
<i>Symbiodinium</i> spp. "Y106"	(Shoguchi et al., 2018)	http://sampgr.org.cn/downloads/syma_aug_37.aa.longest.fa.tar.gz
<i>Breviolum minutum</i>	(Shoguchi et al., 2013)	https://marinegenomics.oist.jp/symb/download/symbB.v1.2.augustus.prot.fa.gz
<i>Cladocopium goreau</i>	(Chen et al., 2020)	https://espace.library.uq.edu.au/view/UQ:8279c9a
<i>Cladocopium goreau</i>	(Chen et al., 2022)	https://espace.library.uq.edu.au/view/UQ:fba3259
<i>Cladocopium</i> spp., "C92"	(Chen et al., 2020)	https://espace.library.uq.edu.au/view/UQ:8279c9a
<i>Cladocopium</i> spp., "C15"	(Messer et al., 2024)	http://plutreefgenomics.org/cladocopium/download/
<i>Cladocopium</i> spp., "Y103"	(Shoguchi et al., 2018)	http://sampgr.org.cn/downloads/symC_aug_40.aa.longest.fa.tar.gz
<i>Durusdinium trenchii</i>	(Shoguchi et al., 2021)	https://marinegenomics.oist.jp/symbd/viewer/download?project_id=102
<i>Durusdinium trenchii</i> "CCMP2556"	(Dougan et al., 2024)	https://espace.library.uq.edu.au/view/UQ:27da3e7/Dtrenchii_CCMP2556_PROT_fasta.gz?dsi_version=be68147c743657b4e92ef4d7eb8012ce
<i>Durusdinium trenchii</i> "SCF082"	(Dougan et al., 2024)	https://espace.library.uq.edu.au/view/UQ:27da3e7/Dtrenchii_SCF082_PROT_fasta.gz?dsi_version=6d08446e7b3c2b35ddb5a3edce528200

Prior to alignment, raw read files were first processed using the program fastp to remove adapters and filter out poor-quality reads (Chen, 2023; 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. Next, these quality-filtered reads were run through BBsplit (Bushnell) to separate host and Symbiodiniaceae reads. Specifically, reads were filtered against

a single transcriptome for each of the four dominant genera of Symbiodiniaceae known to commonly associate with Caribbean stony corals (**Table 1**, denoted by asterisk), and their respective generated *de novo* coral host transcriptome that had previously been assembled. The resulting reads which matched to the host transcriptome were then fully aligned to that same *de novo* transcriptome using Salmon (Patro et al., 2017).

2.4.2. Ortholog Analyses- Identification of species specific markers of SCTLD resistance

Leveraging our polished and high quality *de novo* transcriptomes, we then sought to identify single copy orthologs for downstream multi-species analyses. Single copy orthologs (SCOs) refer to single copy transcripts which are conserved across a group of species, identified based on transcript similarity. The combined sequence similarity and lack of copy number variation across species makes SCOs an ideal subset for comparing gene expression datasets across species. We identified single copy orthologs following well established pipelines. Specifically we first generated predicted proteomes for each transcriptome using Transdecoder (Haas) and collapsed similar sequences with CD-HIT (Fu et al., 2012). We then analyzed the resultant proteomes with Orthofinder (v. 2.5.2; Emms & Kelly, 2019) using default parameters. The resultant data was used to create a transcript to gene file for importing reads corresponding only to SCOs into R for downstream analysis.

To identify species-independent markers of SCTLD resistance we combined our set of identified single copy orthologs (SCO; present in all species) with previously generated read count matrices, isolating expression of SCOs only. We then focused on two main types of markers: 1) SCOs whose expression prior to disease onset predicted eventual disease susceptibility (predictive models) and 2) SCOs which were differentially expressed between healthy tissue from resistant colonies and apparently healthy tissue from susceptible colonies at either disease onset timepoint (June or August; epidemic models). We also identified those genes which were significantly differentially expressed between apparently healthy and diseased tissue within susceptible colonies to gain insight regarding the mechanisms of response to SCTLD (response models). In each case we used the R package dream (within the variancePartition package) to fit linear mixed models to our data, allowing for the addition of random effects of colony ID and site where needed. Dream is built upon existing R packages limma and voom, but adds necessary steps to fit random effects (Hoffman & Roussos, 2021). In each grouping of models we ran two sets of models, one which incorporated and interaction with species to identify species-specific responses, and a second model without the species interaction term to identify species independent responses (main effects cannot be interpreted when included in an interaction term, necessitating this approach). **Table 3** details models used. Additionally, we used the makeContrasts() function for hypothesis testing contrasts of interest. To facilitate hypothesis testing and contrast specification, we set the intercept to 0 in all models and, where necessary, collapsed factors into one term. For example, in the epidemic & response models, sampling time, tissue type, and species were collapsed into one group and contrasts specified to identify SCOs with species specific differences between tissues of interest within a given month (“interaction” model). The main effect model for this category used a combination of timepoint and tissue type to investigate species independent differences between tissue types of interests within a given month. Contrasts run are listed in **Table 4**.

Table 3: Summary of Model Groups and models used to identify species independent markers of resistance. In each case models with interaction terms were followed up with main effect models to confidently identify species independent markers. category=disease susceptible of resistant; timepoint_tis_spec = combination of timepoint collected, tissue type, and species; timepoint_tis_spec = combination of timepoint collected and tissue type

Model Group	Type	Model	Factors of Interest
Predictive Models	Interaction	~0+category*species+(1 site)	category*species
	Main Effect	~0+category+species+(1 site)	category
Epidemic & Response Models	Interaction	~0+timepoint_tis_spec+(1 site)+(1 ID)	month_tis_spec
	Main Effect	~0+timepoint_tis +species+(1 site)+(1 ID)	month_tis_type

Once our models were run, we then compared results of the interaction and main effect models within a group to identify candidate SCOs marking resistance. Specifically, an SCO was classified as a species independent marker if it was significant for the main effect factor of interest (category or month-tissue type comparison) and not significant for any factor of interest in our interaction model. SCOs which were significant for contrasts in our interaction models were deemed species dependent markers and flagged for later analysis. In the case of Epidemic and Response Models, an additional filtering step was used; our list of best marker SCOs only included those SCOs which were identified as species independent markers in either June and August (i.e. consistently differentially expressed between healthy and apparently healthy or apparently healthy and disease margin tissues) with congruent log-fold changes in both months. Candidate markers with immune functions were identified by search associated annotations, specifically assigned gene ontology terms, for a breadth of immune keywords. Representative graphs were constructed in R using ggplot to demonstrate patterns of interest.

Table 4: Summary of contrasts specified within each model group/type and the hypotheses tested. category=disease susceptible of resistant; timepoint_tis_spec = combination of timepoint collected, tissue type, and species; timepoint_tis_spec = combination of timepoint collected and tissue type.

Model Group	Type	Contrast	Hypothesis Tested
Predictive Models	Interaction	RES-SUS X MCAV-CNAT RES-SUS X MCAV-OFVAV RES-SUS X MCAV-OFRA RES-SUS X OFAV-CNAT RES-SUS X OFAV-OFRA RES-SUS X OFRA-CNAT	species specific differences between resistant and susceptible colonies prior to disease arrival
	Main Effect	RES-SUS	species independent differences between resistant and susceptible colonies prior to

			disease arrival
Epidemic Model	Interaction	June_AH_MCAV-June_H_MCAV vs. June_AH_CNAT-June_H_CNAT June_AH_MCAV-June_H_MCAV vs. June_AH_OFAV-June_H_OFAV June_AH_MCAV-June_H_MCAV vs. June_AH_OFRA-June_H_OFRA June_AH_OFAV-June_H_OFAV vs. June_AH_CNAT-June_H_CNAT June_AH_OFAV-June_H_OFAV vs. June_AH_OFRA-June_H_OFRA June_AH_OFRA-June_H_OFRA vs. June_AH_CNAT-June_H_CNAT Aug_AH_MCAV-Aug_H_MCAV vs. Aug_AH_CNAT-Aug_H_CNAT Aug_AH_MCAV-Aug_H_MCAV vs. Aug_AH_OFAV-Aug_H_OFAV Aug_AH_MCAV-Aug_H_MCAV vs. Aug_AH_OFRA-Aug_H_OFRA Aug_AH_OFAV-Aug_H_OFAV vs. Aug_AH_CNAT-Aug_H_CNAT Aug_AH_OFAV-Aug_H_OFAV vs. Aug_AH_OFRA-Aug_H_OFRA Aug_AH_OFRA-Aug_H_OFRA vs. Aug_AH_CNAT-Aug_H_CNAT	species specific differences between healthy tissue on resistant and susceptible colonies during disease outbreak in each time point (June or August)
	Main Effect	June_AH-June_H Aug_AH-Aug_H	species independent differences between healthy tissue on resistant and susceptible colonies during disease outbreak in each time point (June or August)

Response Models	Interaction	June_AH_MCAV-June_DM_MCAV vs. June_AH_CNAT-June_DM_CNAT June_AH_MCAV-June_DM_MCAV vs. June_AH_OFAV-June_DM_OFAV June_AH_MCAV-June_DM_MCAV vs. June_AH_OFRA-June_DM_OFRA June_AH_OFAV-June_DM_OFAV vs. June_AH_CNAT-June_DM_CNAT June_AH_OFAV-June_DM_OFAV vs. June_AH_OFRA-June_DM_OFRA June_AH_OFRA-June_DM_OFRA vs. June_AH_CNAT-June_DM_CNAT Aug_AH_MCAV-Aug_DM_MCAV vs. Aug_AH_CNAT-Aug_DM_CNAT Aug_AH_MCAV-Aug_DM_MCAV vs. Aug_AH_OFAV-Aug_DM_OFAV Aug_AH_MCAV-Aug_DM_MCAV vs. Aug_AH_OFRA-Aug_DM_OFRA Aug_AH_OFAV-Aug_DM_OFAV vs. Aug_AH_CNAT-Aug_DM_CNAT Aug_AH_OFAV-Aug_DM_OFAV vs. Aug_AH_OFRA-Aug_DM_OFRA Aug_AH_OFRA-Aug_DM_OFRA vs. Aug_AH_CNAT-Aug_DM_CNAT	species specific differences between apparently healthy and diseased tissue on susceptible colonies in each time point (June or August)
	Main Effect	June_AH-June_DM Aug_AH-Aug_DM	species dependent differences between apparently healthy and diseased tissue on susceptible colonies in each time point (June or August)

2.4.3. Integrative Analyses: Linking Transcriptomic, Microbial, Viral, and Histological Data

Next, we shifted to focus on identification of SCOs which were highly correlated to microbial/viral community composition or histological traits of interest. This was conducted via integrative correlative analyses following established pipelines adapted from other systems (Fuess et al., 2021). To identify microbial taxa associated with SCOs, we first filtered generated 16S data to retain only those ASVs

which accounted for a minimum average of .1% abundance across samples, yielding 120 ASVs. Generated proportions of each of these taxa were then correlated to normalized read counts for our 7,462 SCOs using a pairwise kendall correlation. The top 5% of associations based on *p-value* were retained for downstream analyses, specifically gene ontology enrichment analysis with GOMWU (Wright et al., 2017).

A similar approach was used for viral community analyses. First, we filtered generated sequences to retain only those with the maximum rank support score at the domain level to ensure retained sequences corresponded to putative viruses. A total of 145,783 out of a starting 621,896 sequences fit this filtering qualification. We then collapsed sequences to the Order level, or highest level of taxonomy if not assigned to an order, by combining TPM normalized read counts for all IDs within the same Taxonomic group. Any sequences which were unclassified at the order level were removed at this step (106,939 out of 146,783 sequences retained). It was necessary to collapse reads to the Order level as all sequences passing confidence filters were unclassified at the Family, Genus, and Species levels. In the end, we generated TPM normalized read counts for a total of 235 viral groups. This data was then correlated to normalized read counts for our 7,462 SCOs using a pairwise kendall correlation. The top 5% of associations based on *p-value* were retained for downstream analyses, specifically gene ontology enrichment analysis with GOMWU (Wright et al., 2017).

Finally, to correlate histological features of interest with transcriptomic data, we used a Network-based approach. A single co-expression network was generated from our normalized ortholog data using Weighted Gene Correlation Network Analysis (Langfelder & Horvath, 2008). We then correlated expression of resultant groups of orthologs to our top microbial families and mean histological traits of interest (vacuolization, max vacuole size, and exocytosis). Modules of interest were further analyzed using gene ontology enrichment analysis with GOMWU (Wright et al., 2017).

2.5. Immunological Analysis of samples from project C21169

Fragments of corals used in experiments described in FDEP project C21169 (Lead PI Ushijima) were received by the team at Texas State University (PI Fuess) for processing and analysis for immunological activity. Full details of experimental methods can be found in the relevant project reports for C21169. Herein we describe processing and analysis of these samples for immunological activity only. Assays are modified versions of those established for tropical corals (Changsut et al., 2022; Fuess et al., 2016; Mydlarz & Palmer, 2011).

Tissue was removed from frozen coral fragments using a Paasche airbrush and 100 mM Tris + 0.05mM DTT (pH 7.8) buffer. This tissue slurry was then be homogenized for 1 min. and placed on ice for 7 – 10 min. For melanin analyses, a 1 mL aliquot of tissue slurry was taken at this step and placed into a pre-weighed 1.5 mL tube and flash frozen, then stored at -20° C until analysis. For all other protein analyses, remaining tissue slurry was centrifuged at 3500 rpm for 5 min. and two aliquots of the supernatant (aka protein extract) will be transferred into 2 mL tubes and flash frozen, then stored at -80 °C until analysis.

All of the following assays were conducted in triplicate, including negative controls using the Tris+DTT buffer used to homogenize, and were measured on the BioTek Cytation 1 imaging reader, unless specified otherwise. Protein concentration was measured for each sample before conducting any assays using a Red660 assay. Protein extract (10 uL) was combined with 150 uL of G-Biosciences Red660. Sample absorbance was read at 660 nm and compared to a bovine serum albumin (BSA) standard curve to determine sample protein concentration. Measurements from subsequent assays were all standardized by protein concentrations unless mentioned otherwise.

To measure phenoloxidase (PO) activity, 20 uL of protein extract, 45 uL of 50 mM PBS pH 7.0, and 30 uL of 10 mM dopamine were combined and the changes in absorbance measured every 45 seconds for 15 min. at 490 nm. PO activity was then calculated using the most linear part of the curve in the first 1-5 minutes of the reaction. To measure total phenoloxidase activity, total phenoloxidase (TPO) activity was measured. TPO was measured by combining 20 uL of protein extract, 20 uL of 50 mM PBS pH 7.0, 25 uL of 0.1 mg/mL trypsin, and allowing this mixture to incubate at room temperature for 30 minutes. Then 30

uL of 10 mM dopamine was added, and changes in absorbance were immediately measured every 45 seconds for 15 min. at 490 nm. TPO activity was then calculated using the most linear part of the curve in the first 1-5 minutes of the reaction (Mydlarz & Palmer, 2011).

Melanin concentrations were measured using the 1.0 mL of protein slurry aliquoted during airbrushing into a pre-weighed tube. These samples were first dried in a vacufuge for 12+ hours until completely dehydrated. The tubes containing the dried samples were then weighed, and 250 uL of 1 mm glass beads and 400 uL of 10M sodium hydroxide were added. A 2 mg/mL melanin standard stock was made using melanin and 10M sodium hydroxide. This standard and the prepared samples were then incubated for 48 hours in the dark for digestion of tissues. Finally, absorbance at 410 nm of a serial dilution using the melanin standard and 40 uL of each sample was measured. The measurements of each sample were standardized using the curve generated from the serial dilution. Melanin is reported standardized by sample dry tissue weight (Mydlarz & Palmer, 2011).

To measure catalase activity, 2.5 uL of protein extract was combined with 47.5 uL of 50 mM phosphate-buffered saline pH 7.0 (PBS) and 75 uL of 25 mM H₂O₂. Negative controls of Tris+DTT and a set of serial dilution wells using the PBS and H₂O₂ were included as well. Changes in absorbance (catalase activity) were then measured every 45 seconds for 15 minutes at the 240 nm wavelength. Catalase activity was then calculated using the most linear part of the curve in the first 1-5 minutes of the reaction and standardized using the serial dilution curve (Mydlarz & Palmer, 2011).

To measure peroxidase activity, 10 uL of protein extract was combined with 20 uL of 10 mM phosphate-buffered saline pH 6.0 (PBS), 25 uL of 5 mM of guaiacol, and 20 uL of 20 mM H₂O₂. Negative controls of Tris+DTT. Changes in absorbance (catalase activity) were then measured every 45 seconds for 15 minutes at the 470 nm wavelength. Peroxidase activity was then calculated using the entire curve (Mydlarz & Harvell, 2007)

To measure bacterial killing ability, bacteria doubling time/hour using a *Vibrio coralliilyticus* strain isolated from an *O. faveolata* colony in Florida will be used, as this species has been identified in relation to several stony coral diseases and will therefore serve as a good estimate of antibiotic activity (Ushijima et al., 2020). Specifically, we are using a plasmid-transformed strain, Oft6-21 pBU164, which expresses a yellow-fluorescent protein (YFP). This bacteria strain was cultured in sterile glycerol artificial seawater (GASW) broth media at 27° C until it reached an RFU (relative fluorescence units) of approximately 750. Then, 60 uL of protein extract (diluted to a standard protein concentration) and 140 uL of the diluted bacterial stock will be combined, incubated at 27° C for 12 hours, measuring YFP fluorescence every ten minutes. Bacterial doubling time was then calculated for each sample from the logarithmic growth phase.

To identify signatures of changes in immunological activity we conducted a two-way repeated measures ANOVA on each immunological assay independently. We incorporated fixed factors of interest: time point and treatment, and accounted for repeated sampling at the level of host genotype. Data was checked against appropriate statistical assumptions (normality, no outliers, etc) prior to analyses and transformed when necessary. When factors of interest were identified as significant we conducted post hoc tests in the form of paired t-tests to determine significance of groups. All statistical analyses and graphing were conducted in R.

2.6. Analysis of TEM images from coral sampled under projects C1E0A5 and C21169

Transmission electron microscopy (TEM) image datasets from projects C1E0A5 and C21169 were processed using the same approaches. To minimize observer bias and assess the reproducibility of observations, multiple individuals independently analyzed the same images. Three independent pairs were established to review and analyze TEM images of coral tissue samples acquired at magnifications of 10,000× and 50,000×, saved as uncompressed TIFF files. Each pair was assigned different subsets of TEM image folders, allowing for a systematic pairing of analyses. However, to further prevent systematic bias, individuals working on the same day did not cross-check the same images. Instead, each analyst used an overlapping set of images distributed on different days, which helped to strengthen the reliability of the analyses. All images were processed using ImageJ version 1.53c. Analysis began at lower magnifications

(ranging from 45.9% to 150%). Analysts uniformly adjusted brightness and contrast settings and applied an unsharp mask with a radius of 2 pixels for edge enhancement. Subsequently, each analyst manually reviewed the images to annotate icosahedral virus-like particles (diameters of 20–200 nm with smooth perimeters) and filamentous virus-like particles (thread-like structures up to several μm in length), using the Region of Interest Manager for accuracy (Work et al., 2021).

From these standardized annotations, we derived four complementary summary tables. First, a distribution-by-morphotype table aggregated the total number of images, mean particle count per image, standard deviation, morphotype richness, and dominant morphotypes for each time point: pre-outbreak, during outbreak, and post-outbreak. Second, a site-specific occurrence table cross-tabulated morphotype combinations against baseline and outbreak reefs, reporting both the number of images in which each combination appeared and the corresponding cluster counts. Third, a morphology-by-health-status table compared each morphotype in terms of total cluster events, number of images surveyed, and mean particle abundance across diseased, apparently healthy, and baseline samples. Finally, a health-by-site table tallied total virus-like particle (VLP) counts by health status at each reef. When replication was sufficient for statistical analysis, we looked to test for influence of health on virus-like particle rate of detection (i.e., percentage of photos with VLPs detected). To do this, VLP presence proportions were arcsine-square-root transformed to control for variances. We then used a linear mixed-effects model in R (packages: *lme4* and *lmerTest*) where Health was the main effect tested, and Colony (nested within Species) was included as a random effect to account for repeated measurements within the same groups. Next, a type III ANOVA (Satterthwaite's method) was used to get p-values for the Health effect. If significant, pairwise comparisons between health states were done with Tukey adjustment using the *emmeans* package. All tests used a significance level of $\alpha = 0.05$.

3. RESULTS

3.1. Metagenomic Analyses

3.1.1. Analyses identifying bacteria and microeukaryotes from metagenomic data

Additional bacterial, archaeal, and microeukaryotic taxa were classified from metagenomic data using the program PhyloFlash. This tool produced html files summarizing taxonomic composition of identified SSU rRNA sequences as well as alignments of read data to reference databases and assemblies of full-length 16S and 18S rRNA sequences from metagenomic data. Text summaries of top taxa (Bacteria, Archaea, Eukaryota) found in each sample group per species were produced from taxon read count data. These taxa assignments were used for downstream analysis. Bacterial taxonomy was used for alpha and beta analyses below. Annotation of Eukaryota was very poor; >80% of eukaryotic reads were only annotated to Opisthokonta. Dominant archaea in samples included Woesearchaeales and Nitrosopumilales, and no clear patterns emerged in archaeal abundance based on coral species or health state.

3.1.2. Assessments of alpha and beta diversity from metagenomic data

Alpha diversity analyses were performed on count tables and taxa assignments generated by phyloFlash. Health status (disease state) had a significant impact on microbial diversity as quantified by Shannon diversity (omnibus test $p = 1.623\text{e-}06$, $\chi^2 = 29.665$) and numbers of observed bacterial taxa from shotgun DNA sequencing data ($p = 5.67\text{e-}11$, $\chi^2 = 50.699$). Corals that were naive to disease (i.e., sampled before disease arrived to the region) and samples taken from active disease lesions had the highest microbial diversity. Although bacterial richness was significantly higher (at $p < 0.05$) in both naive and lesion samples from *Colpophyllia natans* and *Orbicella faveolata* than in apparently healthy sample types (samples taken from unaffected sections of diseased corals and healthy corals sampled from sites with disease), bacterial richness trended highest in naive samples (**Figure 1A**). No significant differences in microbial diversity were observed in *Montastraea cavernosa* and *Orbicella franksi* samples by health status, though in both species within-sample diversity in naive samples appeared to be higher, though highly variable between samples (**Figure 1A and B**). Across all

diversity metrics, diversity was equally high in naive samples as in disease lesion samples, and other sample types generally had lower bacterial diversity. There were no significant differences found in Shannon diversity or species richness by coral species ($p = 0.3574$, $\chi^2 = 3.23$) or by site ($p = 0.4735$, $\chi^2 = 1.50$).

Bacterial community composition as assessed from metagenomic data was significantly influenced by health status (Disease Lesion, Diseased – Unaffected, Apparently Healthy, Naïve) ($R^2 = 0.079$, $F = 3.668$, $p = 0.001$). All health status categories were significantly different from one another ($p < 0.01$) except for diseased unaffected and apparently healthy ($R^2 = 0.017$, $F = 1.092$, $p = 0.1990$) (**Figure 2A**). Additionally, community composition significantly differed by sampling month ($R^2 = 0.055$, $F = 2.510$, $p = 0.001$) and species ($R^2 = 0.055$, $F = 2.495$, $p = 0.001$, **Figure 2B**), but F values were small for these variables, indicating that variability between and within groups was similar. Site was not found to be a significant driver of community composition ($R^2 = 0.018$, $F = 1.104$, $p = 0.14$).

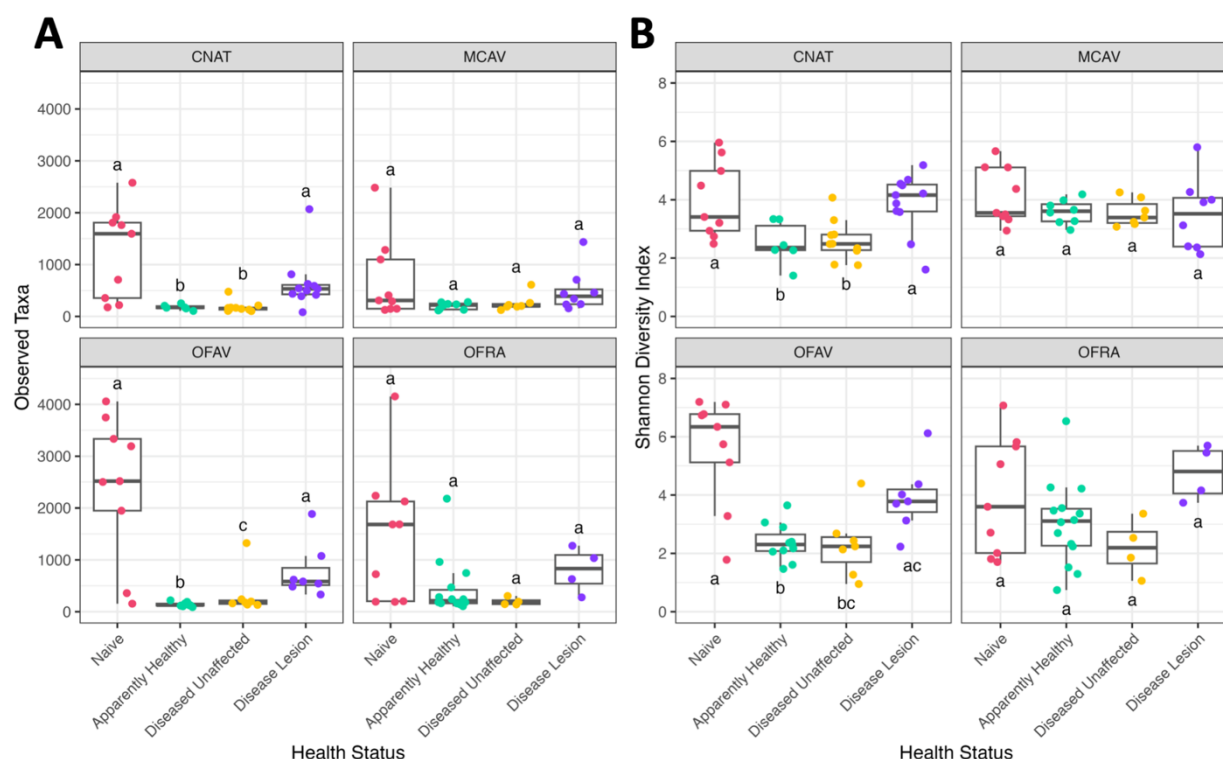


Figure 1: Differences in taxon richness (A) and Shannon-Wiener diversity (B) by health status (Naive, Apparently Healthy, Diseased – Unaffected, and Disease Lesion). Boxes sharing a letter are not significantly different from each other using an FDR corrected significance level of $p < 0.05$. Pairwise comparisons were only performed within each coral species.

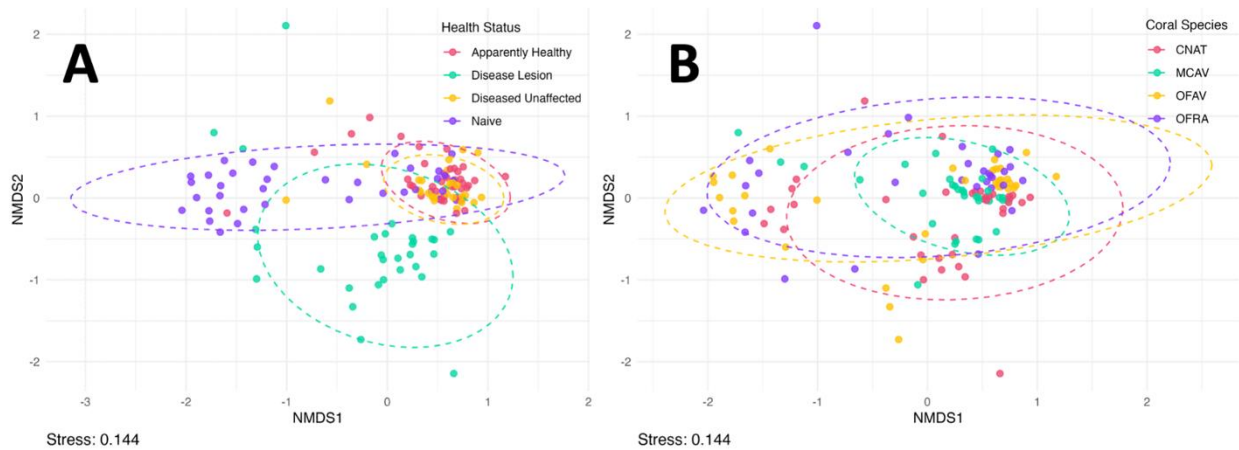


Figure 2: NMDS ordinations of bacterial taxa as identified by phyloFlash in samples using Bray–Curtis distance. A) ellipses plotted by health status, B) ellipses plotted by coral species.

3.1.3. Species-independent markers of disease status

Random forest classification was used to identify microbial taxa that best discriminate among coral disease classes (Disease Lesion, Diseased – Unaffected, Apparently Healthy, Naïve) based on taxon abundance profiles from phyloFlash aggregated to the genus level. The random forest model, trained on data produced by phyloFlash, achieved an overall out-of-bag (OOB) error rate of 36.09%, indicating moderate classification accuracy across all disease classes. For each class in the dataset, a binary classifier was trained to distinguish that class from all others (one-vs-all approach). Error rates varied by class, with the lowest error observed in Disease Lesion samples (13.33%) and the highest in Disease – Unaffected samples (85.71%). This was likely due to similarities in bacterial community composition

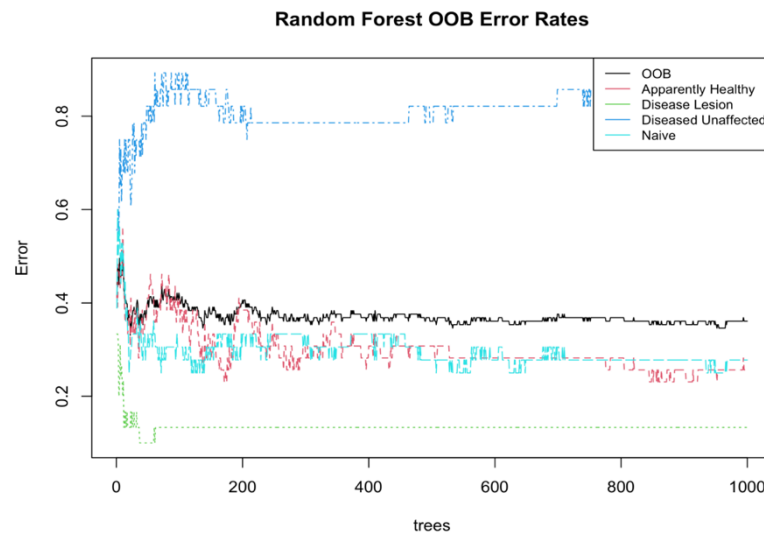


Figure 3: Out-of-bag (OOB) error rates for the Random Forest classification model predicting disease classes. The plot shows the overall OOB error rate (black line) as well as the class-specific OOB error rates for each disease class, represented by colored lines. These error rates indicate the proportion of samples misclassified during model training, providing an internal estimate of model accuracy without the need for a separate test set.

observed in Disease Unaffected and Apparently Healthy samples (see alpha and beta diversity analyses).

Variable importance was assessed using both Mean Decrease in Accuracy (MDA) and Mean Decrease in Gini (MDG). The taxa with the highest MDA across all health categories were *Pseudomonas* (MDA = 4.851367), *Aestuariibacter* (MDA = 3.827359), *Fusibacter* (MDA = 3.658612), and an unclassified genus in the family Cyclobacteriaceae (MDA = 3.535318). Similarly, the taxa with the highest MDG across all health categories were *Pseudomonas* (MDG = 0.9959390), *Alkalispirochaeta* (MDG = 0.7627235), *Endozoicomonas* (MDG = 0.9200709), and *Desulfatitalea* (MDG = 0.7503214). These taxa are important across all classes with *Pseudomonas* consistently ranked highly across both importance metrics. The importance of these taxa indicates that these taxa respond strongly to disease exposure and progression.

In contrast, taxa that were associated strongly with a single disease condition may be better predictors of health status. For example, unclassified taxa within the family Caulobacteraceae were found to have high MDA and MDG within Naïve coral samples across all species. This indicates that this taxon may be a member of the naïve coral microbiome that was lost upon sitewide exposure to disease (all disease classes after SCTLD arrived to the region). Mean Decrease in Accuracy (MDA) is the best metric to identify potential predictors of a class, as this metric measures how much the model's overall accuracy decreases when a specific feature (e.g., a taxon) is randomly permuted. Taxa that were predictors of the Disease Lesion health status based on MDA included *Halarcobacter*, *Fusibacter*, *Amphritea*, and *Desulfocella* (**Figure 4**). These taxa may play a potential role in lesion formation or persistence. *Ferrimonas* was most predictive of Diseased Unaffected samples, possibly indicating microbiome adaptations that help maintain tissue integrity despite adjacent lesions. Notably, *Fusibacter*, *Ferrimonas*, and the unclassified taxon from Family Caulobacteraceae were associated with Apparently Healthy samples, suggesting that these samples represent a transitional stage between Naïve and Diseased Unaffected.

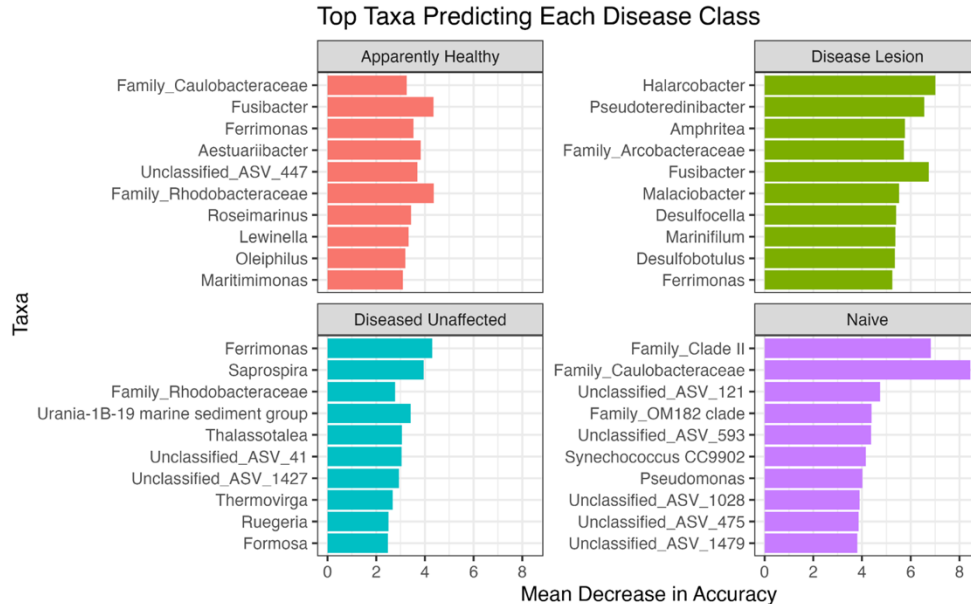


Figure 4: Top microbial taxa predictive of each disease class based on Mean Decrease in Accuracy (MDA) from Random Forest classification. Each bar represents the MDA value of a taxon, indicating its importance in predicting the corresponding disease class. Higher MDA values suggest greater contribution to model accuracy when predicting that class. Taxa were ordered by importance within each facet. Only the top 10 predictors per class are shown. Disease classes include Diseased Lesion, Diseased Unaffected, Apparently Healthy (exposed), and Naive (unexposed) samples.

3.1.4. Integrative analyses of metagenomic data with transcriptomic data

To investigate the functional relationships between host Single Copy Orthologs (SCOs) and microbial metabolic pathways, correlation analysis and network analysis were performed separately on each disease class. Host genes and bacterial metabolic pathways were first clustered across all samples to identify co-occurring genes/pathways that would influence network modularity. Spearman correlation with hierarchical clustering was used to find biologically relevant modules of co-occurring genes and pathways (i.e., host gene to host gene correlations greater than 0.9) that were persistent across disease states. This assumes that biological functions within the host and the bacterial community are not completely restructured by disease development. Then, abundance of pathways within each module was averaged (as read count data were already normalized) to produce a summary metric (mean abundance) for each module. Samples were then split by disease class (e.g., Diseased Lesion, Naïve) and Spearman correlations (positive and negative) were identified between host genes and bacterial pathways within disease classes.

A total of 188 clusters of bacterial pathways and 24 clusters of SCOs were produced from hierarchical clustering. Data were then examined by disease class to identify how interactions between host gene groups and bacterial metabolic pathways were influenced by disease development. For disease lesion samples, correlation analysis between pathways and SCOs found a total of 229 significant correlations ($|r| \geq 0.8$) FDR-corrected $p < 0.05$, involving 24 unique SCOs and 87 unique bacterial pathways. The resulting bipartite network consisted of 107 nodes (24 SCO nodes and 83 Pathway nodes) and 106 edges after the removal of 2 dyads (nodes only connected to one another).

The average degree for the disease network was 1.981308, meaning that each node was connected on average to ~2 other nodes. Density of the network was low at 0.01869159, where a density of 1 means that every node is connected to every other node. Modularity for the network was 0.7992376 (scale of 0 to 1) suggesting that there were groups of genes/pathways that were highly connected and that there was a higher degree of internal connectivity within these groups than between them.

Cluster 5, identified through network analysis, exhibited the highest connectivity among all modules, with a degree of 23, indicating it was the most functionally integrated cluster in the host-microbe network. This was a complex module of SCOs associated with DNA integration, repair, recombination, and synthesis (indicative of tissue damage and attempted regeneration) clustered with other SCOs associated with immune and inflammatory response, regulation of apoptosis and mitochondrial ADP transmembrane transport. Interestingly, host cluster 5 also contained genes involved in ossification and osteoblast-related genes, which although derived from vertebrate annotations, may represent analogs of extracellular matrix production, calcification, or structural remodeling in coral tissue. Lastly, this cluster contained genes involved in Wnt signaling (including canonical and planar cell polarity pathways), which governs cell fate, polarity, and tissue organization, likely reflects a breakdown and attempted re-patterning of host tissue architecture under stress. Cluster 5 was strongly associated with bacterial pathways involved in core aerobic energy metabolism (bacterial TCA cycle, ubiquinol biosynthesis) but also stress-associated fermentative and anaerobic pathways (ethanol degradation, nitrate reduction, proline-to-cytochrome electron transfer), potentially indicative of hypoxic microenvironments. Also linked to cluster 5 were genes involved in the degradation of amino acids; fatty acids; and sugars, consistent with nutrient scavenging from host tissue or decaying organic matter.

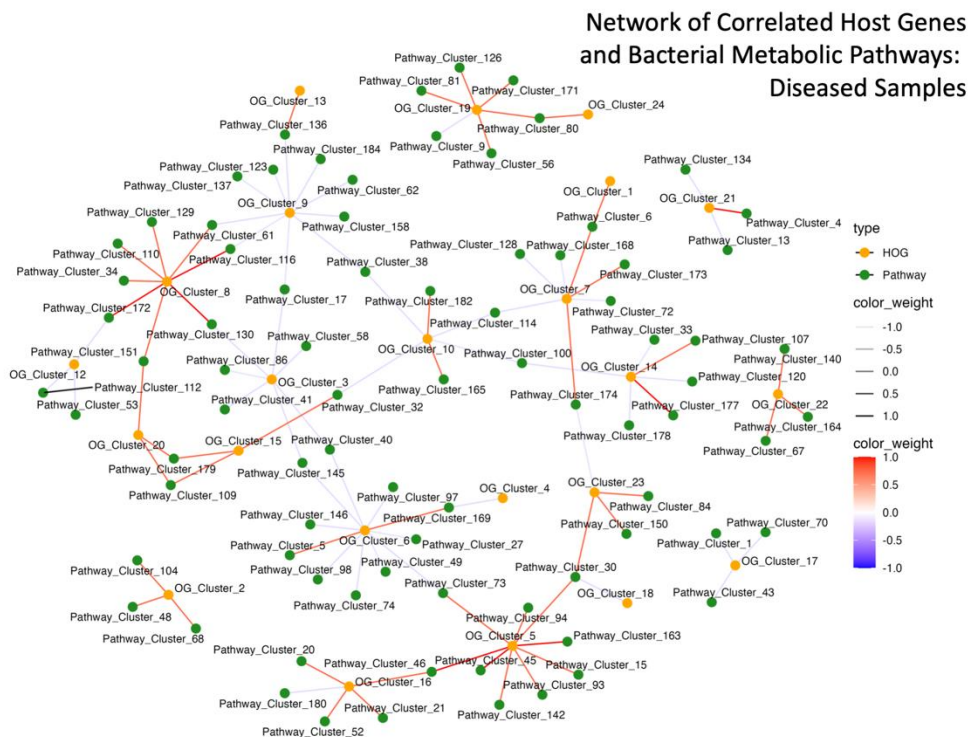


Table 5: Average degree, density, and network modularity for each of the three networks created correlating Single Copy Orthologs (SCOs) and bacterial functional pathways. Average degree is calculated as the average number of connections (edges) each node has in the network. Density is calculated as the proportion of all possible edges in the network that are actually present. Modularity is a measure of how well the network can be divided into distinct modules or communities.

	Avg_Degree	Density	Modularity
Apparently Healthy	6.191335740	0.0111959	0.53721668
Naïve	4	0.17391304	0.48460614
Disease Lesion	1.981308	0.01869159	0.7992376

3.2. Viral Community Analyses

3.2.1. Virus sequence detection and classification success

Figure 5: Network visualization of statistically significant correlations between Single Copy Orthologs (SCOs) (orange nodes) and bacterial functional pathways (green nodes). Edges represent Spearman correlations between nodes, with edge color and transparency indicating the strength and direction of the correlation (blue = strong negative, red = strong positive, white = neutral). Node placement is determined by the Fruchterman-Reingold layout algorithm, which positions strongly connected nodes closer together to emphasize modular structure. Only correlations above $|r| \geq 0.8$ are shown. Node labels indicate GO terms or pathway names; edge transparency reflects correlation magnitude.

Viral sequences were identified independently from RNA and DNA sequencing libraries to generate a final non-redundant set of putative viral contigs. Evidence of viral origin for each sequence was supported by one or more detection or classification tools. In the metatranscriptomic (RNA sequencing) dataset, the program deep6 predicted 3,058,884 sequences as potentially viral. Of these ~3 million putative virus sequences, 621,896 sequences were determined to be likely viral representing whole or partial RNA virus genomes or DNA/RNA virus transcripts (herein referred to as metatV sequence set) that were the focus of downstream analyses. From the metagenomic data, there was a starting total of 22,881,581 non-redundant contigs sourced from all sample libraries. Of these, 17,747,362 were determined to be of cellular origin either via alignment to known coral/dinoflagellate genome sequences or by the program CAT. From the remainder of the sequences, post length filtering (<1000 bp was removed), there were 1,311,959 putative non-cellular sequences, of which 10,774 had substantial evidence indicating viral origin (herein referred to as metagV sequence set). These 10,774 putative virus sequences represent whole or partial DNA virus genomic sequences.

Classifying viral sequences from non-model systems like coral reefs with sparse reference databases makes confident assignments at lower ranks (Order–Species) challenging. For that reason, we only report taxonomic results at ranks where classification confidence is highest (e.g., Realm–Class). Putative viral sequences from both metatranscriptome and metagenome sequence sets (metatV and metagV, respectively) spanned all known viral realms. However, only the metatV set contained sequences belonging to the viral Realm Adnaviria, which is a double-stranded DNA virus realm that includes filamentous archaea-infected viruses. In both sequence sets, the top three most abundance classes were Caudoviricetes (tailed, prokaryotic dsDNA viruses), Retraviricetes (eukaryote-infecting retroviruses), and Megaviricetes (eukaryote-infecting giant dsDNA viruses). Of the 535,006 metatV transcripts classified at the class rank, 338,084 (64.4%) fell into six classes - five of which are eukaryotic virus classes: Megaviricetes (n = 216,121), Revtraviricetes (n = 45,421), Pokkesviricetes (n = 33,754), Naldaviricetes (n = 26,625), and Herviviricetes (n = 16,163).

3.2.2. Core virome characterization and evidence for phyllosymbiosis

Core virome analysis revealed 32 different viral sequence taxa (i.e., sequences grouped into a recognized virus class or sequences with a provided classification but lacking a class designation) that were present in at least 95% of samples within a species. Of these 32 groups, 17 corresponded to recognized viral classes spanning diverse DNA and RNA virus taxa (**Table 6**). These included 10 DNA virus classes, some of which had the highest number of recovered transcripts, such as Megaviricetes (giant viruses), Caudoviricetes (tailed bacteriophages), Herviviricetes (herpesviruses), Naldaviricetes, and Pokkesviricetes (poxviruses) (**Table 6A**). The remaining seven classes represented RNA virus taxa, including Duplopiviricetes, Pisoniviricetes, Revtraviricetes, and most notably Stelpaviricetes, which includes +ssRNA filamentous viruses (**Table 6B**). Hierarchical clustering based on Bray–Curtis distances of core virome taxa revealed evidence of phyllosymbiosis (the tendency of closely related species to share similar viral communities), with *O. faveolata* and *O. franksi* forming a distinct cluster, while *M. cavernosa* and *C. natans* were more dissimilar (Figure 6).

Table 6 (next pages). Summary of viral classes identified across all coral species assessed. (A) DNA virus classes and (B) RNA virus classes detected in coral viromes. For each class, the genome type is provided based on information within the International Committee on Taxonomy of Viruses (ICTV) taxonomy database (2024 release; Walker et al., 2022). Genome type includes double-stranded (ds) or single-stranded(ss) DNA/RNA, as well as reverse-transcribing genomes. *Indicates viral class present in all core viromes except *C. natans*. **Indicates viral class present in all core viromes except *M. cavernosa*.

A) DNA Virus class	Genome type
Maveriviricetes	dsDNA genomes (~17–30 kb, circular or linear)
Megaviricetes	dsDNA genomes (often >100 kb)
Naldaviricetes	Large circular dsDNA genomes
Pokkesviricetes	Linear dsDNA genomes
Polintoviricetes	dsDNA genomes
Quintoviricetes*	Single-stranded DNA (ssDNA) genomes
Tectiliviricetes	Linear dsDNA genomes (non-enveloped icosahedral virions)
Caudoviricetes	dsDNA genomes
Faserviricetes	Circular ssDNA genomes (positive-sense)
Herviviricetes	dsDNA genomes

B) RNA Virus class	Genome type
Alsuviricetes	Positive-sense ssRNA genomes
Duplopiviricetes**	Double-stranded RNA (dsRNA) genomes
Ellioviricetes	Negative-sense ssRNA genomes (segmented)
Leviviricetes	Positive-sense ssRNA genomes
Pisoniviricetes	Positive-sense ssRNA genomes
Revtraviricetes	Reverse-transcribing viral genomes (RNA or DNA)
Stelpaviricetes	Positive-sense ssRNA genomes

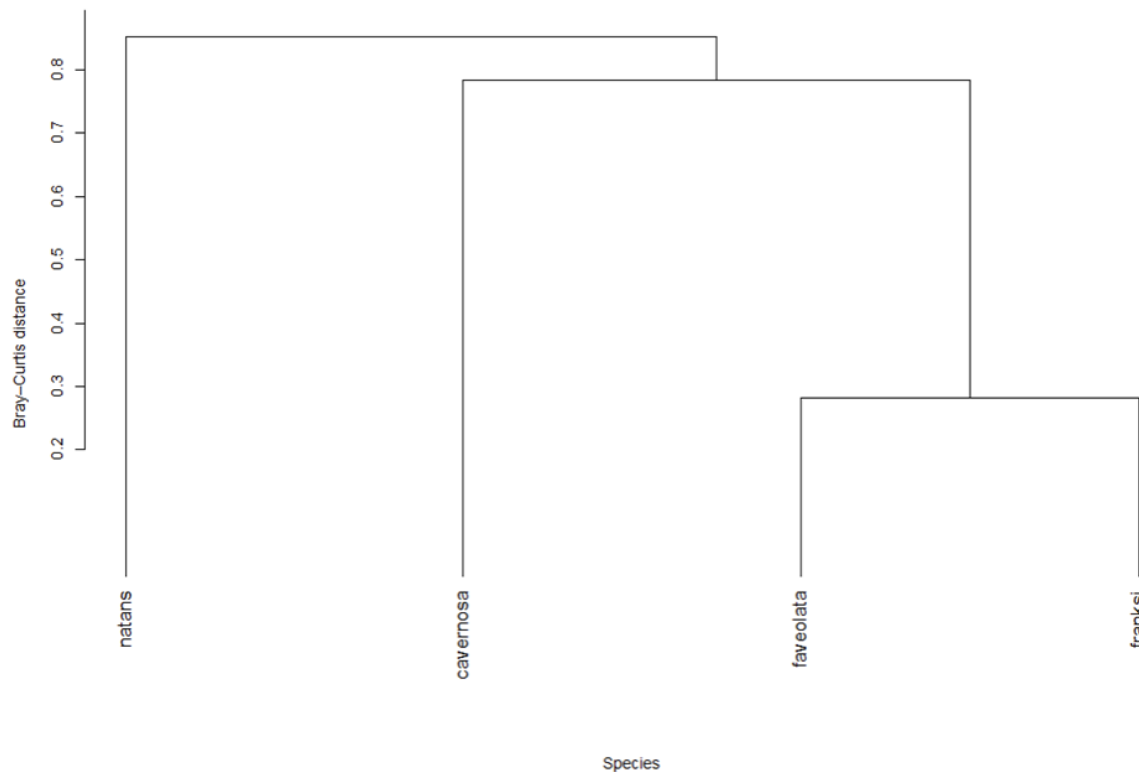


Figure 6. Hierarchical clustering of coral species based on core viral class transcript presence/abundance using Bray-Curtis dissimilarities. The dendrogram illustrates differences in core virome composition and expression profiles across species, with *O. faveolata* and *O. franksi* forming a distinct cluster, while *M. cavernosa* and *C. natans* are more dissimilar. Distances on the y-axis represent pairwise Bray-Curtis dissimilarity values, indicating relative divergence in viral community structure among host species.

3.2.3 Virus community dynamics across time and health state

One major objective for Phase II of this project was to assess virus community variability over time, before and during the SCTL outbreak at Dry Tortugas National Park. As a first step, we assessed alpha diversity (Shannon's H' index) across sampling time points (T0: Pre-outbreak, T1: Early outbreak, T2: Late outbreak). When considering all species, we observed a statistically significant difference in alpha diversity between time points (Kruskal-Wallis chi-squared = 20.604, $df = 2$, $p = <0.01$). Pairwise comparisons between time points revealed that virus alpha diversity at T0 (pre-outbreak) was significantly different from both the early outbreak (T1; Wilcoxon rank-sum exact test, $p = 0.03$) and late outbreak (T2; Wilcoxon rank-sum exact test, $p = 0.01$). Additionally, T1 community alpha diversity also differed significantly from T2 (Wilcoxon rank-sum exact test, $p = 0.01$). When assessing whether this pattern was also observed within species, only *M. cavernosa* (Wilcoxon rank-sum exact test: T0 vs. T1, $p < 0.01$; T0 vs. T2, $p < 0.01$) and *O. faveolata* (Wilcoxon rank-sum exact test: T0 vs. T2, $p < 0.01$) virus communities showed significant differences across time points (**Figure 7**).

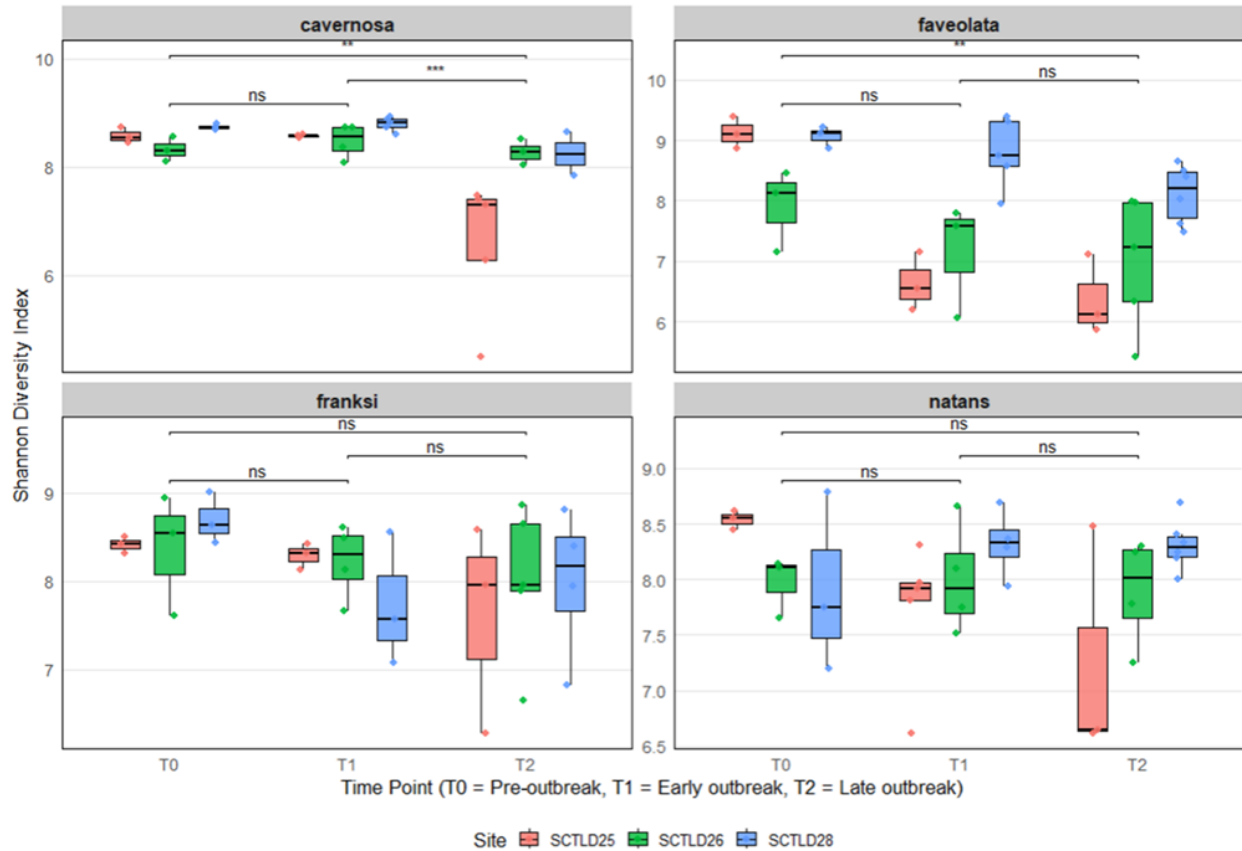


Figure 7. Viral alpha diversity (Shannon H' Index) across time points, coral species, and sites. Boxplots show the Shannon diversity index of viral communities associated with four coral species (*Montastraea cavernosa*, *Orbicella faveolata*, *Orbicella franksi*, and *Colpophyllia natans*) sampled at three time points relative to a stony coral tissue loss disease (SCTLD) outbreak: pre-outbreak (T0), early outbreak (T1), and late outbreak (T2). Colors represent samples from three reef sites (SCTLD25: red, SCTLD26: green, SCTLD28: blue). Asterisks indicate statistically significant pairwise comparisons between time points (Wilcoxon rank-sum exact test: $p < 0.05$, $p < 0.01$). Differences in alpha diversity patterns were observed across species and time points, with some species showing significant shifts in viral diversity during early and/or late outbreak

Next, we assessed virus community beta diversity across coral species, health status, and timepoint. We used PERMANOVA (adonis2, Bray-Curtis distance, 999 permutations) to assess the contribution of host species, time point, and health status to virus community composition. All three factors had statistically significant effects: host species explained the largest proportion of variance ($R^2 = 0.480$, $F = 41.69$, $p = 0.001$), followed by time point ($R^2 = 0.016$, $F = 2.02$, $p = 0.021$) and health status ($R^2 = 0.013$, $F = 1.65$, $p = 0.047$). Within-species PERMANOVAs revealed that only *O. faveolata* showed a clear health effect: dispersions were homogeneous ($p > 0.05$), and β -diversity differed significantly between healthy and diseased ($R^2 \approx 0.112$, $p = 0.001$). In *M. cavernosa*, although PERMANOVA flagged a strong time-point effect ($R^2 \approx 0.118$, $p = 0.003$), dispersion among T0/T1/T2 was unequal ($p = 0.003$), so that result is likely driven by spread rather than a true centroid shift; its health effect was marginal ($p = 0.06$). For *O. franksi* and *C. natans*, health status groups had significantly different dispersions ($p = 0.001$), making their borderline PERMANOVA health p-values ($p \approx 0.05$ – 0.08) uninterpretable. Neither species showed a reliable time-point effect (all $p > 0.3$). Thus, only *O. faveolata* exhibits a β -diversity shift by health status. Visualization of sample clustering based on

computed Bray-Curtis distances with a Principal Coordinates Analysis (PCoA) did not reveal any obvious clustering by health status or time point across species (**Figure 8**).

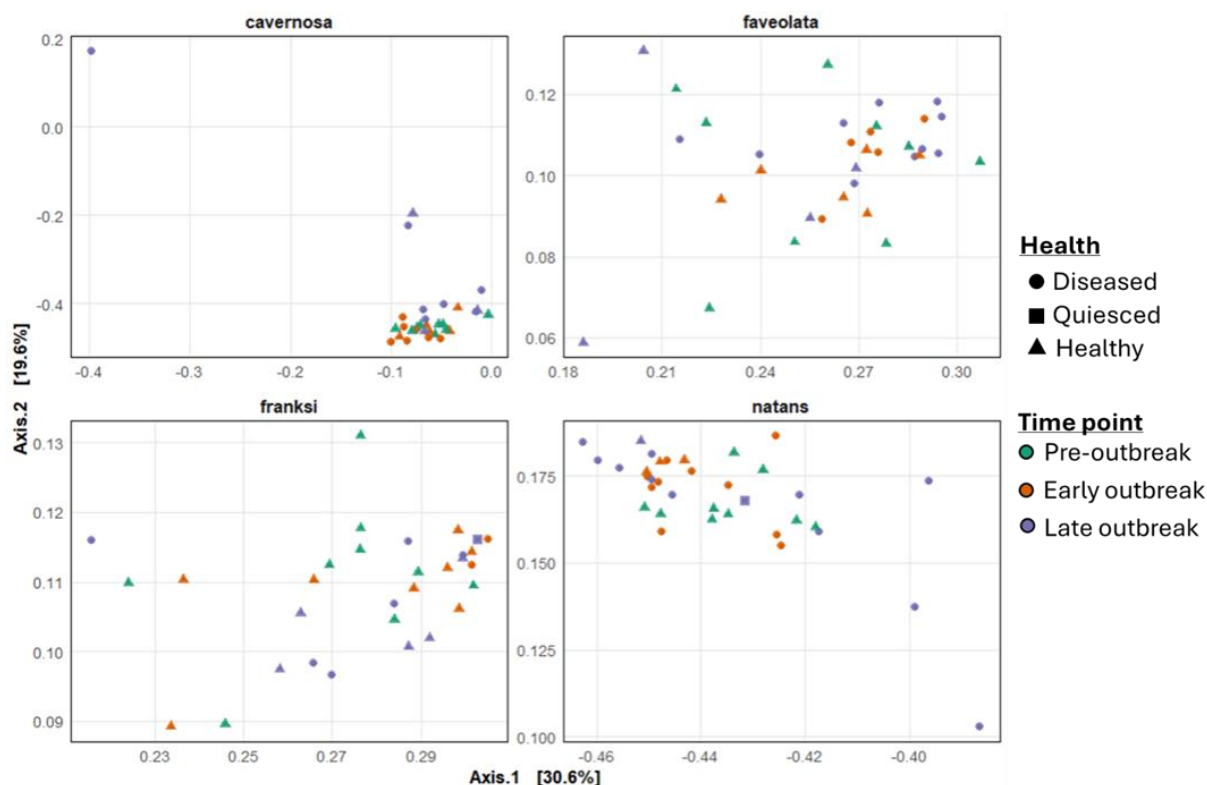


Figure 8. Principal Coordinates Analysis (PCoA) of viral community composition within each coral species, based on Bray-Curtis dissimilarities. Each panel represents one species (*M. cavernosa*, *O. faveolata*, *O. franksi*, *C. natans*). Points are colored by time point (T0 = pre-outbreak, T1 = early outbreak, T2 = late outbreak) and shaped by health status (diseased, healthy, quiesced). The percentage of variation explained by each axis is indicated in brackets.

3.2.3 Upregulated virus groups and genes within disease margin tissues

Differential abundance analysis revealed 6,588 sequences that were upregulated in diseased margin tissue samples across species. The six most upregulated taxa correspond to viral classes identified as part of the core virome: Megaviricetes, Caudoviricetes, Naldaviricetes, Pokkesviricetes, Retraviricetes, and Herviviricetes (**Figure 9**). CheckV results indicate that 6,535 of the upregulated sequences are likely transcripts or genome fragments. The remaining upregulated sequences were classified as high-quality ($n = 52$, showing >90% genome completeness) or complete ($n = 2$, with evidence of being complete genomes). These complete and near-complete genomes were representative of members within the orders Ortervirales ($n = 46$; retroviruses) and Picornavirales ($n = 2$), while two transcripts remained unclassified. From the 6,588 sequences, functional annotations were found for just 930 open-reading frames (ORFs) by the program MetaCerberus. A large proportion of these ORFs were annotated as polyproteins or accessory factors ($n=464$) and hypothetical/unknown proteins ($n=142$). The remaining ORFs were annotated as proteins involved in reverse transcription ($n=107$), metabolic processes ($n=84$), virus replication/transcription ($n=84$), and viral structural proteins ($n=61$).

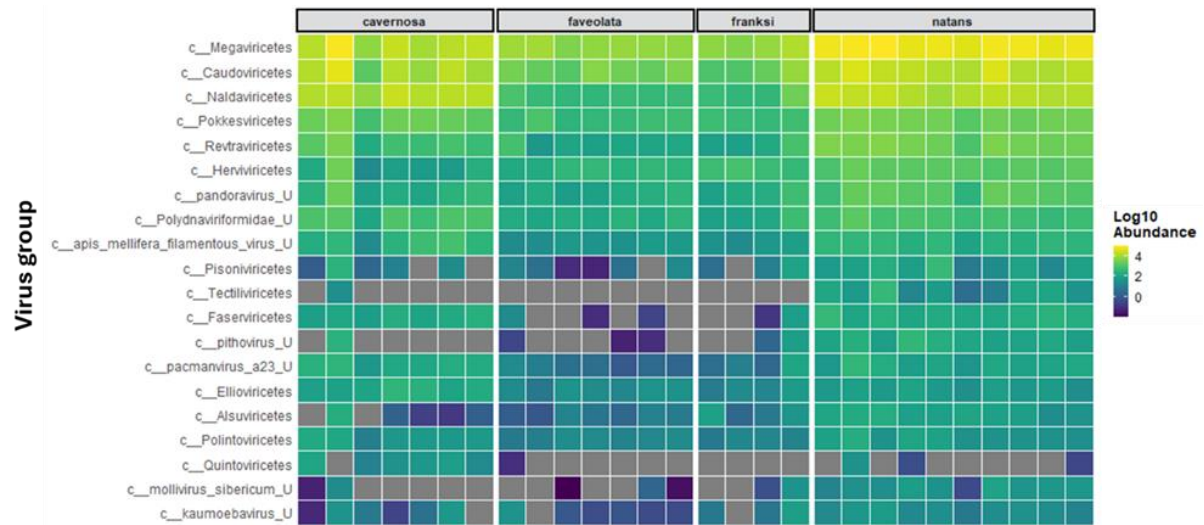


Figure 9. Heatmap of top 20 viral groups within significantly increased transcript abundance in diseased margin coral tissue samples (Wald test, DESeq2, adjusted $p < 0.1$). Abundances are based on TMM-normalized counts and log10-transformed for visualization. Each column represents a diseased sample, grouped by coral species (*M. cavernosa*, *O. faveolata*, *O. franksi*, *C. natans*), and each row corresponds to a viral class. Color intensity reflects transcript abundance, with gray indicating absence. Taxa labels containing “_U” indicate that that group of sequences were classified as a virus that has yet to be assigned to a recognized class.

Table 7. Virus-derived proteins predicted from differentially abundant transcripts in diseased coral tissues, grouped by functional category based on annotations from virus focused databases PVOG, VOG, PHROG, and GVDB, as well as potential auxiliary metabolic genes identified by NFixDB and AMRFinder.

Category	Count	Insight
Other	464	Diverse polyproteins and accessory factors, such as RNA replication polyproteins, enzymatic polyproteins, and unclassified enzymatic domains that may bundle multiple functions in a single ORF.
Hypothetical/Unknown	142	Uncharacterized proteins may represent novel viral functions or unannotated domains.
Reverse transcription	107	Reverse transcriptase proteins suggest retrotransposition or genome integration events by retroviruses.
Metabolic (AMG-like)	84	Auxiliary metabolic genes manipulate host metabolic pathways to support viral replication influencing host interactions within tissues.
Replication/Transcription	72	These enzymes drive viral genome replication and transcription, fueling viral proliferation.
Structural	61	Structural proteins form the virus capsid or mediate host attachment, essential for infectivity. Indicative of later infection stage (virus particle production).

3.3. Histological Analyses

Histopathology analysis of the diseased and healthy corals from the Dry Tortugas National Park revealed a similar pattern for diseased *M. cavernosa* and *O. faveolata* corals as previous studies (Beavers et al., 2023) Rossin et al in review) from St. Thomas, USVI and Florida. For the Dry Tortugas samples, an interesting discrepancy from previous studies was in exocytosis, where healthy colonies had high exocytosis. This is unusual compared to other studies but could be connected to a higher turnover of symbionts in this region. Methodology and results are included in a public Github repository (https://github.com/ashleyrossin/dry_tortugas). The goal of this repository is to provide a public reference for how to collect quantitative histological data and how it can effectively be used in disease identification.

3.4. Transcriptomic & Associated Integrative Analyses

3.4.1. Species independent markers of disease resistance

Ortholog analyses comparing refined transcriptomes for the four species identified a set of 7,462 conserved single copy orthologs. Analysis of this set of orthologs did not identify any species independent resistance markers; no orthologs were consistently significantly differentially expressed between resistant and susceptible colonies across all species either before or during the disease outbreak. We did however identify 533 transcripts which were consistently responsive to SCLTD (differentially expressed between apparently healthy and disease margin tissue) across all species. Of these only 4 were significant in both June and August, 491 were differentially expressed in June only, and 38 in August only. A total of 61 of these orthologs (~12%) were identified as putatively associated with immunity, all but 4 of which were significantly differentially expressed in June only. Immune orthologs were roughly evenly divided between up and down regulation in response to disease (29 up, 32 down).

Notable trends of differential expression include strong upregulation of orthologs associated with oxidative stress and immune receptors in June only (**Figure 10**). Furthermore, orthologs associated with inflammation and antiviral response were frequently downregulated in response to disease (**Figure 11**)

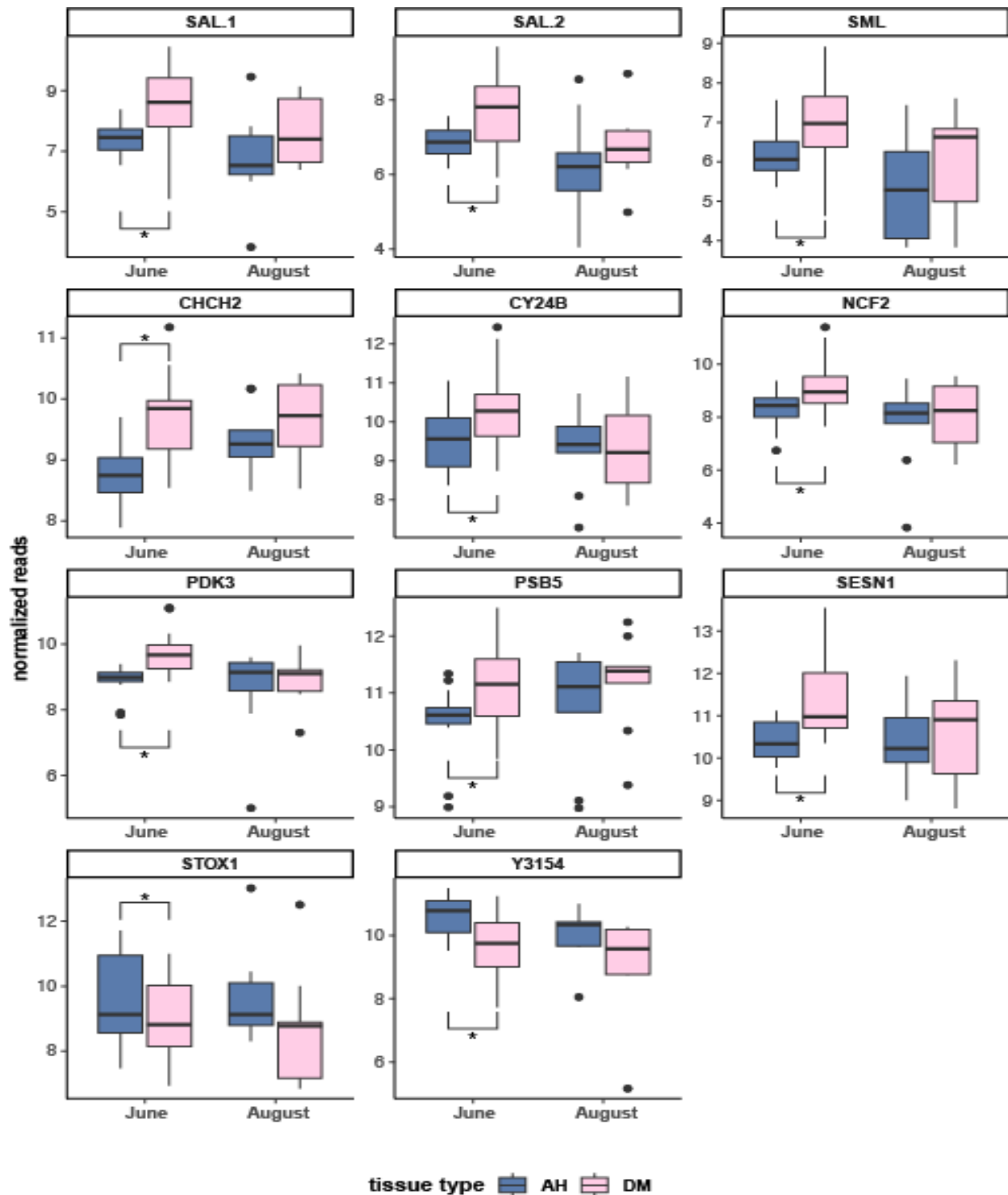


Figure 10: Box plots of expression of immune related orthologs involved in immune recognition (top row) and oxidative stress (bottom three rows), the majority of which were significantly upregulated in disease margin tissue compared to apparently healthy portions of colonies. Stars indicate significant differences between tissues. AH- apparently healthy, DM- disease margin

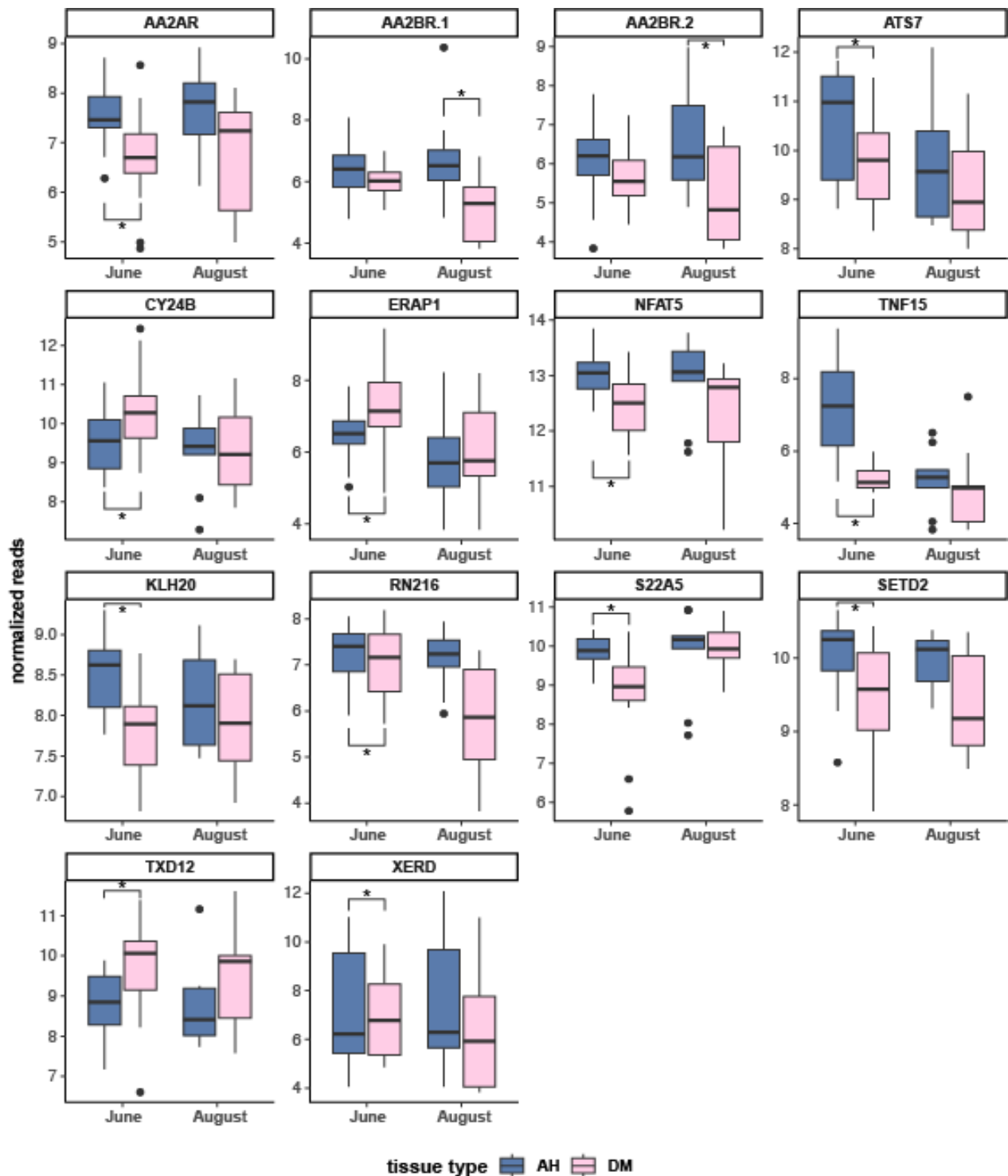


Figure 11: Box plots of expression of immune related orthologs involved in inflammation (top two rows) and viral response (bottom two rows), most of which were significantly downregulated in disease margin tissue compared to apparently healthy tissue. Stars indicate significant differences between tissues. AH- apparently healthy, DM- disease margin

3.4.2. Species independent correlations between microbiome/viral community composition and histological markers and host gene expression

Investigation of associations between normalized ortholog expression and 16S microbial data revealed strong correlations between microbial abundance and host gene expression. A total of 356 out of 390 identified bacterial families were correlated to at least one ortholog; every ortholog was correlated to at least one bacterial family. Herein we focus on the top 5% most correlated bacterial families (~20; **Table 8**) and their associations with the top 5% most correlated orthologs (~376), following established methods from other systems. The top 20 most correlated bacterial families included common coral-associated microbes including Endozoicomonadaceae.

Table 8: Summary of microbial families associated with the highest number of orthologs. Listed are number of associations with the top 5% most correlated orthologs, and number of enriched biological process GO terms.

Microbial Family	Associated Orthos (top 5%)	Associated GO Terms
Blastocatellaceae	259	20
Caminicellaceae	260	22
Chloroflexaceae	225	32
Christensenellaceae	293	25
Chromatiaceae	138	5
Desulfobacteraceae	179	26
Desulfotomaculales Incertae Sedis	280	16
Dethiobacteraceae	306	35
Endozoicomonadaceae	228	15
Halobacteroidaceae	139	15
Inquilinaceae	161	16
Kiloniellaceae	201	20
Lentimicrobiaceae	157	4
Lentisphaeraceae	240	30
Nisaeaceae	257	20
Oligoflexaceae	276	27
Prolixibacteraceae	229	9
Rhodocyclaceae	68	264

Spirochaetaceae	230	34
Terasakiellaceae	238	27

We then considered gene ontology enrichment of the associations between top families and top genes at the level of family. Fifteen microbial families had associations with top genes which were enriched for processes associated with immunity, seven of which were associated with two or more terms (**Figure 12**). Of these, four families were positively associated with immunity, two were neutrally, and one was negatively associated. Of all families, Rhodocyclaceae was associated with the most immune terms, though no clear trend of positive or negative association with immunity could be ascertained.

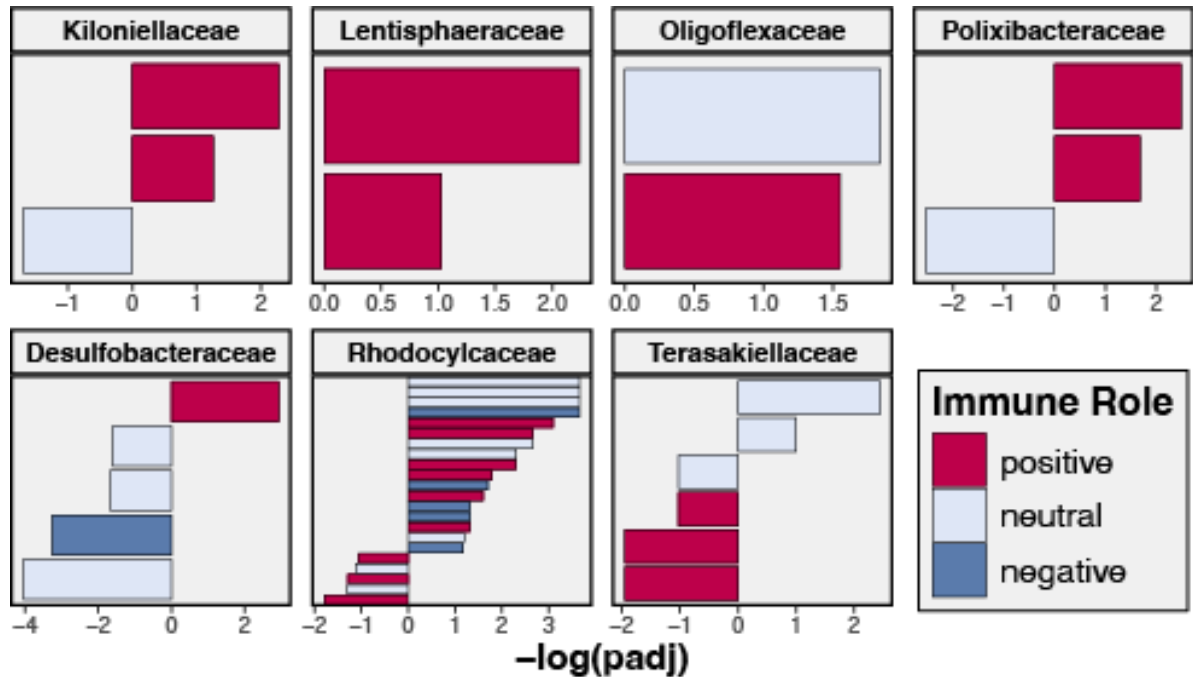


Figure 12: Significance and enrichment of biological process go terms of interest associated with 7 of the microbial families of interest. Only those families which were associated with at least two terms of interest are shown. Families are ordered based on associations: the top row includes families with positive associations and the bottom row is split between those with neutral and negative associations. Each bar indicates a significantly enriched biological GO term; color indicates the association of the term with immunity; the direction of bar indicates positive or negative enrichment; the magnitude of bars indicates the negative log of the adjusted p value.

Next, we considered associations between normalized viral sequence abundance (aggregated at the level of viral Order or similar taxonomic groups) and normalized ortholog abundance. In general viral sequence abundance was highly correlated with normalized ortholog abundance (**Table 9**). Every viral group was significantly associated with at least one ortholog and vice versa. Furthermore, of the top 22 orders, 12 (over half) were correlated to the full set of top genes. Viral groups were consistently associated with a core set of biological process GO terms, including a set of six immune-related terms mostly related to antigen presentation. Nearly every group was positively associated with these terms, and negatively associated with melanin biosynthesis (**Figure 13**).

Table 9: Summary of viral group associated with the highest number of orthologs. Listed are number of associations with the top 5% most correlated orthologs, and number of enriched biological process GO terms.

Viral Group	Associated Orthos (top 5%)	Associated GO Terms
Bingvirus	383	24
Brochothrix phage BL3	383	25
Burrovirus	336	28
Carmenvirus	383	24
Clostridium phage phiCT19406C	383	24
Decurrovirus	383	25
Godonkavirus	328	23
Halcyonevirus	383	24
Klebsiella phage phiKO2	240	11
Lacusarxvirus	184	17
Lightbulbvirus	355	25
Nonagvirus	344	25
Pseudoalteromonas phage H103	383	25
Pseudomonas phage JBD44	379	24
Schitoviridae	295	32
Spizizenvirus	383	24
Streptococcus phage phiARI0746	383	24
Streptococcus phage phiNJ2	383	25
Uetakevirus	318	22
Vertoviridae	317	18
Vividuovirus	383	25

Zobellviridae	383	24
---------------	-----	----

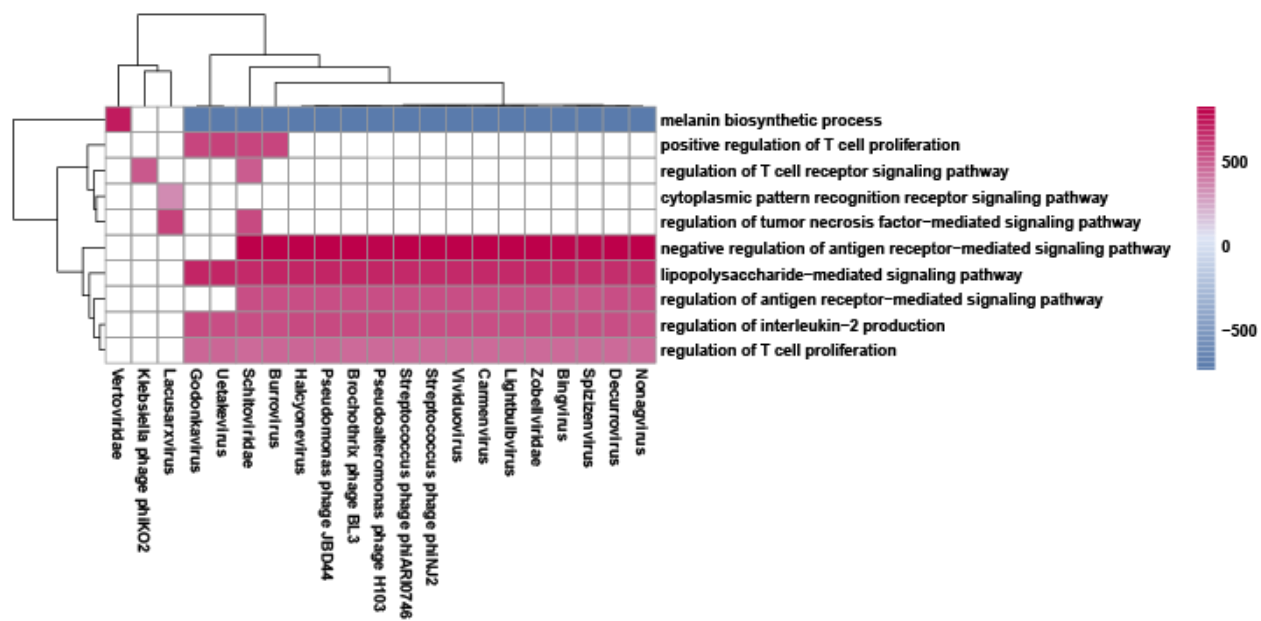


Figure 13: Heatmap displaying enrichment of significantly enriched immune-related GO terms associated with each viral order of interest. Colors are representative of delta rank (metric of positive or negative enrichment); blank/white cells indicate nonsignificant enrichment. Viral orders (columns) and GO terms (rows) are hierarchically clustered based on patterns of enrichment.

Finally, to further consider associations between host gene expression and histological markers of interest as well as our identified microbial families of interest (**Table 8**) we constructed a gene coexpression network with our normalized ortholog read count matrix, and the correlated average expression of resultant modules (groups of orthologs) to histological traits of interest and relative abundance of microbial families of interest. Our coexpression network consisted of 15 modules ranging in size from 33 to 1,727 transcripts. Overall, module expression was highly correlated to traits of interest (**Figure 14**). We saw strong patterns of association between most microbial families of interest and module gene expression, with the exception of Halobacteroidaceae, which was not correlated to any modules, and Lentimicrobiaceae, which was only correlated to two modules. Histological traits vacuolization (mean_symb_vac) and max vacuole size (mean_max_vac) were highly correlated to gene expression, but exocytosis (mean_prop_exo) was only correlated to one module of interest. Likewise, the red and brown modules were most correlated to bacterial families and histological traits of interest, though neither had clear roles in immune function. The brown module was enriched for orthologs related to sensory and nervous processes, and the red module showed no significant biological process enrichment, though it did contain several hub (highly connected) orthologs related to apoptosis. Gene ontology enrichment analysis revealed one clear immune module, dark green, which was enriched for over twenty biological process GO terms related to immunity, and several more related to response to biotic stimuli/stress. This module was positively associated with Lentisphaeraceae, Nisaeaceae, and Oligoflexaceae abundance, but negatively associated with Endozoicomonadaceae abundance.

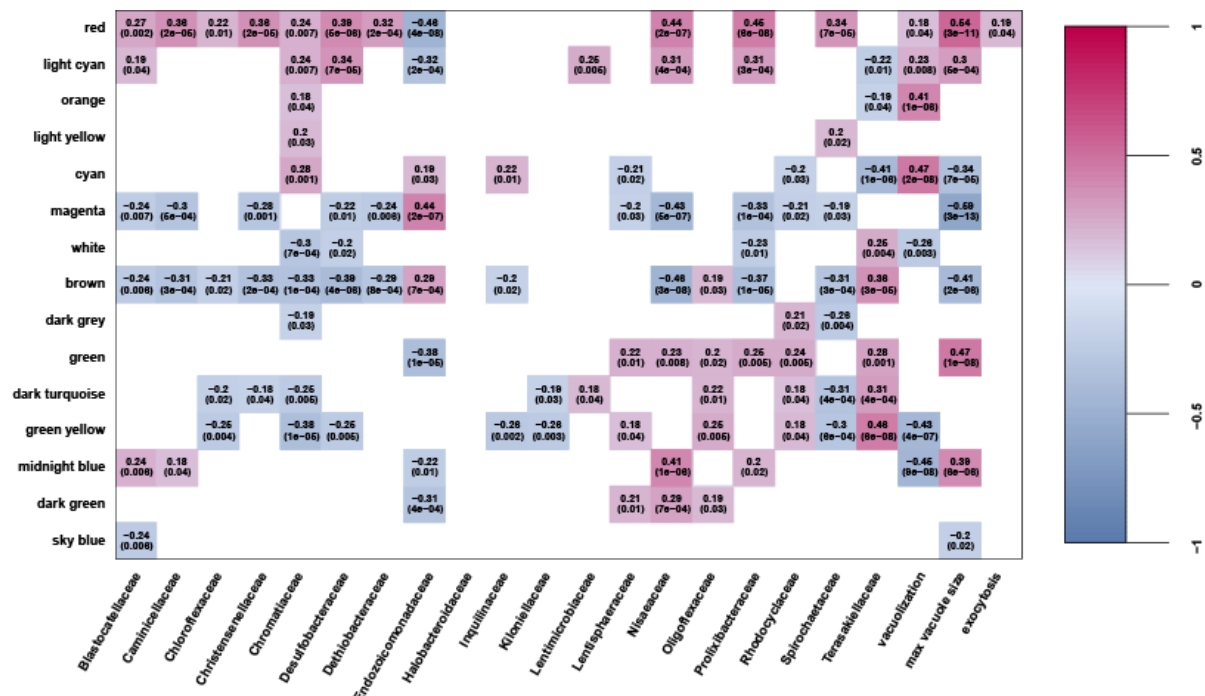


Figure 14: Heatmap displaying associations between average expression of each coexpression network module (rows) and traits of interest: microbial family abundance and histological traits (columns). Only significant correlations ($p < 0.05$) are displayed. Correlation value (R^2) and p value are shown for each association. Boxes are colored based on correlation value in accordance with displayed scale (pink for positive associations, blue for negative associations).

3.5. Immunological Analysis of samples from project C21169

Differences in immunological activity were largely driven by time, regardless of SCTLD exposure or immune pathway in question (**Table 9**). Specifically, peroxidase, melanin abundance, and antibacterial activity all significantly changed over the course of the experiment regardless of treatment. Catalase and total phenoloxidase were not significantly different as a result of any measured factor.

Table 9. Results of two-way repeated measures ANOVA testing the effects of treatment, timepoint, and their interaction on each of the measured immune activity assays independently. Bold font indicates significant *p* value.

Catalase					
Effect	DFn	DFd	F	<i>p</i> value	ges
Treatment	1	2	8.483	0.100	0.406
Time Point	3	6	3.305	0.099	0.315
Treatment:TimePoint	3	6	1.511	0.305	0.127
Peroxidase					
Effect	DFn	DFd	F	<i>p</i> value	ges
Treatment	1	3	7.493	0.071	0.078
Time Point	3	9	4.170	0.042*	0.444
Treatment:TimePoint	1	3.58	0.665	0.493	0.064
Total Phenoloxidase					
Effect	DFn	DFd	F	<i>p</i> value	ges
Treatment	1	3	1.457	0.314	0.041
Time Point	3	9	0.606	0.628	0.046
Treatment:TimePoint	3	9	0.649	0.603	0.077
Melanin					
Effect	DFn	DFd	F	<i>p</i> value	ges
Treatment	1	3	2.576	0.207	0.042
Time Point	3	9	3.922	0.048	0.283
Treatment:TimePoint	3	9	0.290	0.832	0.014
Antibacterial Activity					
Effect	DFn	DFd	F	<i>p</i> value	ges
Treatment	1	3	7.553	.071	0.206
Time Point	3	9	70.509	< 0.001**	0.871
Treatment:TimePoint	3	9	0.281	0.838	0.037

For the three immunological metrics which did change over time, the effect of time seemed to be driven by shifts midway through exposure, around timepoint 4 (**Figure 15, Table 10**). For example, peroxidase spiked at timepoint 4, when lesions were first appearing on exposed colonies. Similarly, antibacterial activity was significantly lower at timepoints 4 & 5, when lesions were apparent, compared to early exposure. Melanin showed similar patterns though no posthoc comparisons were significantly different.

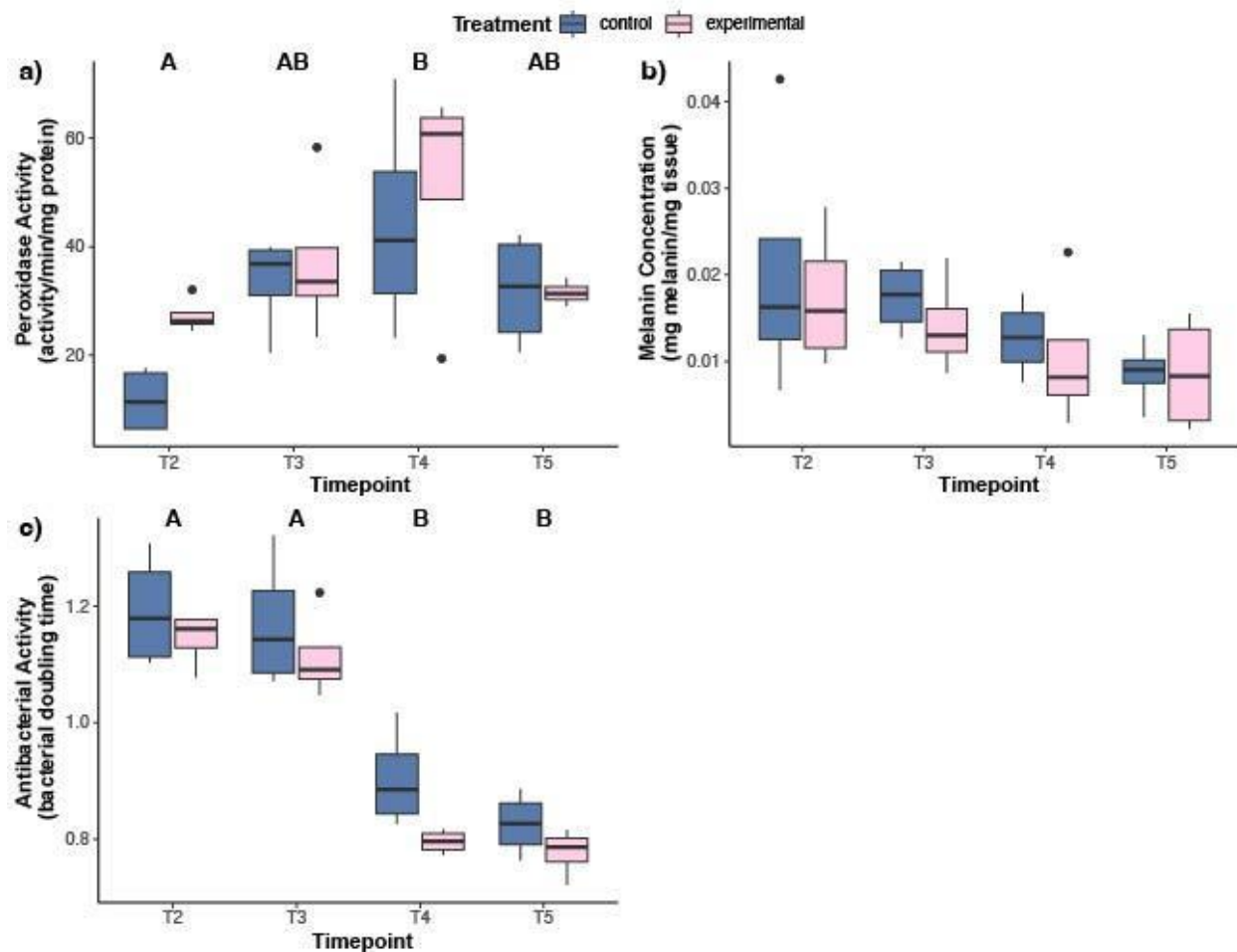


Figure 15: Box plot displaying immunological metric assay results for those assays where a significant effect of time was detected: **a)** peroxidase activity, **b)** melanin concentration, and **c)** antibacterial activity. Letters represent significant groups across timepoints, regardless of treatment; no pairwise differences were significant for melanin concentration after multiple test correction.

Table 10: Results of post-hoc pairwise t-tests comparing immunological metrics of interest across sampling timepoints. Results only shown for those metrics where the 2-way ANOVA identified a significant effect of timepoint. Significance (*p*) values are adjusted for multiple comparisons using a Bonferroni correction. Bold font indicates significant *padj* value.

Peroxidase				
Comparison	Statistic	df	<i>p</i> value	<i>padj</i>
T2-T3	-3.25	7	0.014	0.084
T2-T4	-3.79	7	0.007	0.041*
T2-T5	-3.55	7	0.009	0.056
T3-T4	-2.11	7	0.73	0.44
T3-T5	0.810	7	0.444	1
T4-T5	2.51	7	0.04	0.242
Melanin				
Comparison	Statistic	df	<i>p</i> value	<i>padj</i>
T2-T3	0.375	7	0.719	1
T2-T4	2.12	7	0.071	0.427
T2-T5	2.87	7	0.024	0.143
T3-T4	1.69	7	0.136	0.816
T3-T5	2.96	7	0.021	0.127
T4-T5	1.55	7	0.165	0.99
Antibacterial Activity				
Comparison	Statistic	df	<i>p</i> value	<i>padj</i>
T2-T3	0.729	7	0.489	1
T2-T4	7.46	7	< 0.001	< 0.001**
T2-T5	9.86	7	< 0.001	< 0.001**
T3-T4	6.53	7	< 0.001	0.002**
T3-T5	8.16	7	< 0.001	< 0.001**
T4-T5	2.03	7	0.082	0.49

3.6. Analysis of TEM images from coral sampled under projects C1E0A5 and C21169

3.6.1. TEM image analysis from Dry Tortugas corals sampled under C1E0A5

A total of 432 TEM micrographs from Dry Tortugas corals were analyzed across six species–health combinations. We quantified the proportion of images showing virus-like particles (VLPs) in coral tissue and within intracellular Symbiodiniaceae (zooxanthellae or “zoox”) at the level of overall prevalence and by VLP morphotype. In *C. natans* tissue, the median percentage of VLP-positive images was 92.9% in healthy samples (range 85.7–100%), 86.7% in apparently healthy (73.3–100%), and 81.8% in diseased colonies (0–100%; **Figure 16**). For *M. cavernosa*, tissue-level prevalence reached 94.1% in healthy samples, 69.2% in apparently healthy, and 25.0% in diseased samples (range 0–50%; **Figure 16**). Within Symbiodiniaceae cells, VLPs were observed in every group except one diseased outlier. In *C. natans*, intracellular medians were 100% in healthy, 60.0% in apparently healthy (range 50–100%), and 92.9% in diseased symbionts (0–100%; **Figure 17**). In *M. cavernosa*, medians were 100% in healthy, 46.7% in apparently healthy (33.3–60.0%), and 81.7% in diseased cells (16.7–100%; **Figure 17**). Both icosahedral and filamentous VLPs were detected in every *C. natans* tissue sample, with median prevalences of 85.7%, 100.0% and 90.9% (filamentous) and 92.9%, 92.9% and 92.9% (icosahedral) in healthy, apparently healthy, and diseased states, respectively (**Figure 18**). In *M.*

cavernosa tissues, icosahedral VLPs appeared at medians of 94.1% (healthy), 69.2% (apparently healthy) and 50.0% (diseased), whereas filamentous forms were present only in healthy and diseased tissues (medians 94.1% and 50.0%, respectively) and absent in the apparently healthy group (Figure TEM3). Across all comparisons (where sample representation allowed), no statistically significant effect of health state on VLP detection was found.

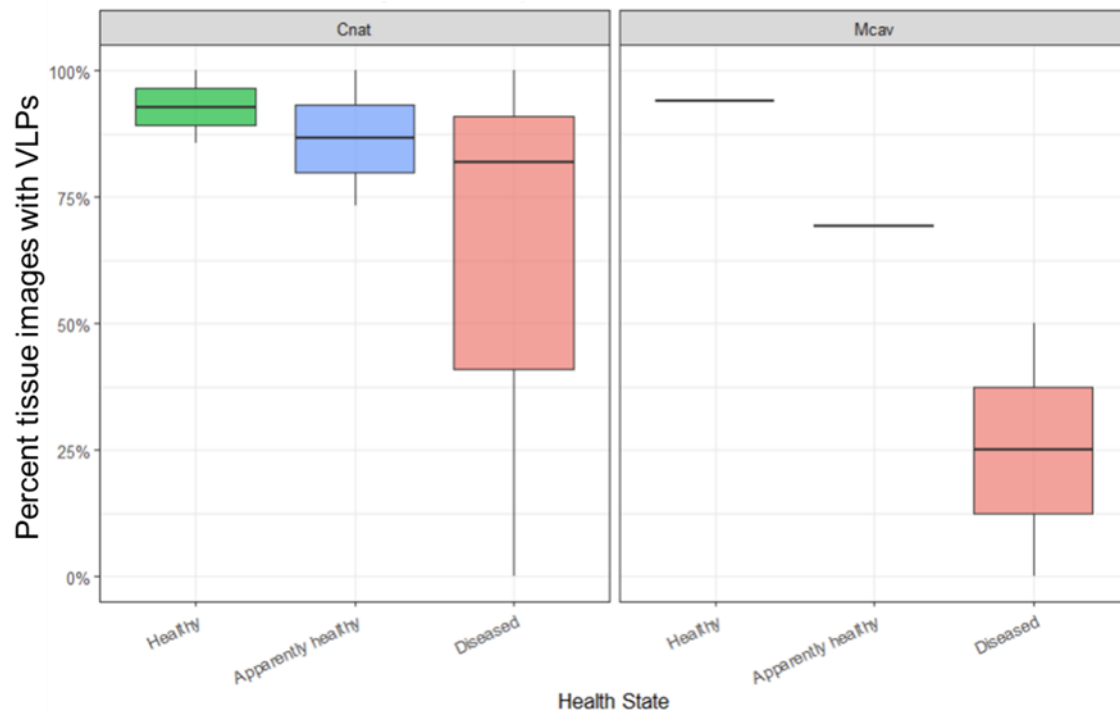


Figure 16. Prevalence of virus-like particles (VLPs) in coral tissue by species and health state. Boxplots show the percentage of transmission electron microscopy (TEM) tissue images containing VLPs in *Colpophyllia natans* (Cnat, left) and *Montastraea cavernosa* (Mcav, right) across three health states: Healthy, Apparently Healthy, and Diseased. Each point represents the proportion of images showing VLPs for an individual coral tissue sample, based on a set of TEM images collected per sample. Each box represents the interquartile range, the horizontal line indicates the median, and whiskers denote the full data range excluding outliers

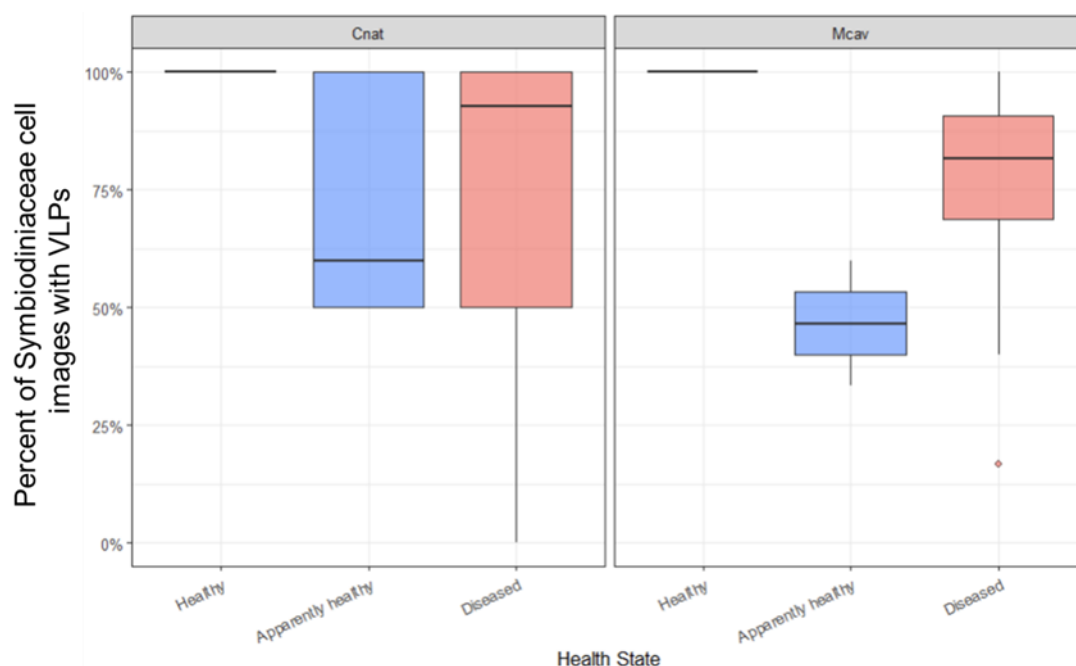


Figure 17. Prevalence of virus-like particles (VLPs) in symbiont (zooxanthellae) cells by coral species and health state. Boxplots show the percentage of symbiotic algal (zooxanthellae) cell images containing virus-like particles (VLPs) in *Colpophyllia natans* (Cnat, left) and *Montastraea cavernosa* (Mcav, right), across three health states: Healthy, Apparently Healthy, and Diseased. Each point represents the proportion of VLP-positive cells within the set of zooxanthellae images analyzed from a single coral tissue sample. Boxes represent the interquartile range; horizontal lines indicate medians, and whiskers denote the full data range excluding outliers.

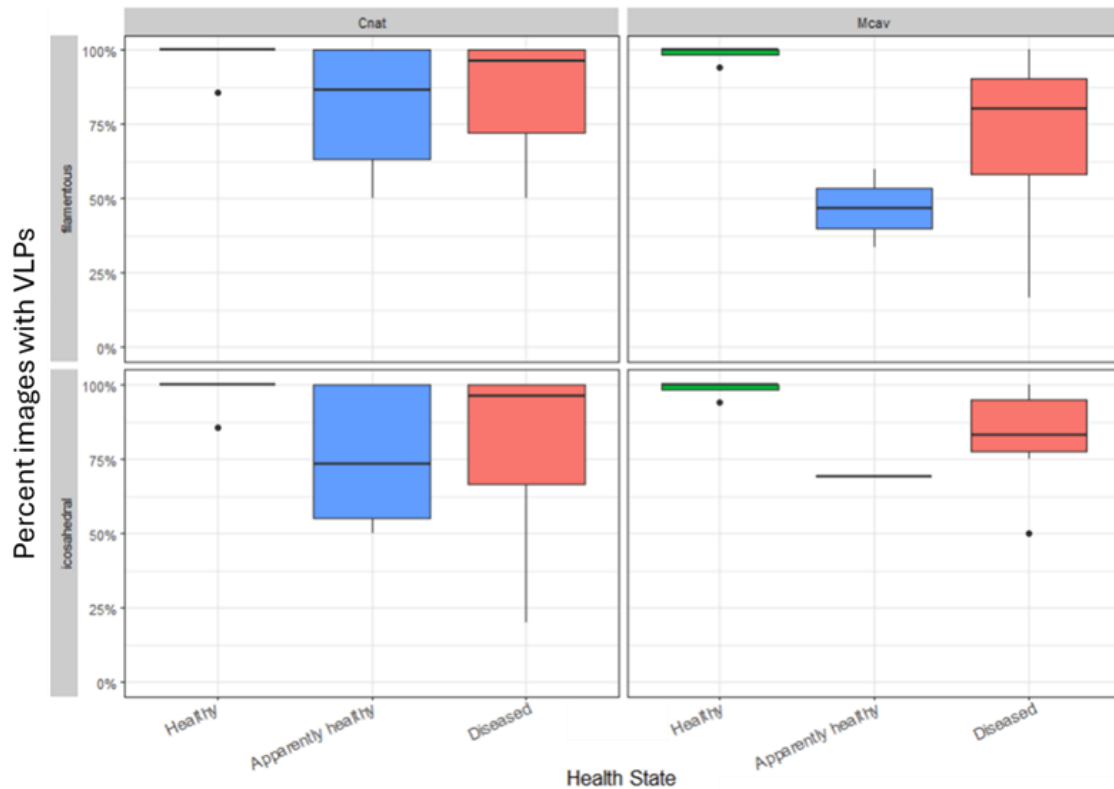


Figure 18. Detection frequency of virus-like particles (VLPs) in coral tissues by health state, species, and VLP morphology. Boxplots show the percentage of transmission electron microscopy (TEM) images containing virus-like particles (VLPs) in *Colpophyllia natans* (Cnat, left panels) and *Montastraea cavernosa* (Mcav, right panels), separated by VLP morphology: filamentous (top row) and icosahedral (bottom row). Coral health states are color-coded as Healthy (H, blue), Apparently Healthy (AH, blue), and Diseased (D, red). Each point represents the percentage of images showing VLPs from a single coral sample, calculated across both host tissue and symbiont (zooxanthellae) image sets.

3.6.2. TEM image analysis from time series experiment under project C21169

A total of 387 transmission electron microscopy (TEM) images were analyzed from the time series experiment conducted under project C21169, encompassing both coral tissue and symbiotic dinoflagellate (zooxanthellae) compartments. Images were derived from *Colpophyllia natans* ($n = 47$) and *Montastraea cavernosa* ($n = 340$), and categorized into four treatment groups: (i) Healthy-Control (HC), (ii) Healthy-Healthy (HH), (iii) Early-Disease (ED), and (iv) Disease-Diseased (DD). Virus-like particle (VLP) detection was quantified using two complementary metrics: the number of TEM fields in which each morphotype (filamentous or icosahedral) was observed, and the percentage of total fields per condition containing that morphotype. For *C. natans*, all 47 images were from DD tissues; within this group, filamentous VLPs were rarely detected (~5% of images), while icosahedral VLPs were present in ~30% of images. In contrast, *M. cavernosa* was represented across three conditions (HH ($n = 163$), HC ($n = 67$), and ED ($n = 110$)) with no images available for DD tissues. In *M. cavernosa*, filamentous VLPs were detected in 40% of HH images, dropped to 10% in HC, and rose slightly to 25% in ED. Icosahedral VLPs showed a similar but more pronounced trend: present in 80% of HH images, declining to 5% in HC, and detected in 12% of ED images.

4. DISCUSSION AND MANAGEMENT RECOMMENDATIONS

4.1. Discussion

4.1.1. Metagenomics

We found that the bacterial taxa *Halarcobacter*, *Fusibacter*, *Amphritea*, and *Desulfocella* were predictive of a diseased state with random forest analysis. *Halarcobacter* and the closely-related *Malaciobacter*, as well as many other indicator taxa of SCTLD, were shed by corals with SCTLD into filtered seawater in Evans et al. (2023). *Malaciobacter* was also found in higher abundances in the seawater surrounding corals affected by SCTLD (Bloomberg et al., 2025). *Arcobacter*, with 99% BLAST identity to the sequence in this study, was also identified as an indicator taxon of SCTLD (Becker et al., 2022). *Fusibacter* has been identified as an indicator of SCTLD across several studies (Becker et al., 2022; Evans et al., 2023; Huntley et al., 2022; Rosales et al., 2022), but we found that *Fusibacter* was not sufficiently discriminatory, as it was also a predictive taxon of an apparently healthy (but not naïve) state. We found that the genus *Ferrimonas* was a predictor of apparently healthy and diseased unaffected states, but there is no known association of *Ferrimonas* with either healthy or diseased corals: *Ferrimonas* are facultatively anaerobic, Fe(III)-reducing Gammaproteobacteria typically found in marine sediments and aquatic environments. The family Caulobacteraceae was found to be a strong predictor of naïve coral states. Bacteria in this family have been found to be tightly associated with the algal symbiont in the upper layers of coral tissue (Ainsworth et al., 2015) and it has been suggested that they may play a role in host nutrient cycling, though bacteria in this family harbor widespread roles.

Metagenomic functional analyses supported conclusions from community composition analyses, and revealed consistently high levels of aerobic microbial activity across all coral species in all health states except disease lesion, with elevated abundance of pathways related to aerobic respiration via cytochrome c (PWY-3781) and the TCA cycle. These functions may be attributed to aerobic bacteria such as *Endozoicomonas*, *Tistlia*, and members of *Terasakiellaceae*, consistent with patterns observed in 16S rRNA data and prior findings (Rosales et al., 2022). We also detected abundant pathways involved in sulfur cycling (sulfate reduction and oxidation) and nitrogen utilization (ammonia assimilation and nitrate reduction), alongside high relative abundance of sulfate-reducing and anaerobic taxa including *Desulfocella*, *Desulfovibrio*, *Halodesulfovibrio*, *Roseimarinus*, *Fusibacter*, and *Halanaerobium*. While sulfate reduction pathways were elevated in diseased corals, this difference was not statistically significant, potentially due to sample size imbalance.

These microbial functional shifts align with host gene expression patterns showing upregulation of lectins and oxidative stress genes during disease progression, suggesting a host response to both elevated microbial respiration and increased reactive oxygen species. Concurrent downregulation of antiviral genes may indicate suppression of viral defense pathways, potentially facilitating viral or microbial colonization and exacerbating disease. We hypothesize that the healthy coral microbiome is dominated by aerobic taxa including *Endozoicomonas*, with high abundance of aerobic respiration pathways (TCA cycle, cytochrome c). The loss of these pathways likely reflects the disruption of beneficial host-microbe interactions in which products of nutrient cycling may be transferred to the host. As disease develops, tissue damage and mucus overproduction may lead to localized hypoxia, favoring the proliferation of anaerobic and sulfate-reducing bacteria such as *Desulfovibrio* and *Halodesulfovibrio*, as evidenced by both taxonomic and functional profiling data. The byproducts of sulfate reduction and nitrate respiration include toxic compounds such as hydrogen sulfide that may harm coral tissue and inhibit healing (Jørgensen, 1982; Philippot & Højberg, 1999). These anaerobes may not be primary pathogens but amplify disease impacts through opportunistic colonization of diseased tissue. Coral gene expression data suggest that the host attempts to recognize and neutralize bacterial opportunists through the upregulation of host lectins, however, pattern recognition receptors such as lectins launch a non-specific immune response to both beneficial and harmful bacteria taxa (Kvennefors et al., 2008), potentially leading to the observed decline in symbionts such as *Endozoicomonas*. Indeed, lectins play a role in maintaining healthy symbiosis with the algal symbiont

(Wood-Charlson et al., 2006), and this relationship may also be impacted by the dramatic upregulation of lectins observed. Overexpression of host lectins can lead to tissue damage and comes at a considerable energetic cost to the coral, possibly contributing to the observed downregulation of antiviral responses.

The Fruchterman–Reingold layout used to produce a network visualization of interactions between host genes and bacterial metabolic pathways helped to emphasize modular structure by placing strongly correlated nodes in closer proximity, aiding interpretation of functional clusters. Overall, the network revealed structured, non-random associations between host gene expression and microbial functional capacity that may underlie key disease processes. The network shows numerous clusters of host genes that were strongly associated with microbial pathways during disease development. A total of 210 relationships were found to be statistically significant, and 24 host gene clusters and 186 microbial pathways were part of these significant associations. Clustering within the network highlighted several tightly connected modules, suggesting groups of co-regulated GO terms and pathways. 10 clusters were identified with 10 or greater edges; all were host genes.

The high degree of connectivity observed in Cluster 5 suggests that this group of host genes and associated bacterial pathways plays a central role in shaping the coral's physiological response to disease. The co-enrichment of host genes related to DNA repair, inflammation, apoptosis, and tissue remodeling indicates an active, but potentially overwhelmed, host attempt to manage cellular damage and maintain tissue integrity. The concurrent association with bacterial pathways involved in central metabolism (e.g., TCA cycle), anaerobic respiration (e.g., nitrate reduction, PWY0-1584), and degradation of amino acids and lipids points to a metabolically flexible microbial community exploiting host-derived substrates in a hypoxic, nutrient-rich environment. The presence of bacterial nitrate reduction and ethanol degradation pathways (ETOH-ACETYLCOA-ANA-PWY)—both capable of producing cytotoxic intermediates such as nitric oxide and acetaldehyde—suggests that microbial metabolism may contribute directly to tissue degradation and immune dysregulation (Philippot & Hojberg, 1999). The correlation between host developmental signaling (e.g., Wnt pathways, osteoblast-like gene expression) and microbial metabolism may reflect an attempted compensatory regeneration response that is being disrupted or hijacked by opportunistic microbes. Overall, the high interconnectivity of Cluster 5 supports the hypothesis that disease progression involves a tightly coupled but maladaptive interaction between host stress responses and microbial functional shifts. The observed correlation may reflect a pathological feedback loop, in which the host attempts to repair and re-pattern damaged tissue, while microbes exploit the disrupted environment through elevated metabolic activity, including fermentation, anaerobic respiration, and nutrient scavenging. This interaction likely contributes to sustained inflammation, impaired healing, and further tissue degradation in SCTLD-affected corals.

4.1.2. *Virus sequence analyses*

The large volume of sequencing data produced per coral colony in this study has allowed for a deep investigation into virus community diversity and dynamics in the context of SCTLD. The goal for Year 2 was to continue testing the current working hypothesis that viruses contribute to SCTLD etiology through community-level dynamics in the form of opportunistic infections (Klinges et al., 2024; Vega Thurber & Correa, 2023; Veglia, 2023). To do this successfully, we set out to 1. Develop a virus classification framework that integrates multiple bioinformatic tools to improve classification success and confidence; 2. Determine the core viral groups per species and perform a preliminary assessment of phyllosymbiosis as a first step toward understanding whether resident viral communities may influence coral disease susceptibility or severity; 3. Assess virus community dynamics across time and health by measuring and comparing alpha and beta diversity differences; and 4. Identify viral taxa with upregulated transcripts in disease margin tissues and characterize the genes they encode.

Inferring viral pathogens from ‘omics data remains a major challenge in environmental virology, particularly in understudied systems such as Florida’s coral reefs, where baseline data for most coral-

associated viral lineages are lacking (Vega Thurber et al., 2025). In addition, public databases are incomplete and biased toward viruses from well-studied hosts (e.g., humans, plants, and model organisms), leaving reef-associated viral diversity significantly underrepresented. As a result, assigning confident taxonomic classifications to putative viral sequences (especially at lower taxonomic ranks like order, family, genus, species) remains difficult. This complicates efforts to determine whether the same virus has been detected across studies or to track consistent signatures of viral community change. To address these limitations and improve classification success and confidence, we developed a standardized bioinformatic pipeline tailored to deal with the novelty of coral-associated viromes. The pipeline integrates multiple detection and classification tools, enforces consistent annotation practices, and mitigates the impact of inconsistent or outdated taxonomies in public repositories. In its initial application in this study, the pipeline provided taxonomic classifications for 632,670 putative viral sequences from both DNA and RNA sequencing libraries. Importantly, it applied a conservative approach to lower-rank assignments below “class”, allowing for the future implementation of lineage-specific rules for species-level designation. This framework is designed to be reproducible and adaptable as coral virome references improve, forming the basis for a coral reef-specific virus classification tool that will facilitate more effective virus studies in Florida.

To identify disease-specific coral virus communities or a potential viral pathogen, it is first necessary to establish which virus taxa are consistently present in individuals and which may appear sporadically. This study’s sampling design, which includes samples collected across multiple time points from the same coral individuals, provides a unique opportunity to address this knowledge gap in the Dry Tortugas by characterizing the core virome specific to each species sampled. Prior to this study, the core viromes (inclusive of both DNA and RNA eukaryotic- and prokaryotic-infecting virus groups) had not been characterized for any of the four coral species. Focusing on high-confidence viral sequences, we identified 17 virus classes (officially recognized by the International Committee on Taxonomy of Viruses) as core components in at least one of the four coral species’ viromes (**Table 6**). These included ten DNA virus classes and seven RNA virus classes, each present in at least 95% of samples within a given species (**Table 6**). Several core classes identified in this study, which include the eukaryotic virus groups Megaviricetes (DNA), Pokkesviricetes (DNA), Herviviricetes (DNA), and Revtraviricetes (RNA), as well as the prokaryotic virus group Caudoviricetes, have previously been reported as core coral virome taxa and were identified to be a core group for each species (Ambalavanan et al., 2021; Thurber et al., 2017). We also identified six additional DNA virus classes and six RNA virus classes that have not yet been recognized as part of the core coral virome. Two notable core RNA virus classes for all species are Stelpaviricetes and Alsuviricetes. Stelpaviricetes and Alsuviricetes contain the orders Patatavirales and Tymovirales, respectively, both of which are comprised of positive-sense single-stranded RNA viruses with filamentous morphologies matching virus-like particles identified in TEM images by Work et al. (2021) and Howe-Kerr et al. (2023) in Florida corals. Hierarchical clustering analysis revealed evidence consistent with phylosymbiosis, wherein more closely related coral species shared more similar core virome sequence diversity at the virus class level (**Figure 6**). Given the current hypothesis proposing a community-wide contribution to SCTLTD, this observation provides a foundation for future investigations aimed at elucidating the role of core virome diversity in shaping coral holobiont responses to SCTLTD.

Assessment of virus community variability across sampling time points and colony health status revealed notable patterns. When all coral species were pooled, we observed a significant difference in virus alpha diversity across the pre-outbreak (T0), early outbreak (T1), and late outbreak (T2) periods (Kruskal-Wallis chi-squared = 20.604, df = 2, $p < 0.01$) (**Figure 7**). However, because time point and health status are inherently confounded in our sampling design, this observed difference may reflect changes in coral health or stress states, similar to those documented under other stress conditions (Grupstra et al., 2022; Howe-Kerr, Grupstra, et al., 2023). Beta diversity analyses (PERMANOVA) indicated significant effects of host species, sampling time point, and health status on virus community composition, with host species explaining the largest proportion of variance (**Figure 9**). The

statistically significant contributions of sampling timepoint and health status warrant careful interpretation given the confounding nature of these factors. Nonetheless, collectively these results suggest a potential impact of disease on virus community dynamics, consistent with a scenario of holobiont dysbiosis (Egan & Gardiner, 2016). To understand how coral viromes respond to disease, future studies should first assess how viral communities vary over time under non-diseased conditions, providing a baseline for normal community dynamics.

Next, we looked to determine the virus taxa with upregulated transcripts specifically within tissue sampled at the disease margin on a diseased coral across species. DESeq2 analyses revealed a total of 6,587 sequences that were differentially more abundant within disease margin tissue across all species. These differentially abundant sequences represented 32 different virus sequence groups of which 19 represented recognized virus classes (**Figure 9**). All core virus classes besides class Maveriviricetes (viruses that infect giant viruses), including the two groups of filamentous RNA viruses (Alsuviricetes and Stelpaviricetes), had increased transcript abundance in disease margin samples (Figure V4). Three additional classes identified with increased transcript abundance included Monjiviricetes (negative-strand RNA viruses), Arfiviricetes (single-stranded circular DNA viruses – ssDNA viruses have been identified in diseased coral tissues previously (Soffer et al., 2014)), and Chrymotiviricetes (double-stranded RNA viruses). The specific target hosts of these viruses within coral holobionts remain unclear. As these virus groups were not identified as core classes in any of the studied coral species, their increased transcription within disease margin tissues could reflect several possibilities: i. Some or all these virus groups may already be present in healthy coral tissues at abundances below detection thresholds, becoming detectable only upon increased viral productivity in diseased tissues; ii. Some or all these virus groups might be directly associated with a cellular pathogen, becoming detectable due to pathogen invasion or proliferation within coral tissues; or (iii) Some or all these virus groups could represent pathogenic agents directly responsible for SCTLD symptoms. Lacking substantial baseline information for these groups impedes our ability to make strong declarations regarding their role in SCTLD, and future research should focus on producing foundational information regarding coral virus communities outside the context of disease.

Finally, we aimed to examine the functional annotation of differentially expressed transcripts to identify genes upregulated in disease margin tissues (**Table 7**). While we had a low annotation rate for the 6,587 sequences (likely driven by the novelty of the genomic information recovered), one observation is the presence of genes related to reverse transcription indicating the increased activity of retroviruses (class Retraviricetes) in disease margin tissues. Increased reverse transcriptase activity could indicate retroviral integration events into host genomes, potentially disrupting host genes critical for immune responses or activating the transcription of neighboring genes typically silenced under normal conditions (Jern & Coffin, 2008) Further investigation is needed to determine how retroviruses might influence gene expression patterns in coral or their associated symbionts. Additionally, a notable observation is the elevated expression of viral structural protein genes within disease margin tissues, suggesting that these viral infections are at an advanced replication stage involving active particle production and eventual viral release via lysis or budding. Specifically, lysis of coral or dinoflagellate host cells could directly contribute to previously observed cytopathic effects in symbiont cells and coral tissues (Landsberg et al., 2020) Additional baseline data on the temporal variability of virus gene expression landscapes are needed to identify reliable signatures of coral disease or to reveal mechanisms potentially driving disease progression.

4.1.3. Histology

This study is unique in itself as corals are shown to recover from disease which is unusual in the case of SCTLD, but also provides for a broader reach. Samples collected for histology in this project were analyzed with the same methodology as samples across two (potentially three) other diseases and three other locations. This approach allows researchers to visually assess tissue parameters without any internal bias for their own data and compare to other diseases across other regions. This has the

potential to create a bank of knowledge across users for a faster verification of disease state. As the body of data grows, so will the verification. Moving these data and the methodology behind it allows for streamlining these techniques across labs.

Moving forward, this technique could be applied to outplanting and *ex situ* facilities, marking what new “healthy” colonies look like, how species differ from one another, and potentially in the future, how tissue markers change seasonally and over time. A public repository allows for collaboration across disciplines and networks to create more holistic studies.

4.1.4. Transcriptomics

Our species-independent analyses of markers of SCTLD resistance and disease response provided mixed results in identifying unifying potential diagnostic markers. While we were able to identify over 7000 shared single copy orthologs across our four species, analysis of collated gene expression of these orthologs failed to identify any reliable markers of disease resistance, neither predictive nor epidemic. This result was consistent regardless of whether or not highly-susceptible species *Colpophyllia natans* was included in our analyses. The presence of predictive resistance markers, i.e. genetic variation which can be used to predict variation in response of host corals to a particular stressor, has been confirmed previously using species-specific approaches (Jin et al., 2016; Kelley et al., 2021; Vollmer et al., 2023). Furthermore, clear transcriptomic differences have also been observed between resistant and susceptible individuals during other disease outbreaks, again using species-specific approaches (Libro & Vollmer, 2016; Wright et al., 2017), including our own previous work in year one of this project (Klinges et al., 2024). Our approach here is highly unique; no studies to date have applied multi-species approaches to identify species independent predictive markers, and only two have taken a multi-species approach to consider transcriptomic differences between resistant and susceptible individuals during disease outbreaks (Beavers et al., 2023; MacKnight et al., 2022). These studies have similarly found few or no species-independent markers of resistance (Beavers et al., 2023; MacKnight et al., 2022). It is possible that mechanisms of resistance are highly species specific, as indicated by previous studies (Beavers et al., 2023; MacKnight et al., 2022). However, all of these studies have been limited by sample size and existing genetic resources. More robust genomic sequencing of species of interest will certainly improve identification of orthologs and statistical power for identifying any potential species independent markers.

While we did not identify clear predictive markers of disease, we were successful in identifying species-independent markers of disease response by comparing active lesion tissue to apparently healthy tissue on the same colony. Notably, these markers showed high temporal variability, with only four shared contigs consistently differentially expressed between these two tissue types over both sampling points (June and August). Generally, more species-independent response markers were observed early in the disease outbreak, in June, suggesting stronger initial disease stages. These patterns are important when considering management practices as they suggest corals may mount strong initial responses to disease, which cannot be sustained without intervention. The waning of responses over time may be a direct contributor to mortality, suggesting intervention to bolster and maintain initial immune responses may be a helpful strategy.

This pattern of early responses was particularly true of immunological changes, wherein 57/61 differentially expressed orthologs were only responsive in June. Specifically, we see strong upregulation of a number of lectins (putative immunological receptors; Kvennefors et al., 2008; Zhou et al., 2017) and oxidative stress genes in June, coupled with down regulation of antiviral and inflammatory genes. The upregulation of lectins is indicative of increased capacity to recognize and respond to pathogens (Kvennefors et al., 2008; Zhou et al., 2017). Immunological receptors are highly diversified across cnidarian lineages (Emery et al., 2021), though our patterns observed here suggest a role for some evolutionarily conserve receptors in initiating responses to SCTLD. The paired upregulation of oxidative stress orthologs and down regulation of inflammatory orthologs may be indicative of general pathways mitigating immunopathology, or self-damage inflicted by immune

responses. Both excessive oxidative stress and inflammation can be damaging to hosts during immune responses, but production of antioxidants and anti-inflammatory compounds can mitigate these responses (Hasnain, 2018; Knight, 2000; Marshall et al., 2018).

Most notably, we observed downregulation of antiviral genes, which is surprising given the widely proposed viral causative agent hypothesis. Some of these genes, like RN216, are negative regulators of antiviral response (Evankovich et al., 2020), and others, like XERD, are indicative of viral integration into host genomes (Yeh, 2020), thus explaining their downregulation. We also observed clear upregulation of antiviral ortholog TXD12 (Hanchapola et al., 2023). Combined, our results provide clear evidence for the roles of antiviral processes in response to SCTLD, though the directionality of these responses is not abundantly clear.

In addition to patterns of gene expression indicative of resistance and response to disease we also considered species-independent patterns of association between host gene expression and other generated metrics of interest, specifically microbial/viral community composition and histological traits of interest. As expected, host gene expression was highly correlated to viral and microbial community composition, though the magnitude of effects of viral abundance on host gene expression was shocking. Viral abundance was exceptionally strongly correlated to host gene expression and had strong positive impacts on host immunity. Notably, many of the viral groups most associated with host gene expression are associated with viral phages including Streptococcus phage phiNJ2 (Tang et al., 2013), Zobellviridae (Gorodnichev et al., 2023), Streptococcus phage phiARI0746 (Abril et al., 2020), Uetakevirus (Li et al., 2022), Schitoviridae (Lokareddy et al., 2024), and Brochothrix phage BL3 (Kilcher et al., 2010). Several of these, including both Streptococcus phages, Zobellviridae, and Uetakevirus are phages of disease-causing bacteria (Abril et al., 2020; Gorodnichev et al., 2023; Li et al., 2022; Tang et al., 2013), hence the associations with immunity are likely indicative of higher abundance of these bacteria rather than the viruses themselves. We did also identify a *Pseudoalteromonas* phage (H103; Zheng et al., 2023) which was also positively associated with host immunity. *Pseudoalteromonas* bacteria are frequently associated with corals, and some may have probiotic effects which aid in SCTLD defense (Ushijima et al., 2023). Our results suggest these beneficial effects may be mediated through manipulation of host immunity.

When considering bacterial community composition, we saw similar strong associations with host gene expression generally, and immunity specifically. We were able to identify several microbial families with effects of host immunity, the most notable of which was Rhodocyclaceae, an incredibly diverse group of microbes spanning ecological niches (Oren, 2014). Nitrogen fixing bacterial family Terasakiellaceae, was also highly associated with host gene expression generally, and negatively associated with host immune processes, despite previous observations of associations with healthy host tissues (Moynihan et al., 2022) and evidence for roles in thermal adaptation (Wei et al., 2024). Finally, we also observed strong effects of putative microbial symbionts of both host and Symbiodiniaceae on host gene expression and immune function. Putative Symbiodiniaceae symbiont Oligoflexaceae (Aguirre et al., 2023), was positively associated with both immunological GO terms, and our identified immunological co-expression module. Core Symbiodiniaceae microbiome member Chromatiaceae (Lawson et al., 2018) was also highly associated with host gene expression, though no effects on host immunity were observed. Notably, we also observed negative associations between putative beneficial microbial family Endozoicomonadaceae (Pogoreutz & Ziegler, 2024), and expression of our immunological module, suggesting more complex effects of members of this family on host function than currently considered. Combined, our results highly complex, evolutionary conserved relationships between microbial abundance and host gene expression.

Finally, we did observe significant associations between both vacuolization and maximum vacuole size determined by histology and host gene expression. However, there were no clear trends of association of these processes and host immunity, nor was exocytosis (which can be important for SCTLD response) associated strongly with host immunity. These findings potentially indicate species-specific pathways controlling direct immunological functions important for SCTLD defense,

including vacuolization.

4.1.5. Immunological Analysis of samples from project C21169

Our immunological assays failed to find any notable differences in immune metrics as a result of exposure to stony coral tissue loss disease, despite gross observations of disease progression during the experimental period. Instead, we identified strong signatures of the effects of experiment duration on immune metrics, revealing potential effects of general experimental stress which may have swamped out any disease signatures. The lack of feeding and frequent water changes during the duration of the experiment most likely induced these changes, evidenced by a general decline in costly melanin and antibacterial activity, and a reciprocal spike in antioxidant protein at the start of these declines (T3). Still, visually our data shows trends towards decreased antibacterial and melanin activity during early disease progression, and heightened peroxidase activity. Improved sample size, and more careful experimentation to reduce stress would help parse out the effects of general stress from disease response to provide a better picture of host immunological responses to SCTLD.

4.1.6. Analysis of TEM images from coral sampled under projects C1E0A5 and C21169

The preliminary analysis of 819 transmission electron microscopy (TEM) images from both healthy and diseased coral tissues (collected *in situ* and *in vitro*) revealed the widespread presence of virus-like particles (filamentous and spherical) in both the coral host and endosymbiotic dinoflagellate compartments. Consistent with sequencing results and previous findings from diverse studies (reviewed in Vega Thurber et al., 2017; Vega Thurber et al., 2024), these observations confirm the high prevalence and apparent diversity of viruses across all coral health states. When statistical comparisons were possible, no significant differences in VLP detection rates were observed across health conditions, reinforcing the idea that viral presence in coral tissues is not restricted to disease. These findings underscore a key limitation of using TEM as a standalone diagnostic tool for identifying viral pathogens in corals. Morphological similarity among unrelated viral taxa, combined with the potential for pleomorphic viruses (those capable of adopting multiple capsid structures depending on environmental conditions), complicates efforts to assign taxonomy or infer pathogenicity based on particle structure alone. Moreover, given the persistent presence of VLPs in healthy tissues and the fact that many virus groups are transcriptionally upregulated in diseased samples, it remains difficult to distinguish between normal components of the resident virome and true disease-causing agents. Until more background data on coral-associated cellular components and viruses (including their morphological diversity, their within-colony spatial variability, their infection dynamics/temporal variability in particle production in non-diseased contexts) become available, TEM is best used as a complementary tool in coral disease research. While TEM cannot reliably distinguish between pathogenic and non-pathogenic viruses, it can provide valuable spatial context, including evidence of cytopathic effects, viral replication structures, virus-associated structural changes, and host cell degradation, offering insight into the functional impact of viral activity at the cellular level within coral tissues (Papke et al., 2024).

4.2. Future Steps

While integrative analyses between different ‘omics data streams have begun to identify connections between host immune response, bacterial and viral community composition and function, and disease outcomes and histology, there is still considerable work to be done with this dataset that may help identify bacterial and viral functional pathways that are critical to disease development and the corresponding response by the coral host and by the algal symbiont. Gene expression of the algal symbiont has not been assessed for these samples and may provide the missing link with histological data that suggests a role of the algal symbiont in disease histopathology. We will build on our correlation-based analyses of host gene expression, bacterial taxonomic groups, and bacterial function by grouping host genes and microbial pathways into broader functional categories (e.g., apoptosis,

anaerobic metabolism) and using multivariate models to refine key associations and identify cofactors (symbiont composition/clade, site, coral species) that influence interactions. We will integrate these associations into new random forest models to assess whether functional groups or bacterial community member identity are better predictors of disease state. Similar analyses should also be performed to integrate microbial pathways with viral sequence abundance to assess potential microbial response to viruses and phage. These results will inform targeted experiments to test whether microbial byproducts like nitric oxide or sulfide contribute to host immune activation, oxidative stress, or cell death. The interaction between broader bacterial functional categories such as sulfate and nitrate reduction and breakdown of host immunity could also be tested through coral inoculation with bacteria cultivated in nitrate- or sulfate-rich media.

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.

1. 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 SCTLD, 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.
2. Synergies between bacterial metabolic pathways and host immune response should be further explored to identify targeted treatments that block bacterial function, such as targeted antibiotic and biocides that directly influence sulfate reduction.
3. Similarities in response to SCTLD across species indicate a strong, multi-faceted immunological response during early disease stages, which wanes over time perhaps due to declining coral fitness. Interventions to mitigate this immunological waning may prevent mortality.
4. 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.
5. Our understanding of viruses and their roles in promoting coral health or contributing to diseases like SCTLD remains underdeveloped. To improve the utility of coral virome data for management decisions, future efforts should prioritize foundational research on the temporal and spatial dynamics of the apparently healthy coral virome to generate valuable baseline information. In addition, our initial findings highlight the need for research on the diversity and ecology of core virome members under varying environmental conditions, to better interpret their activity and potential impacts under biotic or abiotic stressors. This knowledge would support the identification of virome signatures associated with coral health or resilience, aiding managers, researchers, and restoration practitioners in early detection and response strategies.

5. REFERENCES CITED

- Abril, A. G., Carrera, M., Bohme, K., Barros-Velazquez, J., Canas, B., Rama, J. L. R., Villa, T. G., & Calomata, P. (2020). Characterization of Bacteriophage Peptides of Pathogenic Streptococcus by LC-ESI-MS/MS: Bacteriophage Phylogenomics and Their Relationship to Their Host. *Front Microbiol*, 11, 1241. <https://doi.org/10.3389/fmicb.2020.01241>
- Aguirre, E. G., Fine, M. J., & Kenkel, C. D. (2023). Abundance of Oligoflexales bacteria is associated with algal symbiont density, independent of thermal stress in Aiptasia anemones. *Ecol Evol*, 13(12), e10805. <https://doi.org/10.1002/ece3.10805>
- Ambalavanan, L., Iehata, S., Fletcher, R., Stevens, E. H., & Zainathan, S. C. (2021). A Review of Marine Viruses in Coral Ecosystem. *Journal of Marine Science and Engineering*, 9(7). <https://doi.org/10.3390/jmse9070711>
- 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). Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. *Sci Rep*, 6, 39734. <https://doi.org/10.1038/srep39734>
- Aylward, F. O., Moniruzzaman, M., Ha, A. D., & Koonin, E. V. (2021). A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol*, 19(10), e3001430. <https://doi.org/10.1371/journal.pbio.3001430>
- 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). 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). 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). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife*, 10. <https://doi.org/10.7554/eLife.65088>
- Bellanger, M., Figueroa, J. L., 3rd, Tiemann, L., Friesen, M. L., & Iii, R. A. W. (2024). NFixDB (Nitrogen Fixation DataBase)-a comprehensive integrated database for robust 'omics analysis of diazotrophs. *NAR Genom Bioinform*, 6(2), lqae063. <https://doi.org/10.1093/nargab/lqae063>
- Bellantuono, A. J., Dougan, K. E., Granados-Cifuentes, C., & Rodriguez-Lanetty, M. (2019). Free-living and symbiotic lifestyles of a thermotolerant coral endosymbiont display profoundly distinct transcriptomes under both stable and heat stress conditions. *Mol Ecol*, 28(24), 5265-5281. <https://doi.org/10.1111/mec.15300>
- 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., . . . Segata, N. (2023). 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>
- Bloomberg, J., Hutchinson, S., Becker, C., Meiling, S., Bloomberg, M., Brandt, M., & Apprill, A. (2025). Longitudinal Microbiomes of Coral and Near-Coral Seawater are Influenced by Disease Phase. *SSRN*. <https://doi.org/10.2139/ssrn.5200743>
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). 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/
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421. <https://doi.org/10.1186/1471-2105-10-421>
- Camargo, A. P., Roux, S., Schulz, F., Babinski, M., Xu, Y., Hu, B., Chain, P. S. G., Nayfach, S., & Kyrpides, N. C. (2024). Identification of mobile genetic elements with geNomad. *Nat Biotechnol*, 42(8), 1303-1312. <https://doi.org/10.1038/s41587-023-01953-y>
- Changsut, I., Womack, H. R., Shickle, A., Sharp, K. H., & Fuess, L. E. (2022). Variation in symbiont density is linked to changes in constitutive immunity in the facultatively symbiotic coral, *Astrangia poculata*. *Biol Lett*, 18(11), 20220273. <https://doi.org/10.1098/rsbl.2022.0273>
- Chen, S. (2023). Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta*, 2(2), e107. <https://doi.org/10.1002/imt2.107>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884-i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Chen, Y., Gonzalez-Pech, R. A., Stephens, T. G., Bhattacharya, D., & Chan, C. X. (2020). Evidence That Inconsistent Gene Prediction Can Mislead Analysis of Dinoflagellate Genomes. *J Phycol*, 56(1), 6-10. <https://doi.org/10.1111/jpy.12947>
- Chen, Y., Shah, S., Dougan, K. E., van Oppen, M. J. H., Bhattacharya, D., & Chan, C. X. (2022). Improved *Cladocopium goreau* Genome Assembly Reveals Features of a Facultative Coral Symbiont and the Complex Evolutionary History of Dinoflagellate Genes. *Microorganisms*, 10(8). <https://doi.org/10.3390/microorganisms10081662>
- 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>
- Dougan, K. E., Bellantuono, A. J., Kahlke, T., Abbriano, R. M., Chen, Y., Shah, S., Granados-Cifuentes, C., van Oppen, M. J. H., Bhattacharya, D., Suggett, D. J., Rodriguez-Lanetty, M., & Chan, C. X. (2024). Whole-genome duplication in an algal symbiont bolsters coral heat tolerance. *Sci Adv*, 10(29), eadn2218. <https://doi.org/10.1126/sciadv.adn2218>
- Egan, S., & Gardiner, M. (2016). Microbial Dysbiosis: Rethinking Disease in Marine Ecosystems. *Front Microbiol*, 7, 991. <https://doi.org/10.3389/fmicb.2016.00991>
- Emery, M. A., Dimos, B. A., & Mydlarz, L. D. (2021). Cnidarian Pattern Recognition Receptor Repertoires Reflect Both Phylogeny and Life History Traits. *Front Immunol*, 12, 689463. <https://doi.org/10.3389/fimmu.2021.689463>
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*, 20(1), 238. <https://doi.org/10.1186/s13059-019-1832-y>
- Evankovich, J., Lear, T., Baldwin, C., Chen, Y., White, V., Villandre, J., Londino, J., Liu, Y., McVerry, B., Kitsios, G. D., Mallampalli, R. K., & Chen, B. B. (2020). Toll-like Receptor 8 Stability Is Regulated by Ring Finger 216 in Response to Circulating MicroRNAs. *Am J Respir Cell Mol Biol*, 62(2), 157-167. <https://doi.org/10.1165/rcmb.2018-0373OC>
- Evans, J. S., Paul, V. J., Ushijima, B., Pitts, K. A., & Kellogg, C. A. (2023). Investigating microbial size classes associated with the transmission of stony coral tissue loss disease (SCTLD). *PeerJ*, 11, e15836. <https://doi.org/10.7717/peerj.15836>
- Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J. G., Haendiges, J., Haft, D. H., Hoffmann, M., Pettengill, J. B., Prasad, A. B., Tillman, G. E., Tyson, G. H., & Klimke, W. (2021). AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci Rep*, 11(1), 12728. <https://doi.org/10.1038/s41598-021-91456-0>
- Figuerola III, J. L., Dhungel, E., Bellanger, M., Brouwer, C. R., & White Iii, R. A. (2024). MetaCerberus: distributed highly parallelized HMM-based processing for robust functional annotation across the tree of life. *Bioinformatics*, 40(3). <https://doi.org/10.1093/bioinformatics/btae119>

- Finke, J. F., Kellogg, C. T. E., & Suttle, C. A. (2023). 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>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150-3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Fuess, L. E., den Haan, S., Ling, F., Weber, J. N., Steinel, N. C., & Bolnick, D. I. (2021). Immune Gene Expression Covaries with Gut Microbiome Composition in Stickleback. *mBio*, 12(3). <https://doi.org/10.1128/mBio.00145-21>
- Fuess, L. E., Klings, G., Veglia, A. J., Rossin, A., Studivan, M., & Ushijima, B. (2025). *An Integrative 'Omic Approach Leveraging Historical SCTLD Data to Identify Predictive Markers of SCTLD Resistance*.
- Fuess, L. E., Pinzomicronn, C. J., Weil, E., & Mydlarz, L. D. (2016). 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>
- Gonzalez-Pech, R. A., Stephens, T. G., Chen, Y., Mohamed, A. R., Cheng, Y., Shah, S., Dougan, K. E., Fortuin, M. D. A., Lagorce, R., Burt, D. W., Bhattacharya, D., Ragan, M. A., & Chan, C. X. (2021). Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus Symbiodinium. *BMC Biol*, 19(1), 73. <https://doi.org/10.1186/s12915-021-00994-6>
- Gorodnichev, R. B., Kornienko, M. A., Malakhova, M. V., Bespiatykh, D. A., Manuvera, V. A., Selezneva, O. V., Veselovsky, V. A., Bagrov, D. V., Zaychikova, M. V., Osnach, V. A., Shabalina, A. V., Goloshchapov, O. V., Bespyatykh, J. A., Dolgova, A. S., & Shitikov, E. A. (2023). Isolation and Characterization of the First Zobellviridae Family Bacteriophage Infecting Klebsiella pneumoniae. *Int J Mol Sci*, 24(4). <https://doi.org/10.3390/ijms24044038>
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., . . . Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, 29(7), 644-652. <https://doi.org/10.1038/nbt.1883>
- Grazziotin, A. L., Koonin, E. V., & Kristensen, D. M. (2017). Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res*, 45(D1), D491-D498. <https://doi.org/10.1093/nar/gkw975>
- Gruber-Vodicka, H. R., Seah, B. K. B., & Pruesse, E. (2020). phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems*, 5(5). <https://doi.org/10.1128/mSystems.00920-20>
- Grupstra, C. G. B., Howe-Kerr, L. I., Veglia, A. J., Bryant, R. L., Coy, S. R., Blackwelder, P. L., & Correa, A. M. S. (2022). Thermal stress triggers productive viral infection of a key coral reef symbiont. *ISME J*, 16(5), 1430-1441. <https://doi.org/10.1038/s41396-022-01194-y>
- Guo, J., Bolduc, B., Zayed, A. A., Varsani, A., Dominguez-Huerta, G., Delmont, T. O., Pratama, A. A., Gazitua, M. C., Vik, D., Sullivan, M. B., & Roux, S. (2021). VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, 9(1), 37. <https://doi.org/10.1186/s40168-020-00990-y>
- Haas, B. J. *TransDecoder*. <https://github.com/TransDecoder/TransDecoder>
- Hanchapola, H., Liyanage, D. S., Omeka, W. K. M., Lim, C., Kim, G., Jeong, T., & Lee, J. (2023). Thioredoxin domain-containing protein 12 (TXNDC12) in red spotted grouper (*Epinephelus akaara*): Molecular characteristics, disulfide reductase activities, and immune responses. *Fish Shellfish Immunol*, 132, 108449. <https://doi.org/10.1016/j.fsi.2022.11.037>
- Hasnain, S. Z. (2018). Endoplasmic reticulum and oxidative stress in immunopathology: understanding the crosstalk between cellular stress and inflammation. *Clin Transl Immunology*, 7(7), e1035. <https://doi.org/10.1002/cti2.1035>

- Hauptfeld, E., Pappas, N., van Iwaarden, S., Snoek, B. L., Aldas-Vargas, A., Dutilh, B. E., & von Meijenfeldt, F. A. B. (2024). Integrating taxonomic signals from MAGs and contigs improves read annotation and taxonomic profiling of metagenomes. *Nat Commun*, 15(1), 3373. <https://doi.org/10.1038/s41467-024-47155-1>
- Hoffman, G. E., & Roussos, P. (2021). Dream: powerful differential expression analysis for repeated measures designs. *Bioinformatics*, 37(2), 192-201. <https://doi.org/10.1093/bioinformatics/btaa687>
- Hofmeyr, S., Egan, R., Georganas, E., Copeland, A. C., Riley, R., Clum, A., Eloë-Fadrosch, E., Roux, S., Goltsman, E., Buluc, A., Rokhsar, D., Olikar, L., & Yelick, K. (2020). Terabase-scale metagenome coassembly with MetaHipMer. *Sci Rep*, 10(1), 10689. <https://doi.org/10.1038/s41598-020-67416-5>
- Howe-Kerr, L. I., Grupstra, C. G. B., Rabbitt, K. M., Conetta, D., Coy, S. R., Klings, J. G., Maher, R. L., McConnell, K. M., Meiling, S. S., Messyas, A., Schmeltzer, E. R., Seabrook, S., Sims, J. A., Veglia, A. J., Thurber, A. R., Thurber, R. L. V., & Correa, A. M. S. (2023). Viruses of a key coral symbiont exhibit temperature-driven productivity across a reefscape. *ISME Commun*, 3(1), 27. <https://doi.org/10.1038/s43705-023-00227-7>
- Howe-Kerr, L. I., Knochel, A. M., Meyer, M. D., Sims, J. A., Karrick, C. E., Grupstra, C. G. B., Veglia, A. J., Thurber, A. R., Vega Thurber, R. L., & Correa, A. M. S. (2023). Filamentous virus-like particles are present in coral dinoflagellates across genera and ocean basins. *ISME J*, 17(12), 2389-2402. <https://doi.org/10.1038/s41396-023-01526-6>
- Huntley, N., Brandt, M. E., Becker, C. C., Miller, C. A., Meiling, S. S., Correa, A. M. S., Holstein, D. M., Muller, E. M., Mydlarz, L. D., Smith, T. B., & Apprill, A. (2022). Experimental transmission of Stony Coral Tissue Loss Disease results in differential microbial responses within coral mucus and tissue. *ISME Commun*, 2(1), 46. <https://doi.org/10.1038/s43705-022-00126-3>
- Jern, P., & Coffin, J. M. (2008). Effects of retroviruses on host genome function. *Annu Rev Genet*, 42, 709-732. <https://doi.org/10.1146/annurev.genet.42.110807.091501>
- Jin, Y. K., Lundgren, P., Lutz, A., Raina, J. B., Howells, E. J., Paley, A. S., Willis, B. L., & van Oppen, M. J. (2016). Genetic markers for antioxidant capacity in a reef-building coral. *Sci Adv*, 2(5), e1500842. <https://doi.org/10.1126/sciadv.1500842>
- Jørgensen, B. B. (1982). Mineralization of organic matter in the sea bed—the role of sulphate reduction. *Nature*, 296(5858), 643-645. <https://doi.org/10.1038/296643a0>
- Kelley, E. R., Sleith, R. S., Matz, M. V., & Wright, R. M. (2021). Gene expression associated with disease resistance and long-term growth in a reef-building coral. *R Soc Open Sci*, 8(4), 210113. <https://doi.org/10.1098/rsos.210113>
- Kieft, K., Adams, A., Salamzade, R., Kalan, L., & Anantharaman, K. (2022). vRhyme enables binning of viral genomes from metagenomes. *Nucleic Acids Res*, 50(14), e83. <https://doi.org/10.1093/nar/gkac341>
- Kilcher, S., Loessner, M. J., & Klumpp, J. (2010). Brochothrix thermosphacta bacteriophages feature heterogeneous and highly mosaic genomes and utilize unique prophage insertion sites. *J Bacteriol*, 192(20), 5441-5453. <https://doi.org/10.1128/JB.00709-10>
- Klimes, G., Fuess, L. E., Veglia, A. J., Rossin, A., & Holstein, D. (2024). *Tracking Bacterial, Viral, and Gene Expression Shifts throughout SCTL Development in the Dry Tortugas*.
- Knight, J. A. (2000). Review: Free radicals, antioxidants, and the immune system. *Ann Clin Lab Sci*, 30(2), 145-158.
- Kvennefors, E. C., Leggat, W., Hoegh-Guldberg, O., Degnan, B. M., & Barnes, A. C. (2008). An ancient and variable mannose-binding lectin from the coral *Acropora millepora* binds both pathogens and symbionts. *Dev Comp Immunol*, 32(12), 1582-1592. <https://doi.org/10.1016/j.dci.2008.05.010>
- 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>

- Langfelder, P., & Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9, 559. <https://doi.org/10.1186/1471-2105-9-559>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9(4), 357-359. <https://doi.org/10.1038/nmeth.1923>
- Lawson, C. A., Raina, J. B., Kahlke, T., Seymour, J. R., & Suggett, D. J. (2018). Defining the core microbiome of the symbiotic dinoflagellate, Symbiodinium. *Environ Microbiol Rep*, 10(1), 7-11. <https://doi.org/10.1111/1758-2229.12599>
- Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12, 323. <https://doi.org/10.1186/1471-2105-12-323>
- Li, D., Luo, R., Liu, C. M., Leung, C. M., Ting, H. F., Sadakane, K., Yamashita, H., & Lam, T. W. (2016). 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). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, 37(23), 4572-4574. <https://doi.org/10.1093/bioinformatics/btab705>
- Li, P., Yong, S., Zhou, X., & Shen, J. (2022). Characterization of a New Temperate Escherichia coli Phage vB_EcoP_ZX5 and Its Regulatory Protein. *Pathogens*, 11(12). <https://doi.org/10.3390/pathogens11121445>
- Libro, S., & Vollmer, S. V. (2016). Genetic Signature of Resistance to White Band Disease in the Caribbean Staghorn Coral *Acropora cervicornis*. *PLoS One*, 11(1), e0146636. <https://doi.org/10.1371/journal.pone.0146636>
- Lokareddy, R. K., Hou, C. D., Forti, F., Iglesias, S. M., Li, F., Pavlenok, M., Horner, D. S., Niederweis, M., Briani, F., & Cingolani, G. (2024). Integrative structural analysis of Pseudomonas phage DEV reveals a genome ejection motor. *Nat Commun*, 15(1), 8482. <https://doi.org/10.1038/s41467-024-52752-1>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, 15(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
- MacKnight, N. J., Dimos, B. A., Beavers, K. M., Muller, E. M., Brandt, M. E., & Mydlarz, L. D. (2022). Disease resistance in coral is mediated by distinct adaptive and plastic gene expression profiles. *Sci Adv*, 8(39), eabo6153. <https://doi.org/10.1126/sciadv.abo6153>
- Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Curr Protoc*, 1(12), e323. <https://doi.org/10.1002/cpz1.323>
- Marshall, J. S., Warrington, R., Watson, W., & Kim, H. L. (2018). An introduction to immunology and immunopathology. *Allergy Asthma Clin Immunol*, 14(Suppl 2), 49. <https://doi.org/10.1186/s13223-018-0278-1>
- Martinez Arbizu, P. (2020). *pairwiseAdonis: Pairwise multilevel comparison using adonis*. In (Version R package version 0.4)
- 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>
- Messer, L. F., Bourne, D. G., Robbins, S. J., Clay, M., Bell, S. C., McIlroy, S. J., & Tyson, G. W. (2024). A genome-centric view of the role of the *Acropora kenti* microbiome in coral health and resilience. *Nat Commun*, 15(1), 2902. <https://doi.org/10.1038/s41467-024-46905-5>
- Moynihán, M. A., Goodkin, N. F., Morgan, K. M., Kho, P. Y. Y., Lopes Dos Santos, A., Lauro, F. M., Baker, D. M., & Martin, P. (2022). Coral-associated nitrogen fixation rates and diazotrophic diversity on a nutrient-replete equatorial reef. *ISME J*, 16(1), 233-246. <https://doi.org/10.1038/s41396-021-01054-1>
- Mydlarz, L. D., & Harvell, C. D. (2007). Peroxidase activity and inducibility in the sea fan coral exposed to a fungal pathogen. *Comp Biochem Physiol A Mol Integr Physiol*, 146(1), 54-62. <https://doi.org/10.1016/j.cbpa.2006.09.005>

- Mydlarz, L. D., & Palmer, C. V. (2011). The presence of multiple phenoloxidases in Caribbean reef-building corals. *Comp Biochem Physiol A Mol Integr Physiol*, 159(4), 372-378. <https://doi.org/10.1016/j.cbpa.2011.03.029>
- Nayfach, S., Camargo, A. P., Schulz, F., Elie-Fadrosh, E., Roux, S., & Kyrpides, N. C. (2021). 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>
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Res*, 27(5), 824-834. <https://doi.org/10.1101/gr.213959.116>
- Oksanen, J., Simpson, G., Blanchet, F., Kindt, R., Legendre, P., Minchin, P., O'Hara, R., Solymos, P., Stevens, M., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., . . . Weedon, J. (2025). *vegan: Community Ecology Package*. In (Version R package version 2.8-0) <https://vegandevs.github.io/vegan/>
- Oren, A. (2014). The Family Rhodocyclaceae. In E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, & F. Thompson (Eds.), *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria* (pp. 975-998). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30197-1_292
- Parkinson, J. E., Baumgarten, S., Michell, C. T., Baums, I. B., LaJeunesse, T. C., & Voolstra, C. R. (2016). 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>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, 14(4), 417-419. <https://doi.org/10.1038/nmeth.4197>
- Philippot, L., & Hojberg, O. (1999). Dissimilatory nitrate reductases in bacteria. *Biochim Biophys Acta*, 1446(1-2), 1-23. [https://doi.org/10.1016/s0167-4781\(99\)00072-x](https://doi.org/10.1016/s0167-4781(99)00072-x)
- Pogoreutz, C., & Ziegler, M. (2024). Frenemies on the reef? Resolving the coral-Endozoicomonas association. *Trends Microbiol*, 32(5), 422-434. <https://doi.org/10.1016/j.tim.2023.11.006>
- Reich, H. G., Kitchen, S. A., Stankiewicz, K. H., Devlin-Durante, M., Fogarty, N. D., & Baums, I. B. (2021). Genomic variation of an endosymbiotic dinoflagellate (Symbiodinium 'fitti') among closely related coral hosts. *Mol Ecol*, 30(14), 3500-3514. <https://doi.org/10.1111/mec.15952>
- 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>
- Shen, W., Sipos, B., & Zhao, L. (2024). SeqKit2: A Swiss army knife for sequence and alignment processing. *Imeta*, 3(3), e191. <https://doi.org/10.1002/imt2.191>
- Shoguchi, E., Beedesse, G., Hisata, K., Tada, I., Narisoko, H., Satoh, N., Kawachi, M., & Shinzato, C. (2021). A New Dinoflagellate Genome Illuminates a Conserved Gene Cluster Involved in Sunscreen Biosynthesis. *Genome Biol Evol*, 13(2). <https://doi.org/10.1093/gbe/evaa235>
- Shoguchi, E., Beedesse, G., Tada, I., Hisata, K., Kawashima, T., Takeuchi, T., Arakaki, N., Fujie, M., Koyanagi, R., Roy, M. C., Kawachi, M., Hidaka, M., Satoh, N., & Shinzato, C. (2018). Two divergent Symbiodinium genomes reveal conservation of a gene cluster for sunscreen biosynthesis and recently lost genes. *BMC Genomics*, 19(1), 458. <https://doi.org/10.1186/s12864-018-4857-9>
- Shoguchi, E., Shinzato, C., Kawashima, T., Gyoja, F., Mungpakdee, S., Koyanagi, R., Takeuchi, T., Hisata, K., Tanaka, M., Fujiwara, M., Hamada, M., Seidi, A., Fujie, M., Usami, T., Goto, H., Yamasaki, S., Arakaki, N., Suzuki, Y., Sugano, S., . . . Satoh, N. (2013). Draft assembly of the Symbiodinium minutum nuclear genome reveals dinoflagellate gene structure. *Curr Biol*, 23(15), 1399-1408. <https://doi.org/10.1016/j.cub.2013.05.062>
- Soffer, N., Brandt, M. E., Correa, A. M., Smith, T. B., & Thurber, R. V. (2014). Potential role of viruses in white plague coral disease. *ISME J*, 8(2), 271-283. <https://doi.org/10.1038/ismej.2013.137>
- Stankiewicz, K. H., Guiglielmoni, N., Kitchen, S. A., Flot, J. F., Barott, K. L., Davies, S. W., Finnerty, J. R., Grace, S. P., Kaufman, L. S., Putnam, H. M., Rotjan, R. D., Sharp, K. H., Peters, E. C., &

- Baums, I. B. (2025). Genomic comparison of the temperate coral *Astrangia poculata* with tropical corals yields insights into winter quiescence, innate immunity, and sexual reproduction. *G3 (Bethesda)*, 15(4). <https://doi.org/10.1093/g3journal/jkaf033>
- Su, S., & Yajima, M. (2015). *R2jags: Using R to Run "JAGS"*. In <https://CRAN.R-project.org/package=R2jags>
- Tang, F., Bossers, A., Harders, F., Lu, C., & Smith, H. (2013). Comparative genomic analysis of twelve *Streptococcus suis* (pro)phages. *Genomics*, 101(6), 336-344. <https://doi.org/10.1016/j.ygeno.2013.04.005>
- Terzian, P., Olo Ndela, E., Galiez, C., Lossouarn, J., Perez Bucio, R. E., Mom, R., Toussaint, A., Petit, M. A., & Enault, F. (2021). PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genom Bioinform*, 3(3), lqab067. <https://doi.org/10.1093/nargab/lqab067>
- Thurber, R. V., Payet, J. P., Thurber, A. R., & Correa, A. M. (2017). 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>
- Ushijima, B., Gunasekera, S. P., Meyer, J. L., Tittl, J., Pitts, K. A., Thompson, S., Sneed, J. M., Ding, Y., Chen, M., Jay Houk, L., Aeby, G. S., Hase, C. C., & Paul, V. J. (2023). Chemical and genomic characterization of a potential probiotic treatment for stony coral tissue loss disease. *Commun Biol*, 6(1), 248. <https://doi.org/10.1038/s42003-023-04590-y>
- Ushijima, B., Meyer, J. L., Thompson, S., Pitts, K., Marusich, M. F., Tittl, J., Weatherup, E., Reu, J., Wetzell, R., Aeby, G. S., Hase, C. C., & Paul, V. J. (2020). Disease Diagnostics and Potential Coinfections by *Vibrio coralliilyticus* During an Ongoing Coral Disease Outbreak in Florida. *Front Microbiol*, 11, 569354. <https://doi.org/10.3389/fmicb.2020.569354>
- 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* (Final Report, Issue.
- Vega Thurber, R. L., Silva, D., Speare, L., Croquer, A., Veglia, A. J., Alvarez-Filip, L., Zaneveld, J. R., Muller, E. M., & Correa, A. M. S. (2025). Coral Disease: Direct and Indirect Agents, Mechanisms of Disease, and Innovations for Increasing Resistance and Resilience. *Ann Rev Mar Sci*, 17(1), 227-255. <https://doi.org/10.1146/annurev-marine-011123-102337>
- Veglia, A. J. (2023). *Detecting and interpreting viral dynamics in marine invertebrate holobionts* [Rice University]. <https://hdl.handle.net/1911/115214>
- Vollmer, S. V., Selwyn, J. D., Despard, B. A., & Roesel, C. L. (2023). Genomic signatures of disease resistance in endangered staghorn corals. *Science*, 381(6665), 1451-1454. <https://doi.org/10.1126/science.adi3601>
- Vosloo, S., Huo, L., Anderson, C. L., Dai, Z., Sevillano, M., & Pinto, A. (2021). Evaluating de Novo Assembly and Binning Strategies for Time Series Drinking Water Metagenomes. *Microbiol Spectr*, 9(3), e0143421. <https://doi.org/10.1128/Spectrum.01434-21>
- Walker, P.J., Siddell, S.G., Lefkowitz, E.J., Mushegian, A.R., Adriaenssens, E.M., Alfenas-Zerbini, P., Dempsey, D.M., Dutilh, B.E., García, M.L., Curtis Hendrickson, R., Junglen, S., Krupovic, M., Kuhn, J.H., Lambert, A.J., Łobocka, M., Oksanen, H.M., Orton, R.J., Robertson, D.L., Rubino, L., Sabanadzovic, S., Simmonds, P., Smith, D.B., Suzuki, N., Van Doorslaer, K., Vandamme, A.M., Varsani, A., Zerbini, F.M. (2022) [Recent changes to virus taxonomy ratified by the International Committee on Taxonomy of Viruses](https://doi.org/10.1007/s00705-022-05516-5). Arch Virol. 2022 Aug 23. doi: 10.1007/s00705-022-05516-5. Epub ahead of print. PMID: 35999326.
- Wei, Y., Chen, B., Yu, K., Liao, Z., Yu, X., Qin, Z., Bao, Z., Xu, L., & Wang, Y. (2024). Evolutionary radiation and microbial community dynamics shape the thermal tolerance of Fungiidae in the southern South China Sea. *Microbiol Spectr*, 12(2), e0243623. <https://doi.org/10.1128/spectrum.02436-23>
- Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol*, 20(1), 257. <https://doi.org/10.1186/s13059-019-1891-0>

- Wood-Charlson, E. M., Hollingsworth, L. L., Krupp, D. A., & Weis, V. M. (2006). Lectin/glycan interactions play a role in recognition in a coral/dinoflagellate symbiosis. *Cell Microbiol*, 8(12), 1985-1993. <https://doi.org/10.1111/j.1462-5822.2006.00765.x>
- 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). 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>
- Yeh, T. Y. (2020). XerD-dependent integration of a novel filamentous phage Cf2 into the *Xanthomonas citri* genome. *Virology*, 548, 160-167. <https://doi.org/10.1016/j.virol.2020.06.010>
- Zheng, K., Dong, Y., Liang, Y., Liu, Y., Zhang, X., Zhang, W., Wang, Z., Shao, H., Sung, Y. Y., Mok, W. J., Wong, L. L., McMinn, A., & Wang, M. (2023). Genomic diversity and ecological distribution of marine *Pseudoalteromonas* phages. *Mar Life Sci Technol*, 5(2), 271-285. <https://doi.org/10.1007/s42995-022-00160-z>
- Zheng, K., Sun, J., Liang, Y., Kong, L., Paez-Espino, D., McMinn, A., & Wang, M. (2025). VITAP: a high precision tool for DNA and RNA viral classification based on meta-omic data. *Nat Commun*, 16(1), 2226. <https://doi.org/10.1038/s41467-025-57500-7>
- Zhou, Z., Yu, X., Tang, J., Zhu, Y., Chen, G., Guo, L., & Huang, B. (2017). Dual recognition activity of a rhamnose-binding lectin to pathogenic bacteria and zooxanthellae in stony coral *Pocillopora damicornis*. *Dev Comp Immunol*, 70, 88-93. <https://doi.org/10.1016/j.dci.2017.01.009>