Meta-transcriptomics to determine if and how viruses are involved in SCTLD infection status and/or disease susceptibility

Meta-transcriptomics to determine if and how viruses are involved in SCTLD infection status and/or disease susceptibility

Final Report
Prepared By:

Dr. Rebecca Vega Thurber, Dr. Adrienne M.S. Correa, Dr. Samantha Coy, Dr. Alex Veglia, Eddie Fuques, Clark Hamor

Oregon State University, Rice University

08/31/2023

Completed in Fulfillment of C0633A – DEP 3753 ORCP CPR for
Florida Department of Environmental Protection
Coral Protection and Restoration Program
1277 N.E. 79th Street Causeway
Miami, FL 33138

This Final Report should be cited as follows:
Vega Thurber, R.L. and 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. Florida DEP. Miami, FL. 20 pp.

This report was prepared for the Florida Department of Environmental Protection, Office of Resilience and Coastal Protection by Oregon State University and Rice University. Funding was provided by the Florida Department of Environmental Protection Award No. BA002F. The views, statements, findings, conclusions, and recommendations expressed herein are those of the authors and do not necessarily reflect the views of the State of Florida or any of its sub-agencies.

Management Summary

Using meta-transcriptomics, this project aimed to clarify the role of viruses in stony coral tissue loss disease (SCTLD) by evaluating links between RNA or DNA viruses and SCTLD infection status. We generated 428 meta-transcriptomic libraries (4 independent preparations of 107 samples) and acquired deep high throughput sequence data for all. We generated ITS2 rDNA amplicon libraries for all 107 sampled coral colonies to identify Symbiodiniaceae lineages and test for links among symbiont, viral diversity/abundance, and coral disease status. We developed novel informatic methods to quantitatively track what viruses are present and/or variably abundant in our samples. We also tested the application of a dsRNA immunofluorescence staining approach to detect a hallmark of RNA virus infection in corals; this approach did not yield interpretable dsRNA detections due to high background coral and autofluorescence levels; recommendations are provided for future attempts. We have now reviewed and analyzed all the data and have provided the raw data to DEP on external hard drives.

Our preliminary results indicate that 1) most metatranscriptomic preparation methods used here produced similar results in terms of the specific viral groups identified as present in coral holobionts, and 2) viruses may be involved in SCLTD, although it remains unclear whether they constitute the primary cause or a secondary infection that exacerbates the disease. Furthermore, we found 3) variation in the abundance and diversity of viral groups, as well as the dominant Symbiodiniaceae lineages, associated with SCTLD-affected corals. Despite this, some viral groups were commonly and differentially abundant in samples of diseased tissues. Specifically, we present evidence that there are some viral orders including, 2 +ssRNA viruses (Tolivirales and Picornavirales), the NCLDVs (Algavirales and Imitervirales), and a dsRNA virus (Durnavirales) enriched in diseased specimens suggesting a direct or indirect role in SCTLD. Lastly, we found evidence that while filamentous viruses are common in corals they are not associated exclusively with disease animals nor with Symbiodinaceae of specific clades.

The outcomes of this project will be incorporated into an ongoing coral disease response effort which seeks to improve understanding about the scale and susceptibility of the coral disease outbreak on Florida's Coral Reef, identify primary and secondary causes, identify management actions to remediate disease impacts, restore affected resources, and ultimately prevent future outbreaks. Importantly, **our comparative approach will allow us to provide critical advice to the DEP and other coral disease researchers about whether viruses or their abundances are associated with SCTLD.** These efforts will inform disease intervention and management efforts throughout Florida's Coral Reef. The identification of a pathogen or pathogens associated with SCTLD will also facilitate the development of diagnostic methods such as quantitative PCR primers specific to the pathogen, as well as improved intervention strategies such as targeted antibiotic or antiviral treatments.

Executive Summary

This project used previously collected samples (originally provided by Dr. Erinn Muller, Mote Marine Lab), to genomically (meta-transcriptomics) identify signatures of viral infection and determine if they are associated exclusively with stony coral tissue loss disease (SCTLD) disease signs (tissue health state) and/or colony disease susceptibility. This project also characterized the Symbiodiniaceae communities associated with SCTLD-affected and apparently healthy Florida corals in order to determine whether there is an association between Symbiodiniaceae and viral genetic diversity, and SCTLD signs. In particular, there were four primary tasks to be completed sequentially from the samples**: 1) nucleic acid isolation, 2) processing and sequencing of nucleic acids from all samples, 3) completed data curation, and 4) completed high level sequence analysis, interpretation, and dissemination of the resulting data.**

We found variation in the abundance and diversity of viral groups, as well as the dominant Symbiodiniaceae lineages, associated with SCTLD-affected corals. Despite this, some viral groups were commonly and differentially abundant in samples of diseased tissues. Specifically, four viral orders including, two +ssRNA viruses (Tolivirales and Picornavirales), the NCLDVs (Algavirales and Imitervirales), and a dsRNA virus (Durnavirales) were more abundant in diseased specimens suggesting a direct or indirect role in SCTLD. We also tested the application of a dsRNA immunofluorescence staining approach to detect a hallmark of RNA virus infection in corals; this approach did not yield interpretable dsRNA detections due to high background coral and autofluorescence levels; recommendations are provided for future attempts.

Acknowledgements

We acknowledge financial assistance provided by the State of Florida, as administered by DEP's Coral Protection and Restoration Program. We thank Thierry Work (USGS), Erinn Muller (Mote Marine Laboratory) and other members of the SCTLD research community for generating the samples upon which this research is being conducted. We thank Yasu Kiryu and Noretta Perry (Florida Fish and Wildlife Conservation Commission) for their assistance with slide preparation. We thank Rusty Ludwigsen and Joyah Watkins (Rice U.) for their contributions to dsRIF work. We thank Drs. James Spurlin and Peter Lwigale (Rice U.) for their support in the development of immunostaining techniques. We are grateful to Budi Utama and Harshavardhan Deshmukh (Rice U.) for supporting microscopy imaging and analyses, and Shane Smith (Rice U.) for his work on the RdRp gene analyses.

Table of Contents

List of Figures:

List of Tables:

Acronyms:

SCTLD – Stony coral tissue loss disease
DNA – Deoxyribonucleic acid
RNA – Ribonucleic acid
+ssRNA – Positive-strand RNA virus
qPCR – quantitative polymerase chain reaction
dsRNA – Double-stranded RNA viruses
dsRIF – dsRNA-Immunofluorescence
UofO – University of Oregon
ORF – open reading frames
RdRp – RNA-dependent RNA polymerase
OSU – Oregon State University
DEP – Florida Department of Environmental Protection
cDNA – copy DNA
H – Healthy
U – Unaffected
D - Diseased
Kbp – kilo-base pair
CHFV – Coral Holobiont-associated Filamentous Viruses
NCLDV – Nucleocytoplasmic Large DNA Viruses
ISH – In situ hybridization

1 DESCRIPTION

1.1 Project to identify viral sequences and genomes associated with apparently healthy and SCTLD afflicted corals from the Florida Keys.

Viral ecology of corals relies on two primary methods, visual characterization using electron microscopy, and genomic analysis using high throughput sequencing of DNA and RNA. The viruses visually identified by Dr. Thierry Work from SCTLD-affected Florida corals are reminiscent of filamentous RNA-based viruses typically associated with plant diseases (Work et al. 2021). Other researchers have found evidence of similar viral particles in Symbiodiniaceae (see references in Supp. Table 1 of Howe-Kerr et al. *in review*). These signatures suggest SCTLD is associated with some RNA virus yet to be identified. This project aimed to use Dr. Adrienne Correa's previously collected samples (originally provided by Dr. Erinn Muller, Mote Marine Lab), including all those used by Dr. Work, to genomically (meta-transcriptomics) identify signatures of infection and determine if they are associated exclusively with SCTLD infection or if they correspond to disease susceptibility. This project also aimed to characterize the Symbiodiniaceae communities associated with SCTLD-affected and apparently healthy Florida corals in order to determine whether there is an association between Symbiodiniaceae and viral genetic diversity, and SCTLD infection.

We hypothesized that all corals contain some viral signatures, but that SCTLD lesion sites contain more abundant or different viral signatures such as the presence of the +ssRNA viruses. These signatures could include RNA viral genomes similar to other dinoflagellate viruses or ones unique to this coral disease. By comparing coral lesions experiencing active SCTLD with apparently healthy corals, we sought to gain more definitive evidence of the association of SCTLD with particular viral and microbial communities. The identification of a pathogen or pathogens associated with active SCTLD and/or its susceptibility will facilitate the development of diagnostic methods such as qPCR primers and fluorescence in situ hybridization (FISH) probes specific to the pathogen as well as improved intervention strategies including targeted antibiotic or antiviral treatments.

Explicitly our labs conducted four forms of meta-transcriptomic preparation: 1) RNA isolation with ribosomal depletion (i.e., rRNA removal), 2) RNA isolation with small RNA enrichment, RNA isolation with polyA enrichment, and 4) dsRNA immunoprecipitation. Using rRNA removal methods provides us a mechanism to explore both DNA viral and bacterial transcripts to determine if DNA viruses and/or bacteria are directly or opportunistically involved in SCTLD. Furthermore, many viral sequences and genomes are not polyadenylated and many viruses - particularly RNA viruses- are small enough that they are inadvertently eliminated during the sequencing preparation steps. Enrichment for these small RNAs will provide a more comprehensive look at the RNA viral repertoire in corals. PolyA selected libraries and dsRNA immunoprecipitation constituted targeted strategies that might increase recovery of some known Symbiodiniaceae-infecting +ssRNA viruses (e.g., dinoRNAVs) that polyadenylate their transcripts, and many RNA viruses, respectively. We compared all of the four metatranscriptomic libraries generated per sample in this project and found strikingly consistent results across three of the four library preparation methods. Importantly, **this comparative approach allowed us to provide critical advice to the DEP and other coral disease researchers about whether viruses or their abundances are associated with this devastating**

disease. Simultaneously, this comparative work provides novel insights into which methods are optimal for evaluating coral viral infections. Along with these omics-based efforts, we explored the potential for application of immunostaining to identify signals of viral infection from coral thin sections. Although dsRNA immunofluorescence (dsRIF) staining remains a promising approach (Coy et al. 2023), challenges related to the high amount of autofluorescence in both corals and Symbiodiniaceae made detection of dsRNA (a hallmark of infections by many RNA viruses) infeasible; we offer some suggestions for potential next efforts in this line of research.

2 METHODS

2.1 Task 1 Generation and sequencing of viral meta-transcriptomes from corals collected from Middle Keys:

Dr. Correa's previously collected specimens were all extracted for total RNA and sent for processing and sequencing in late 2021 and early 2022; methods were described in a previous report (Vega Thurber et al. 2022) and therefore will not be discussed here. Dr. Correa provided aliquots of sufficient concentration and volume to the Vega Thurber laboratory (OSU). At OSU, for each of 107 samples, classified as healthy (H), Unaffected (U), or Diseased (D), a sub-sample of total RNA was prepared via ribosomal depletion, and another via small RNA enrichment. Each preparation was used to generate an RNASeq library. As only some viruses polyadenlyate their transcripts, and because we aimed to look at RNA virus genomes, the Vega Thurber lab conducted mRNA and small RNA enrichments without oligo-DT reduction to create complete viral meta-transcriptomes for comparison to the polyA enriched and dsRNA immunoprecipitation libraries prepared for each sample by the Correa lab. These rRNA removal and small RNA enrichment techniques have the advantage of providing a broad overview of the holobiont during infection. From these two library preparation approaches, we acquired information about RNA viral diversity and activity and DNA virus activity. Future students can also glean insights regarding microbial gene activity, and host gene activity and regulation, from the meta-transcriptomic sequence datasets resulting from this project. The Vega Thurber lab worked with the Genomics Core (GC3G) at University of Oregon to prepare the libraries and sequence them on the Illumina NovaSeq 3000. Meta-transcriptomic sequencing results were received in August 2022 from the UofO sequencing facility (single flow cell of Illumina NovaSeq 6000, S4 Output Paired-end sequencing, 2x150 nucleotide read length); these 214 libraries were analyzed using a viral metagenomics approach.

2.1.1 Meta-transcriptomic Sequencing Data Summary

Table 1 provides information on the coral species, specimen IDs, location of collection and RNA and DNA extraction results, and meta-transcriptomic sequencing status of each coral sample. A majority of the specimens (with a few exceptions) underwent all four methodologies (across the Vega Thurber and Correa labs) to produce high quality meta-transcriptomic libraries for analysis.

Table 1: Sample, RNA and DNA isolation and sequencing status of the coral colonies analyzed for this project.

Coral Species	**Specimen ID**	Location of Collection	RNA Conc (ng/ul)	DNA Conc (ng/ul)	Meta-transcriptomic Sequencing 4 ways	Sequence reads post QC
Colpophyllia natans	300 D	West Turtle Shoal	377.5	11.6	4 of 4	Y
Colpophyllia natans	300 U	West Turtle Shoal	388.8	8.7	4 of 4	Y
Colpophyllia natans	301 D	West Turtle Shoal	325.3	21.4	4 of 4	Y
Colpophyllia natans	301 U	West Turtle Shoal	730	42.4	4 of 4	Y
Colpophyllia natans	302 D	West Turtle Shoal	1494.7	8.3	4 of 4	Y
Colpophyllia natans	302 U	West Turtle Shoal	325.8	7.6	4 of 4	Y
Colpophyllia natans	303 D	West Turtle Shoal	258.2	7.7	4 of 4	Y
Colpophyllia natans	303 U	West Turtle Shoal	106.4	7.1	4 of 4	Y
Colpophyllia natans	305 H	West Turtle Shoal	113.1	8.3	4 of 4	Y
Colpophyllia natans	306 H	West Turtle Shoal	79.9	20.2	4 of 4	Y
Colpophyllia natans	307 H	West Turtle Shoal	37.3	9.5	4 of 4	Y
Colpophyllia natans	340 D	Lat Long 2 - southernmost	145.3	6.5	4 of 4	Y
Colpophyllia natans	340 U	Lat Long 2 - southernmost	293.7	20.4	4 of 4	Y
Colpophyllia natans	341 D	Lat Long 2 - southernmost	1031.4	13.2	4 of 4	Y
Colpophyllia natans	341 U	Lat Long 2 - southernmost	221.9	16.2	4 of 4	Y
Colpophyllia natans	342 D	Lat Long 2 - southernmost	369	11.2	4 of 4	Y
Colpophyllia natans	342 U	Lat Long 2 - southernmost	370.5	18.7	4 of 4	Y
Colpophyllia natans	343 D	Lat Long 2 - southernmost	411.3	11.3	4 of 4	Y
Colpophyllia natans	343 U	Lat Long 2 - southernmost	337.3	10.7	4 of 4	Y
Colpophyllia natans	345 H	Lat Long 2 - southernmost	184.2	7.5	4 of 4	Y
Colpophyllia natans	346 H	Lat Long 2 - southernmost	533.6	8.7	4 of 4	Y
Colpophyllia natans	347 H	Lat Long 2 - southernmost	102.8	21	4 of 4	Y

Colpophyllia natans	420 D	Lat Long 4 - northernmost	546.6	22.8	4 of 4	Y
Colpophyllia natans	420 U	Lat Long 4 - northernmost	862.5	22.4	4 of 4	Y
Colpophyllia natans	421 D	Lat Long 4 - northernmost	191	12.6	4 of 4	Y
Colpophyllia natans	421 U	Lat Long 4 - northernmost	136.4	14.4	4 of 4	Y
Colpophyllia natans	422 D	Lat Long 4 - northernmost	210.8	12.6	4 of 4	Y
Colpophyllia natans	422 U	Lat Long 4 - northernmost	248.9	15.6	4 of 4	Y
Colpophyllia natans	423 D	Lat Long 4 - northernmost	1112.2	32.4	4 of 4	Y
Colpophyllia natans	423 U	Lat Long 4 - northernmost	700.4	23.8	4 of 4	Y
Colpophyllia natans	425 H	Lat Long 4 - northernmost	551.3	29.5	4 of 4	Y
Colpophyllia natans	426 H	Lat Long 4 - northernmost	284.8	10.4	4 of 4 *	Y
Colpophyllia natans	427 H	Lat Long 4 - northernmost	232.6	31.9	4 of 4	Y
Colpophyllia natans	516 H	Western Sambo Patch 2	??	11.4	4 of 4	Y
Montastraea cavernosa	316 D	West Turtle Shoal	431.9	6.9	4 of 4	Y
Montastraea cavernosa	316 U	West Turtle Shoal	527	9	4 of 4	Y
Montastraea cavernosa	317 D	West Turtle Shoal	300.3	8.8	4 of 4	Y
Montastraea cavernosa	317 U	West Turtle Shoal	330	16.4	4 of 4	Y
Montastraea cavernosa	319 D	West Turtle Shoal	268.1	12.1	4 of 4	Y
Montastraea cavernosa	319 U	West Turtle Shoal	457.4	7.5	4 of 4	Y
Montastraea cavernosa	320 D	West Turtle Shoal	153	16.1	4 of 4	Y
Montastraea cavernosa	320 U	West Turtle Shoal	315.3	15.1	4 of 4	Y
Montastraea cavernosa	321 H	West Turtle Shoal	244.8	10.8	3 of 4 **	Y
Montastraea cavernosa	322 H	West Turtle Shoal	214.4	12.7	4 of 4	Y
Montastraea cavernosa	323 H	West Turtle Shoal	359.3	7.6	4 of 4	Y
Montastraea cavernosa	356 D	Lat Long 2 - southernmost	149.1	6	4 of 4	Y
Montastraea cavernosa	356 U	Lat Long 2 - southernmost	127	7.6	4 of 4	Y

Montastraea cavernosa	357 D	Lat Long 2 - southernmost	371.2	6.1	4 of 4 *	Y
Montastraea cavernosa	357 U	Lat Long 2 - southernmost	498.8	9.5	4 of 4	Y
Montastraea cavernosa	358 D	Lat Long 2 - southernmost	356.7	5.7	4 of 4	Y
Montastraea cavernosa	358 U	Lat Long 2 - southernmost	439.8	6	4 of 4	Y
Montastraea cavernosa	359 D	Lat Long 2 - southernmost	343.6	13.2	4 of 4 *	Y
Montastraea cavernosa	359 U	Lat Long 2 - southernmost	236.6	8.4	4 of 4	Y
Montastraea cavernosa	361 H	Lat Long 2 - southernmost	349.9	5.9	4 of 4	Y
Montastraea cavernosa	362 H	Lat Long 2 - southernmost	111.3	11.5	4 of 4	Y
Montastraea cavernosa	363 H	Lat Long 2 - southernmost	429.5	9.7	4 of 4	Y
Montastraea cavernosa	436 D	Lat Long 4 - northernmost	186.6	8.5	4 of 4	Y
Montastraea cavernosa	436 U	Lat Long 4 - northernmost	748.7	11	4 of 4	Y
Montastraea cavernosa	437 D	Lat Long 4 - northernmost	940.4	5.4	4 of 4	Y
Montastraea cavernosa	437 U	Lat Long 4 - northernmost	688.9	7.1	4 of 4	Y
Montastraea cavernosa	438 D	Lat Long 4 - northernmost	443.7	13.1	4 of 4	Y
Montastraea cavernosa	438 U	Lat Long 4 - northernmost	366.9	10	4 of 4	Y
Montastraea cavernosa	439 D	Lat Long 4 - northernmost	860.6	11.5	4 of 4	Y
Montastraea cavernosa	439 U	Lat Long 4 - northernmost	592.3	5.1	4 of 4	Y
Montastraea cavernosa	441 H	Lat Long 4 - northernmost	706.3	12.8	4 of 4	Y
Montastraea cavernosa	442 H	Lat Long 4 - northernmost	1125.1	6.7	4 of 4	Y
Montastraea cavernosa	443 H	Lat Long 4 - northernmost	315.3	4	4 of 4	Y
Montastraea cavernosa	506 H	Western Sambo Patch 2	54.2	6.9	4 of 4	Y
Pseudodiploria strigosa	332 D	West Turtle Shoal	71.8	7.6	4 of 4	Y
Pseudodiploria strigosa	332 U	West Turtle Shoal	254.3	33.2	4 of 4 *	Y
Pseudodiploria strigosa	333 D	West Turtle Shoal	101.2	24	4 of 4	Y
Pseudodiploria strigosa	333 U	West Turtle Shoal	219.2	28.4	4 of 4	Y

Pseudodiploria strigosa	334 D	West Turtle Shoal	170.7	2.9	4 of 4 *	Y
Pseudodiploria strigosa	334 U	West Turtle Shoal	58.6	9.6	4 of 4	Y
Pseudodiploria strigosa	335 D	West Turtle Shoal	68.5	4.8	4 of 4	Y
Pseudodiploria strigosa	335 U	West Turtle Shoal	140	8.3	4 of 4	Y
Pseudodiploria strigosa	337 H	West Turtle Shoal	115.7	32	4 of 4	Y
Pseudodiploria strigosa	338 H	West Turtle Shoal	93.3	14.7	4 of 4	Y
Pseudodiploria strigosa	339 H	West Turtle Shoal	96.6	6.4	4 of 4	Y
Pseudodiploria strigosa	372 D	Lat Long 2 - southernmost	194.4	19	4 of 4	Y
Pseudodiploria strigosa	372 U	Lat Long 2 - southernmost	151.3	30.8	4 of 4	Y
Pseudodiploria strigosa	373 D	Lat Long 2 - southernmost	161.2	5.7	4 of 4 *	Y
Pseudodiploria strigosa	373 U	Lat Long 2 - southernmost	186.2	12	4 of 4	Y
Pseudodiploria strigosa	374 D	Lat Long 2 - southernmost	95.3	6.8	4 of 4	Y
Pseudodiploria strigosa	374 U	Lat Long 2 - southernmost	20.9	6.2	4 of 4	Y
Pseudodiploria strigosa	375 D	Lat Long 2 - southernmost	92.5	4.9	4 of 4	Y
Pseudodiploria strigosa	375 U	Lat Long 2 - southernmost	89.7	5.7	4 of 4	Y
Pseudodiploria strigosa	377 H	Lat Long 2 - southernmost	194.1	3.7	4 of 4 *	Y
Pseudodiploria strigosa	378 H	Lat Long 2 - southernmost	219	3.8	4 of 4	Y
Pseudodiploria strigosa	379 H	Lat Long 2 - southernmost	373.8	4.5	4 of 4	Y
Pseudodiploria strigosa	512H	Western Sambo Patch 2	180.2	6.1	4 of 4	Y
Pseudodiploria strigosa	452 D	Lat Long 4 - northernmost	163.5	5	4 of 4	Y
Pseudodiploria strigosa	452 U	Lat Long 4 - northernmost	122.6	12	4 of 4	Y
Pseudodiploria strigosa	453 D	Lat Long 4 - northernmost	109.2	45.1	4 of 4	Y
Pseudodiploria strigosa	453 U	Lat Long 4 - northernmost	163.6	60.9	4 of 4	Y
Pseudodiploria strigosa	454 D	Lat Long 4 - northernmost	158.5	14.1	4 of 4	Y
Pseudodiploria strigosa	454 U	Lat Long 4 - northernmost	121.8	6.7	4 of 4	Y

Pseudodiploria strigosa	455 D	Lat Long 4 - northernmost	173.2	11.3	4 of 4	Y
Pseudodiploria strigosa	455 U	Lat Long 4 - northernmost	313.2	16.4	4 of 4	Y
Pseudodiploria strigosa	457 H	Lat Long 4 - northernmost	261	7.7	4 of 4	Y
Pseudodiploria strigosa	458 H	Lat Long 4 - northernmost	167.3	10.1	4 of 4	Y
Pseudodiploria strigosa	459 H	Lat Long 4 - northernmost	129.1	9.2	4 of 4	Y
Siderastrea siderea	324 U	West Turtle Shore	12.7	9.3	4 of 4	Y
Siderastrea siderea	327 D	West Turtle Shore	30.1	3.6	4 of 4	Y
Siderastrea siderea	509 H	Western Sambo Patch 2	30.1		4 of 4	Y
Orbicella faveolata	308 D	West Turtle Shore	35.6	5.7	4 of 4	Y
Orbicella faveolata	308 U	West Turtle Shore	19.5	4.5	4 of 4	Y
Orbicella faveolata	310 D	West Turtle Shore	27	7.6	4 of 4	Y
Orbicella faveolata	310 U	West Turtle Shore	68.7	7.4	4 of 4	Y
Orbicella faveolata	503 H	Western Sambo Patch 2	15.8	6	4 of 4	Y

* Samples that produced less than 1 million paired reads for one of the four sample proccessings. Within the samples subjected to rRNA removal: 321 H, 334 D, and 426 H. Furthermore, among the samples that underwent small RNA enrichment: 332 U, 357 D, 359 D, 373 D, and 377 H. These samples were still included in all the analyses performed.

** Besides the low amount of reads, no viral transcripts were identified for sample 321 H library subjected to rRNA removal.

2.2 Task 1b: Initial quality control and database generation of meta-transcriptomic data libraries:

Once Vega Thurber lab sequence data was fully generated last summer (2022), the research team began conducting sequence quality control, initial processing, and preliminary analysis of the sequence data. This task was led by PhD students Eddie Fuques and Alex Veglia, who together have eight years of experience working on viral metagenomics.

2.2.1 Quality check and trimming of the data

Quality control of the raw reads was performed with FastQC (Andrews 2010), while MultiQC (Ewels et al. 2016) was used for the visualization of multiple results in a single report. Trimming of the raw data (adapter contamination detection/removal, low-quality read removal, and read error correction) was conducted using fastp (v0.20.1, Chen et al., 2018) with default settings. Once QC was complete (Table 1), the remaining sequences were analyzed.

2.3 Task 2a. Independent analysis of the 2 datasets generated in Vega Thurber Lab.

To ensure each lab conducted analysis of the sequence datasets resulting from their respective library preparation methods in an unbiased fashion, sequenced datasets were initially processed and evaluated independently. Each team agreed on shared approaches to be used separately on their respective data; results were not initially compared until respective viral analyses were completed by each lab.

2.4 Task 2b: Comparative analysis of Vega Thurber and Correa lab datasets to holistically assess the role of viruses in SCTLD infection status and disease severity.

From this comparative analysis, we focused on **the primary question of whether SCTLD or its susceptibility is differentially associated with any viral genomes and/or viral transcripts**. As a next step, full comparison of the results and statistical validation from these four metatranscriptomic library approaches can support the development of a comprehensive user guide for future researchers regarding coral viral analysis. This guide/protocol can include molecular and informatic pipelines that can explore past and future samples for viral signatures.

2.4.1 Metatranscriptome processing and virus diversity assessment

An overview of the metatranscriptomic sequence processing pipeline used in this project is provided in Figure 1. In summary, raw sequencing data underwent quality control, normalization, *de novo* assembly, identification of viral genomes, taxonomic classification, and differential expression analysis. Below we provide specifics regarding the methods associated with these informatic analyses:

- **Data normalization and *de novo* assembly**
  Normalization of sequence transcripts/reads was performed with the BBNorm tool from the package BBMap (Bushnell 2014) in order to remove redundant sequences. This step reduces the computational cost of downstream analyses. *De novo* assembly of the reads was carried out using RNASPAdes (Bushmanova et al. 2019) with k-mer sizes of 77, 99, and 127.

- **Identification and quantification of viral sequences**
  Potential viral contigs were selected using Deep6 (Finke, Kellogg, and Suttle 2023) a deep learning model for classifying sequences into prokaryotes, eukaryotes, or one of the four viral realms (Duplodnaviria, Varidnaviria, Monodnaviria, or Riboviria). Contigs with a score of 0.7 or higher for any of the viral realms and a minimum of 500 bp length were selected for further analyses and pooled based on the type of sample processing (PolyA enrichment, dsRNA immunoprecipitation, small RNA enrichment or rRNA removal). These pools of contigs were used to generate four non-redundant databases using the program EvidentialGene (Gilbert 2013).

  Reads were mapped to their corresponding non-redundant database using Kallisto (v0.48.00, Bray et al. 2016) with a total of 100 bootstrap samples, and abundance data was obtained using Trinity (v2.15.1, Grabherr et al. 2013).

Each one of the non-redundant databases generated with EvidentialGene was subjected to five different tools in parallel to obtain the most reliable results possible for taxonomic classification and quality assessment of the potential viral contigs:

- Contigs were mapped to the IMG/VR v4 database (Camargo et al. 2022) using DIAMOND v2.0.15 (Buchfink, Reuter, and Drost 2021) with a minimum bit score of 60, a minimum percentage identity of 40%, and a minimum length of 80 amino acids. The IMG/VR database contains >15 million virus genomes and genome fragments and includes functional, taxonomic, and ecological metadata associated.
- To assess the quality and completeness of the viral genomes we used CheckV v1.0.1 (Nayfach et al. 2020) with default settings.
- A second round of Deep6 was performed to classify the sequences, following the same criteria as described above.
- VirSorter2 v2.2.4 (Guo et al. 2021) was implemented with default settings.
- GeNomad v1.6.1 (Camargo et al. 2023) was employed with default settings.

Results obtained from all these analyses were taken together to obtain an accurate and confident classification of the viral sequences present in the data. A custom BASH script was then developed to evaluate results from all approaches and assign each putative viral sequence a "virus confidence score" and a "classification confidence category". Virus confidence score (*S)* is calculated using the equation $S=(g+V+C+D+d)/(5)$ where the sum of the geNomad score (*g*), VirSorter2 score (*V*), CheckV score (*C*), DIAMOND score (*D*), and Deep6 score (*d*) is divided by the total possible score (5). Virus confidence scores closer to 1 suggest higher confidence that a given contig is viral in nature. Differences in the abundance of the different viral taxa were analyzed via differential expression analysis using the R packages phyloseq and DESeq2. Plots were generated with R packages ggplot2 and pheatmap.

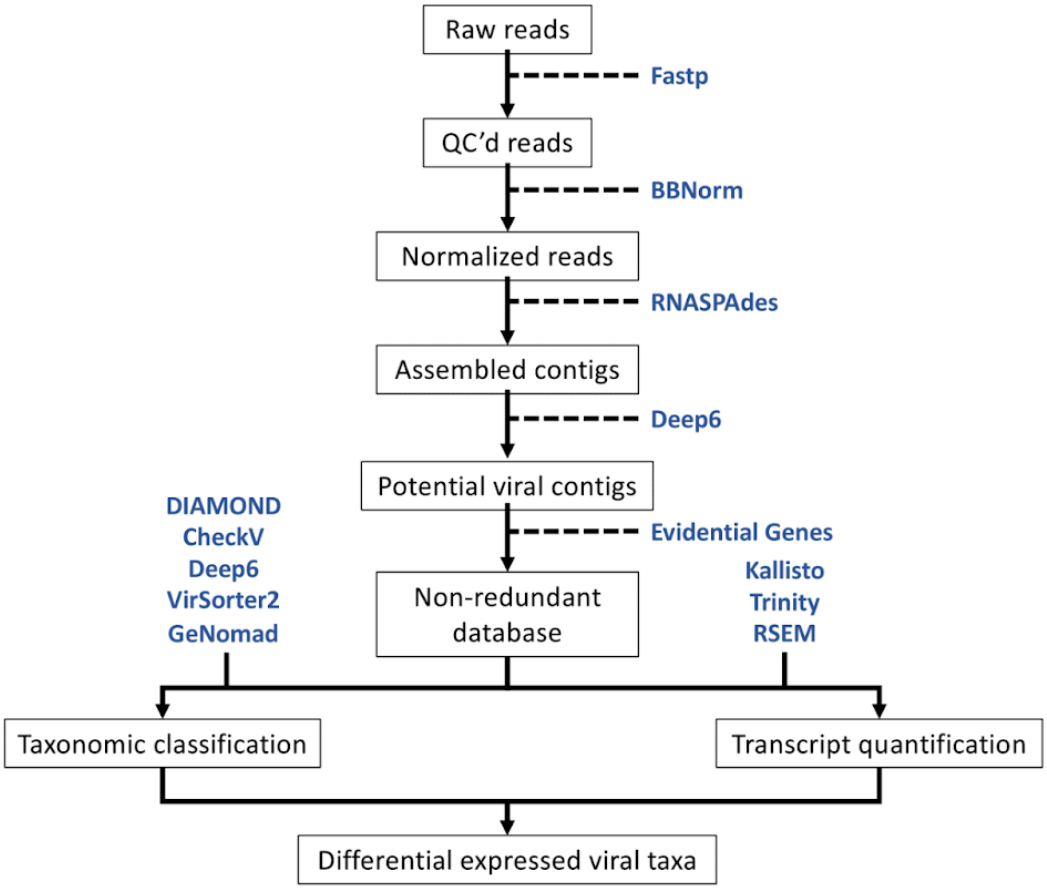

Figure 1. Schematic overview of the meta-transcriptomic sequence processing pipeline. Raw sequencing data underwent quality control, normalization, de novo assembly, identification of viral genomes, taxonomic classification, and differential expression analysis.

2.4.2 Filamentous virus genome-level comparisons

Putative complete filamentous virus genomes were identified using the program CheckV v1.0.1 (Finke, Kellogg, and Suttle 2023; Nayfach et al. 2021) with default settings. Virus contigs that were considered "High quality" by CheckV and classified as Patatavirales or Tymovirales were first extracted from the dsRNA immunoprecipitation and polyA enriched non-redundant reference databases. Sequences were then manually checked to ensure their genome lengths were similar to what has been reported previously for these viral Orders. After this initial check, genome sequence characteristics were compared to identify patterns and describe any variation, focusing on genome length, the number of open reading frames (ORF), and ORF protein identity. ORF prediction was performed with the program, Prodigal (Hyatt et al. 2010), and protein identity was inferred by aligning sequences to the UniProtKB reference proteomes and the Swiss-Prot databases using BLASTp (Camacho et al. 2009) with an e-value cutoff of 0.001. Genomic sequences were also aligned using Mauve v1.1.3 (Camacho et al. 2009; Darling et al. 2004) to visually assess genome similarity.

2.4.3 RNA Virus RdRp Marker Gene Analysis

Virus *RdRp* protein sequences were identified from the translated non-redundant reference databases by aligning sequences to the RdRp-scan sequence database (Charon et al. 2022) using DIAMOND BLASTp with the parameters described previously. Protein sequences exhibiting homology to reference *RdRp* sequences were then extracted from the non-redundant reference file. Taxonomy was assigned to *RdRp* sequences based on their best match within the RdRp-scan database. To provide phylogenetic context to these sequences we combined our *RdRp* sequences with RdRp-scan sequences. All sequences were first aligned using the program muscle v5 (v5; Edgar 2021) then a maximum likelihood tree was generated using the program IQTREE2 (Minh et al. 2020) with the best model determined by ModelFinder (Kalyaanamoorthy et al. 2017).

2.5 Task 3: Sample preparation of ISH slides (n = 630). Immunostaining and microscopy imaging of *O. faveolata* samples resulting in 432 microscopy images (includes raw & processed).

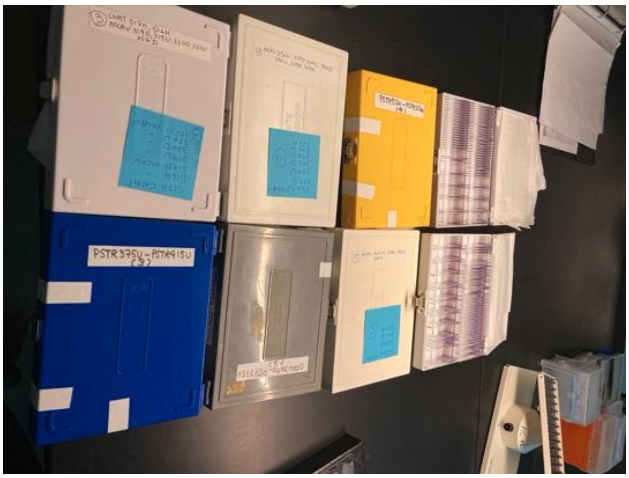

Figure 2. Slide boxes containing the 630 slides of coral tissues sampled from reefs affected by stony coral tissue loss disease (SCTLD) that were produced for this project. The slides underwent analysis for an indicator of viral infections.

For each sample set, a total of fifteen thin sections were generated from three tissue types (n=5 thin sections per tissue type including Disease Lesion, Disease Healthy, and Healthy Healthy) by cutting paraffin-embedded coral holobiont tissues with a microtome at a thickness of four microns, and adhered to positively charged glass slides. Since Symbiodiniaceae cells are 8-12 microns on average, this approach partitioned individual symbiont cells across multiple slides in a series, allowing immunostaining of dsRNA and experimental controls to be applied to different sections of the same symbiont cells across slides. A total of 630 slides were produced (Figure 2).

2.5.1 Dewaxing, rehydrating, permeabilization and tissue encircling of slides

Coral thin section slides were dewaxed in Histosol, then rehydrated with a series of ethanol washes (100% to 50%) and DEPC-treated water, then stored in 1X PBS in the dark at 4C. RNase-treated samples were kept separate from non-RNase treated samples. To permeabilize

the coral tissues, the slides were then incubated for 10 minutes in 1X PBT (PBS + 0.1% tween) in a Copeland jar three times, then placed in 1X PBS. Each slide was then removed from the 1X PBS and a Kimwipe was used to dry the area around the tissue. A hydrophobicity pen was used to draw a circle around the coral tissue, and the slide was then returned to 1X PBS.

2.5.2 Enzyme pre-treatment

Slides were pre-treated with Proteinase-K (20ug/mL) + DTT at 37℃ for 1 hour. Excess solution was then removed from the slides and they were indirectly rinsed using a PBT squeegee. Slides were returned to the Copeland jar and 1X PBT was added for a 10-minute incubation step three times. Non-RNase III-treated slides were then placed in a moisture chamber.

2.5.3 RNase III treatment for a subset of slides per coral tissue sample

For slides undergoing RNase III-treated, 1X RNase III + DTT buffer was then added to cover the encircled tissue. Three Units of RNase III were then added to the buffer and slides were incubated at 37℃ for 2 hours. After the incubation period for the RNase III-treated slides, excess enzyme was removed with a tap/PBT rinse/PBT wash three times for 10 minutes.

2.5.4 Immunostaining to primary antibody

A blocking buffer solution (0.01% w/v BSA; 5% w/v heat inactivated Goat Serum; 95% PBST) was added to the top of the encircled portion of each thin section, which was then placed in a moisture chamber and incubated at room temperature in the dark for 1 hour. Excess blocking solution was then removed and the primary antibody (9D5, 1:1000) was added to the encircled tissue on the thin sections and incubated at 4℃ overnight.

2.5.5 Immunostaining to secondary antibody

Slides were then returned to the Copeland jar and incubated in 1X PBT for 10 minutes at room temperature; this wash was repeated four times in total. Slides were placed in the moisture chamber and incubated in blocking solution for 15 minutes at room temperature, and then excess blocking solution was removed. Encircled tissues on slides were then incubated in the secondary antibody (Alexa Fluor 405, 1:200, i.e. 5µL antibody / 1 mL block) for 2 hours at 4℃; subsequently, excess fluid was removed. Slides were then washed with 1X PBT for 10 minutes at room temperature, then rinsed three times in 1X PBS.

2.5.6 Mounting slides

A #1.5 coverslip with 45uL of Fluoromount-G was placed on each slide and allowed to cure for at least 2 hours, then stored at 4℃ until imaging was conducted on the microscope.

2.5.7 Microscopic imaging

Following immunostaining, samples were imaged on a Nikon A1-Rsi confocal microscope using real-time spectral unmixing after 405nm excitation (for Alexa Fluor 405) and 560nm excitation (for Alexa Fluor 555). Coral tissue morphology was visualized with a PMT transmitted light detector, as well as with 405nm excitation and spectral unmixing of coral and Symbiodiniaceae (Figure X). Images were processed using the Nikon NIS Elements C imaging software using an in-house fluorescence spectra database that was curated by Dr. Coy (Rice U.). To quantify the dsRNA signal from each coral sample, the autofluorescence for the coral and

Symbiodiniaceae for each image (based on values derived from the control RNAse+ slides) was subtracted from the dsRNA-derived fluorescence on non-RNAse treated slides.

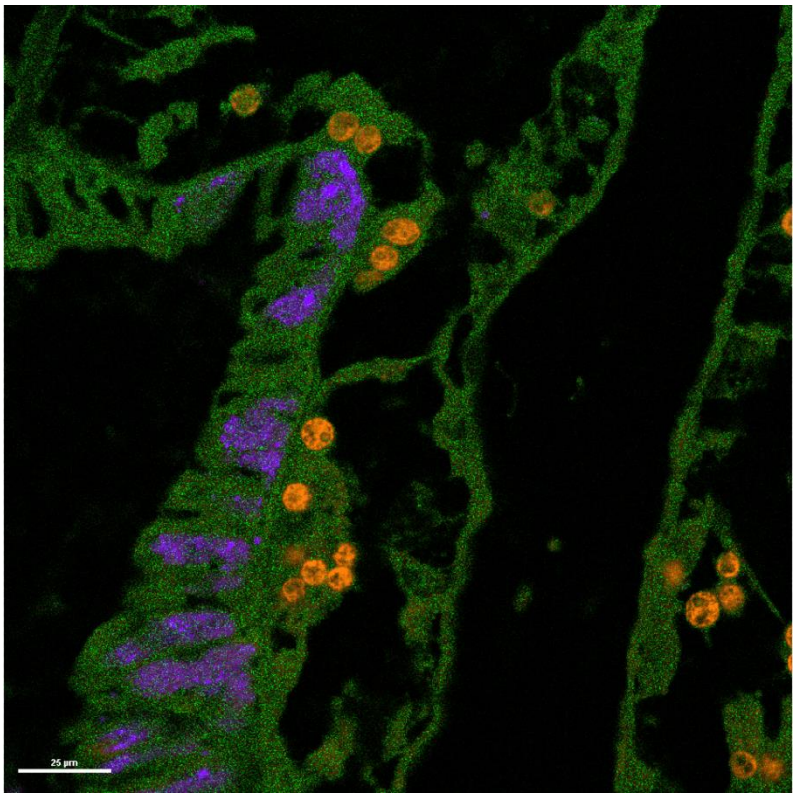

Figure 3. Representative thin-section of Orbicella faveolata coral tissue imaged with confocal microscopy at 40X magnification. Coral tissue is shown in green, dinoflagellate symbionts (Family Symbiodiniaceae) are shown in orange, and an epidermis-localized unknown spectra is shown in purple. Image credit: Dr. Samantha Coy, Rice University.

3 RESULTS

3.1 Task 1 Generation and sequencing of viral meta-transcriptomes from corals collected from Middle Keys:

All sequencing and data curation was completed for the 428 viral metatranscriptomic libraries (Table 1), which can be used to assess the potential role of viruses in SCTLD. The power of metatranscriptomics is that it can achieve several aims in a single run. Viruses can be comprised of either RNA or DNA genomes, and conveniently, metatranscriptomics provides a view of the presence and abundance of RNA viruses (from genome sequence), as well as DNA virus activity (from DNA virus transcripts). These data, as well as transcripts resulting from host and symbiotic microbe gene expression (available but to be analyzed in future work), can provide a holistic view of the changing biology occurring within colonies of different coral species that are affected or apparently unaffected by SCTLD. This study and its comprehensive associated sequence databases will be available to all coral researchers interested in coral disease, and can assist in our overall assessment of host, symbiont, bacterial, and viral dynamics in the Florida Keys, as well as the wider Caribbean.

3.2 Task 1b: Initial quality control and database generation of meta-transcriptomic data libraries:

Two types of primary sequence data, raw and curated, were generated. For the raw data, all sequence information (FASTQ files) from the sequencing centers are stored locally at OSU and at Rice University (where Dr. Correa retains an adjunct research professor position). All raw data are also stored on 5TB LaCie rugged external hard drives for long-term archiving; Florida DEP is in possession of these hard drives.

All processed data, including QC reads (FASTQ files) and viral sequence contigs (FASTA format), will become fully publicly available on the Sequence Read Archive (SRA) and GenBank databases once data analysis is complete. These data, and all the code associated with our analyses, will be made available in public archives. We expect this to be before 1.5 years from now (March of 2025). We are currently generating two manuscripts for peer review from these datasets.

3.3 Task 2a. Independent analysis of the 2 datasets generated in Vega Thurber Lab.

To ensure we conducted our work in an unbiased fashion, we agreed to initially process and evaluate our data streams independently. Each team agreed on shared approaches to be used separately on their respective data; results were not initially compared until respective viral analyses were completed by each lab. During their initial analysis efforts, the Vega Thurber lab found suspiciously high numbers of Coronaviridae contigs in the sequence datasets ('small RNA enrichment' and 'rRNA removal') generated by University of Oregon. Because of the notoriety and focus on coronaviruses in contemporaneous virus research, and because sequences similar to the Coronaviridae have not previously been reported from corals, we immediately suspected the libraries were contaminated with exogenous coronavirus cDNA. After performing our bioinformatic pipeline for the identification of viral genomes (quality check, normalization of reads, de novo assembly, and taxonomic classification of contigs), we were able to identify several contigs that corresponded specifically to variants of the SARS-CoV-2 genome that causes COVID-19 in humans (Table 2).

Table 2. Number of contigs identified as belonging to the SARS-CoV-2 genome for each type of library generated by UofO, and the percentage that those contigs represented within the total number of contigs identified as viral**.** *H = apparently healthy; U = unaffected tissue on a SCTLD-affected colony; D = diseased tissue from a SCTLD-affected colony*.

Type of library	Number of contigs	Percentage of viral contigs
rRNA removal - H	46	23%
rRNA removal - U	51	11%
rRNA removal -D	54	30%
Small RNA - H	45	4%
Small RNA - U	43	15%
Small RNA - D	60	22%

To confirm the presence of COVID-19-related SARS-CoV-2 genomes in our sequencing data, we conducted assemblies using the strain Wuhan-Hu-1 (NC_045512.2) as a reference. When normalized reads from rRNA removal libraries were used to do the mapping, up to 2% of the normalized reads were mapped to the SARS-CoV-2 genome, a very high percentage given the source of these samples (Figure 4). When normalized reads from small RNA enriched libraries were used to do the mapping, up to 3% of the normalized reads were mapped to the SARS-CoV-2 genome (Figure 5).

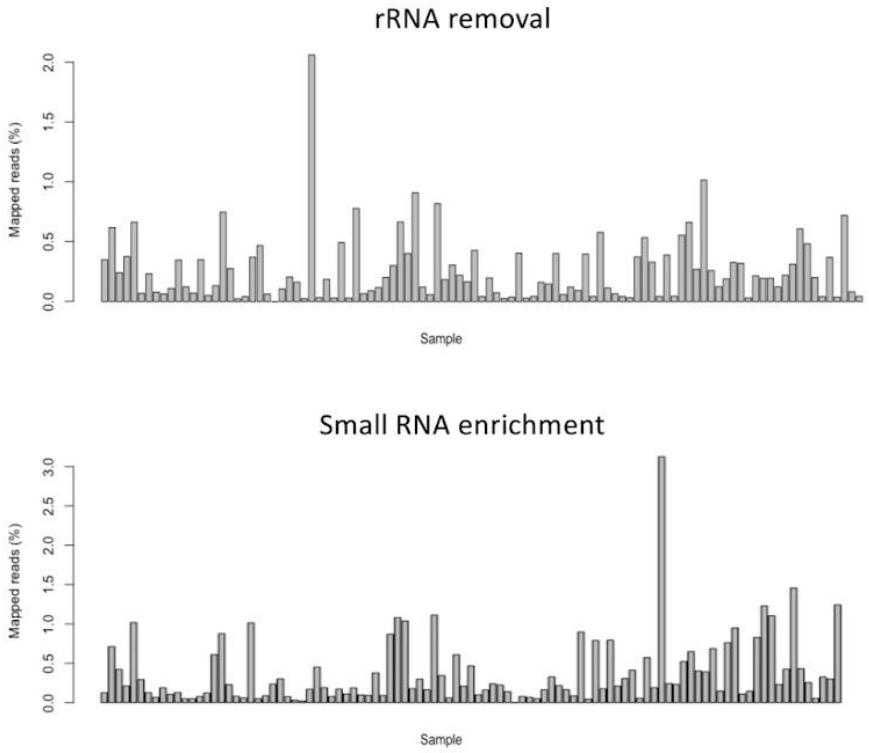

Figure 4. Percentage of normalized reads that mapped to the SARS-CoV-2 reference genome for libraries subjected to rRNA removal (top) and small RNA enrichment (bottom).

Reference-genome assemblies were also performed for two of the quality-checked libraries (one per sample treatment) on quality trimmed (but unnormalized) reads. For library 332-D_ribozero, a total of 108,575 reads covered 87.75% of the genome with an average depth of 200 (Figure X). For library 422-D_Size_selected, a total of 1,276,000 reads covered 99.34% of the genome with an average depth of 6,121. These results confirmed the presence of the complete SARS-CoV-2 genome in the libraries, representing an important portion of the total reads generated during sequencing, particularly considering the established understanding that viral sequences typically constitute a minute fraction of sequencing outcomes during metagenomic analyses. This further demonstrated that all sequences with best hits to SARS-CoV-2 in the UofO-generated libraries constituted exogenous contamination, and did not originate from the coral holobiont samples themselves. **Therefore, we elected to remove all of these contaminating sequences from the UofO-generated datasets prior to conducting downstream analyses of these two libraries**

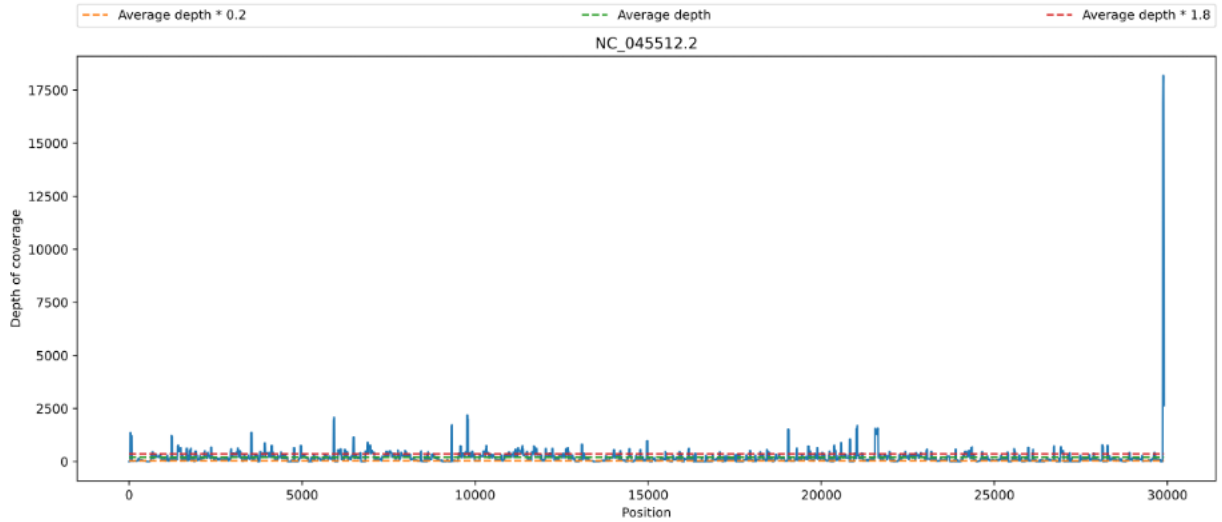

Figure 5. Coverage plot for the SARS-CoV-2 reference genome assembly performed for sample 332-D subjected to rRNA removal.

It is important to note that as part of this project, by design, the Correa lab sent RNA aliquots of the same samples sequenced by the Vega Thurber lab to a non UofO sequencing facility. After performing the same bioinformatic pipeline, no contigs were identified as SARS-CoV-2 in the Correa lab datasets. The absence of coronavirus in Correa lab data (coronaviruses do polyadenylate, thus if they were truly present, they should be in the Rice polyA enriched library) confirms that University of Oregon contaminated the Vega Thurber lab's sequencing runs and, therefore, all resulting sequence datasets. Despite this challenging circumstance, the ability to compare Vega Thurber lab data to Correa lab data for the same samples (but different library preparations), gave the research team high confidence that the correct path forward was for the Vega Thurber lab to remove all best hits to coronaviruses from their sequence datasets and subsequently conduct their independent bioinformatic analyses.

3.4 Task 2b: Comparative analysis of Vega Thurber and Correa lab datasets to holistically access role of viruses in SCTLD infection status and disease severity (joint with Rice University).

The research team collaborated to devise an appropriate and benchmarked method to compare the coral holobiont metatranscriptomes generated across the different library preparation approaches, coral species and colony health states. After development of the informatic pipeline (see Methods), it was applied to all 4 library preparation methods (polyA enrichment, dsRNA immunoprecipitation, small RNA enrichment, and rRNA removal) to identify genomic and transcriptomic sequences representing potential viruses associated with SCTLD-affected and apparently unaffected coral tissues, as well as tissues from apparently healthy corals (Figure 6). At the highest Domain level, few differences were seen in the approaches although some did have changes in the relative abundance of Riboviria (RNA viruses) and Duplodnaviria (dsDNA viruses).

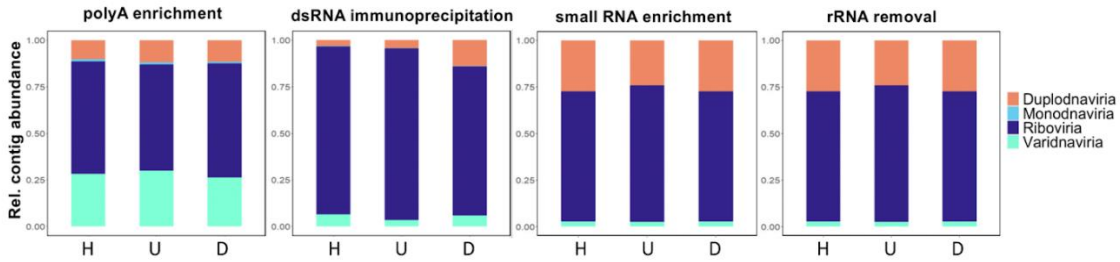

Figure 6. Relative contig abundance across the four viral realms and by library preparation method. PolyA enrichment notably contained more Varidnaviria (DNA viruses with a vertical jelly fold roll in capsid) than the other three library preparation methods, whereas dsRNA immunoprecipitation contained fewer Duplodnaviria (dsDNA viruses) and more Riboviria (RNA viruses) sequences. The methods 'small RNA enrichment' and 'rRNA removal' had similar Domain-level annotations to each other, but fewer Varidnaviria than the other two library preparation methods. H = Apparently Healthy; U = Unaffected tissue from a SCTLD-affected colony; D = Diseased tissue from a SCTLD-affected colony.

3.4.1 Specific Viral Orders Are Associated with SCLTD-affected Corals.

In order to test the hypothesis that viruses are associated with SCTLD, we compared all viral annotations at the Realm and Order levels across the three health states: apparently healthy (H), unaffected tissue from a SCTLD-affected colony (U), and diseased tissue from a SCLTD-affected colony (D). In summary, we found strong evidence that some viral Orders and Families are strongly associated with SCLTD, based on results from three of the four library preparation approaches (polyA enrichment, dsRNA immunoprecipitation and small RNA enrichment). To our surprise the rRNA removal approach, a gold standard in viral work, identified fewer viral groups that were differentially abundant in diseased coral tissues (Figure 7).

Using differential abundance analysis, we found there was a large number of viral Orders that were significantly more abundant in diseased samples (Figure 7), and many of these Orders were consistent across the three coral species for which there was sufficient replication: *Colphyllia natan*s (Figure 8), *Montastrea cavernosa* (Figure 9), and *Pseudodiploria strigosa* (Figure 10). Importantly, three of the four library preparation methods showed that Tolivirales and Picornavirales were the most significantly elevated in diseased coral holobiont tissues. Tolivirales and Picornavirales are both small non-enveloped icosahedral viruses with a +ssRNA genome that can infect plants (Tolivirales) and animals (Tolivirales and Picornavirales).

Several other viral Orders were highly abundant in diseased samples, including three members of the nuclear cytoplasmic large DNA viruses (NCLDVs): Imitervirales (mimiviruses), Pimascovirales, Algavirales (phycoviruses). These Orders are known to infect corals and their symbionts (e.g., (Wood-Charlson et al. 2015) (Howe-Kerr et al. 2023; Thurber et al. 2017) and were previously implicated in worsening bleaching phenotypes ((Correa et al. 2016); (Gunjal and Shinde 2021); (Soffer et al. 2014); Messyasz et al. 2020, Schmeltzer et al., in prep). Two other RNA virus orders were differentially abundant in diseased samples: Durnavirales (dsRNA fungal and plant viruses) and Cryppavirales (capsid-less ssRNA mitochondria-infecting viruses). With the exception of Cryppavirales, these viral groups were differentially abundant in diseased tissues across all four library preparation methods. Interestingly, none of these differentially abundant viral Orders contain viruses with known filamentous morphologies, despite previous documentation of filamentous virus-like particles from these same coral colonies (Work et al. 2021). However, a recent study reported filamentous virus-like particles within Symbiodinaceae

in Pacific coral holobionts (where SCTLD has yet to be observe); filamentous viruses may therefore be common and widespread in stony corals, and not necessarily diagnostic of stony coral tissue loss disease (Howe-Kerr et al., *in review*).

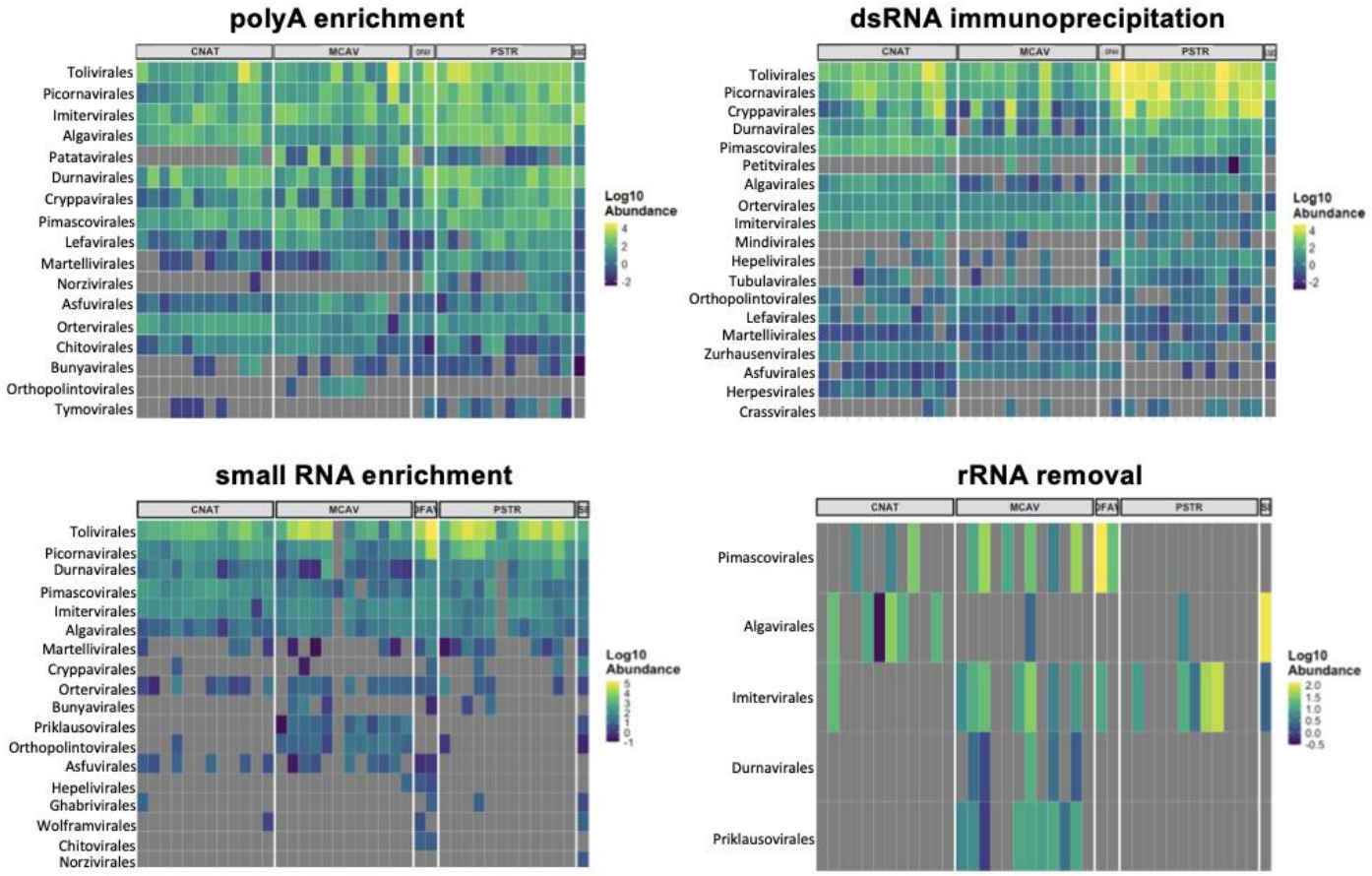

Figure 7. Similar viral Orders are differentially abundant in diseased tissues of SCTLD-affected corals across all four species. Based on annotations across all four library generation approaches and all coral species in this study, there was strong agreement in the viral Orders differentially abundant in SCTLD-affected (diseased) tissues. Three of four methods identify Tolivirales and Picornavirales (+ssRNA viruses), as the top two viral Orders differentially abundant in the diseased tissues of SCTLD-affected colonies. Members of the NCLDVs (Imitervirales, Durnavirales, Algavirales) were also differentially abundant in some coral species; results from diseased tissues of Montastrea cavernosa did not always agree with patterns observed in diseased tissues of Colpophyllia natans and Pseudodiploria strigosa. **Taken together, these data strongly suggests these viral groups play a strong direct or indirect role in SCTLD.**

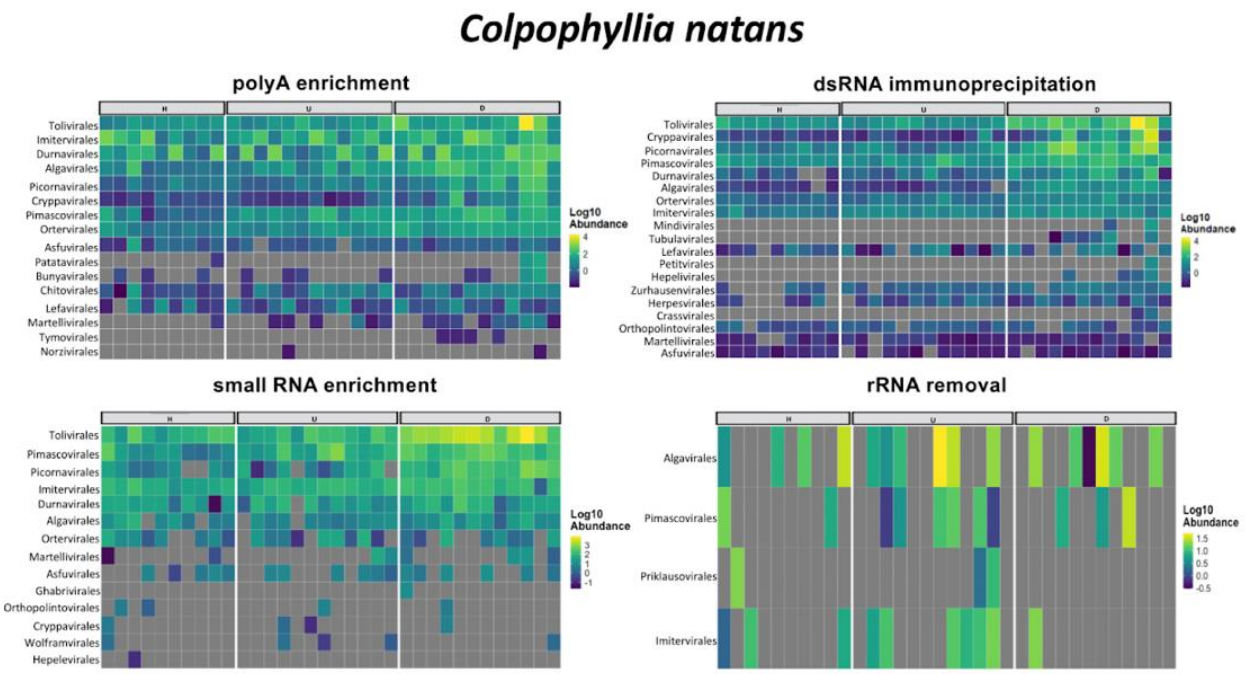

Figure 8. Differentially abundant viral Orders in diseased Copophyllia natans across the four library preparation approaches. In three of the four library preparations, Tolivirales, a +ssRNA virus, was most differentially abundant in diseased tissues. Although Tolivirales was most abundant in diseased (D) tissues, it was also present in lower abundances within the apparently healthy (H) colonies and in the unaffected tissues (U) of diseased colonies. Some other viral Orders, such as Durnavirales and Imitervirales, followed a similar pattern, albeit to a lesser degree.

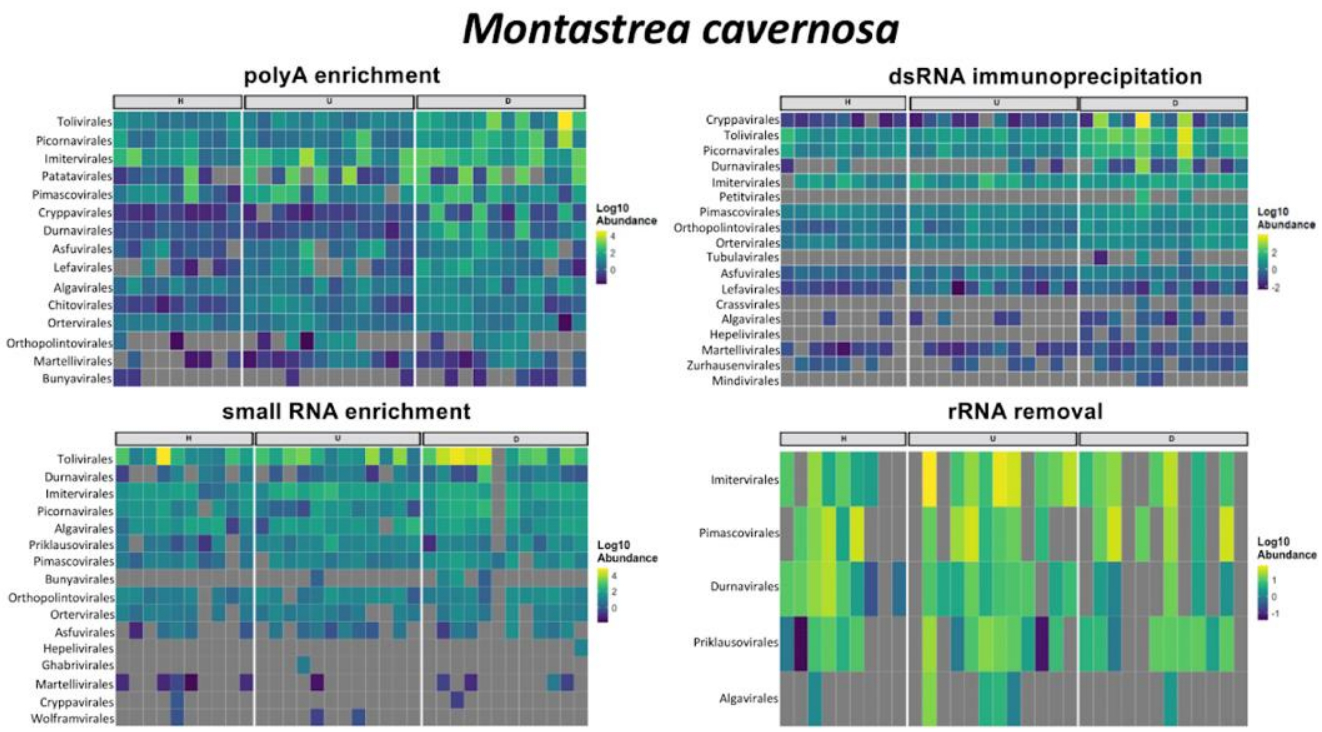

Figure 9. Differentially abundant viral Orders in diseased Montastrea cavernosa across the four library preparation approaches. In two of the four library preparations, Tolivirales, a +ssRNA virus, was most differentially abundant in diseased tissues. Although Tolivirales was most abundant in diseased (D) tissues, it was also present in lower abundances within the apparently healthy (H) colonies and in the unaffected tissues (U) of diseased colonies. While some viral Orders, such as Tolivirales and Imitervirales, were differentially abundant across all health states, other viral Orders were differentially abundant only in diseased tissues, such as the Cryppavirales and Durnavirales.

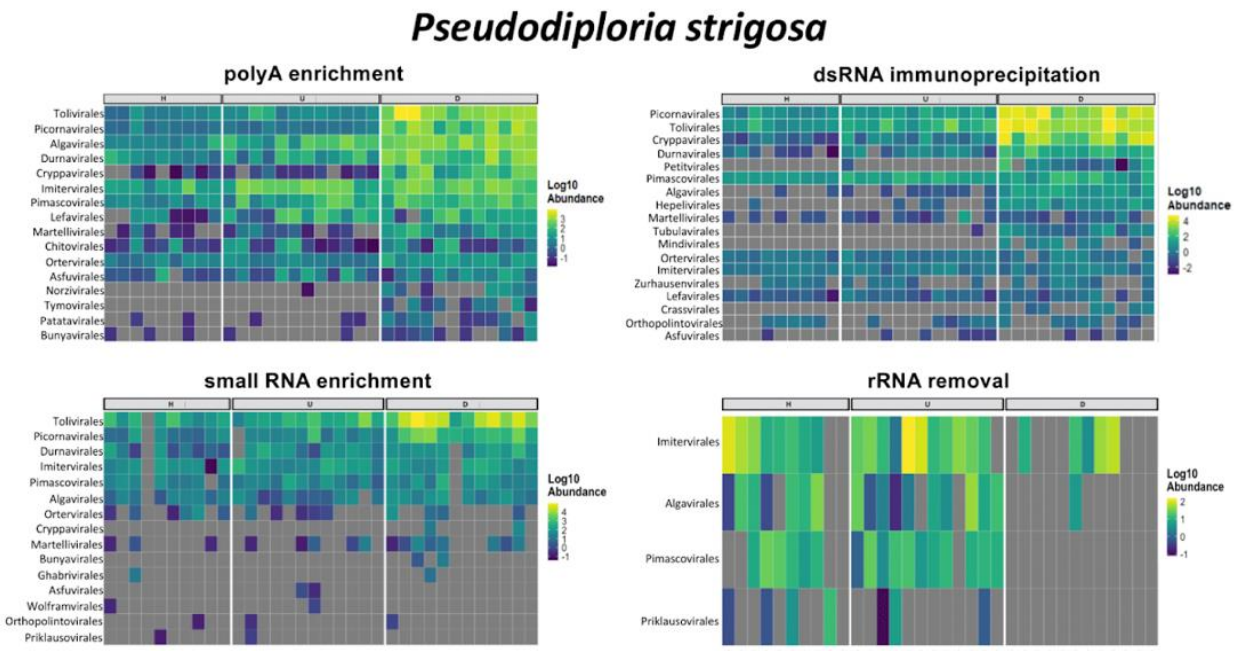

Figure 10. Differentially abundant viral Orders in diseased Pseudodiploria strigosa across the four library preparation approaches. In two of the four library preparations, Tolivirales, a +ssRNA virus, was most differentially abundant in diseased tissues. Although Tolivirales, Cryppavirales and Picornavirales was most abundant in diseased (D) tissues, they were also present in lower abundances within the apparently healthy (H) colonies and in the unaffected tissues (U) of diseased colonies. Interestingly, while some viral Orders were differentially abundant across all health states, other viral Orders, including the filamentous Tymovirales (polyA enriched library preparation only), were differentially abundant only in diseased tissues.

3.4.2 Filamentous virus genome-level comparisons

Given previous identification of filamentous viruses in association with SCTLD-affected corals (Work et al., 2021; Veglia et al. 2022; Veglia 2023), we performed full genome annotation of the filamentous virus-like genomes (Patatavirales and Tymovirales) recovered in this study. Twenty-four virus contigs recovered in this study were identified as Patatavirales, of which four were derived from dsRNA immunoprecipitation sequencing libraries and twenty were derived from polyA enriched RNA sequencing libraries. The 24 Patatavirales genome sequences had lengths ranging from ~8-11.3 Kbp, in agreement with the range reported for the genomes within Family Potyviridae (Inoue-Nagata et al. 2022). All sequences were extremely divergent, impeding accurate genome alignments with Mauve. However, based on visual inspection, alignments indicated four genome "groups" based on similarity in length and the number of ORFs they were predicted to contain. Genomes within Group 1 (n=8) , Group 2 (n=12) and Group 4 (n=1) were predicted to contain a single ORF putatively encoding for a polyprotein. This is a common characteristic of some taxa within Potyviridae (Inoue-Nagata et al. 2022). Interestingly, all 12 genome sequences in Group 3 contained two ORFs, according to Prodigal. One of these ORFs exhibited characteristics consistent with a structural polyprotein, whereas the second identified ORF appeared to encode another polyprotein with a putative RNA-dependent RNA polymerase (*RdRp*) region.

Furthermore, there were eight Tymovirales genomes recovered from the dsRNA immunoprecipitation (n=1) and polyA enriched (n=7) non-redundant sequence databases. The lengths of these genomes ranged from ~6.2-7.5 Kbp, and all but one genome (from the dsRNA immunoprecipitation sequencing data) encoded for three ORFs, resembling the Coral Holobiont-associated Filamentous Viruses (CHFV) we previously reported from the U.S. Virgin Islands (Veglia et al. 2022). The Tymovirales genome recovered from the dsRNA immunoprecipitation sequencing data contained two ORFs, as predicted by Prodigal. One of these ORFs contained an *RdRp* encoding region showing 23% amino acid similarity to the *RdRp* of plant tymoviruses.

3.4.3 Analysis of RNA Dependent RNA Polymerase sequences in apparently healthy and SCTLD-affected corals, with particular attention to members of the Picornavirales

The RNA dependent RNA polymerase (*RdRp*) is a key phylogenetic gene in the study of RNA viruses, as it is found in all of the Ribovaria Domain. Identification and evolutionary analysis of these genes can be used to assess the abundance and relationship among different viruses from a sample. As this analysis requires considerable computational time, we here report results from this approach for two of the four library preparations (dsRNA and polyA enriched); future applications of this approach to the rRNA removal and small RNA enrichment libraries are also likely to yield valuable insights (Figure 11).

In total, 284 *RdRp* sequences from the dsRNA immunoprecipitation libraries and 177 RdRp sequences from the polyA-enriched libraries were analyzed. Overall, most *RdRp* sequences could not be classified to a particular taxonomic group, underscoring that coral holobionts contain high levels of as yet undescribed viral diversity. Specifically, 224 dsRNA immunoprecipitation library sequences and 114 polyA enriched library sequences were unclassifiable, constituting 78.9% and 64.4% of sequences in the respective libraries. This result likely stems (in part) from the fact that even some *RdRp* sequences in the reference database

itself lack classification. It is additionally possible that some *RdRp* sequences in our libraries are highly unique.

Of the 60 dsRNA immunoprecipitation and 63 polyA-enriched library sequences that were able to be classified, 45 dsRNA immunoprecipitation sequences (~15.9% of all dsRNA immunoprecipitation sequences) and 61 polyA-enriched sequences (~35.0% of all polyA enriched sequences) belonged to the Pisuviricota, of which Picornavirales is a member (Figure 11). These sequences are of particular interest because they are among the most differentially abundant viral groups in diseased tissues across all coral species in this study (Figure 7), and are present in the tissues of apparently healthy corals, as well as the unaffected and diseased tissues of SCTLD affected corals for *C. natans* (Figure 8), *M. cavernosa* (Figure 9) and *P. strigosa* (Figure 10). For dsRNA immunoprecipitated viruses, five of the remaining sequences were classified as Lenaviricota, six were classified as Kitrinoviricota, and four were classified as Duplornaviricota, representing approximately 1.8%, 2.1%, and 1.4% of all dsRNA immunoprecipitated sequences analyzed, respectively. For polyA enriched sequences, one sequence was classified as Lenaviricota and one was classified as Duplornaviricota; each represented approximately 0.6% of all polyA enriched viruses studied, respectively (Figure 11). Where possible, finer resolution classifications revealed that some sequences in the dsRNA immunoprecipitated libraries within Picornavirales were most similar to the Discistroviridae (Figure 11). Best sequence similarities to the Discistroviridae were also observed in the polyA enriched libraries, as well as best sequence similarities to the Picornaviridae (another family within Picornavirales, Figure 11).

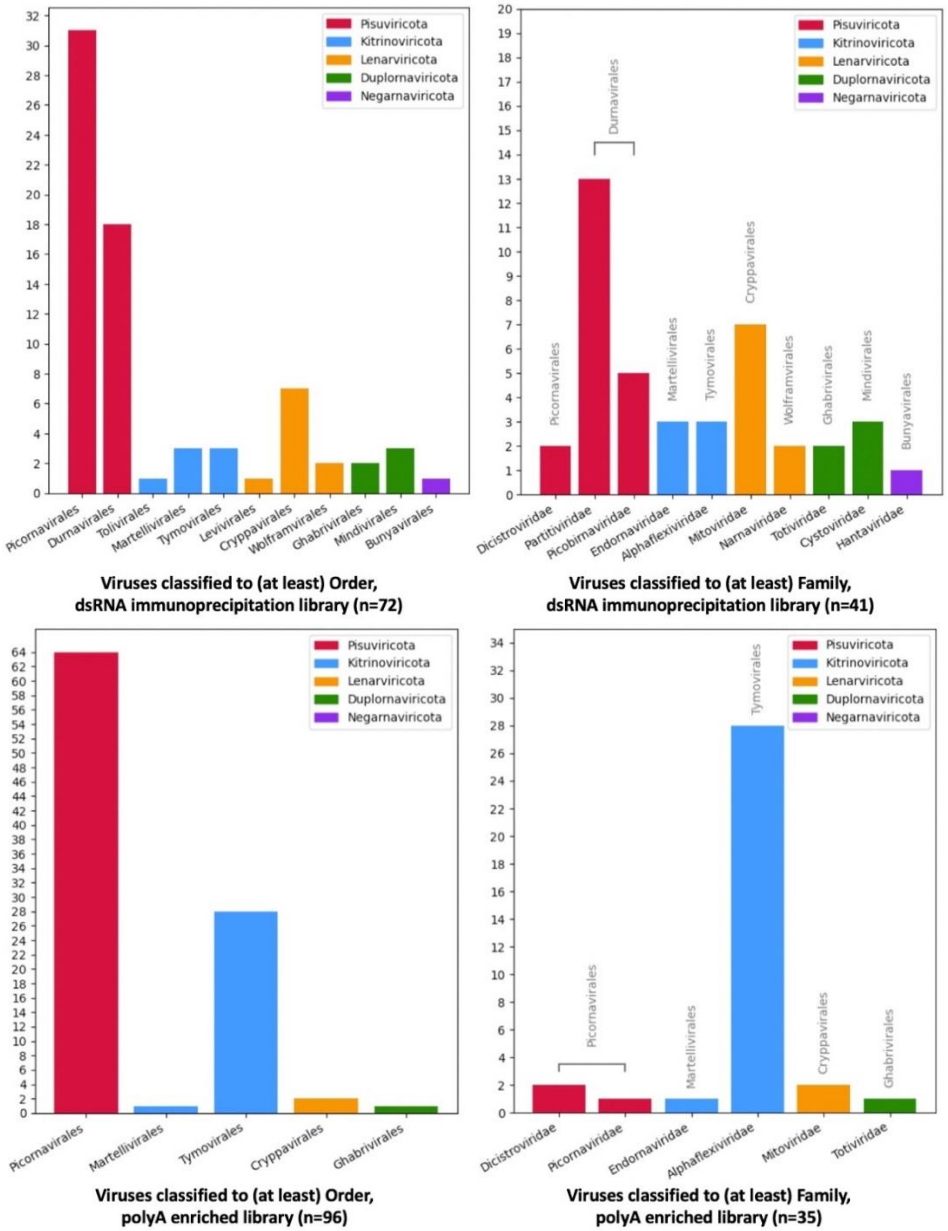

Figure 11. Order and Family-level classification results for identified RdRp sequences based on best match to RdRP-scan database sequences. Interestingly, many of the classifiable RdRp sequences had best hits to the Picornavirales, a group of viruses that was differentially abundant in diseased tissues, and found in all tissue types for Colpophyllia natans, Montastrea cavernosa and Pseudodiploria strigosa, in this study. Some RdRp sequences were also classified to Tymovirales, a filamentous virus Order. The successful classification of RdRp sequences by this approach is heavily dependent on the quality of the reference RdRp database used; lack of classification of RdRps to some viral Orders or Families should not be taken as proof of their absence from samples.

Further insights on the classification and relatedness of these viruses can be inferred from phylogenetic reconstructions of the *RdRp* gene. Best models have currently been selected for RdRp sequences from both the dsRNA immunoprecipitation and the polyA enriched libraries. At the time of writing this report, IQTREE2 is still in the process of constructing the phylogenies.

3.5 Task 3a. Sample preparation of ISH slides (n = 630). Immunostaining and microscopy imaging of *O. faveolata* samples resulting in 432 microscopy images (includes raw & processed).

Along with the genomic work, we aimed to conduct immunostaining and microscopy to identify possible viral particles in SCTLD corals. All slides for all coral species and sample sets were prepared (n=630 slides). A total of 108 of these slides (including sample sets from *Colpophyllia natans, Montastrea cavernosa* and *Orbicella faveolata)* underwent immunostaining (dsRIF) and microscopy imaging (at three different wavelengths), generating a total of 324 raw images. Spectral unmixing was performed at a single wavelength for each raw image (either the 561 or 405 channel, depending on which fluorophore was used), resulting in a total of 108 processed images (e.g., Figure 12). Thus, 432 microscopy images (324 raw + 108 processed) were generated. Three species (*C. natans, M. cavernosa* and *O. faveolata*) were imaged (rather than just *O. faveolata* as originally proposed) to prioritize samples in which Dr. Thierry Work (USGS) and colleagues had previously observed AVLPs (Work et al. 2021).

Spectral unmixing was attempted, but there were no discernible differences in the spectral emission profiles between RNase III-treated and non-RNase III-treated thin sections for any given coral sample analyzed (Figure 12). Thus, when we attempted to subtract autofluorescence for the coral and Symbiodiniaceae for each image (based on values derived from the control RNAse+ slides, 'Middle Section' of Figure 12) from the dsRNA-derived fluorescence on non-RNAse treated slides ('Top' and 'Bottom' Sections of Figure 12), no dsRNA-derived signal was ever detected from any sample (spectral profiles in rightmost column of Figure 12). Given this, we cannot determine whether the RNase III treatment actually degraded dsRNA on the coral thin section system, as it has been shown to successfully do in free-living dinoflagellates (Coy et al. 2023) and/or whether dsRNA was present in the coral holobiont tissues to provide a signal. Coral holobionts are notoriously difficult to work with using ISH approaches because of the high levels of autofluorescence they can exhibit (Loram et al. 2007). Thus, this outcome, although disappointing, was not entirely surprising.

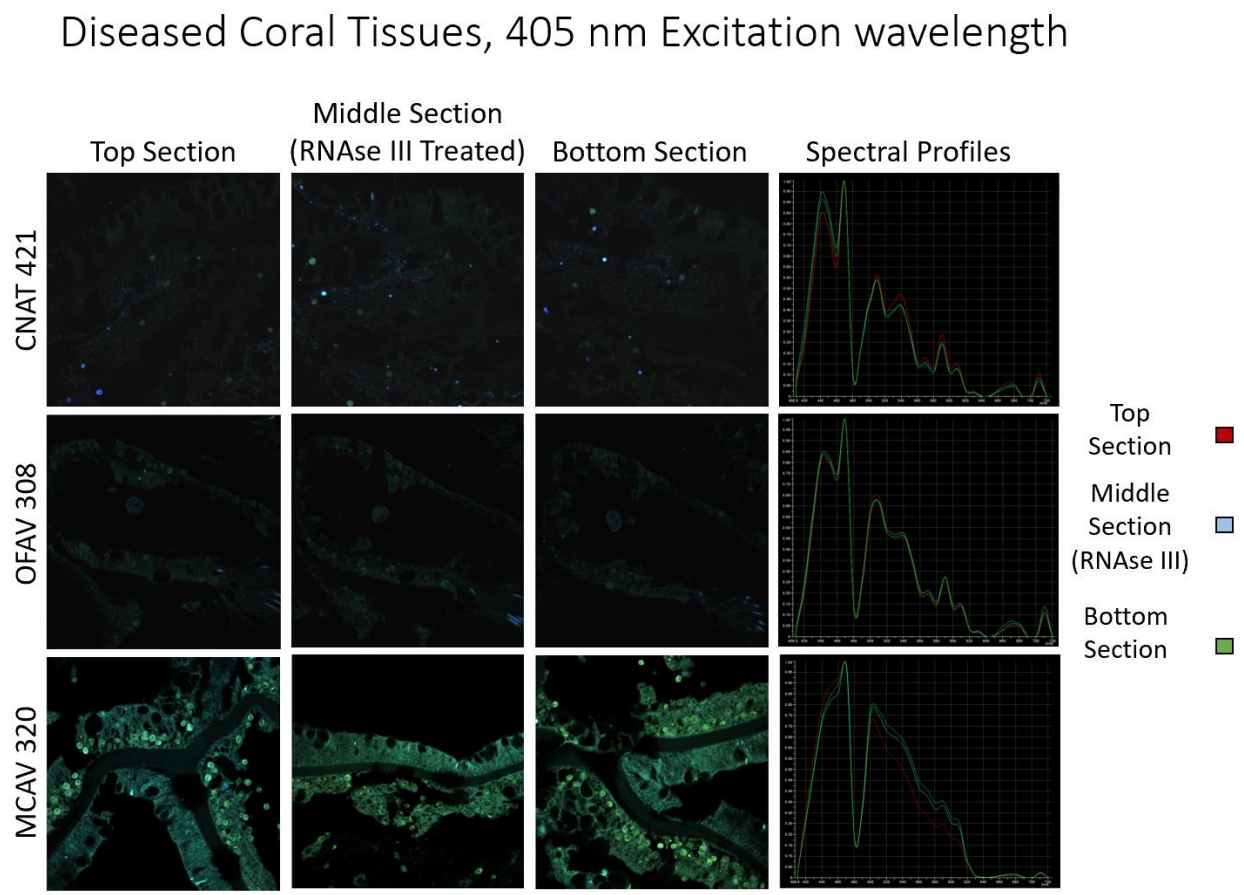

Figure 12. 40X images and spectral profiles of diseased Colpophyllia natans (CNAT), Orbicella faveolata (OFAV) and Montastrea cavernosa (MCAV) coral tissues at an excitation wavelength of 405 nm. Top, middle and bottom sections represent sections of coral tissue that were taken in tandem. Only the middle section was treated with RNAse III. Spectral profiles show the intensity of signal (Y-axis, arbitrary units) at a given wavelength (X-axis, nanometers) for the whole image of each of the tissue sections represented in this figure.

4 DISCUSSION

This project created a remarkable comparative and holistic dataset of 428 coral meta-transcriptomes, 107 Symbiodiniaceae ITS2 amplicon libraries, 630 coral thin section slides, and 324 slide images (raw and processed), to determine the interacting roles of coral host and symbiont identifies, viral consortia, and SCTLD susceptibility and infection status. We have completed all the primary tasks of 1) isolating nucleic acids, 2) generating four comparative metatranscriptomic libraries per sample, based on four different library preparation approaches, 3) created Symbiodiniaceae ITS amplicon libraries and profiles for all 107 samples, and 4) curated and 5) compared all metatranscriptomic sequence data. We also attempted a novel approach for detecting a molecular biomarker of RNA virus infections in coral holobiont tissues, by immunostaining 630 thin section slides of coral tissues for dsRNA signals.

In this report, we present evidence that there are some viral Orders including two +ssRNA viruses (Tolivirales and Picornavirales), the NCLDVs (Algavirales and Imitervirales),

and a dsRNA virus (Durnavirales) enriched in diseased specimens suggesting they play a direct or indirect role in SCTLD. Interestingly, Beavers et al. (Beavers et al. 2023), who assessed cellular organism expression in SCTLD-affected corals, identified across SCTLD disease state and coral species, the upregulation of *Abce1*, a negative regulator of RNase L, a protein shown to reduce virus propagation via genome degradation (Drappier & Michiels, 2015). Human-infecting +ssRNA viruses have been shown to inhibit and evade RNase L antiviral mechanisms through the upregulation of host *Abce1 (*Drappier and Michiels 2015). One such virus, the encephalomyocarditis-virus (Martinand et al. 1998) is a +ssRNA virus that belongs to Picornavirales, an Order that was consistently differentially abundant in diseased tissues in all species investigated here.

Many of the viral groups recovered here have been previously implicated in disease and bleaching, but to date, no study has been as comprehensive as this one, in which we conducted extremely deep sequencing of 107 libraries using 4 different preparation approaches per sample. We additionally identified differentially abundant contigs associated with filamentous viruses, but they were not among the most enriched viral groups in this study. Although genome-level analysis of these sequences expanded our understanding of the diversity and potential characteristics of filamentous virus genomes associated with apparently healthy and SCTLD-affected corals, the highly divergent nature of the recovered Patatavirales and Tymovirales sequences recovered in this study, compounded with the current overall paucity of information about these viral groups, made them difficult to fully resolve. It is thus essential for researchers to continue gathering presence/absence and prevalence information for these filamentous viruses, as well as the Tolivirales, Picornavirales, Algalvirales, Imitervirales and Durnavirales, particularly under "normal" conditions on the reef. This will provide context to better understand the potential role of these viruses in diseases such as SCTLD, and ultimately, inform coral conservation and mitigation strategies.

Along with our omics-based work, we explored a new technique, immunostaining thin sections of coral tissues to identify viral dsRNA signals. Although dsRNA immunofluorescence (dsRIF) staining remains a promising approach (Coy et al. 2023), challenges related to the high amount of autofluorescence in both corals and Symbiodiniaceae prohibited the successful application of of this technique to assessing the potential role of viral infections in SCTLD. We found that another challenge in applying this approach to corals was that a small subset of thin-sectioned tissue must be selected for analysis across the RNAse-treated and untreated slides for each tissue sample. It is time-consuming and sometimes difficult to track a given subset of thin section across treated and untreated slides during confocal imaging. This approach also leaves the vast majority of the coral tissue unanalyzed. Thus, if dsRNA signal is present anywhere outside of the small, 40X image that is taken for dsRIF, that dsRNA signal will remain undetected. Future efforts to develop an immunohistochemistry approach for the detection of biomarkers of viral infection could circumvent these two challenges. Immunohistochemistry would detect dsRNA through chemical deposition of brown dye that is clearly visible in standard bright-field microscopy (circumventing autofluorescence), and would allow whole-slide scanning of each coral sample (enabling all tissue on each slide to be imaged and analyzed).

In summary, the work describes here provides important early information about the potential role of viruses in SCTLD. With these data confirming that some viral Orders (a high

level annotation) are increased in abundance in the diseased tissues of SCTLD-affected corals, and sometimes prevalent across coral health states, we aim to continue digging deeper and refining our analyses to determine if any one or two types of viral genomes associated with the disease can be fully assembled (e.g., Veglia et al 2022). With these kinds of future data we can 1) generate markers to track absolute abundance of viral targets, 2) localize viral targets within tissues using staining and microscopy techniques, and potentially 3) isolate and culture these viruses for more conclusive experiments regarding their role in coral health and disease. While metatranscriptomics can be an extremely valuable tool to identify potential virus sequences that are associated with the disease, follow up experiments are needed to confirm the etiology of the disease. Viral activity can increase as secondary infections during disease and may exacerbate disease rather than induce it. Regardless, it is clear from these data that increased viral activity is strongly associated with diseased tissues in SCTLD-affected colonies, and this work is a critical aspect of defining a potential etiological agent for SCTLD.

Importantly, **our comparative approach will allow us to provide critical advice to the DEP and other coral disease researchers about whether viruses or their abundances are associated with SCTLD.** These efforts will inform disease intervention and management efforts throughout Florida's Coral Reef. The identification of a pathogen or pathogens associated with SCTLD will also facilitate the development of diagnostic methods such as quantitative PCR primers specific to the pathogen, as well as improved intervention strategies such as targeted antibiotic or antiviral treatments.

References

Beavers, Kelsey M., Emily W. Van Buren, Ashley M. Rossin, Madison A. Emery, Alex J. Veglia, Carly E. Karrick, Nicholas J. MacKnight, et al. 2023. "Stony Coral Tissue Loss Disease Induces Transcriptional Signatures of in Situ Degradation of Dysfunctional Symbiodiniaceae." *Nature Communications* 14 (1): 2915.

Buchfink, Benjamin, Klaus Reuter, and Hajk-Georg Drost. 2021. "Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND." *Nature Methods* 18 (4): 366–68.

Bushnell, Brian. BBMap: A Fast, Accurate, Splice-Aware Aligner. United States: N. p., 2014.

Clarke, K.R. and Gorley, R.N. (2015) PRIMER v7: User Manual/Tutorial. PRIMER-EPlymouth.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10 (December): 421.

Charon, Justine, Jan P. Buchmann, Sabrina Sadiq, and Edward C. Holmes. 2022. "RdRp-Scan: A Bioinformatic Resource to Identify and Annotate Divergent RNA Viruses in Metagenomic Sequence Data." *Virus Evolution* 8 (2): veac082.

Correa, Adrienne M. S., Tracy D. Ainsworth, Stephanie M. Rosales, Andrew R. Thurber, Christopher R. Butler, and Rebecca L. Vega Thurber. 2016. "Viral Outbreak in Corals Associated with an In Situ Bleaching Event: Atypical Herpes-Like Viruses and a New Megavirus Infecting Symbiodinium." *Frontiers in Microbiology* 7 (February): 127.

Coy, Samantha R., Budi Utama, James W. Spurlin, Julia G. Kim, Harshavardhan Deshmukh, Peter Lwigale, Keizo Nagasaki, and Adrienne M. S. Correa. 2023. "Visualization of RNA Virus Infection in a Marine Protist with a Universal Biomarker." *Scientific Reports* 13 (1): 5813.

Darling, Aaron C. E., Bob Mau, Frederick R. Blattner, and Nicole T. Perna. 2004. "Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements." *Genome Research* 14 (7): 1394–1403.

Drappier, Melissa, and Thomas Michiels. 2015. "Inhibition of the OAS/RNase L Pathway by Viruses." *Current Opinion in Virology* 15 (December): 19–26.

Edgar, R. C. (2021). MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. BioRxiv, 2021-06.

Finke, Jan F., Colleen T. E. Kellogg, and Curtis A. Suttle. 2023. "Deep6: Classification of Metatranscriptomic Sequences into Cellular Empires and Viral Realms Using Deep Learning Models." *Microbiology Resource Announcements* 12 (2): e0107922.

Gunjal, Aparna, and Sonali Shinde. 2021. *Microbial Diversity and Ecology in Hotspots*. Academic Press.

Guo, Jiarong, Ben Bolduc, Ahmed A. Zayed, Arvind Varsani, Guillermo Dominguez-Huerta, Tom O. Delmont, Akbar Adjie Pratama, et al. 2021. "VirSorter2: A Multi-Classifier, Expert-Guided Approach to Detect Diverse DNA and RNA Viruses." *Microbiome* 9 (1): 37.

Howe-Kerr, Lauren I., Carsten G. B. Grupstra, Kristen M. Rabbitt, Dennis Conetta, Samantha R. Coy, J. Grace Klinges, Rebecca L. Maher, et al. 2023. "Viruses of a Key Coral Symbiont Exhibit Temperature-Driven Productivity across a Reefscape." *ISME Communications* 3 (1): 27.

Hyatt, Doug, Gwo-Liang Chen, Philip F. Locascio, Miriam L. Land, Frank W. Larimer, and Loren J. Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11 (March): 119.

Inoue-Nagata, A. K., Jordan, R., Kreuze, J., Li, F., López-Moya, J. J., Mäkinen, K., ... & ICTV Report Consortium. (2022). ICTV virus taxonomy profile: Potyviridae 2022. Journal of General Virology, 103(5), 001738.

Kalyaanamoorthy, Subha, Bui Quang Minh, Thomas K. F. Wong, Arndt von Haeseler, and Lars S. Jermiin. 2017. “ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates.” *Nature Methods* 14 (6): 587–89.

Loram, J. E., N. Boonham, P. O'Toole, H. G. Trapido-Rosenthal and A. E. Douglas (2007). "Molecular quantification of symbiotic dinoflagellate algae of the genus *Symbiodinium*." Biological Bulletin 212: 259-268.

Martinand, C., T. Salehzada, M. Silhol, B. Lebleu, and C. Bisbal. 1998. “The RNase L Inhibitor (RLI) Is Induced by Double-Stranded RNA.” *Journal of Interferon & Cytokine Research: The Official Journal of the International Society for Interferon and Cytokine Research* 18 (12): 1031–38.

Minh, Bui Quang, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear. 2020. “IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.” *Molecular Biology and Evolution* 37 (5): 1530–34.

Nayfach, Stephen, Antonio Pedro Camargo, Frederik Schulz, Emiley Eloe-Fadrosh, Simon Roux, and Nikos C. Kyrpides. 2021. “CheckV Assesses the Quality and Completeness of Metagenome-Assembled Viral Genomes.” *Nature Biotechnology* 39 (5): 578–85.

Soffer, Nitzan, Marilyn E. Brandt, Adrienne M. S. Correa, Tyler B. Smith, and Rebecca Vega Thurber. 2014. “Potential Role of Viruses in White Plague Coral Disease.” *The ISME Journal* 8 (2): 271–83.

Thurber, Rebecca Vega, Jérôme P. Payet, Andrew R. Thurber, and Adrienne M. S. Correa. 2017. “Virus-Host Interactions and Their Roles in Coral Reef Health and Disease.” *Nature Reviews. Microbiology* 15 (4): 205–16.

Vega Thurber, Rebecca, Emily R. Schmeltzer, Andréa G. Grottoli, Robert van Woesik, Robert J. Toonen, Mark Warner, Kerri L. Dobson, et al. 2022. “Unified Methods in Collecting, Preserving, and Archiving Coral Bleaching and Restoration Specimens to Increase Sample Utility and Interdisciplinary Collaboration.” *PeerJ* 10 (November): e14176.

Veglia, A. J., K. Beavers, E. W. Van Buren, S. S. Meiling, E. M. Muller, T. B. Smith, D. M. Holstein, et al. 2022. “Alphaflexivirus Genomes in Stony Coral Tissue Loss Disease-Affected, Disease-Exposed, and Disease-Unexposed Coral Colonies in the U.S. Virgin Islands.” *Microbiology Resource Announcements* 11 (2): e0119921.

Veglia, Alex Joseph. "Detecting and interpreting viral dynamics in marine invertebrate holobionts." (2023) Diss., Rice University. https://hdl.handle.net/1911/115214.

Wood-Charlson, Elisha M., Karen D. Weynberg, Curtis A. Suttle, Simon Roux, and Madeleine J. H. van Oppen. 2015. “Metagenomic Characterization of Viral Communities in Corals: Mining Biological Signal from Methodological Noise.” *Environmental Microbiology* 17 (10): 3440–49.

Work, T. M., T. M. Weatherby, J. H. Landsberg, Y. Kiryu, S. M. Cook and E. C. Peters (2021). "Viral-Like Particles Are Associated With Endosymbiont Pathology in Florida Corals Affected by Stony Coral Tissue Loss Disease." Frontiers in Marine Science 8 (1651)