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, Alex Veglia, Eddie Fuques

Oregon State University, Rice University

06/06/2022

Completed in Fulfillment of BA002F – 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. 2022. 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 linkages between RNA or DNA viruses and SCTLD infection status. We generated 428 meta-transcriptomic libraries (4 independent preparations of 107 samples) and acquired sequence data for half. We generated ITS2 amplicon libraries for all 107 specimens to identify Symbiodiniaceae lineages and test for linkages among symbiont, viral diversity/abundance, and coral disease status.

We began development of novel informatic methods to quantitatively track what viruses are present and/or variably abundant in our samples. Our preliminary results indicate variability in the abundance and diversity of different viral groups and Symbiodiniaceae in SCTLD susceptible corals. Diseased specimen tended to have higher viral diversity. Samples previously shown to contain filamentous virus-like particles (VLPs) were dominated by mixtures of Symbiodiniaceae. Symbiont lineages within *Breviolum* also varied with SCTLD susceptibility. Although not conclusive, these early data provide foundational knowledge about connections between Floridian coral species SCTLD status and viral pathogens.

The outcomes of this project will be incorporated into an on-going 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 (max 1 page)

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-infected and apparently healthy Florida corals in order to determine whether there is an association between Symbiodiniaceae and viral genetic diversity, and SCTLD infection. In particular, there were three primary tasks to be completed sequentially from the samples**: 1) nucleic acid isolation, 2) processing and sequencing of nucleic acids from all samples, 3) preliminary data curation, analysis, interpretation, and dissemination of the resulting data.** In the Correa lab (Rice University), all of the samples have undergone RNA and DNA isolation. Both teams processed their total RNA aliquots using independent methods. The Vega Thurber Lab (OSU) conducted rRNA removal and small RNA enrichment metatranscriptomic library preparation approaches, whereas the Correa Lab conducted polyA enrichment and immunoprecipitation of dsRNA. The Correa Lab isolated, sequenced and analyzed the Symbiodiniaceae ITS2 amplicon data. We have thus generated 428 meta-transcriptomic libraries, 214 at each university. Further, the Correa lab isolated DNA from each specimen and generated ITS2 amplicon libraries for sequencing at OSU.

From these high throughput sequencing libraries, Symbiodiniaceae ITS2 profiles, as well as the polyA enrichment and immunoprecipitated dsRNA meta-transcriptomic libraries were fully sequenced. Preliminary analyses are now underway. However, while all samples were prepared for rRNA removal and small RNA enrichment at OSU, delays at the sequencing center in Oregon resulted in only 2 of the 214 samples being sequenced, although we have received confirmation that they should be sequenced very soon (end of June). While waiting for the data, we developed informatic and visualization methodologies to accelerate the project. We aim to use the genomes of the corals and algal symbionts to accelerate and streamline our viral analysis.

Results are highly preliminary, but include the identification of Symbiodiniaceae genera dominating sampled coral colonies. Colonies previously identified to contain signatures of filamentous virus-like particles (VLPs) based on transmission electron microscopy (TEM) were dominated by either *Breviolum*, *Cladocopium*, *Durusdinium* or mixtures thereof. Therefore, filamentous VLPs are not exclusive to corals dominated by a particular Symbiodiniaceae genus. A preliminary phylogenetic analysis identified 18 nonredundant sequences with strong similarity to RNA viruses, including 8 filamentous viruses (e.g., alphaflexiviruses, betaflexiviruses), based on the RNA-dependent RNA Polymerase (RdRp) gene. An additional Kmer mapping approach demonstrated that the polyA and dsRNA immunoprecipitation library preparation approaches each enrich for different types of sequence reads. The polyA enrichment method recovered 10-fold more hits to virus-like sequences than the dsRNA immunoprecipitation approach. PolyA enrichment resulted in higher detected viral richness, but dsRNA immunoprecipitation resolved stronger differences in richness between tissue health states (disease lesion or DL, disease healthy or DH, healthy healthy or HH), with DL having the highest richness and HH having the lowest. Additional comparative analyses for each metatranscriptome approach, coral species, tissue type and symbiont identity should be applied in the future.

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.

Table of Contents

List of Figures:

List of Tables:

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 are reminiscent of filamentous RNA-based viruses typically associated with plant diseases. Other researchers have found evidence of similar viral particles in Symbiodiniaceae. 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-infected and apparently healthy Florida corals in order to determine whether there is an association between Symbiodiniaceae and viral genetic diversity, and SCTLD infection.

We hypothesize that while all corals will contain some viral signatures, SCTLD lesion sites will contain more abundant or different viral signatures such as the presence of the +ssRNA viruses. These signatures may include RNA viral genomes similar to other dinoflagellate viruses or ones unique to this coral disease. However, using this approach we are also comparing viral signatures to host physiology by preparing and sequencing RNA in ways that will provide gene transcripts from all members of the coral holobiont (bacterial, viral, animal, dinoflagellate, fungi and others). By comparing coral lesions experiencing active SCTLD with apparently healthy corals, we gain more definitive evidence on 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 FISH probes specific to the pathogen as well as improved intervention strategies including targeted antibiotic or antiviral treatments.

To date we have isolated nucleic acids (RNA and DNA) for library preparation (Table 1) and sequencing to explicitly analyze for the presence of viral signatures (genomes and transcripts) that may provide further evidence that viruses play a significant role in disease induction or exacerbation (Figures 12-14). Symbiodiniaceae in each sample were also identified in order to test for links between symbiont identity and viral presence/absence and differential abundance (Figure 11). This report details preliminary findings for symbiont identity, and viral diversity recovered from two library preparation methods (polyA and dsRNA immunoprecipitation).

2. METHODS

2.1. Task 1

In order to address which viral sequences and genomes are associated with SCTLD afflicted corals we processed 107 previously collected samples for the isolation of both viral RNA and symbiont DNA. We recovered significant amounts of high-quality RNA and DNA as measured by bioanalyzer (Table 1 for details). Below is a description of these nucleic acid isolation and purification methods.

RNA extraction and isolation (viruses):

At Rice University, coral tissue slurries were quickly thawed. A 15 ml falcon tube containing sterile glass beads was used to create a 1:1:1:1 mixture of coral tissue slurry, RNA lysis buffer containing the RNase-deactivating compound 2-mercaptoethanol, acid-phenol, and chloroform. Samples were bead-beaten at max speed for 10 minutes. Phase separation was achieved via centrifugation, and the RNA from the aqueous layer was precipitated using molecular grade ethanol (final: 70%). Following resuspension in DNA lysis buffer, the concentrated RNA sample was Turbo DNase treated, and re-purified with two rounds of acid-phenol chloroform treatment, followed by two chloroform treatments. The RNA-containing aqueous layer was again purified by ethanol precipitation, with the RNA pellet resuspended in 100ul of RNase/DNase free water. These samples were stored at -80^{o}C with Ribolock Inhibitor according to manufacturer's directions. Aliquots of 25ul were sent to OSU for metatranscriptome library preparation and sequencing as defined in the grant; another 25ul per sample was retained at Rice U. for polyA enrichment or immunoprecipitation of dsRNA prior to library preparation, as defined in the scope of work.

DNA extraction and isolation (Symbiodiniaceae):

At Rice University coral tissue slurries were quickly thawed, and DNA was extracted using a ZymoBIOMICs DNA/RNA Kit (ZymoResearch) following the manufacturer's instructions. To identify dominant Symbiodiniaceae lineages, the internal transcribed spacer-2 (ITS-2) region of nuclear ribosomal RNA gene was amplified from the unfractionated DNA. Amplification was conducted by OSU's Center for Genome Research and Biotechnology using the SYM_VAR_5.8SII and SYM_VAR_REV primer set (Hume et al. 2018). PCR clean-up was completed using Agencourt AMPure XP Magnetic Beads. The resulting PCR product was purified with Agencourt AMPure XP Magnetic Beads, quantified via qPCR using the KAPA library quantification kit (Roche Sequencing Solutions, Pleasanton, CA), pooled in equal molar amounts, and sequenced on the Illumina MiSeq platform with PE300 chemistry.

2.2. Task 2

2.2.1. Sequencing Library Preparation:

Aliquots of all total RNA preparations were shipped from the Correa lab to the Vega Thurber lab at Oregon State University. We agreed in the proposal that each lab would independently prepare 2 sets of libraries using different approaches; the Vega Thurber Lab would conduct rRNA removal and small RNA enrichment approaches to contrast with Dr. Correa's lab's approach of polyA enrichment and immunoprecipitation. In this way we could holistically cover all approaches avoiding bias in the outcomes (Figure 1).

The Correa lab conducted dsRNA immunoprecipitation in-house using the J2 antibody (Absolute Antibody Ab01299-2.0) paired with a DynabeadsA Immunoprecipitation Kit. Aliquots of total RNA and enriched dsRNA were sent to Genohub to be sequenced for RNA viruses. Sequencing has been completed and sequence libraries have been received.

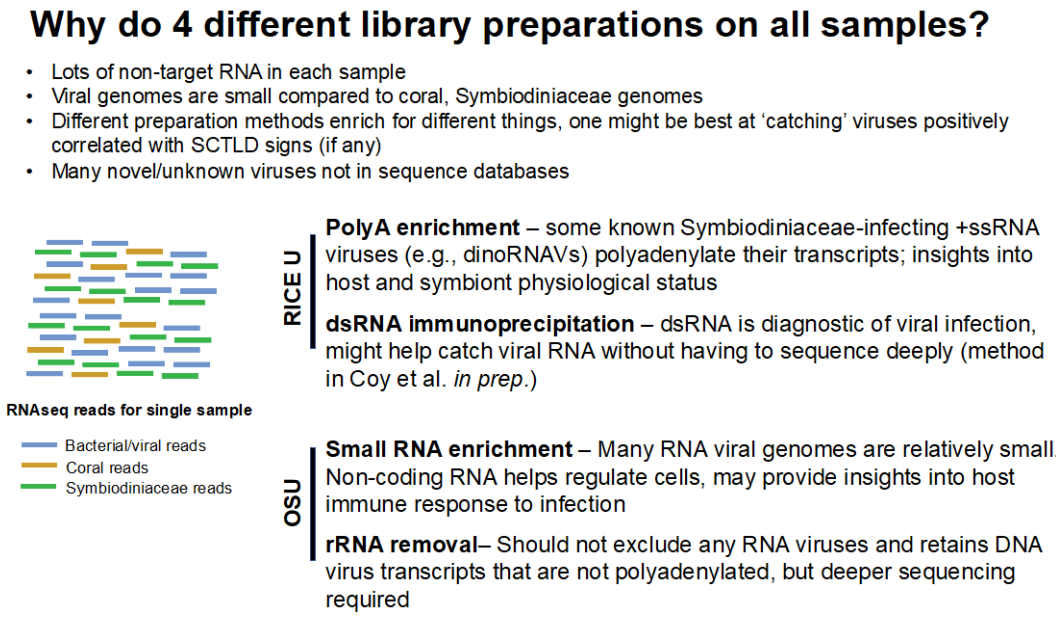


Figure 1 Rationale for holistic RNA preparation. We chose a comprehensive 4 methodological approach to explore the potential role of viruses in SCTLD. The two labs selected 2 different approaches that can acquire similar but unique fractions of sequences from the same specimen.

2.2.2. Meta-transcriptomic Sequencing Approaches

Aside from dsRNA enrichment, which was performed in house at Rice University, RNA-Seq library prep and sequencing was conducted by Quick Biology Inc. (Pasadena, CA) as an affiliate of Genohub. For dsRNA enriched samples, RNA-Seq libraries were prepared using the KAPA RNA Hyper Prep Kit. PolyA selected libraries were enriched from total RNA and prepared using a KAPA Stranded mRNA-seq Kit. Paired-end sequencing was conducted on the NovaSeq platform with a targeted length of 150 base-pairs for each library set, with a targeted read depth of 30 and 50 million reads per read direction for dsRNA and polyA prepared libraries, respectively. This read length and depth is standard for proper assembly and detection of viral sequences in RNA libraries, given viruses normally comprise a small fraction of the total library.

In summary, the dsRNA approach resulted in a total of 9,243,071,100 reads for the entire 107 libraries (Supplemental Table 1). After initial quality control there remained a total of 9,142,655,870 (99%) reads for further analysis. The polyA selected sequence libraries resulted in 12,527,174,192 number of reads (Supplemental Table 1), and after initial quality control there remained 12,090,807,628 reads (97%).

The Vega Thurber lab sent their RNA preparations to the Genomics Core (GC3G) at University of Oregon on February 8th, and processing began on Feb 28th. All 110 RNA samples were processed by two different approaches: rRNA depletion and small RNA enrichment. rRNA removal was performed by using the Illumina Ribo-Zero Plus rRNA Depletion kit. Size selection for small fragments and quality assessment of the libraries was done in a Fragment Analyzer equipment (Advanced Analytical Technologies). Library preparation was performed with Illumina Stranded Total RNA Prep kit. The

concentration of the prepared libraries was measured with a Qubit 3.0 Fluorometer. To evaluate the quality of these libraries, two random samples were selected for a MiSeq Nano testing run (PE 2x150), with an estimated depth of ~1 million total reads per sample. Data obtained is being analyzed to evaluate the rate of PCR-duplicates, as well as running them through the pipeline originally designed. To date the samples have thus been processed to the metatranscriptomic library stage but are awaiting sequencing.

2.2.3. ITS-2 Amplicon Generation and Sequencing Approaches:

Aliquots of all DNA extractions were sent from Rice to Oregon State University's Center for Genome Research and Biocomputing for amplification and sequencing of the Symbiodiniaceae ITS2 gene region, a marker often used to compare symbiont communities in coral specimen. We used our previously successful approach (Howe-Kerr et al., 2020). In brief, the ITS-2 region of Symbiodiniaceae rDNA was amplified using symbiont-specific primers: SYM_VAR_5.8SII (5′-GAATTGCAGAACTCCGTGAACC-3′) and SYM_VAR_REV (5′-CGGGTTCWCTTGTYTGACTTCATGC-3′).

The target amplicon was approximately 234–266 bp (Hume et al., 2018). The PCR reaction contained 5 µl of DNA (5 ng/µl), 2.5 µl of SYM_VAR_5.8SII + MiSeq Adapter (5′TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GAATTGCAGAACTCCGTGAACC-3′; 2 µM), 2.5 µl of SYM_VAR_REV + MiSeq Adapter (5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG CGGGTTCWCTTGTYTGACTTCATGC-3′; 2 µM), 12.5µl 2× KAPA HiFi HotStart ReadyMix and 2.5 µl of molecular grade water for a total reaction volume of 25 µl. PCR cycles were as follows: 95°C for 3 min, 15 cycles of 95°C for 30 s, 56°C for 30 s, and 72°C for 30 s, and 72°C for 4 min. PCR clean-up was completed using Agencourt AMPure XP Magnetic Beads. Illumina indexing primers were added to 50 µl of purified PCR product, and a new PCR was run to incorporate unique barcodes for each sample. The PCR reaction contained 5 µl of cleaned PCR product, 5 µl of Illumina Indexed Primer 1 (i5), 5 µl of Illumina Indexed Primer 2 (i7), 25 µl 2× KAPA HiFi HotStart, and 10 µl molecular grade water for a total reaction volume of 50 µl. PCR cycles were as follows: 95°C for 3 min, 20 cycles of 95°C for 30 s, 56°C for 30 s, and 72°C for 30 s, and lastly 72°C for 4 min. The resulting PCR product was purified with Agencourt AMPure XP Magnetic Beads. Samples were quantified via qPCR using the KAPA library quantification kit and normalized and pooled in equal molar amounts.

Sequencing of amplicon pools has been completed and all libraries have been received (Supplemental Table 1). In summary, a total of 19,478,672 reads were generated and after initial quality control 18,041,092 (92.6%) were used for downstream analysis. On average, each ITS-2 amplicon library had 173,703 reads for analysis.

2.3. Task 3 Data Analytic Methodologies and Pipeline Generation

2.3.1. Overall Approach to Data Analytics

Each team has begun independent curation and analysis of their resulting data. While awaiting the sequences from the University of Oregon genomics core, the Vega Thurber lab has spent a majority of its time curating genomic resources and developing and benchmarking its approaches using other coral meta-transcriptomic datasets. We created a flexible and adaptive approach to the analysis that can use different portions of

the data to identify unique aspects of disease etiology. In the simplest version (Figure 2) of our approaches, we first quality control our data. Quality control of the raw reads is typically performed with FastQC (Andrews 2010). To visualize all the samples in a single report, FastQC results are run through MultiQC (Ewels et al. 2016). Trimming of the raw data was conducted using Trim Galore (https://github.com/FelixKrueger/TrimGalore), which is a wrapper around Cutadapt (Martin 2011) and FastQC. Trim Galore can be employed for both adapter (Illumina's universal adapter automatically detected by this tool) and quality trimming, with a quality Phred score cutoff and a minimum sequence length requirement. Quality control of the fastq files obtained after the trimming process are also assessed using FastQC and MultiQC. Once sequences pass QC, we then can take two basic informatic approaches where we remove any sequences similar to host and symbiont genes using their respective reference genomes and then either a) analyze only the remaining sequences through de novo assembly or use targeted viral genomes (e.g., the CHFV coral associated genome) or genes (e.g., the RDRP) for reference-based assembly. Both of these approaches are discussed in the following results sections.

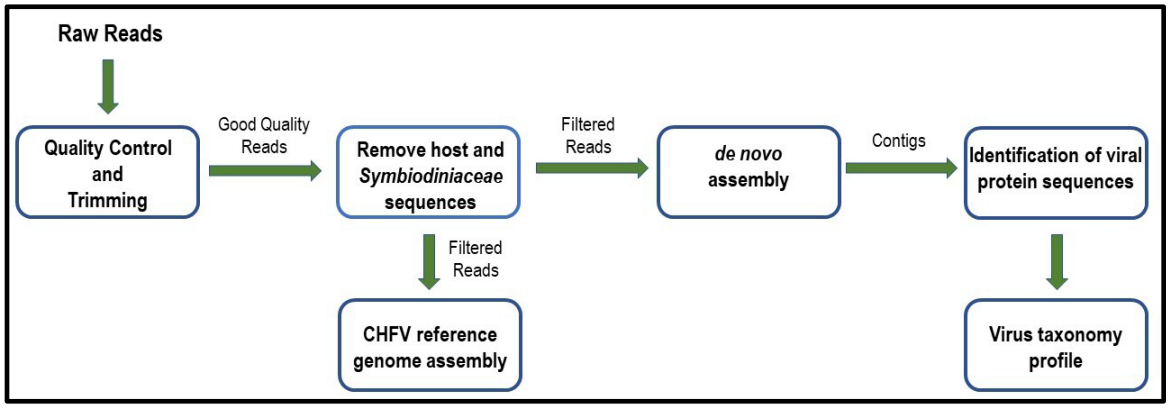


Figure 2. Summary of Meta-transcriptomic Sequence Processing & Analysis. Raw sequence reads undergo extensive quality control prior to any downstream steps such as the removal of off target (i.e. non-viral) sequences using host and symbiont genomes, assembly, and sequence annotation and profiling. QC reads can also be used for reference-based genome assembles to viral targets like the CHFV genome shown to be associated with SCTLD infected corals.

2.3.2. Symbiodiniaceae amplicon sequence analyses

Symbiodiniaceae ITS-2 amplicon sequencing libraries were analyzed using the SymPortal analytical framework (Hume et al. 2018; Figure 3). The multicopy nature of the Symbiodiniaceae ITS2 marker gene can lead to over estimations of community diversity within coral tissue. To circumvent this, SymPortal generates ITS-2 type profiles that bin marker sequences to most accurately represent Symbiodiniaceae community composition. The SymPortal-produced ITS2 type profile counts tables were processed (e.g., converted to relative abundance values) in R using the phyloseq (McMurdie and Holmes 2013). Symbiodiniaceae community composition was then visualized via relative abundance bar plots. The R package DEseq2 (Love et al. 2014) was used to identify ITS2 profiles significantly associated with the diseased coral health state.

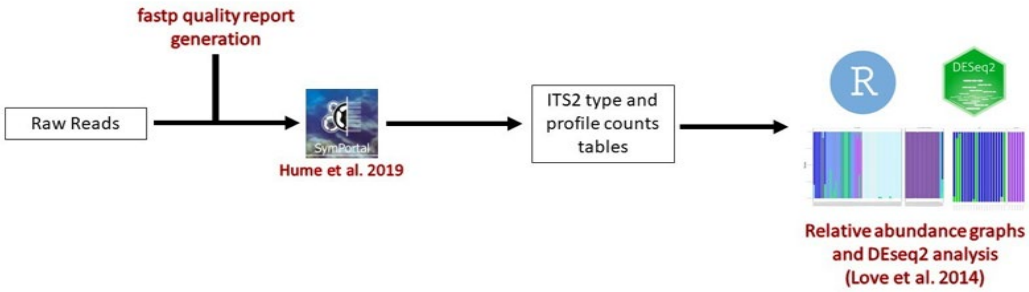
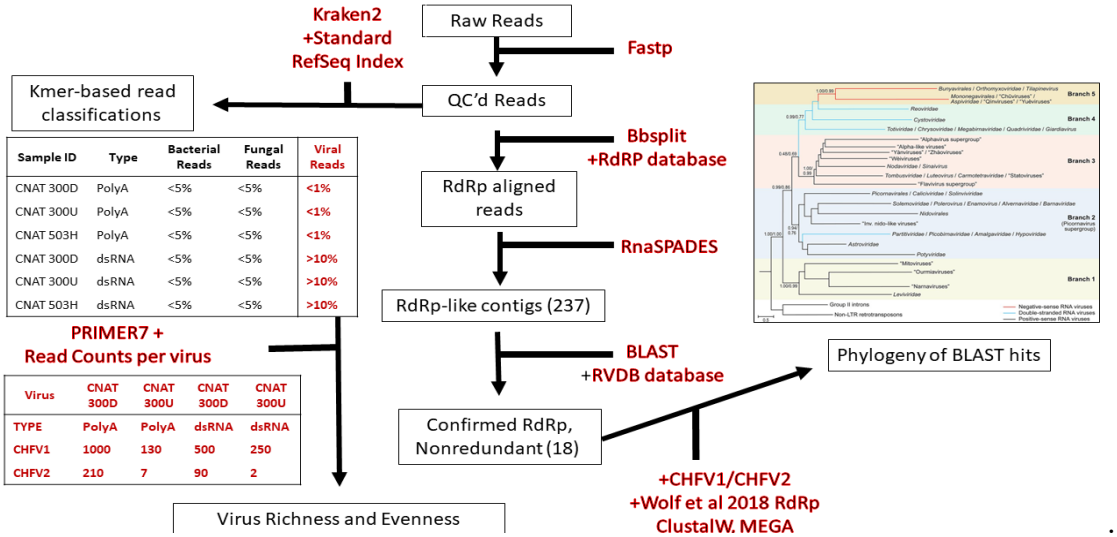


Figure 3. Summary of ITS2 Data Processing and Analysis. To analyze coral specimen Symbiodiniaceae profiles, raw ITS2 reads undergo quality control and then analysis on Symportal which provides both ITS type data and profile counts data for differential abundance analysis.

2.3.3. Metatranscriptome processing and virus diversity assessment

Upon being received from the sequencing facility, Correa lab meta-transcriptomes were quality checked using the program FastQC (v0.11.9, Andrews 2010). Any libraries showing signs of low quality (>30 phred score) were removed from downstream analyses. Adapter contamination detection/removal, over-representation/duplicate analysis, low quality read removal, and read error correction was then performed with the program fastp (v0.20.1, Chen et al., 2018) with default settings. The final quality checked and cleaned libraries were then used in preliminary assessments of virus diversity and abundance. Preliminary virus diversity assessments were done using two different approaches: 1. Both a kmer-based approach to classify reads as cellular or viral and a phylogeny-based approach generating virus marker gene sequences and assessing homology of these sequences to marker genes of known RNA viruses (Figure 4).



Sample ID	Type	Bacterial Reads	Fungal Reads	Viral Reads
CNAT 300D	PolyA	<5%	<5%	<1%
CNAT 300U	PolyA	<5%	<5%	<1%
CNAT 503H	PolyA	<5%	<5%	<1%
CNAT 300D	dsRNA	<5%	<5%	>10%
CNAT 300U	dsRNA	<5%	<5%	>10%
CNAT 503H	dsRNA	<5%	<5%	>10%

Virus	CNAT 300D	CNAT 300U	CNAT 300D	CNAT 300U
TYPE	PolyA	PolyA	dsRNA	dsRNA
CHFV1	1000	130	500	250
CHFV2	210	7	90	2

Figure 4. Summary of ongoing informatic pipeline for analysis of metatranscriptomes using phylogenetic-based and Kmer-based approaches. Red text denotes bioinformatic processes and tools used, while black text in boxes denotes output of the bioinformatic processes. RdRp = RNA Dependent RNA Polymerase, with parentheses denoting the number of RdRp-like sequences that were detected at each

step in the process. Red text highlighted in the kmer and read counts tables were used as specific inputs in paired bioinformatic processes.

2.3.3.1. Kmer-based read classification

Kmer-based read classification was done using the program Kraken2 (Wood et al 2019) and the Standard Kraken2 RefSeq index and default parameters. Kraken2 results summary reports and counts tables were then generated with the R-based program Pavian (Breitweiser and Salzberg 2020). Virus diversity and relative abundance were assessed using the RNA virus counts (identified as Riboviria) in Primer7 (Clark and Gorley 2015) using Margalef's Species Richness and Pielou's evenness index. The results were visualized using ggplot2 in R Studio.

2.3.3.2. Phylogeny-based virus diversity assessment

While there is no universal marker gene for viruses, it is known that there are shared genes within specific virus groups. For RNA viruses, the ideal marker gene is the RNA-dependent RNA polymerase (RDRP) gene, as this single gene is ubiquitously across and unique to the Ribovaria. Given this, for this project, RDRP sequences were identified, annotated, and phylogenetically compared to homologous RDRP sequences identified using BLAST. To start, RDRP-like reads were extracted from fastp cleaned metatranscriptome libraries using the program bbsplit.sh (Bushnell 2014) and a custom RDRP database containing all RDRP sequences on GenBank. The extracted RDRP reads were then concatenated into a single file and assembled into contigs using rnaSPAdes (Bushmanova et al. 2019). To check if these sequences were truly most homologous to RDRP, contigs were aligned to the protein version of the Reference Virus DataBase (RVBD) (v21; Bigot et al 2021). Sequences confirmed to be of viral RDRP origin were then used in downstream phylogenetic analyses.

Due to extreme differences in contig length and sequence divergence of viral RDRPs, we elected to build a reference tree of viral RDRP sequences that could be used as a scaffold to phylogenetically place smaller contigs. This reference tree was built using the best blast similarities to the viral RDRP sequences checked against the RVBD database. We also added RDRP sequences for CHFV1/CHFV2 identified from previous SCTLD metatranscriptomes (Veglia et al. 2022). Specifically, RDRP sequences were aligned using ClustalW (Thompson et al. 1994) and assessed for appropriate substitution tests. The recommended WAG+G3 substitution parameter was used for maximum likelihood phylogenetic reconstruction with 50 bootstraps in MEGAX (Kumar et al 2018), which was visually rendered in iTOL (Letunic and Bork 2021).

3. RESULTS

3.1. Task 1 RNA Extraction and Specimen Allocation

DNA and RNA were successfully extracted from the 107 samples (Table 1). Aliquots of RNA from each sample were sent to OSU; the other half of each extraction was retained at Rice U. DNA extractions from each sample were retained at Rice U. Yields for all extractions can be found in Table 1.

3.2. Task 2 Metatranscriptomic Sequence Library Generation and Curation

We generated a combined 428 meta-transcriptomic libraries that can be used for analysis of the role of viruses in SCTLD. The power of transcriptomics is that it can achieve several aims in a single run. First, transcriptomics can relay differences among host and symbiont physiological states based on quantification of their differentially abundant transcripts (Figure 1). Second, because viruses can contain either RNA or DNA genomes, transcriptomics provides a view of both the presence and abundance of RNA viruses (from genome sequence) and DNA virus activity (from DNA virus transcripts). Lastly, if other pathogens or microbial members of the holobiont are present, meta-transcriptomics can quantify changes in their community dynamics, and provide context to different disease states. Together these data can provide a holistic view of the changing biology occurring within different coral species during SCTLD.

3.3. Task 3

3.3.1. Symbiodiniaceae Results

We found that *Colpophyllia natans, Pseudodiploria strigosa*, and *Siderastrea siderea* coral colonies tended to be dominated Symbiodiniaceae in the genus *Breviolum*, but sometimes also contained (or occasionally were dominated by) symbionts in the genus *Durusdinium* (Figure 5). *Montastraea cavernosa* and *Orbicella faveolata* colonies tended to be dominated by *Cladocopium*. *Orbicella faveolata* contained symbionts in *Breviolum* and/or *Durusdinium*. Colonies of coral species considered to be highly susceptible to SCTLD tended to be dominated by *Breviolum*, whereas coral species considered to be moderately susceptible to SCTLD tended to be dominated by *Cladocopium*. This could indicate that some symbionts in *Breviolum* are more susceptible to SCTLD than some symbionts in *Cladocopium*. However, it is also possible that susceptibility to SCTLD is driven by coral genotype or some other factor, and the observed trends in Symbiodiniaceae dominance merely reflect specificity in the symbiotic associations formed by different coral species or populations.

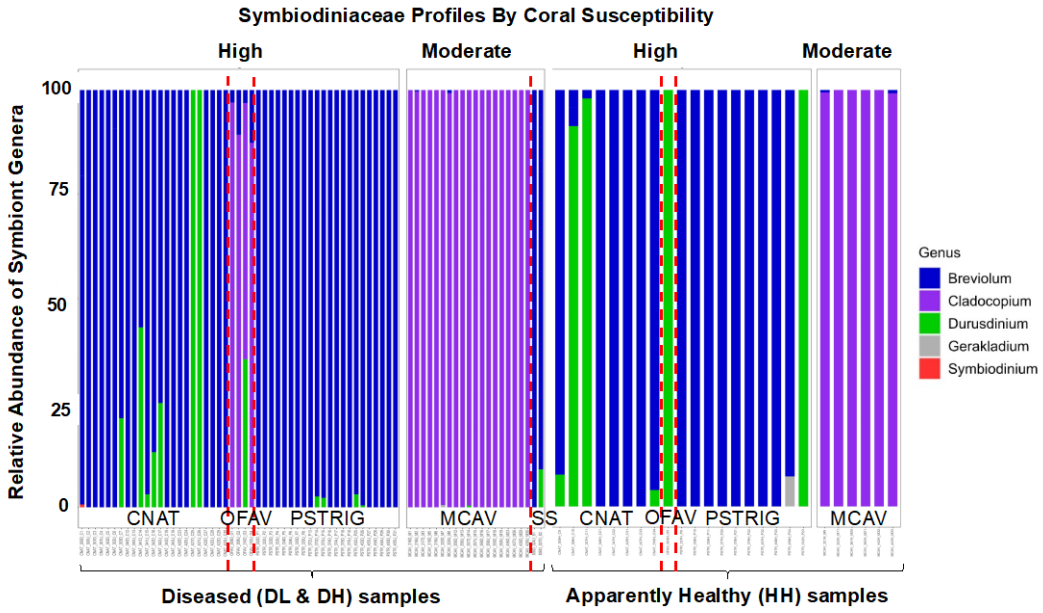


Figure 5. Relative abundance of symbiont genera across species, host disease susceptibility, and disease status. Using ITS-2 profiles 5 main genera were identified among the samples, with *Breviolum* dominating many highly disease susceptible coral hosts and all health states. *Dursusdinium* was also found among the hosts (with the exception of *M. cavernosa*) and across both diseased and apparently healthy samples.

3.3.2. Viral Metatranscriptome Results

To date we have only had a few weeks to process and access our meta-transcriptomic data which can take considerable time to analyze completely. However, below is a short preliminary assessment of viral signatures within a subset of the libraries (dsRNA and polyA selected libraries from Rice). Please note that all of the below analysis requires additional statistical and informatic confirmation and must be interpreted with caution.

Also, tangential but germane to this project, Dr. Correa's lab recently assembled and disseminated 2 new alphaflexiviruses from SCTLD afflicted corals (CHFV1; CHFV2). The data are publicly available (Veglia et al., 2022) and provide context and additional genomic resources for us to use in our current research aims (Figure 6). These genomic data will be used as a reference genome for all additional analysis (see below).

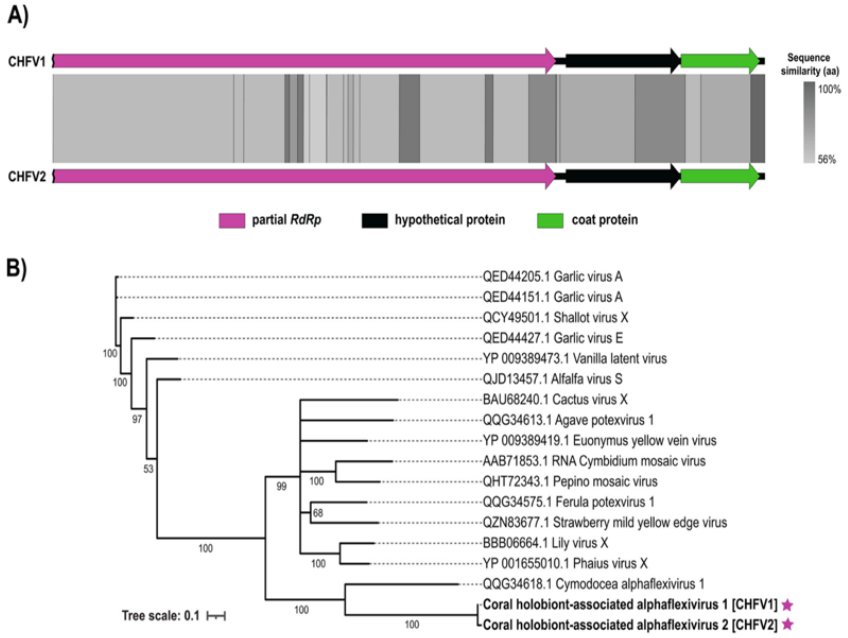


Figure 6: Genomes and phylogenetic relationships of two SCTLD-associated alphaflexivirus (CHFV) genomes
from Veglia et al., 2022. Arrows represent predicted genes, the RNA Dependent RNA polymerase and the major coat protein. Gray-scale shading between the two genomes represents the percent amino acid (aa) sequence similarity. (B) Maximum likelihood phylogeny generated from translated alphaflexivirus RdRd amino acid sequences from the CHFVs (purple stars) as well as previously described plant-associated alphaflexiviruses.

3.3.2.1. RDRP Phylogenetic Results

RNA viruses (Baltimore classifications III, IV, and IV) universally share RDRP genes, which has inspired gallant efforts to unify their evolutionary origins (Wolf et al. 2018). In consequence, RDRP can be used to survey and phylogenetically assess total

RNA virus diversity. Assembly initially produced 237 proposed RDRP contigs (Figure 7), though some of these initial hits were considered non-viral following BLASTX verification against the RVDB, thus reducing potential hits to 222 sequences (Figure 7). The resulting 222 sequences contained a variety of functional annotations including those to polyproteins (which contain multiple domains in addition to RDRP), genes involved in ubiquitin synthesis, and RDRP.

To accelerate our preliminary analysis, we only chose to work on clearly annotated, non-redundant RDRP annotated sequences for further analysis. CHFV1/2, a +ssRNA flexivirus which we previously detected in SCTLD metatranscriptomes (Figure 6), was not identified in these new datasets thus far. However, the vast majority of RDRP sequences match were to viruses with +ssRNA genomes (Figure 7). Moreover, half of the best similarities not only have a positive sense genome, but they appear to be closely related filamentous viruses. This is in stark contrast to the other half of RDRP sequences that are not represented by filamentous particles. Indeed, most of the other viral RDRP references include those with putative icosahedral virions, though viruses with naked and unknown structures (i.e., environmentally sequenced) were also found.

While this analysis is incomplete, it does indicate that RNA viruses are highly prevalent in coral metatranscriptomes. Moreover, rather than focusing on one lineage, it might be necessary to broaden directed analyses at viruses collapsed into broader taxonomic groups (i.e., family or genus level) to delineate whether filamentous viruses are significantly correlated with specific coral health states.

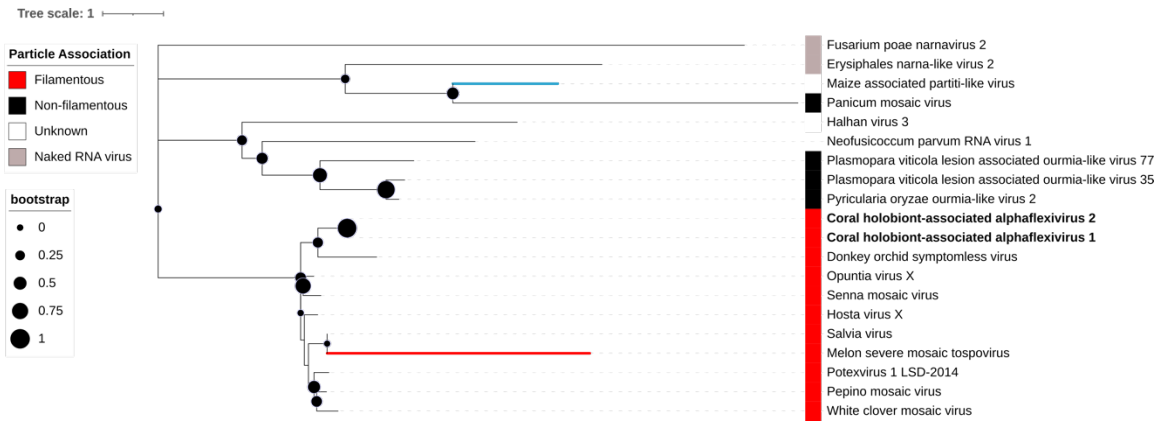


Figure 7: Phylogenetic reconstruction of reference viral RDRP sequences representing best blastx hits to RdRp contigs generated in this study. Bolded, Coral-holobiont-associated alphaflexivirus 1/2 were not best hits in this study, but were added as a reference due to their detection in previous SCTLD metatranscriptomes (Veglia et al 2022). Tree branch colors follow the same structure in Wolf et al. 2018; red = -ssRNA viruses; blue = dsRNA viruses; black = +ssRNA viruses.

3.3.2.2. Kmer-based Results

Kmer mapping methods aimed to reveal viral diversity and taxonomic profiles of the samples using more sequences than the RDRP reads alone. Based exclusively on only the 33 *Colpophyllia natans* libraries, polyA enrichment resulted in higher detected viral richness than dsRNA approaches which could be due to overall higher sequence similarities. However, while the number of similarities were lower, dsRNA immunoprecipitation resolved stronger differences in viral richness between tissue health states (Figure 8). Although preliminary, these data first suggest that the methodologies

will provide unique answers to the questions. Yet excitingly, both approaches showed that the patterns in viral diversity are the same. Both methods clearly demonstrated that viral diversity is different among health states (disease lesion or DL, disease healthy or DH, healthy healthy or HH). Diversity is both higher and more variable in the diseased lesion (DL) than in the healthy portion of the disease animals (DH). Remarkably, viral diversity, as estimated by our kmer approach, was lowest in the healthy healthy (HH) specimen. This trend must be validated, but this initial observation in the metatranscriptomic data further suggests that viruses may play some role in SCTLD.

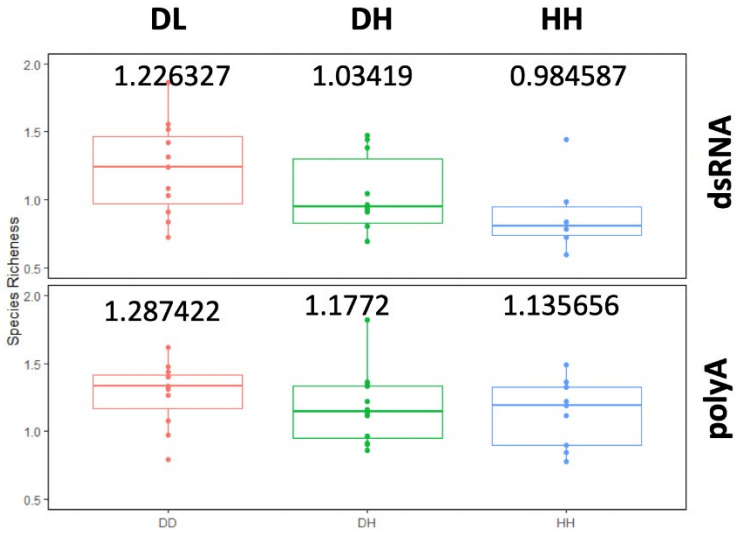


Figure 8. Viral species richness estimates from Kmer based analysis. Based on 33 *Colpophyllia natans* libraries, polyA enrichment resulted in higher detected viral richness, but dsRNA immunoprecipitation resolved stronger differences in richness between tissue health states (disease lesion or DL (red), disease healthy or DH (green), healthy healthy or HH (blue), with DL having the highest richness and HH having the lowest).

4. DISCUSSION

This project thus far has created a remarkable comparative and holistic dataset of 428 coral meta-transcriptomes and 107 ITS2 amplicon libraries to determine the interacting roles of coral host and symbiont identifies, viral consortia, and SCTLD susceptibility and infection status. We have completed the primary tasks of 1) isolating nucleic acids, 2) generating 4 comparative metatranscriptomic library sets from 4 approaches, 3) created ITS amplicon libraries and profiles for all 107 samples, and 4) have received a majority of the meta-transcriptomic sequence data itself.

The data analysis reported herein are only a cursory and preliminary exploration at the large amount of data generated. In this report we present evidence that there is a myriad of viral signatures across the libraries, but that a majority of the coral associated RDRP genes fall within +ssRNA viruses that have been described as filamentous, adding evidence that filamentous viruses play a potential role in the disease, although what that role is undetermined. We also showed that viral diversity is higher in diseased lesions compared to health portions of diseased colonies and apparently healthy colonies. However, viral production can be stimulated by other pathogens or physiological diseases

such as bleaching, and thus evidence of their presence and elevated diversity does not necessarily confirm they are the causative agent of a given disease sign. We will continue to evaluate if the abundance of certain viral taxa (such as these putative filamentous +ssDNA viruses) are positively associated with disease state to more conclusively link them to SCTLD.

Further, we also showed here that corals with the Symbiodiniaceae genus *Breviolum* tend to be more susceptible (needs to be confirmed statistically) to SCTLD, whereas coral species considered to be moderately susceptible to SCTLD tended to be dominated by *Cladocopium*. We also, however, found *Breviolum* within all disease states (healthy and diseased) ruling out that they are diagnostic of disease susceptibility. We suspect that corals that contain primarily *Breviolum* may show disease signs more prominently or earlier in the infection cycle. Also, it has been suggested that corals containing the thermotolerant symbiont *Durusdinium* are not-susceptible to SCTLD, but we showed here that corals previously shown to have filamentous viruses also contain these *Durusdinium* symbionts, potentially ruling out this hypothesis.

For our next year's proposal, we will complete the sequencing and focus on data analysis, statistical validations, interpretations, and disseminations of our findings. Regardless, the generated data from this first project represents a foundational dataset for us to uncover the role (if any) of viruses in this coral disease.

In summary, these early data provide important information about the potential role of viruses in SCTLD. While preliminary, early analysis suggests that viruses related to other known filamentous viruses are present in the analyzed specimen. Further viral diversity was highest in diseased lesions suggesting at minimum an associative role for viruses in the disease.

The outcomes of this project will be incorporated into an on-going 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.

***Table 1**: Details concerning the coral specimen RNA and DNA isolation and sequencing as of June 15th 2022.*

Coral Species	**Specimen ID**	Location of Collection	RNA Conc (ng/ul)	DNA Conc (ng/ul)	Meta-transcriptomic Sequencing 4 ways	Meta-barcoding of ITS2
Colpophyllia natans	300 D	West Turtle Shoal	377.5	11.6	2 of 4	Y

Colpophyllia natans	300 U	West Turtle Shoal	388.8	8.7	2 of 4	Y
Colpophyllia natans	301 D	West Turtle Shoal	325.3	21.4	2 of 4	Y
Colpophyllia natans	301 U	West Turtle Shoal	730	42.4	2 of 4	Y
Colpophyllia natans	302 D	West Turtle Shoal	1494.7	8.3	2 of 4	Y
Colpophyllia natans	302 U	West Turtle Shoal	325.8	7.6	2 of 4	Y
Colpophyllia natans	303 D	West Turtle Shoal	258.2	7.7	2 of 4	Y
Colpophyllia natans	303 U	West Turtle Shoal	106.4	7.1	2 of 4	Y
Colpophyllia natans	305 H	West Turtle Shoal	113.1	8.3	2 of 4	Y
Colpophyllia natans	306 H	West Turtle Shoal	79.9	20.2	2 of 4	Y
Colpophyllia natans	307 H	West Turtle Shoal	37.3	9.5	2 of 4	Y
Colpophyllia natans	340 D	Lat Long 2 - southernmost	145.3	6.5	2 of 4	Y
Colpophyllia natans	340 U	Lat Long 2 - southernmost	293.7	20.4	2 of 4	Y
Colpophyllia natans	341 D	Lat Long 2 - southernmost	1031.4	13.2	2 of 4	Y
Colpophyllia natans	341 U	Lat Long 2 - southernmost	221.9	16.2	2 of 4	Y
Colpophyllia natans	342 D	Lat Long 2 - southernmost	369	11.2	2 of 4	Y
Colpophyllia natans	342 U	Lat Long 2 - southernmost	370.5	18.7	2 of 4	Y
Colpophyllia natans	343 D	Lat Long 2 - southernmost	411.3	11.3	2 of 4	Y
Colpophyllia natans	343 U	Lat Long 2 - southernmost	337.3	10.7	2 of 4	Y
Colpophyllia natans	345 H	Lat Long 2 - southernmost	184.2	7.5	2 of 4	Y
Colpophyllia natans	346 H	Lat Long 2 - southernmost	533.6	8.7	2 of 4	Y
Colpophyllia natans	347 H	Lat Long 2 - southernmost	102.8	21	2 of 4	Y
Colpophyllia natans	420 D	Lat Long 4 - northernmost	546.6	22.8	2 of 4	Y
Colpophyllia natans	420 U	Lat Long 4 - northernmost	862.5	22.4	2 of 4	Y
Colpophyllia natans	421 D	Lat Long 4 - northernmost	191	12.6	2 of 4	Y
Colpophyllia natans	421 U	Lat Long 4 - northernmost	136.4	14.4	2 of 4	Y
Colpophyllia natans	422 D	Lat Long 4 - northernmost	210.8	12.6	2 of 4	Y
Colpophyllia natans	422 U	Lat Long 4 - northernmost	248.9	15.6	2 of 4	Y

Colpophyllia natans	423 D	Lat Long 4 - northernmost	1112.2	32.4	2 of 4	Y
Colpophyllia natans	423 U	Lat Long 4 - northernmost	700.4	23.8	2 of 4	Y
Colpophyllia natans	425 H	Lat Long 4 - northernmost	551.3	29.5	2 of 4	Y
Colpophyllia natans	426 H	Lat Long 4 - northernmost	284.8	10.4	2 of 4	Y
Colpophyllia natans	427 H	Lat Long 4 - northernmost	232.6	31.9	2 of 4	Y
Colpophyllia natans	516 H	Western Sambo Patch 2	??	11.4	2 of 4	Y
Montastraea cavernosa	316 D	West Turtle Shoal	431.9	6.9	2 of 4	Y
Montastraea cavernosa	316 U	West Turtle Shoal	527	9	2 of 4	Y
Montastraea cavernosa	317 D	West Turtle Shoal	300.3	8.8	2 of 4	Y
Montastraea cavernosa	317 U	West Turtle Shoal	330	16.4	2 of 4	Y
Montastraea cavernosa	319 D	West Turtle Shoal	268.1	12.1	2 of 4	Y
Montastraea cavernosa	319 U	West Turtle Shoal	457.4	7.5	2 of 4	Y
Montastraea cavernosa	320 D	West Turtle Shoal	153	16.1	2 of 4	Y
Montastraea cavernosa	320 U	West Turtle Shoal	315.3	15.1	2 of 4	Y
Montastraea cavernosa	321 H	West Turtle Shoal	244.8	10.8	2 of 4	Y
Montastraea cavernosa	322 H	West Turtle Shoal	214.4	12.7	2 of 4	Y
Montastraea cavernosa	323 H	West Turtle Shoal	359.3	7.6	2 of 4	Y
Montastraea cavernosa	356 D	Lat Long 2 - southernmost	149.1	6	2 of 4	Y
Montastraea cavernosa	356 U	Lat Long 2 - southernmost	127	7.6	2 of 4	Y
Montastraea cavernosa	357 D	Lat Long 2 - southernmost	371.2	6.1	2 of 4	Y
Montastraea cavernosa	357 U	Lat Long 2 - southernmost	498.8	9.5	2 of 4	Y
Montastraea cavernosa	358 D	Lat Long 2 - southernmost	356.7	5.7	2 of 4	Y
Montastraea cavernosa	358 U	Lat Long 2 - southernmost	439.8	6	2 of 4	Y
Montastraea cavernosa	359 D	Lat Long 2 - southernmost	343.6	13.2	2 of 4	Y
Montastraea cavernosa	359 U	Lat Long 2 - southernmost	236.6	8.4	2 of 4	Y
Montastraea cavernosa	361 H	Lat Long 2 - southernmost	349.9	5.9	2 of 4	Y
Montastraea cavernosa	362 H	Lat Long 2 - southernmost	111.3	11.5	2 of 4	Y

Montastraea cavernosa	363 H	Lat Long 2 - southernmost	429.5	9.7	2 of 4	Y
Montastraea cavernosa	436 D	Lat Long 4 - northernmost	186.6	8.5	2 of 4	Y
Montastraea cavernosa	436 U	Lat Long 4 - northernmost	748.7	11	2 of 4	Y
Montastraea cavernosa	437 D	Lat Long 4 - northernmost	940.4	5.4	2 of 4	Y
Montastraea cavernosa	437 U	Lat Long 4 - northernmost	688.9	7.1	2 of 4	Y
Montastraea cavernosa	438 D	Lat Long 4 - northernmost	443.7	13.1	2 of 4	Y
Montastraea cavernosa	438 U	Lat Long 4 - northernmost	366.9	10	2 of 4	Y
Montastraea cavernosa	439 D	Lat Long 4 - northernmost	860.6	11.5	2 of 4	Y
Montastraea cavernosa	439 U	Lat Long 4 - northernmost	592.3	5.1	2 of 4	Y
Montastraea cavernosa	441 H	Lat Long 4 - northernmost	706.3	12.8	2 of 4	Y
Montastraea cavernosa	442 H	Lat Long 4 - northernmost	1125.1	6.7	2 of 4	Y
Montastraea cavernosa	443 H	Lat Long 4 - northernmost	315.3	4	2 of 4	Y
Montastraea cavernosa	506 H	Western Sambo Patch 2	54.2	6.9	2 of 4	Y
Pseudodiploria strigosa	332 D	West Turtle Shoal	71.8	7.6	2 of 4	Y
Pseudodiploria strigosa	332 U	West Turtle Shoal	254.3	33.2	2 of 4	Y
Pseudodiploria strigosa	333 D	West Turtle Shoal	101.2	24	2 of 4	Y
Pseudodiploria strigosa	333 U	West Turtle Shoal	219.2	28.4	2 of 4	Y
Pseudodiploria strigosa	334 D	West Turtle Shoal	170.7	2.9	2 of 4	Y
Pseudodiploria strigosa	334 U	West Turtle Shoal	58.6	9.6	2 of 4	Y
Pseudodiploria strigosa	335 D	West Turtle Shoal	68.5	4.8	2 of 4	Y
Pseudodiploria strigosa	335 U	West Turtle Shoal	140	8.3	2 of 4	Y
Pseudodiploria strigosa	337 H	West Turtle Shoal	115.7	32	2 of 4	Y
Pseudodiploria strigosa	338 H	West Turtle Shoal	93.3	14.7	2 of 4	Y
Pseudodiploria strigosa	339 H	West Turtle Shoal	96.6	6.4	2 of 4	Y
Pseudodiploria strigosa	372 D	Lat Long 2 - southernmost	194.4	19	2 of 4	Y
Pseudodiploria strigosa	372 U	Lat Long 2 - southernmost	151.3	30.8	2 of 4	Y
Pseudodiploria strigosa	373 D	Lat Long 2 - southernmost	161.2	5.7	2 of 4	Y

Pseudodiploria strigosa	373 U	Lat Long 2 - southernmost	186.2	12	2 of 4	Y
Pseudodiploria strigosa	374 D	Lat Long 2 - southernmost	95.3	6.8	2 of 4	Y
Pseudodiploria strigosa	374 U	Lat Long 2 - southernmost	20.9	6.2	2 of 4	Y
Pseudodiploria strigosa	375 D	Lat Long 2 - southernmost	92.5	4.9	2 of 4	Y
Pseudodiploria strigosa	375 U	Lat Long 2 - southernmost	89.7	5.7	2 of 4	Y
Pseudodiploria strigosa	377 H	Lat Long 2 - southernmost	194.1	3.7	2 of 4	Y
Pseudodiploria strigosa	378 H	Lat Long 2 - southernmost	219	3.8	2 of 4	Y
Pseudodiploria strigosa	379 H	Lat Long 2 - southernmost	373.8	4.5	2 of 4	Y
Pseudodiploria strigosa	512H	Western Sambo Patch 2	180.2	6.1	2 of 4	Y
Pseudodiploria strigosa	452 D	Lat Long 4 - northernmost	163.5	5	2 of 4	Y
Pseudodiploria strigosa	452 U	Lat Long 4 - northernmost	122.6	12	2 of 4	Y
Pseudodiploria strigosa	453 D	Lat Long 4 - northernmost	109.2	45.1	2 of 4	Y
Pseudodiploria strigosa	453 U	Lat Long 4 - northernmost	163.6	60.9	2 of 4	Y
Pseudodiploria strigosa	454 D	Lat Long 4 - northernmost	158.5	14.1	2 of 4	Y
Pseudodiploria strigosa	454 U	Lat Long 4 - northernmost	121.8	6.7	2 of 4	Y
Pseudodiploria strigosa	455 D	Lat Long 4 - northernmost	173.2	11.3	2 of 4	Y
Pseudodiploria strigosa	455 U	Lat Long 4 - northernmost	313.2	16.4	2 of 4	Y
Pseudodiploria strigosa	457 H	Lat Long 4 - northernmost	261	7.7	2 of 4	Y
Pseudodiploria strigosa	458 H	Lat Long 4 - northernmost	167.3	10.1	2 of 4	Y
Pseudodiploria strigosa	459 H	Lat Long 4 - northernmost	129.1	9.2	2 of 4	Y
Siderastrea siderea	324 U	West Turtle Shore	12.7	9.3	2 of 4	Y
Siderastrea siderea	327 D	West Turtle Shore	30.1	3.6	2 of 4	Y
Siderastrea siderea	509 H	Western Sambo Patch 2	30.1		2 of 4	Y
Orbicella faveolata	308 D	West Turtle Shore	35.6	5.7	2 of 4	Y
Orbicella faveolata	308 U	West Turtle Shore	19.5	4.5	2 of 4	Y
Orbicella faveolata	310 D	West Turtle Shore	27	7.6	2 of 4	Y
Orbicella faveolata	310 U	West Turtle Shore	68.7	7.4	2 of 4	Y

Orbicella faveolata	503 H	Western Sambo Patch 2	15.8	6	2 of 4	Y

References:

Bigot, Thomas, Sarah Temmam, Philippe Pérot, and Marc Eloit. 2019. "RVDB-Prot, a Reference Viral Protein Database and Its HMM Profiles." F1000Research 8 (April): 530.

Breitwieser, Florian P., and Steven L. Salzberg. 2020. "Pavian: Interactive Analysis of Metagenomics Data for Microbiome Studies and Pathogen Identification." Bioinformatics. https://doi.org/10.1093/bioinformatics/btz715.

Bushmanova, Elena, Dmitry Antipov, Alla Lapidus, and Andrey D. Prjibelski. 2019. "rnaSPAdes: A de Novo Transcriptome Assembler and Its Application to RNA-Seq Data." GigaScience. https://doi.org/10.1093/gigascience/giz100.

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

Davis, Matthew P. A., Stijn van Dongen, Cei Abreu-Goodger, Nenad Bartonicek, and Anton J. Enright. 2013. "Kraken: A Set of Tools for Quality Control and Analysis of High-Throughput Sequence Data." Methods. https://doi.org/10.1016/j.ymeth.2013.06.027.

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

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." Bioinformatics 32 (19): 3047–48.

Howe-Kerr, Lauren I., Benedicte Bachelot, Rachel M. Wright, Carly D. Kenkel, Line K. Bay, and Adrienne M. S. Correa. n.d. "Symbiont Community Diversity Is More Constrained in Holobionts That Tolerate Diverse Stressors." https://doi.org/10.1101/572479.

Hume, Benjamin C. C., Edward G. Smith, Maren Ziegler, Hugh J. M. Warrington, John A. Burt, Todd C. LaJeunesse, Joerg Wiedenmann, and Christian R. Voolstra. 2019. "SymPortal: A Novel Analytical Framework and Platform for Coral Algal Symbiont Next-generation Sequencing ITS2 Profiling." Molecular Ecology Resources. https://doi.org/10.1111/1755-0998.13004.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. "MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms." Molecular Biology and Evolution 35 (6): 1547–49.

Letunic, Ivica, and Peer Bork. 2021. "Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation." Nucleic Acids Research 49 (W1): W293–96.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." Genome Biology 15 (12): 550.

Martin, Marcel. 2011. "Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads." EMBnet.journal. https://doi.org/10.14806/ej.17.1.200.

McMurdie, Paul J., and Susan Holmes. 2013. "Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." PloS One 8 (4): e61217.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. "CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice." Nucleic Acids Research 22 (22): 4673–80.

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.

Wolf, Yuri I., Darius Kazlauskas, Jaime Iranzo, Adriana Lucía-Sanz, Jens H. Kuhn, Mart Krupovic, Valerian V. Dolja, and Eugene V. Koonin. 2018. “Origins and Evolution of the Global RNA Virome.” mBio 9 (6). https://doi.org/10.1128/mBio.02329-18.

Wood, Derrick E., Jennifer Lu, and Ben Langmead. 2019. “Improved Metagenomic Analysis with Kraken 2.” Genome Biology 20 (1): 257.