Quick Look Report

Deliverable

CRCP Project 31313: SCTLD Investigations into diagnostic and treatments: Functional Proteomics

Janet E Saunders[1], Elizabeth Duselis[1], Michael G. Janech[2], Benjamin Neely[3], Cheryl M. Woodley[4]

[1]Consolidated Safety Services, Inc., contractor for National Oceanic and Atmospheric Administration, National Centers for Coastal Ocean Science, Hollings Marine Laboratory, 331 Ft. Johnson Rd., Charleston, SC, 29412 USA

[2]College of Charleston, Department of Biology, 66 George St., Charleston, SC 29424 USA

[3]National Institute of Standards and Technology, Marine Biochemical Sciences Group, Hollings Marine Laboratory, 331 Ft. Johnson Rd., Charleston, SC 29412 USA

[4]National Oceanic and Atmospheric Administration, National Ocean Service, National Centers for Coastal Ocean Science, Hollings Marine Laboratory, 331 Ft. Johnson Rd., Charleston, SC 29412 USA

January 26, 2024

I. Introduction

Florida's coral reefs have been experiencing a multi-year disease-related mortality event, stony coral tissue loss disease (SCTLD), that has resulted in massive die-offs in multiple coral species. Over 24 species of coral, including Endangered Species Act-listed and primary reef-building species, are estimated to have displayed tissue loss lesions which frequently result in whole colony mortality. SCTLD was first observed in late 2014, in Virginia Key (Precht et al. 2016) and has continued to spread affecting corals in 28 countries/territories throughout the Caribbean as of this writing (Aeby et al. 2021).

The causative agent(s) of SCTLD remain unknown, but are suspected to be a combination of abiotic, microbial and/or viral agents as described for other coral diseases (Bourne et al. 2009; Munn 2015; Wear and Thurber 2015; Thurber et al. 2017). Molecular approaches (e.g., 16s rDNA microbiome analysis) have been used in conjunction with field sampling and laboratory experiments to identify and confirm causes of other coral diseases, satisfying the basic tenets of Koch's postulates e.g., coral bleaching associated with *Vibrio coralliilyticus* (Ben-Haim and Rosenberg 2002). Based on two published studies that have comprehensively investigated other stony coral diseases, dominant changes to the microbiome appear to create a signature during the onset and duration of disease (Garcia et al. 2016; Sato et al. 2017). In human disease, molecular changes in the host and microbiome occur well before observable phenotype and gross pathology is observed (Taylor et al. 2016). For this reason, defining the changes in the molecular landscape of the coral holobiont can provide useful information not only in diagnosis, but for prediction and prognosis of disease (Rechenberger et al. 2019). Specifically, in the case of SCTLD, defining molecular changes in the coral holobiont will aid in identifying the causative agent(s) by clearly defining traits of disease progression shared across affected species. We initially focused on the functional response of the coral microbiome because this search space can be defined using metagenomics. This analysis entailed creating metagenomic libraries from these SCTLD positive samples, which are therefore specific to the microbial composition of each sample that has SCTLD, as opposed to selecting publicly available databases that would be less specific to this study as they were likely created from non-diseased samples. In the future, these species-specific metagenomic databases will be searched during metaproteomic analysis.

In an effort to define the molecular changes elicited by SCTLD in the coral microbiome, tools measuring multiple 'omic fields are useful. We focused primarily on protein abundance to capture the functional aspects of the disease phenotype as well as microbial species diversity by way of the metagenomic analysis mentioned previously. Classification of the disease phenotype can be facilitated by determining the expression of virulence factors associated with the microbiome that may underlie infection and spread of disease in addition to measuring protein abundance in the host. Together these can be used to predict the acute phase of the disease process that is not grossly nor histologically visible.

Meta-omic molecular tools have been sparsely applied to the study of coral disease (Garcia et al. 2016; Meyer et al. 2017; Sato et al. 2017; Meyer et al. 2019). Meyer et al. (2019) applied metagenomic techniques to corals affected by SCTLD. This laid the groundwork for defining the pathogen pool in four coral hosts from Florida. Importantly, the study also identified taxonomic groups that are unique to the diseased lesions. Metaproteomics, on the other hand, has largely been ignored despite being a critical aspect to assigning functional operations to a biological system. Many infectious agents have a synergistic etiology with the host microbiome, such as in human respiratory tract infections (Taylor et al. 2016), emphasizing the value of studying functional changes in the microbiome in response to unknown viral/microbial/abiotic stressors. Garcia et al. (2016) addressed coral disease using metaproteomics; however, only 361 proteins were assessed, which is far below expected using current technology, thereby drastically limiting functional interpretation of the data. The first portion of our project aimed to comprehensively understand the microbial and viral community composition through metaproteomics and extend these data to define a conserved functional SCTLD shift in a broad array of coral species.

The second portion of this project encompasses the proteomic portion of an integrated, multi-investigator, multi-disciplinary project (Florida DEP Award No. B8A48D) the goal of which is to understand the genetic, biochemical and physiological underpinnings in the holobiont of *Orbicella faveolata* that are driving intraspecific differences in resistance and susceptibility to SCTLD. Reconnaissance efforts in southeast Florida reefs of moderately susceptible species (*Orbicella* spp., *Montastraea cavernosa*) have identified populations of individuals with differing infection patterns: 1) those with no disease signs, 2) those with infection that arrests with antibiotic treatment, or 3) those with new lesions erupting numerous times despite antibiotic

treatment. This raises the question: why are some conspecific individuals more susceptible to the disease?

Populations of *Orbicella faveolata* in Florida waters have been the subject of intensive disease intervention efforts to stop SCTLD. This study provided the proteomics component of a larger project that was undertaken by a consortium of experts using a trans-disciplinary approach to develop a fundamental understanding of Florida's mountainous star coral (*Orbicella faveolata*) holobiont at gross morphologic, genetic, biochemical and molecular scales at three time points. The larger study has identified differences in endosymbionts, genotypes, metabolites, microbes, biochemical and physiological pathways, transcripts and histopathology, although no differences have been specifically attributed to the presence of SCTLD. The proteomics component has identified differentially abundant proteins across disease states, time and location from the repeated measures sampling design. Based on the putative identities of proteins differentially expressed, functional pathways have been hypothesized in relation to mechanisms likely involved in the pathology that is being expressed as tissue loss. From this aspect we have identified possible candidates for diagnostic markers for resistance and susceptibility. The strength of this data however will reside in the in-depth analysis across data sets that represent different aspects of the holobiont, to erect testable hypotheses and a weight of evidence in a systems analysis approach.

The dynamic nature of this disease and the wide range of individual coral responses led us to hypothesize five "outcomes" for these individual corals to have and utilize these outcomes to re-classify the individual colonies. By treating corals as individual patients and looking at dynamic changes versus static changes, we have expanded the potential conclusions of all the different 'omic analyses, including proteomics.

II. Methods

2.1 Sampling to Construct Specific Metagenomic Libraries of SCTLD+ and SCTLD- corals from Nine Different Species for Future Metaproteomic Analysis

Nine different coral species were selected to represent highly susceptible species with faster and slower onset SCTLD dynamics. Specimen were collected in July and September 2019 in collaboration with the FL Keys National Marine Sanctuary under permit #FKNMS-2019-069 to Dr. Cheryl Woodley of NOAA and KFNMS-2019-001-A1 to Dr. Andy Bruckner of the

Florida Keys National Marine Sanctuary. Samples in July were taken offshore of Key West, FL in a diseased zone and paired with specimens from a SCTLD-free zone. In September, diseased samples were collected offshore of Key West but the disease line was approaching the Marquesas, requiring the specimen from the disease-free zone to be collected in a yet unaffected area of the Marquesas. The permitted species include:

Colpophyllia natans
Dichocoenia stokesii
Diploria labyrinthiformis
Meandrina meandrites
Montastraea cavernosa
Orbicella annularis
Orbicella faveolata
Pseudodiploria strigose
Pseudodiploria clivosa

2.2 DNA Isolation and Constructing Metagenomic Libraries of SCTLD+ and SCTLD- Coral Samples from Nine Different Species

Tissue samples were collected from colonies and immediately frozen in liquid nitrogen and stored at Hollings Marine Laboratory at -80°C. DNA and protein from each species were first extracted by incubating coral skeleton with adherent tissue in 5% SDS for 30 minutes with manual disruption using a bristle brush in a 15 mL falcon tube. Samples were centrifuged at 1,500 rcf and 1 mL of supernatant was removed and stored at -80°C prior to being shipped to Novogene for DNA extractions.

Table 1. Samples with and without SCTLD that passed QC for metagenomic analysis.

Samples that passed QC	SCLTD + or -
CN.92	+
CN.94	+
CN.96	+
DL.73	-
DL.119	-
DL.251	+

DS.23	-
DS.37	+
DS.265	+
DS.10	+
OA.241	+
OA.363	+
OF.35	-
OF.113	-
OF.240	+
OF.261	+
PS.114	-

Forty-six coral samples representing nine species were submitted for metagenome library construction. Only 17 of 46 samples passed QC and were utilized for library preparation and downstream analysis (**Table 1**). Agarose gel electrophoresis was used to determine DNA purity and integrity, and a Qubit 2.0 fluorometer quantification was used to quantify DNA concentration. Physical fractionation was performed by a Covaris Sonicator. Fractionation steps were checked by an Agilent 2100 and Q-PCR to ensure that sufficient enrichment of the target was achieved. End repairing, A-tailing, ligation of sequencing adapters, size selection and PCR enrichment steps were used to produce each library. A total amount of 1 μg DNA per sample was used as input material for the DNA sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Sequencing was performed using the Illumina platform after library clustering with paired-end reads. The clustering of the index-coded samples was performed on a cBot Cluster Generation System according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina HiSeq platform and paired-end reads were generated.

Raw data were filtered for quality and clean data were used for assembly and comparisons. Metagenomes were assembled and used as a quality control of each sample; these showed how consistent the constructed genomes of SCTLD+ and SCTLD- samples were among all species analyzed. The unutilized reads of each sample were put together for mixed assembly to explore the information of low-abundance species of the samples. Gene prediction was carried out by MetaGeneMark based on the scaftigs which were assembled by single and mixed

samples. Predicted genes were pooled together for dereplication to construct the gene catalog. Taxonomy was annotated by comparing metagenomic reads to the NCBI non-reductant database of taxonomically informative gene families to annotate each metagenomic homolog. Abundance of different taxonomic ranks were based on a gene abundance table. The function of the coding sequence was inferred based on its similarity to sequences in the database (KEGG, eggnog, CAZy). Based on the taxonomic abundance table and the function abundance table, clustering analysis, Anosim, PCA and NMDS was carried out across SCTLD+ and SCTLD- samples combined irrespective of species. When grouping information was available, Metastats and LEfSe multivariate statistical analysis used a comparative analysis of metabolic pathways to explore species composition and functional composition differences between groups.

2.3 Protein Isolation from SCTLD+ and SCTLD- Coral Samples from Nine Different Species with Trypsin Digestion

Total protein was quantified using a microBCA protein assay (Pierce). Samples were diluted 1:100 and standards were amended with lysis buffer of the same dilution to replicate the assay matrix. Polyacrylamide gels were run to ensure protein was visible across a wide molecular weight range in a subset of samples. Protein from each coral sample was reduced with dithiothreitol (DTT), alkylated with chloroacetamide (CAA) and digested with trypsin using the micro S-traps protocol (ProtiFi) (ProtiFi, Fairport NY). Peptides were eluted from the S-trap column and assayed using a colorimetric peptide assay kit (Pierce). Peptide samples were stored at -80°C until mass spectrometry analysis.

2.4 Liquid Chromatography / Mass spectrometry of SCTLD+ and SCTLD- Samples from Nine Different Species

Peptide mixtures in 0.1% formic acid (volume fraction) were analyzed using an UltiMate 3000 Nano LC coupled to a Fusion Lumos Orbitrap mass spectrometer (Thermo Fisher Scientific). Using the original sample, randomization yielded a randomized sample order and injection volumes were determined for 0.5 µg loading (between 0.5 and 4.8 µL). Peptide mixtures were loaded onto a PepMap 100 C18 trap columns (75 µm id x 2 cm length; Thermo Fisher Scientific) at 3 µL/min for 10 min with 2% acetonitrile (volume fraction) and 0.05% trifluoroacetic acid (volume fraction) followed by separation on an Acclaim PepMap RSLC 2

µm C18 column (75 µm id x 25 cm length; Thermo Fisher Scientific) at 40°C. Peptides were separated along a 65 min two-step gradient of 5% to 30% mobile phase B (80% acetonitrile volume fraction, 0.08% formic acid volume fraction) over 50 min followed by a ramp to 45% mobile phase B over 10 min and lastly 95% mobile phase B over 5 min, and held at 95% mobile phase B for 5 min, all at a flow rate of 300 nL/min. The Fusion Lumos was operated in a positive polarity with 30% RF (radiofrequency) lens, data-dependent mode (topN, 3 sec cycle time) with a dynamic exclusion of 60 s (with 10 ppm error). Full scan resolution using the orbitrap was set at 60,000, the mass range was m/z 375 to 1500. Full scan ion target value was 4.0e5 allowing a maximum injection time of 50 ms. Monoisotopic peak determination was used, specifying peptides and an intensity threshold of 2.5e4 was used for precursor selection, including charge states 2 to 6. Data-dependent fragmentation was performed using higher-energy collisional dissociation (HCD) at a normalized collision energy of 32 with quadrupole isolation at m/z 1.3 width. The fragment scan resolution using the orbitrap was set at 15,000 m/z 100 as the first mass, ion target value of 2.0e5 and 30 ms maximum injection time. The MS1 data was collected as profile data, the MS2 data was collected as centroid data. Inject all ions for parallelizable time was not used. Raw files were converted to peak lists using ThermoRawFileParserGUI 1.2.1 using "Native Thermo library peak picking". These data were searched using the Mascot algorithm (v2.6.2; Matrix Science). All the searches included a metagenomic database compiled from a subset of the samples: SCTLD_unigenes_protein_cdhit.fasta as well as Symbiodinium (taxonID:2949) from the UniProtKB 2021_01 release (both SwissProt and TrEMBL). The 10 samples from *Orbicella* spp. were searched with the *Orbicella faveolata* RefSeq database (release 100:GCF_002042975.1_ofav_dov_v1), while all the data (including these *Orbicella* spp.) were searched with Scleractinia (taxonID:6125) from the UniProtKB 2021_01 release (both SwissPro and TrEMBL). Searches also included the common Repository of Adventitious Proteins database (cRAP; 2012.01.01; the Global Proteome Machine; 107 sequences). These fasta files are included in the PRIDE dataset. The following search parameters were used: trypsin was specified as the enzyme allowing for one mis-cleavages; carbamidomethyl (C) was fixed and oxidation (M) was variable modification; 10 ppm precursor mass tolerance and 10 ppm fragment ion tolerance; instrument type was specified as ESI-FTICR; the decoy setting was used within Mascot to provide local false discovery rate (FDR). The resulting .dat files were loaded into

Scaffold to enable exploration of the data and export count based on relative quantification (weighted spectral counts). These result files (.sf3) are also included in the PRIDE dataset (a viewer can be downloaded from Proteome Software) and protein inference settings may be changed. The raw data files can be found at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA726962/

2.5 Proteomic Statistics and Clustering of SCTLD+ and SCTLD- Samples from Nine Different Species

Weighted spectral counts were compared across SCTLD+ and SCTLD- samples irrespective of coral species. Fisher's Exact test with Bonferroni correction was used to calculate p-values for each protein group. Minimum count was set to 1.

Clustering and taxonomic grouping of identified peptides from both SCTLD+ and SCTLD- categories were conducted in Unipept (Verschaffelt, Van Den Bossche et al. 2021). Hierarchical clustering of taxa at the class level as well as Gene Ontology classification for the top 15 gene categories were conducted for comparison.

2.6 Sampling of Orbicella faveolata to Investigate SCTLD Proteomic Changes Over Two Time Periods and Between Multiple Regions Along the Florida Coast

Orbicella faveolata species were selected for a long-term study of SCTLD dynamics to address underpinnings of observed resistance and susceptibility among colonies. Collection permits and tissue biopsies were provided to NOAA as part of companion project by Nova Southeastern University, led by Dr. Brian Walker. Tissue biopsies were flash frozen in liquid nitrogen vapor shipper and transported to the Charleston SC, NOAA NOS Hollings Marine Laboratory. Coral biopsy samples were collected from two regions off the Florida coast: northern Broward County, FL in the southeast Florida Coral Reef Ecosystem and Conservation Area (ECA) and southern Florida Keys (FL Keys), comprising of Sand and Looe Keys. Fifteen replicates from each of three infection-rate groups were collected (Susceptible, Medium, Resistant) from each region, for a total of 90 biopsies. Each colony was sampled three times during a one-year period and proteomic analyses were performed on samples collected from two of the three sampling periods. Specimens were collected in May/June 2020 and February/March

2021 for proteomic analysis. A third collection during pre- and post- spawning was made August/September 2020 and included in other analyses by consortium members.

2.7 Protein Isolation of Orbicella faveolata

Frozen tissue biopsies were cryomilled into a fine powder. An aliquot of the tissue homogenate was used to extract a protein fraction with an SDS extraction buffer (5% SDS in 50mM triethylammonium bicarbonate; 1 Roche cOmplete EDTA-free protease inhibitor tablet per 50 mL lysis buffer) and tip sonicated on ice for 10 seconds at a time, 5 times. Samples were then centrifuged 2000 rcf for 5 minutes. The supernatant was decanted to a new vial, centrifuged 10000 rcf for 10 minutes, then the supernatant was again decanted to a new vial. Protein concentrations were determined with Pierce™ BCA Protein Assay Kit.

Protein (100 µg) was added to 80 µl in 5% SDS. Disulfide bonds were reduced (dithiothreitol) at 60 °C for 10 min. The samples were cooled at room temperature for 10 minutes prior to alkylation (calcium acetamide) and incubating in the dark for 30 minutes. Alkylation was quenched by adding 10 µl of 12% phosphoric acid. The proteins were digested with Protifi S-traps per the company protocol (Protifi, Fairport, NY, USA), using a 1:20 trypsin:substrate ratio and incubating for 2 hours at 47 °C. The resulting peptide solution was evaporated to dryness and reconstituted with 100 µl 0.1% formic acid in water. The peptide concentration was determined with Pierce™ Quantitative Fluorometric Peptide Assay kit.

Samples were sent to the proteomic core at the University of Arkansas Medical Sciences (UAMS), Little Rock AR, for nano-LC-MS/MS analysis on a Thermo Orbitrap™ Eclipse. Peptides were separated by reverse phase XSelect CSH C18 using an UltiMate 3000 RSLCnano system (Thermo Scientific). The instrument method used data-dependent acquisition with higher energy collisional fragmentation. Instrument performance and data quality were performed by the UMAS proteomic core facility.

2.8 Protein Identification and Functional Analysis

Peptide and protein identification was directed through Mascot and/or MS Fragger following conversion of raw files via Proteome Discoverer software (Thermo Fisher). The annotated genome of *Orbicella* spp. and coral symbionts from NCBI was used to identify proteins. Modifications were included based on concatenated error tolerant searches with acetyl

and oxidation modifications included in the search parameters. Identifications were accepted at <1% local false detection rate (FDR) for peptides and proteins. In the case of shared peptides, protein families were constructed and proteins containing the majority peptides were used for downstream classification and pathways analysis. Normalized spectral abundance factor (NSAF) and intensity Based Absolute Quantitation (IBAQ) intensity data were calculated to determine relative protein abundance.

The protein abundance data, NSAF or IBAQ, enables statistical analyses to identify significant differentially abundant proteins. Non-parametric tests (e.g., Wilcoxon Rank-Sum) are appropriate given the non-normal distribution of data, and a Benjamini-Hochberg (BH) correction were used to identify proteins with BH corrected p-value < 0.05 (equivalent to 5% false discovery). In addition to differential abundance analysis, ordination analyses (with non-parametric multidimensional scaling and/or principal component analyses) and hierarchical clustering were performed on all identified proteins. These analyses of identified proteins and their relative quantification were used to elucidate enriched pathways of expression, identifying molecular factors and mechanisms of susceptibility and resistance to SCTLD.

III. Results

3.1 Metagenomics of Coral from Nine Different Species

Of the 46 samples sent to Novogene, only 17 passed QC which required a minimal amount of DNA of 1 µg. Samples passing QC were assembled initially using MEGAHIT for Soil and Water (K-mer=55); parameter: --presets meta-large. The Scaffolds were cut off at "N" to get fragments without "N", called Scaftigs, which are small fragments of run reads that are overlapped to aligned together to form a complete run readClean data of all samples were mapped to assembled Scaftigs and unutilized reads were collected and combined to identify low abundance proteins. Summary statistics for gene assembly, gene prediction, taxonomic annotation, functional annotation, and antibiotic resistant genes are listed in **Table 2**. Following assembly and removal of redundancy by CD-HIT, 4,285,204 open-reading frames were identified from the metagenomic sequences and compiled into a protein FASTA database to use for a metaproteomic search. Of those 2,146,221 (50.08%) contain start and stop codons; whereas 90,448 (2.11%) contain neither a start or stop codon (only had reading frames from the middle). Open-reading frame length was relatively small for a majority of compiled sequences.

Table 2. Top-level data for metagenomic sequencing of 17 coral samples. Values describe the sequencing data following QC, assembly and gene prediction.

Data Clean		Assembly and Mix-Assembly		Gene Prediction	
Total Raw Data	122.28 Mbp	Scaffolds (Average)	425,289	Total ORFs	6,481,125
Average Raw Data	7.19 Mbp	Total length (nt)	9,150,373,596 bp	Average ORFs	381,243
Total Clean Data	122.07 Mbp	Average length (nt)	1,265.63 bp	Gene catalogue	4,285,204
Average Clean Data	7.18 Mbp	Longest length (nt)	1,172,196 bp	Complete ORFs	2,146,221(50.08%)
Effective percent	99.83%	N50 length (nt)	1,784.76 bp	Total length (Mbp)	1,336.42
Total Nohost Data	121.83 Mbp	N90 length (nt)	618.29 bp	Average length (bp)	311.87
Average Nohost Data	7.17 Mbp	Scaftigs (Average)	425,289	GC percent	49.57%
Effective rate	99.80%	Total length (nt)	9,150,373,596 bp		
		Average length (nt)	1,266 bp		
		N50 length (nt)	1,785 bp		
		N90 length (nt)	618 bp		

In a comparison of unique gene and common genes between corals with SCTLD+ versus SCTLD- coral samples, 1,208,206 genes were common between both groups. SCTLD- corals had 174,911 unique genes and SCTLD+ corals had 718,296 unique genes. Unique genes most likely reflect differences in species included in each group and not causative agents of disease or response to disease. Taxonomic distribution across the coral samples is shown. There was no statistical difference in taxonomic distribution between SCTLD+ and SCTLD- samples, although it should be noted that the metagenomic study was not designed to investigate differences between SCTLD+ and SCTLD- samples. The distribution only indicates a high level of diversity in the FASTA sequence database that was created for metaproteomics. With regards to distinct differences between diseased and non-diseased corals, a list of top-level functions representing the study population are listed in **Figure 1**.

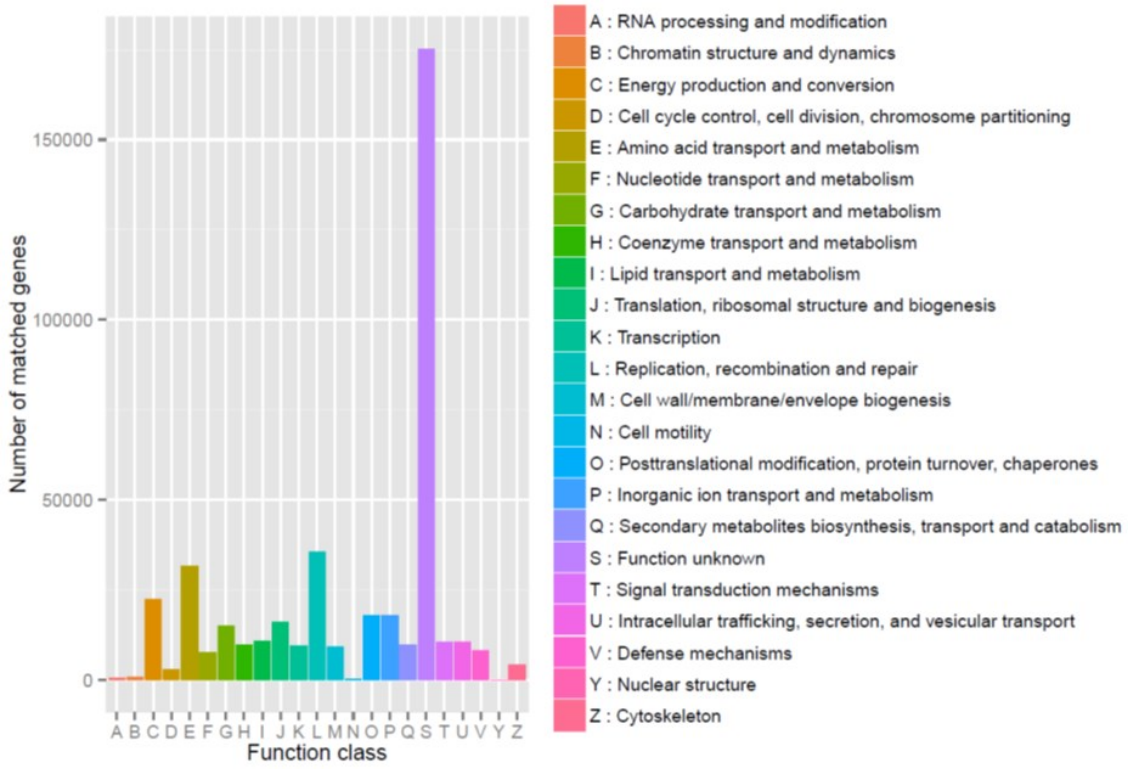

Figure 1. Functional grouping of genes across all sequenced taxa using EggNOG mapper (A database of orthology relationships, functional annotation, and gene evolutionary histories).

Although the majority of the genes are assigned the function of uncategorized (Figure 1, Column S), there remains a large distribution across 25 high-level functional categories. A list of viral DNA sequences matched to NCBI nr database were collated and provided in **Table 3**, which is often excluded in 16s rRNA studies.

Table 3. Species level classification of viruses identified from metagenome data. MetaGeneMark was used to predict open-reading frames, followed by reducing redundant sequences via CDHIT. Taxonomic classification was predicted by a similarity search against NCBI NR using MEGAN. Metagenome FASTA numbers are included for reference to indicate sample origin and sequence identifier. The major phages discovered include: Yellowstone Lake virophage 6 (17%), Organic Lake virophage (13%), and Klosneuvirus (7%) of all unique phage DNA sequences examined. Vibriophages comprise 4% of total phages and were only found in two coral taxa (*Orbicella annularis* and *Dichocoenia stokesii*). 9.6% of phages were identified as "uncultured Mediterranean phage".

Name	Metagenome FASTA number	No. of sequences
Ab18virus	OF.240_180148;OF.240_274484	2
Acaryochloris phage A-HIS2	DL.119_117063;DL.119_117063	2

Acinetobacter phage Acj61	OA.363_330496;OA.363_330496	2
Acinetobacter phage Acj9	OA.363_593584;OA.363_593584	2
African swine fever virus	DS.23_104946;DS.23_282133;DS.23_66887	3
Agrobacterium phage Atu_ph07	CN.94_198292;CN.94_218489;CN.94_275825;CN.94_319358;CN.94_60050;CN.94_8360;CN.94_198292;CN.94_218489;CN.94_275825;CN.94_319358;CN.94_60050;CN.94_8360	12
Ambidensovirus	DS.23_311367	1
Anopheles minimus irodovirus	DL.119_213888;DS.10_187629;DS.10_337338;DS.10_373887;DS.23_129199;DS.23_98082;DS.265_117141;DS.265_265282;DL.119_213888;DS.10_187629;DS.10_337338;DS.10_373887;DS.23_129199;DS.23_98082;DS.265_117141;DS.265_265282	16
Asfivirus	DS.23_104946;DS.23_282133;DS.23_66887	3
Bacillus phage v_B-Bak1	DL.73_142784;DL.73_142784	2
Bacillus thuringiensis phage MZTP02	OA.363_207167;OA.363_207167	2
Badnavirus	OA.241_51765	1
Barns Ness breadcrumb sponge narna-like virus 6	DS.10_239253;DS.37_233465;DS.10_239253;DS.37_233465	4
Barns Ness breadcrumb sponge weivirus-like virus 1	DS.37_508905;OF.261_151852;DS.37_508905;OF.261_151852	4
Bathycoccus sp. RCC1105 virus BpV	DS.10_243729;DS.10_421555;DS.37_398599;DS.37_50824;DS.37_584050;OA.241_45438	6
Bcep22virus	OA.363_408730	1
Beihai narna-like virus 6	OF.240_430680;OF.240_430680	2
Beihai narna-like virus 8	DS.37_574227;DS.37_574227	2
Beihai narna-like virus 9	DS.10_361826;DS.37_359004;DS.10_361826;DS.37_359004	4
Beihai sobemo-like virus 7	DS.37_277626;DS.37_734111;DS.37_277626;DS.37_734111	4
Bovine mastadenovirus A	DL.119_301101	1
Bracovirus	DS.37_93989	1
Bradyrhizobium phage BDU-MI-1	OA.363_60476;OA.363_60476	2

Brazilian marseillevirus	OF.261_176292	1
Burkholderia phage BcepNazgul	OA.363_876379;OA.363_876379	2
Burkholderia virus Bcepil02	OA.363_408730	1
Campylobacter phage vB_CjeM_Los1	OA.363_7042	1
Catovirus	OF.113_87239	1
Catovirus CTV1	OF.113_87239	1
Cherax quadricarinatus iridovirus	OF.261_227285;OF.261_227285	2
Chlamys acute necrobiotic virus	OF.113_63664;OF.113_63664	2
Chlorovirus	DS.10_496629	1
Chrysochromulina ericina virus	OF.113_150630;OF.113_150630	2
Cotesia congregata bracovirus	DS.37_93989	1
Cp8virus	OA.363_7042	1
Croceibacter phage P2559Y	DL.119_185676;DL.119_185676	2
Cronobacter phage vB_CsaP_Ss1	OA.363_471569;OA.363_471569	2
Cvm10virus	OA.363_960076	1
Diaphorina citri densovirus	DS.23_311367	1
Dinornavirus	OF.261_108583;OF.261_57047	2
environmental Halophage eHP-31	OA.363_353879;OA.363_353879	2
Flavobacterium phage 11b	DL.251_188954;DL.251_191065;DL.251_191066;DS.265_108503;DL.251_188954;DL.251_191065;DL.251_191066;DS.265_108503	8
Golden Marseillevirus	DS.265_491428	1
Haemophilus phage Aaphi23	OA.363_241730;OA.363_241730	2
Heterocapsa circularisquama RNA virus 01	OF.261_108583;OF.261_57047	2
Hokovirus	DL.119_388219	1

Hokovirus HKV1	DL.119_388219	1
Hot spring virus BHS1	OA.363_349661;OA.363_349661	2
Hubei narna-like virus 11	OF.35_362846;OF.35_362846	2
Ictalurid herpesvirus 2	CN.94_49490	1
Ictalurivirus	CN.94_49490	1
Indivirus	DS.10_248153;DS.23_216661;DS.265_290150;DS.265_408681	4
Indivirus ILV1	DS.10_248153;DS.23_216661;DS.265_290150;DS.265_408681	4
Invertebrate iridescent virus 6	DS.265_401950	1
Invertebrate iridescent virus 9	OF.113_220750	1
Iridovirus	DS.265_401950;OF.113_220750	2
Kiln Barn virus	DS.37_364062;DS.37_364062	2
Klosneuvirus	DL.119_243706;DL.119_393222;DL.251_115884;DL.251_185509;DS.10_212482;DS.10_23755;DS.10_288272;DS.10_307272;DS.10_351139;DS.10_368152;DS.10_397891;DS.10_462000;DS.23_94260;DS.265_13562;DS.265_183035;DS.265_217902;DS.265_222663;DS.265_371079;DS.265_490264;DS.265_73141;DS.265_86771;DS.37_141618;DS.37_318707;DS.37_342459;DS.37_351457;DS.37_395947;DS.37_546732;DS.37_709051;DS.37_753584;OA.241_209917;OA.241_290057;OA.241_297778;OA.241_334802;OA.241_65831;OA.241_66065;OA.363_714510;OA.363_773449;OF.240_508275;OF.261_111705	39
Lactobacillus virus LP65	OA.241_118560;OA.241_118560	2
Likavirus	OA.363_410880	1
Marseillevirus	DS.265_491428;OF.261_176292	2
Mastadenovirus	DL.119_301101	1
Megavirus chiliensis	DS.23_276989	1
Methanosarcina spherical virus	OA.241_7794;OA.241_7794	2
Microbacterium phage Min1	OA.363_295386;OA.363_325457;OA.363_295386;OA.363_325457	4
Micromonas sp. RCC1109 virus MpV1	CN.92_132083	1
Mimivirus	DS.10_329889;DS.23_276989	2
Mimivirus-dependent virus Sputnik	DS.37_335430	1

Mimivirus AB-566-O17	CN.94_56583;CN.94_56583	2
Moumouvirus goulette	DS.10_329889	1
Np1virus	OF.113_55024	1
Organic Lake phycodnavirus 2	CN.94_11835;CN.94_11835	2
Organic Lake virophage	CN.92_36356;CN.96_195725;CN.96_252080;DL.73_286777;DS.10_120928;DS.10_162865;DS.10_191645;DS.10_255715;DS.10_282296;DS.10_367135;DS.10_424870;DS.10_58167;DS.10_75222;DS.23_159992;DS.265_381004;DS.265_503030;DS.265_59472;DS.265_79129;DS.37_102061;DS.37_186038;DS.37_277932;DS.37_347084;DS.37_351517;DS.37_433685;DS.37_439312;DS.37_462447;DS.37_553594;DS.37_635112;DS.37_659561;DS.37_677883;DS.37_68899;DS.37_740739;DS.37_93030;OA.241_384704;OA.363_416198;OF.261_344664;OF.35_378096;OF.35_41825;PS.114_180551;CN.92_36356;CN.96_195725;CN.96_252080;DL.73_286777;DS.10_120928;DS.10_162865;DS.10_191645;DS.10_255715;DS.10_282296;DS.10_367135;DS.10_424870;DS.10_58167;DS.10_75222;DS.23_159992;DS.265_381004;DS.265_503030;DS.265_59472;DS.265_79129;DS.37_102061;DS.37_186038;DS.37_277932;DS.37_347084;DS.37_351517;DS.37_433685;DS.37_439312;DS.37_462447;DS.37_553594;DS.37_635112;DS.37_659561;DS.37_677883;DS.37_68899;DS.37_740739;DS.37_93030;OA.241_384704;OA.363_416198;OF.261_344664;OF.35_378096;OF.35_41825;PS.114_180551	78
Orthopoxvirus	OF.240_43605	1
Ostreavirus	OF.113_31856;OF.113_41052	2
Ostreid herpesvirus 1	OF.113_31856;OF.113_41052	2
Pacific flying fox faeces associated circular DNA virus-3	OF.113_175889;OF.113_175889	2
Pacmanvirus A23	DL.251_189462;DL.73_353477;DS.23_218322;PS.114_109055;DL.251_189462;DL.73_353477;DS.23_218322;PS.114_109055	8
Paenibacillus phage PG1	OA.363_842507;OA.363_842507	2
Pandoravirus	DS.37_89586;OF.261_173183;OF.261_346738	3
Pandoravirus dulcis	OF.261_346738	1
Pandoravirus salinus	DS.37_89586	1
Phaeocystis globosa virus	OA.363_334611;OA.363_974637	2

Phaeocystis globosa virus virophage	DS.265_489168;DS.265_489168	2
Piper yellow mottle virus	OA.241_51765	1
Pis4avirus	DS.10_213681;DS.37_22661	2
Prasinovirus	CN.92_132083;DS.10_243729;DS.10_421555;DS.37_398599;DS.37_50824;DS.37_584050;OA.241_45438	7
Prymnesiovirus	OA.363_334611;OA.363_974637	2
Rdjlvirus	OF.240_534442	1
Rhizobium phage vB_RleM_P10VF	OF.240_127479;OF.240_127479	2
Rhizobium phage vB_RleS_L338C	DS.265_424241;DS.265_482091;DS.265_424241;DS.265_482091	4
Saccharomonospora phage PIS 136	OA.363_368276;OA.363_961201;OA.363_368276;OA.363_961201	4
Salicola phage SCTP-2	CN.94_123584;CN.94_15242;CN.94_28856;CN.94_123584;CN.94_15242;CN.94_28856	6
Salmonella phage Skate	DS.10_213681;DS.37_22661	2
Sk1virus	DS.10_323387	1
Sputnikvirus	DS.37_335430	1
Symbiodinium +ssRNA virus TR74740 c13_g1_i1	CN.94_305793;DS.10_488343;DS.265_361621;DS.37_121930;DS.37_121931;DS.37_499682;CN.94_305793;DS.10_488343;DS.265_361621;DS.37_121930;DS.37_121931;DS.37_499682	12
Symbiodinium +ssRNA virus TR74740 c13_g1_i2	DS.10_493117;DS.10_493117	2
Synechococcus phage S-CAM7	OA.363_918661;OA.363_918661	2
Synechococcus phage S-CBM2	OA.363_443137;OA.363_443137	2
Tetraselmis virus 1	CN.94_37707;DS.265_116605;DS.37_320867;DS.37_438892;CN.94_37707;DS.265_116605;DS.37_320867;DS.37_438892	8
Thermobifida phage P1312	OA.363_100168;OA.363_578254;OA.363_669080;OA.363_918676;OF.113_205116;OF.113_279362;OF.113_67267;OA.363_100168;OA.363_578254;OA.363_669080;OA.363_918676;OF.113_205116;OF.113_279362;OF.113_67267	14
Tupanvirus	CN.92_223268;CN.94_196476;CN.94_3674;CN.96_232195;CN.96_24214;DS.10_126044;DS.10_428689;DS.10_493163;DS.10_88493;DS.23_186248;DS.23_201736;DS.37_144008;DS.37_439380;DS.37_501923;DS.37_522451;DS.37_689447;OF.35_119810;OF.35_223349	18

Tupanvirus soda lake	CN.92_223268;CN.94_196476;CN.94_3674;CN.96_232195;DS.10_126044;DS.10_88493;DS.23_186248;DS.37_144008;DS.37_439380;DS.37_501923;DS.37_522451;OF.35_223349	12
uncultured marine virus	DL.119_376360;DL.119_376360	2
uncultured Mediterranean phage	OA.363_158119;OA.363_424888;OA.363_459538;OA.363_482239;OA.363_570762;OA.363_572894;OA.363_691541;OA.363_761348;OF.113_783;OA.363_158119;OA.363_424888;OA.363_459538;OA.363_482239;OA.363_570762;OA.363_572894;OA.363_691541;OA.363_761348;OF.113_783	18
uncultured Mediterranean phage uvDeep-CGR0-AD1-C123	OA.363_495305;OA.363_495305	2
uncultured Mediterranean phage uvDeep-CGR2-KM18-C269	OF.35_86548;OF.35_86548	2
uncultured Mediterranean phage uvDeep-CGR2-KM23-C198	OA.363_364094;OA.363_827249;OA.363_364094;OA.363_827249	4
uncultured Mediterranean phage uvDeep1-CGR2-KM23-C896	OF.261_19877;OF.261_19877	2
uncultured Mediterranean phage uvMED	DL.119_10637;DL.119_378575;DL.73_33798;OA.363_114476;OA.363_276016;OA.363_369037;OA.363_561334;OA.363_615068;OA.363_808939;OF.113_10981;OF.113_144329;OF.113_185278;OF.113_83866;OF.240_280464;DL.119_10637;DL.119_378575;DL.73_33798;OA.363_114476;OA.363_276016;OA.363_369037;OA.363_561334;OA.363_615068;OA.363_808939;OF.113_10981;OF.113_144329;OF.113_185278;OF.113_83866;OF.240_280464	28
uncultured virus	OA.363_118048;OF.113_205787;OA.363_118048;OF.113_205787	4
Vibrio phage 1.117.O._10N.261.45.E9	OA.363_530744;OA.363_530744	2
Vibrio phage 1.161.O._10N.261.48.C5	DS.10_79051;DS.23_116376;DS.23_141729;DS.23_250983;DS.265_145526;DS.265_414694;DS.37_511810;DS.10_79051;DS.23_116376;DS.23_141729;DS.23_250983;DS.265_145526;DS.265_414694;DS.37_511810	14
Vibrio phage 1.215.A._10N.222.54.F7	OA.363_633944;OA.363_633944	2
Vibrio phage VvAW1	OA.363_119221;OA.363_557335;OA.363_119221;OA.363_557335	4

Wenzhou weivirus-like virus 1	OF.240_420284;OF.240_420284	2
Wizardvirus	OA.363_29130	1
Wuchan romanomermis nematode virus 2	DS.37_27348;DS.37_27348	2
Xanthomonas phage XacN1	OA.363_272014;OA.363_272014	2
Yellowstone Lake virophage 6	CN.92_23675;CN.94_139908;CN.94_334010;CN.94_78344;CN.96_112484;CN.96_258115;DL.119_172621;DL.119_72384;DL.251_156984;DL.251_183016;DL.251_88242;DS.10_265698;DS.10_316498;DS.10_339852;DS.10_365483;DS.10_382826;DS.10_448654;DS.10_58004;DS.10_98442;DS.23_12449;DS.265_15696;DS.265_323366;DS.265_362324;DS.265_415820;DS.37_148278;DS.37_204703;DS.37_270161;DS.37_345785;DS.37_376427;DS.37_463945;DS.37_504968;DS.37_542328;DS.37_644373;DS.37_73969;OA.241_108669;OA.241_14145;OA.241_158105;OA.241_23889;OA.241_26197;OA.241_310082;OA.241_315795;OA.241_320250;OA.363_278335;OF.240_190295;OF.240_200944;OF.240_232526;OF.240_562500;OF.261_239766;OF.261_511633;OF.261_593674;OF.35_349502;OF.35_355451;CN.92_23675;CN.94_139908;CN.94_334010;CN.94_78344;CN.96_112484;CN.96_258115;DL.119_172621;DL.119_72384;DL.251_156984;DL.251_183016;DL.251_88242;DS.10_265698;DS.10_316498;DS.10_339852;DS.10_365483;DS.10_382826;DS.10_448654;DS.10_58004;DS.10_98442;DS.23_12449;DS.265_15696;DS.265_323366;DS.265_362324;DS.265_415820;DS.37_148278;DS.37_204703;DS.37_270161;DS.37_345785;DS.37_376427;DS.37_463945;DS.37_504968;DS.37_542328;DS.37_644373;DS.37_73969;OA.241_108669;OA.241_14145;OA.241_158105;OA.241_23889;OA.241_26197;OA.241_310082;OA.241_315795;OA.241_320250;OA.363_278335;OF.240_190295;OF.240_200944;OF.240_232526;OF.240_562500;OF.261_239766;OF.261_511633;OF.261_593674;OF.35_349502;OF.35_355451	104

3.2 Proteomics of SCLTD+ and SCTLD- Coral Samples from Nine Different Species

We initially examined protein alterations in corals that had varying degrees of susceptibility to SCTLD, comparing SCTLD+ and SCTLD- samples. Proteins were isolated from 46 samples that spanned nine species with and without SCTLD. These nine species were selected to represent species that are susceptible to SCTLD with varying disease dynamics. There are six species that were characterized as highly susceptible which were, *Colpophyllia natans, Dichocoenia stokesii, Diploria labyrinthiformis, Meandrina meandrites, Pseudodiploria strigosa,* and *Pseudodiploria clivosa,* and there were three species that demonstrate intermediary

susceptibility, which were *Montastraea cavernosa, Orbicella annularis,* and *Orbicella faveolata* (agrra.org).

Across the 46 samples included in the study, 8,552 proteins were identified (1% FDR), where only 18 proteins were statistically higher and 5 proteins lower in the SCTLD+ group compared to the SCTLD- group (**Figure 2**, p<0.00016, Fisher's Exact Test).

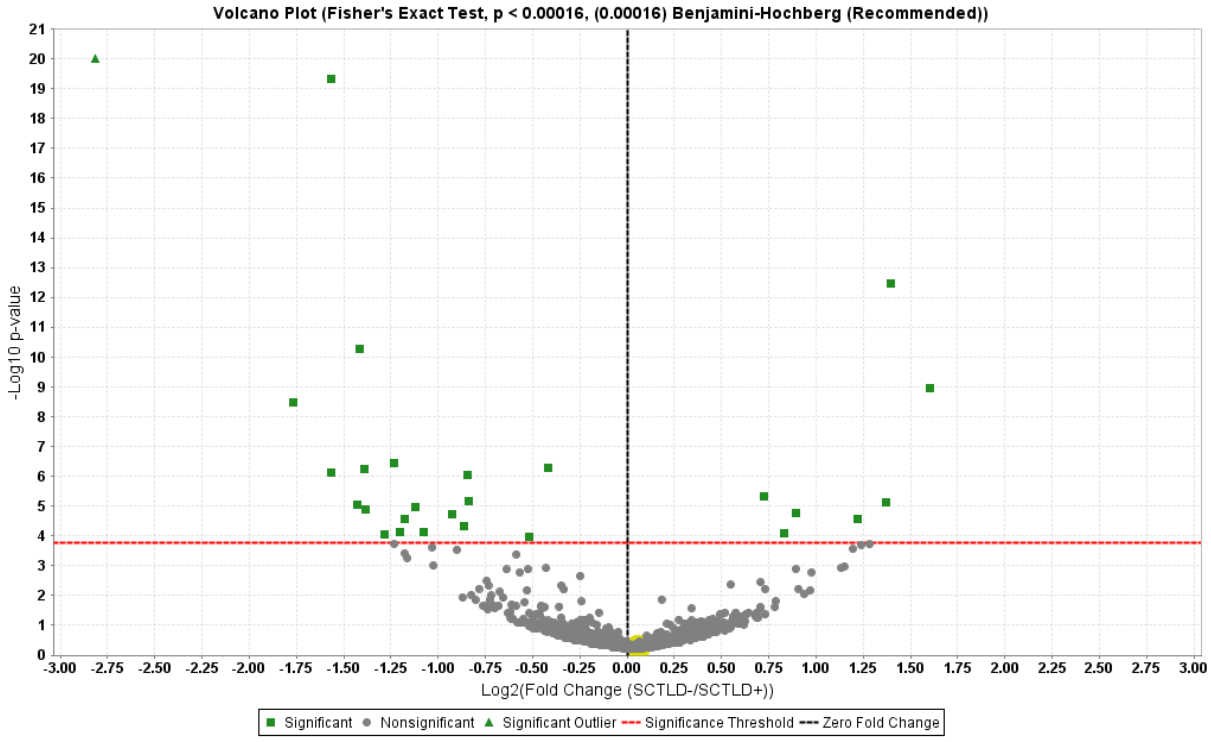

Figure 2. Volcano plot [log10 p-value vs fold-change] of proteins identified in this study. Proteins above the red line indicate significance. Dotted vertical line at 0 indicates unity.

Proteins were putatively identified from combined Uniprot and metagenome databases. Further annotation of the differential metaproteome data were made using batch Blast search and are displayed in (**Table 4**). Across taxa, there were no single proteins that allowed SCTLD+ to be perfectly classified from SCTLD- groups. These data concur with metagenome data; although, the metagenome data contained only an abbreviated number of samples which did not allow a complete comparison, as did the proteomic analysis. In short, there is no pan-taxa protein classifier of SCTLD according to this analysis.

Table 4. Comparison of differentially abundant proteins between SCTLD+ vs SCTLD- coral groups. Identified proteins are listed as protein families or clusters. Accession number or metagenome FASTA number from the translated CDHIT metagenome library are indicated in the second column. 18 proteins were elevated and 5 proteins were depressed.

Identified Proteins	Accession Number or Metagenome Fasta Number	Molecular Weight	Fisher's Exact Test ($p < 0.00016$)	Fold Change (SCTLD+ vs SCTLD-)
chitinase-3-like protein 1	OF.261_242309	10 kDa	< 0.00010	7
N-acetyl-D-galactosamine binding lectin precursor	CN.92_12282	24 kDa	< 0.00010	4
glycoside hydrolase family 18	DL.119_176235	34 kDa	< 0.00010	3
tumor necrosis factor ligand superfamily member 11-like	OA.241_364575	13 kDa	< 0.00010	3
cyan fluorescent protein	A7UAL1	26 kDa	< 0.00010	2.7
chitinase-3-like protein 1	OF.261_29881	13 kDa	< 0.00010	2.7
ubiquitin-like modifier-activating enzyme 5	CN.92_32059	48 kDa	< 0.00010	2.6
uncharacterized protein LOC110066892 [*Orbicella faveolata*]	CN.92_140525	26 kDa	< 0.00010	2.3
putative cysteine desulfurase	OA.241_248375	29 kDa	< 0.00010	2.2
cysteine dioxygenase family	DS.10_64707	53 kDa	< 0.00010	2.2
tyrosinase-like	OF.240_519741	66 kDa	< 0.00010	2.2
argininosuccinate synthase	A0A2B4S7F2	94 kDa	< 0.00010	2
choloylglycine hydrolase	DS.10_190139	45 kDa	< 0.00010	1.9
tumor necrosis factor ligand superfamily member 11-like	OF.35_88657	13 kDa	< 0.00010	1.8
uncharacterized protein LOC110066892 [*Orbicella faveolata*]	CN.92_140526	34 kDa	< 0.00010	1.8
uncharacterized protein LOC107328915 isoform X2 [*Acropora digitifera*]	DL.119_55130	38 kDa	< 0.00010	1.7
glutamate dehydrogenase	CN.92_9756	49 kDa	< 0.00010	1.3
Actin, cytoplasmic	A0A2B4S283	42 kDa	< 0.00010	1.2
sporulation-specific protein 15	DS.23_139989	329 kDa	< 0.00010	0.6
protocadherin-like protein	DS.10_14288	148 kDa	< 0.00010	0.5
Green fluorescent protein-like	Q86LV7	24 kDa	< 0.00010	0.4
uncharacterized protein LOC117107688 [*Anneissia japonica*]	DS.23_199250	30 kDa	< 0.00010	0.4
DELTA-actitoxin-Ate1a-like	OF.35_18244	14 kDa	< 0.00010	0.3

Peptides that matched to amino acid sequences in the FASTA databases were exported for all coral taxa with SCTLD+ and SCTLD- groups. Exported peptides were uploaded into Unipept and classified for taxonomy and function (Verschaffelt et al. 2021). Domain level unique peptide membership from 8519 classified peptides in SCTLD+ samples was distributed as: eukaryotes, 5004 peptides; bacteria, 708 peptides; archaea, 14 peptides; viruses, 1 peptide. Domain level unique peptide membership from 8441 classified peptides in SCTLD- samples was distributed as: eukaryotes, 5014 peptides; bacteria, 659 peptides; archaea, 15 peptides; viruses, 1 peptide. Approximately 14% of peptide spectral matches were classified as non-eukaryote peptides in both samples.

Gene ontology mapping of peptides from both SCTLD+ and SCTLD- corals revealed similar enrichment of biological processes: SCTLD+, carbohydrate metabolic process, intracellular protein transport, vesicle-mediated transport, and cell adhesion; SCTLD-, carbohydrate metabolic process, intracellular protein transport, vesicle-mediated transport, and bioluminescence. Cell adhesion was represented more in SCTLD+ corals while bioluminescence was represented more in SCTLD- corals. Finer comparisons of peptide distributions for both groups using heatmaps revealed very few differences between SCTLD+ and SCTLD- corals based on taxonomy. At the class level, peptides from red algae (Florideophyceae) were more abundant in SCTLD+ corals; whereas, peptides from Alphaproteobacteria were more abundant in the SCTLD- corals.

Top GO terms associated with both groups were ATP binding, integral component of membrane, and cytoplasm; neither of which were discriminatory. Across all GO categories, both groups were largely similar in distribution. Only biological processes showed slight variation in peptide association. Biological processes membership was interrogated deeper to include the top 16-30 categories. Several categories showed differences in peptide membership between SCTLD+ and SCTLD- corals. For example, more peptides from SCTLD+ corals were included in the ubiquitin-dependent protein catabolic process category than were found for SCTLD-. Bioluminescence continued to show differences in peptide membership where SCTLD+ corals have fewer peptides related to bioluminescence compared to SCTLD- corals.

Noting how many of the differences were driven by taxon-specific protein identification, we resubmitted .raw files to search spectra against taxon-specific translated metagenome files, except for *Meandrina meandrites* which was searched against Uniprot taxa ID 6125

(Scleractinia, stony corals) because metagenome data was not available for that set of samples. Six species were included in the taxon-specific analysis: *Dichocoenia labyrinthiformis, Orbicella annularis, Meandrina meandrites, Orbicella faveolata, Colpophyllia natans,* and *Dichocoenia stokesii*, all of which had balanced groups of SCTLD+ and SCTLD- individuals. Significantly different proteins at the taxon-specific level were collated and categorized using gene ontology terms: molecular function and biological process. Top categories were exported based on majority membership across species e.g., if ontology terms were associated with different proteins ($p<0.05$) in 6 out of 6 species, this category was prioritized based on function. No ontology term was associated with 6 out of 6 species, but several terms were common across multiple species comparisons. The categories "bioluminescence" and "generation of precursor metabolites and energy" had the highest membership and consequently the results were driven by a reduction in green fluorescent protein-like proteins.

In reanalyzing the previous data of nine different species of SCTLD+ and SCTLD- coral, we identified 5,804 protein clusters within all samples, only 97 were significantly different ($p<0.05$) between SCTLD+ and SCTLD- samples across all species; however, not all peptides were identified in each sample. Of all the clusters identified, only 25 clusters were present in all 46 samples, of which only one was significant between SCTLD+ and SCTLD-, Histone H2B. The significant data was next evaluated for a significant fold change between SCTLD+ and SCTLD- samples, which was determined to be greater than 25% in either direction and did not have to have a peptide match in all 46 samples. In total, 891 protein clusters had an increase in protein expression with SCTLD+, while 393 protein clusters had a decrease in protein expression with SCTLD+. Of these, only 19 protein clusters had a significant decrease in protein expression ($p<0.05$) with SCTLD+, while 73 protein clusters had a significant increase in protein expression ($p<0.05$) with SCTLD+. The proteins were associated with serine/threonine-protein kinases, ubiquitin proteins, lipid metabolism, chlorophyll-binding proteins, histone H2B, protocadherin-like proteins, metabolic proteins, and numerous uncharacterized protein clusters.

3.3 Proteomics of Orbicella faveolata Samples Over Two Time Periods and Two Different Regions

SCTLD was evaluated in *Orbicella faveolata* samples that had varying degrees of SCTLD. These samples were collected from two regions across two time periods. These

samples were grouped initially based on susceptibility to SCTLD, which was determined by the number of visits in which SCTLD was present on the colony over a year. Initially, 5,623 peptides were identified in Period 1 samples from all the different regions, while 5,619 peptides were identified in Period 3 samples. Due to the large number of peptides and multiple sampling combinations (time and location comparisons), we decided to limit the analysis initially to include only proteins that were present in all samples within each collection period. There were 169 proteins that appeared in all 84 samples from Period 1, and 129 proteins that appeared in all 87 samples from Period 3.

Of the 169 proteins identified in all samples collected in Period 1, there were 25 proteins whose relative abundance changed significantly ($p<0.10$) in the "SCTLD susceptible" group compared to "SCTLD resistant" group. Interestingly, there were less changes overall in Period 3 compared to changes seen during Period 1. In Period 3, when comparing the "SCTLD susceptible" group with the "SCTLD resistant" group, none of the original 129 proteins identified in all samples were significantly different. This would suggest that changes in these periods might be influenced by environmental factors rather than disease.

The samples were collected from two general regions, ECA (Broward County) and the Lower Keys (Looe and Sand Key). Due to the two broad locations, we wanted to evaluate to determine whether changes were consistent between both areas when analyzed separately within each sampling period. ECA samples generally had a decrease in relative protein abundances of the changing proteins among the "SCTLD susceptible" samples compared to the "SCTLD resistant" samples, while Looe and Sand Keys had the opposite pattern. In Looe and Sand Keys, the "SCTLD susceptible" samples had a higher protein abundance compared to the "SCTLD resistant" samples. This observation raised the questions: are there regional differences in response to SCTLD? and if so, what factors might be contributing to the coral's SCTLD variation in susceptibility?

When the ECA were separated from the Keys samples, we found 40 proteins during Period 1 samples from the ECA that were significantly different between "susceptible" and "resistant" groups. In the Keys, there were 115 proteins in Period 1 samples that were significantly different between the "susceptible" and "resistant" groups. The ECA and Lower Keys samples had 23 protein changes in common (based on % fold change). However, in these regions only two of the 23 proteins had a significantly different relative abundance during Period

1, indicating that these regions might be responding to disease differently due to disease severity or perhaps the protein changes were dictated by environmental factors rather than disease.

Interestingly in Period 3, there were no common protein abundance changes between either ECA or Lower Keys "susceptible" versus "resistance" groups, mostly due to the fact that there were few significant protein changes. There was one significant protein change in ECA "susceptible" and "resistant" groups, while when comparing "susceptible" and "resistant" groups in the Lower Keys, there were nine significant protein changes.

ECA and the Lower Keys regions had varying protein changes between susceptibility groups and collection periods. These changes could be attributed to location of the samples and all differences related to that (weather, temperature, salinity, pH, etc). Therefore, we asked if the changes between the susceptibility groups in each region were consistent between collection periods. We found that they were not; the changes in Period 1 were consistently greater (40 v 1 in ECA and 115 v 9 in Keys) and didn't share any common proteins with Period 3. This result was perplexing as it was unclear why there was such variation in protein changes between Period 1 and Period 3 sampling. It's unknown if the variations are due to normal cyclic physiological changes, weather, environmental factors, toxins, or perhaps antibiotic exposure because there was no apparent correlation to disease state.

3.4 Developing New Dynamic Groups for Orbicella faveolata SCTLD-affected Corals

The antibiotic treatment history was available for these colonies, so next we investigated whether there were any patterns in the protein expression that corresponded to antibiotic treatments, as the frequency and number of treatments varied between regions and collection periods.

During Period 1, there were minimal treatments (based on number and frequency) in ECA "susceptible" group that were done prior to the samples being collected, while in Looe and Sand Keys "susceptible" groups, there were ongoing antibiotic treatments for infections during collection (**Figure 3,** top, middle, bottom panels, red arrow). When evaluating Period 3 (**Figure 3**, blue arrow), note that most samples in Sand Key had not been treated in almost 6 months prior collection, whereas ECA and Looe Key had both received treatments one to two months prior to sampling. Similarly, the number of Looe Key samples treated was far greater than either of the other two areas. Because of the significant difference in Looe Key and Sand Key history, it was

decided to split these into two distinct regions, rather than keeping them combined as “Lower Keys”.

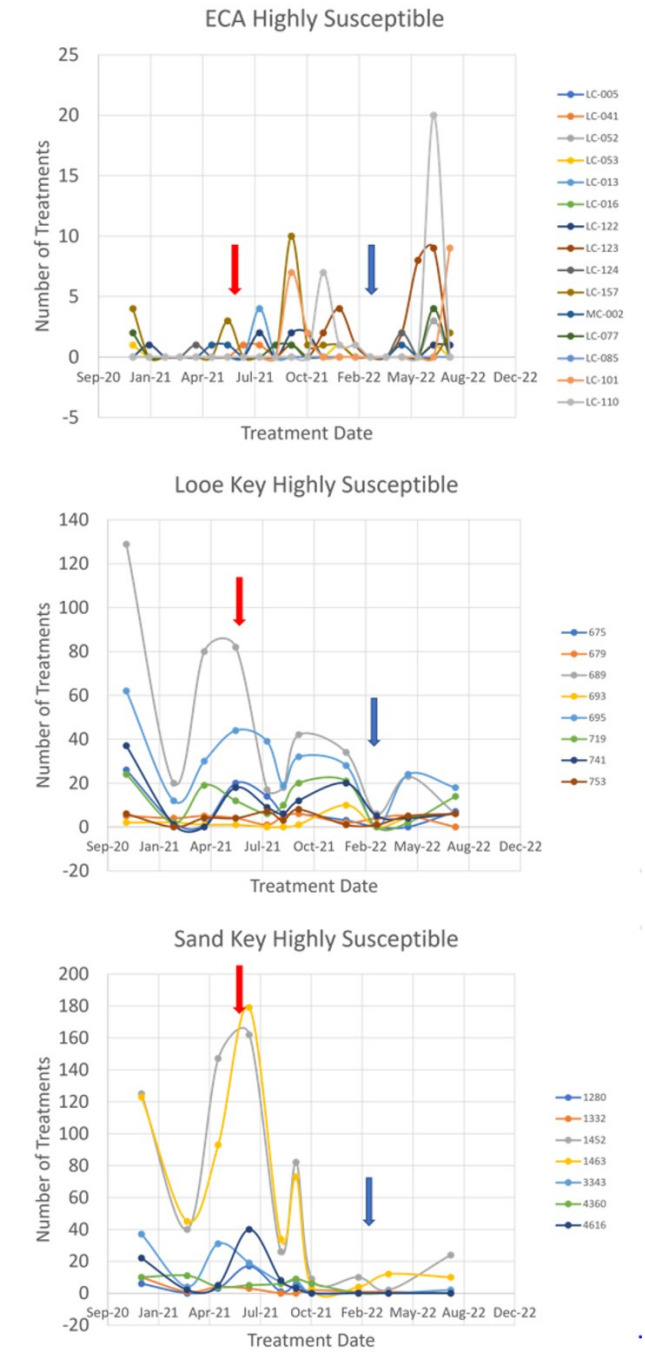

Figure 3. These graphs show the history of antibiotic treatments of individual “susceptible” corals in the different regions. Red arrows indicate when Period 1 samples were taken. Blue arrow indicates when Period 3 samples were collected.

Based on these differences among individual colonies in the study, the groupings (classifications) were restructured based on disease treatment history relative to sampling (similar to an anamnesis). While samples may have had similar numbers of treatments, those that had disease and treatments close to the sampling were different from those that had disease and treatments more than 6 months prior to sampling. Therefore, five new clustering groups (classifications) based on the coral's "patient" history were constructed, providing a new perspective for data analysis.

The data was re-organized into five groups: (1) Active (disease was present during sampling), (2) 2-6 months prior (disease was present 2-6 months prior to sampling), (3) >6 months prior (disease was present greater than 6 months prior to sampling), (4) After (disease developed after sampling), (5) Never (disease never developed). The diseased >6 months prior to sampling was later split into two separate groups: diseased >6 months prior to sampling and never reinfected, diseased >6 months prior to sampling and later reinfected after sampling.

Initially, t-tests were performed to find significance between each comparison of the first five groups. The number of proteins that significantly ($p<0.05$) differed between each group in just the ECA are shown in **Table 5**.

Table 5. Number of significant protein changes between different susceptibility groups in the ECA colonies during Period 1.

ECA	Active	2-6 months Prior	>6 months Prior	After	Never
Active	X	0	**25**	1	9
2-6 months Prior		X	5	1	8
>6 months Prior			X	**60**	5
After				X	**44**
Never					X

In the ECA, there were a total of 158 proteins that changed only between two distinct outcome groups. The greatest changes occurred between the >6 months Prior v After, the Active vs >6 months Prior, and the After vs Never (bolded in **Table 5**).

Table 6. Significant protein changes between different susceptibility groups in Looe Key colonies during Period 1.

Looe Key	Active	>6 months prior	After	Never
Active	X	**45**	10	**157**
>6 months Prior		X	17	6
After			X	**36**
Never				X

In Looe Key samples from Period 1, there were 271 proteins that only changed between one pair of groups. The greatest changes were between the Active and Never groupings, where 157 proteins were significantly different ($p<0.05$) (bolded in **Table 6**).

Table 7. Significant protein changes between different susceptibility groups in Sand Key colonies during Period 1.

Sand Key	Active	>6 months Prior	Never
Active	X	**17**	**19**
>6 months Prior		X	2
Never			X

Sand Key had the fewest significant changes of all three regions. In Sand Key samples from Period 1, there were 19 proteins that differed significantly between Active vs Never and 17 proteins that differed significantly between Active and >6 months Prior (**Table 7**).

Table 8. Significant protein changes between different susceptibility groups in the ECA colonies during Period 3.

ECA	Active	2-6 months Prior	>6 months prior	After	Never
Active	X	0	0	0	1
2-6 months Prior		X	0	1	3
>6 months Prior			X	4	1
After				X	4
Never					X

Table 9. Similar protein changes between different susceptibility groups in Looe Key colonies during Period 3.

Looe Key	Active	2-6 months Prior	Prior	Never
Active	X	13	1	1
2-6 months Prior		X	**158**	0
Prior			X	0

Never		X

Fewer proteins were changed in Period 3 among all three regions, ECA, Looe Key and Sand Key (**Tables 8, 9, 10, respectively**). This could be due to seasonal environmental or physiological changes. It's interesting that Sand Key also had the fewest treatments prior to sampling, which might suggest that the changes seen can be teased apart from antibiotic treatment, whereas Looe Key showed the most protein changes among all three regions and also had multiple treatments prior to Period 3 sampling, so the drastic increase in protein changes might be attributed to antibiotic treatment.

Table 10. Similar protein changes between different susceptibility groups in Sand Key colonies during Period 3.

Sand Key	Active	Just Prior	Prior	Never
Active	X	1	3	1
Just Prior		X	3	2
Prior			X	2
Never				X

A general trend showed protein expression decreased in the colonies that had active disease >6 months prior to sampling, while it was increased when disease was active during sampling or the colonies were diseased post-sampling. This would indicate that there might be an overall upregulation in protein machinery prior to and during active infection that subsides and decreases when the disease is absent or inactive.

Given the fluidity of these groupings, we asked whether any of the different outcome groups shared proteins changes. In evaluating the proteins that changed similarly between two different outcome comparisons in Period 1 ECA, Never and the >6 months Prior outcome groups individually shared 13 common protein changes that were significantly different when compared to the colonies with active disease during sampling. In Looe Key, among the >6 months Prior and Never compared to Active, there were 46 proteins that differed significantly and were seen in both comparisons. In Sand Key, shared protein changes occurred between Active vs >6 months Prior and Active vs Never with 6 common proteins that changed. The similarities between Never and >6 months Prior compared to Active would indicate proteins that are associated most with a non-diseased or inactive disease state compared to Active disease state.

Table 11. Significant protein changes between Active and Inactive Disease States in ECA, Looe Key and Sand Key.

	Active vs Never & Active vs >6 months Prior	Never and >6months Prior vs Active
ECA	13	222
Looe	46	664
Sand	6	221

When >6 months Prior and Never samples were combined together and compared against the Active group, there were 222 proteins that differed in ECA, 664 that differed in Looe Key and 221 proteins that differed in Sand Key (**Table 11**). Next, we asked how many of the changes in these groups were similar across all three regions and found that 47 protein changes were shared across all three regions indicating that these proteins were most likely associated with an active vs nonactive disease state (**Table 12**).

Table 12. List of common protein changes among all three regions

FASTA Headers	KEGG Pathway
XP_020613781.1 tyrosine-protein kinase SRK2-like	Adherens Junction
XP_020627630.1 RAC-gamma serine/threonine-protein kinase-like	AKT Pathway
XP_020600494.1 glutamine--tRNA ligase-like	Aminoacyl-trna biosynthesis
XP_020601817.1 calcium/calmodulin-dependent protein kinase type II delta chain-like isoform X2 ;XP_020601810.1 calcium/calmodulin-dependent protein kinase type II delta chain-like isoform X1	AMPK signaling
XP_020606200.1 serine/threonine-protein phosphatase 2A catalytic subunit beta isoform ;XP_020606199.1 serine/threonine-protein phosphatase 2A catalytic subunit beta isoform	Autophagy/AMPK signaling
XP_020629422.1 coadhesin-like isoform X2 ;XP_020629421.1 coadhesin-like isoform X1	Calcifying processes
XP_020609557.1 adenosylhomocysteinase B-like	Cysteine and methionine metabolism
XP_020602177.1 plastin-3-like	Cytoskeleton
XP_020610235.1 LOW QUALITY PROTEIN: AP-2 complex subunit mu-like	Endocytosis
XP_020621799.1 LOW QUALITY PROTEIN: vacuolar protein sorting-associated protein 13A-like	Endocytosis
XP_020630784.1 phosphatidylinositol-binding clathrin assembly protein LAP-like	Endocytosis
XP_020632450.1 WASH complex subunit 5-like	Endocytosis
XP_020602569.1 extended synaptotagmin-2-A-like	ER membrane
XP_020611683.1 LOW QUALITY PROTEIN: long-chain-fatty-acid--CoA ligase 1-like	Fatty acid degradation

XP_020603515.1 UTP--glucose-1-phosphate uridylyltransferase-like isoform X2 ;XP_020603514.1 UTP--glucose-1-phosphate uridylyltransferase-like isoform X1	Galactose Metabolism
XP_020600978.1 peroxiredoxin-1-like isoform X1 ;XP_020600979.1 peroxiredoxin-1-like isoform X2 ;XP_020613504.1 peroxiredoxin-4-like	Glutathione metabolism
XP_020609612.1 LOW QUALITY PROTEIN: isocitrate dehydrogenase [NADP] cytoplasmic-like	Glutathione metabolism
XP_020609846.1 gamma-glutamyltranspeptidase 1-like	Glutathione metabolism
XP_020623983.1 5-oxoprolinase-like	Glutathione metabolism
XP_020630097.1 glutathione S-transferase 1-like	Glutathione metabolism
XP_020601996.1 propionyl-CoA carboxylase beta chain, mitochondrial-like	Glyoxylate and dicarboxylate metabolism
XP_020621203.1 melanotransferrin-like	Iron binding
XP_020611142.1 LOW QUALITY PROTEIN: myoferlin-like	Membrane trafficking
XP_020615759.1 adenylyl cyclase-associated protein 2-like ;XP_020601827.1 adenylyl cyclase-associated protein 2-like	Membrane trafficking
XP_020602400.1 coatomer subunit delta-like	Membrane trafficking / ER
XP_020603952.1 ADP-ribosylation factor-like protein 8B-A	Membrane trafficking / GTP-binding
XP_020619369.1 tropomyosin-2-like	Motor proteins
XP_020626162.1 rho GDP-dissociation inhibitor 1-like	Neurotrophin signaling pathway
XP_020607893.1 guanylate-binding protein 6-like ;XP_020607892.1 guanylate-binding protein 6-like	Nod-like
XP_020607894.1 guanylate-binding protein 4-like	Nod-like
XP_020609885.1 14-3-3-like protein C	Nod-like
XP_020611664.1 14-3-3 protein zeta-like	Nod-like
XP_020619571.1 AP-1 complex subunit mu-1-like	Nod-like
XP_020622387.1 GTP-binding nuclear protein Ran	Nucleocytoplasmic transport
XP_020628861.1 cytochrome c oxidase subunit 5A, mitochondrial-like	Oxidative Phosphorylation
XP_020615854.1 band 4.1-like protein 3	P13K-AKT Signaling
XP_020616092.1 guanine nucleotide-binding protein G(q) subunit alpha-like	P13K-AKT Signaling
XP_020624905.1 transitional endoplasmic reticulum ATPase	Protein processing in endoplasmic reticulum
XP_020628074.1 protein disulfide isomerase-like 2-2	Protein processing in endoplasmic reticulum

XP_020619087.1 malate dehydrogenase, cytoplasmic-like	pyruvate metabolism
XP_020612721.1 myosin regulatory light chain 12A-like	Regulation of actin cytoskeleton
XP_020614368.1 dehydrogenase/reductase SDR family member 7-like	Retinol Metabolism
XP_020615918.1 putative all-trans-retinol 13,14-reductase	Retinol Metabolism
XP_020618614.1 UDP-glucuronosyltransferase 2B15-like	Retinol Metabolism
XP_020611262.1 hexosaminidase D-like	Various types of N-glycan biosynthesis and degradation
XP_020619026.1 uncharacterized protein LOC110056820 ;XP_020619025.1 uncharacterized protein LOC110056819 ;XP_020619024.1 uncharacterized protein LOC110056818 ;XP_020619023.1 uncharacterized pro	
XP_020619491.1 uncharacterized protein LOC110057246	

The protein changes involve endocytosis, protein processing, membrane trafficking and metabolic pathways, such as glutathione metabolism. These pathways are most consistent with cellular response to bacterial infection and could indicate a possible mechanism of bacteria being absorbed via endocytosis, disrupting protein processing and trafficking downstream ultimately leading to actin remodeling and eventually apoptosis. A review by Posor et al. (2022), discussed that some bacteria effector proteins interfere with the host cell's autophagy processes causing an inability to rid the host of bacteria. This could result in the upregulation of a host cell's production of autophagy related proteins to try to clear the invasive bacteria (Posor et al. 2022). Phosphoinositides play an important role in membrane trafficking via endocytosis and are likely being altered with disease. Based on the pathway changes observed, one hypothesis is that the bacteria are interfering with Rho GTPase signaling. Fiorentini et al. (2003) described bacterial toxins that could interfere with Rho GTPases via multiple pathways, such as directly modifying Rho proteins or indirectly acting on Rho proteins by signaling regulatory proteins.

These proteomic studies have provided a good direction for future SCTLD research. Next steps include comparing the proteomic data with other 'omic data to investigate connections that might help explain some of the changes seen in these datasets, such as transcriptomic data that could validate transcription factors associated with downstream effects of Rho changes. Given the changes in cytoskeletal proteins and lesions that are present, some changes in histology would be expected. Although traditional histological techniques are useful initially to look at overall structural changes, it would be appropriate to use different specialty stains to highlight and verify protein changes that are predicted by the mass spectrometry data.

In particular gross structural changes with Movat's pentachrome stain which differentiates between collagen, elastin, muscle, mucin and fibrin. This information could provide new insights on connective tissue restructuring due to disease and clues to what structural changes are occurring prior to lesion formation.

Our continuing project will incorporate glycomic analyses of the *O. faveolata* active/inactive disease samples to evaluate differential changes in glycan profiles of the disease versus healthy tissues. Many of the proteins that were changing with disease are also involved upstream of glycosylation pathways. Glycan changes are likely occurring downstream of the initial bacterial uptake in response to protein processing differences. Other changes could result from environmental factors that alter nutrient uptake which may influence disease progression in the different regions, accounting for the difference in other proteomic profiles. Understanding how glycans are changing in both tissue and mucus can lead to more specific treatments and/or diagnostics with our goal of ultimately developing a liquid biopsy for diagnostic purposes to reduce the need for the more invasive biopsies.

IV. References

Aeby GS, Ushijima B, Bartels E, Walter C, Kuehl J, Jones S, Paul VJ. (2021). Changing stony coral tissue loss disease dynamics through time in *Montastraea cavernosa*. *Front. Mar. Sci. 8*. https://doi.org/10.3389/fmars.2021.699075

Ben-Haim Y. and Rosenberg E. (2002). A novel *Vibrio* sp. pathogen of the coral *Pocillopora damicornis. Marine Biology* **141**(1): 47-55. https://doi.org/10.1007/s00227-002-0797-6.

Bourne DG, Garren M, Work TM, Rosenberg E, Smith GW, Harvell CD. (2009). Microbial disease and the coral holobiont. *Trends Microbiol.* **17**(12): 554-562. https://doi.org/10.1016/j.tim.2009.09.004.

Fiorentini C, Falzano L, Travaglione S, Fabbri A. (2003). Hijacking Rho GTPases by protein toxins and apoptosis: molecular strategies of pathogenic bacteria. *Cell Death Differ.* **10**(2): 147-152. https://doi.org/10.1038/sj.cdd.4401151.

Garcia, GD, Santos, Ede O, Sousa, GV, Zingali, RB, Thompson, CC, Thompson, FL. (2016). Metaproteomics reveals metabolic transitions between healthy and diseased stony coral, *Mussismilia braziliensis*. *Mol. Ecol.* **25**(18): 4632-4644. https//doi.org//10.1111/mec.13775.

Meyer JL, Castellanos-Gell J, Aeby GS, Häse CC, Ushijima B, Paul VJ. (2019). Microbial community shifts associated with the ongoing stony coral tissue loss disease outbreak on the Florida Reef Tract. *Front. Microbiol.* **10**: 2244. https://doi.org/10.3389/fmicb.2019.02244.

Meyer JL, Paul VJ, Raymundo LJ, Teplitski M. (2017). Comparative metagenomics of the polymicrobial black band disease of corals. *Front. Microbiol.* **8**: 618. https://www.frontiersin.org/articles/10.3389/fmicb.2017.00618

Munn CB. (2015). The role of Vibrios in diseases of corals. *Microbiol. Spectr.* **3**(4). https//doi.org/10.1128/microbiolspec.VE-0006-2014.

Posor Y, Jang W, Haucke V. (2022). Phosphoinositides as membrane organizers. *Nat. Rev. Mol. Cell Biol.* **23**(12): 797-816. https://doi.org/10.1038/s41580-022-00490-x

Precht WF, Gintert BE, Robbart ML, Fura R, van Woesik R. (2016). Unprecedented disease-related coral mortality in Southeastern Florida. *Sci. Rep. 6*(1). https://doi.org/10.1038/srep31374.

Protifi S-Trap micro protocol long 4.7. *Protifi Protocols* https://files.protifi.com/protocols/s-trap-micro-long-4-7.pdf accessed 1-26-24.

Rechenberger J, Samaras P, Jarzab A, Behr J, Frejno M, Djukovic A, Sanz J, González-Barberá EM, Salavert M, López-Hontangas JL, Xavier KB, Debrauwer L, Rolain JM, Sanz M, Garcia-Garcera M, Wilhelm M, Ubeda C, Kuster B. (2019). Challenges in clinical metaproteomics highlighted by the analysis of acute leukemia patients with gut colonization by multidrug-resistant Enterobacteriaceae. *Proteomes* **7**(1):2. https://doi.org//10.3390/proteomes7010002.

Sato Y, Ling, EYS, Turaev D, Laffy P, Weynberg KD, Rattei T, Willis BL, Bourne DG. (2017). Unraveling the microbial processes of black band disease in corals through integrated genomics. *Sci. Rep.* **7**: 40455. https://doi.org/10.1038/srep40455.

Taylor SL, Wesslingh S, Rogers GB. (2016). Host-microbiome interactions in acute and chronic respiratory infections. *Cell. Microbiol.* **18**(5): 652-662. https://doi.org.//10.1111/cmi.12589.

Thurber RV, Payet JP, Thurber AR, Correa AMS. (2017). Virus-host interactions and their roles in coral reef health and disease. *Nat. Rev. Microbiol.* **15**(4):205-216. https://doi.org/10.1038/nrmicro.2016.176.

Verschaffelt P, Van Den Bossche T, Martens L, Dawyndt P, Mesuere B. (2021). Unipept desktop: A faster, more powerful metaproteomics results analysis tool. *J. Proteome Res.* **20**(4): 2005-2009. https://doi.org//10.1021/acs.jproteome.0c00855.

Wear SL, Thurber RV. (2015). Sewage pollution: mitigation is key for coral reef stewardship. *Ann. N. Y. Acad. Sci.***1355**:15-30. https://doi.org//10.1111/nyas.12785.