Metaproteomic Analysis of Stony Coral Tissue Loss Disease

Final report prepared by:
Michael G. Janech, Ph.D.[1], Benjamin Neely, Ph.D.[2], Frank Mari, Ph.D.[2], Elizabeth Duselis, Ph.D.[3], Cheryl Woodley, Ph.D.[3]

[1]Department of Biology, College of Charleston, Charleston, SC 29424
[2]Marine Biochemical Sciences Group (CSD/MML), National Institute of Standards and Technology, U.S. Department of Commerce, NIST Charleston, Charleston, SC 29412
[3]NOAA NOS NCCOS, Hollings Marine Laboratory, Charleston, SC 29412

June 30, 2021

Completed in Partial Fulfillment of PO B67079

for Florida Department of Environmental Protection Coral Reef Conservation Program 1277 N.E. 79th Street Causeway Miami, FL 33138

Original Project Dates:
December 1, 2019-June 24, 2020

Revised Project Dates:
December 1, 2019-June 30, 2021

Table of Contents

1. Background

Florida's Coral Reef is currently experiencing a multi-year disease-related mortality event, that has resulted in massive die-offs in multiple coral species. Approximately 21 species of coral, including both Endangered Species Act-listed and the primary reef-building species, have displayed tissue loss lesions which often result in whole colony mortality. First observed near Virginia Key in late 2014, Stony Coral Tissue Loss Disease (SCTLD) has since spread to the northernmost extent of the Florida Reef Tract, and south to the eastern side of the Dry Tortugas. The best available information indicates that the disease outbreak is continuing to spread southwest and throughout the Caribbean.

Determining the causative agent(s) of coral disease relies on a multidisciplinary approach since the causation may be a combination of abiotic, microbial or viral agents [1-4]. Molecular approaches such 16s rDNA microbiome analysis have been used in conjunction with field sampling and laboratory experiments to identify and confirm causes, e.g. *Vibrio coralliilyticus* [5], satisfying the basic tenets of Koch's postulates in a few cases of other coral diseases. Based on two published studies that have more comprehensively investigated other stony coral diseases, dominant changes to the microbiome appear to create a signature during the onset and duration of disease [6, 7] that results in sulfate-reducing communities around the site of disease. It is known in human disease that molecular changes both in the host and microbiome occur well before observable phenotype and gross pathology is observed[8]. For this reason, defining the changes in the molecular landscape in the coral holobiont can provide useful information not only in diagnosis, but for prediction and prognosis [9]. Specifically, in the case of SCTLD, defining molecular changes in the coral holobiont will help define disease progression and aid in identifying the causative agent by clearly defining traits of disease progression shared across affected species. We focused on the functional response of the coral microbiome because this search space can be defined using metagenomics and provide answers in the period of the statement of work. The analysis will be specific to the microbial composition of each sample, as opposed to selecting publicly available databases that may or may not be relevant. In the future, these species databases will be searched during metaproteomic analysis.

In an effort to define the molecular changes elicited by SCTLD in the coral microbiome, tools measuring presence of genes, expressed transcripts and abundance of proteins (i.e., genetics, transcriptomics, and proteomics) are useful. For this proposed research, we will be focused on the abundance of proteins to capture the functional aspects of the disease phenotype as well as microbial species diversity. Functional characterization includes determining the expression of virulence factors associated with the microbiome that may underly infection and spread of disease. Secondly, the functional picture provided by measuring protein abundance can be utilized to classify the disease phenotype to predict the acute phase of the disease process that is not yet histologically visible.

Meta-omic molecular tools have been sparsely applied to the study of coral disease[6, 7, 10, 11]. Only a single recent publication has applied metagenomic techniques to corals affected by SCTLD[10], which 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[8], emphasizing the value of studying functional changes in the microbiome in response to unknown viral/microbial/abiotic stresses. Of the one study to address coral disease

using metaproteomics, only 361 proteins were assessed [6], which is far below expected using current technology, thereby drastically limiting functional interpretation of the data. The proposed study aims to more 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 hosts.

1.1 Project Goals and Objectives:

1) Create metagenomic assemblies for each set of healthy and diseased corals.

1a) Isolate DNA from holobiont specimens
1b) Assemble sequence reads and create a Fasta file for proteomic searches.

2) Acquire proteome data for up to 9 coral holobionts from uninfected and actively infected diseased corals.

2a) Isolate and digest proteins from for each set of healthy and diseased corals.

2b) Analyze mass spectrometry data and identify holobiont proteins.

3) Associate functional changes in the microbiome in diseased corals and compare across hosts.

4) Synthesize data into information needed by managers and DAC to communicate the project's findings and possible recommendation for further actions.

2. Methodology

2.1 Samples

Nine coral species were selected to represent highly susceptible species with fast onset and slower onset disease dynamics with five replicates of each condition (i.e., healthy, SCTLD). Specimens were collected in July and September 2019 in collaboration with the Florida Keys National Marine Sanctuary under permit # FKNMS-2019-069 to Dr. Cheryl Woodley of NOAA and FKNMS-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 specimens from the disease-free zone to be collected in a yet unaffected area of the Marquesas. The permitted species include:

1. *Colpophyllia natans*
2. *Dichocoenia stokesii*
3. *Diploria labyrinthiformis*
4. *Meandrina meandrites*
5. *Montastraea cavernosa*
6. *Orbicella annularis*
7. *Orbicella faveolata*
8. *Pseudodiploria strigosa*
9. *Pseudodiploria clivosa*

2.2 DNA and Protein Extraction

Samples were extracted from diseased and non-diseased coral polyps that were harvested in fragments that included whole skeleton from the colony, immediately frozen in liquid nitrogen, and was stored at Hollings Marine Laboratory at -80°C. For diseased samples, fragments including skeletal material with polyps that included areas of active disease.

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 to pellet loose tissue and 1mL was removed and stored at -80°C for DNA extraction. The remaining sample with pellet was tip sonicated to further lyse refractory cells. Cell debris was removed by centrifugation at 1,500 RCF. Lysis buffer (1mL) containing DNA was shipped to Novogene for DNA extraction

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.

2.3 Trypsin Digestion

Protein from each coral sample was digested with trypsin following reduction (DTT) and alkylation (CAA) using micro S-traps (Protofi). Peptides were eluted and assayed using a colorimetric peptide assay kit (Pierce). Peptide samples are currently stored at -80°C at the Grice Marine Laboratory awaiting mass spectrometry analysis.

2.4 Metagenomic libraries

Forty-seven coral samples representing nine species were submitted for metagnome library construction. Only 17 of 46 samples passed QC and we utilized for library preparation and downstream analysis.

Agarose gel electrophoresis was used to determine DNA purity and integrity, and a Qubit 2.0 fluorometer quantitation was used for accurate measurement of DNA concentration. Physical fractionation was performed by a Covaris Sonicator. Fractionation steps were checked by an Agilent2100 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 1μg 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 quality control of each sample and put the unutilized reads of each sample 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-redundant 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 databases (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 Normal samples combined irrespective of species. When grouping information was available, Metastats and LEfSe multivariate statistical analysis and comparative analysis of metabolic pathways was carried out to explore species composition and functional composition differences between groups.

2.5 Liquid Chromatography/Mass spectrometry

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 column (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 (75µm 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 to 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 positive polarity with 30 % RF 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 SwissProt and TrEMBL). Searches also included the common Repository of Adventitious Proteins database (cRAP; 2012.01.01; the Global Proteome Machine; 107 sequences). These fasta 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 FDR. The resulting .dat files were loaded into Scaffold to enable users to explore

the data and export count based relative quantification (weighted spectral counts). These result files (.sf3) are also included (a viewer can be downloaded from Proteome Software) and protein inference settings may be changed.

2.6 Proteomic Statistics and Clustering

Weighted spectral counts were compared across SCTLD+ vs 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. Hierarchical clustering of taxa at the class level as well as Gene Ontology classification for the top 15 gene categories were conducted for comparison.

3. Results

3.1 Protein sample preparation quantification

Protein extraction using 5% SDS resulted in a range of concentrations between 0.53-7.32 µg/µL as estimated by BCA assay (Table 1). Protein integrity was assessed in a subset of samples for each species by polyacrylamide gel electrophoresis (PAGE) followed by Coomasie blue staining (Figure 1). Every sample examined shows a wide distribution of protein molecular weight. Tryptic peptide concentrations range between 0.13 – 0.87 µg/µL following preparation with the S-Trap protocol. These concentrations are within historical ranges for the lab when starting with 100µg total protein. A minimum of 0.1µg/µL is required for mass spectrometry peptide analysis, of which all samples exceed the minimum concentration.

3.2 Metagenomics

Of 46 samples sent to Novogene, only 17 passed QC which required a minimal amount of DNA of 1µg. 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 (i.e., continuous sequences within scaffolds). Clean data of all samples were mapped to assembled Scaftigs and unutilized reads were collected. Sequencing statistics are displayed in Table 2 (section 5.2) for all 17 samples. Summary statistics for gene assembly, gene prediction, taxonomic annotation, functional annotation, and antibiotic resistant genes are listed in Tables 3A and 3B (Section 5.3). Following assembly and removal of redundancy by CD-HIT, 4,285,204 open-reading frames were identified and compiled into a protein FASTA database for metaproteomic search. Of those, 2,146,221(50.08%) contain start and stop codons; whereas, 90,448(2.11%) contain neither a start nor stop codon. Open-reading frame length was relatively small for a majority of compiled sequences (Figure 2).

Figure 3, in a comparison of unique gene and common genes between corals with SCTLD *versus* unaffected normal coral samples, 1,208,206 genes were common between both groups. Normal 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 in in Figure 3. There was no statistical difference in taxonomic distribution between SCTLD and Normal samples, although it should be noted that the metagenomic study was not

designed to investigate differences between SCTLD and Normal. 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 taxa, principle components analysis and analysis of similarity show considerable overlap between groups suggesting a strong overlap in species composition between Normal and SCTLD samples (Figure 4).

Functional differences between metagenome samples were not statistically different. Again, the rationale underlying metagenomics was to develop a widespread database for protein searching and not to determine differences between diseased and non-diseased corals. A list of top-level functions representing the study population are listed in Figure 5. Although the majority of genes are assigned the function of uncategorized, there remains a large distribution across 25 high-level functional categories. For reference, a list of viral DNA sequences matched to NCBI nr database were collated and provided in Table 4 which is often excluded in 16s rRNA studies.

3.3 Metaproteomics

8552 proteins were identified across 46 samples included in the study (1% FDR). 18 proteins were statistically higher and 5 proteins lower in the SCTLD+ group compared to the SCTLD- group (Figure 6, $p<0.00016$, Fisher's Exact Test). 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 only contained an abbreviated number of samples that did not allow a more complete comparison like the proteomic analysis. In short, there is no pan-taxa protein classifier of SCTLD according to this analysis.

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[12]. 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 (Figure 79). 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 (Figure 10). 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: (Figure 12) SCTLD+, carbohydrate metabolic process, intracellular protein transport, vesicle-mediated transport, and cell adhesion; (Figure 13) SCTLD-, carbohydrate metabolic process, intracellular protein transport, vesicle-mediated transport, and bioluminescence. Cell adhesion was more represented in SCTLD+ corals and Bioluminescence was more represented 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 (Figure 13). 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.

Fine comparisons of gene ontology terms are displayed in Figure 14 for both SCTLD+ and SCTLD- corals for the top 15 categories based on peptide membership. 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 (Figure 15). 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 (Figure 15 red box) continued to show difference in peptide membership which was also indicated in data displayed in Figures 11 and 12. 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 (Figure 7 and 8), we resubmitted .raw files to search spectra against taxon-specific translated metagenome files, except for *Montastrea 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, Montastrea 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 (Figure 16). 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 protein (Figure 17).

4. Conclusions

Protein extraction from corals was successful for trypsin digestion using 5% SDS lysis buffer, combined with manual and sonic disruption. Protein yields were sufficient for tryptic digestion using the S-trap methodology and downstream proteomic analysis for all 46 coral specimens.

Metagenomic sequencing was attempted to generate custom FASTA libraries for all 46 specimens. Only 17 specimens were ultimately included due to DNA yield. Following sequencing, assembly and characterization, a fasta file was generated containing 4,285,204 open-reading frames which will be utilized to conduct a proteomic search. Comparisons between SCTLD and Normal corals are not recommended using metagenomic data due to species distribution bias and uneven case/control pairing. However, the metagenome data did yield some interesting ancillary information that could be utilized in future projects. 1) Viral DNA results

were collated into a table for all samples included in the metagenomic analysis. At the strain level, 121 different viruses associated with the samples were identified. About 10% of the relative abundance was due to uncultured Mediterranean phage. 2) About 70 Antibiotic resistance genes (ARGs) were identified from the dataset using The Comprehensive Antibiotic Resistance Database. Given recent research efforts focused on antibiotic treatment[13] of SCTLD affected corals, the potential for antibiotic resistance should be considered provided the presence of ARGs in the sequence data. Prospective monitoring of ARGs could also provide some guidance on antibiotic selection. 3) Raw sequence data are made available through NCBI BioProject ID: PRJNA726962 Submission ID: SUB9574423 to support identification of potential causative biological agents.

Metaproteomic comparisons across all 46 SCTLD+ and SCTLD- coral tissue samples were made using proteins predicted from the metagenome as well as inclusion of extant protein data from common repositories. Metaproteomics did not result in the identification of a perfect classifier for SCTLD+ and SCTLD- corals based on individual protein comparisons, functional annotation, nor taxonomic comparison. Taxonomic similarities based on NCBI classification by Unipept of the holobiont corresponded with recent conclusions based on 16s DNA studies [14] where similarities greatly outweighed any differences. Notably, when gene ontology classification was applied to both groups, the category “bioluminescence” was highlighted as a pathway associated with SCTLD- corals and not SCTLD+ corals.

Of the proteins that were significantly elevated or reduced in abundance between SCTLD+ and SCTLD- corals, differences were largely due to the ability of the search engine to match peptides to specific species which were then compared versus grouping homologous proteins across species for statistical comparisons. This led to a more focused approach to data analysis, where individual species were compared to each other (SCTLD+ vs SCTLD-) and commonalities were assessed between statistically different protein groups. Sub-group analysis of individual taxa for six different species identified a general reduction in green-fluorescent protein-like proteins for SCTLD+ samples in 4 of 6 species examined. These data were corroborated Sub-group analysis suggested that intraspecific comparisons may be more appropriate due to differences in interspecific responses to SCTLD.

The reduction in green fluorescent protein-like proteins was the most consistent response found across all species examined. Green fluorescent proteins represent chromoproteins that may or may not be fluorescent and have been reported to provide a photoprotective function[15, 16]. The detection and distribution of fluorescent proteins in Hawaiian coral, *Montipora capitata*, was able to discriminate between tissues with active tissue loss disease induced by inoculation with *Vibrio coralliilyticus*[17]. Elevations in both mean edge area and mean edge to area ratio based on laser scanning confocal microscopy were able to classify pathogen treated corals. Given the noted disease manifestations affecting Zooxanthellae in SCTLD+ corals[18] and the suspected photoprotective role of fluorescent proteins, the reduction in green-fluorescent protein-like proteins in SCTLD+ corals could be an indicator and mechanism of SCTLD+ progression.

5. Figures

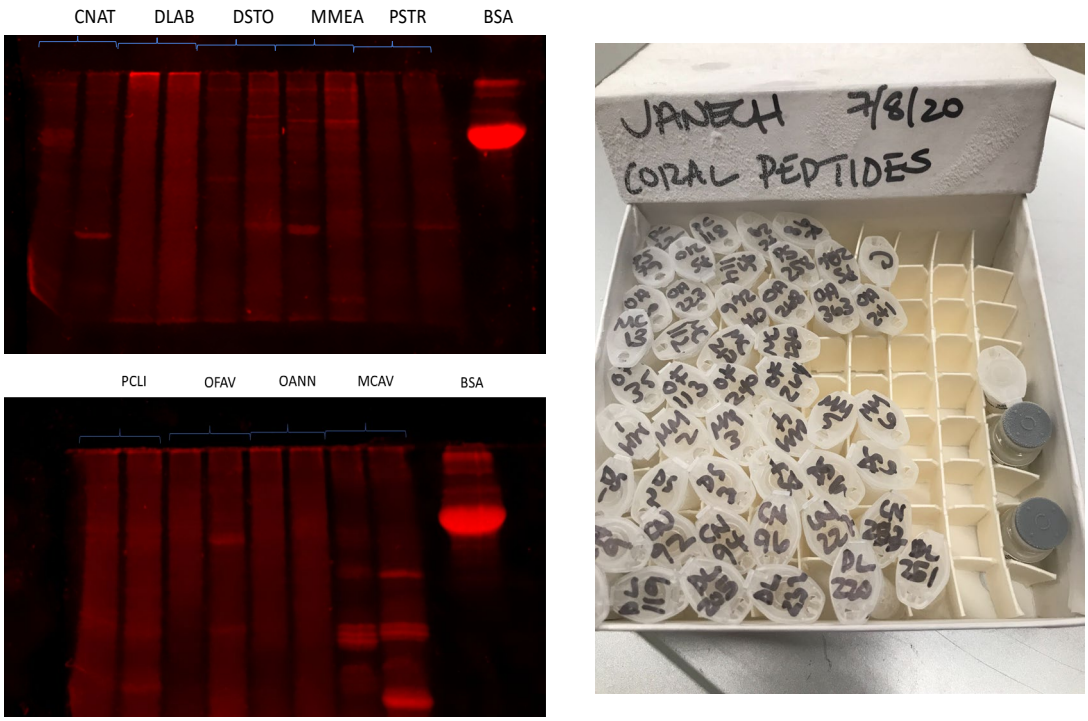


Figure 1. (Left) Protein PAGE of a subset of proteins for different coral species to assess quality. Lanes are labelled using abbreviations listed in Table 1. 20µg protein was loaded per lane. BSA = bovine serum albumin. (Right) protein digests in -80 freezer at Grice Marine Laboratory.

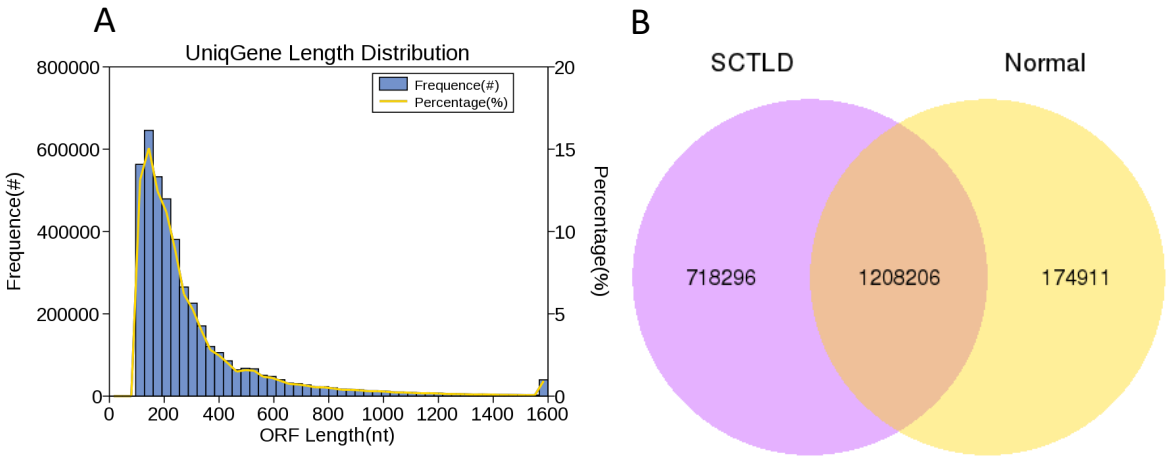


Figure 2. A) Open-reading frame length frequency for all 17 corals combined. Most open reading frames were less than 250 base pairs. B) Venn diagram of common and unique genes included in the analysis for both normal and SCTLD corals.

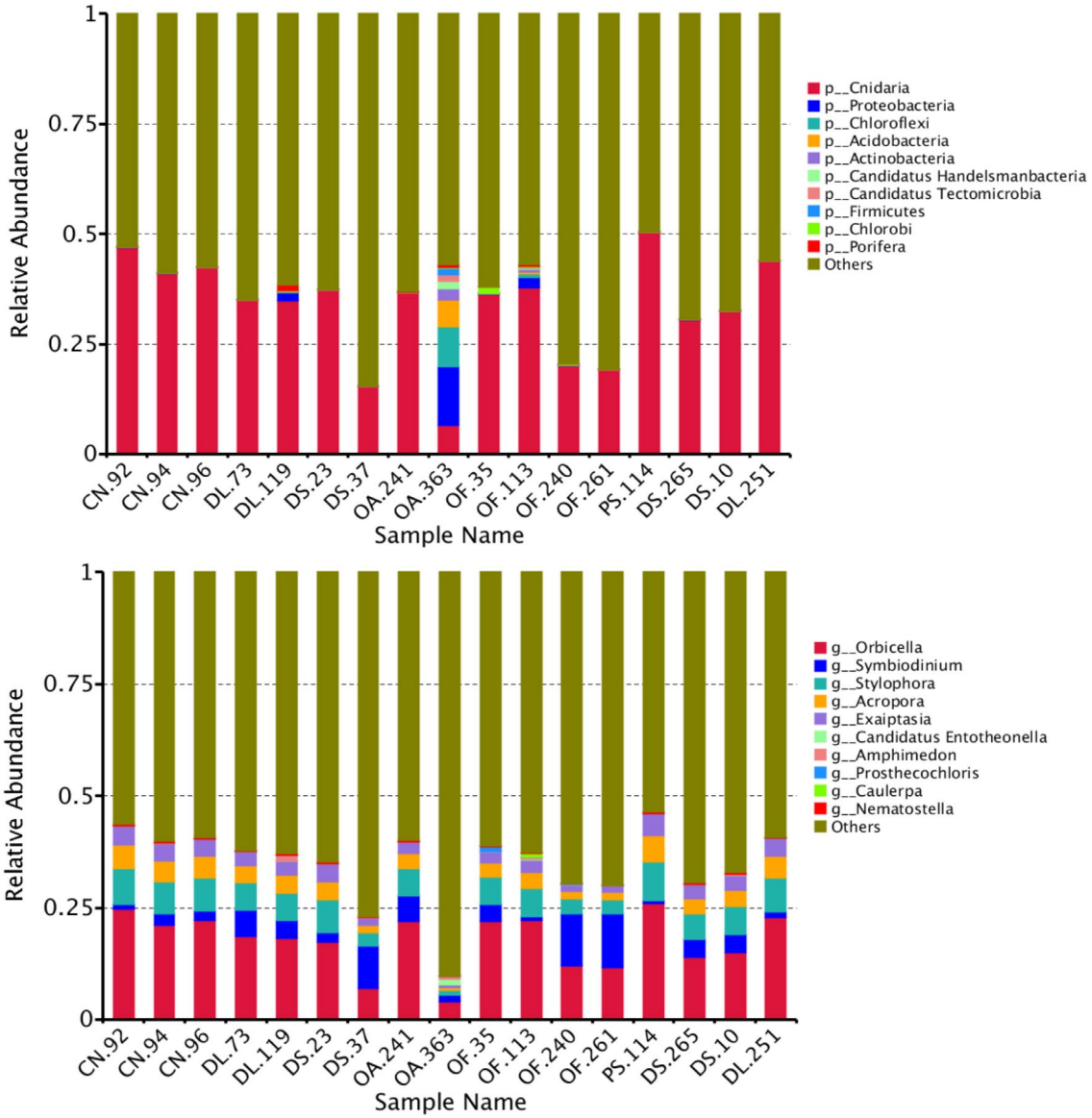


Figure 3. Taxonomic distribution across taxa. Based on the abundance of each taxonomic level, the top 10 taxa were selected and the other taxa were set as "Others". Bar charts show the relative taxonomy abundance of each sample in different taxonomic level. (Top) Taxonomic distribution by phyla. (Bottom) Taxonomic distribution by genera.

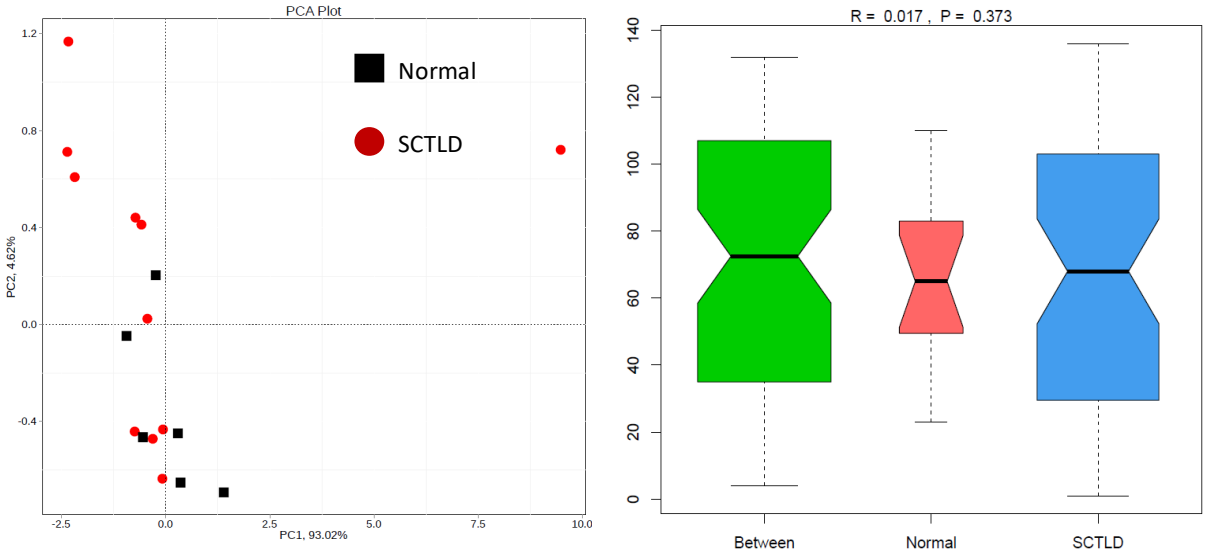


Figure 4. (Left) Principle component analysis of taxonomy at the phylum level. The percentage stands for the contribution of the principle components to the variation in samples. Each point in the graph stands for a sample. Samples belong to the same group that are in the same color. (Right) Analysis of similarities (ANOSIM) displays an R value of -0.05 indicative of inner-group variation is greater than inter-group.

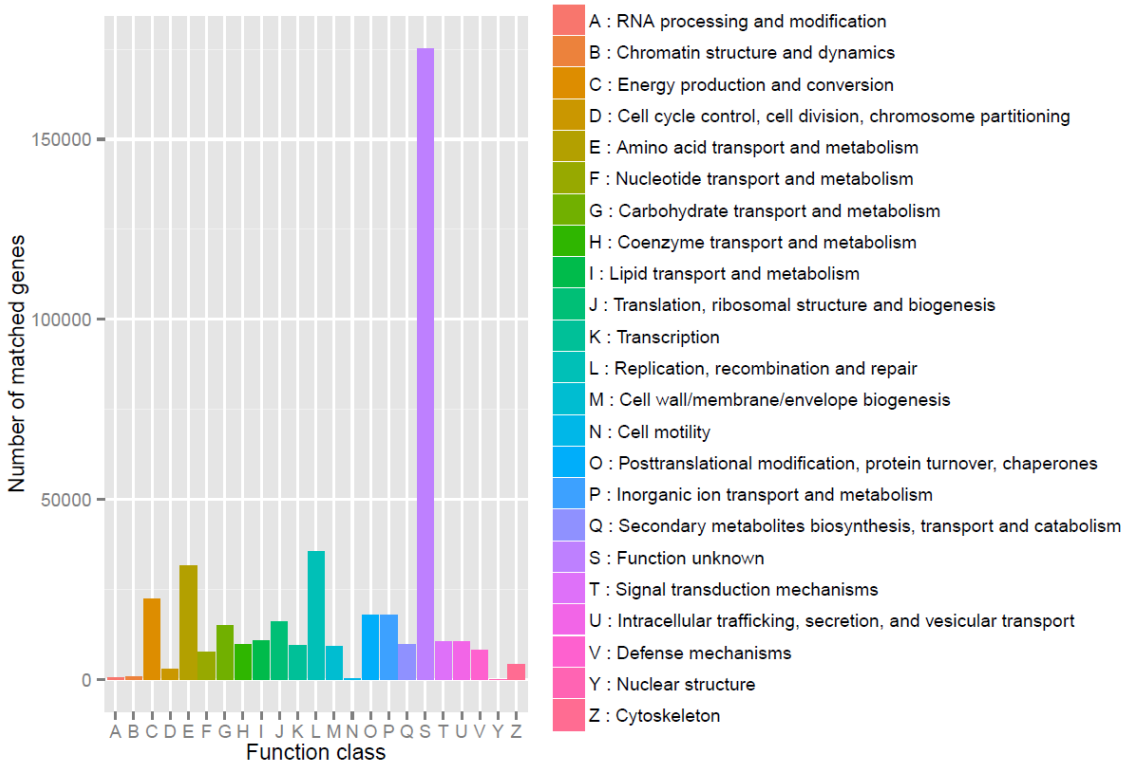


Figure 5. Functional grouping of genes across all sequenced taxa using EggNOG mapper (A database of orthology relationships, functional annotation, and gene evolutionary histories).

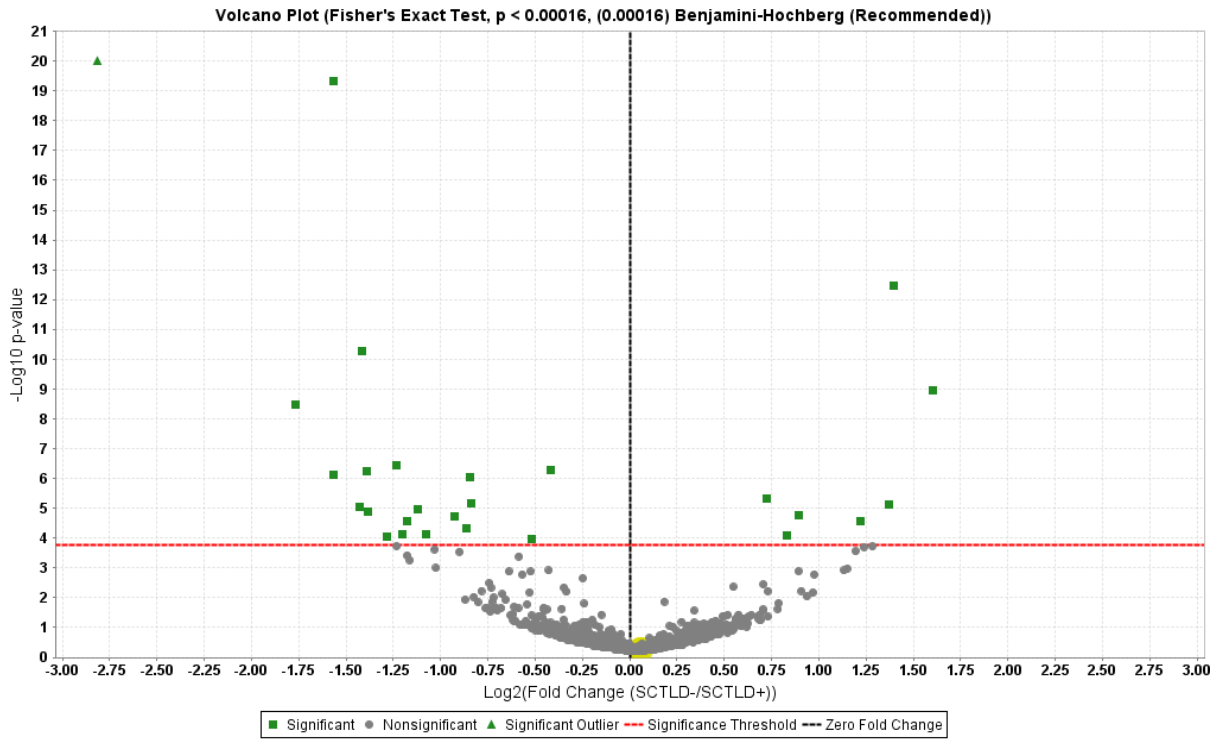


Figure 6. Volcano plot [log10 p-value vs fold-change] of proteins identified in this study. Proteins above the red line indicate significance. Dotted vertical line at 0 indicates unity.

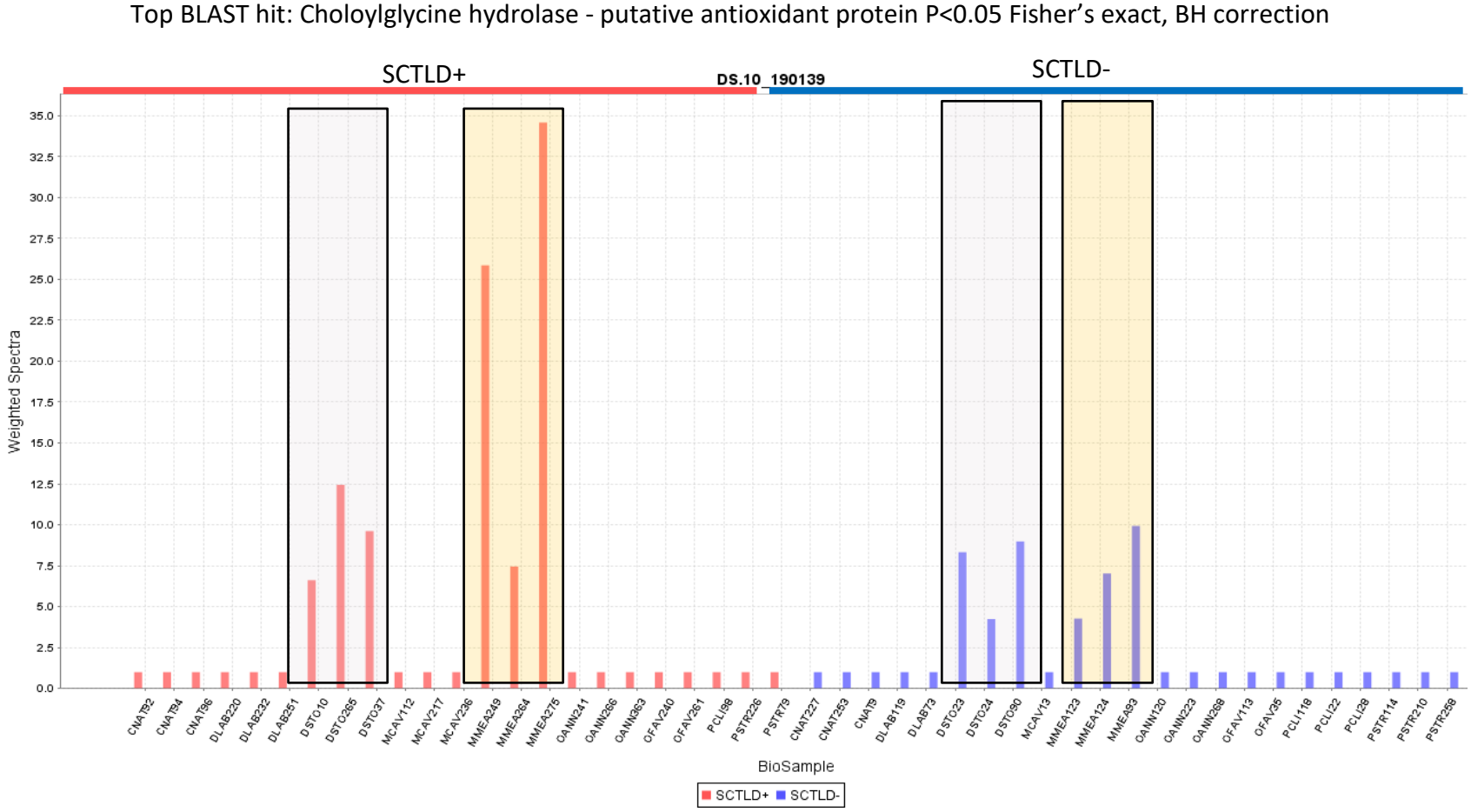


Figure 7. Weighted spectral counts for Choloylglycine hydrolase across coral taxa in SCTLD+ (red bars) or SCTLD- (blue bars) groups. Individual coral samples are indicated on the X-axis (e.g CNA 792, *Colpophyllia natans*). Choloylglycine hydrolase was elevated in *Dichocoenia stokesii* (light purple shaded boxes) and *Montastrea meandrites* (yellow shaded boxes) Protein differences were often driven by protein differences between distinct taxonomic groups rather than protein abundance across all taxonomic groups.

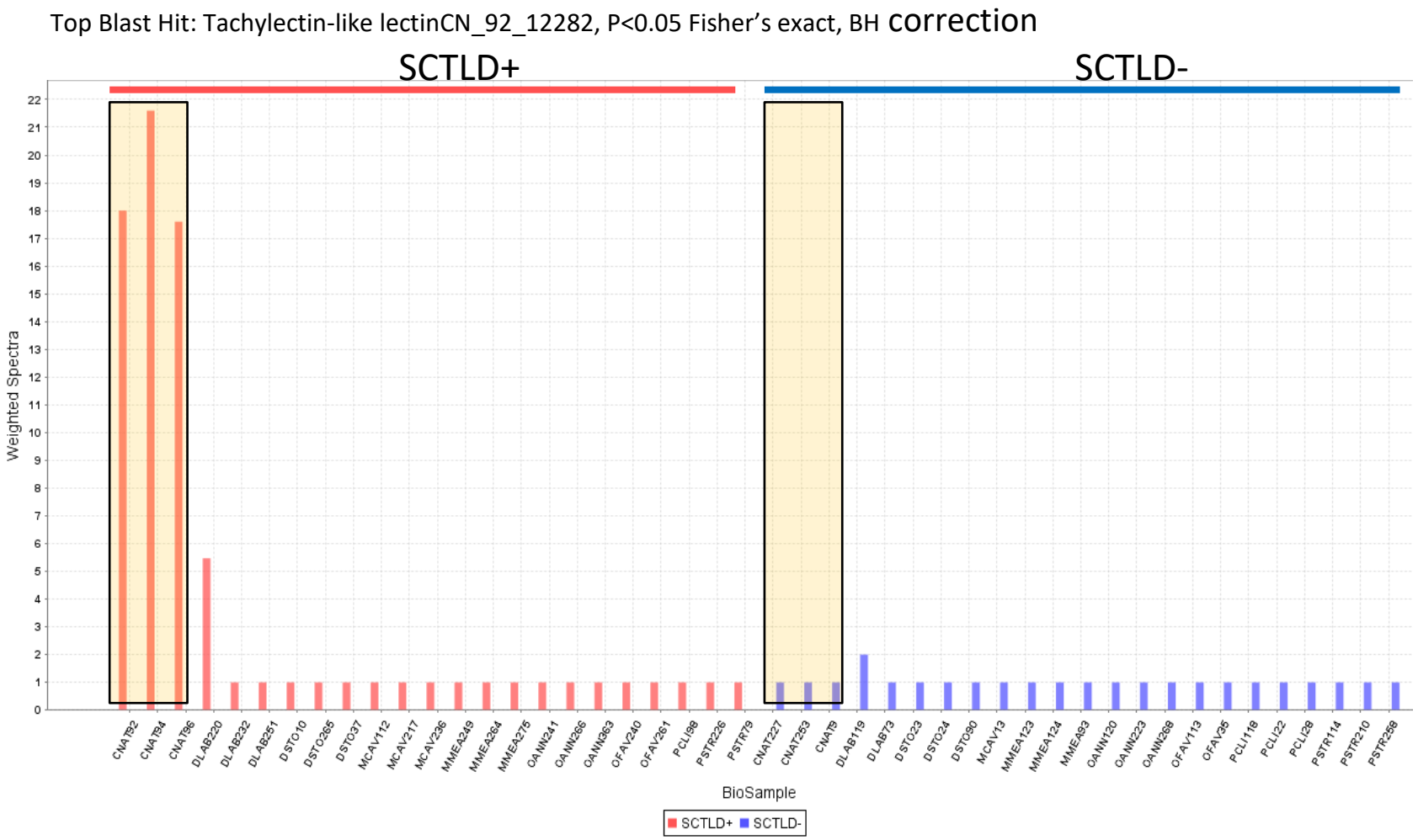
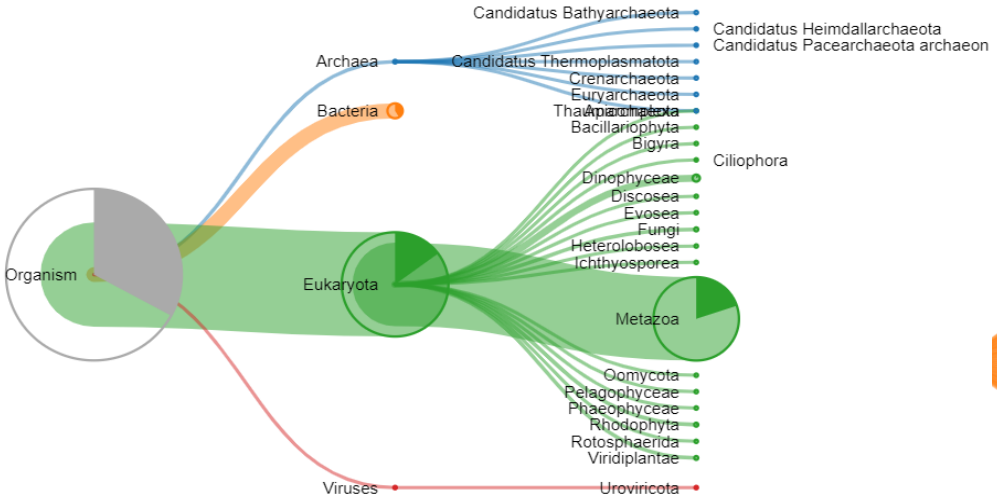


Figure 8. Weighted spectral counts for Tachylectin-like lectin across coral taxa in SCTLD+ (red bars) or SCTLD- (blue bars) groups. Individual coral samples are indicated on the X-axis (e.g CNA 792, *Colpophyllia natans*). Tachylectin-like- lectin was elevated in SCTLD+ samples from *Colpophyllia natans* (yellow shaded boxes).

SCTLD positive

Bacteria retracted



Bacteria expanded

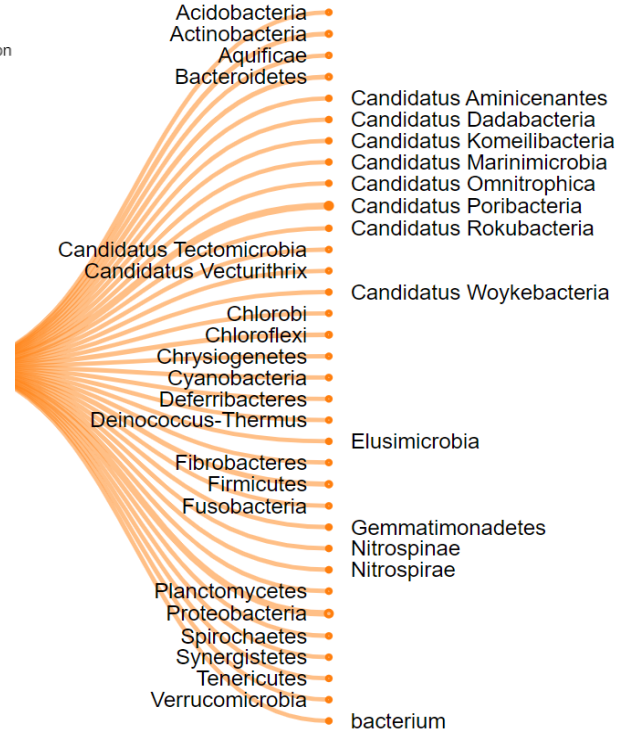


Figure 9. Treeview of taxonomic distribution from all SCTLD+ samples. Left, the domain bacteria was collapsed to permit visualization of the metazoan and archaeal phyla identified by unique peptide sequences. Right, the domain bacteria was expanded for inspection of bacterial phyla.

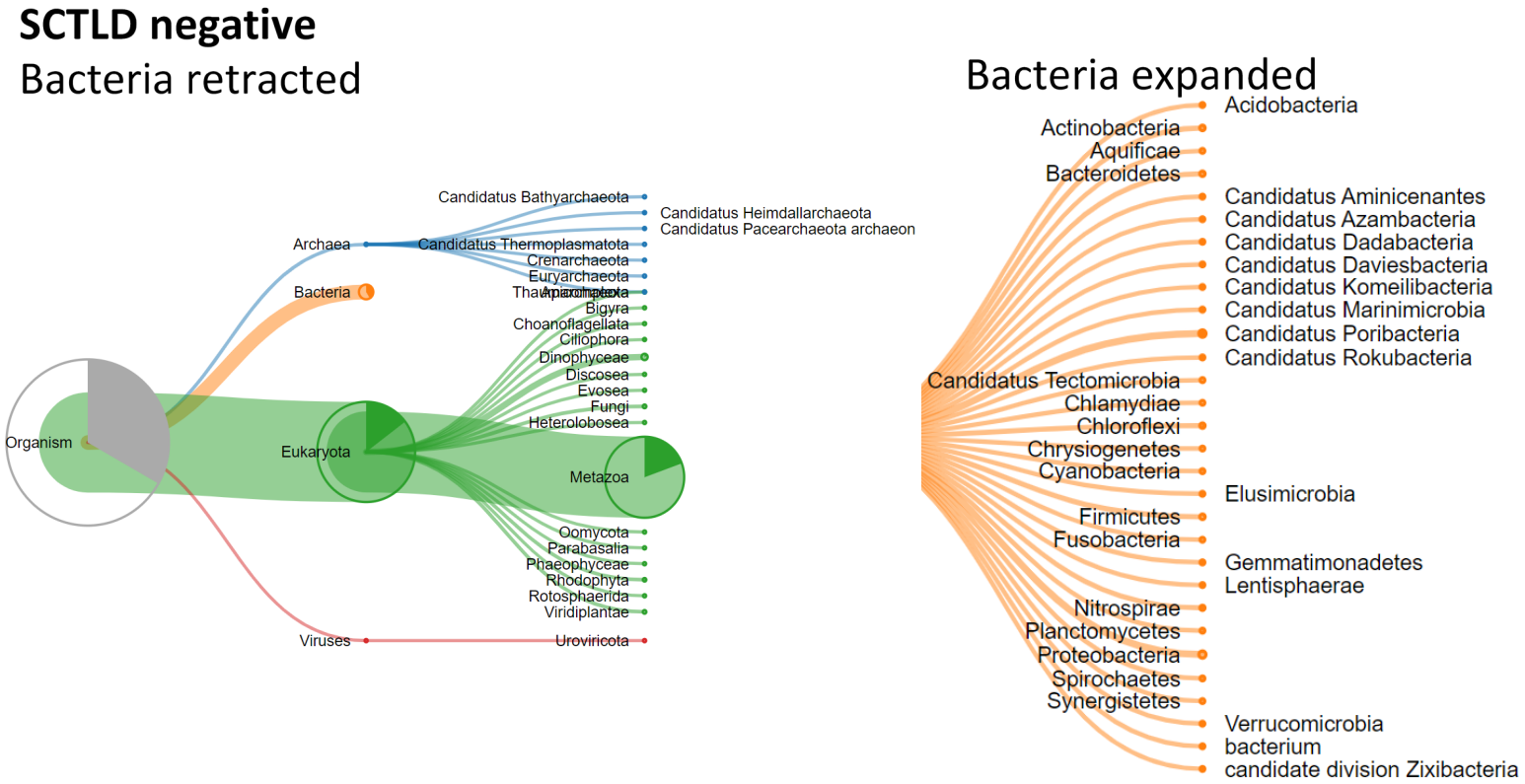


Figure 10. Treeview of taxonomic distribution from all SCTLD- samples. Left, the domain bacteria was collapsed to permit visualization of the metazoan and archaeal phyla identified by unique peptide sequences. Right, the domain bacteria was expanded for inspection of bacterial phyla.

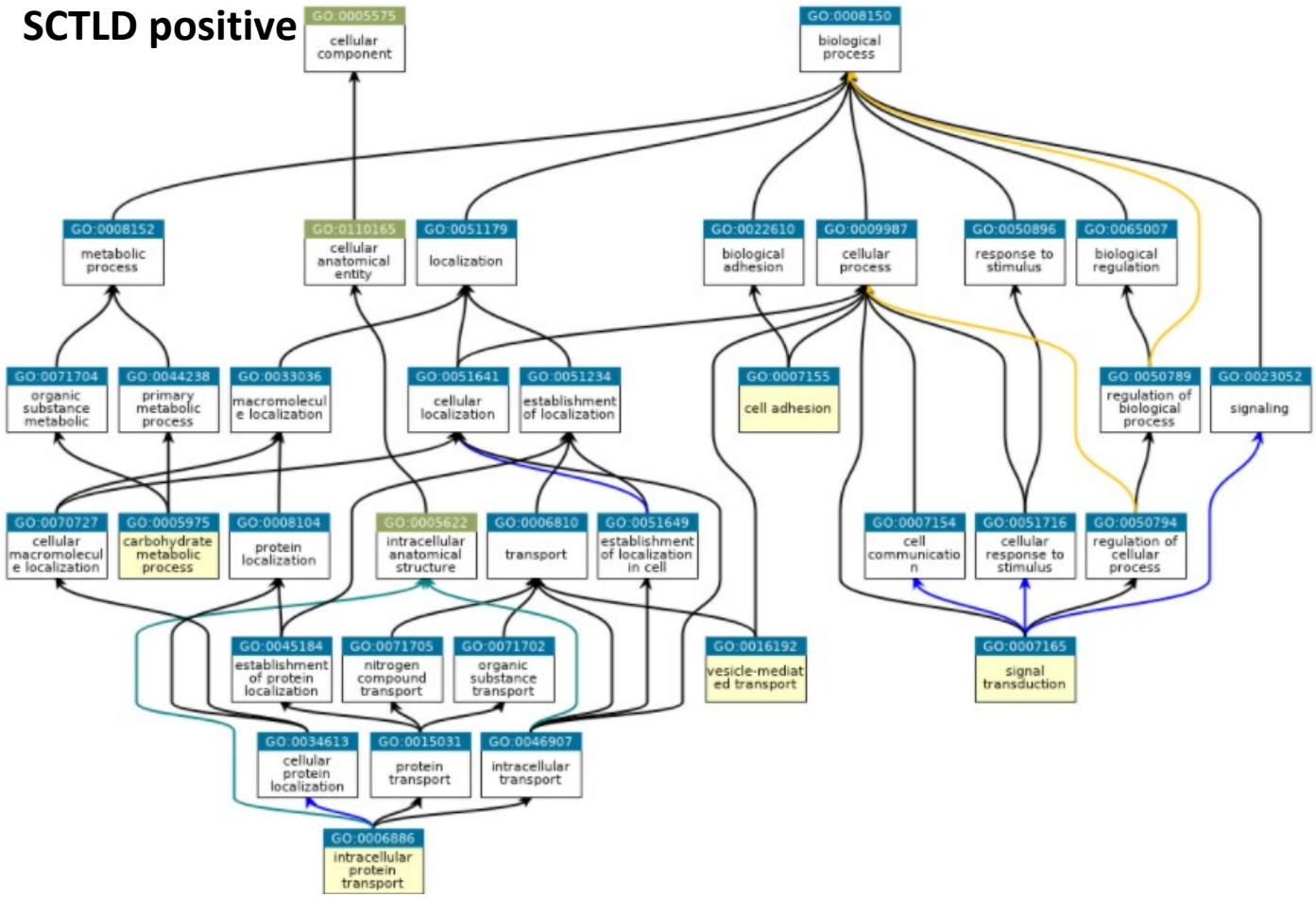


Figure 11. Gene Ontology map of SCTLD+ proteins. Over-representation is indicated by highlighted yellow boxes.

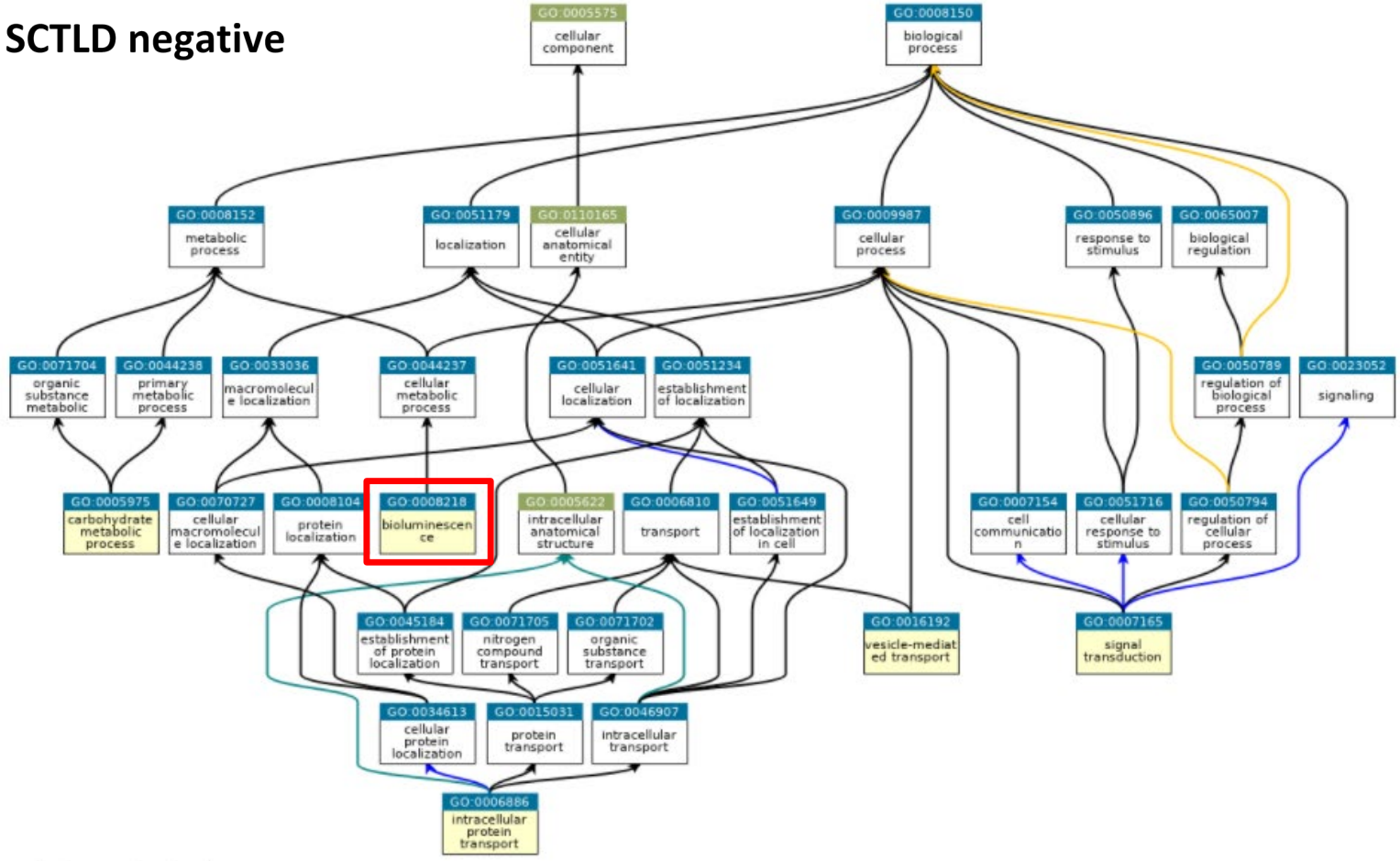


Figure 12. Gene Ontology map of SCTLD- proteins. Over-representation is indicated by highlighted yellow boxes. Bioluminescence is indicated by a red box.

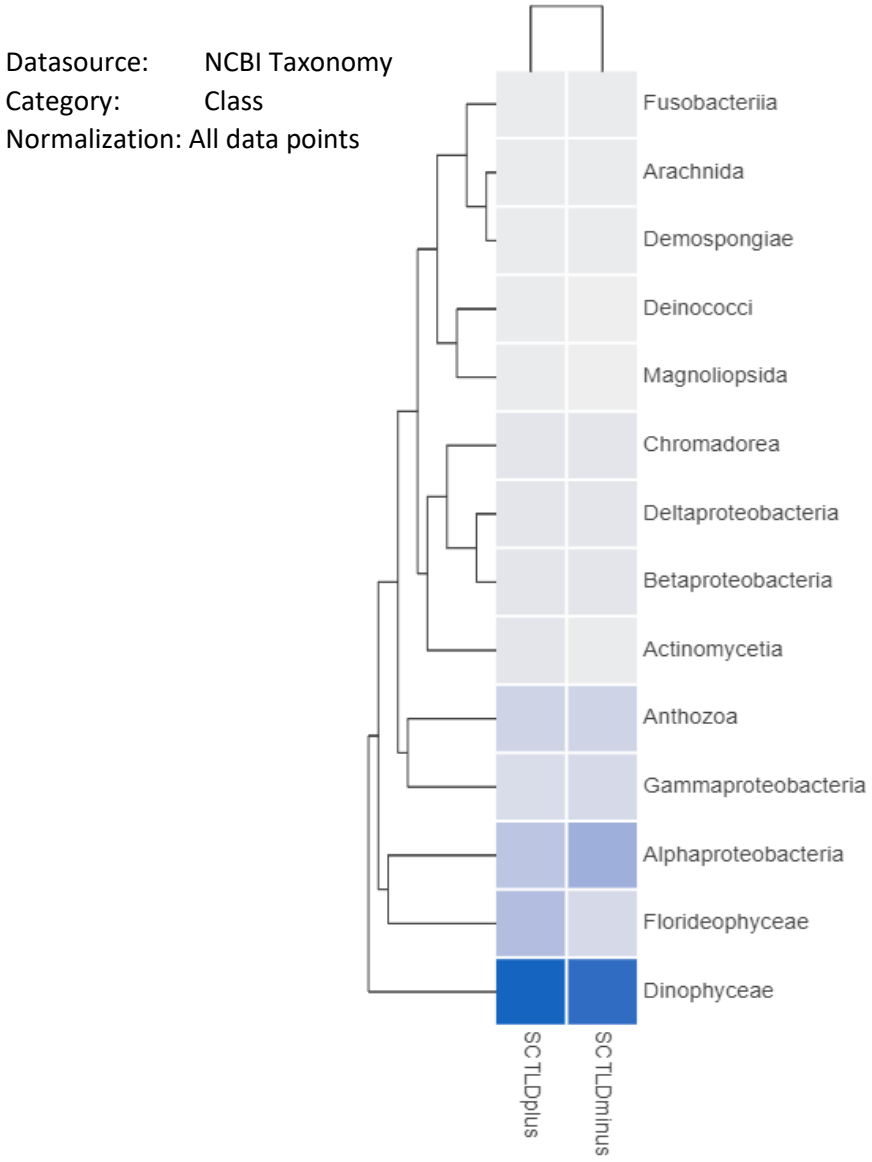
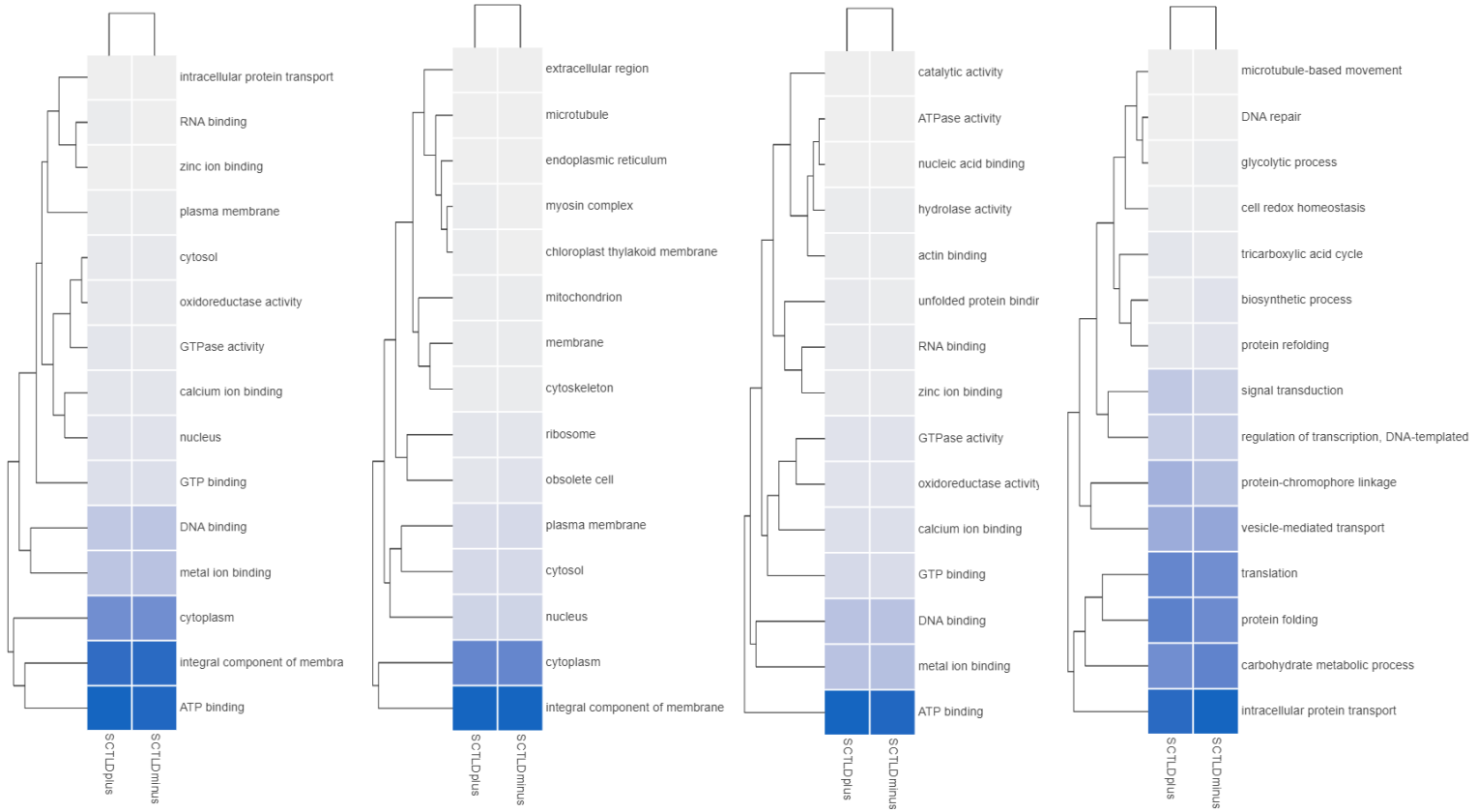


Figure 13. Heat map of SCTLD+ vs SCTLD- peptides classified by taxonomy in Unipept. Taxonomy was limited to Class-level. Normalization was set to All Points. Intensity of blue color indicates number of peptides associated with a given category for a particular group. Dinoflagellates contained the most peptides based on taxonomic classification. SCTLD+ contained slightly more peptides belonging to red algae; whereas, SCTLD- contained slightly more peptides belonging to Alphaproteobacter.



Datasource: Gene Ontology
Category: All
Normalization: All data points
Top 15 categories

Gene Ontology
Cell component
All data points
Top 15 categories

Gene Ontology
Molecular Function
All data points
Top 15 categories

Gene Ontology
Biological Process
All data points
Top 15 categories

Figure 14. Heat map of SCTLD+ vs SCTLD- peptides classified by Gene Ontology in Unipept. Displayed are results for all categories, cell component, molecular function, and biological process. The heatmap is limited to the top 15 categories for display. Normalization was set to All Points. Intensity of blue color indicates number of peptides associated with a given category for a particular group. Categorical classification demonstrated good similarity between both coral groups for the most abundant peptide groups. No striking differences were observed between SCTLD+ and SCTLD-.

Datasource: Gene Ontology
Category: Biological Process
Normalization: All data points
Top 16-30 categories

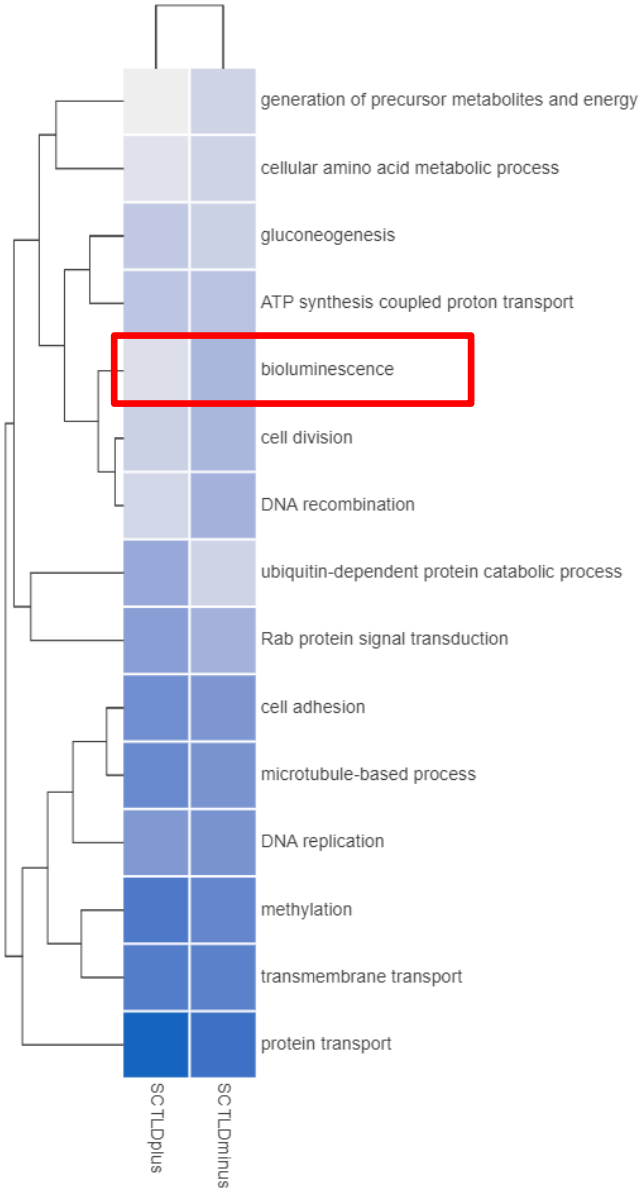


Figure 15. Heat map of SCTLD+ vs SCTLD- peptides classified by Gene Ontology in Unipept. Displayed are results for the category biological process. The heatmap displays the top 16-30 categories. Normalization was set to All Points. Intensity of blue color indicates number of peptides associated with a given category for a particular group. Biological category classification demonstrated discrepancy between several categories based on peptides identified e.g. ubiquitin-dependent protein catabolic process, dna recombination, cell division etc. Red box highlights bioluminescence, which contains fewer peptides associated with the SCTLD+ group suggesting a decrease in proteins associated with this term.

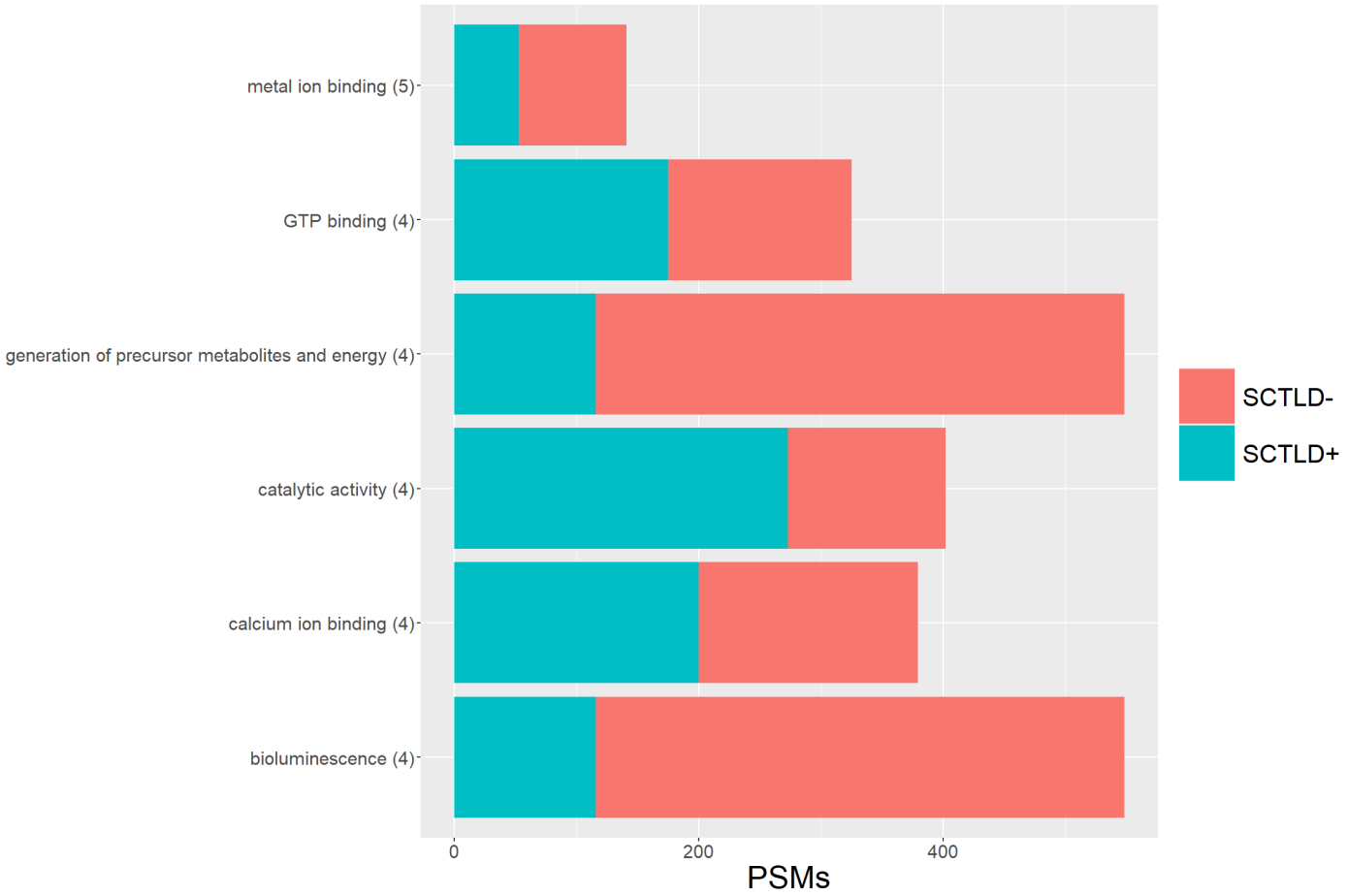


Figure 16. Top Gene Ontology categories (molecular function and biological process) of differential proteins for 6 coral species. Individual species were searched against their respective metagenome FASTA database and differential proteins assessed via intraspecific comparison to reduce variability inherent to species-specific peptide matches. GO category names are on the left. The number in parentheses indicates number of species out of 6 that contained differential proteins per GO category. X-axis indicates the number of peptide spectral matches (PSMs) which are a measure of protein abundance. There were no common differential proteins across all species. Bioluminescence and Generation of precursor metabolites and energy are indicative of higher number of PSMs related to Green-Fluorescent protein in 4 of 6 species.

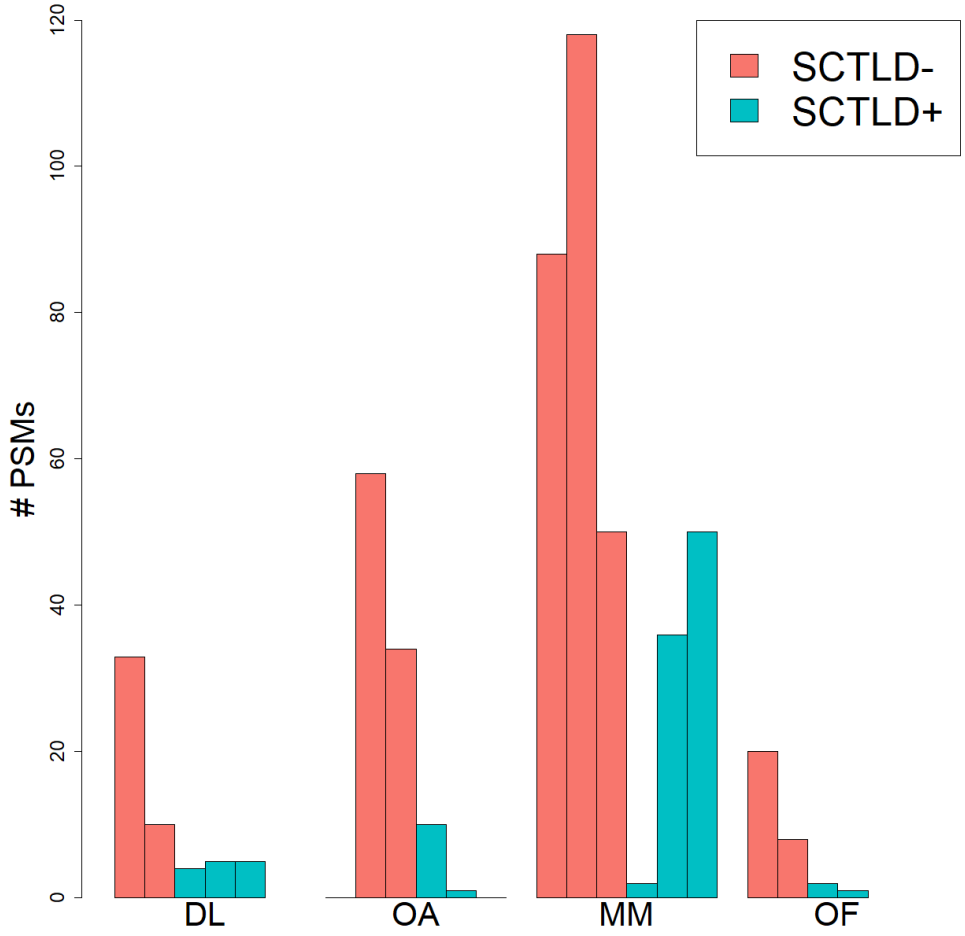


Figure 17. Peptide spectral matches (PSMs) of Green-Fluorescent protein-like protein across 4 of 6 taxa identified in Figure 16 representing Bioluminescence and Generation of precursor metabolites and energy. Fisher's Exact test, $P<0.05$ for all species listed. DL = *Dichocoenia labyrinthiformis*; OA = *Orbicella annularis;* MM= *Montastrea meandrites;* OF = *Orbicella faveolata.* A consensus reduction in GFP-like protein was visualized. Blue bars are SCTLD+ and red bars are SCTLD-. Individual PSMs are plotted for each sample.

6. Tables

Table 1. Protein and tryptic peptide concentrations for 46 coral specimens representing 9 species. Unaffected coral specimens are labeled as SCTLD-. SCTLD+ = stony coral tissue loss disease.

SPECIES	SCTLD-				SCTLD+		
	sample number	protein (µg/µl)	peptide (µg/µl)		sample number	protein (µg/µl)	peptide (µg/µl)
Pseudodiploria strigosa	PSTR 114	5.37	0.36		PSTR 79	3.52	0.18
	PSTR 210	1.48	0.24		PSTR 226	1.29	0.34
	PSTR 258	2.93	0.53				
Pseudodiploria clivosa	PCLI 22	2.65	0.46		PCLI 98	2.48	0.48
	PCLI 28	2.62	0.55				
	PCLI 118	2.95	0.41				
Orbicella faveolata	OFAV 35	2.06	0.43		OFAV 240	1.65	0.18
	OFAV 113	3.09	0.24		OFAV 261	2.42	0.40
Orbicella annularis	OANN 120	2.20	0.33		OANN 241	1.03	0.54
	OANN 223	1.57	0.37		OANN 266	1.49	0.66
	OANN 268	0.53	0.36		OANN 363	1.35	0.18
Montastrea cavernosa	MCAV 13	4.31	0.13		MCAV 112	1.88	0.39
					MCAV 217	1.97	0.33
					MCAV 236	3.89	0.38
Colpophyllia natans	CNAT 9	3.51	0.24		CNAT 92	1.77	0.15
	CNAT 227	1.99	0.25		CNAT 94	1.34	0.26
	CNAT 253	1.87	0.32		CNAT 96	4.64	0.10
Dichocoenia labyrinthiformis	DLAB 73	3.88	0.60		DLAB 232	2.33	0.87
	DLAB 119	5.23	0.50		DLAB 220	4.50	0.49
	DLAB 205	2.90	0.62		DLAB 251	4.88	0.65
Dichocoenia stokesii	DSTO 23	2.60	0.52		DSTO 10	1.64	0.37
	DSTO 24	1.93	0.39		DSTO 37	2.72	0.67
	DSTO 90	0.85	0.35		DSTO 265	1.97	0.48
Montastrea meandrites	MMEA 93	2.00	0.24		MMEA 264	2.44	0.44
	MMEA 123	4.20	0.20		MMEA 249	3.11	0.37
	MMEA 124	7.32	0.21		MMEA 275	2.09	0.58

Table 2. Statistics of scaftigs (>=500bp). Total Len. (bp) stands for length of all the Scaftigs. Num. stands for the total number of Scaftigs. Average Len. (bp) stands for the average length of all the Scaftigs. N50 or N90 length are defined as the shortest sequence length at 50% or 90% of the genome. Maximum Length means the max length of Scaftigs.

SampleID	Total length(bp)	Number	Average length(bp)	N50 Length(bp)	N90 Length(bp)	Maximum length(bp)
CN.92	361,784,409	197,134	1,835.22	2,858	713	68,123
CN.94	434,747,625	299,694	1,450.64	2,084	606	68,097
CN.96	384,526,684	245,651	1,565.34	2,360	629	45,236
DL.73	498,660,692	370,104	1,347.35	1,792	591	64,344
DL.119	585,904,926	446,942	1,310.92	1,634	590	65,891
DS.23	475,030,550	348,803	1,361.89	1,738	621	45,116
DS.37	784,424,796	662,475	1,184.08	1,342	640	63,271
OA.241	494,169,944	386,359	1,279.04	1,670	585	33,912
OA.363	772,173,189	764,178	1,010.46	1,008	561	1,172,196
OF.35	484,516,051	349,102	1,387.89	1,924	604	105,609
OF.113	445,791,822	369,744	1,205.68	1,481	581	142,011
OF.240	722,149,967	664,349	1,087.00	1,175	604	45,387
OF.261	804,316,669	673,446	1,194.33	1,340	633	28,479
PS.114	371,561,924	212,269	1,750.43	2,550	710	69,892
DS.265	571,735,559	523,530	1,092.08	1,176	582	26,724
DS.10	596,495,882	512,482	1,163.94	1,300	590	84,330
DL.251	362,382,907	203,647	1,779.47	2,909	671	73,157

Table 3A. 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		

Table 3B. Top-level data for metagenomic sequencing of 17 coral samples continued. Values describe taxonomic annotation, functional annotation, and antibiotic resistance (CARD – comprehensive antibiotic disease resistance database).

Taxonomic Annotation		Functional Annotation		CARD Annotation	
Gene catalogue	4,285,204	Gene catalogue	4,285,204	Gene catalogue	4285204
Annotated on NR	890,769(20.79%)	Annotated on KEGG	409,324(9.55%)	Annotated on CARD	672
Annotated on Unclassified	36.02%	Annotated on KO	197,755(4.61%)/9,300	Annotated ARGs	79
Annotated on Kingdom level	63.98%	Annotated on EC	118,425(2.76%)/2,598		
Annotated on Phylum level	56.68%	Annotated on pathway	128,000(2.99%)/416		
Annotated on Class level	53.86%	Annotated on eggNOG	417,528(9.74%)		
Annotated on Order level	51.24%	Annotated on OG	417,528(9.74%)/18,646		
Annotated on Family level	47.56%	Annotated on CAZy	10,603(0.25%)		
Annotated on Genus level	46.62%				
Annotated on Species level	45.45%				
Assigned Phyla(top 5)	Cnidaria, Proteobacteria, Chloroflexi, Acidobacteria, Actinobacteria				

Table 4. Species level classification of viruses identified from metagenome data. MetaGeneMark was used to predict open-reading frames, followed by reducing redundant seqeunces via CDHIT. Taxonomic classification was predicted by 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 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

Table 5. 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

7. Data Sharing and Raw Files

Metagenomic files

Raw fastq files and metagenomic data have been uploaded to the Sequence Read Archive located at the National Center for Biotechnology Information (NCBI) BioProject ID
SubmissionID: SUB9574423
BioProject ID: PRJNA726962
The locus_tag prefixes for each linked BioSample are included in the locustagprefix.txt file that can accessed from this BioProject in the submission portal:
https://submit.ncbi.nlm.nih.gov/subs/bioproject/SUB9574423/overview

Metaproteomic files

Raw data files (.raw), FASTA search databases (.fasta), Mascot generic files (.mgf), and Mascot search files (.dat), and Scaffold analysis files (.SF3) are accessible via the EMBL-EBI Proteomics Identification Database (Pride).

http://www.ebi.ac.uk/pride/archive/projects/PXD026192

8. Literature cited

1. Bourne, D.G., et al., *Microbial disease and the coral holobiont.* Trends Microbiol, 2009. **17**(12): p. 554-62.
2. Munn, C.B., *The Role of Vibrios in Diseases of Corals.* Microbiol Spectr, 2015. **3**(4).
3. Thurber, R.V., et al., *Virus-host interactions and their roles in coral reef health and disease.* Nat Rev Microbiol, 2017. **15**(4): p. 205-216.
4. Wear, S.L. and R.V. Thurber, *Sewage pollution: mitigation is key for coral reef stewardship.* Ann N Y Acad Sci, 2015. **1355**: p. 15-30.
5. Ben-Haim, Y. and E. Rosenberg, *A novel Vibrio sp pathogen of the coral Pocillopora damicornis.* Marine Biology, 2002. **141**(1): p. 47-55.
6. Garcia, G.D., et al., *Metaproteomics reveals metabolic transitions between healthy and diseased stony coral Mussismilia braziliensis.* Mol Ecol, 2016. **25**(18): p. 4632-44.
7. Sato, Y., et al., *Unraveling the microbial processes of black band disease in corals through integrated genomics.* Scientific Reports, 2017. **7**(1): p. 40455.
8. Taylor, S.L., S. Wesselingh, and G.B. Rogers, *Host-microbiome interactions in acute and chronic respiratory infections.* Cellular Microbiology, 2016. **18**(5): p. 652-662.
9. Rechenberger, J., et al., *Challenges in Clinical Metaproteomics Highlighted by the Analysis of Acute Leukemia Patients with Gut Colonization by Multidrug-Resistant Enterobacteriaceae.* Proteomes, 2019. **7**(1): p. 2.
10. Meyer, J.L., et al., *Microbial Community Shifts Associated With the Ongoing Stony Coral Tissue Loss Disease Outbreak on the Florida Reef Tract.* Frontiers in Microbiology, 2019. **10**(2244).
11. Meyer, J.L., et al., *Comparative Metagenomics of the Polymicrobial Black Band Disease of Corals.* Frontiers in Microbiology, 2017. **8**(618).
12. Verschaffelt, P., et al., *Unipept Desktop: A Faster, More Powerful Metaproteomics Results Analysis Tool.* J Proteome Res, 2021. **20**(4): p. 2005-2009.
13. Neely, K.L., et al., *Effectiveness of topical antibiotics in treating corals affected by Stony Coral Tissue Loss Disease.* PeerJ, 2020. **8**: p. e9289-e9289.
14. Iwanowicz, D., et al., *Exploring the Stony Coral Tissue Loss Disease Bacterial Pathobiome*. 2020.
15. Smith, E.G., et al., *Screening by coral green fluorescent protein (GFP)-like chromoproteins supports a role in photoprotection of zooxanthellae.* Coral Reefs, 2013. **32**(2): p. 463-474.
16. Salih, A., et al., *Fluorescent pigments in corals are photoprotective.* Nature, 2000. **408**(6814): p. 850-853.
17. Caldwell, J.M., et al., *Intra-colony disease progression induces fragmentation of coral fluorescent pigments.* Scientific Reports, 2017. **7**(1): p. 14596.
18. Landsberg, J.H., et al., *Stony Coral Tissue Loss Disease in Florida Is Associated With Disruption of Host–Zooxanthellae Physiology.* Frontiers in Marine Science, 2020. **7**(1090).