Metaproteomic Analysis of Stony Coral Tissue Loss Disease

Interim report prepared by:
Michael G. Janech, Ph.D[1]., Benjamin Neely, Ph.D.[2], Frank Mari, Ph.D.[2], 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

August 31, 2020

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-August 31, 2020

Table of Contents

1. Background

Florida's coral reefs are 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, the disease has since spread to the northernmost extent of the Florida Reef Tract, and south to the Marquesas Keys in the Lower Florida Keys. 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 not 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 Continuing 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 FL 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 was 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 47 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.

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 47 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 or 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 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.

4. Preliminary conclusions and limitations

Protein extraction with 5% SDS lysis buffer resulted in protein extracts of reasonable quality and concentration for trypsin digestion. Following protein digestion by trypsin using the S-trap method, tryptic peptide concentrations for all 47 coral specimens exceed the lower limit required for mass spectrometry analysis.

Metagenomic sequencing was attempted for all 47 specimens, but was successful for only 17 specimens due to low DNA concentration. 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 because differences are driven by species included in the sequencing effort and not disease.

FASTQ and FASTA files are deposited in a public domain at:

https://drive.google.com/drive/folders/1LekZaNAsATB6R0mywBw5lZKi-igLe2JZ?usp=sharing.

Alternatively, files can be requested from the PI (Janech) at janechmg@cofc.edu.

Covid-19 greatly impacted this project and its ability to pivot when issues were encountered. The complete shutdown of the Hollings Marine Laboratory from March 30 to present day did not permit peptide analysis or re-extraction of DNA from alternate samples due to lab and office closure. The Janech laboratory was relocated to Grice Marine Laboratory, College of Charleston, to allow some wet lab work to be completed. Personal computers were largely utilized to generate this report and for data analysis.

5. Figures

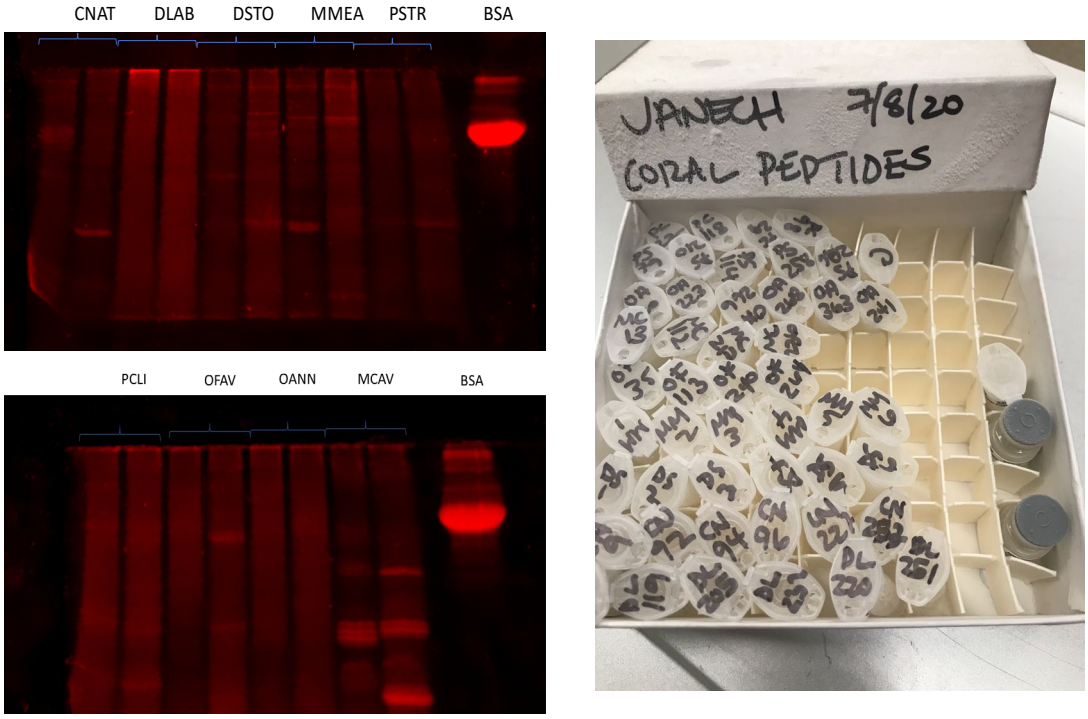

Figure 1. (Left) Protein PAGE of a subset of proteins for different coral species to assess quality. Lanes are labelled using abbreviations listed in Table 1. 20µg protein was loaded per lane. BSA = bovine serum albumin. (Right) protein digests in -80 freezer at Grice Marine Laboratory.

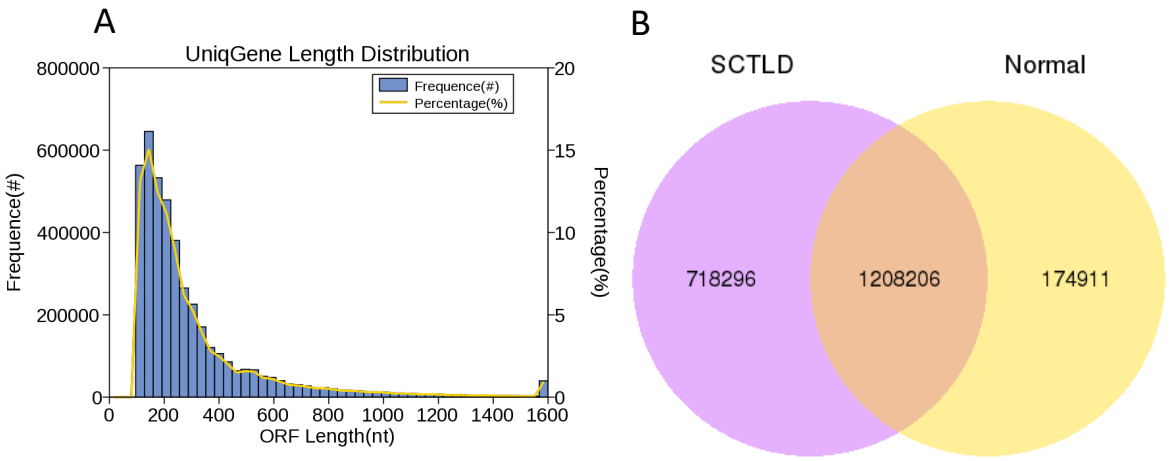

Figure 2. A) Open-reading frame length frequency for all 17 corals combined. Most open reading frames were less than 250 base pairs. B) Venn diagram of common and unique genes included in the analysis for both normal and SCTLD corals.

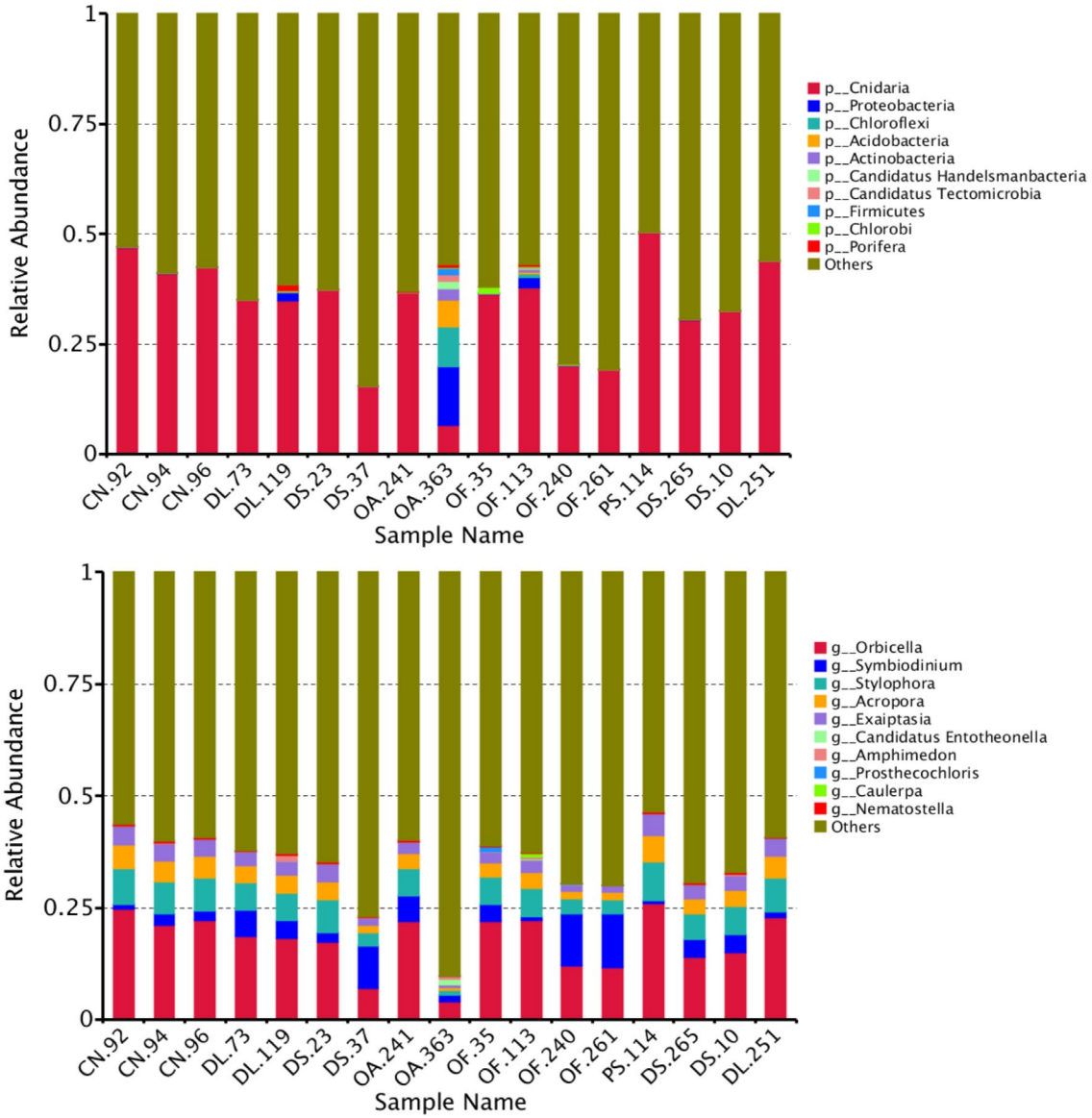

Figure 3. Taxonomic distribution across taxa. Based on the abundance of each taxonomic level, the top 10 taxa were selected and the other taxa were set as "Others". Bar charts show the relative taxonomy abundance of each sample in different taxonomic level. (Top) Taxonomic distribution by phyla. (Bottom)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

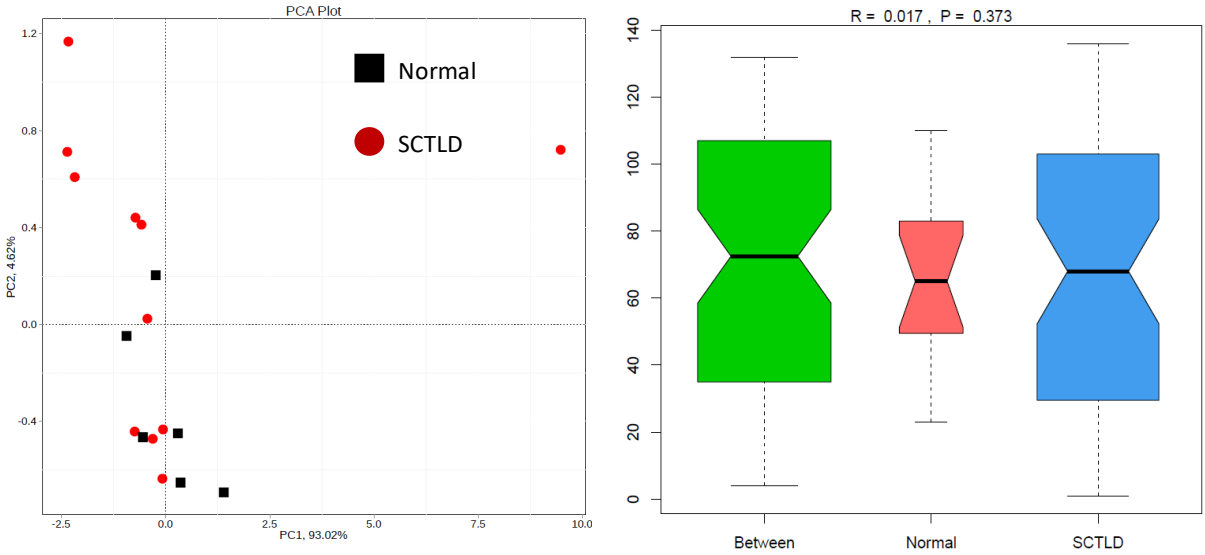

Figure 4. (Left) Principle component analysis of taxonomy at the phylum level. The percentage stands for the contribution of the principle components to the variation in samples. Each point in the graph stands for a sample. Samples belong to the same group that are in the same color. (Right) Analysis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

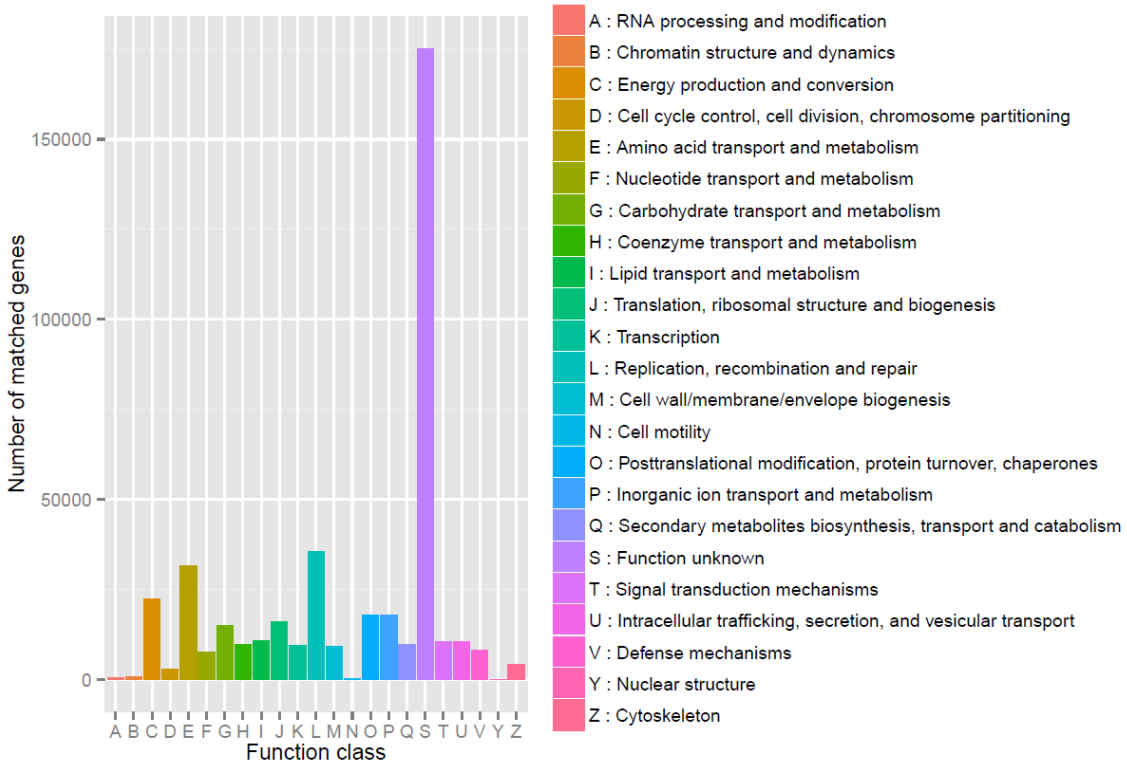

Figure 5. Functional grouping of genes across all sequenced taxa using EggNOG mapper (A database of orthology relationships, functional annotation, and gene evolutionary

6. Tables

Table 1. Protein and tryptic peptide concentrations for 47 coral specimens representing 9 species. Unaffected coral specimens are labeled as Normal. SCTLD = stony coral tissue loss disease.

SPECIES	Normal				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				

7. 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. Liao, B., et al., *iMetaLab 1.0: a web platform for metaproteomics data analysis.* Bioinformatics, 2018. **34**(22): p. 3954-3956.