Stephen F. Austin State University
SFA ScholarWorks

Electronic Theses and Dissertations

5-2018

Changes in the microbial community of Lubomirskia baicalensis affected by Red Sponge Disease

Colin Rorex
Stephen F Austin State University, colinrorex@gmail.com

Follow this and additional works at: https://scholarworks.sfasu.edu/etds

Part of the Bioinformatics Commons, Biology Commons, Environmental Microbiology and Microbial Ecology Commons, Marine Biology Commons, Other Genetics and Genomics Commons, and the Other Immunology and Infectious Disease Commons

Tell us how this article helped you.

Repository Citation

Rorex, Colin, "Changes in the microbial community of Lubomirskia baicalensis affected by Red Sponge Disease" (2018). *Electronic Theses and Dissertations*. 189.
https://scholarworks.sfasu.edu/etds/189

This Thesis is brought to you for free and open access by SFA ScholarWorks. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of SFA ScholarWorks. For more information, please contact cdsscholarworks@sfasu.edu.

Changes in the microbial community of Lubomirskia baicalensis affected by Red Sponge Disease

Creative Commons License

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.

This thesis is available at SFA ScholarWorks: https://scholarworks.sfasu.edu/etds/189

Changes in the microbial community of *Lubomirskia baicalensis* affected by Red Sponge Disease

By

Colin Boyd Rorex, Bachelor of Arts

Presented to the Faculty of the Graduate School of

Stephen F. Austin State University

In Partial Fulfillment

Of the Requirements

For the Degree of

Masters of Science

STEPHEN F. AUSTIN STATE UNIVERSITY

May, 2018

Changes in the microbial community of *Lubomirskia baicalensis* affected by Red Sponge Disease

By

Colin Boyd Rorex, Bachelor of Arts

APPROVED:

Alexandra Martynova-Van Kley, Thesis Director

James Van Kley, Committee Member

Kenneth Farrish, Committee Member

Matthew Kwiatkowski, Committee Member

Pauline Sampson, Ph.D.
Dean of Research and Graduate Studies

ABSTRACT

Lake Baikal is the oldest known lake and a unique ecosystem, home to several species of fresh water sponge. A disease outbreak affecting the dominant species, Lubormirskia baialensis, was recently reported. The cause of the disease has not been determined but one of the current hypothesis is that the increase in methane concentration is correlated to the disease outbreak. This pilot study characterized the microbiomes of sick and healthy sponges through the use of 16S rRNA sequencing. Sick sponge microbiomes shared a conserved group of taxa while the healthy sponge microbiomes had greater diversity. Indicator species analysis identified two significant taxa in the sick sponges that are acidophilic. There was also a decrease of methanotroph taxa in the sick sponge samples compared to the healthy sponge samples which suggested methane was not associated with the disease outbreak. However decreased pH may be a factor related to Red Sponge Disease.

ACKNOWLEDGEMENTS

There are many people whom I wish to thank for their help, support and kindness. Without them I would not be where I am today. Dr. Alexandra Van Kley, without your knowledge, support and enthusiasm this project would not have turned out as well as it has. You went out of your way these last two years to make me feel welcome both on campus as a student and also a friend to you and your family. I will always be in your debt.

To my committee members, your time and insight on this project has been invaluable. To Armen Nalian, without your expertise in coding this project would have taken easily twice as long. To my fellow students that I have met over these last two years, you have made my time in Nacogdoches truly enjoyable.

A special thanks to Kerri Ridenour, for all that you have done. You always managed to cheer me up and your care packages were always appreciated.

TABLE OF CONTENTS

ABSTRACT i
ACKNOWLEDGEMENTS..... iii
LIST OF FIGURES iv
LIST OF TABLES v
INTRODUCTION 1
Sponge Bioremediation 2
Historical Sponge Disease 4
Environmental Factors 7
Microbiome Profiling 10
HYPOTHESIS 14
MATERIALS AND METHODS 15
Samples 15
Primers 16
Amplification and Sequencing 18
Cleaning of FASTA Files 19
Classification 20
Data Analysis 22
RESULTS 27
Taxonomic Data 27
Samples and Sample Nomenclature 28
Comparison of the Bacterial Primers 28
Species Richness 30
Multivariate Analysis 32
DISCUSSION 44
LITERATURE CITED 52
APPENDIX 64
VITA 79

LIST OF FIGURES

Figure 1: Dissimilarity of samples using Bray-Curtis index and clustered by UPGMA 31

Figure 2: Species Richness 32

Figure 3: PCA of all samples 33

Figure 4: NMDS using multiple distance formula 35

Figure 5: NMDS using Bray-Curtis distance with non-significant environmental fit results 36

Figure 6: NMDS using Bray-Curtis distance with a joint plot based on disease state of the organism 37

Figure 7: NMDS using Bray-Curtis distance with taxonomic vector data superimposed 39

LIST OF TABLES

Table 1: Summary of samples 17

Table 2: Primers used for microbiome profiling 18

Table 3: Descriptive maxtrix used for multivariat analysis 25

Table 4: Sample names and a general summary of the sample 28

Table 5: A two way ordered table generated by TWINSPAN 40

Table 6: Summary of the significant indicator species from all data sets 42

APPENDIX

Table 1: Species list of all recovered bacterial taxa for healthy and untreated sponge samples 64

Table 2: Species list of all recovered bacterial taxa for diseased and treated sponge samples 69

Table 3: Indicator Species Analysis of samples grouped together based on health of the organisms they were harvested from 74

INTRODUCTION

Lake Baikal in Siberia is the largest freshwater lake by volume on the Earth and is the source of 20% of all free fresh water. It is a rift lake formed by extensional tectonics between the Eurasian and Amurian plates. The lake reaches a depth of 1600 meters with a sediment layer up to 7,000 meters between the lake bed and rift bed. It is thought to be approximately 25 million years old and contains a vibrant and complex ecosystem.

The first reports of a disease outbreak in the endemic freshwater sponge *Lubomirskia baicalensis* of Lake Baikal were in 2016 (Denikina et al., 2016). Diseased sponges show atypical pink lesions suggesting this is a Red Sponge disease. Prior to this, there was no historical record of disease outbreaks in this sponge population. However, disease outbreaks and the corresponding loss of sponge populations have occurred in other ecosystems and can have severe consequences for the ecosystem. Sponges are filter feeders, most consuming microorganisms in their environment. They also play roles in nutrient cycling in the water column and are an important element in the food web of aquatic ecosystems (Bell, 2008).

This focus on sponges and their ability to consume microorganisms is becoming an important topic in research. Put into the context of human pollution

of ecosystems, an organism able to consume bacteria can play a powerful role as a bioremediator. Sponges can function by consuming microorganisms to protect important aquatic systems from toxic plankton blooms and fecal bacteria from sewage runoff entering the water (Milanese et al., 2003). They also can be directly applied in fisheries to treat harmful microbial byproducts of industry (Zhang et al., 2010).

Sponge Bioremediation

A study of the subtropical Atlantic Ocean found bacterial loads of 2.3×10^5 per ml and virus-like particle loads of 5.58×10^6 per ml in the epipelagic layer (De Corte et al., 2010). Sponges can directly utilize both bacteria and virus-like particles as food sources, resulting in reduction in microbial loads (Bell, 2008). Research conducted in the Florida Bay linked cyanobacteria and phytoplankton blooms with large scale loss of sponge populations (Bradley et al., 2006). This is significant because phytoplankton blooms can have many effects, ranging from fish death due to decreased oxygen levels to loss of sea grasses via a light reduction mechanism to toxic blooms that can affect humans and fisheries (Bradley et al., 2006).

This concept of bioremediation has been further explored in recent aquaculture studies. One study investigated the use of *Hymeniacidon perlevis* to

reduce bacterial loads in a commercial *Scophthalmus maximus* fishery (Zhang et al., 2010). In this study *H. perlevis* was able to reduce microbial levels of fecal coliform, vibro, and other bacterial species in effluent from the fishery. Another study used a controlled laboratory environment to quantify the filtration capacity of *Chondrilla nucula* as up to $7x10^{10}$ *Escherichia coli* cells per hour per square meter of sponge mass (Milanese et al., 2003). Microbial filtration is not a species directed mechanism. Sponges will deplete the surrounding water of all microbial organisms as they feed (Patterson et al., 1997). While sponges are able to recognize symbiotic bacteria, they do not selectively eat one non-symbiotic species over another (Nguyen et al., 2014). So in the *S. maximus* fishery example, the waste waters were rich in microbes from fish excrement**,** presumably if the same sponges were cultivated near the waste output of a brewery**,** they would remove yeast or other microbial effluent with similar efficiency.

The relationship between sponge bioremediation and microorganism loads in aquatic systems drives the importance of understanding the nature of sponge disease. Identification of the causative agent(s) for the sponge disease will describe the type of environmental pollution that needs to be rectified to curb the disease. A simple example would be if a microbe identified as a causative agent is also in untreated water from a sewage treatment plant**.** Then repairs or upgrades to the plant would resolve the disease outbreak.

Historical Sponge Disease

Sponge disease is a global occurrence given that; sponges are found in aquatic ecosystems across the globe. Like diseases in other organisms, sponge diseases are quite varied in appearance and progression. Numerous studies have been conducted trying to identify the causative agent(s), and the results describe a complex system between sponges, microbes, and the surrounding environment.

Initial studies focused on identifying specific microorganisms responsible for disease, viewing diseases in Porifera as mechanistically similar to disease in humans with pathogen A causing disease A. For example, disease in *Rhopaloeides odorabile* in at the Great Barrier Reef was associated with a novel α-proteobacteria (Nicole et al., 2002). In the Red Sea, microbiota recovered from diseased sponge tissue at two sites 30 km apart was assayed. Both sites showed colonization of verrucomicrobia despite their separation, indicating one potential causative agent (Gao et al., 2015).

Alternately**,** disease may be caused by a more global microbial shift in the ecosystem. Nutrient enrichment of an ecosystem, such as coral reefs, may be the direct cause of disease (Vega Thurber et al., 2014). While that study focuses on coral reefs**,** a similar situation could occur in sponge disease. Since bacterial replication is limited by nutrient availability, if nutrient constraints are lifted the

bacterial species will replicate until secondary mechanisms limit their growth (Schaechter et al., 1958). This change in microbial population, where formerly low concentrations of potentially pathogenic bacteria increase in abundance, may cause outbreaks of disease (Gochfeld et al., 2012). These would be situations where not a single bacterium is the causal agent, but a superinfection is responsible for disease. A homologous situation in humans would be a superinfection of Epstein-Barr Virus and the malaria parasite. When both infections occur at the wrong time, there is a synergistic inhibition of the patient's immune system resulting in death (Matar et al., 2015). While both pathogens cause distinct diseases with different symptoms**,**
the co-infection results in a specific morbidity and mortality that only occurs during co-infection.

This concept of multiple causative agents can be addressed using massive parallel sequencing and other bacterial diversity techniques. Sponge White Patch disease affecting the Caribbean Sea sponge *Amphimedon compressa*, originally thought to be caused by the introduction of a sponge boring bacteria, was re-evaluated utilizing these techniques. Diseased sponges had a different microbiota than healthy sponges, with some of the species detected previously implicated in other sponge and coral diseases (Angermeier et al., 2012). This result suggests that multiple causative agents may be playing a role in this disease. Both studies were conducted in the Caribbean Sea,

allowing for direct links between nutrient enrichment, microbial proliferation and disease.

A sponge disease outbreak off the coast of Papua New Guinea illustrates this multiple organism mechanism of disease. Five bacterial species were isolated from diseased tissue and used to inoculate healthy sponges under laboratory conditions (James et al., 2006). Using each of the five bacterial isolates alone to inoculate healthy sponge tissue did not cause disease. It was only when all five bacterial isolates were combined into one inoculum that the same disease was observed (James et al., 2006). In the sponge disease Sponge Necrosis Syndrome, two bacterial and four fungal species were identified in the diseased tissue. It was found that inoculation of healthy sponge tissue with a mixture of one bacterial species and one fungal species together was the minimum required to cause disease (Sweet et al., 2015).

Nutrient and microbial pollution may not be the only factors in disease outbreaks. Several studies linked increasing water temperatures from climate change to disease outbreaks (Blanquer et al., 2016; Cebrian et al., 2011). Interestingly, one commonality in sponge disease outbreaks is environmental change. This can be either through a stress mechanism on the sponge weakening the immune system or a temperature dependent sponge-symbiote interaction. An example of this in *L.baicalensis* is its symbiotic relationship with a dinoflagellate species which regulates the heat shock protein pathway as a

defense against cold (Müller et al., 2007). If warming of the lake would disrupt this symbiotic relationship, there would be sponge stress or death later when temperatures drop in winter.

Since this project is focusing on the disease of *L.baicalensis*, the environment of Lake Baikal must be taken into account. As direct access to the lake for water testing is not possible; this study can at best use the changes in microbiome to describe ecological changes that are promoting this disease.

Environmental Factors

Lake Baikal reaches depths over 1.5 km with a sediment layer approximately seven km deep. This sedimentary layer is rich in methane and other hydrocarbons that escape into the water and contribute to its methane load. This can be directly observed by oily sheens on the surface of the lake and methane gas bubbling to the surface (Kapitanov et al., 2007; Schmid et al., 2007).

Several long-term studies of Lake Baikal have described the effects of climate change on this ecosystem. Over the last 60 years there has been an average increase in surface temperature, down to depths of 25 m, of 0.2^{o}C per decade (Hampton et al., 2008). The greatest temperature increases have been during the summer, 0.38^{o}C per decade, and these changes are large enough to

impact winter temperatures despite it being a sub-arctic lake (Hampton et al., 2008). These changes in average temperature are not linear but have cycles of warming and cooling with data from the last decade indicate the start of a new warming period (Hampton et al., 2011). This warming, especially during winter, can place stress on organisms such as *L. baicalensis* that have adapted well to the seasonal variation in lake temperatures. From 1999 to 2015 Lake Baikal and the Selenga River, the main tributary of Lake Baikal, has seen a reduction in rainfall ranging from 30 to 14 mm compared agaist a wet period from 1980 to 1998 (Dabaeva et al., 2016). This has seen both a reduction in lake water level to the lowest levels in 100 years and an increase in forest fire activity in the Lake Baikal basin (Dabaeva et al., 2016).

Methane concentration in Lake Baikal has been increasing (Zakharenko et al., 2015). There are two mechanisms proposed to explain this increase: either decreasing lake levels are reducing the pressure on the sediment layer or the increasing temperature of the lake is allowing for more disassociation of gas hydrates into the water (Granin et al., 2012; Zakharenko et al., 2015). It is also possible that both mechanisms are acting synergistically. This increase in methane concentration is hypothesized to be a contributing factor to the disease outbreak in *L. baicalensis* (Denikina et al., 2016).

Tourism is playing an increasing economic role to the economies surrounding Lake Baikal. In 2013 the Agency on Tourism for the Irkutsk reported

over 1,000,000 Russian and foreign visitors (Kirillov et al., 2014). Listvyanka is one of the primary tourist destinations on the western shore of the lake and part of one of two special economic zones designated in the Baikal Valley. Insufficient sewage processing infrastructure has resulted in green algae growth in the waters off the coast of Listvyanka (Kravtsova et al., 2014). This is a significant point source of nutrient and microbial pollution. This increased nutrient availability causes an increase in algae growth (algae blooms). When an algae bloom ends the aerobic decay of the algae results localized oxygen depletion. This in turn kills off aerobic species in the affected region.

There are also industrial factors to consider. There are plans to convert a recently closed pulp mill on the southern shore of Lake Baikal into a tourist resort center (Staff, 2014). The mill site still houses over six million tons of industrial waste that is awaiting remediation. The primary tributary into Lake Baikal, the Selenga River flows through several cities and industrial sites and also may contribute to pollution in Lake Baikal. For example, extraction of coal, gold, and wolfram along with refining sites along the Selenga River result in phosphorous and cyanide compounds entering the water system (Kasimov et al., 2017).

Each of these environmental factors could be related to the recent disease outbreak. Some of the factors would correlate to specific changes in bacterial composition in Lake Baikal. Both fecal waste contamination and increased methane concentrations result in altered microbial populations. If those

populations are changing in the microbiome of sick and healthy sponges, it implicates specific environmental changes with the disease mechanism.

Microbiome Profiling

Taxonomic identification of complex multi-cellular organisms classically uses morphological features and a dichotomous key. There are some drawbacks to this approach as features can change based on season, or life cycle stage, or among isolated populations of the same species. The dichotomous key itself may be limited or overly complex (Walter et al., 2007). For example, with bacteria the best characterized organisms are those of medical or economic significance with other organisms being less well-characterized (Emerson et al., 2008). .

Classically, bacteria are identified by a combination of morphology and metabolism (Emerson et al., 2008). This approach requires single organisms to be isolated from samples and then grown clonally to sufficient number for analysis. Once sufficient copy numbers have been reached, samples can be serially plated on a range of diagnostic culture media to characterize metabolism. Samples can be mounted on microscope slides and stained against a number of phenotypical traits. The Identification is then done with a dichotomous key using the sum total of the previous results. There are a number of drawbacks to this

approach. Initial culturing may be difficult as the nutritional requirements and optimal environment for the bacteria is unknown until identification is complete. Also, some of the critical characteristics may be coded on transposable elements which can result in misidentification.

These drawbacks can be best illustrated with the historic classification of *Shigella* as a separate species from *E. Coli*. *Shigella*, through expression of shigatoxin, causes dysentery and is a significant human pathogen. *Escherichia. coli* is also an intestinal colonizer; but depending on horizontal gene transfer, it can have varying levels of pathogenicity. Classically, due to morphological and metabolic differences, *Shigella* and *E. Coli* were taxonomically distinct. It wasn't until technological advancements allowed for rapid and lower-cost whole genome sequencing that a wider analysis of both organisms was made, and it was found that *Shigella* is a clade of *E. coli* that has acquired pathogenic elements from horizontal gene transfer (Pettengil et al., 2016).

Genetic approaches offer alternatives to standard phenotypical taxonomic classification. In general, this approach relies on a DNA sequence that has a low degree of variability among members of the same species, but significant differences at a species or genus level. This sequence can then be amplified by polymerase chain reaction (PCR), sequenced, and then compared against reference databases for a species identification. The ribosome RNA genes are currently one of the principle genes targeted for phylogenetics (Woese et al.,

1990). But there are several options with rich databases that can be used. For complex multicellular eukaryotic organisms, the mitochondrial cytochrome *c oxidase* subunit 1 (COI) is often used (Hebert et al., 2003). For bacterial identifications, the 16S rRNA subunit is used. Fungal identification targets the two internal transcribed spacers (ITS) located between the 18S and 5.8S and 5.8S and 28S genes. For general eukaryotic identification, the 18S rRNA subunit, itself a homologue to the 16S in bacteria, is used. The 16S and 18S rRNA genes contain multiple constant and variable regions, with the constant regions being highly conserved across all phyla of bacteria and the variable regions being conserved at only a genus or species level.

The advent of next generation sequencing (NGS) technologies allow for a mixed pool of DNA to be sequenced concurrently. The drawback to this parallel sequencing is dramatically reduced read lengths when compared against Sanger-style sequencers or third generation NGS platforms. The Illumina MiSeq used in this study offers a maximum read length of 300 bp per direction. Sequencing of amplicons larger than 300 bp is possible. There is a 600 bp theoretical maximum reading window, but a practical limit of 500-520 bp allows enough overlap to ensure alignment between the forward and reverse read. This size constraint plays a role in primer selection or primer design for microbiome profiling.

The full length rRNA genes are too large to be sequenced with a single primer pair on the Illumina platform. The 16S rRNA gene is over 1700 bp, and a typical primer pair would span two variable regions. To sequence the entire gene, multiple reactions would need to be run generating overlapping amplicons, and the resulting data would need to be aligned to recover the full length sequences. For microbiome profiling, a relatively small amplicon covering at most two variable regions is sufficient to recover organism data down to the genus or species level (Wang et al., 2014). However, not all variable regions will yield the same results. The specific variable regions still need to be considered, as different variable regions can yield differences in data (Rintala et al., 2017). In this study, it was found the choice of primer had the largest impact on the type of data, even more than the DNA extraction method. Because of this, multiple primers and variable regions were used in this study.

HYPOTHESIS

1) Changes in the microbiome, rather than a single agent, is linked to the outbreak of the pink lesion disease occurring in the *L.baicalensis* outbreak in Lake Baikal.
2) The microbiome changes, based on metabolic choice or organism type, will be indicative of an environmental change.

Objectives

1) Evaluate four sets of primers used in high-throughput sequencing for their ability to detect bacterial, fungal and eukaryotic populations of the *L. baicalensis* microbiome.
2) Determine the composition of the microbiome in healthy sponges.
3) Determine the composition of the microbiome in diseased sponges.
4) Analyze the recovered microbiome data to see what variations exist between healthy and diseased sponges.
5) Is there an increase in methanotrohps in sick sponges compared to healthy sponges.

MATERIALS AND METHODS

Samples

There were two sets of samples used in this study. The first set was three DNA samples extracted from sponge tissue collected at the Lake Baikal Research Station, part of Irkutsk State Technical University located in Siberia, Russia, and kindly provided by our collaborators. These samples were DNA extracted directly from sponge tissue, so they contained a mixture of both sponge DNA and microorganism DNA. Two of these samples were collected from healthy sponges prior to the reported outbreak of disease (PI and PII). The third sample was collected after the outbreak from a sponge that displayed the red phenotype indicative of disease (PIII). Three additional samples were provided as FASTA files sequenced with the Roche 454 system (Roche, currently discontinued). These three samples were collected after the disease outbreak and were DNA extracted from sponge tissue and amplified using a bacterial V3-V4 primer. One of these samples was from a healthy sponge (Healthy), another from the same sponge that sample PIII was collected from (PIII-454), and the last a sample of diseased sponge tissue cultured in the laboratory and rescued with

anti-microbial treatments (Rescue). A general summary of all samples is included in Table 1.

Primers

Based on information from prior studies of sponge disease, four primers sets were selected to identify both organisms in both the bacterial and eukaryotic domains (Table 2). The rRNA variable regions 3, 4 and 5 (V3, V4, V5) are the most commonly sequenced. A review comparing data recovered from a single variable region to the entire 16S sequence identified V4 as yielding the most representative data (Yang et al., 2016). Due to the constraints of the maximum read size of the amplicon, only one adjacent variable region was sequenced with V4 in a single reaction run. The 357wF/785R primer set was selected. This amplified V3 and V4. Because our samples contained a mixed population of sponge and microorganism DNA, several primer sets were selected to profile the eukaryotic species. The ITS1F/ITS2aR primer set was selected to classify the fungal community. A specific fungal primer set was used rather than simply a general eukaryote primer because the ITS genes are the target for fungal barcoding. This allows identification of fungal organisms through the fungal barcoding database. This is a specialized database containing only fungal sequences, so any sponge sequences should not be returned when classifying

Table 1

Summary of samples. Samples sharing the same name were collected from the same organism. Disease state indicates if the sampled organism was healthy or diseased at the time of collection. Sequencing platform refers to the platform used to generate sequencing data. The primer sets used for sequencing are indicated by their names from RTL Genomics assay listing (http://rtlgenomics.com/s/Amplicon-Diversity-Assay-List.pdf) when appropriate. Amplicon is the brief description of the region of the gene amplified.

Sample name	Disease state	Sequencing platform	Primer (Forward/Reverse)	Amplicon
PI	Healthy	Illumina MiSeq	357wF/785R	16S V3/V4
	Healthy	Illumina MiSeq	515yF/926pfR	16S V4/V5 Universal primer
	Healthy	Illumina MiSeq	ITS1F/ITS2aR	ITS 1
	Healthy	Illumina MiSeq	TAReukF/785R	18S V3/V4
PII	Healthy	Illumina MiSeq	357wF/785R	16S V3/V4
	Healthy	Illumina MiSeq	515yF/926pfR	16S V4/V5 Universal primer
	Healthy	Illumina MiSeq	ITS1F/ITS2aR	ITS 1
	Heallthy	Illumina MiSeq	TAReukF/785R	18S V3/V4
PIII	Diseased	Illumina MiSeq	357wF/785R	16S V3/V4
	Diseased	Illumina MiSeq	515yF/926pfR	16S V4/V5 Universal primer
	Diseased	Illumina MiSeq	ITS1F/ITS2aR	ITS 1
	Diseased	Illumina MiSeq	TAReukF/785R	18S V3/V4
	Diseased	Roche-454	Bacteria	V3/V4
Healthy	Healthy	Roche-454	Bacteria	V3/V4
Rescue	Diseased culture recovered with treatment	Roche-454	Bacteria	V3/V4

this dataset. But there may be other non-fungal eukaryotic organisms of significance, so a general 18S primer set TAReukF/TAReukR was also selected.

The last primer set selected was the 515yF/926pfR. This primer is nominally a V4-V5 16S primer, but it was also described as a universal primer in a study using DNA extracted from sponge tissue (Wang et al., 2014). Universal primers amplify rRNA sequences from all three domains, potentially reducing costs of follow-up sequencing projects by up to 2/3rds, as a single reaction will yield data for bacteria, archaea, and eukaryotes.

Table 2

Primers used for microbiome profiling. All primer sequences run 5' to 3'. The type refers to the target organism for the primer. Region refers to the rRNA regions amplified by the primer set

Type	Primer name	Sequence	Region
Bacterial	357wF	CCTACGGGNGGCWGCAG	V3-V4
	785R	GACTACHVGGGTATCTAATCC	
Fungal	ITS1F	CTTGGTCATTTAGAGGAAGTAA	ITS1
	ITS2aR	GCTGCGTTCTTCATCGATGC	
Eukaryote	TAReukF	CCAGCASCYGCGGTAATTCC	V4-V5
	TAReukR	ACTTTCGTTCTTGATYRA	
Universal/ Bacterial	515yF	GTGYCAGCMGCCGCGGTAA	V4-V5
	926pfR	CCGYCAATTYMTTTRAGTTT	

Amplification and Sequencing

Samples were submitted to RTL Genomics of Lubbock, Texas, (rtlgenomics.com) for amplification and sequencing. Each of the three samples was amplified and sequenced using each of the four primer sets (Table 2).

Sequencing of the amplicons was conducted on the Illumina MiSeq. Data output was two FASTA files.

Cleaning of FASTA Files

Sequencing data was received as FASTA formatted files. The first several nucleotides from each sequence downstream from the primer represent a barcode region of typically six to eight nucleotides which is used to generate the label for the header. The example below illustrates the general format. Here, the nucleotides in parenthesis would correspond to the barcode.

**>SAMPLE DATA HEADER
**

(GCAATTAA)ATTAACNGYCCDVAT

Several steps need to be taken to clean the raw FASTA data prior to any analysis. The barcode region, which is not part of the amplicon, needed to be removed from all samples, as this sequence will not align with any organismal database. Also, the header needed to be converted to a meaningful sample name as this will be the eventual name of the data set containing all of the sequences for that sample. This was done in Python using a script that parses the beginning of each line of the FASTA file. Once loaded as an object, the header was truncated to “Sample.forwardprimer-”, e.g. PI.357wF by deleting everything after the first “-“ symbol until the line break. The barcode regions

were removed by deleting the first eight characters following the line break if they corresponded to the barcode sequences provided with the data files. Each cleaned line of FASTA data was then combined in a new file. The two-file format was preserved with one containing the bacterial and universal primers since both are 16S rRNA primers and one eukaryotic file with the ITS1 and 18S rRNA primers. Sequencing data provided to us by our collaborators was already cleaned and handled as a separate file.

Classification

Cleaned FASTA files were classified using the Ribosomal Database Project's RDP Classifier (Wang et al., 2007). There is a web-based application (http://rdp.cme.msu.edu/classifier/classifier.jsp) which is suitable for short sets of sequences and a downloadable Java environment classifier (http://github.com/rdpstaff/RDPTools) used for the full length files. This is a program that will align FASTA data to one of several reference databases of rRNA sequences and return taxonomic information along with a confidence level. There are several databases available: a general 16S bacterial, a 16S archaea, and two ITS databases. For general 18S classification, the SILVA database was used (https://www.arb-silva.de/) in the QIIME environment (http://qiime.org/1.6.0/index.html).

The resulting output files retained the header data from the FASTA files with taxonomic rank information from domain to species along with a confidence value for each level of classification. Each line of the file represented a single sequence. The output files were run through another Python script (provided by Dr. Armen Nalian) that checks each level of classification for the confidence value starting from species until a confidence value of 80% or greater is found and all other taxonomic data is discarded. This results in each sequence being reduced to the lowest level of taxonomic identification. Counts for each unique taxa were done and duplicates removed resulting in an organism profile with abundance data. Counts were converted to frequency by dividing each count by the sum of the number of counts for that sample. The resulting output files were matrices with samples as columns and taxonomic data as rows that were used for data analysis.

The Basic Local Alignment Search Tool (BLAST, https://blast.ncbi.nlm.nih.gov/Blast.cgi) was also used to check small sets of FASTA data for accuracy. Unlike the RDP classifier or SILVA, the BLAST tool allows for taxa or organisms to be selected or excluded from the alignment query. BLAST was used with small subsets of data to verify results of the ribosome classification.

Data Analysis

R-studio (rstudio.com), an integrated development environment for R, was used for statistical analysis. In addition to the species matrices described above, a descriptive matrix was constructed consisting of information about the samples, including primer type, health state of sample organism, and sequencing platform, and year collected. This matrix is analogous to an environmental matrix for plant community and contained variables that could be tested for during analysis. Multiple R libraries were used, as needed, for the following analysis.

Species richness, the number of unique taxa present in each sample, was plotted where frequency was greater than zero. Cluster dendrograms were used to compare samples to determine the amount of dissimilarity between each sample. The Bray-Curtis index was used from the vegan package to calculate the dissimilarity then samples were clustered using unweighted pair group method with arithmetic mean (UPGMA). This resulted in a value between 1, data sets share no values, and 0, where the data sets are identical.

Several multivariate ordination methodologies were used for this study to visualize patterns in the data. Each sample can considered a collection of vectors or dimensional data, with each representing one of the unique taxonomic classifications whose magnitude corresponds to the frequency of sequences that occurred for each taxa. Multivariate ordination takes these poly-dimensional

objects and collapses them into lower order visual space with an ultimate result of having similar samples cluster together and more dissimilar samples separate farther apart (Paliy et al., 2016). Two different types of analysis were performed, principle component analysis (PCA) and non-metric multidimensional scaling (NMDS).

PCA is an eigenanalysis technique that calculates the variance of the dataset into two orthogonal vectors (Paliy et al., 2016). These vectors, called eigenvectors, have associated eigenvalues that describe the amount of variance in the dataset. Multiple eigenvectors were calculated and the first axis was assigned to the vector with the highest eigenvalue and the second axis to the second highest. Samples were then plotted in the resulting two-dimensional space.

NMDS uses an alternate approach that is iterative. After the number of axes is selected, typically two or three for visualization restrictions, distances between samples are determined, and the resulting fit of the data is calculated. This is repeated multiple times until the best fit, which minimizes a value called stress, is reached or the permutation limit is reached. Stress is reported as a fractional number with 0.3 random, 0.1 an acceptable fit and values less than 0.05 are considered good fits. However, exceptionally low stress values that are reached rapidly may indicate that sample sizes are too small for the method

selected. This becomes obvious when looking at the ordination as many samples will be overlapped.

Multiple distance formulae were tested, including Euclidian ($d_{jk} = \sqrt{\sum_i (x_{ij} - x_{ik})^2}$), Manhattan($d_{jk} = \sum_i |x_{ij} - x_{ik}|$), Gower($d_{jk} = (\frac{1}{M}) \sum_i \frac{|x_{ij} - x_{ik}|}{\max x_i - \min x_i}$), Bray-Curtis ($d_{jk} = \frac{\sum_i |x_{ij} - x_{ik}|}{\sum_i |x_{ij} + x_{ik}|}$) and Raup-Crick$d(_{jk} = (1 - prob(j))$. Both two-dimensional and three-dimensional solutions of the NMDS were generated to compare differences in distribution. There was a decrease in stress as the number of dimensions was increased.

The resulting ordinations will be evaluated to determine what best represented the data. Once an ordination technique is selected, it was tested against various descriptive factors to see which, if any, could be responsible for driving the differences between samples. These descriptive factors were assembled into a matrix and include the following values: sequencing platform, organism the sample was collected from, primer, if the sample was collected before or after the disease outbreak was first observed (Table 3). These variables were fit to the resulting NMDS using the envfit function which returned a p value for statistical significance and visualized with a joint plot linking samples with the same environmental variables together.

Table 3. The descriptive matrix used for data analysis. Sample name is the identifier for this study. Organism ID is an arbitrary label used when the same DNA pool is shared or when both samples are taken from the same organism. The primer refers to the 16S variable region used to sequence the sample. Outbreak status indicates whether the sample was collected before or after the first signs of disease were

detected. The sequencing platform indicates which of the two NGS sequencing platforms was used to generate sequencing data.

Sample name	Organism ID	Primer	Health	Pre/Post Outbreak	Sequencing platform
PI-357wF	PI	V3/V4	Healthy	Pre	Illumina MiSeq
PI-515yF	PI	V4/V5	Healthy	Pre	Illumina MiSeq
PII-357wf	PII	V3/V4	Healthy	Post	Illumina MiSeq
PII-515yF	PII	V4/V5	Healthy	Post	Illumina MiSeq
PIII-357wF	PIII	V3/V4	Sick	Post	Illumina MiSeq
PIII-515yF	PIII	V4/V5	Sick	Post	Illumina MiSeq
PIII-454	PIII	V3/V4	Sick	Post	Roche 454
Healthy-454	PIV	V3/V4	Healthy	Post	Roche 454
Rescue	PV	V3/V4	Treated	Post	Roche 454

Several methods were used to explore the taxa that drive the differences between the samples. First, species data was added to an NMDS. This visually shows the taxa that are driving the position of the data points in ordination space. Second, Indicator species analysis (ISA) was used to determine the statistically significant species that make up different groupings of data. This technique uses a prior clustering of the samples, which was determined by the multivariate analysis, to determine which taxa drive the differences between meaningful sample groups. In short, this technique calculated the relative abundance and frequency of taxa in the samples, and then Monte Carlo permutations were run on the data to calculate the statistical strength of the indicator for each species. The third method used was two-way indicator species analysis (TWINSPAN).

This is a divisive method using reciprocal averaging to determine a division between two groups of samples and then repeated iteratively until a threshold is met. When constructing the final table taxa that occurred within a single sample were excluded.

RESULTS

Taxonomic Data

A total of 217 unique taxa were recovered from the samples amplified with the general bacterial primer set, the universal primer set, and the Roche-454 data provided by our collaborators (Appendix, Table 1).

Fungal taxonomic data, using both fungal datasets in RDP classifier, classified samples into the kingdom Metazoa or subphylum Ecdysozoa and for all samples. The eukaryotic classification using the SILVA database also failed to yield any meaningful taxonomic data outside of the phylum Porifera. BLAST was used to validate these negative results by submitting random subsets of FASTA sequences from this data file and either excluding Porifera (taxid:6040) or including fungi (taxid:4751) to constrain the alignment. There was no improvement in classification, thus sequencing data for both eukaryotic primer sets was excluded from further analysis. In addition, no nonbacterial taxonomic data were recovered from the 515yF/926pfR primer set described as universal.

Samples and Sample Nomenclature

The remaining samples were used for further analysis. Samples were identified either by the organism they were collected from (PI, PII, PIII) or by a general descriptor (Healthy and Rescue), along with either the forward primer (357wF or 515yF) or the sequencing platform used (454). A summary of the samples is provided in Table 4.

Table 4

Sample names and a general summary of the health of the organism the sample was collected from, the sequencing platform and the variable region amplified for sequencing.

Sample name	Organism health	Sequencing platform	16S rRNA region
PI-357wF	Healthy	Illumina	V3-V4
PI-515yF	Healthy	Illumina	V4-V5
PII-357wF	Healthy	Illumina	V3-V4
PII-515yF	Healthy	Illumina	V4-V5
PIII-357wF	Sick	Illumina	V3-V4
PIII-515yF	Sick	Illumina	V4-V5
PIII-454	Sick	Roche 454	V3-V4
Healthy-454	Healthy	Roche 454	V3-V4
Rescue	Cured *in vitro*	Roche 454	V3-V4

Comparison of the Bacterial Primers

The results for the V3/V4 and V4/V5 primers with samples PI, PII and PIII were compared using Bray-Curtis indexing. This evaluated the differences in populations between samples to address whether the different primers were recovering the same data and should be treated as duplicates or different enough that the data sets should be treated as distinct data points. It was

expected that the samples would cluster in one of two ways, either by organism the DNA was collected from or by primer. If clustering by organism, there would be three groups, PI, PII and PIII, each of which would have two elements of primers, V3-V4 and V4-V5. Clustering by primer resulted in two clusters of primer, V3-V4 and V4-V5, each with three elements of organism, PI, PII, and PIII. The Y-axis is dissimilarity, with a lower dissimilarity value indicating higher similarity.

Comparing only samples PI, PII and PIII sequenced on the Illumina platform, they clustered into three groups of two (Figure 1A). One cluster was by sample, PIII, which showed the least overall dissimilarity between the two primers. Interestingly, the other two clusters were grouping by primer with PI-515yF being most similar to PII-515yF and PI-357wF grouping with PII-357wF. However, there was little similarity between clusters with the PIII being >70% dissimilar to the 357wF cluster and the 515yF cluster being an outlier. This unexpected hybrid result with the healthy samples grouping by primer rather than sample could be due to the healthy samples having more diverse microbiomes than the sick samples. With a more diverse microbiome the V3-V4 and V4-V5 primer sets could then recover different sets of taxonomic data resulting in the observed clustering. Another explanation is that a labeling error occurred. The latter explanation was ruled out by having two samples (PI_2-357wF and PII_2-357wF) re-sequenced with the 357wF/785R primer set. The re-sequenced

samples were >80% similar to the initial samples, and were within the same cluster as the original samples meaning that this group is indeed most similar based on primer (Figure 1B). By extension, this means the initial cluster pattern was correct (Figure 1A). Lastly, both initial data sets were combined (Figure 1C). The data from the Roche 454 platform showed a low amount of similarity between the sick (PIII) and healthy samples. The rescue sample appears as an outlier to this group. There was little overlap between the two groups based on sequencing platform.

Species Richness

Species richness, the number of unique taxonomic groups in each data set, ranged from 37 taxonomic groups to 89 taxonomic groups. The Rescue sample showed highest diversity with 89 unique taxa. The samples amplified with the universal primers (515yF/926pfR) showed the lowest species richness with 37 unique taxa in sample PI , 40 with PII and 61 in the PIII sample. The samples amplified with the 357wF primer recovered more diverse bacterial populations with both PII and PIII recovering 77 unique taxa and PI recovering 88. The Roche 454 samples recovered 74, 78 and 89 for the PIII-454, Healthy-454 and Rescue samples respectively (Figure 2).

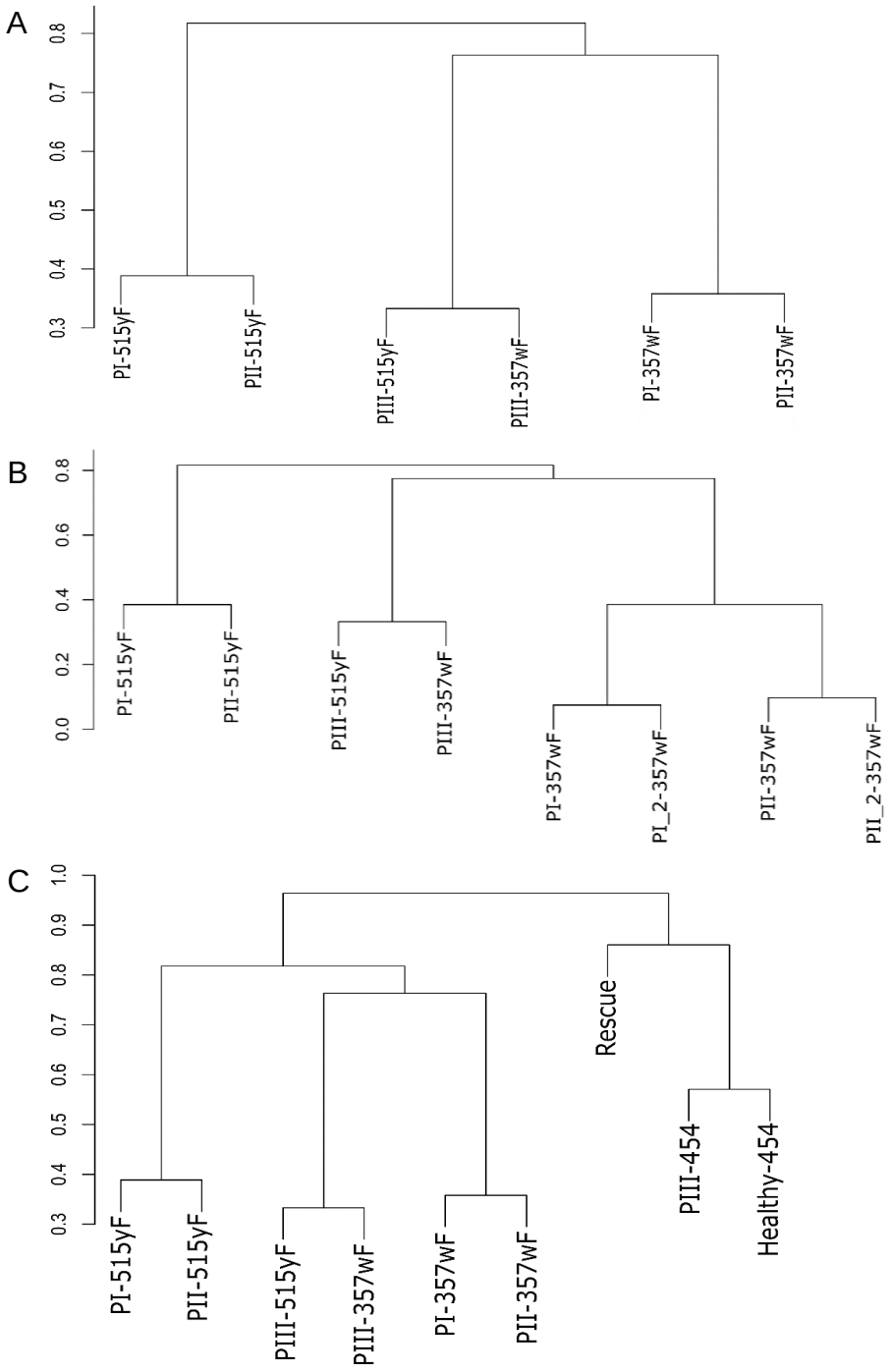

Figure 1. Dissimilarity of samples using Bray-Curtis index and clustered by UPGMA. (a) Initial comparison of Illumina amplified samples. (b) Initial samples plus re-sequenced PI-357wF (PI_2-357wF) and PII-357wF (PII_2-357wF). (c) Combined Illumina and Roche 454 data sets.

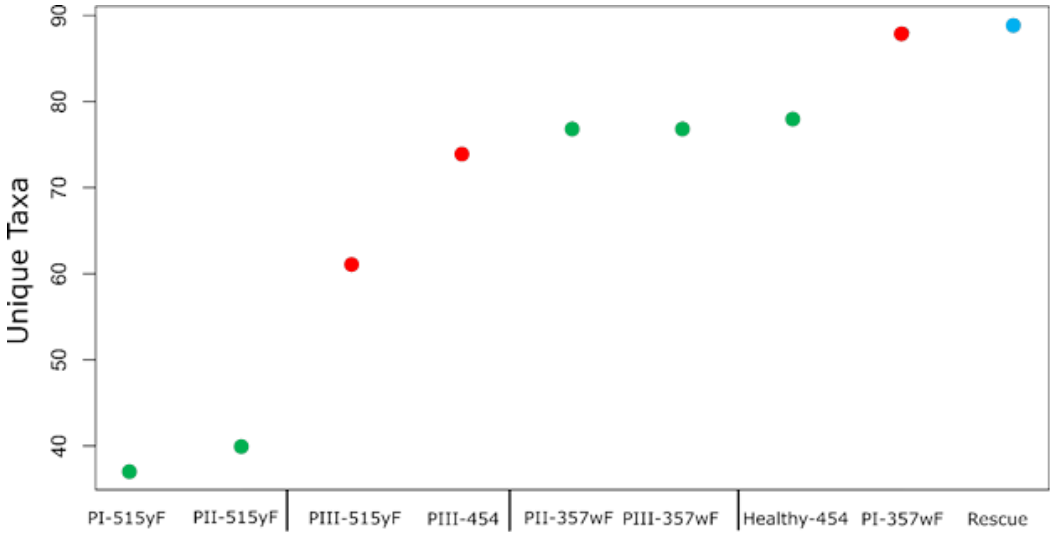

Figure 2. Species richness for all microbiome profiles. PI, PII and Healthy were samples obtained from uninfected sponge tissue. PIII were samples from an infected sponge. Rescue is a diseased sample that was treated and cured under laboratory conditions. Suffixes denote primer set used or sequencing platform.

Multivariate Analysis

The PCA determined the primary axis (x-axis) accounted for 31.6% of the sample variation with 16.2% accounted for in the second axis (y-axis). This separated the data points into two major groups along the primary axis. The laboratory-treated rescue sample was an outlier to the closely-related samples collected from untreated sponges. Variation in the second group was primarily described by the second axis with the PIII-357wF and PI-357wF samples as outliers to this group (Figure 3). Multiple different distance methodologies were used to construct NMDS plots to evaluate which best represented the data both on final stress value and clustering of samples (Figure 4). The Euclidean distance method resulted in a low stress ranging from 0.2 to 0.0414, below the 0.05

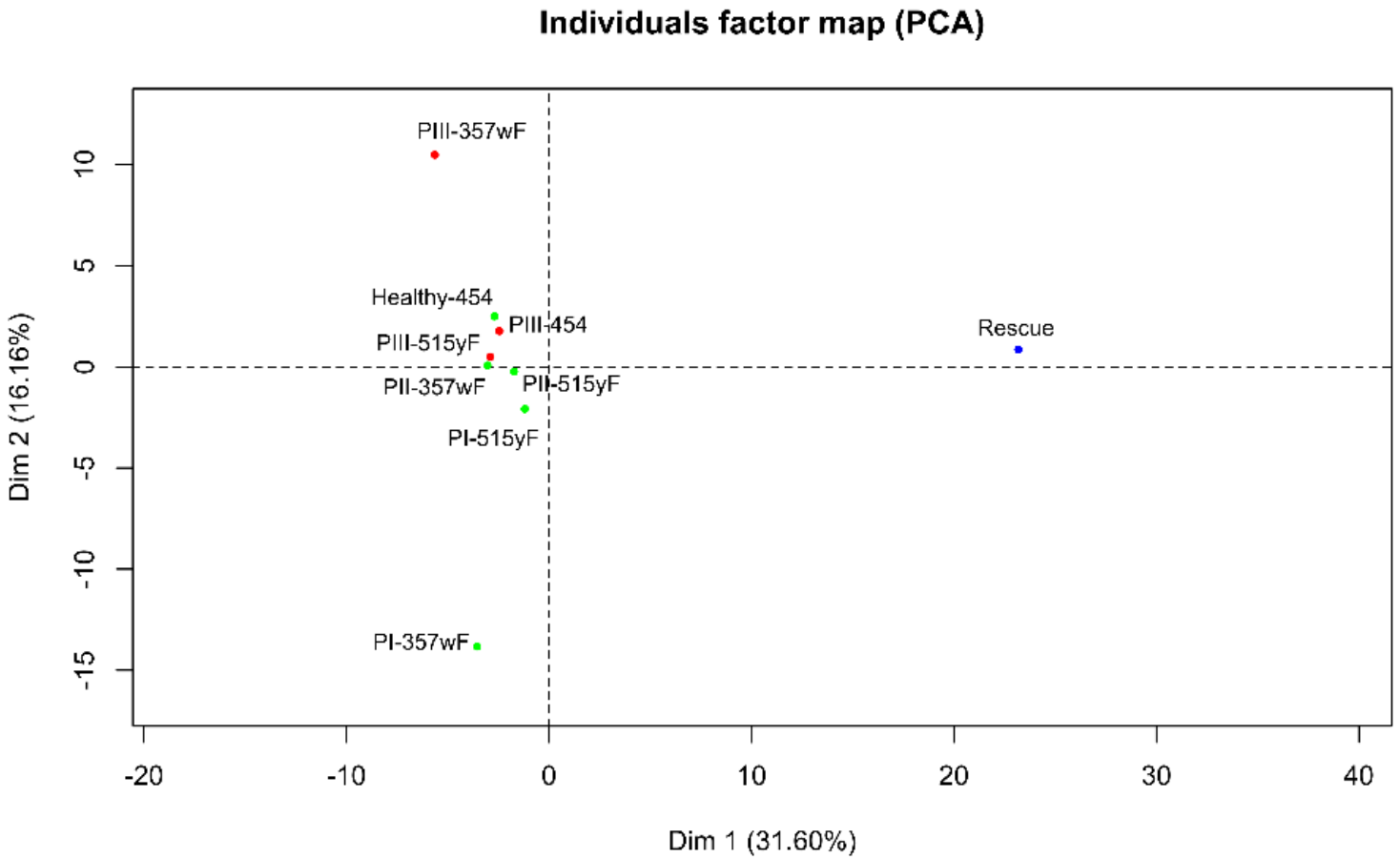

Figure 3. A PCA of all samples. Sick samples are marked in red, healthy in green and the rescue sample in blue.

threshold with all runs resulting in very similar ordinations (Figure 4a). Samples clustered in two groups. The samples from the sick sponge clustered together at the top of ordination space in a relatively tight cluster. The remaining samples, both healthy and rescue, formed a broad group across the bottom first axis. Interestingly, the rescue sample with this analysis is not an outlier group as with the PCA analysis and appeared to be more of a mid-point between the untreated healthy samples. Both the Gower and Raup-Crick distance methods had exceptionally low stress with all results as low as 0.0002 and 0.0005, respectively (Figure 4b). Both methods resulted in identical ordinations that separated the data points into two groups with the healthy and sick samples (PI, PII, PIII and Healthy) as identical with the rescue sample as an outlier. These methods were excluded as they did not result in a separation between multiple points. Bray-Curtis and Manhattan dissimilarities yielded similar stress to the Euclidean, ranging from 0.2 to 0.04 and identical ordination plots to each other after each run (Figure 4c). This yields three groups that appear to organize along the health status of the samples. The PIII samples form a group in the top right of ordination space with the rescue sample being most unrelated to any other sample and pushed to the far side of ordination space. The healthy samples formed a large group between the sick and rescue samples.

A three-dimensional solution using the Bray-Curtis dissimilarity was performed (Figure 4d). This did reduce stress to being consistently under 0.05

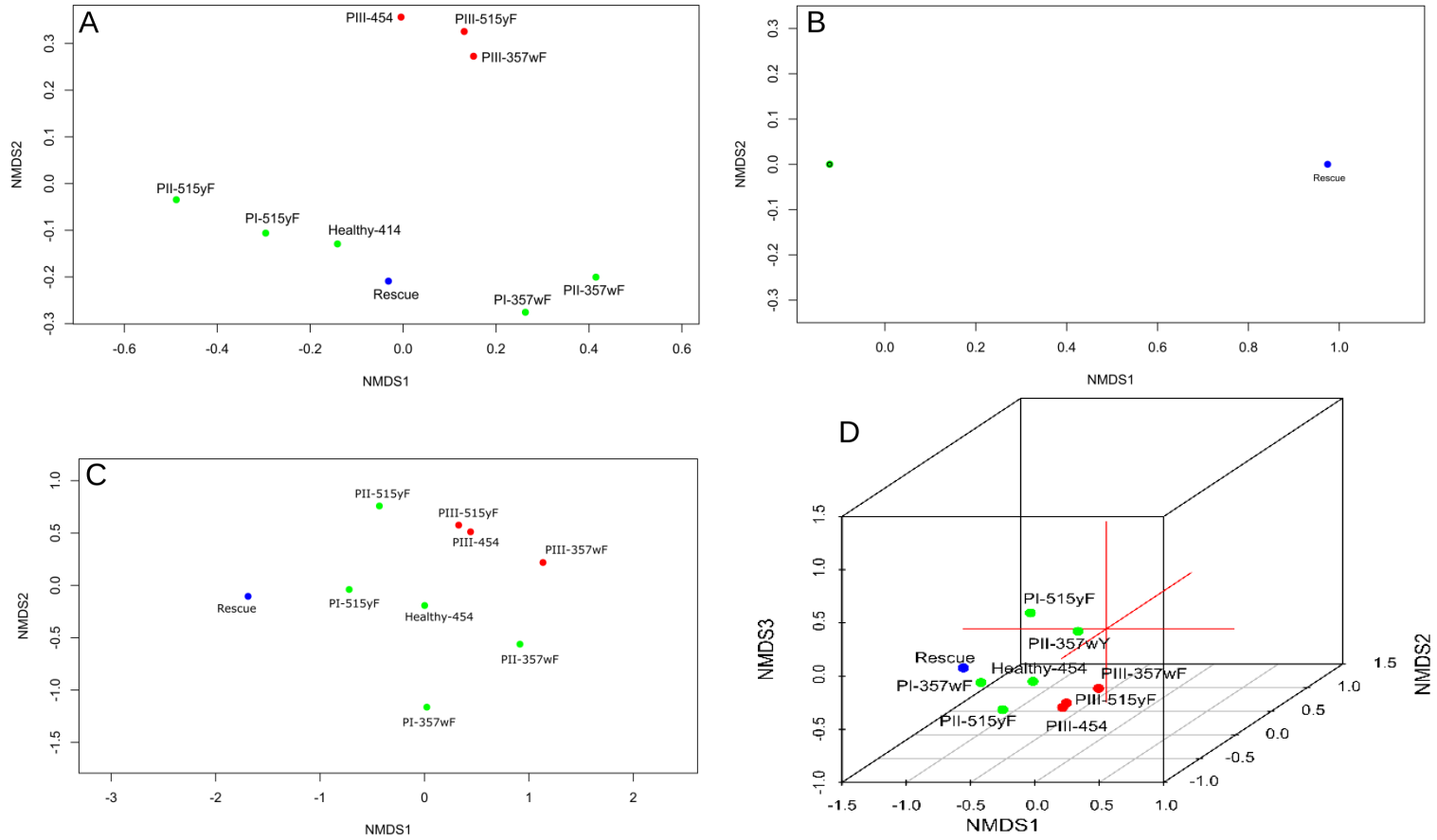

Figure 4. NMDS analysis of samples utilizing different distribution calculations and axis number. Samples are labeled using standard nomenclature and color coded based on health of sample organism. Red are samples from sick sponges, green from healthy and blue the sample treated *in vivo*. (a) Distances calculated using the Euclidean formula in two dimensions. (b) Distances calculated with both Gower and Raup-Crick formulae. All samples except for the rescue sample centered on the same location in ordination space. (c) Distances calculated using Bray-Curtis and Manhattan formulae. (d) A three dimensional NMDS using the Bray-Curtis distance formula.

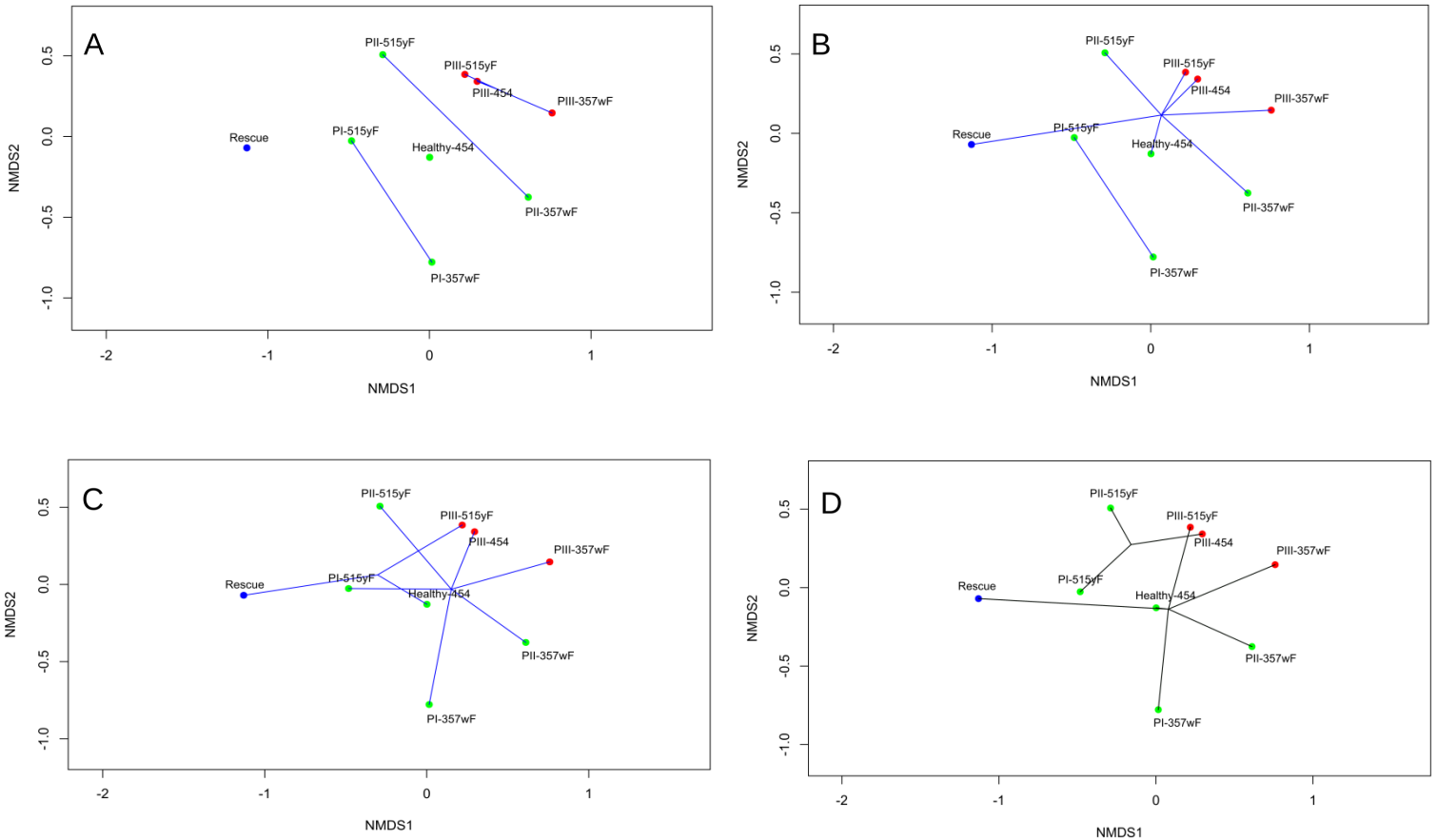

Figure 5. An NMDS using Bray-Curtis distance with non-significant environmental fit results. (a) By organism origin of samples. (b) By samples collected pre or post disease outbreak. (c) By sequencing platform. (d) By variable region sequenced. Samples are color coded by disease state of the organism they were collected from. Red for diseased, green for healthy and blue for the *in vitro* treated and rescued sample

again with invariable distribution of samples between trials and continuance of the separation into three groups. However, determining where the data points are in three-dimensional space was more difficult than the solutions for two dimensions. Going forward, the sample variables will be tested using a two-dimensional Bray-Curtis solution. Using this solution as the underlying pattern for the data, the next step was to see what, if any, characteristics of the samples were driving the grouping (Figure 5). A summary of descriptive variables is listed in Table 3. No statistical significance was found when fitting organism ID ($p = 0.1931$, Figure 5a) by samples collect prior to or after disease outbreak occurance ($p = 0.2955$, Figure 5b), by variable region sequenced ($p = 0.4306$,

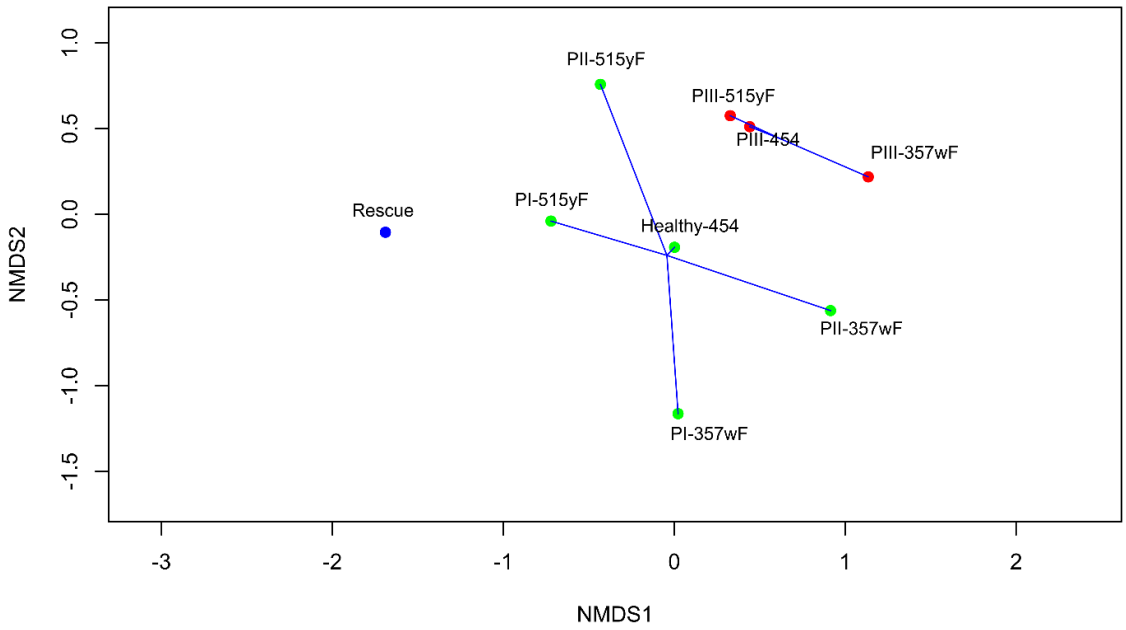

Figure 6. An NMDS using Bray-Curtis distance with a joint plot based on disease state of the organism DNA samples were collected from ($p < 0.05$). Samples are color coded by disease state of the organism they were collected from. Red for diseased, green for healthy and blue for the *in vitro* treated and rescued sample.

Figure 5c), or by sequencing platform (p = 0.4417, Figure 5d). Statistical significance (p = 0.0243) was reached when the health of the sample was used as the descriptive variable (Figure 6). This result indicates that the microbiome of healthy sponge samples (PI and PII) is different from diseased sponge samples (PIII) due to their diseased state. The rescue sample is grouped as an outlier.

The contributions of the various taxa that drive the grouping of samples in ordination space were represented by plotting each taxa as a point in ordination space (Figure 7). These points represent vectors from the origin with their length measuring the magnitude of the contribution of that factor. There sick samples have a strong association with several taxa in this ordination. The methanotroph taxa Methylocystaceae and Methlyocystis both associate with the sick samples, however both occur in higher abundance in healthy samples (Appendix Table 1 and 2). For example, Methylocystaceae has only a single sequence recovered across all three sick samples. *Legionella*, a pathogenic genus of note, also groups with the sick samples though it is found in higher abundance in healthy samples.

Based on the distribution of samples in ordination space samples were organized into three groups for ISA and TWINSPAN analysis. PI-515yF and PII-515yF were clustered into the healthy 1 group. The healthy 2 group consisted of

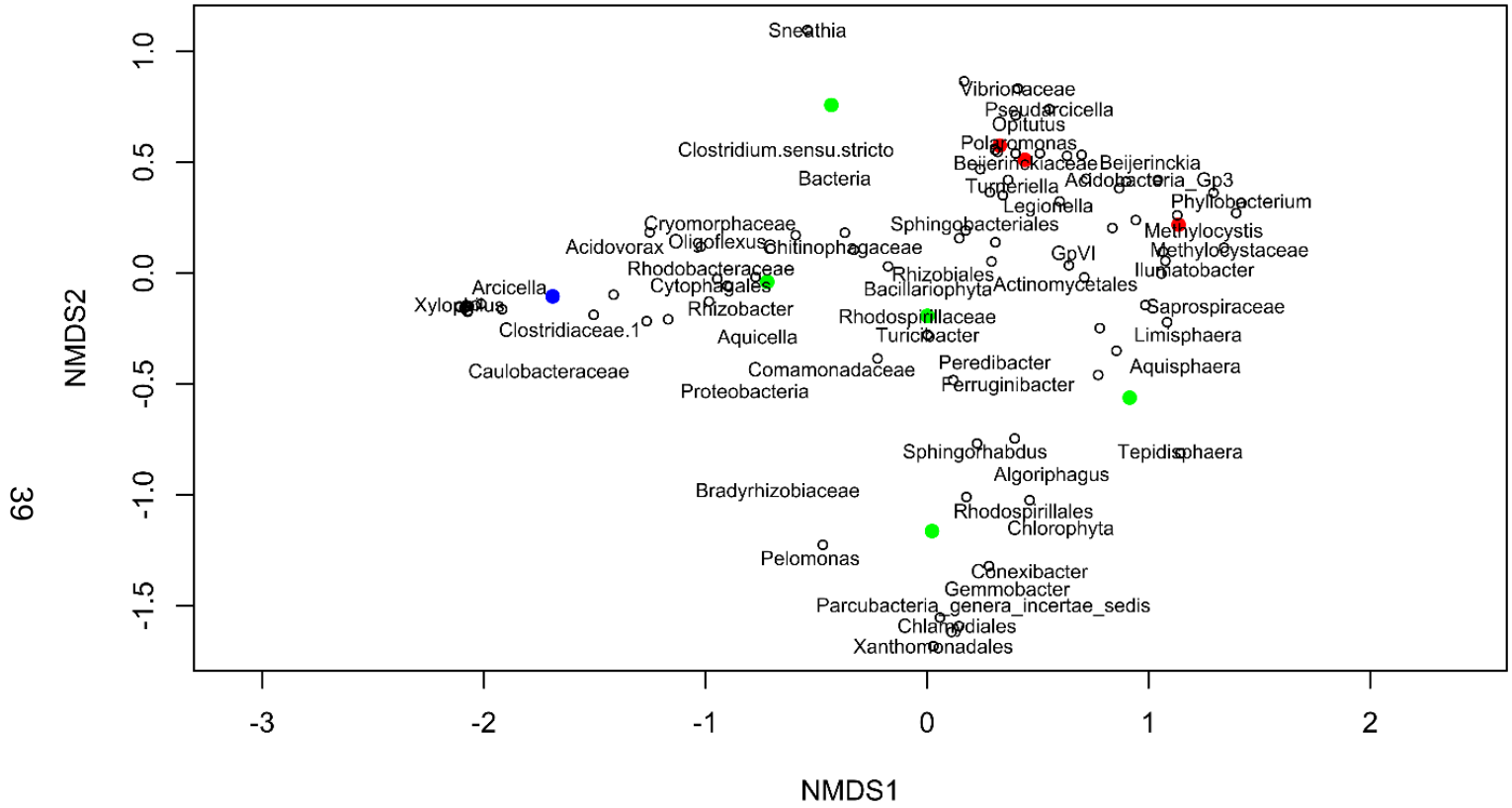

Figure 7. An NMDS using Bray-Curtis distance with taxonomic vector data overlaid. Sample names were removed for clarity but this uses the same underlying sample point structure as prior figures. Taxa points represent the end point of a vector starting from the origin.

PI-357wF, PII-357wF and Healthy-454. The sick group consisted of the PIII samples.

TWINSPAN analysis revealed that there is a conserved group of taxa across all disease states (Table 5). Among the healthy samples the V3-V4 and V4-V5 primers recovered distinct sets of taxonomic data. The greatest differences between samples within a group occurred within the healthy 1 group.

Table 5. An ordered two-way table generated by TWINSPAN. Samples are designated by number: 1 is PI-515yF, 2 is PI-357wF, 3 is PII-515yF, 4 is PII-357wF, 5 is PIII-515yF, 6 is PIII-357wF, 7 is PIII-454, and 8 is Healthy-454. These are combined into higher order groups sick, healthy 1 (H1) or healthy 2 (H2).

	Samples								Group
	Sick			H2			H1		
	6	7	5	8	4	2	3	1	
Cytophagaceae	1	1	-	-	-	-	-	-	00000
Fluviicola	1	-	-	1	-	-	-	-	00000
Heliimonas	-	1	-	1	-	-	-	-	00000
Legionellales	-	1	-	1	-	-	-	-	00000
Nitrosospira	1	1	-	1	-	-	-	-	00000
Pseudorhodoferax	-	1	-	1	-	-	-	-	00000
Verrucomicrobiaceae	1	1	-	1	-	-	-	-	00000
Acetobacteraceae	1	1	1	1	-	-	-	-	00001
Family II	-	1	-	1	-	1	-	-	00001
Flavobacterium	1	1	-	-	1	-	-	-	00001
Methylocystis	1	-	-	1	-	1	-	-	00001
Opitutus	1	1	1	-	-	-	-	-	00001
Phycisphaera	-	1	-	1	-	1	-	-	00001
Polaromonas	1	-	1	1	-	-	-	-	00001
Sphingomonadaceae	-	1	-	1	-	1	-	-	00001
Turneriella	1	1	-	1	-	-	1	-	00001
Deltaproteobacteria	1	1	1	1	-	1	-	-	00010
Limnohabitans	1	1	-	1	1	1	-	-	00010
Methylophilaceae	1	1	-	1	1	1	-	-	00010
Phycisphaerae	1	1	-	1	1	1	-	-	00010
Planctomycetes	1	1	-	1	1	1	-	-	00010
Saccharibacteria incertae sedis	1	1	-	1	1	1	-	-	00010
Acidobacteria Gp3	1	1	1	-	-	1	-	-	00011
Actinobacteria	1	1	1	1	1	1	-	-	00011
Alcaligenaceae	1	1	1	1	1	1	-	-	00011
Algoriphagus	1	-	-	-	1	-	-	-	00011
Candidatus Pelagibacter	1	1	1	1	1	1	-	-	00011
Chlorophyta	1	1	1	1	1	1	1	-	00011
Chloroplast	1	1	1	1	1	1	1	-	00011
Comamonadaceae	1	1	1	1	1	-	1	-	00011
Cyanobacteria	1	1	1	1	1	1	1	-	00011
Ferruginibacter	1	1	1	-	1	-	-	-	00011
Flavobacteriaceae	-	-	-	1	1	-	-	-	00011
Gammaproteobacteria	1	1	1	1	1	1	1	-	00011
Gemmata	1	1	1	1	-	1	1	-	00011
GpVI	1	1	1	-	1	-	-	-	00011
Legionella	1	1	1	1	1	1	1	-	00011
Methylocystaceae	1	-	-	-	-	1	-	-	00011
Methylotenera	-	1	-	1	1	1	-	-	00011
Myxococcales	-	1	-	-	1	-	-	-	00011
Oligoflexus	1	-	1	1	-	1	1	-	00011

Parachlamydiaceae	1	-	-	-	1	-	-	-	00011
Planctomycetia	-	-	1	1	-	-	-	-	00011
Polynucleobacter	1	1	1	1	1	1	1	-	00011
Rhodoferax	1	1	1	1	1	1	1	-	00011
Saprospiraceae	1	-	-	-	-	1	-	-	00011
Spartobacteria	1	1	1	1	1	1	1	-	00011
Spartobacteria incertae sedis	1	1	1	1	1	1	1	-	00011
Terrimicrobium	1	1	1	1	1	1	-	-	00011
Undibacterium	-	1	-	-	1	-	-	-	00011
Xanthomonadaceae	-	1	-	-	1	-	-	-	00011
Acidimicrobiaceae	1	-	-	-	1	1	-	-	0010
Acidimicrobiales	1	1	1	-	1	1	1	-	0010
Aquisphaera	1	-	-	-	-	1	1	-	0010
Beijerinckia	1	-	1	-	-	-	1	-	0010
Beijerinckiaceae	-	-	1	1	1	1	1	-	0010
Clostridiaceae 1	-	-	-	1	1	1	-	-	0010
Cryomorphaceae	-	-	1	1	1	-	-	-	0010
Flavobacteriales	-	-	-	1	1	1	-	-	0010
Ilumatobacter	1	-	1	-	1	1	-	-	0010
Limisphaera	1	-	-	-	1	1	-	-	0010
Opitutae	-	-	1	1	1	1	-	-	0010
Oxalobacteraceae	-	1	-	-	1	1	-	-	0010
Roseomonas	1	-	1	-	-	1	-	-	0010
Schlesneria	-	-	1	1	1	1	1	-	0010
Subdivision3	1	-	-	-	1	1	-	-	0010
Acidovorax	-	-	1	-	-	-	1	-	0011
Arenimonas	-	-	1	-	1	-	-	-	0011
Bradyrhizobiaceae	-	-	-	-	1	1	-	-	0011
Chlamydiales	-	-	-	-	1	1	-	-	0011
Clostridium sensu stricto	-	-	-	-	1	1	1	-	0011
Conexibacter	-	-	-	-	1	1	-	-	0011
Diplorickettsia	-	-	-	-	1	1	-	-	0011
Gemmobacter	-	-	-	-	1	1	-	-	0011
Lacibacterium	-	-	1	-	-	-	1	-	0011
Neochlamydia	-	-	-	-	1	1	-	-	0011
Parcubacteria incertae sedis	-	-	-	-	1	1	-	-	0011
Peredibacter	-	-	1	-	1	-	-	-	0011
Polymorphobacter	-	-	1	-	-	1	-	-	0011
Sphingorhabdus	-	-	1	-	1	-	-	-	0011
Actinomycetales	1	1	1	1	1	1	1	1	01
Alphaproteobacteria	1	1	1	1	1	1	1	1	01
Bacillariophyta	1	-	1	-	1	1	1	1	01
Bacteria	1	1	1	1	1	1	1	1	01
Betaproteobacteria	1	1	1	1	1	1	1	1	01
Chitinophagaceae	1	1	1	1	1	1	1	1	01
Cyanobacteria Chloroplast	1	1	1	1	1	1	1	1	01
GpIIa	1	1	1	1	1	1	1	1	01
Planctomycetaceae	1	1	1	1	1	1	1	1	01
Proteobacteria	1	1	1	1	1	1	1	1	01
Rhodospirillaceae	1	-	1	-	1	1	1	1	01
Sediminibacterium	1	1	1	1	1	1	1	1	01
Verrucomicrobia	1	1	1	1	1	1	1	1	01
Aquicella	-	-	-	-	1	-	-	1	100
Rhizobacter	-	-	-	-	1	-	-	1	100
Rhodospirillales	1	1	1	-	1	1	1	1	100
Subdivision 3 genera incertae sedis	1	1	1	-	1	1	-	1	100
Bacteriovorax	-	-	-	1	1	-	-	1	101
Enterobacteriaceae	-	1	-	-	-	-	-	1	1100
Propionibacterium	1	-	-	-	-	-	-	1	1100
Pseudomonadaceae	-	1	-	-	-	-	-	1	1100
Bacteroidetes	1	1	1	1	1	1	-	1	1101
Burkholderiales	1	1	1	1	1	1	-	1	1101
Luteolibacter	1	1	-	1	1	1	1	1	1101
Rhizobiales	1	1	-	1	-	1	1	1	1101
Rhodobacteraceae	1	1	-	1	1	1	1	1	1101
Sphingobacteriales	1	1	-	1	-	-	-	1	111

ISA was performed to obtain statistical information regarding which taxa best represented the three groups found from the environmental fitting (Table 6 all ISA data is reported in the Appendix Table 3). The ISA analysis used returned three primary values: the p value for statistical significance; A, which measures the percentage of times a taxonomic result occurs within a group, and B, which measures the percentage of times that taxonomic result occurs in the samples that made up that group. Higher order groups were also included: healthy and sick, sick and rescue, and rescue and healthy. All indicator species were present in all samples that made up their respective groups, B = 1. Two

Table 6. Summary of the significant indicator species from all data sets. Sick is a combined group of all PIII samples. Rescue is the rescue data set. A is the proportion of times a taxonomic group occurred within that group. B is the frequency the taxonomic group occurred in the samples that make up that group.

Healthy 2				
	A	B	Stat	P value
Clostridiaceae 1	1	1	1	0.0396
Flavobacteriales	1	1	1	0.0396
Sick				
	A	B	Stat	P value
Opitutus	1	1	1	0.0344
Acidobacteria Gp3	0.9778	1	0.989	0.023
Acetobacteraceae	0.84	1	0.916	0.0428
Sick and Healthy 2				
	A	B	Stat	P value
Actinobacteria	1	1	1	0.0368
Alcaligenaceae	1	1	1	0.0368
Candidatus pelagibacter	1	1	1	0.0368
Terrimicrobium	1	1	1	0.0368

taxa within the healthy 1 group were significant with $p < 0.05$: Clostridiaceae 1, Flavobacteriales with both found exclusively within this group, A = 1. The sick group had three taxa with $p < 0.05$: Opitutus, Acidobacteria Gp3 and Acetobacteraceae. Of these taxonomic groups Opitutus occurred exclusively in the sick samples, A = 1, and Acidobacteria Gp3 and Aceotobacteraceae were found with low abundance in the healthy samples, with A values of 0.9778 and .84 respectively. The last statistically significant indicator taxa were present in both the sick and healthy 2: Actinobacteria, Alcaligenaceae, Candidatus Pelagibacter and Terrimucrobium. These taxa all had equal P values of 0.0368 and occurrences of these taxa were within these two groups, A = 1 and B = 1.

DISCUSSION

The evaluation of the primers and the bacterial microbiome profiles recovered suggested there was a complex interaction between sample and primer set used for sequencing (Figure 1). The initial assumption, based on the clustering pattern of three groups of two samples, is that the sample was the primary driving force for similarity between profiles. We expected to see smaller differences in data when using the V3/V4 and V4/V5 primers on the same sample and larger differences when looking across samples as these primers generate overlapping data. The initial data showed that PI and PII samples showed the highest similarity by primer while the PIII group showed the highest similarity based on sample origin. The PIII group was also more similar to the PI and PII samples sequenced with the 515yF primer than the PI and PII samples sequenced with the 357wF primer. This means that the healthy samples contained a more diverse microbiome and that each primer set, V3-V4 and V4-V5, recovered a distinct range of taxonomic data. The sick samples had a much less diverse microbiome that was recovered equally with both primer sets.

Examining the species list, there are a large number of sequences that identified only to the bacterial domain for the 515yF primer set in the healthy samples. This is unlikely to be a function of the primer itself as the PIII-515yF data set has a two log reduction in the number of sequences identified to the

bacterial domain. This could mean the domain level classifications represent organisms not present in the RDP 16S database. This database does contain data from global sources, but Lake Baikal may be poorly represented in that database. More accurate taxonomic classification for the unclassified bacteria should be obtainable if longer sequence reads could be generated (Shin et al., 2016). Future exploration of these unclassified taxa may be an important avenue of further research since this occurs in high abundance within healthy organisms but at much lower abundance in sick samples.

The general lack of recoverable eukaryotic data was disappointing, but not unexpected. The DNA samples were generated from harvested sponge tissue, the bulk of which would be sponge cells. This high amount of sponge DNA would then mask non-sponge eukaryotic rRNA sequences. Our sequencing project only generated 100,000 sequences per reaction. That is also the most likely explanation for why only bacterial data was recovered by the 515yF/928pfR primer set. If more sequences were obtained in the reaction then non-bacterial sequences could be recovered, de-multiplexed and analyzed using other databases. Other eukaryotic primers were not tested due to constraints of DNA samples sizes.

The V4/V5 primer set does appear to recover fewer unique taxa compared to the V3/V4 primer (Figure 2). There were too few samples to perform meaningful statistical analysis on the species richness data. However, looking at

the species data there were a number of instances where rare taxa, fewer than five sequences recovered, appeared in only one sample and account for a large portion of species richness. So the differences in species richness may be simply if a taxonomic group is rare enough it is either missed in sequencing or could not be present in the local area where the sample was collected. Samples with lower species richness did not cluster together with multivariate analysis, so this is unlikely to be a significant factor. The disparity in richness between samples can be addressed with access to water samples collected from sample sites. Microorganisms can be filtered from the water samples, profiled, and compared against the existing sponge data to see if the species richness reflectes the local environment. An alternate explanation is a product of primer choice (Rintala et al., 2017). The two primer sets used both amply V4. If the bulk of the sequencing data from endemic Lake Baikal bacteria was generated with V3-V4 primers, then the V4-V5 sequences recovered would contain a variable region not in any of the reference databases resulting in poor alignments at low taxonomic levels. Either explanation underscores the need for expanded sampling and use of newer sequencing platforms that enable longer reads.

Multivariate analysis was conducted with the core nine samples, omitting the samples PI and PII re-sequenced with 357wF due to their exceptional degree of similarity. Multiple ordination techniques were used to visualize the data because, while there were certain objective metrics that needed to be satisfied,

e.g.low stress in NMDS, the resulting ordination still needs to be explainable. NMDS using Raup-Crick distribution yielded an exceptionally low stress, but the resulting ordination shows the rescue sample is so unrelated to the remaining samples that there were essentially no differences between them.

The PCA analysis showed how the large differences between the rescue sample and the other samples may have obscured the relationships between the non-rescue samples. One of the sick samples, PIII-357wF, was clearly distinct, yet the other sick samples show high similarity to healthy samples (Figure 3). So other ordination techniques need to be examined. NMDS is the current most favored for ordination techniques and was used next. There are multiple different methodologies available to calculate the differences between samples when constructing the ordination (Figure 4). This is a small data set, so a range of methodologies were tested. The different methodologies have different applications with some being appropriate for ecological data sets where species exist as a gradient. For example, the Raup-Crick method looks at the probability of a taxa being present or absent in a data set and it was unexpected that it yielded one of the worst distributions. The ordinations produced using Manhattan and Bray-Curtis dissimilarity functions were identical when overlaid on top of each other (Figure 4c). After selecting the Bray-Curtis method for generating the NMDS, a two-dimensional solution was compared to a three-dimensional one. Both generated low stress solutions, but the two-dimensional

solution had the advantage of readability over the three-dimensional when species plots were superimposed on the ordinations. The inclusion of the species information (not shown) resulted in a plot that was too cluttered to be of use.

With an ordination that shows an ordered structure to the data, the descriptive variables were fit to the ordination and tested statistically to see if they explain the observed grouping (Figure 5). The grouping of samples in ordination space cannot be explained by factors such as the organism the DNA was collected from, the variable regions used for sequencing, the sequencing platform used to generate data or the timing of samples collected before or after the disease outbreak occurred. These results have several important conclusions. It is appropriate in this project to compare data from different sequencers. A concern arose looking at the dissimilarity dendrograms, and there was not a profound change in the microbiome composition in the sponges following disease outbreak.

The best explanation for the grouping of samples is the health of the organism the samples were collected from (Figure 6). This is an expected conclusion as healthy sponges would be functioning as filter feeders and would have a microbiome population that most likely reflects the microbiome of Lake Baikal. The sick sponges show signs of disease by color change and have invasive organism(s) that overcame their immune system and opportunistic

organisms that colonized compromised tissue. Of note, the distribution of sick samples was much smaller, meaning greater similarity, than the wide distribution of healthy samples. But there was only one sick organism sampled; and many more samples, both sick and healthy, would need to be sequenced to see if this pattern holds true.

TWINSPAN revealed a core group of taxa that probably contains the core microbiome reported to be transmitted vertically among conspecific sponges (Reveillaud et al., 2014). The presence of these shared taxa in samples of different disease states suggests that they are unrelated to the disease outbreak. Analysis of samples collected prior to 2010 would allow for further clarification of the core microbiome of *L. baicalensis*.

Several taxa were found to be significant by ISA. These taxonomic units must be considered as independent to each other as they represent different DNA sequences that are aligned to a reference database. Thus a phylum level taxa would represent a distinct organism from a sequence that identified to the family level within that phylum and not be considered together as a group.

The significant species in the healthy 2 sample include Flavobacteriales, which is an order of bacteria that contains members that are pathogenic to humans, fish or other aquatic life, while others are non-pathogenic. Clostridiaceae 1 is an unclassified family group within the Clostridiales order and also encompasses organisms of very different pathogenicities and environmental

niches. Organisms within class Clostridiales are associated with coral mucus but little is known as to the nature of that relationship (McKew et al., 2012). Given that these are significant to one of the healthy groups these are most likely unclassified bacteria that are not pathogenic.

The significant taxa for the sick samples are more interesting. The presence of Opitutus as an indicator species is surprising as this is a genus of Verrucomicrobia found in rice paddy soil. It is an obligate anerobe that produces propionate as a fermentation byproduct but very little else is known about this organism (Chin et al., 2002). Acetobacteraceae is a family of oxidative fermenters that can tolerate low acid environments but grow optimally around pH 5-6. They ferment sugars and ethanol and produce acetic acid. Acidobacteria Gp3 is a subdivision of the phylum Acidobacteria which also consists of acidophiles. Without species level identification, these are still broad enough groups that more definite conclusions cannot be made. This suggests that there may be a correlation between pH and disease that should be explored in future work.

The combined healthy 2 and sick group had several significant taxa. Acidobacteria, which is a phylum level taxonomic group at a higher order than Acidobacteria Gp3, also includes acidophiles. The healthy 2 group contains samples collected from sponges both prior to and after the disease outbreak which supports the need to collect pH data both in sponge tissues and in the

surrounding water. Alcaligenaceae is a family of bacteria found in all non-extreme enviroments some of which are known to be pathogenic. Terrimicrobium, another Verrucomicrobia, is another poorly characterized bacteria that was originally isolated in rice paddies. *Pelagibacter unqiue*, is the most abundant marine and freshwater bacterium on Earth (Morris et al., 2002). Its inclusion as an indicator species is most likely due to its absence in the V4-V5 data set.

An interesting result from both the species data and the ISA is that two of the main groups of methanophile; Methylococcaceae and Methylocystaceae, are not indicator species nor found in high numbers in any samples. One of the hypotheses proposed for the cause of the sponge disease outbreak is an increase in methane composition in Lake Baikal (Denikina et al., 2016). This is unlikely as there would be an expected increase in abundance of methanophiles in sick samples. However, this cannot be ruled out completely because one mechanism of sponge disease is that disease-causing organisms can suppress sponge immunity to create an ecological niche for themselves. This immune suppression allows other microorganisms to invade the sponge where they could outcompete the disease-causing agents. Thus, there could be a situation where the red appearance of disease is a byproduct of disease and the actual causative organisms are no longer present by the time sponge becomes symptomatic.

This pilot study has demonstrated for the first time that sick sponges have a distinct community of bacteria that is separate from healthy sponges. The identification of two poorly classified acidophilic organisms as indicator species suggest that low pH may correlate to disease state. This hypothesis needs to be further explored by expanding the number of samples collected and to collect data on local water conditions and pH of sponge tissue. Also, classifying the microbiome profile of the surrounding water to see if the sponge microbiome is similar to its environment or if it is distinct would give additional understanding as to what is happening during this course of disease.

LITERATURE CITED

Angermeier, H., V. Glöckner, J.R. Pawlik, N.L. Lindquist, and U. Hentschel. 2012. Sponge white patch disease affecting the Caribbean sponge Amphimedoncompressa. Diseases of Aquatic Organisms *99*, 95-102.

Bell, J.J. 2008. The functional roles of marine sponges. Estuarine, Coastal and Shelf Science *79*, 341-353.

Blanquer, A., M.J. Uriz, E. Cebrian, and P.E. Galand. 2016. Snapshot of a Bacterial Microbiome Shift during the Early Symptoms of a Massive Sponge Die-Off in the Western Mediterranean. Frontiers in Microbiology *7*, 752.

Bradley, J.P., M.C. Charles, J.J. Frank, and W.F. James. 2006. Potential role of sponge communities in controlling phytoplankton blooms in Florida Bay. Marine Ecology Progress Series *328*, 93-103.

Cebrian, E., M.J. Uriz, J. Garrabou, and E. Ballesteros. 2011. Sponge mass

mortalities in a warming Mediterranean Sea: are cyanobacteria-harboring species worse off? PLoS One *6*, e20211.

Chin, K.-J., and P.H. Janssen. 2002. Propionate Formation by *Opitutus terrae* in Pure Culture and in Mixed Culture with a Hydrogenotrophic Methanogen and Implications for Carbon Fluxes in Anoxic Rice Paddy Soil. *Applied and Environmental Microbiology*, *68*(4), 2089–2092.

Dabaeva, D. B., B.Z. Tsydypov, A.A Ayurzhanaev, S.G. Andreev, and Y.Z. Garmaev. 2016 Peculiarities of Lake Baikal water level regime IOP Conf. Ser.: Earth Environ. Sci. 48 012014

De Corte, D., E. Sintes, C. Winter, T. Yokokawa, T. Reinthaler, and G.J. Herndl. 2010. Links between viral and prokaryotic communities throughout the water column in the (sub)tropical Atlantic Ocean. ISME J *4*, 1431-1442.

Denikina, N.N., E.V. Dzyuba, N.L. Bel'kova, I.V. Khanaev, S.I. Feranchuk, M.M. Makarov, N.G. Granin, and S.I. Belikov. 2016. The first case of disease of the sponge *Lubomirskia baicalensis*: Investigation of its microbiome. Biology Bulletin *43*, 263-270.

Emerson, D., L. Agulto, L. Henry, L. Liping. 2008. Identifying and Characterizing Bacteria in an Era of Genomics and Proteomics. BioScience *58*(10), 925-936.

Gallo, R.C., and L. Montagnier. 2003. The Discovery of HIV as the Cause of AIDS. New England Journal of Medicine *349*, 2283-2285.

Gao, Z.-M., Y. Wang, R.-M. Tian, O.O. Lee, Y.H Wong, Z.B. Batang, A. Al-Suwailem, F.F. Lafi, V.B. Bajic, and P.-Y. Qian. 2015. Pyrosequencing revealed shifts of prokaryotic communities between healthy and disease-like tissues of the Red Sea sponge Crellacyathophora. PeerJ *3*, e890.

Gochfeld, D.J., C.G. Easson., C.J. Freeman, R.W. Thacker, and J.B. Olson. 2012. Disease and nutrient enrichment as potential stressors on the Caribbean sponge Aplysinacauliformis and its bacterial symbionts. Marine Ecology Progress Series *456*, 101-111.

Granin, N.G., S.I. Muyakshin, M.M. Makarov, K.M. Kucher, I.y.A.Aslamov, L.Z. Granina, and I.B. Mizandrontsev. 2012. Estimation of methane fluxes from bottom sediments of Lake Baikal. Geo-Marine Letters *32*, 427-436.

Hampton, S.E., L.R. Izmest'eva, M.V. Moore, S.L. Katz, B. Dennis, and E.A. Silow. 2008. Sixty years of environmental change in the world's largest freshwater lake – Lake Baikal, Siberia. *Global Change Biology*, *14*(8), 1947–1958.

Hampton, S.E., D.K. Gray, L.R. Izmest'eva, M.V. Moore, and T. Ozersky. 2014. The Rise and Fall of Plankton: Long-Term Changes in the Vertical Distribution of Algae and Grazers in Lake Baikal, Siberia. *PLoS ONE*, *9*(2), e88920.

Hebert, P. D. N., S. Ratnasingham, and J.R. deWaard. 2003. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society B: Biological Sciences, 270(Suppl 1), S96–S99.

James, M.C., W.-C. Kathryn, W.P. Shawn, G. Thomas, W.S. and Garriet. 2006. Identification of bacteria associated with a disease affecting the marine sponge anthellabasta in New Britain, Papua New Guinea. Marine Ecology Progress Series *324*, 139-150.

Kapitanov, V.A., I.S. Tyryshkin, N.P. Krivolutskii, Y.N. Ponomarev, M. De Batist, , and R.Y. Gnatovsky. 2007. Spatial distribution of methane over Lake Baikal surface. SpectrochimActa A MolBiomolSpectrosc *66*, 788-795.

Kasimov, N., D. Karthe, and S. Chalov. 2017. Environmental change in the Selenga River—Lake Baikal Basin. Reg Environ Change 17: 1945.

Kirillov, S., N. Sedova, E. Vorobyevskaya. T. Zengina. 2014. Problems and prospects for tourism development in the Baikal Region, Russia. 14th GeoConference on ecology, economics, education and legislation, At Bulgaria, Volume: 2

Kravtsova, L.S., L.A. Izhboldina, I.V. Khanaev, G.V. Pomazkina, E.V. Rodionova, , V.M. Domysheva, M.V. Sakirko, I.V. Tomberg, T.Y. Kostornova, O.S. Kravchenko, and A.B. Kupchinsky. 2014. Nearshore benthic blooms of filamentous green algae in Lake Baikal. Journal of Great Lakes Research *40*, 441-448.

Luter, H.M., S. Whalan, and N.S. Webster. 2010. Exploring the Role of Microorganisms in the Disease-Like Syndrome Affecting the Sponge *Ianthellabasta*. Applied and Enviromental Microbiology *76*, 5736-5744.

Matar, C.G., N.R. Anthony, B.M. O'Flaherty, N.T. Jacobs, L. Priyamvada, C.R. Engwerda, S.H. Speck, and T.J. Lamb. 2015. Gammaherpesvirus Co-infection with Malaria Suppresses Anti-parasitic Humoral Immunity. PLoSPathog *11*, e1004858.

McKew, B. A., A.J. Dumbrell, S.D. Daud, L. Hepburn, E. Thorpe, L. Mogensen, and C. Whitby. 2012. Characterization of Geographically Distinct Bacterial Communities Associated with Coral Mucus Produced by Acropora spp. and Porites spp. *Applied and Environmental Microbiology*, *78*(15), 5229–5237.

Milanese, M., E. Chelossi, R. Manconi, A. Sarà, M. Sidri, and R. Pronzato. 2003. The marine sponge Chondrillanucula Schmidt, 1862 as an elective candidate for bioremediation in integrated aquaculture. Biomolecular Engineering *20*, 363-368.

Morris R.M., M.S. Rappé, S.A. Cannon, K.L. Vergin, W.A. Siebold, C.A. Carlson, and S.J. Giovannoni. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature *420*, 806–810

Müller, W.E.G., S.I. Belikov, O.V. Kaluzhnaya, S. Perović-Ottstadt, E. Fattorusso, H. Ushijima, A. Krasko, and H.C. Schröder. 2007. Cold stress defense in the freshwater sponge Lubomirskia baicalensis. FEBS Journal *274*, 23-36.

Nicole, S.W., P.N. Andrew, I.W. Richard, and T.H. Russell. 2002. A spongin-boring α-proteobacterium is the etiological agent of disease in the Great Barrier Reef sponge Rhopaloeidesodorabile. Marine Ecology Progress Series *232*, 305-309.

Nguyen, M.T.H.D., M. Liu, and T. Thomas. 2014. Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Molecular Ecology *23*, 1635-1645.

Oulas, A., C. Pavloudi, P. Polymenakou, G.A. Pavlopoulos, N. Papanikolaou, G. Kotoulas, C. Arvanitidis, and I. Iliopoulos. 2015. Metagenomics: Tools and Insights for Analyzing Next-Generation Sequencing Data Derived from Biodiversity Studies. Bioinformatics and Biology Insights *9*, 75-88.

Paliy, O., and V. Shankar. 2016. Application of multivariate statistical techniques in microbial ecology. Molecular Ecology, *25*(5), 1032–1057.

Parfenova, V.V., I.A. Terkina, T.Y. Kostornova, I.G. Nikulina, V.I. Chernykh, and E.A. Maksimova. 2008. Microbial community of freshwater sponges in Lake Baikal. Biology Bulletin *35*, 374-379.

Patterson, M.R., V.I. Chernykh, V.A. Fialkov, and M. Savarese. 1997. Trophic effects of sponge feeding within Lake Baikal's littoral zone. 1. Insitu pumping rates. Limnology and Oceanography *42*, 171-178.

Pettengill, E. A., J.B. Pettengill, and R. Binet. 2015. Phylogenetic Analyses of Shigella and Enteroinvasive Escherichia coli for the Identification of Molecular Epidemiological Markers: Whole-Genome Comparative Analysis Does Not Support Distinct Genera Designation. Frontiers in Microbiology, 6, 1573.

Ramette, A. 2007. Multivariate analyses in microbial ecology. Fems Microbiology Ecology *62*, 142-160.

Reveillaud, J., L. Maignien, M.A. Eren, J.A. Huber, A. Apprill, M.L. Sogin and A. Vanreusel. 2014. Host-specificity among abundant and rare taxa in the sponge microbiome. The ISME Journal, *8*(6), 1198–1209.

Rintala, A., S. Pietilä, E. Munukka, E. Eerola, J.-P. Pursiheimo, A. Laiho, S. Pekkala, and P. Huovinen. 2017. Gut Microbiota Analysis Results Are Highly Dependent on the 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor. Journal of Biomolecular Techniques *28*(1), 19–30.

Schaechter, M., O. MaalØe, and N.O. Kjeldgaard. 1958. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. Microbiology *19*, 592.

Schmid, M., M.D. Batist, N.G. Granin, V.A. Kapitanov, D.F. McGinnis, I.B. Mizandrontsev, A.I. Obzhirov, and A. Wüest. 2007. Sources and sinks of methane in Lake Baikal: A synthesis of measurements and modeling. Limnology and Oceanography *52*, 1824-1837.

Shin, J., S. Lee, M.-J. Go, S.Y. Lee, S.C. Kim, C.-H. Lee, and & B.-K. Cho. 2016. Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. Scientific Reports, *6*, 29681.

Staff, T. (2014). Paper Mill That Polluted Baikal to Become Russian Disneyland. The Moscow Times.

Sweet, M., M. Bulling, and C. Cerrano. 2015. A novel sponge disease caused by a consortium of microorganisms. Coral Reefs *34*, 871-883.

Wang, Q., G.M. Garrity, J.M. Tiedje, and J.R. Cole. 2007. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy . Applied and Environmental Microbiology, *73*(16), 5261–5267.

Wang, Y., R.M. Tian, Z.M. Gao, S. Bougouffa, and P.-Y. Qian. 2014. Optimal Eukaryotic 18S and Universal 16S/18S Ribosomal RNA Primers and Their Application in a Study of Symbiosis. PLoS ONE *9*, e90053.

Walter, D. E. and S. Winterton. 2007. Keys and the Crisis in Taxonomy: Extinction or Reinvention? Annual Review of Entomology *52*:1, 193-208

Woese, C. R., O. Kandler, and M.L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences of the United States of America, *87*(12), 4576–4579.

Yang, B., Y. Wang, and P.-Y. Qian. 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics, *17*, 135.

Zaneveld, J.R. 2014. Chronic nutrient enrichment increases prevalence and severity of coral disease and bleaching. Global Change Biology *20*, 544-554.

Zakharenko, A.S., N.V. Pimenov, V.G. Ivanova, and T.I. Zemskaya. 2015. Detection of methane in the water column at gas and oil seep sites in central and southern Lake Baikal. Microbiology *84*, 90-97.

Zhang, X., W. Zhang, L. Xue, , B. Zhang, M. Jin, and W. Fu. 2010. Bioremediation of bacteria pollution using the marine sponge Hymeniacidonperlevis in the intensive mariculture water system of turbot Scophthalmus maximus. Biotechnology and Bioengineering *105*, 59-68.

APPENDIX

Table 1

A species list of all recovered bacterial taxa for healthy and untreated sponge samples. Samples PI and PII were sequenced with the Illumina platform using either the 515yF V4/V5 primer set or the 357wF V3/V4 primer set. Both the PI and PII 357wF samples were resequenced (r). The Healthy 454 sample was sequenced in the Roche 454 platform. Values represent the number of sequences recovered that identify with that taxonomic group with a confidence >80%.

	PI 515yF	PI 357wF	PI r 357wF	PII 515yF	PII 357wF	PII r 357wF	Healthy 454
Acetobacteraceae	0	0	0	0	0	0	22
Acidimicrobiaceae	0	2	7	0	6	0	0
Acidimicrobiales	0	137	259	34	119	0	0
Aciditerrimonas	0	0	0	0	2	0	0
Acidobacteria Gp3	0	0	0	0	4	0	0
Acidovorax	0	0	644	10	0	11	0
Actinobacteria	0	688	683	0	223	0	180
Actinomycetales	119	860	0	154	189	0	120
Afipia	0	0	0	0	0	86	0
Alcaligenaceae	0	14	4698	0	12	1023	5
Algoriphagus	0	27	1	0	0	0	0
Alkanindiges	0	0	93	0	0	0	0
Alphaproteobacteria	83	5225	0	34	2516	431	3333
Anaerolineaceae	0	0	7	0	0	0	5
Aquicella	83	14	3	0	0	0	0
Aquisphaera	0	0	0	10	199	0	0
Archaea	0	0	0	0	0	0	5
Arenimonas	0	58	9	0	0	0	0
Armatimonadetes	0	0	0	0	0	0	5
Armatimonadetes gp5	0	0	145	0	0	22	0
Armatimonas Armatimonadetes gp1	892	0	0	0	0	0	0
Arthrobacter	239	0	3145	0	0	2672	0
Bacillales	18	0	325	0	0	0	0
Bacillariophyta	404	219	0	337	298	22	0
Bacilli	9	0	105	0	0	11	0
Bacteria	48888	2145	28	76677	3729	0	32137
Bacteriovorax	18	229	0	0	0	0	11
Bacteroidales	0	0	1	0	0	0	0
Bacteroides	0	0	498	0	0	43	5
Bacteroidetes	625	208	0	0	2	0	55
Bangiophyceae	0	0	3	0	0	0	5

Bdellovibrio	0	0	2	0	0	0	0
Beijerinckia	0	0	16	19	0	0	0
Beijerinckiaceae	0	3	875	5	4	237	5
Betaproteobacteria	1039	322	0	284	82	0	191
Blastopirellula	0	0	62	0	0	0	0
Bradyrhizobiaceae	0	17	0	0	4	0	0
Brevifollis	0	0	1732	0	0	603	0
Bryocella	0	0	67	0	0	0	0
Burkholderiales	3548	614	1	0	121	0	38
Burkholderialesincertaesedis	0	0	39687	0	0	58597	0
CandidatusPelagibacter	0	71	5899	0	12	5559	82
Catenococcus	0	0	40	0	0	0	5
Caulobacteraceae	0	2	0	0	0	0	0
Chitinophagaceae	13097	1139	1425	2227	479	32	3230
Chlamydiales	0	19	17	0	2	0	0
Chlorarachniophyceae	0	0	18	0	0	0	16
Chlorophyta	0	45187	0	241	54201	0	4948
Chloroplast	0	4895	24	87	4264	0	34406
Chryseobacterium	0	0	46	0	0	0	0
Chryseolinea	0	9	13	0	0	0	0
Clostridiaceae 1	0	2	11	0	2	22	5
Clostridiales	0	0	33	0	0	140	5
Clostridium sensustricto	0	5	864	34	2	1411	0
Cnuella	0	0	76	0	0	0	0
Comamonadaceae	0	1471	39	48	0	0	22
Conexibacter	0	16	9	0	10	0	0
Corynebacterium	0	0	34	0	16	0	0
Coxiellaceae	0	2	0	0	0	0	0
Cryomorphaceae	0	5	0	0	0	0	5
Cryptomonadaceae	0	13	0	0	0	32	0
Curvibacter	0	5	0	0	0	0	0
Cyanobacteria	0	19	2	48	14	108	464
Cyanobacteria Chloroplast	1443	499	20	1571	569	0	2057
Cytophagaceae	0	0	1	0	0	0	0
Cytophagales	101	0	823	0	0	0	0
Deltaproteobacteria	0	0	30	0	6	0	27
Dermacoccaceae	0	0	4	0	0	0	0
Diplorickettsia	0	35	192	0	4	786	0
Dorea	0	0	2	0	0	0	0
Duganella	0	0	10	0	0	0	0
Emticicia	0	0	5	0	0	0	0
Enterobacteriaceae	138	0	803	0	0	7972	0
Erysipelotrichaceae	0	0	9	0	0	0	0
Escherichia.Shigella	18	0	0	0	0	0	0
Family.II	0	0	1	0	2	0	5
Ferruginibacter	0	139	6	0	0	0	0

Firmicutes	0	0	141	0	0	0	0
Flavobacteriaceae	0	2	0	0	0	0	11
Flavobacteriales	0	11	0	0	2	0	5
Flavobacterium	0	398	0	0	0	0	0
Fluviicola	0	0	1	0	0	11	11
Frigidibacter	0	0	5	0	0	0	0
Gallionellaceae	0	0	15	0	0	0	5
Gammaproteobacteria	0	268	4	323	227	32	87
Gemmata	0	0	71	24	8	0	213
Gemmobacter	0	9	5	0	4	0	0
Gimesia	0	0	1957	0	0	0	16
GpIIa	221	871	24	7924	11867	0	7234
GpVI	0	27	0	0	0	0	0
Heliimonas	0	0	7	0	0	0	191
Herminiimonas	0	0	353	0	0	0	0
Humitalea	0	0	133	0	0	43	5
Hyphomicrobium	0	6	0	0	0	0	0
Ilumatobacter	0	39	21	0	137	0	0
Intrasporangiaceae	0	0	13	0	0	0	0
Janthinobacterium	0	0	47	0	0	0	0
Labrys	0	11	0	0	0	0	0
Lachnospiraceae	0	0	35	0	0	0	0
Lacibacterium	0	0	0	5	0	0	0
Lacihabitans	0	0	3	0	0	0	0
Lactococcus	0	0	0	0	12	43	0
Legionella	0	13	30	48	48	0	38
Legionellales	0	0	5	0	0	0	16
Limisphaera	0	5	15	0	16	0	0
Limnohabitans	0	172	10	0	48	0	71
Luteolibacter	18	1436	0	10	50	0	16
Methylococcaceae	0	0	10	0	0	0	0
Methylocystaceae	0	0	0	0	2	129	0
Methylocystis	0	0	0	0	12	0	22
Methylophilaceae	0	401	1	0	2	0	5
Methylosoma	0	0	34	0	0	172	5
Methylotenera	0	144	1	0	16	0	22
Methylovirgula	0	3	0	0	0	0	0
Micrococcaceae	377	0	0	0	0	0	0
Micrococcus	0	17	345	0	0	291	0
Microvirgula	0	0	160	0	0	237	5
Myxococcales	0	3	9	0	0	0	0
Namhaeicola	0	0	0	0	0	22	0
Neochlamydia	0	25	52	0	2	22	0
Nevskia	0	0	1	0	0	0	0
Nitrosomonadaceae	0	0	0	0	0	32	0
Nitrosospira	0	0	4	0	0	0	5
Nitrospira	0	68	4564	0	0	280	0
Nocardia	0	0	5	0	4	0	0

Oligoflexus	0	0	1	48	18	0	22
Opitutae	0	8	4	0	2	0	5
Opitutus	0	0	0	0	0	22	0
Oxalobacteraceae	0	33	7	0	4	0	0
Paludibaculum	0	0	16	0	0	0	0
Parachlamydiaceae	0	38	11	0	0	0	0
Parasegetibacter	0	0	126	0	0	0	0
Parcubacteria genera incertaesedis	0	8	0	0	2	0	0
Parvibaculum	0	0	67	0	0	0	0
Paucibacter	0	0	27	0	0	0	5
Pedobacter	0	0	1903	0	0	1013	0
Pedomicrobium	0	0	4002	0	0	452	0
Pelomonas	0	77	12	0	0	0	0
Peredibacter	0	8	1	0	0	0	0
Perlucidibaca	0	0	0	0	0	22	0
Phycisphaera	0	0	2	0	2	0	436
Phycisphaerae	0	39	0	0	153	0	147
Phyllobacterium	0	0	1	0	0	0	0
Pirellula	0	0	18	10	0	776	0
Planctomycetaceae	938	312	0	2015	565	0	1604
Planctomycetes	0	133	2	0	121	0	44
Planctomycetia	0	0	59	0	0	0	5
Polaromonas	0	0	19565	0	0	1875	5
Polymorphobacter	0	0	2	0	18	0	0
Polynucleobacter	0	79	0	164	52	0	142
Propionibacterium	37	0	3	0	0	0	0
Propionigenium	0	0	3	0	0	399	0
Proteobacteria	2858	4092	78	217	103	2252	742
Proteus	0	0	1	0	0	0	0
Pseudomonadaceae	18	0	34	0	0	0	0
Pseudomonadales	0	0	3	0	0	0	0
Pseudomonas	175	0	24	0	0	0	0
Pseudorhodoferax	0	0	1	0	0	0	5
Rhizobacter	257	17	0	0	0	0	0
Rhizobiaceae	0	0	3	0	0	0	0
Rhizobiales	55	0	0	5	14	0	507
Rhodobacter	0	0	2338	0	2	11129	0
Rhodobacteraceae	28	14	16	48	20	0	5
Rhodoferax	0	2	20	19	6	0	5
Rhodospirillaceae	2050	1832	0	766	579	0	0
Rhodospirillales	110	3967	11	72	1138	0	0
Roseiarcus	0	0	0	0	0	0	5
Roseomonas	0	0	0	0	4	0	0
Saccharibacteria genera incertaesedis	0	104	0	0	847	0	278
Saprospiraceae	0	0	0	0	12	0	0
Schlesneria	0	5	0	130	16	0	235

Sediminibacterium	20882	18460	0	5528	1921	0	5
Sneathia	0	0	0	24	0	0	0
Solirubrobacterales	0	2	0	0	0	0	0
Spartobacteria	0	2	0	116	195	0	202
Spartobacteria genera incertaesedis	0	69	0	617	2985	0	595
Sphingobacteriales	9	0	0	0	0	0	5
Sphingomonadaceae	0	0	0	0	2	0	22
Sphingorhabdus	0	63	0	0	0	0	0
Staphylococcus	836	0	0	0	0	0	0
Streptophyta	239	0	0	0	0	0	0
Subdivision3	0	2	0	0	2	0	0
Subdivision3 genera incertaesedis	110	21	0	0	6	0	0
Tabrizicola	0	2	0	0	0	0	0
Tepidisphaera	0	0	0	0	2	0	0
Terrimicrobium	0	47	0	0	18	0	38
Turicibacter	0	0	0	0	0	0	11
Turneriella	0	0	0	19	0	0	22
Undibacterium	0	2	0	0	0	0	0
Verrucomicrobia	18	2312	0	48	11638	0	5493
Verrucomicrobiaceae	0	0	0	0	0	0	16
Xanthomonadaceae	0	5	0	0	0	0	0
Xanthomonadales	0	2	0	0	0	0	0

Table 2

A species list of all recovered bacterial taxa for diseased and treated sponge samples. Samples PIII was sequenced using the Illumina platform using either the 515yF V4/V5 primer set or the 357wF V3/V4 primer set. Samples PIII 454 and Rescue were sequenced using the Roche 454 platform. The Rescue sample was a diseased tissue sample that was treated in the lab and cured. Values represent the number of sequences recovered that identify with that taxonomic group with a confidence >80%.

	PIII 515yF	PIII 357wF	PIII 454	Rescue
Acetobacteraceae	3	7	2	0
Acidimicrobiaceae	0	4	0	0
Acidimicrobiales	13	79	1	0
Acidobacteria Gp1	0	0	0	1
Acidobacteria Gp3	5	12	1	1
Acidovorax	4	0	0	14
Actinobacteria	1	95	11	0
Actinomycetales	108	120	9	0
Afipia	0	0	1	0
Aggregicoccus	2	0	0	0
Alcaligenaceae	18	16	1	0
Algoriphagus	0	2	0	0
Alkanindiges	0	0	0	18
Alphaproteobacteria	13	156	95	5
Aquisphaera	0	8	0	0
Archaea	0	0	0	1
Arcicella	0	0	1	14
Arenimonas	4	0	0	0
Armatimonadetes gp5	0	2	0	0
Bacillariophyta	16	8	0	0
Bacteria	539	406	2543	1226
Bacteriovorax	0	0	0	7
Bacteroidales	0	0	0	1
Bacteroides	0	0	0	1
Bacteroidetes	12	38	5	132
Bdellovibrio	0	0	0	1
Beijerinckia	2	9	0	0
Beijerinckiaceae	5	0	0	0
Betaproteobacteria	57	30	5	147
Blastopirellula	1	0	0	0
Bradyrhizobiaceae	0	0	0	1
Brevifollis	0	0	0	2
Bryocella	0	0	0	1
Burkholderiales	11	7	5	341
Burkholderialesincertaesedis	0	1	0	29
Candidatus pelagibacter	5	15	10	0

Caulobacter	0	0	0	17
Caulobacteraceae	0	0	0	1
Chitinophagaceae	438	27	149	58
Chlorophyta	7	2	1	0
Chloroplast	1	9	3	1
Chryseobacterium	0	0	0	8
Clostridiaceae 1	0	0	0	3
Clostridiales	0	0	0	5
Clostridium sensustricto	0	0	0	1
Cnuella	0	0	0	1
Comamonadaceae	116	3	4	114
Coriobacteriaceae	0	0	0	1
Cryomorphaceae	26	0	0	33
Cyanobacteria	1	4	149	0
Cyanobacteria Chloroplast	11	5	41	0
Cytophagaceae	0	2	1	2
Cytophagales	0	0	0	1
Deltaproteobacteria	2	16	1	0
Dermacoccaceae	0	2	0	0
Dorea	0	0	0	1
Duganella	0	0	0	1
Dyadobacter	0	0	0	4
Emticicia	0	0	0	1
Enterobacteriaceae	0	0	1	1
Erysipelotrichaceae	0	0	0	1
Escherichia.Shigella	0	0	0	1
Family II	0	0	7	0
Ferruginibacter	11	3	1	0
Firmicutes	0	0	1	1
Flavobacteriaceae	0	0	0	1016
Flavobacteriales	0	0	0	105
Flavobacterium	0	2	1	2501
Fluviicola	0	2	0	190
Frigidibacter	0	0	0	2
Gaiella	2	0	0	0
Gammaproteobacteria	191	132	8	429
Gemmata	34	3	9	0
GpIIa	5763	5135	6267	0
GpVI	4	3	1	0
Heliimonas	0	0	14	0
Herminiimonas	0	0	0	1
Hydrocarboniphaga	0	0	0	1
Hymenobacter	0	7	0	0

Ilumatobacter	6	45	0	0
Intrasporangiaceae	0	1	0	0
Janibacter	0	4	0	0
Janthinobacterium	0	0	0	98
Lachnospiraceae	0	0	0	1
Lacibacterium	1	0	0	0
Lacihabitans	4	0	0	0
Legionella	1	8	11	0
Legionellales	0	0	5	0
Limisphaera	0	3	0	0
Limnohabitans	0	85	9	0
Luteolibacter	0	2	1	1
Methylococcaceae	0	0	1	0
Methylocystaceae	0	1	0	0
Methylocystis	0	17	0	0
Methylophilaceae	0	18	1	0
Methylotenera	0	0	4	0
Moraxellaceae	0	0	0	3
Myxococcales	0	0	1	0
Namhaeicola	0	0	0	1
Nevskia	0	0	0	2
Nitrosomonadaceae	12	0	0	0
Nitrosospira	0	35	4	0
Oligoflexus	5	3	0	20
Opitutae	12	0	0	0
Opitutus	31	3	1	2
Oxalobacteraceae	0	0	1	541
Paludibaculum	0	0	1	0
Parachlamydiaceae	0	1	0	0
Parasegetibacter	0	1	0	0
Parvibaculum	0	0	0	1
Paucibacter	0	0	0	39
Pedobacter	0	0	0	26
Pedomicrobium	0	0	1	0
Pelomonas	0	0	0	3
Peredibacter	1	0	0	0
Perlucidibaca	0	0	0	2
Phycisphaera	0	0	29	1
Phycisphaerae	0	17	5	0
Phyllobacterium	0	12	0	0
Planctomycetaceae	1114	171	123	0
Planctomycetes	0	27	4	1
Planctomycetia	1	0	0	0

Polaromonas	8	1	0	1
Polymorphobacter	12	0	0	0
Polynucleobacter	81	73	11	0
Propionibacterium	0	14	0	0
Propionigenium	0	0	0	1
Proteobacteria	35	38	27	450
Proteus	0	0	1	0
Pseudarcicella	5	0	0	0
Pseudomonadaceae	0	0	1	127
Pseudomonadales	0	0	0	115
Pseudomonas	0	0	0	273
Pseudorhodoferax	0	0	1	1
Rhizobiaceae	0	0	0	1
Rhizobiales	0	5	16	2
Rhizobialesincertaesedis	0	0	0	1
Rhodobacteraceae	0	3	1	16
Rhodoferax	42	25	1	122
Rhodospirillaceae	177	18	0	0
Rhodospirillales	20	102	1	0
Roseomonas	3	3	0	0
Rugamonas	0	0	0	50
Runella	0	0	0	1
Saccharibacteria genera incertaesedis	0	116	4	0
Saprospiraceae	0	2	0	0
Schlesneria	46	0	0	0
Sediminibacterium	755	167	4	0
Serpens	0	0	0	1
Singulisphaera	0	0	1	0
Sinobacteraceae	0	0	0	1
Soonwooa	0	0	0	1
Spartobacteria	13	59	9	0
Spartobacteria genera incertaesedis	128	649	27	0
Sphingobacteriaceae	0	0	0	6
Sphingobacteriales	0	1	1	0
Sphingomonadaceae	0	0	4	0
Sphingorhabdus	4	0	0	0
Subdivision3	0	12	0	0
Subdivision3 genera incertaesedis	21	23	1	0
Telmatocola	0	0	1	0
Terrimicrobium	4	10	4	0

Turneriella	0	3	1	0
Undibacterium	0	0	2	1641
Variovorax	0	0	0	2
Verrucomicrobia	33	1853	328	3
Verrucomicrobiaceae	0	3	1	1
Vibrionaceae	0	0	1	0
Xanthomonadaceae	0	0	1	1
Xylophilus	0	0	0	1

Table 3

Indicator Species Analysis of samples grouped together based on health of the organisms they were harvested from. Healthy 1 is a group consisting of PI-515yF and PII-515yF. Healthy 2 consists of PI-357wF, PII357wF and Healthy-454. The sick group consists of PIII-515yF, PIII-357wF and PIII-454. Significance codes: 0.05 '*'

Healthy 1		
	Stat	P value
Armatimonas Armatimonadetes gp1	0.707	0.252
Arthrobacter	0.707	0.252
Bacillales	0.707	0.252
Bacilli	0.707	0.252
Cytophagales	0.707	0.252
Escherichia.Shigella	0.707	0.252
Micrococcaceae	0.707	0.252
Pirellula	0.707	0.245
Pseudomonas	0.707	0.252
Sneathia	0.707	0.245
Staphylococcus	0.707	0.252
Streptophyta	0.707	0.252
Enterobacteriaceae	0.695	0.465
Rhizobacter	0.692	0.465
Aquicella	0.67	0.465

Healthy 2		
	Stat	P value
Clostridiaceae 1	1	0.0396*
Flavobacteriales	1	0.0396*
Methylotenera	0.901	0.1273
Bradyrhizobiaceae	0.816	0.2523
Chlamydiales	0.816	0.2523
Conexibacter	0.816	0.2523
Diplorickettsia	0.816	0.2523
Flavobacteriaceae	0.816	0.2455
Gemmobacter	0.816	0.2523
Neochlamydia	0.816	0.2523
Parcubacteria genera incertae sedis	0.816	0.2523
Aciditerrimonas	0.577	1
Anaerolineaceae	0.577	1
Archaea	0.577	1
Armatimonadetes	0.577	1
Bacteroides	0.577	1
Bangiophyceae	0.577	1

Catenococcus	0.577	1
Caulobacteraceae	0.577	1
Chlorarachniophyceae	0.577	1
Chryseolinea	0.577	1
Clostridiales	0.577	1
Corynebacterium	0.577	1
Coxiellaceae	0.577	1
Cryptomonadaceae	0.577	1
Curvibacter	0.577	1
Gallionellaceae	0.577	1
Gimesia	0.577	1
Humitalea	0.577	1
Hyphomicrobium	0.577	1
Labrys	0.577	1
Lactococcus	0.577	1
Methylosoma	0.577	1
Methylovirgula	0.577	1
Micrococcus	0.577	1
Microvirgula	0.577	1
Nitrospira	0.577	1
Nocardia	0.577	1
Paucibacter	0.577	1
Pelomonas	0.577	1
Rhodobacter	0.577	1
Roseiarcus	0.577	1
Solirubrobacterales	0.577	1
Tabrizicola	0.577	1
Tepidisphaera	0.577	1
Turicibacter	0.577	1
Xanthomonadales	0.577	1

Sick

	Stat	P value
Opitutus	1	0.0344*
Acidobacteria Gp3	0.989	0.023*
Deltaproteobacteria	0.92	0.204*
Acetobacteraceae	0.916	0.0428
GpVI	0.88	0.1897
Cytophagaceae	0.816	0.2496
Nitrosospira	0.811	0.2129
Polaromonas	0.792	0.2141
Roseomonas	0.79	0.2121

Afipia	0.577	1
Aggregicoccus	0.577	1
Arcicella	0.577	1
Armatimonadetes Gp5	0.577	1
Blastopirellula	0.577	1
Burkholderiales incertae sedis	0.577	1
Dermacoccaceae	0.577	1
Firmicutes	0.577	1
Gaiella	0.577	1
Hymenobacter	0.577	1
Intrasporangiaceae	0.577	1
Janibacter	0.577	1
Lacihabitans	0.577	1
Methylococcaceae	0.577	1
Nitrosomonadaceae	0.577	1
Paludibaculum	0.577	1
Parasegetibacter	0.577	1
Pedomicrobium	0.577	1
Phyllobacterium	0.577	1
Proteus	0.577	1
Pseudarcicella	0.577	1
Singulisphaera	0.577	1
Telmatocola	0.577	1
Vibrionaceae	0.577	1

Healthy 1& Healthy 2	3	
	Stat	P value
Luteolibacter	0.99	0.445
Bacteriovorax	0.775	0.57
Clostridium sensu stricto	0.775	0.453

Healthy 1 & Sick		
	Stat	P value
Subdivision3 genera incertae sedis	0.876	0.568
Beijerinckia	0.775	0.351
Sphingobacteriales	0.711	0.521
Acidovorax	0.632	0.671
Lacibacterium	0.632	0.671
Propionibacterium	0.632	0.678
Pseudomonadaceae	0.632	0.682

Healthy 2 & Sick

	Stat	P value
Actinobacteria	1	0.0368*
Alcaligenaceae	1	0.0368*
Candidatus.Pelagibacter	1	0.0368*
Terrimicrobium	1	0.0368*
Chloroplast	0.999	0.3001
Chlorophyta	0.998	0.3573
Cyanobacteria	0.983	0.5154
Rhodoferax	0.98	0.5289
Spartobacteria genera incertae sedis	0.963	0.3239
Gammaproteobacteria	0.943	0.4789
Polynucleobacter	0.942	0.4832
Spartobacteria	0.936	0.2467
Limnohabitans	0.913	0.2419
Methylophilaceae	0.913	0.2419
Phycisphaerae	0.913	0.1403
Planctomycetes	0.913	0.1403
Saccharibacteria genera incertae sedis	0.913	0.2014
Comamonadaceae	0.901	0.49
Acidimicrobiales	0.894	0.4351
Gemmata	0.889	0.3526
Ferruginibacter	0.816	0.4089
Ilumatobacter	0.816	0.422
Opitutae	0.816	0.4339
Acidimicrobiaceae	0.707	0.6818
Cryomorphaceae	0.707	0.6828
Family II	0.707	0.6755
Flavobacterium	0.707	0.6882
Limisphaera	0.707	0.6818
Methylocystis	0.707	0.6772
Oxalobacteraceae	0.707	0.6814
Phycisphaera	0.707	0.6755
Sphingomonadaceae	0.707	0.6755
Subdivision3	0.707	0.6818
Verrucomicrobiaceae	0.707	0.6707
Algoriphagus	0.577	1
Arenimonas	0.577	1
Fluviicola	0.577	1
Heliimonas	0.577	1
Legionellales	0.577	1
Methylocystaceae	0.577	1

Myxococcales	0.577	1
Parachlamydiaceae	0.577	1
Peredibacter	0.577	1
Planctomycetia	0.577	1
Polymorphobacter	0.577	1
Pseudorhodoferax	0.577	1
Saprospiraceae	0.577	1
Sphingorhabdus	0.577	1
Undibacterium	0.577	1
Xanthomonadaceae	0.577	1

VITA

In 1997, Colin Rorex entered Carleton College at Northfield, Minnesota. He received the degree of Bachelor of Arts in Biology from Carleton College in June of 2001. That summer he joined the lab of Jason T. Kimata, Ph.D. at the Southwestern Foundation for Biomedical Research at San Antonio, Texas. In 2003, he joined the lab of Patrick Smith, Ph.D. at Louisiana State University Health Science Center Shreveport as a Research Associate. In 2004, he enrolled in the Department of Immunology and Microbiology's Ph.D. program at Louisiana State University Health Science Center Shreveport before separating from the program in 2005. He joined the lab of Alexander Huk at University of Texas at Austin in 2008 as a laboratory manager. In August 2016, he entered the Graduate School of Stephen F. Austin State University, and received the degree Master of Science in May of 2018.

Permanent Address: 116 Jack Pine Dr.
Kalispell, MT 59901

The journal Ecology was used as the style guide for the presentation of this thesis

This thesis was typed by Colin Rorex