NSU Florida
NOVA SOUTHEASTERN UNIVERSITY

Nova Southeastern University
NSUWorks

All HCAS Student Capstones, Theses, and Dissertations

HCAS Student Theses and Dissertations

8-4-2020

Metagenomes of the Dominant Microbial Symbionts of Sponge Genus Cinachyrella Display Common Sulfur Metabolic and Quorum Sensing Functions

Elizaveta Shmakova

Follow this and additional works at: https://nsuworks.nova.edu/hcas_etd_all

Share Feedback About This Item

This Thesis has supplementary content. View the full record on NSUWorks here: https://nsuworks.nova.edu/hcas_etd_all/14

NSUWorks Citation

Elizaveta Shmakova. 2020. *Metagenomes of the Dominant Microbial Symbionts of Sponge Genus Cinachyrella Display Common Sulfur Metabolic and Quorum Sensing Functions.* Master's thesis. Nova Southeastern University. Retrieved from NSUWorks, . (14) https://nsuworks.nova.edu/hcas_etd_all/14.

This Thesis is brought to you by the HCAS Student Theses and Dissertations at NSUWorks. It has been accepted for inclusion in All HCAS Student Capstones, Theses, and Dissertations by an authorized administrator of NSUWorks. For more information, please contact nsuworks@nova.edu.

Thesis of
Elizaveta Shmakova

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science
Biological Sciences

Nova Southeastern University
Halmos College of Arts and Sciences

August 2020

Approved:
Thesis Committee

Committee Chair: Jose Victor Lopez

Committee Member: Cory C Padilla

Committee Member: Julie Torruellas Garcia

This thesis is available at NSUWorks: https://nsuworks.nova.edu/hcas_etd_all/14

NOVA SOUTHEATERN UNIVERSITY

HALMOS COLLEGE OF ARTS AND SCIENCES

METAGENOMES OF THE DOMINANT MICROBIAL SYMBIONTS OF SPONGE GENUS *CINACHYRELLA* DISPLAY COMMON SULFUR METABOLIC AND QUORUM SENSING FUNCTIONS

By

Elizaveta Shmakova

Submitted to the Faculty of
Halmos College of Arts and Sciences
in partial fulfillment of the requirements for
the degree of Master of Science with a specialty in:

Biological Sciences

Nova Southeastern University

09.08.2020

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere appreciation to my major supervisor, Professor Dr. Lopez, who guided and encouraged me to be professional and curious about the hidden world around us. His patience cannot be underestimated. I really want to thank him for believing in me and trusting me as a research assistant in his lab. This position helped me gain more experience in genomics and microbiology, which opened up a new world for me.

Also, I would like to recognize the invaluable assistance that Dr. Cory Padilla provided me during my studies. Without his guidance and persistent help, this thesis would not have been possible. In addition, advice and comments given by Dr. Julie Garcia have been a great help for me in questioning the "big picture" of my work. I would like to give my special thanks to Lauren Krausfeldt, for her useful presentation and bioinformatics tools and suggestions.

Furthermore, I would also like to thank my labmates past and present-Eric Fortman, Denise Swack, Yvain Desplat, Shelby Cain, Colleen Mcmaken, and Hyo Won Lee. Eric, thank you for showing me and guiding me through major genomics techniques. Denise, thank you for helping me with my first extraction. Yvain, thank you for your positive attitude and lab help. Shelby, you've been such a big help during our metagenomics discussions. Coleen, thank you for making this such a great presentation experience together. Hyo Won, I will never forget your meticulous approach in finding answers, which showed me great examples for how to problem solve. I truly believe that these people represent the future of microbiology.

I am incredibly thankful to my best friend and boyfriend, Maxim Gralnik, who showed incredible support while I was studying and working on this thesis. Moreover, I learned how to be dedicated and hardworking from his personal example. Also, I would like to thank my good friend Kaci Gregory who has helped me improve my English and has supported me from the very beginning. I want to thank my friend Karthik, who always believed in my potential and never doubt it. Moreover, I want to thank my friend Moishe, because of who I started this career and who showed me that everything is possible if we really want it.

Lastly, the completion of my thesis would not have been possible without the constant support and nurturing of my family. My dad always tells me that I can achieve everything that I dream about and he truly believes in me. My mother is always ready to discuss any issue with me to help me to become a better person. My brother, Sergey, who has taught me how to stay positive in any situation.

I would like to dedicate this thesis to my grandmother, Larisa, who showed me the importance of education.

ABSTRACT

Host associated microbes play important roles in animal health. The symbiotic relationship between sponges and the microorganisms, prokaryotes-in particular, forms a complex mutualistic system, or community called the “microbiome”. Prokaryotes play a dynamic roll in these symbioses including providing nutritional sources, digestion, development, metabolism, immune defenses, while host sponge provides an environment for the symbionts. However, the whole metabolism of a sponge, holobioint, and the symbiont diversity and respective contributions to the holobtiont remains unclear. Metagenomics is an approach to study these systems by analyzing the gene content of the microorganisms for both taxonomic affiliation and biochemical function. The primary objective of this study is to describe local reef sponge *Cinachyrella* microbiome diversity, composition, and functional roles they occupy. Sponge metagenome sequences were assembled and binned into twenty-six high-quality Metagenomically Assembled Genomes (MAGs) and were taxonomically and functionally characterized. These two MAGs have been identified as *Ardenticatena* (Phylum Chloroflexi), and *Thioalkalivibrio* (Phlyum Proteobacteria). The MAG characterization includes inter/intraspecific diversity, composition, and functional roles they occupy. These two MAGs represent over 20% abundance of all sequence reads, suggesting their important functions for the holobiont. A MAG matching Thioalkivibrio has been shown to occur in other sponge symbioses. Conserved function in several genomes included quorum sensing and sulfur metabolism.

Keywords: symbiosis, *Cinachyrella,* microbiome, MAGs

LISTS

Lists of figures

List of tables

INTRODUCTION

The symbiotic relationship between sponges and the microorganisms they host, prokaryotes-in particular, forms a complex system called the "microbiome". Sponge microbiome composition is predominantly host-driven, with only a small degree of overlap with seawater microbial communities. The host-microbiome relationships can uncover the coevolutionary changes that they went through to survive in current environment. These functional relationships between sponge *Cinachyrella* and its microbiome are still not well described, including sulfur metabolism and quorum sensing.

Symbiosis

Microbiome plays an important role in the host life, including providing nutritional sources, digestion, development, metabolism, and immune defenses. Many host-prokaryotic interactions are classified as symbiotic. This can be contrasted with commensalism or pathogenesis. Symbiosis is a partnership between two or more organisms, which may, but does not necessarily benefit each member. Microorganisms are fundamentally important to many aspects of a host's phenotype, adaptations, and evolution. A host with all of its microbial symbionts was named "holobiont" by Lynn Margulis in 1991 (Gilbert, Sapp & Tauber, 2012). The importance of holobiont is evident in many studies of host metabolism, development, immunity and many other processes (Lee & Mazmanian, 2010; McCutcheon & von Dohlen, 2011; McFall-Ngai et al., 2011). A study by McCutcheon & von Dohlen (2011) found that hosts and their microbial community (also known as "microbiomes") genomic data showed remarkable levels of metabolic complementary functional roles between the bacterial symbionts of several insect lineages. Moreover, host-microbiome relationships provide the holobiont with many other advantages including evolutionary adaptation to parasites (Rigaud & Juchault, 1993). The hologenome theory of evolution explains the importance of the host-microbiome interactions, emphasizing the importance of phenotypic selection of the host and microbial members of the holobiont. It was discovered that community heritability and phenotypic characteristics of the holobiont can be analyzed by its hologenome (Shropshire & Bordenstein, 2016).

Microbiomes

At a microscopic level trillions of microorganisms of thousands of species live on the surface and inside of our body and other live creatures. However, because of their small size these microorganisms might make up only a small percent of the host's the body mass. For example, Human Microbiome Project (HMP) calculated that microorganisms occupy 1 to 3 percent of the body's mass (Lloyd-Price et al., 2017). The microbial network includes bacteria, fungi, parasites, and viruses. Each organism has a unique community coexisting with the host. Nobel Laureate and Microbiologist, Joshua Lederberg, first coined the term "microbiome" in 2001 (Prescot, 2017).

In order to analyze these diverse communities, standard culturing techniques were used for many years. However, these techniques can identify only 1% or less of the bacterial diversity in most environmental samples (Amann, Ludwig & Schleifer, 1995). It was found that nearly 99% of the microorganisms cannot be cultured by standard techniques (Riesenfeld, Schloss & Handelsman, 2004). The uncultured fraction is composed of diverse microorganisms that are only distantly related to the cultured ones and therefore have to be analyzed. The solution is culture-independent methods that were developed to examine genetic diversity, population structure, and ecological roles of many uncultured organisms. One method is metagenomics, or the culture-independent genomic analysis of assembled microorganisms, which includes the functional and sequence-based analysis of microbial genomes directly from the environmental sample (Riesenfeld, Schloss & Handelsman, 2004). The concept of analyzing DNA directly from its environment was initially proposed by Pace et al. (1985) and first implemented by Schmidt, DeLong & Pace (1991). Sequencing technologies for genome technologies and cost are rapidly improving over time. Therefore, it has become feasible to sequence near-complete genomes from a metagenome sample.

The human microbiome plays an important role in maintenance of health and of the immune system. Today, the microbiome is considered to be one of the life-supporting organs, which plays an extremely important role in maintaining one's health. In addition, certain adverse modifications of the microbiome may be responsible for various disorders of physical and mental functions. The National Institutes of Health Human Microbiome Project has provided a broad characterization of microbial and functional diversity using 2,355 total human microbiomes (Lloyd-Price et al., 2017). This project describes inter- and intra-personal microbiome variability

to compare the effects of personal microbiomes. For example, the Human Microbiome Project described microbial and functional variation over time of diverse body sites in 265 individuals. It was discovered that the human gut community (Bacteroidetes species in particular) is highly personalized compared to other sites. Even though the project already discovered crucial characteristics of the human microbiome, many aspects still remain to be characterized by this project.

The Earth Microbiome Project (EMP, http://www.earthmicrobiome.org) established a framework to catalogue microbiota globally. This endeavor provides a reference database of global DNA context and a framework for integrating data from future studies, encouraging complete characterization of Earth's microbiome diversity on this planet (Thompson et al., 2017). The project promised to analyze 200,000 samples from diverse communities with collaborative effort using sequencing, metagenomics, and metabolomics. This will help to generate a global Gene Atlas describing protein space, a global metabolic model, and a data-analysis portal for visualization of processed information.

These two projects are systematizing and characterizing the planet's microbial taxonomic and functional diversity for the well-being of the planet and humankind. The results of the EMP and similar consortia have accelerated our understanding of microbial taxonomic and functional diversity and how they interact with their surrounding environments, including host interactions. Microbial communities live in constant interaction with other organisms, whether it be with the host or a different microbe. One of these types of interaction is parasitism, this microorganism is harmful to other organisms (Wilkins et al., 2019). Some bacteria depend upon other organisms but can also be beneficial to these organisms, or neutral - symbiotic bacteria. Symbiotic relationships are present if some microorganisms help support the life of others. Usually in symbiotic relationships substances produced by one microorganism in the process of metabolism are used or consumed by other microorganisms.

One such area in which metagenomic sequencing of microbiomes has yielded rich findings is the sponge-microbe interaction. The most ancient living Metazoa - sponges, are in constant interaction with microbes and have evolved together with them. Therefore, sponges form symbiotic relationships with complex communities of microorganisms (Hentschel et al., 2012). In general, sponges have diverse but specific symbiont communities, even though their filter-feeding activities constantly filter out seawater microorganisms (Taylor et al., 2013). The

microorganisms which can inhabit sponge include bacteria, archaea, eukaryotes, and viruses (Pita et al., 2018). The microbial communities are composed of generalist microbes which are found in many sponge species and specialists that are enriched in this specific sponge. They are constantly exposed to the free-living microbes in the seawater which are their primary food source and at the same time they are able to harbor distinct symbiotic microbial communities. A study by Thomas et al., (2016) compared 81 different sponge species and found that their microbiome complexity (as assessed by the number of OTUs) ranges from 50 to 3,820 genetically distinct symbionts per host. However, the ecological and functional significance of these microbiomes in many sponges remain largely unknown.

High-throughput (HT) DNA sequencing and informatics enable the characterization microbiomes. HT DNA sequencing is often applied to microbiome systems, such as sponge microbiomes, in two ways, 1) generating amplicons originated from the subunit of the ribosome as indicator of taxonomic affiliation and abundance referred to as 16s rRNA sequencing and 2) whole genome shotgun (WGS) where whole DNA is extracted converted into a sequencing library.

The use of 16s rRNA sequence is a powerful approach to characterizing taxonomic composition. In this approach amplicon sequences are collapsed into a representative sequence based on shared sequence identity known as an operational taxonomic unit (OTU). The similarity threshold for defining OTUs has changes with our understand of microbial species differentiation at the genetic level, often ranging from 97 – 99% shared sequence identity. While, highly utilized and an effective starting point in microbiome analyses, these data fall short in characterizing the functional capacity of a microbiome system, due to a high degree of hozintal gene transfer that occurring in microbial communities where two populations share the same 16S sequence, but one has acquired a functional gene laterally from a different, non-related taxa. In addition, environmental samples often lack robust full-genome homologues in reference databases, where a even strong match to a ribosomal subunit does not reflect the full range of genomic diversity associated the source of the query sequence.

In previous studies different metagenomic and metatranscriptomic sequencing approaches were used in order to explain the functional roles of the sponge microbiome. The Illumina MiSeq platform with a high sequence output revealed abundant bacterial diversity in metagenomic studies. Illumina MiSeq can produce more sequence reads than other methods, for

example Roche 454, at a lower cost (Caporaso et al., 2012). A study by Hirai, Nagai & Hidaka (2017) performed metagenomic analysis with Illumina MiSeq and Roche 454 and found Illumina MiSeq to be more cost-effective and have higher quality (2017). MiSeq is usually used for smaller projects, where the sample can be processed quickly (Quail, Swerdlow & Turner, 2009). There are a few ways to query microbial communities: a) specific marker genes analysis and b) a deeper metagenomic analysis of the entire genomic information contained in the system (shotgun metagenomics). The first method analyzes phylogenetic marker genes. Usually the 16S rRNA gene markers of bacteria are used to find the taxonomic composition of the community (Thomas et al., 2016). The marker gene analysis approach focuses only on a single gene. These taxonomic indicator sequences by themselves cannot identify metabolic and other functional capabilities of the microbiome (Langille et al., 2013).

When applied to environmental samples, WGS data can assembled into contiguous genomic units (contigs) and the contigs can be further clustered to into metagenome-assembled genomes (MAGs) (Albertsen et al., 2013). These genomic fragments of varying completeness reflect both the taxonomic and functional diversity contained in a microbiome. Comparing these assembled sequences to known genomes references (derived as whole genomes derived from a culturable material) enables the assignment of these environmental sequences to known functions. Of particular interested is the recent ability to produce MAGs by binning contigs by similar abundance and tetra-nucleotide frequency profiles (Wu et al., 2014). MAGs can be viewed as a proxy or substitute for reference genomes defined above but not necessarily culture derived and more based on bioinformatics.

An important step of understanding how microbiomes function within a system is the process of gene annotation of the assembled sequences. Genome annotation is the process by which open-reading frames (ORFs) are generated and queried against known genes to find matches or homologues, thereby assigning function to particular regions of the genome (Frishman et al, 1998). Simply, genome annotation is a process of making nucleotide sequence meaningful. Annotation also begins with finding open reading frames (Frishman et al, 1998). It is important to perform annotation, because DNA sequencing produces sequences with unknown functions. Genome annotation is a multi-level process that includes structural and functional annotation and prediction of protein-coding genes, as well as other functional genome units including pseudogenes, control regions, structural RNAs, tRNAs, small RNAs, insertion

sequences, and other elements. There are many annotation servers available online (Richardson and Watson 2013). The most commonly used and known are the NCBI, which provides a Prokaryotic Genomes Automatic Annotation Pipeline service via email, RAST (Aziz et al., 2008), xBASE2 (Chaudhuri et al., 2008), and PROKKA (Seemann, 2014).

A study by Burns et al. (2018) used metagenomic analysis for four sulfur prill DNA samples to understand the microbial basis of sulfur-driven denitrification (SDN) reactors. Most abundant taxa and their metabolic genes were analyzed and found to be potentially important contributors to SDN metabolism. The data was sequenced on an Illumina MiSeq, de-multiplexed, trimmed, quality controlled, and organized into metagenome- assembled genomes (MAGs) similarly to our proposed research. The metagenomic approach helped to identify the diversity of specific taxa and their potential roles for the host. Another study by Slaby et al. (2017) revealed a symbionts nutritional specialization and a significant enrichment of genes involved in bacterial defense and host colonization by using metagenomic analysis.

Sponges and their microbiome are an exemplar in showcasing hologenome functionality. This symbiotic relationship is considered to be a major contributor to the evolutionary success of sponges (Hentschel et al., 2012). Marine sponges are known for hosting dense and highly diverse microbial communities. The specific characteristics of marine habitats are reflected in the colossal variety of physiological capabilities of marine microorganisms. The coexistence of sponges and the sponge associated prokaryotes (SAPs) that inhabit them (against the background of changing external conditions) lead to the formation of many interactions between them. These interactions ultimately affect the uniqueness of their metabolism. Microbes are the driving force of many biogeochemical cycles, such as carbon, nitrogen, sulfur cycles (Falkowski, Fenchel & DeLong, 2008). Up to 40% of sponge mass consists of microbes (Schmitt et al., 2012). A “core microbial community” is when major microorganisms inhabiting a specific host. It was found in recent studies that a “core microbial community” has the most abundant microbial community in phylogenetically distant hosts (Hentschel et al., 2002). It should be mentioned that sponges also can be described by its “pan microbial genome", which consist of its ‘core genome’ with all the genes present in all strains, a ’dispensable genome’ with genes present in two or more strains, and genes which are unique to single strains (Tettelin et al., 2005). A study by Thomas et al. (2016) compared 81 sponge species and found 39 predominant microbial phyla. This study found few shared features in species composition across the phylum. Furthermore, symbionts were

characterized as specialists and generalists. It suggests that the microbiome should provide many functional roles for the host. However, many functional roles are still unclear. There needs to be a comprehensive study of microbial associates of sponges, which can help determine the possible functions of their coexistence.

Cinachyrella sponge

This study assesses the microbiome community of *Cinachyrella* sponge with the aims to characterize their functional role within the host. Many metagenomic studies analyzed host metagenomic data, which have potential for scientific discoveries (Wooley, Godzik & Friedberg, 2010; Thompson et al., 2017). The sponge genus *Cinachyrella* is a common resident of Western Atlantic and Caribbean reefs, including the Florida Reef Tract (Schmitt et al., 2012). According to Lifemap, genus *Cinachyrella* is a member of the Tetillidae family, as well as its 88 descendants. This taxon represents a genus of sponges that have 45 species (Fig. 1). This map depicts the evolutionary relationships as a tree-like structure. Interestingly enough, genus *Cinachyrella* shows all 45 species' branches stemming from a single point, indicating that *Cinachyrella* is a monophyletic clade. This tree is created based on the NCBI taxonomy updates from published phylogenetic trees.

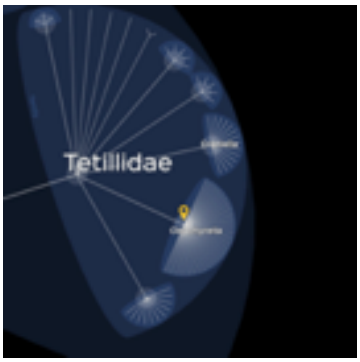


Fig.1: Genus *Cinachyrella* shows a taxon of sponges with 45 species (Lifemap NCBI).

At least four, and perhaps up to ten, *Cinachyrella* species (*Cinachyrella kuekenthali* 1-4, *C. alloclada1*/2 & 3, *C. apion* 1& 2, and *C. arenosa*) occur or overlap in the Western Atlantic area (Cárdenas & Schuster, unpublished results; Schuster et al, 2017). Several traits support the genus for use as an experimental system: this sponge can be maintained for weeks and months in

aquaculture, can reproduce via viviparous propagation or asexually (albeit at irregular times and unknown cues), and appears resistant to fouling. To date, a draft *Cinachyrella* transcriptome (Smith 2013), metagenome and multiple microbiomes from various individuals have been sequenced and characterized. *Cinachyrella* sp. was studied in the field and in laboratory experiments to investigate changes in holobiont physiology and microbial community structure in response to stressors such as crude oil and antibiotics (Baquiran et al., 2020). Electron microscopy of the holobiont ultrastructure has revealed both low and high microbial abundances in various *Cinachyrella* sp. For example, *Cinachyrella kuekenthali* and perhaps other congeners could be considered high microbial abundance (HMA) sponges via TEM. A recent study has shown that the presence of mitochondrial group I introns has divided sympatric *Cinachyrella* individuals in Florida into at least two *C. alloclada* species; there is also a strong correlation of intron presence/absence with divergence into two distinct microbiomes (group 1 and 2) of these same individuals. This finding suggests that host genetics (i.e. host divergence) can have a strong influence on microbiome structure, via currently unknown factors, even for visually identical, sympatric sponges which appear to have slight genetic differences.

The Lopez laboratory where I work, has been developing Cinachyrella sp sponges, as an experimental and evolutionary model. For example, the sponge was chosen for oil dosing experiments after the Deepwater Horizon oil spill in 2020 (Smith MS thesis; Vijayan MS thesis; Cuvelier et al 2014; Desplat in preparation).

HYPOTHESES

It is known that the host phylogeny can have a strong influence on microbiome structure and its genetics (Lutz et al., 2019; Youngblut et al., 2019). More specific questions regarding *Cinachyrella* sp. and other potential model sponges focus on the specificity of their microbial symbiont communities (microbiomes), and these potential effects on holobiont divergence and speciation. This characterizes the total microbial diversity through 16S sequencing and through the generation of MAGs assess the functional capacity of bacterial genomes found in *Cinachyrella* sp. host by using metagenomic methods. An underlying thread of the thesis addresses the question of whether there is an underlying “core” metabolism of this *Cinachyrella* sponge species. Statistical comparison of the genomic repertoire of sponge symbionts with reference genomes will provide information about the involved genes and their functional roles

within the host. A deeper understanding of these functions will provide greater insight into the involved molecular processes that underlie microbial-sponge symbiosis.

H1 – MAG generation will reveal novel genomes of sponge associated microbes providing insight into microbial speciation and *Cinachyrella* sponge microbiome. It will increase our knowledge about the microbiome and expand the genome database.
H2 – Annotation and comparative genomics will define the functional roles for key members of the *Cinachyrella* sponge microbiome, which might help to identify the functional roles in the microbial-sponge symbiosis.

MATERIALS AND METHODS

Sample Collection

Cinachyrella sponges studied in this project were originally collected as part of a sponge whole genome sequencing project in collaboration with Dovetail Genomics of California. They would apply shotgun, Chicago, and/or Dovetail Hi-C sequencing libraries. Focus was therefore on just a few sponge individuals collected off the coast of Dania Beach FL.

DNA extraction

Cells in the X fraction from the percoll isolation were selected as the Sponge Associated Prokayotes (SAPs). SAP DNA was isolated with the DNeasy PowerSoil Kit (Qiagen) as recommended by Earth Microbiome Project. DNA extracts were stored at -20°C until downstream analysis.

16S rRNA amplicon sequencing and analysis

The hypervariable V3-V4 region of the 16S rRNA gene was amplified and sequenced to assess SAP taxonomic composition on diversity metrics. Library preparation and analysis followed the workflow described in Padilla et al., 2015. Briefly, amplicons were synthesized using Platinum® PCR SuperMix (Life Technologies) with primers F515 and R806 (Caporaso et al., 2011). Both forward and reverse primers were barcoded and appended with Illumina-specific adapters according to Kozich et al. (2013). PCR reactions were carried out with 1 ng of DNA.

Thermal cycling involved: Denaturation at 94°C (3 min), followed by 30 cycles of denaturation at 94°C (45 s), primer annealing at 55°C (45 s) and primer extension at 72°C (90 s), followed by extension at 72°C for 10 min. Amplicons were analyzed by gel electrophoresis to verify size (~400 bp, including barcodes and adaptor sequences) and purified using the QIAquick PCR purification (Qiagen). Amplicons were sequenced on an Illumina MiSeq using a 500 cycle Nano kit with 5% PhiX to increase read diversity.

Amplicons were analyzed using QIIME (Caporaso et al., 2010). Barcoded sequences were de-multiplexed and trimmed (length cutoff 100 bp) and filtered to remove low quality reads (average Phred score <25) using Trim Galore!. Paired-end reads were then merged using FLASH (Magoč and Salzberg, 2011), with criteria of average read length 250, fragment length 300, and fragment standard deviation 30. Chimeric sequences were detected by reference-based searches using USEARCH (Edgar, 2010). Identified chimeras were filtered from the input dataset, and merged non-chimeric sequences were clustered into Operational Taxonomic Units (OTUs) at 97% sequence similarity using open reference picking with the UCLUST algorithm (Edgar, 2010) in QIIME. Taxonomy was assigned to representative OTUs from each cluster using the SILVA database. Data were visualized with the R package Phyloseq.

SAP Metagenome sequencing and assembly

A single shotgun library was generated from the same DNA extract that was used to generate the 16S rRNA data. The library was prepared using the NEBNext ® Ultra II DNA library Prep (New England Biolabs) as described above. This library was sequenced on an Illumina NextSeq resulting in 2 X 150 bp read pairs. Sequence statistics can be viewed in Supplemental Table 2.

The metagenomic assembly was generated after adaptor trimming and quality filtering. SPAdes (v 3.11.1) was used to assemble the short reads into contigs using the –meta flag (Nurk et al., 2013). Contigs were then clustered into metagenomically assembled genomes (MAGs) following the avin'o workflow (Eren et al., 2015). MAG contamination and completeness were assessed with checkm (Parks et al., 2015).

SAP MAG analysis

The resulting MAGs were annotated to assess both taxonomic affiliation and gene content. The tool Prodigal (PROkaryotic DYnamic programming Gene-finding ALgorithm) was used to predict open reading frames (ORFs) for each MAG (Hyatt, 2010). This program showed a greater sensitivity in identifying existing genes. It was designed to minimize the number of false positive predictions.

Annotation

Contigs in each genome cluster were annotated using PROKKA (v.1.7), which produces standards-compliant output files (Seemann, 2014). The software uses preassembled genomic DNA sequences in FASTA format, a set of scaffold sequences produced by *de novo* assembly software. *De novo* method does not rely on the reference genome, instead it predicts the genome structure based on the statistical models (Al-Nakeeb, Peterson & Sicheritz-Pontén, 2017). To decrease running time on multicore computers PROKKA uses parallel processing. The comparison between RAST, xBase2, and PROKKA on an *E.coli* genome showed that Prokka produced an overall better annotation (Seemann, 2014).

Coverage

The coverage analysis is a very straight forward and unsophisticated approach. It is quite reliable and hard to misinterpret or refute. The coverage per bin was calculated by mapping the number of reads back to the genome, per size of the genome. The mapped stats were calculated with samtools. To find # of reads per Mbp of genome, mapped number of reads (12244319) were divided by MAG size in Mbp.

LCA Taxonomic assignment of SAP MAGs

To deduce the taxonomic profiles of the SAP MAGs, a lowest common ancestor (LCA) was applied per each MAG. ORFs generated during gene prediction were queried against the NCBI-nr database (BLASTn). BLAST matches with a bit score > 50 were considered for further analyses. The resulting matches were processed through MEGAN (V6.18.5) for LCA taxonomic assignment (Huson et al., 2007; Hudson et al., 2011).

Functional analysis

Metabolic reconstruction of MAGs was performed by mapping of MAG-encoded gene products to KEGG pathways/modules. Metabolic pathways were predicted with the Kyoto Encyclopedia of Genes and Genomes (KEGG) tool BlastKOALA (Kanehisa, Sato & Morishima, 2016). Mapping of gene products encoded by assembled genes to KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways revealed 282 functional categories. Each functional category for each MAG was added to Supplemental Table 5.

Intraspecific Diversity within the Sponge Microbiome

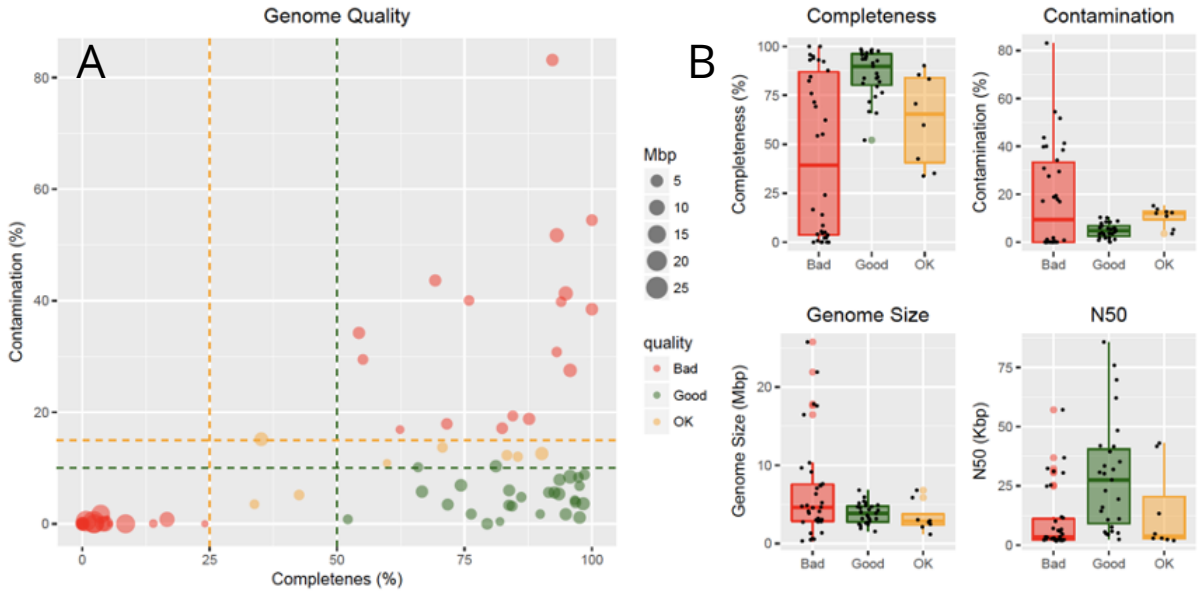
Population structure for the community drivers was assessed using recruitment plots (Konstantinidis and DeLong, 2008). The most abundant MAGs were assigned for genetic heterogeneity through the recruitment of sequence data to the binned genomes. To determine the extent of collapsed heterogeneity, trimmed and merged sequence data were sub-sampled to 1 million reads and aligned to the reference genomes through BLASTn, retaining the top match and requiring a bit score > 50. Recruitment plots, per MAG, were generated by plotting the percent identity against the alignment position in the reference genome. Counting for the amount of genomic information ensures that abundance is scaled to each genome size, so the data won't be skewed toward the longest MAG.

RESULTS

Sequencing Results

In order to acquire genome sequence information of abundant *Cinachyrella* symbiotic microorganisms, contigs assembled from metagenome reads were binned to MAGs (Metagenomically Assembled Genomes) providing the basis for further characterization of abundant sponge microbiome members and reconstruction of the important characteristics of their metabolism. A total of 69 MAGs were generated, with mean completeness of 62.62 % and mean length of 5.15 Mbp. MAG size equals to the number of bases in the MAG. Normalizing for MAG size is to ensure that the abundance is not skewed by the size of the genetic information contained in the MAG. Further analysis was extended to MAGs passed acceptable qualities for binned genomes as defined by the GOLD (REF) standards (>50% complete; <10% contamination), which resulted in 27 good quality MAGs. At least 42 samples with poor overall

sequence quality were excluded due to high contamination (more that 10%) or poor sample completeness (less that 50%) The highest number of contigs, 1419, belongs to MAG8. The longest MAG47 has length of 6,799,620 bp. The average length of all good quality MAGs is 3.84 Mbp. Also, the average of the minimum contig length needed to cover 50% of the genomes was found to be 783,863 for all good quality MAGs. Moreover, the average N50, the minimum contig length needed to cover 50% of the genome, is 18.056kbps (Fig. 2C).



C

Quality	Completeness	Contamination	No.Contigs	Mbp	N50.kbps
Bad :34	Min. : 0.00	Min. : 0.00	Min. : 43	Min. : 0.3132	Min. : 1.724
Good:27	1st Qu.: 33.81	1st Qu.: 0.80	1st Qu.: 342	1st Qu.: 2.7302	1st Qu.: 2.890
OK : 8	Median : 79.47	Median : 5.62	Median : 619	Median : 4.0046	Median : 6.999
NA	Mean : 62.62	Mean :11.95	Mean : 1510	Mean : 5.1649	Mean :18.056
NA	3rd Qu.: 93.10	3rd Qu.:15.17	3rd Qu.: 1489	3rd Qu.: 5.3828	3rd Qu.:30.695
NA	Max. :100.00	Max. :83.17	Max. :11000	Max. :25.7809	Max. :85.829

Fig. 2: A) Genome Quality graph of 69 MAGs shows all good quality MAGs in green, OK quality MAGs in yellow, and bad quality MAGs in red. B) bar plots of completeness, contamination, genome size, and N50 of bad, good and OK quality MAGs. C) Statistical analysis table of all the MAGs and their qualitative characteristics, including quality, completeness, contamination, number of contigs, mbp, and N50.

Unraveling the Sponge's Microbial Community Composition

To deduce the taxonomic profiles of the sponge *Cinachyrella* microbiomes analyzed, the bioinformatics tool MEGAN V6.18.5 for metagenomic sequence classification was applied. In total 27 MAGs were classified, identifying 8 bacterial phyla (Table 1). The following phyla with descending number of MAGs were identified: *Proteobacteria*, *Chloroflexi, Thaumarchaeota, Gemmatimonadetes, Nitrospirae*, *Verrucomicrobia, Actinobacteria*, and *Acidobacteria*. Taxonomic classifications for all MAGs are provided in Table 1. Overall, the composition of this microbiome is in accordance with microbiomes from other *Cinachyrella* sponge studies, which are dominated by *Proteobacteria* and *Actinobacteria* (Cleary et al., 2019; Cuvelier et al., 2014). However, the majority of the species were only analyzed and described with 16S data (Cuvelier et al., 2014)

Five MAGs are taxonomically affiliated with known sulfur utilizing microbes: *Opitutaceae bacterium*, *Thioalkalivibrio paradoxus*, *Desulfobacterium autotrophicum*, *Thioalkalivibrio sulfidiphilus*, *Sulfurifustis variabilis*.

Table 1: Summary of found species and phylum of high-quality Metagenomically Assembled Genomes (MAGs) of *Cinachyrella* sponge sample. MAGs are sorted by Phylum.

BinID	SpeClosest species matchcies	PhyPhylumlum:
MAG27	*Opitutaceae bacterium TAV5*	*Verrucomicrobia*
MAG11	*Cenarchaeum symbiosum A*	*Thaumarchaeota*
MAG18	*Cenarchaeum symbiosum A*	*Thaumarchaeota*
MAG26	*Cenarchaeum symbiosum A*	*Thaumarchaeota*
MAG2	*"Thioalkalivibrio paradoxus"*	*Proteobacteria*
MAG3	*Nitrosomonas communis*	*Proteobacteria*
MAG9	*Desulfobacterium autotrophicum*	*Proteobacteria*
MAG10	*Rhodoplanes sp.*	*Proteobacteria*
MAG12	*Azospirillum brasilense*	*Proteobacteria*
MAG13	*Acidithiobacillus ferrivorans*	*Proteobacteria*
MAG14	*Haliangium ochraceum*	*Proteobacteria*
MAG16	*Thioalkalivibrio sulfidiphilus*	*Proteobacteria*
MAG20	*Pseudomonas citronellolis*	*Proteobacteria*
MAG23	*Rhizobiales bacterium*	*Proteobacteria*
MAG25	*Pseudomonas citronellolis*	*Proteobacteria*
MAG5	*Sulfurifustis variabilis*	*Proteobacteria*
MAG8	*Polymorphum gilvum*	*Proteobacteria*
MAG6	*Nitrospira defluvii*	*Nitrospirae*
MAG15	*Gemmatirosa kalamazoonesis*	*Gemmatimonadetes*
MAG22	*Gemmatirosa kalamazoonesis*	*Gemmatimonadetes*
MAG1	*Ardenticatena*	*Chloroflexi*
MAG21	*Caldilinea aerophila*	*Chloroflexi*
MAG19	*Simkania negevensis*	*Chlamydiae*
MAG4	*Ilumatobacter coccineus*	*Actinobacteria*

MAG17	*Mycolicibacterium chubuense*	*Actinobacteria*
MAG7	*Acidobacteria bacterium*	*Acidobacteria*
MAG24	*Chloracidobacterium thermophilum*	*Acidobacteria*

Abundance

The relative contribution of the MAGs to the entire metagenomic pool was assessed to determine abundance and ecologic impact on the microbiome system. These abundances ranged from 0.067% to 34.34% per Mb of genome (Table 2). The major driver of the community is MAG_2 with 34.34% real-time abundance, which is Proteobacteria (*Thioalkalivibrio paradoxus*). It's a small size genome (0.48 Mb) with relatively small number of reads mapped (3803428). The second most abundant genome in the total dataset is MAG_1. It showed 16.64% abundance per Mb of genome. MAG_1 represents *Chloroflexi* (*Ardenticatena*). Similarities were observed when we compared taxonomic patterns of the MAG relative abundance with previously published 16S rRNA gene amplicon-based surveys.

Table 2. The relative abundance of each MAG per Mb of genome. To estimate the relative abundance of each MAG, the total number of reads mapped to a MAG was divided by the total number of reads in the metagenome sample and later by the total genome size in Mb for the corresponding MAG.

Bin ID	Number reads mapped	Genome Size	Genome Size (mb)	Abundance per Mb of genome (%)
MAG_2	3803428	476295	0.476295	34.34%
MAG_1	12244319	3165170	3.16517	16.63%
MAG_3	3048256	2692900	2.6929	4.87%
MAG_20	657484	2277246	2.277246	1.24%
MAG_5	1445159	5203527	5.203527	1.19%
MAG_9	781205	2822265	2.822265	1.19%
MAG_14	647010	2441330	2.44133	1.14%
MAG_6	782632	3202857	3.202857	1.05%
MAG_10	1139182	4685360	4.68536	1.05%
MAG_7	589806	2456476	2.456476	1.03%
MAG_21	506869	2277246	2.277246	0.96%

MAG_15	621708	2898325	2.898325	0.92%
MAG_4	1086408	5149290	5.14929	0.91%
MAG_22	1369924	7376007	7.376007	0.80%
MAG_8	906089	4904957	4.904957	0.79%
MAG_16	810765	5382816	5.382816	0.65%
MAG_18	376686	4744898	4.744898	0.34%
MAG_24	427804	6326817	6.326817	0.29%
MAG_12	927703	17842486	17.842486	0.22%
MAG_19	219072	4744898	4.744898	0.20%
MAG_27	353051	9660623	9.660623	0.16%
MAG_26	332142	10309494	10.309494	0.14%
MAG_25	513534	16471589	16.471589	0.13%
MAG_17	539472	17601912	17.601912	0.13%
MAG_13	501869	17842486	17.842486	0.12%
MAG_11	571674	21917315	21.917315	0.11%
MAG_23	401038	25780865	25.780865	0.07%

In order to find the major community drivers at the Phylum level, we combined individual abundance for each phylum. Table 3 indicates that Proteobacteria is the major system driver, followed by *Chloroflexi, Acidobacteria*, and *Gammaproteobacteria*.

Table 3. The relative abundance of each Phylum per Mb of genome. To estimate the relative abundance, the total number of reads mapped to a MAG was divided by the total number of reads in the metagenome sample and later by the total genome size in Mb for the corresponding MAG and then the abundances for each phylum were summed up.

Phylum	Abundance per Mb of Phylum
Proteobacteria	0.4700
Chloroflexi	0.1759
Acidobacteria	0.0236
Gemmatimonadetes	0.0172
Nitrospirae	0.0105
Chlamydiae	0.0020
Verrucomicrobia	0.0016
Thaumarchaeota	0.0011

Intraspecific Diversity within the Sponge Microbiome

Recruitment plots can reveal genetic heterogeneity within discrete population. Recruitment of reads can identify sequence-discrete populations (sequence-discrete "clusters") that are not clonal but instead are composed of highly similar, co-occurring genotypes that contain some degree of genetic diversity with highly conserve regions, represented by horizontal lines separated by percent similarity. The most abundant three MAGs, based on the relative abundance, where selected for speciation analysis. Recruitment plot for MAG_ 2, LCA classified as *Thioalkalivibrio paradoxus* (Fig. 3 A), consists of several sequence-discrete population with high similarity to the MAG (>95 percent identity). The plot identified at least three sub-populations with high percent identity (horizontal pile-up of reads with each distinct cluster >97 percent identity). The second most abundant MAG_1 – *Ardenticatena* (*Chloroflexi*) also consists of a high degree of genetic heterogeneity (Fig. 3 B). At least 5 sequence-discrete "clusters" (> 95 %ID) were associated with the MAG. This bin identified the highest number of discrete populations. The third most abundant sample MAG_3 *Nitrosomonas communis* (also

Proteobacteria), identified only one sequence-discrete "cluster" (> 95 %ID) in the metagenome sample (Fig. 3 C).

Roughly 65% of the genome MAG_2 consists of regions that share a high degree of similarity below 95 %ID, derived of sequence fragments from other taxa. These sequences are not associate with the population-level genome, suggesting a high rate of shared homologous regions with other taxa. On the other hand, a completely different pattern appears for MAG_1. Only a few reads cluster at genomic position in MAG_1. With less than 15% of the genome consisting of homologous regions shared across non-related (<95% ID) taxa. Interestingly, conserved regions of MAG_3 were in low richness as its cluster density. Meaning that for this MAG represents a unique population with low number of closely related populations whose reads mapped with <90% identity.

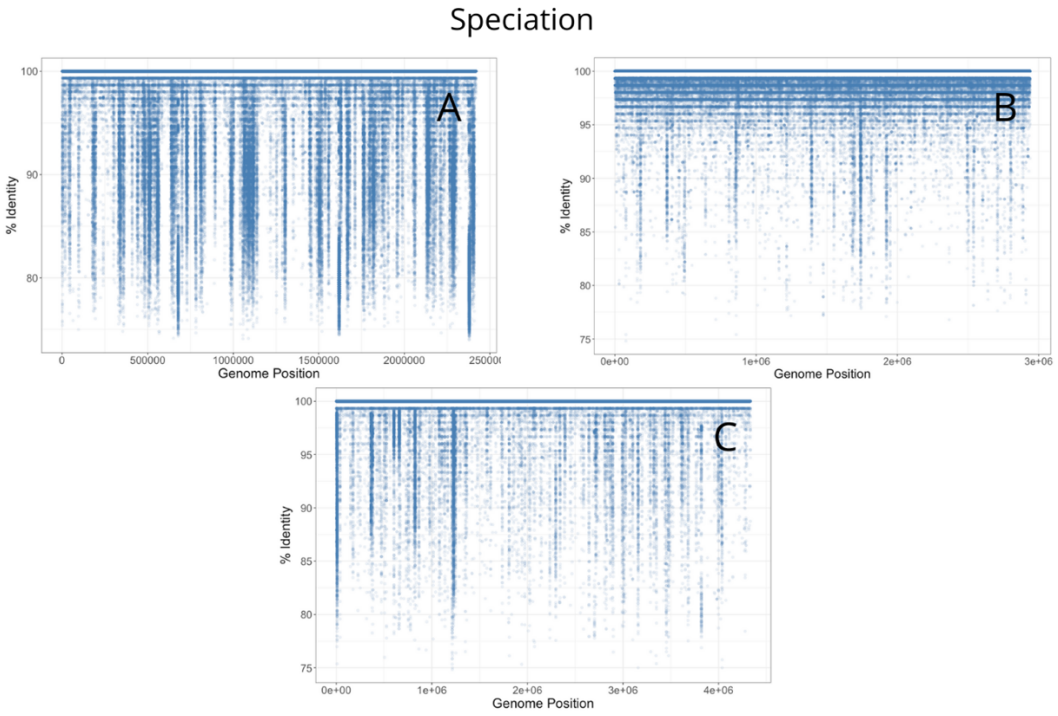


Fig. 3. Recruitment plots of the most abundant three MAGs, based on the relative abundance, represent the reads, placed by genome position (X-axis) and identity percentage (Y-axis). All short reads from the sample metagenomes were searched against all contigs of the bin in a NCBI blast search. The recruitment plot displays the total number of short read-derived base-positions at given percent identities. Each dot represents a read. For the recruitment bit score >50 was used. A) MAG_2 B) MAG_1 C) MAG_3.

Functional Characterization of MAGs

To describe the functional potential of reconstructed MAGs, metabolic reconstruction was performed by mapping MAG-encoded gene products to KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. Functional assignments to KEGG categories revealed 282 functional categories (see Supplement table 5).

Central metabolism was shared functional pathway. Mapping of gene products encoded by MAGs to KEGG pathways revealed that the categories ABC transporters, ribosome, purine metabolism, oxidative phosphorylation, aminoacyl-tRNA biosynthesis, two-component system, glycine, serine, and threonine, pyruvate metabolism, and quorum sensing received the most assignments (see Supplement table 5). The category 'ATP-binding cassette (ABC) transporters' is the most abundant category suggesting that basic metabolism is of prominent importance for sponge microbiome analysis.

Sulfurous compounds play essential roles in both metabolic and catabolic biochemical pathways. Assimilating sulfur as biomass occurs through amino acids, vitamins, secondary metabolite generation. Additionally, many prokaryotes use reduced sulfur as electron sources to generate energy and assimilate inorganic carbon. In total 27 MAGs showed 270 sulfur metabolism genes. More than 75% of the MAGs generated from the sponge microbiome contains PAPSS gene (2.7.7.4), which is the universal sulfonate donor's synthase. It is one of the major player in both dissimilatory and assimilatory sulfate reduction reactions. The Fig. 4 shows that all of the MAGs have some of the sulfur metabolism genes.

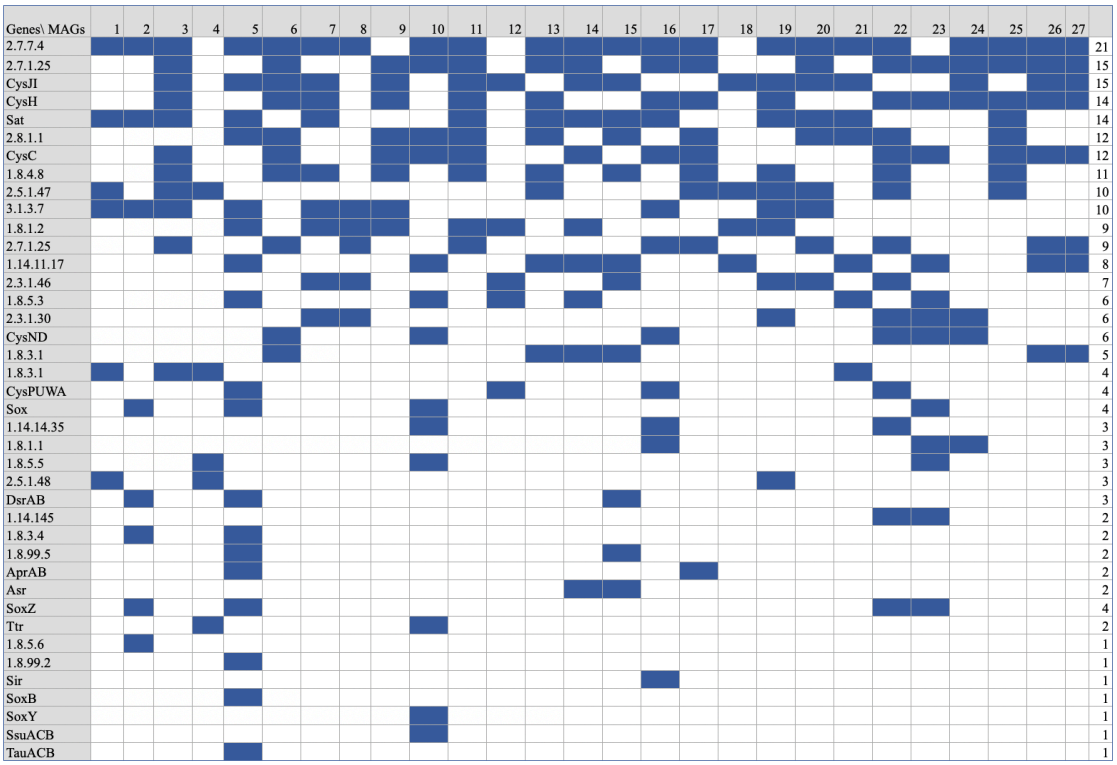


Fig. 4. Sulfur metabolism functional potential of the 27 high quality metagenomically assembled genomes (MAGs) as deciphered by means of KEGG (Kyoto Encyclopedia of Genes and Genomes) modules. Colums represent MAGs; KEGG sulfur metabolism genes are arranged in rows. Blue color of the heatmap indicates the presence of identified genes. Last column shows the total number of represented gene in all MAGs.

Nineteen out of 27 MAGs have Opp/Ami gene, which is an oligopeptide transport system substrate-binding protein involved in the internalization of signaling peptides. Eighteen out of 27 MAGs show the presence of PhnB and Sec genes (Fig. 5). PhnB gene codes for the glutamine-binding beta subunit (PhnB) of anthranilate synthase (AS) provides the glutamine amidotransferase activity to produce anthranilate, a precursor for *Pseudomonas* quinolone signal (PQS). PQS catalyses a bacterial anti-stress response.

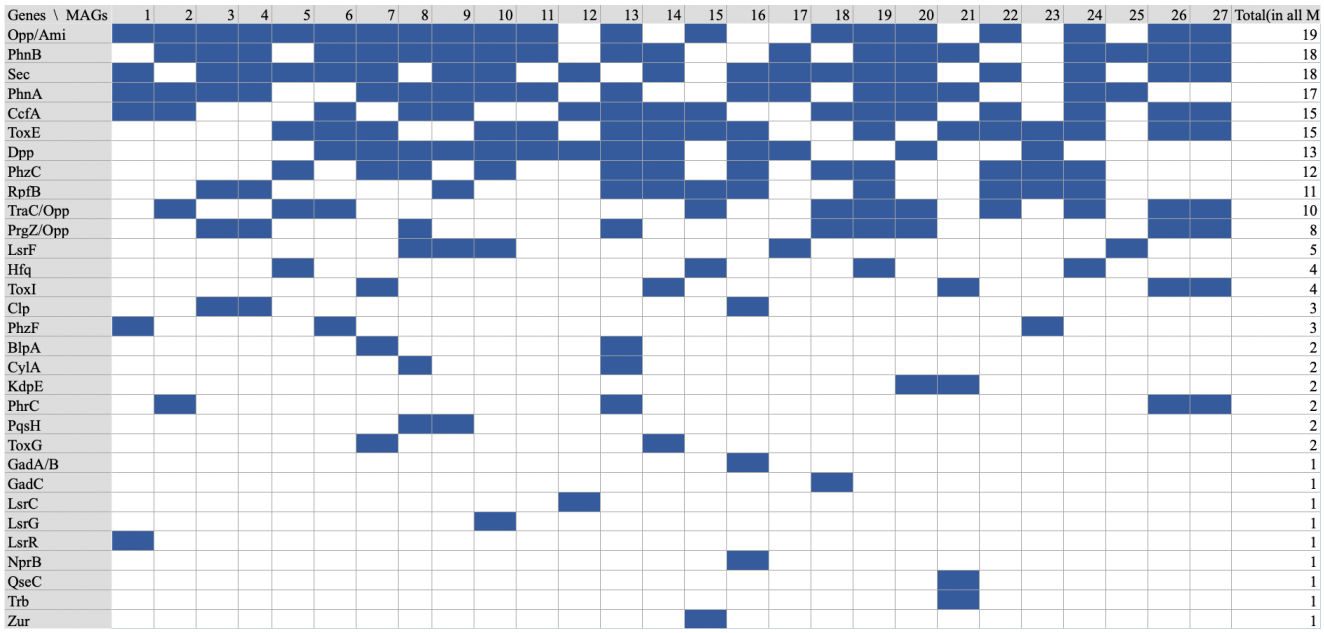


Fig 5. Quorum sensing functional potential of the 27 high quality metagenomically assembled genomes (MAGs) as deciphered by means of KEGG (Kyoto Encyclopedia of Genes and Genomes) modules. Colums represent MAGs; KEGG quorum sensing genes are arranged in rows. Blue color of the heatmap indicates the presence of identified genes. Last column shows the total number of represented gene in all MAGs.

Individual MAGs analysis

MAG_23

MAG_23 assigned to the phylum *Proteobacteria* species *Rhizobiales* bacterium NR represents the largest N50 score. N50 is a measure to describe the quality of assembled genomes. MAG_21 has the highest N50 of 85,829 with 97 contigs. MAG_23 is the largest found genome, which has a total length of 4,004,567bp. The completeness of the genome bin was assigned to 79.47% and features an extremely low contamination rate of 0%, as determined based on presence and copies of single copy marker genes. The average GC-content is 58.353% (std.deviation 3.599) and GC-ratio 0.720 (std.deviation 0.111). The phylum *Proteobacteria* is amongst the most abundant genera identified in the metagenome (Table 2). *Rhizobiales* bacterium is known for its ability to fix nitrogen when in symbiosis with leguminous plants and

pathogenic bacteria to animals and plants. Hence, they are of interest for application in agriculture and plant cultivation concepts.

Interestingly, MAG_23 also identified S16, which is very rare for a whole genome sequencing. KEGG annotation showed three complete energy metabolism pathways: Methane metabolism, sulfur matabolism, and ATP synthesis. MAG_23 uses methane oxidation, where methanotroph and methane oxidizes to formaldehyde. For the sulfur metabolism it showed assimilatory sulfate reduction, in which the sulfate reduces to H2S. On the other hand, complete nitrogen metabolism wasn't identified. KEGG identified 9 genes coding for nitrogen metabolism enzymes, including nitric oxide reductase subunit B (norB), formamidase, and nitrite reductase (NADH) large subunit (nirB). Moreover, a few complete pathways for metabolism of cofactors and vitamins were recognized, including NAD biosynthesis (from aspartate to NAD), Coenzyme A biosynthesis from pantothenate, molybdenum cofactor biosynthesis from GTP, and siroheme biosynthesis from glutamate. During the gene analysis, tlyC putative hemolysin (K03699) was identified, which is a pore-forming toxin.

GhostKOALA showed that the majority of the functional genes were responsible for general metabolism, including carbohydrate metabolism, protein families: genetic information processing, protein families: signaling and cellular processes, and amino acid metabolism. Also, MAG_23 had an abundant prokaryotic defense system, which included CRISPR-Cas system, Restriction and modification system, Toxin-antitoxin system (TA system), and DNA phosphothiolation system. ResFinder-3.2 Server analysis showed no acquired antimicrobial resistance gene to any of the following: beta-lactam, trimethoprim, nitroimidazole, aminoglycoside, quinolone, macrolide, fosfomycin, fusidicacid, oxazolidinone, glycopeptide, tetracycline, rifampicin, phenicol, colistin, sulphonamide.

MAG_14

MAG_14 assigned to the phylum *Proteobacteria* (species Haliangium ochraceum DSM 14365) represents the most complete genome bin (98.29%) and features a very low contamination rate of 3.6%. The phylum *Proteobacteria* is the most abundant phylum identified in this study and was supported in other studies as well for *Cinachyrella* sponge. MAG_13 is one of the longest found genomes with length 5.9 Mb and relatively high N50 (33,426). This genome showed to have 342 contigs. KEGG annotation showed three complete energy metabolism

pathways: carbon fixation (phosphate acetyltransferase-acetate kinase pathway), sulfur metabolism (assimilatory sulfate reduction), and ATP synthesis. Also, six complete cofactor and vitamin metabolism pathways were identified, including NAD biosynthesis, coenzyme A biosynthesis, biotin biosynthesis, molybdenum cofactor biosynthesis, C1-unit interconversion, and siroheme biosynthesis.

MAGs 15 and 22

Both samples MAG_15 and MAG_22 were identified as *Gemmatirosa kalamazoonesisa*. This bacterium was not previously identified in the *Cinachyrella* sponge, and therefore attracts interest. *Gemmatimonadetes*, a poorly described phylum that has not received much attention so far. It has been detected primarily in low moisture soil, making up 2% of the total soil bacteria (DeBruyn et al., 2011). The genome was previously isolated from soil. However, one recent study found this bacterium in lake environments (Cabello-Yeves et al., 2017). Previously sequenced genome of *Gemmatirosa kalamazoonesis* has a size of 5.32 Mb with 4,558 genes an N50 equal 5,311,527. Both MAGs 15 and 22 showed different quality and sequencing statistics. MAG_22 showed total length of 6,799,620 with N50 equal to 22,940 and 815 contigs. This exceeds the published genome. The genome completeness was assigned to 95.83% with contamination rate of 8.44%. The genome showed two complete ATP synthesis pathways for energy metabolism: Cytochrome c oxidase and F-type ATPase. Interestingly, only two metabolisms of cofactors and vitamins complete pathways were found: coenzyme A biosynthesis and heme biosynthesis. For this MAG many antimicrobial resistance genes were found, including Tetracycline, Macrolide resistance, and a few efflux pump multi-drug resistance modules. MAG_15 showed total length of 3,167,017 and N50 was found to be 15,902 with 319 contigs in total. MAG_15 has a smaller completeness rate 76.32% with a very low contamination rate 1.75%. Even though the completeness is smaller, the MAG will be used to compare the results of the same species. KEGG identified that the majority of the found genes belong to purine metabolism, glycine, serine and threonine metabolism, aminoacyl-tRNA biosynthesis, and the two-component system. The only complete energy metabolism pathway was ATP synthesis (F-type ATPase), which was also found in MAG_22 Moreover, MAG_15 during the analysis only showed one complete cofactor and vitamin metabolism pathway - coenzyme A biosynthesis, which was also identified in the previous MAG. When the antimicrobial resistance

genes were analyzed, it was found that MAG_15 has a few gene variants for tetracycline resistance beta-Lactam, and macrolide resistance. However, KEEG analysis on the previously sequenced genome identified beta-Lactam and Vancomycin resistance.

MAG_27

MAG_27 assigned to the phylum *Verrucomicrobia* (species *Opitutaceae bacterium*). This MAG showed completeness of 91.52% with 5.62% contamination rate. The genome has 934 contigs and the total length of 3,321,892 bp with N50 equal to 4,738. The first sequenced genome was isolated from the wood-feeding termite hindgu with chromosome length of 7,317,842 bp. The genome, which was found in 2015 revealed genes for methylotrophy, lignocellulose degradation, and ammonia and sulfate assimilation (Kotak et al., 2015). Interestingly, *Verrucomicrobia* represent the first phylum outside of the *Proteobacteria* to be characterized as methanotrophs. An analysis of the genome revealed the presence of two complete energy metabolism pathways: carbon fixation (phosphate acetyltransferase-acetate kinase pathway) and ATM synthesis (cytochrome c oxidase and F-type ATPase). KEGG analysis also identified three complete cofactor and vitamin metabolism pathways, including coenzyme A biosynthesis, biotin biosynthesis, and molybdenum cofactor biosynthesis. The genome revealed tetracycline resistance genes, ancomycin resistance modules, and multi drug efflux pump resistance modules.

MAG_1

MAG_1 assigned to the phylum *Chloroflexi* represents the second most abundant genome bin (16.63%), with completeness 84.34% and features a very low contamination rate of 3.18%, as determined based on presence and copies of single copy marker genes. MAG_1 has a total length of 3.2 Mb and N50 of 5.6 kbps. N50 is a measure to describe the quality of assembled genomes. This genome showed to have 576 contigs. Interestingly, MAG_1 also identified S16, which is very rare for a whole genome sequencing. BlastKOALA identified Carbohydrate metabolism and Protein families: signaling and cellular process as the most abundant annotated entries. This MAG identified seven complete carbohydrate metabolism pathways, which was different from MAG_3. MAG_1 included complete Glycolysis (Embden-Meyerhof pathway), Glycolysis (core module involving three-carbon compounds), Citrate cycle, Pentose phosphate

pathway (oxidative phase), PRPP biosynthesis, Galactose degradation, Nucleotide sugar biosynthesis. Moreover, two complete cofactor and vitamin metabolism pathways were identified, including coenzyme A biosynthesis and Molybdenum cofactor biosynthesis. Also, some KEGG annotation showed seven energy metabolism pathways: oxidative phosphorylation, photosynthesis, fixation in photosynthetic organisms, carbon fixation pathways in prokaryotes, methane metabolism, nitrogen metabolism, and sulfur metabolism. Two of these pathways were identified as complete: carbon fixation (crassulacean acid metabolism) and cytochrome c oxidase. Several signal transduction pathways were recognized, such as two-component system, MAPK signaling pathways, HIF-1 signaling pathway, FoxO signaling pathway, Phosphatidylinositol signaling system, Phospholipase D signaling pathway, PI3K-Akt signaling pathway, AMPK signaling pathway. Interestingly, functional analysis identified antimicrobial drug resistance modules including beta-Lactam resistance, vancomycin resistance, and cationic antimicrobial peptide (CAMP) resistance.

MAG_3

MAG_3 assigned to the phylum *Proteobacteria* represents the most abundant phylum and the third most abundant genome bin (4.87%), with very high completeness 97.58% and features a very low contamination rate of 1.12%. The phylum Proteobacteria is the most abundant phylum identified in this study and was supported in other studies as well for *Cinachyrella* sponge. MAG_3 has a total length of 3.7 Mb and N50 of 2.8 kbps. This genome showed to have 375 contigs. BlastKOALA identified Carbohydrate metabolism and Protein families: signaling and cellular process as the most abundant annotated entries. KEGG annotation showed five complete energy metabolism pathways: carbon fixation (phosphate acetyltransferase-acetate kinase pathway), nitrogen metabolism (Assimilatory nitrate reduction), sulfur metabolism (assimilatory sulfate reduction), and ATP synthesis (Cytochrome c oxidase and F-type ATPase). Moreover, for this bin one additional complete signature module was identified: nitrate assimilation (NRT, narK, nrtP, nasA - major facilitator superfamily of nitrate/nitrite transporters). Also, KEGG showed two complete metabolism pathways of cofactors and vitamins: NAD biosynthesis and Molybdenum cofactor biosynthesis. Moreover, this MAG shows the same antimicrobial drug resistance modules as MAG_1, including beta-

Lactam resistance, vancomycin resistance, and cationic antimicrobial peptide (CAMP) resistance.

DISCUSSION

Cinachyrella sp. microbiome

The sponge-associated microbiota is host-specific (Reveillaud et al., 2014; Thomas et al., 2016). Host identity is a major factor in explaining the microbiome of the sponges as other studies showed before (e.g. Steinert et al., 2017). Thus, this study provides a deeper understanding of the metagenome of an individual *Cinachyrella sp.* sponge. For the analysis we used one sample, and we did not expect to describe the metagenome of divergent *Cinachyrella* taxa. The taxonomic analysis of this sample correlates with previous *Cinachyrella* metagenome research (Cuvelier et al., 2014). On the other hand, this paper presents the first *Cinachyrella* metagenome functional analysis with MAGs generation. Thus, additional samples with other *Cinachyrella* species would be needed to confirm the functional annotation of these sponge individuals, their similarities or differences.

Holobiont

As was discovered previously, microbiome can influence the host's behavior, metabolite production, reproduction, and immunity in either "microbe-specific" or "microbe-assisted" ways (Shropshire & Bordenstein, 2016). The "holobiont" and "hologenome" - interconnected compositions of animal and microbes, concepts gain more attention, since without this whole picture the host itself cannot be studied. Some elements of host immunity genes are in constant competition with microbiota genes. Host immunity needs to manage the symbionts community, while defending pathogenic infections. This coexistence provokes coevolutionary changes between the host genetics and its microbiome (Obbard et al., 2006). Holobionts can be colonized with microorganisms, by contacting with environmental and social sources (Tung et al., 2015), by host genetic and metabolic traits (McKnite et al., 2012), and diet and age (Yatsunenko et al., 2012). Some studies claim that the microbiota is an extra genetic component that stimulates species formation (Brucker & Bordenstein, 2012). All of these factors have to be considered in order to explain the microbial communities of holobionts (Brucker & Bordenstein, 2013). Our

functional analysis established that all of the analyzed MAGs have antimicrobial resistance genes (eg. beta-lactam, trimethoprim, nitroimidazole, aminoglycoside, quinolone, macrolide, fosfomycin, fusidicacid, oxazolidinone, glycopeptide, tetracycline, rifampicin, phenicol, colistin, sulphonamide). This fact might result in the overall hosts coevolutionary changes. Full genome sequencing of the host as well as its functional analysis for these genes would aid in clarifying this observation.

The major system driver

This study indicates that *Proteobacteria* is the major system driver at the phylum level, followed by *Chloroflexi*, *Acidobacteria*, and *Gammaproteobacteria*. Previous studies indicated that the *Cinachyrella* sp. microbiome was dominated by *Proteobacteria, Bacteroidetes,* and *Cyanobacteria* (Cuvelier et al., 2014; Thomas et al., 2016: Baquiran et al., 2020). This finding suggests that *Chloroflexi* bacterium was underestimated in previous studies, since the whole genome analysis indicated major abundance of *Ardenticatena* genome. High abundance of *Chloroflexi* members are more typical of high microbial abundance HMA species (Schmitt et al., 2012). Moreover, previously major community drivers were already identified in other sponge species, and they correlate with found MAGs in *Cinachyrella sp.* sponge metagenome. *Nitrospiraceae* is dominant in *Rhabdastrella globostellata* (Steinert et al., 2016), *Gammaproteobacteria* - in *Petrosia ficiformis* (Burgsdorf et al., 2014), *Betaproteobacteriais* the community driver in *Callyspongia* sp. (Steinert et al., 2016).

Many previous 16S rRNA-based studies found that 'sponge-specific' microbes are often absent from seawater and other (non-sponge) marine habitats (Simister et al., 2012). Similarly, to what we see in our study. Previous comprehensive phylogenetic analyses with 7546 sponge-derived 16S and 18S rRNA sequences discovered that 27% of all sequences identified as monophyletic, sponge-specific sequence clusters. These clusters mainly including *Chloroflexi*, *Cyanobacteria,* '*Poribacteria'*, *Betaproteobacteria*, and *Acidobacteria* (Simister et al., 2012). Another study discovered that sponge *Ectyoplasia coccinea*, showed a relatively high abundance of *Chloroflexi* members, which is more typical of HMA species (Cleary et al., 2020).

Intraspecific Diversity within the Sponge Microbiome

Recruitment plots suggest that different models of bacterial speciation may apply to the same host environment. Recruitment plots can reveal genetic heterogeneity within discrete population. Recruitment of reads can identify sequence-discrete populations (sequence-discrete "clusters") that are not clonal but instead are composed of highly similar, co-occurring genotypes that contain some degree of genetic diversity with highly conserve regions, represented by horizontal lines separated by percent similarity (Caro-Quintero & Konstantinidis, 2012). Moreover, some research suggests that these similar members of the community may have highly similar, if not identical, ecological roles (Caro-Quintero & Konstantinidis, 2012). Each bin identified a different amount of discrete populations (horizontal clusters). Especially the top two bins (MAG_2 and MAG_1) showed high number of these populations. MAG_1 has an average genome size, comparing to others, with lots of vertical and horizontal clusters. It might suggest that this MAG represents a group of discrete populations that have streamlined genome to core functions with each discrete population occupying a niche based on that small regions of unique genomic information. Possibly, this bin represents the most abundant MAG, because it supports the core functions, which are needed in order to survive in this environment, including the metabolism, defense and energy generation mechanisms. It's the most abundant organism with a high degree of genetic heterogeneity are likely to be able to occupy a larger number of niches, since it performs many core roles in the hologenome. If we take into the consideration MAG_1, which is the second most abundant bin, it represents a large genome with a high number of sequence-discrete populations and a very low rate of shared homologous regions with other population. Therefore, the MAG represents a single population with minimal heterogeneity, and is likely responsible for a particular function within this hologenome, since it is so abundant. Moreover, since other taxa has a very low rate of shared homologous regions, this bin might occupy a particular niche in the environment, which is not feasible for others. On the other hand, the third most abundant MAG_3 and MAG_2 didn't show higher intraspecific diversity speciation in the sample. MAG_3 is a large genome, which has only a few sequence-discrete "clusters", might suggest that the bin contains a wide range of functions that are not parsed over subpopulations, which are unique for the community. It might suggest its unique role in the community, as an indispensable community member.

Functional Characterization of MAGs

ABC-transporters are crucial for importing available substrates, such as inorganic and organic ions, mono- and oligosaccharides, amino-acids and peptides. High numbers of purine metabolism assignments represent the susceptibility of early steps in cytokinesis to nitric oxide toxicity. Many two-component systems show the basic necessity for sponge symbionts community members to respond to changing environmental conditions for quick adaptation. Quorum sensing plays a crucial role in bacterial communication and density regulation. However, the exact roles in sponges' metagenome have yet to be described.

Sulfur metabolism

Sulfur is central to life because the element composes proteins and vitamins. Sulfur metabolism is a complex cycle with chemical reactions that can be microbial-mediated. The dissimilation of sulfur compounds can serve as an energy source for some prokaryotes and plays an important role for microbial growth as well (Vavourakis et al., 2019). Hydrogen sulfide and other reduced sulfur compounds generally serve as electron donors for sulfur-oxidizing microorganisms (SOM). Stable intermediary inorganic sulfur compounds are produced during sulfur cycle, for example free dissolved sulfide ($\Sigma H2S = H2S + HS^- + S2^-$) and sulfate ($SO42^-$) are generated during sulfur oxidation from hydrogen sulfide (Fig 2). In general, the majority of the sulfide is deoxidized back to sulfate, by geochemical or microbial reactions. Then, some of the sulfide precipitates with metals or it can react with organic matter.

Sulfate reducing microorganisms (SRM) include a very diverse group of anaerobic microorganisms with catabolic properties of fermentation products (Jørgensen, Findlay & Pellerin, 2019). A recent study shows that 11.3 tera moles of sulfate are reduced to hydrogen sulfide in marine sediments every year (Bowles *et al*., 2014). Recent genomic data from marine and terrestrial subsurface environments have revealed many SRB, which were not previously identified (Anantharaman et al., 2018). Moreover, it states that many of these types of bacteria still have to be identified. Even with a growing interest in the sulfur cycle, the genetic makeup of the involved microbes is not yet fully understood (Vavourakis et al., 2019).

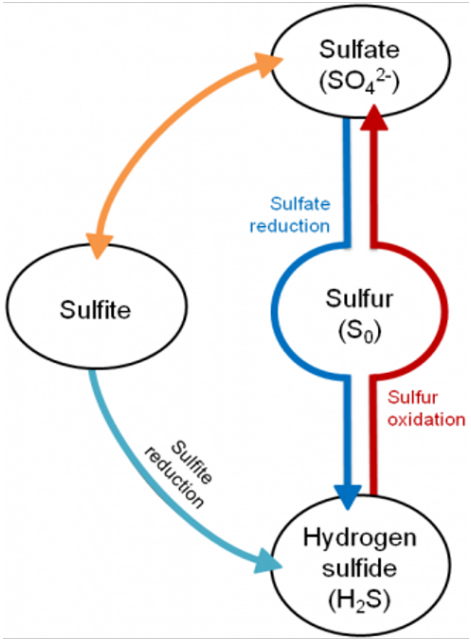


Fig 6. Sulfur cycle. Free dissolved sulfide (ΣH2S = H2S + HS− + S2−) and sulfate (SO42−) are generated during sulfur oxidation from hydrogen sulfide.

Five MAGs are taxonomically affiliated with known sulfur utilizing microbes: *Opitutaceae bacterium*, *Thioalkalivibrio paradoxus*, *Desulfobacterium autotrophicum*, *Thioalkalivibrio sulfidiphilus*, *Sulfurifustis variabilis*, suggesting an important role of sulfur metabolism in sponge-microbiome.

Other research identified microorganisms involved in sulfur cycling by the functional gene *aprA* (adenosine-5′-phosphosulfate reductase) (Jensen et al., 2017). Samples with the *aprA* gene were found to be 87 % affiliated with sulfur-oxidizing bacteria. MAG_5 and MAG_17 showed the presence of this gene. Because more than 75% of the MAGs generated from the sponge microbiome contains PAPSS gene (2.7.7.4), which is the universal sulfonate donor's synthase and the Fig 5. shows that all of the MAGs have some of the sulfur metabolism genes. The findings indicate that assemalatory pathways are highly conservered, oxidataive and reductions pathways ways are less conserved across the MAGs.

Many articles state the importance of dimethylsulfoniopropionate (DMSP) as an osmoprotectant (Yoch D. 2002; Conway et al., 2012). This metabolite is degraded by marine bacteria to dimethylsulfide (DMS). On the other hand, out KEGG analysis did not identify any of the metabolites related to this pathway (dmdA, dddL).

Quorum Sensing

Bacteria–bacteria interactions may play a crucial role in shaping the sponge microbial community. Several sponge-associated bacteria can inhibit the growth of other bacteria through the production of various compounds and regulatory signals (Gutiérrez-Barranquero et al., 2019). Microbes use broad range of defense mechanisms, including physical and chemical strategies to survive in nature. This explains why microbiomes synthesize quorum sensing chemicals with targeted- and broad-spectrum activities. Moreover, detailed metabolic pathway analysis showed abundant prokaryotic defense system, which included CRISPR-Cas system, restriction and modification system, and toxin-antitoxin system. As was discussed previously, relatively little is still known about quorum sensing in sponges. Quorum sensing molecules are involved in the inter-microbial communication with their own species, and also can possibly communicate directly or indirectly with their host (Verbeke et al., 2017). Population density is regulated in four steps: (I) synthesis of signal molecules (autoinducers), (II) secretion of the autoindusers, (III) at a threshold concentration (when the population reached the highest population density possible for optimal survival)a specific receptor is activated and (IV) it activates or suppress gene expression (Yada, 2015). Moreover, the quorum sensing genes are responsible for producing and regulation activities for large number of species. It can include bioluminescence, antibiotic production, biofilm formation, and virulence factors production.

Bacteria use quorum sensing to regulate population density and control collective behavior. Quorum sensing is cell-to-cell communication through the production and release of chemical ques into the surrounding environment. Extracellular signaling molecules called autoinducers are produced and detected. Quorum sensing leads to a group-level response and change of the behavior. Autoinducers accumulate in the community space as the population density increases. Many processes are regulated by quorum sensing, including bioluminescence, the secretion of virulence factors, the formation of biofilms (Papenfort & Bassler, 2016). Quorum sensing plays an important role in successful establishment of symbiotic and pathogenic relationships with eukaryotic hosts (González & Keshavan, 2006)

A few recent articles focused on identification of quorum sensing activators and inhibitors in the marine sponges (Saurav et al., 2020). For example, sponge species Sarcotragus spinosulus expressed both QS signal molecules as well as QS inhibitory (QSI) molecules. Some

articles show that some sponge species like *Sarcotragus sp.* have constant presence of *N*-acyl homoserine lactones (AHLs), quorum sensing molecules mediators (Britstein et al., 2018). Only several studies attempted to examine quorum sensing systems in sponge metagenome (Reen et al., 2019), and therefore more research needs to be done.

It was found that PQS plays a crucial role in adapting the population density and structure while populations adjust to the conditions of increased stress, which are accompanied by bacterial persistence, often a result of chronic infectious diseases (Häussler & Becker 2008). Opp/Ami is an ABC transporter (peptide importer). It plays a crucial role in internalization of different peptides and signal molecules during the population growth phase. This study provides a support of quorum mechanism importance and presence in *Cinachyrella* metagenome, including Opp/Ami gene identification (signaling peptide importer). Previous studies suggest that eukaryotes have evolved defense mechanisms to manipulate bacterial quorum sensing and by doing so protect themselves from pathogenic attack (González & Keshavan, 2006). Therefore, it is possible that this quorum sensing mechanism is used by metagenome to compete with each other for space and food sources. It was also found that some bacteria evolve gene-regulation with incorporates host quorum-sensing signal molecules. Further studies need to examine if it's true for *Cinachyrella* metagenome by analyzing host's genome.

Drug Discovery

Found sponge microbiomes may play a key role in drug discovery. For example, MAG_14, was identified as Haliangium ochraceu. Haliangium ochraceu is known for its ability to produce secondary metabolites. Moreover, this bacterium produces fruiting bodies and develops myxospores, that is based on cell-to-cell signaling among the single cells of the population in a swarm. Interestingly enough, many modules for MAG_14 were assigned as antimicrobial resistant, based on the found antimicrobial resistance gene, which identifies its ability to survive in highly competitive environments. Some of the modules included methicillin, beta-Lactam, vancomycin, tetracycline, and a variety of multi drug resistance of efflux pumps. Priority pathogens by WHO assigned as critical are 1. Acinetobacter baumannii, carbapenem-resistant 2. Pseudomonas aeruginosa, carbapenem-resistant 3. Enterobacteriaceae, carbapenem-resistant, ESBL-producing. Therefore, further studies need to examine if this finding possess the desirable pharmacokinetic properties required for clinical development.

Host genome

Previous attempts at sequencing *Cinachyrella* sample it was found that the majority of the reads belonged to its metagenome, now it would be possible to exclude found MAGs reads and construct its host genome. Therefore, further research needs to analyze *Cinachyrella* holobiont.

CONCLUSIONS

In this study, we report the identification of 27 genomes of sponge associated microbes with at least 50% completeness, and contamination less than 10%. Providing insight into microbial speciation and *Cinachyrella* sponge microbiome. The results identified underestimated in previous studies Chloroflexi bacterium (*Ardenticatena*) as a major community driver.

Results underlined abundant prokaryotic defense system for each MAG, including quorum sensing mechanism and whole metabolic pathways: CRISPR-Cas system, restriction and modification system, and toxin-antitoxin system. Some of the functional roles for key members of the *Cinachyrella* sponge microbiome identified the functional roles in the microbial-sponge symbiosis. Microbial populations exemplified by MAG_14 (*Proteobacteria*) and MAG_27 (*Verrucomicrobia*) probably are of global importance for their host sponges since they were predicted to function in carbon fixation. Moreover, this study provided a detailed *Cinachyrella* sponge microbiome functional characterization.

KEGG analysis showed that 21 out of 27 MAGs have PAPSS gene (2.7.7.4), which is the universal sulfonate donor's synthase. Moreover, detailed pathway analysis identified that MAG_14 and MAG_23 contain complete energy metabolism pathway for sulfur metabolism. Symbiotic bacteria in previous research show elevated expression of genes related to sulfur uptake and metabolism in symbiotic sample compared to the free-living state (Tsukada et al., 2009). Our research underlined abundance of sulfur metabolism genes, what represents an important trait in symbiotic bacteria.

Three identified MAGs (5, 22, 23) showed many complete pathways for metabolism of cofactors and vitamins were recognized, including NAD biosynthesis, Coenzyme A biosynthesis, molybdenum cofactor biosynthesis, and siroheme biosynthesis. This finding suggests an importance of these mechanisms in survival of the microbial symbionts. Previous research

suggests that vitamins and cofactors produced by diverse symbionts could be beneficial to the sponge host (Thomas et al., 2010). This finding correlates with other research and suggests that the sponges' nutrition is supplemented by symbiont-derived vitamins and cofactors (Bayer et al., 2018). Also, many modules were assigned as antimicrobial resistant, which might have notable potent therapeutic activities, and have the desirable properties required for drug development.

APPENDIX 1

Supplemental Tables

Supplemental Table 1. Sequencing Statistics.

Supplemental Table 2. Quality Statistics for all 69 sequenced samples (SAPs).

Supplemental Table 3. Quality Statistics for samples, which were marked with "good quality" and "good completeness" flags. MAGs that were at least 50% complete with contamination less than 10% were used, which resulted in 27 good quality MAGs. MAGs are sorted based on the completeness.

Supplemental Table 4. Sequenced samples names and their MAGs number

Supplemental Table 5. Functional potential of reconstructed MAGs with KEGG pathways. Functional assignments to KEGG categories revealed 282 functional categories.

Please refer to the Additional Files section in this thesis' record for supplemental data tables

Supplemental Code

Supplemental Code 1. Speciation

```
$git clone https://github.com/lh3/bwa.git
$cd bwa; make
$./bwa index *.fasta.fna
#### Aligning
$bwa mem MAG.fa read1.fastq read2.fastq -t
#threads alignment.sam
```

####Converting sam to bam:

```
$wget https://github.com/samtools/samtools/releases/download/1.9/samtools-1.9.tar.bz2
$tar -vxjf samtools-1.9.tar.bz2
$cd samtools-1.9
$make
$export PATH=$PATH:/YourPath/samtools-1.10
$samtools view -bS *.sam > alignment.bam
$for z in *.sam; do samtools view -bS $z > ${z/.sam/.bam}; done
```

####Sorting alignment

```
$for z in *.sam; do samtools sort -o $z ${z/.bam/sorted.alignment.bam}; done
```

####Indexing the alignment

```
$samtools index *.alignment.bam
```

####Geting alignment stats:

```
$samtools flagstat sorted.alignment.bam
```

Supplemental Code 2. Creating Recruitment plots

Linux

####converting multi-sequence fasta SAP reference genome into concacinated single sequence reference

$cat SAP2.001.fasta.fna | grep -v '^>' | grep '^.' | tr -d '[:blank:]' | tr -d '\n' | cat <(echo '>seq_name') - > SAP2.001_concat.fasta

####converting single reference into blast db
$makeblastdb -in SAP2.001_concat.fasta -dbtype nucl -out SAP2.001.ref

####blasting
$blastn -db SAP2.001.ref -query sub1.fa -out SAP2.001vsub1 -outfmt 6 -max_target_seqs 1

R
Load libraries
>library(ggplot2)

Import data
>dat <- read.table("SAP2.001vsub1")

selecting data with bitscore greater than 50
>dat <- dat[which(dat$V12 > 50),]

ploting genome position on x and %ID on y
>ggplot(dat, aes(x=V9, y=V3)) + geom_point(alpha=0.1, color="steelblue") + theme_bw() + xlab("Genome Position") + ylab("% Identity") + theme(text = element_text(size=20))

APPENDIX 2

Assemblies:
https://drive.google.com/drive/folders/13lj9zCSE4IYHhVcKMKBNevsxWKjDkN0C?usp=sharing

APPENDIX 3

Draft manuscript in PeerJ format.

Metagenomes of the Dominant Microbial Symbionts of Sponge Genus *Cinachyrella* Display Common Sulfur Metabolic and Quorum Sensing Functions.

Elizaveta Shmakova1, Cory Cruz Padilla3, Julie Torruellas Garcia 4, Jose Victor Lopez2,
1 Halmos College of Arts and Sciences, Nova Southeastern University, Dania Beach, Florida, USA
2 Halmos College of Arts and Sciences, Nova Southeastern University, Dania Beach, Florida, USA
3 School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA, USA
4 Halmos College of Arts and Sciences, Nova Southeastern University, Fort Lauderdale, Florida, USA
Corresponding Author:
Elizaveta Shmakova1
713 SW 7th St, Dania, Fl, 33004, USA
Email address: es1542@mynsu.nova.edu

Abstract

Host associated microbes play important roles in animal health. The symbiotic relationship between sponges and the microorganisms, prokaryotes-in particular, forms a complex mutualistic system, or community called the “microbiome”. Prokaryotes play a dynamic roll in these symbioses including providing nutritional sources, digestion, development, metabolism, immune defenses, while host sponge provides an environment for the symbionts. However, the whole metabolism of a sponge, holobioint, and the symbiont diversity and respective contributions to the holobtiont remains unclear. Metagenomics is an approach to study these systems by analyzing the gene content of the microorganisms for both taxonomic

affiliation and biochemical function. The primary objective of this study is to describe local reef sponge *Cinachyrella* microbiome diversity, composition, and functional roles they occupy. Sponge metagenome sequences were assembled and binned into twenty-six high-quality Metagenomically Assembled Genomes (MAGs) and were taxonomically and functionally characterized. These two MAGs have been identified as *Ardenticatena* (Phylum Chloroflexi), and *Thioalkalivibrio* (Phlyum Proteobacteria). The MAG characterization includes inter/intraspecific diversity, composition, and functional roles they occupy. These two MAGs represent over 20% abundance of all sequence reads, suggesting their important functions for the holobiont. A MAG matching Thioalkivibrio has been shown to occur in other sponge symbioses. Conserved function in several genomes included quorum sensing and sulfur metabolism.

Introduction

Microbiome plays an important role in the host life, including providing nutritional sources, digestion, development, metabolism, and immune defenses. Many host-prokaryotic interactions are classified as symbiotic. This can be contrasted with commensalism or pathogenesis. Symbiosis is a partnership between two or more organisms, which may, but does not necessarily benefit each member. Microorganisms are fundamentally important to many aspects of a host's phenotype, adaptations, and evolution. A host with all of its microbial symbionts was named "holobiont" by Lynn Margulis in 1991 (Gilbert, Sapp & Tauber, 2012). . The importance of holobiont is evident in many studies of host metabolism, development, immunity and many other processes (Lee & Mazmanian, 2010; McCutcheon & von Dohlen, 2011; McFall-Ngai et al., 2011). A study by McCutcheon & von Dohlen (2011) found that hosts and their microbial community (also known as "microbiomes") genomic data showed remarkable levels of metabolic complementary functional roles between the bacterial symbionts of several insect lineages. Moreover, host-microbiome relationships provide the holobiont with many other advantages including evolutionary adaptation to parasites (Rigaud & Juchault, 1993). The hologenome theory of evolution explains the importance of the host-microbiome interactions, emphasizing the importance of phenotypic selection of the host and microbial members of the holobiont. It was discovered that community heritability and phenotypic

characteristics of the holobiont can be analyzed by its hologenome (Shropshire & Bordenstein, 2016).

At a microscopic level trillions of microorganisms of thousands of species live on the surface and inside of our body and other live creatures. However, because of their small size these microorganisms might make up only a small percent of the host's the body mass. For example, Human Microbiome Project (HMP) calculated that microorganisms occupy 1 to 3 percent of the body's mass (Lloyd-Price et al., 2017). The microbial network includes bacteria, fungi, parasites, and viruses. Each organism has a unique community coexisting with the host. Nobel Laureate and Microbiologist, Joshua Lederberg, first coined the term "microbiome" in 2001 (Prescot, 2017).

In order to analyze metagenomes, culture-independent techniques have to be used. High throughput (HT) DNA sequencing and informatics are just a tool to understand microbiomes and reference genomes (derived as whole genomes derived from a culturable bacterium from the same habitat). Recently, improved sequencing technology and coverage have enabled reconstruction of fragments into metagenome-assembled genomes (MAGs) by process of assembly and "binning"(Albertsen et al., 2013).

When applied to environmental samples, WGS data can assembled into contiguous genomic units (contigs) and the contigs can be further clustered to into metagenome-assembled genomes (MAGs) (Albertsen et al., 2013). These genomic fragments of varying completeness reflect both the taxonomic and functional diversity contained in a microbiome. Comparing these assembled sequences to known genomes references (derived as whole genomes derived from a culturable material) enables the assignment of these environmetnal sequences to known functions. Of particular interested is the recent ability to produce MAGs by binning contigs by similar abundance and tetra-nucleotide frequency profiles (Wu et al., 2014). MAGs can be viewed as a proxy or substitute for reference genomes defined above but not necessarily culture derived and more based on bioinformatics.

The Earth Microbiome Project (EMP, http://www.earthmicrobiome.org) established a framework to catalogue microbiota globally. This endeavor provides a reference database of global DNA context and a framework for integrating data from future studies, encouraging complete characterization of Earth's microbiome diversity on this planet (Thompson et al., 2017). The project promised to analyze 200,000 samples from diverse communities with collaborative effort using sequencing, metagenomics, and metabolomics. This will help to generate a global Gene

Atlas describing protein space, a global metabolic model, and a data-analysis portal for visualization of processed information. The results of the EMP and similar consortia have accelerated our understanding of microbial taxonomic and functional diversity and how they interact with their surrounding environments, including host interactions. Microbial communities live in constant interaction with other organisms, whether it be with the host or a different microbe. One of these types of interaction is parasitism, this microorganism is harmful to other organisms (Wilkins et al., 2019). Some bacteria depend upon other organisms but can also be beneficial to these organisms, or neutral - symbiotic bacteria. Symbiotic relationships are present if some microorganisms help support the life of others. Usually in symbiotic relationships substances produced by one microorganism in the process of metabolism are used or consumed by other microorganisms.

Sponges and their microbiome are an exemplar in showcasing hologenome functionality. This symbiotic relationship is considered to be a major contributor to the evolutionary success of sponges (Hentschel et al., 2012). Marine sponges are known for hosting dense and highly diverse microbial communities. The specific characteristics of marine habitats are reflected in the colossal variety of physiological capabilities of marine microorganisms. The coexistence of sponges and the sponge associated prokaryotes (SAPs) that inhabit them (against the background of changing external conditions) lead to the formation of many interactions between them. These interactions ultimately affect the uniqueness of their metabolism. Microbes are the driving force of many biogeochemical cycles, such as carbon, nitrogen, sulfur cycles (Falkowski, Fenchel & DeLong, 2008). Up to 40% of sponge mass consists of microbes (Schmitt et al., 2012). A “core microbial community” is when major microorganisms inhabiting a specific host. It was found in recent studies that a “core microbial community” has the most abundant microbial community in phylogenetically distant hosts (Hentschel et al., 2002). It should be mentioned that sponges also can be described by its “pan microbial genome", which consist of its ‘core genome’ with all the genes present in all strains, a ’dispensable genome’ with genes present in two or more strains, and genes which are unique to single strains (Tettelin et al., 2005). A study by Thomas et al. (2016) compared 81 sponge species and found 39 predominant microbial phyla. This study found few shared features in species composition across the phylum. Furthermore, symbionts were characterized as specialists and generalists. It suggests that the microbiome should provide many functional roles for the host. However, many functional roles are still unclear. There needs to be

a comprehensive study of microbial associates of sponges, which can help determine the possible functions of their coexistence.

This study assesses the microbiome community of *Cinachyrella* sponge with the aims to characterize their functional role within the host. At least four, and perhaps up to ten, *Cinachyrella* species (*Cinachyrella kuekenthali* 1-4, *C. alloclada1*/2 & 3, *C. apion* 1& 2, and *C. arenosa*) occur or overlap in the Western Atlantic area (Cárdenas & Schuster, unpublished results; Schuster et al, 2017). Several traits support the genus for use as an experimental system: this sponge can be maintained for weeks and months in aquaculture, can reproduce via viviparous propagation or asexually (albeit at irregular times and unknown cues), and appears resistant to fouling. To date, a draft *Cinachyrella* transcriptome (Smith 2013), metagenome and multiple microbiomes from various individuals have been sequenced and characterized. *Cinachyrella* sp. was studied in the field and in laboratory experiments to investigate changes in holobiont physiology and microbial community structure in response to stressors such as crude oil and antibiotics (Baquiran et al., 2020). Electron microscopy of the holobiont ultrastructure has revealed both low and high microbial abundances in various *Cinachyrella* sp. For example, *Cinachyrella kuekenthali* and perhaps other congeners could be considered high microbial abundance (HMA) sponges via TEM. A recent study has shown that the presence of mitochondrial group I introns has divided sympatric *Cinachyrella* individuals in Florida into at least two *C. alloclada* species; there is also a strong correlation of intron presence/absence with divergence into two distinct microbiomes (group 1 and 2) of these same individuals. This finding suggests that host genetics (i.e. host divergence) can have a strong influence on microbiome structure, via currently unknown factors, even for visually identical, sympatric sponges which appear to have slight genetic differences.

It is known that the host phylogeny can have a strong influence on microbiome structure and its genetics (Lutz et al., 2019; Youngblut et al., 2019). More specific questions regarding *Cinachyrella* sp. and other potential model sponges focus on the specificity of their microbial symbiont communities (microbiomes), and these potential effects on holobiont divergence and speciation. This characterizes the total microbial diversity through 16S sequencing and through the generation of MAGs assess the functional capacity of bacterial genomes found in *Cinachyrella* sp. host by using metagenomic methods. An underlying thread of the research addresses the question of whether there is an underlying “core” metabolism of this *Cinachyrella*

sponge species. Statistical comparison of the genomic repertoire of sponge symbionts with reference genomes will provide information about the involved genes and their functional roles within the host. A deeper understanding of these functions will provide greater insight into the involved molecular processes that underlie microbial-sponge symbiosis.

Materials & Methods

Sample Collection

Cinachyrella sponges studied in this project were originally collected as part of a sponge whole genome sequencing project in collaboration with Dovetail Genomics of California. They would apply shotgun, Chicago, and/or Dovetail Hi-C sequencing libraries. Focus was therefore on just a few sponge individuals collected off the coast of Dania Beach FL.

DNA extraction

Cells in the X fraction from the percoll isolation were selected as the Sponge Associated Prokayotes (SAPs). SAP DNA was isolated with the DNeasy PowerSoil Kit (Qiagen) as recommended by Earth Microbiome Project. DNA extracts were stored at -20°C until downstream analysis.

16S rRNA amplicon sequencing and analysis

The hypervariable V3-V4 region of the 16S rRNA gene was amplified and sequenced to assess SAP taxonomic composition on diversity metrics. Library preparation and analysis followed the workflow described in Padilla et al., 2015. Briefly, amplicons were synthesized using Platinum® PCR SuperMix (Life Technologies) with primers F515 and R806 (Caporaso et al., 2011). Both forward and reverse primers were barcoded and appended with Illumina-specific adapters according to Kozich et al. (2013). PCR reactions were carried out with 1 ng of DNA. Thermal cycling involved: Denaturation at 94°C (3 min), followed by 30 cycles of denaturation at 94°C (45 s), primer annealing at 55°C (45 s) and primer extension at 72°C (90 s), followed by extension at 72°C for 10 min. Amplicons were analyzed by gel electrophoresis to verify size (~400 bp, including barcodes and adaptor sequences) and purified using the QIAquick PCR purification (Qiagen). Amplicons were sequenced on an Illumina MiSeq using a 500 cycle Nano kit with 5% PhiX to increase read diversity.

Amplicons were analyzed using QIIME (Caporaso et al., 2010). Barcoded sequences were de-multiplexed and trimmed (length cutoff 100 bp) and filtered to remove low quality reads (average Phred score <25) using Trim Galore!. Paired-end reads were then merged using FLASH (Magoč and Salzberg, 2011), with criteria of average read length 250, fragment length 300, and fragment standard deviation 30. Chimeric sequences were detected by reference-based searches using USEARCH (Edgar, 2010). Identified chimeras were filtered from the input dataset, and merged non-chimeric sequences were clustered into Operational Taxonomic Units (OTUs) at 97% sequence similarity using open reference picking with the UCLUST algorithm (Edgar, 2010) in QIIME. Taxonomy was assigned to representative OTUs from each cluster using the SILVA database. Data were visualized with the R package Phyloseq.

SAP Metagenome sequencing and assembly

A single shotgun library was generated from the same DNA extract that was used to generate the 16S rRNA data. The library was prepared using the NEBNext ® Ultra II DNA library Prep (New England Biolabs) as described above. This library was sequenced on an Illumina NextSeq resulting in 2 X 150 bp read pairs. Sequence statistics can be viewed in Supplemental Table 1.

The metagenomic assembly was generated after adaptor trimming and quality filtering. SPAdes (v 3.11.1) was used to assemble the short reads into contigs using the –meta flag (Nurk et al., 2013). Contigs were then clustered into metagenomically assembled genomes (MAGs) following the avin'o workflow (Eren et al., 2015). MAG contamination and completeness were assessed with checkm (Parks et al., 2015).

SAP MAG analysis

The resulting MAGs were annotated to assess both taxonomic affiliation and gene content. The tool Prodigal (PROkaryotic DYnamic programming Gene-finding ALgorithm) was used to predict open reading frames (ORFs) for each MAG (Hyatt, 2010). This program showed a greater sensitivity in identifying existing genes. It was designed to minimize the number of false positive predictions.

Annotation

Contigs in each genome cluster were annotated using PROKKA (v.1.7), which produces standards-compliant output files (Seemann, 2014). The software uses preassembled genomic DNA sequences in FASTA format, a set of scaffold sequences produced by *de novo* assembly software. *De novo* method does not rely on the reference genome, instead it predicts the genome structure based on the statistical models (Al-Nakeeb, Peterson & Sicheritz-Pontén, 2017). To decrease running time on multicore computers PROKKA uses parallel processing. The comparison between RAST, xBase2, and PROKKA on an *E.coli* genome showed that Prokka produced an overall better annotation (Seemann, 2014).

Coverage

The coverage analysis is a very straight forward and unsophisticated approach. It is quite reliable and hard to misinterpret or refute. The coverage per bin was calculated by mapping the number of reads back to the genome, per size of the genome. The mapped stats were calculated with samtools. To find # of reads per Mbp of genome, mapped number of reads (12244319) were divided by MAG size in Mbp.

LCA Taxonomic assignment of SAP MAGs

To deduce the taxonomic profiles of the SAP MAGs, a lowest common ancestor (LCA) was applied per each MAG. ORFs generated during gene prediction were queried against the NCBI-nr database (BLASTn). BLAST matches with a bit score > 50 were considered for further analyses. The resulting matches were processed through MEGAN (V6.18.5) for LCA taxonomic assignment (Huson et al., 2007; Hudson et al., 2011).

Functional analysis

Metabolic reconstruction of MAGs was performed by mapping of MAG-encoded gene products to KEGG pathways/modules. Metabolic pathways were predicted with the Kyoto Encyclopedia of Genes and Genomes (KEGG) tool BlastKOALA (Kanehisa, Sato & Morishima, 2016). Mapping of gene products encoded by assembled genes to KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways revealed 282 functional categories. Each functional category for each MAG was added to Supplemental Table 2.

Intraspecific Diversity within the Sponge Microbiome

Population structure for the community drivers was assessed using recruitment plots (Konstantinidis and DeLong, 2008). The most abundant MAGs were assigned for genetic heterogeneity through the recruitment of sequence data to the binned genomes. To determine the extent of collapsed heterogeneity, trimmed and merged sequence data were sub-sampled to 1 million reads and aligned to the reference genomes through BLASTn, retaining the top match and requiring a bit score > 50. Recruitment plots, per MAG, were generated by plotting the percent identity against the alignment position in the reference genome. Counting for the amount of genomic information ensures that abundance is scaled to each genome size, so the data won't be skewed toward the longest MAG.

Results

Sequencing Results

In order to acquire genome sequence information of abundant *Cinachyrella* symbiotic microorganisms, contigs assembled from metagenome reads were binned to MAGs (Metagenomically Assembled Genomes) providing the basis for further characterization of abundant sponge microbiome members and reconstruction of the important characteristics of their metabolism. A total of 69 MAGs were generated, with mean completeness of 62.62 % and mean length of 5.15 Mbp. MAG size equals to the number of bases in the MAG. Normalizing for MAG size is to ensure that the abundance is not skewed by the size of the genetic information contained in the MAG. Further analysis was extended to MAGs passed acceptable qualities for binned genomes as defined by the GOLD (REF) standards (>50% complete; <10% contamination), which resulted in 27 good quality MAGs. At least 42 samples with poor overall sequence quality were excluded due to high contamination (more that 10%) or poor sample completeness (less that 50%) The highest number of contigs, 1419, belongs to MAG8. The longest MAG47 has length of 6,799,620 bp. The average length of all good quality MAGs is 3.84 Mbp. Also, the average of the minimum contig length needed to cover 50% of the genomes was found to be 783,863 for all good quality MAGs. Moreover, the average N50, the minimum contig length needed to cover 50% of the genome, is 18.056kbps (Fig. 1C).

Unraveling the Sponge's Microbial Community Composition

To deduce the taxonomic profiles of the sponge *Cinachyrella* microbiomes analyzed, the bioinformatics tool MEGAN V6.18.5 for metagenomic sequence classification was applied. In total 27 MAGs were classified, identifying 8 bacterial phyla (Table 1). The following phyla with descending number of MAGs were identified: *Proteobacteria*, *Chloroflexi, Thaumarchaeota, Gemmatimonadetes, Nitrospirae*, *Verrucomicrobia, Actinobacteria*, and *Acidobacteria*. Taxonomic classifications for all MAGs are provided in Table 1. Overall, the composition of this microbiome is in accordance with microbiomes from other *Cinachyrella* sponge studies, which are dominated by *Proteobacteria* and *Actinobacteria* (Cleary et al., 2019; Cuvelier et al., 2014). However, the majority of the species were only analyzed and described with 16S data (Cuvelier et al., 2014)

Five MAGs show that found species previously were indicated having sulfur metabolism activity: *Opitutaceae bacterium*, *Thioalkalivibrio paradoxus*, *Desulfobacterium autotrophicum*, *Thioalkalivibrio sulfidiphilus*, *Sulfurifustis variabilis*.

Abundance

The relative contribution of the MAGs to the entire metagenomic pool was assessed to determine abundance and ecologic impact on the microbiome system. These abundances ranged from 0.067% to 34.34% per Mb of genome (Table 2). The major driver of the community is MAG_2 with 34.34% real-time abundance, which is *Chloroflexi* (*Ardenticatena*). It's a small size genome (0.48 Mb) with relatively small number of reads mapped (3803428). The second most abundant genome in the total dataset is MAG_1. It showed 16.64% abundance per Mb of genome. MAG_1 represents Proteobacteria (*Thioalkalivibrio paradoxus*). Similarities were observed when we compared taxonomic patterns of the MAG relative abundance with previously published 16S rRNA gene amplicon-based surveys.

Intraspecific Diversity within the Sponge Microbiome

Recruitment plots can reveal genetic heterogeneity within discrete population. Recruitment of reads can identify sequence-discrete populations (sequence-discrete “clusters”) that are not clonal but instead are composed of highly similar, co-occurring genotypes that contain some degree of genetic diversity with highly conserve regions, represented by horizontal lines separated by percent similarity. The most abundant three MAGs, based on the relative

abundance, where selected for speciation analysis. Recruitment plot for MAG_ 2, LCA classified as *Thioalkalivibrio paradoxus* (Fig. 2 A), consists of several sequence-discrete population with high similarity to the MAG (>95 percent identity). The plot identified at least three sub-populations with high percent identity (horizontal pile-up of reads with each distinct cluster >97 percent identity). The second most abundant MAG_1 – *Ardenticatena* (*Chloroflexi*) also consists of a high degree of genetic heterogeneity (Fig. 2 B). At least 5 sequence-discrete "clusters" (> 95 %ID) were associated with the MAG. This bin identified the highest number of discrete populations. The third most abundant sample MAG_3 *Nitrosomonas communis* (also *Proteobacteria*), identified only one sequence-discrete "cluster" (> 95 %ID) in the metagenome sample (Fig. 2 C).

Roughly 65% of the genome MAG_2 consists of regions that share a high degree of similarity below 95 %ID, derived of sequence fragments from other taxa. These sequences are not associate with the population-level genome, suggesting a high rate of shared homologous regions with other taxa. On the other hand, a completely different pattern appears for MAG_1. Only a few reads cluster at genomic position in MAG_1. With less than 15% of the genome consisting of homologous regions shared across non-related (<95% ID) taxa. Interestingly, conserved regions of MAG_3 were in low richness as its cluster density. Meaning that for this MAG represents a unique population with low number of closely related populations whose reads mapped with <90% identity.

Functional Characterization of MAGs

To describe the functional potential of reconstructed MAGs, metabolic reconstruction was performed by mapping MAG-encoded gene products to KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways. Functional assignments to KEGG categories revealed 282 functional categories (see Supplement table 2). Central metabolism was shared functional pathway. Mapping of gene products encoded by MAGs to KEGG pathways revealed that the categories ABC transporters, ribosome, purine metabolism, oxidative phosphorylation, aminoacyl-tRNA biosynthesis, two-component system, glycine, serine, and threonine, pyruvate metabolism, and quorum sensing received the most assignments (see Supplement table 2). The category 'ATP-binding cassette (ABC) transporters' is the most abundant category suggesting that basic metabolism is of prominent importance for sponge microbiome analysis.

Sulfurous compounds play essential roles in both metabolic and catabolic biochemical pathways. Assimilating sulfur as biomass occurs through amino acids, vitamins, secondary metabolite generation. Additionally, many prokaryotes use reduced sulfur as electron sources to generate energy and assimilate inorganic carbon. In total 27 MAGs showed 270 sulfur metabolism genes. More than 75% of the MAGs generated from the sponge microbiome contains PAPSS gene (2.7.7.4), which is the universal sulfonate donor's synthase. It is one of the major player in both dissimilatory and assimilatory sulfate reduction reactions. The Fig. 3 shows that all of the MAGs have some of the sulfur metabolism genes.

Nineteen out of 27 MAGs have Opp/Ami gene, which is an oligopeptide transport system substrate-binding protein involved in the internalization of signaling peptides. Eighteen out of 27 MAGs show the presence of PhnB and Sec genes (Fig. 4). PhnB gene codes for the glutamine-binding beta subunit (PhnB) of anthranilate synthase (AS) provides the glutamine amidotransferase activity to produce anthranilate, a precursor for *Pseudomonas* quinolone signal (PQS). PQS catalyses a bacterial anti-stress response.

Individual MAGs analysis

MAG_1

MAG_1 assigned to the phylum *Chloroflexi* represents the second most abundant genome bin (16.63%), with completeness 84.34% and features a very low contamination rate of 3.18%, as determined based on presence and copies of single copy marker genes. MAG_1 has a total length of 3.2 Mb and N50 of 5.6 kbps. N50 is a measure to describe the quality of assembled genomes. This genome showed to have 576 contigs. Interestingly, MAG_1 also identified S16, which is very rare for a whole genome sequencing. BlastKOALA identified Carbohydrate metabolism and Protein families: signaling and cellular process as the most abundant annotated entries. This MAG identified seven complete carbohydrate metabolism pathways, which was different from MAG_3. MAG_1 included complete Glycolysis (Embden-Meyerhof pathway), Glycolysis (core module involving three-carbon compounds), Citrate cycle, Pentose phosphate pathway (oxidative phase), PRPP biosynthesis, Galactose degradation, Nucleotide sugar biosynthesis. Moreover, two complete cofactor and vitamin metabolism pathways were identified, including coenzyme A biosynthesis and Molybdenum cofactor biosynthesis. Also, some KEGG annotation showed seven energy metabolism pathways: oxidative phosphorylation,

photosynthesis, fixation in photosynthetic organisms, carbon fixation pathways in prokaryotes, methane metabolism, nitrogen metabolism, and sulfur metabolism. Two of these pathways were identified as complete: carbon fixation (crassulacean acid metabolism) and cytochrome c oxidase. Several signal transduction pathways were recognized, such as two-component system, MAPK signaling pathways, HIF-1 signaling pathway, FoxO signaling pathway, Phosphatidylinositol signaling system, Phospholipase D signaling pathway, PI3K-Akt signaling pathway, AMPK signaling pathway. Interestingly, functional analysis identified antimicrobial drug resistance modules including beta-Lactam resistance, vancomycin resistance, and cationic antimicrobial peptide (CAMP) resistance.

MAG_3

MAG_3 assigned to the phylum *Proteobacteria* represents the most abundant phylum and the third most abundant genome bin (4.87%), with very high completeness 97.58% and features a very low contamination rate of 1.12%. The phylum Proteobacteria is the most abundant phylum identified in this study and was supported in other studies as well for *Cinachyrella* sponge. MAG_3 has a total length of 3.7 Mb and N50 of 2.8 kbps. This genome showed to have 375 contigs. BlastKOALA identified Carbohydrate metabolism and Protein families: signaling and cellular process as the most abundant annotated entries. KEGG annotation showed five complete energy metabolism pathways: carbon fixation (phosphate acetyltransferase-acetate kinase pathway), nitrogen metabolism (Assimilatory nitrate reduction), sulfur metabolism (assimilatory sulfate reduction), and ATP synthesis (Cytochrome c oxidase and F-type ATPase). Moreover, for this bin one additional complete signature module was identified: nitrate assimilation (NRT, narK, nrtP, nasA - major facilitator superfamily of nitrate/nitrite transporters). Also, KEGG showed two complete metabolism pathways of cofactors and vitamins: NAD biosynthesis and Molybdenum cofactor biosynthesis. Moreover, this MAG shows the same antimicrobial drug resistance modules as MAG_1, including beta-Lactam resistance, vancomycin resistance, and cationic antimicrobial peptide (CAMP) resistance.

MAG_23

MAG_23 assigned to the phylum *Proteobacteria* species *Rhizobiales* bacterium NR represents the largest N50 score. N50 is a measure to describe the quality of assembled genomes. MAG_21 has the highest N50 of 85,829 with 97 contigs. MAG_23 is the largest found genome, which has a total length of 4,004,567bp. The completeness of the genome bin was assigned to 79.47% and features an extremely low contamination rate of 0%, as determined based on presence and copies of single copy marker genes. The average GC-content is 58.353% (std. deviation 3.599) and GC-ratio 0.720 (std.deviation 0.111). The phylum *Proteobacteria* is amongst the most abundant genera identified in the metagenome (Table 2). *Rhizobiales* bacterium is known for its ability to fix nitrogen when in symbiosis with leguminous plants and pathogenic bacteria to animals and plants. Hence, they are of interest for application in agriculture and plant cultivation concepts.

Interestingly, MAG_23 also identified S16, which is very rare for a whole genome sequencing. KEGG annotation showed three complete energy metabolism pathways: Methane metabolism, sulfur matabolism, and ATP synthesis. MAG_23 uses methane oxidation, where methanotroph and methane oxidizes to formaldehyde. For the sulfur metabolism it showed assimilatory sulfate reduction, in which the sulfate reduces to H2S. On the other hand, complete nitrogen metabolism wasn't identified. KEGG identified 9 genes coding for nitrogen metabolism enzymes, including nitric oxide reductase subunit B (norB), formamidase, and nitrite reductase (NADH) large subunit (nirB). Moreover, a few complete pathways for metabolism of cofactors and vitamins were recognized, including NAD biosynthesis (from aspartate to NAD), Coenzyme A biosynthesis from pantothenate, molybdenum cofactor biosynthesis from GTP, and siroheme biosynthesis from glutamate. During the gene analysis, tlyC putative hemolysin (K03699) was identified, which is a pore-forming toxin.

GhostKOALA showed that the majority of the functional genes were responsible for general metabolism, including carbohydrate metabolism, protein families: genetic information processing, protein families: signaling and cellular processes, and amino acid metabolism. Also, MAG_23 had an abundant prokaryotic defense system, which included CRISPR-Cas system, Restriction and modification system, Toxin-antitoxin system (TA system), and DNA phosphothiolation system. ResFinder-3.2 Server analysis showed no acquired antimicrobial resistance gene to any of the following: beta-lactam, trimethoprim, nitroimidazole,

aminoglycoside, quinolone, macrolide, fosfomycin, fusidicacid, oxazolidinone, glycopeptide, tetracycline, rifampicin, phenicol, colistin, sulphonamide.

MAG_14

MAG_14 assigned to the phylum *Proteobacteria* (species Haliangium ochraceum DSM 14365) represents the most complete genome bin (98.29%) and features a very low contamination rate of 3.6%. The phylum *Proteobacteria* is the most abundant phylum identified in this study and was supported in other studies as well for *Cinachyrella* sponge. MAG_13 is one of the longest found genomes with length 5.9 Mb and relatively high N50 (33,426). This genome showed to have 342 contigs. KEGG annotation showed three complete energy metabolism pathways: carbon fixation (phosphate acetyltransferase-acetate kinase pathway), sulfur metabolism (assimilatory sulfate reduction), and ATP synthesis. Also, six complete cofactor and vitamin metabolism pathways were identified, including NAD biosynthesis, coenzyme A biosynthesis, biotin biosynthesis, molybdenum cofactor biosynthesis, C1-unit interconversion, and siroheme biosynthesis.

MAGs 15 and 22

Both samples MAG_15 and MAG_22 were identified as *Gemmatirosa kalamazoonesisa*. This bacterium was not previously identified in the *Cinachyrella* sponge, and therefore attracts interest. *Gemmatimonadetes*, a poorly described phylum that has not received much attention so far. It has been detected primarily in low moisture soil, making up 2% of the total soil bacteria (DeBruyn et al., 2011). The genome was previously isolated from soil. However, one recent study found this bacterium in lake environments (Cabello-Yeves et al., 2017). Previously sequenced genome of *Gemmatirosa kalamazoonesis* has a size of 5.32 Mb with 4,558 genes an N50 equal 5,311,527. Both MAGs 15 and 22 showed different quality and sequencing statistics. MAG_22 showed total length of 6,799,620 with N50 equal to 22,940 and 815 contigs. This exceeds the published genome. The genome completeness was assigned to 95.83% with contamination rate of 8.44%. The genome showed two complete ATP synthesis pathways for energy metabolism: Cytochrome c oxidase and F-type ATPase. Interestingly, only two metabolisms of cofactors and vitamins complete pathways were found: coenzyme A biosynthesis and heme biosynthesis. For this MAG many antimicrobial resistance genes were found,

including Tetracycline, Macrolide resistance, and a few efflux pump multi-drug resistance modules. MAG_15 showed total length of 3,167,017 and N50 was found to be 15,902 with 319 contigs in total. MAG_15 has a smaller completeness rate 76.32% with a very low contamination rate 1.75%. Even though the completeness is smaller, the MAG will be used to compare the results of the same species. KEGG identified that the majority of the found genes belong to purine metabolism, glycine, serine and threonine metabolism, aminoacyl-tRNA biosynthesis, and the two-component system. The only complete energy metabolism pathway was ATP synthesis (F-type ATPase), which was also found in MAG_22 Moreover, MAG_15 during the analysis only showed one complete cofactor and vitamin metabolism pathway - coenzyme A biosynthesis, which was also identified in the previous MAG. When the antimicrobial resistance genes were analyzed, it was found that MAG_15 has a few gene variants for tetracycline resistance beta-Lactam, and macrolide resistance. However, KEEG analysis on the previously sequenced genome identified beta-Lactam and Vancomycin resistance.

MAG_27

MAG_27 assigned to the phylum *Verrucomicrobia* (species *Opitutaceae bacterium*). This MAG showed completeness of 91.52% with 5.62% contamination rate. The genome has 934 contigs and the total length of 3,321,892 bp with N50 equal to 4,738. The first sequenced genome was isolated from the wood-feeding termite hindgu with chromosome length of 7,317,842 bp. The genome, which was found in 2015 revealed genes for methylotrophy, lignocellulose degradation, and ammonia and sulfate assimilation (Kotak et al., 2015). Interestingly, *Verrucomicrobia* represent the first phylum outside of the *Proteobacteria* to be characterized as methanotrophs. An analysis of the genome revealed the presence of two complete energy metabolism pathways: carbon fixation (phosphate acetyltransferase-acetate kinase pathway) and ATM synthesis (cytochrome c oxidase and F-type ATPase). KEGG analysis also identified three complete cofactor and vitamin metabolism pathways, including coenzyme A biosynthesis, biotin biosynthesis, and molybdenum cofactor biosynthesis. The genome revealed tetracycline resistance genes, ancomycin resistance modules, and multi drug efflux pump resistance modules.

Discussion

Cinachyrella sp. microbiome

The sponge-associated microbiota is host-specific (Reveillaud et al., 2014; Thomas et al., 2016). Host identity is a major factor in explaining the microbiome of the sponges as other studies showed before (e.g. Steinert et al., 2017). Thus, this study provides a deeper understanding of the metagenome of an individual *Cinachyrella sp.* sponge. For the analysis we used one sample, and we did not expect to describe the metagenome of divergent *Cinachyrella* taxa. The taxonomic analysis of this sample correlates with previous *Cinachyrella* metagenome research (Cuvelier et al., 2014). On the other hand, this paper presents the first *Cinachyrella* metagenome functional analysis with MAGs generation. Thus, additional samples with other *Cinachyrella* species would be needed to confirm the functional annotation of these sponge individuals, their similarities or differences.

Holobiont

As was discovered previously, microbiome can influence the host's behavior, metabolite production, reproduction, and immunity in either "microbe-specific" or "microbe-assisted" ways (Shropshire & Bordenstein, 2016). The "holobiont" and "hologenome" - interconnected compositions of animal and microbes, concepts gain more attention, since without this whole picture the host itself cannot be studied. Some elements of host immunity genes are in constant competition with microbiota genes. Host immunity needs to manage the symbionts community, while defending pathogenic infections. This coexistence provokes coevolutionary changes between the host genetics and its microbiome (Obbard et al., 2006). Holobionts can be colonized with microorganisms, by contacting with environmental and social sources (Tung et al., 2015), by host genetic and metabolic traits (McKnite et al., 2012), and diet and age (Yatsunenko et al., 2012). Some studies claim that the microbiota is an extra genetic component that stimulates species formation (Brucker & Bordenstein, 2012). All of these factors have to be considered in order to explain the microbial communities of holobionts (Brucker & Bordenstein, 2013). Our functional analysis established that all of the analyzed MAGs have antimicrobial resistance genes (eg. beta-lactam, trimethoprim, nitroimidazole, aminoglycoside, quinolone, macrolide, fosfomycin, fusidicacid, oxazolidinone, glycopeptide, tetracycline, rifampicin, phenicol, colistin, sulphonamide). This fact might result in the overall hosts coevolutionary changes. Full genome

sequencing of the host as well as its functional analysis for these genes would aid in clarifying this observation.

The major system driver

This study indicates that *Proteobacteria* is the major system driver at the phylum level, followed by *Chloroflexi*, *Acidobacteria*, and *Gammaproteobacteria*. Previous studies indicated that the *Cinachyrella* sp. microbiome was dominated by *Proteobacteria, Bacteroidetes,* and *Cyanobacteria* (Cuvelier et al., 2014; Thomas et al., 2016: Baquiran et al., 2020). This finding suggests that *Chloroflexi* bacterium was underestimated in previous studies, since the whole genome analysis indicated major abundance of *Ardenticatena* genome. High abundance of *Chloroflexi* members are more typical of high microbial abundance HMA species (Schmitt et al., 2012).

Intraspecific Diversity within the Sponge Microbiome

Recruitment plots suggest that different models of bacterial speciation may apply to the same host environment. Recruitment plots can reveal genetic heterogeneity within discrete population. Recruitment of reads can identify sequence-discrete populations (sequence-discrete "clusters") that are not clonal but instead are composed of highly similar, co-occurring genotypes that contain some degree of genetic diversity with highly conserve regions, represented by horizontal lines separated by percent similarity (Caro-Quintero & Konstantinidis, 2012). Moreover, some research suggests that these similar members of the community may have highly similar, if not identical, ecological roles (Caro-Quintero & Konstantinidis, 2012). Each bin identified a different amount of discrete populations (horizontal clusters). Especially the top two bins (MAG_2 and MAG_1) showed high number of these populations. MAG_1 has an average genome size, comparing to others, with lots of vertical and horizontal clusters. It might suggest that this MAG represents a group of discrete populations that have streamlined genome to core functions with each discrete population occupying a niche based on that small regions of unique genomic information. Possibly, this bin represents the most abundant MAG, because it supports the core functions, which are needed in order to survive in this environment, including the metabolism, defense and energy generation mechanisms. It's the most abundant organism with a high degree of genetic heterogeneity are likely to be able to occupy a larger number of niches, since it performs many core roles in the hologenome. If we take into the

consideration MAG_1, which is the second most abundant bin, it represents a large genome with a high number of sequence-discrete populations and a very low rate of shared homologous regions with other population. Therefore the MAG represents a single population with minimal heterogeneity, and is likely responsible for a particular function within this hologenome, since it is so abundant. Moreover, since other taxa has a very low rate of shared homologous regions, this bin might occupy a particular niche in the environment, which is not feasible for others. On the other hand, the third most abundant MAG_3 and MAG_2 didn't show higher intraspecific diversity speciation in the sample. MAG_3 is a large genome, which has only a few sequence-discrete "clusters", might suggest that the bin contains a wide range of functions that are not parsed over subpopulations, which are unique for the community. It might suggest its unique role in the community, as an indispensable community member.

Sulfur metabolism

Five MAGs are taxonomically affiliated with known sulfur utilizing microbes: *Opitutaceae bacterium*, *Thioalkalivibrio paradoxus*, *Desulfobacterium autotrophicum*, *Thioalkalivibrio sulfidiphilus*, *Sulfurifustis variabilis*, suggesting an important role of sulfur metabolism in sponge-microbiome.

Other research identified microorganisms involved in sulfur cycling by the functional gene *aprA* (adenosine-5′-phosphosulfate reductase) (Jensen et al., 2017). Samples with the *aprA* gene were found to be 87 % affiliated with sulfur-oxidizing bacteria. MAG_5 and MAG_17 showed the presence of this gene. Because more than 75% of the MAGs generated from the sponge microbiome contains PAPSS gene (2.7.7.4), which is the universal sulfonate donor's synthase and the Fig 5. shows that all of the MAGs have some of the sulfur metabolism genes. The findings indicate that assemalatory pathways are highly conservered, oxidataive and reductions pathways ways are less conserved across the MAGs.

Many articles state the importance of dimethylsulfoniopropionate (DMSP) as an osmoprotectant (Yoch D. 2002; Conway et al., 2012). This metabolite is degraded by marine bacteria to dimethylsulfide (DMS). On the other hand, out KEGG analysis did not identify any of the metabolites related to this pathway (dmdA, dddL).

Quorum Sensing

Bacteria use quorum sensing to regulate population density and control collective behavior. Quorum sensing is cell-to-cell communication through the production and release of chemical ques into the surrounding environment. Extracellular signaling molecules called autoinducers are produced and detected. Quorum sensing leads to a group-level response and change of the behavior. Autoinducers accumulate in the community space as the population density increases. Many processes are regulated by quorum sensing, including bioluminescence, the secretion of virulence factors, the formation of biofilms (Papenfort & Bassler, 2016). Quorum sensing plays an important role in successful establishment of symbiotic and pathogenic relationships with eukaryotic hosts (González & Keshavan, 2006)

A few recent articles focused on identification of quorum sensing activators and inhibitors in the marine sponges (Saurav et al., 2020). For example, sponge species Sarcotragus spinosulus expressed both QS signal molecules as well as QS inhibitory (QSI) molecules. Some articles show that some sponge species like *Sarcotragus sp.* have constant presence of *N*-acyl homoserine lactones (AHLs), quorum sensing molecules mediators (Britstein et al., 2018). Only several studies attempted to examine quorum sensing systems in sponge metagenome (Reen et al., 2019), and therefore more research needs to be done.

It was found that PQS plays a crucial role in adapting the population density and structure while populations adjust to the conditions of increased stress, which are accompanied by bacterial persistence, often a result of chronic infectious diseases (Häussler & Becker 2008). Opp/Ami is an ABC transporter (peptide importer). It plays a crucial role in internalization of different peptides and signal molecules during the population growth phase. This study provides a support of quorum mechanism importance and presence in *Cinachyrella* metagenome, including Opp/Ami gene identification (signaling peptide importer). Previous studies suggest that eukaryotes have evolved defense mechanisms to manipulate bacterial quorum sensing and by doing so protect themselves from pathogenic attack (González & Keshavan, 2006). Therefore, it is possible that this quorum sensing mechanism is used by metagenome to compete with each other for space and food sources. It was also found that some bacteria evolve gene-regulation with incorporates host quorum-sensing signal molecules. Further studies need to examine if it's true for *Cinachyrella* metagenome by analyzing host's genome.

Drug Discovery

Found sponge microbiomes may play a key role in drug discovery. For example, MAG_14, was identified as Haliangium ochraceu. Haliangium ochraceu is known for its ability to produce secondary metabolites. Moreover, this bacterium produces fruiting bodies and develops myxospores, that is based on cell-to-cell signaling among the single cells of the population in a swarm. Interestingly enough, many modules for MAG_14 were assigned as antimicrobial resistant, based on the found antimicrobial resistance gene, which identifies its ability to survive in highly competitive environments. Some of the modules included methicillin, beta-Lactam, vancomycin, tetracycline, and a variety of multi drug resistance of efflux pumps. Priority pathogens by WHO assigned as critical are 1. Acinetobacter baumannii, carbapenem-resistant 2. Pseudomonas aeruginosa, carbapenem-resistant 3. Enterobacteriaceae, carbapenem-resistant, ESBL-producing. Therefore, further studies need to examine if this finding possess the desirable pharmacokinetic properties required for clinical development.

Conclusion

In this study, we report the identification of 27 genomes of sponge associated microbes with at least 50% completeness, and contamination less than 10%. Providing insight into microbial speciation and *Cinachyrella* sponge microbiome. The results identified underestimated in previous studies Chloroflexi bacterium (*Ardenticatena*) as a major community driver.

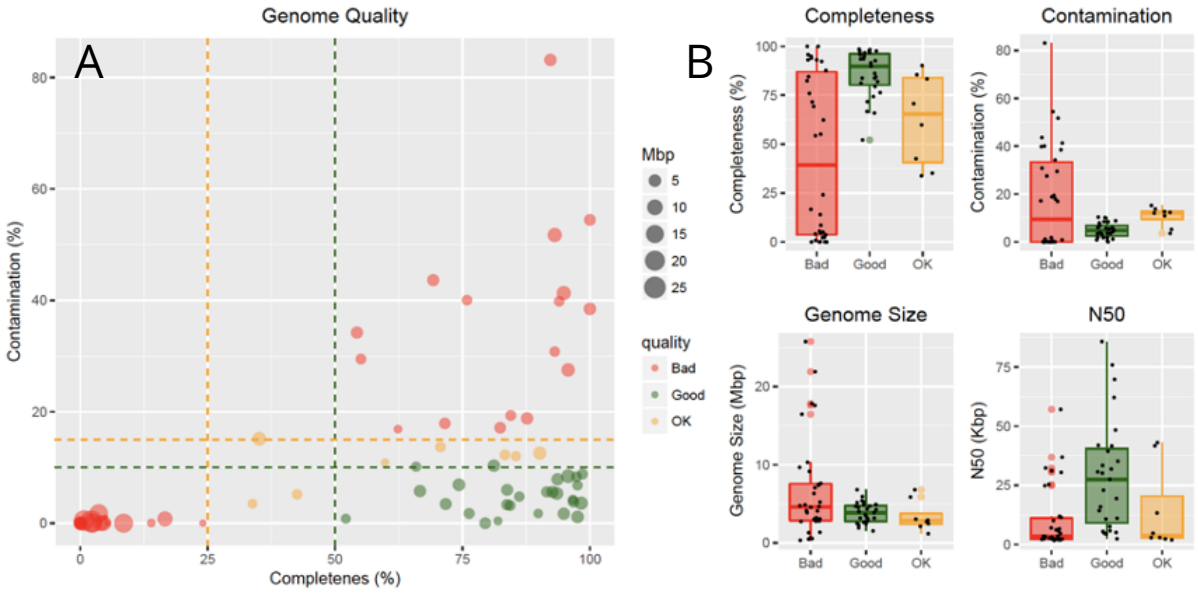
Results underlined abundant prokaryotic defense system for each MAG, including quorum sensing mechanism and whole metabolic pathways: CRISPR-Cas system, restriction and modification system, and toxin-antitoxin system. Some of the functional roles for key members of the *Cinachyrella* sponge microbiome identified the functional roles in the microbial-sponge symbiosis. Microbial populations exemplified by MAG_14 (*Proteobacteria*) and MAG_27 (*Verrucomicrobia*) probably are of global importance for their host sponges since they were predicted to function in carbon fixation. Moreover, this study provided a detailed *Cinachyrella* sponge microbiome functional characterization.

KEGG analysis showed that 21 out of 27 MAGs have PAPSS gene (2.7.7.4), which is the universal sulfonate donor's synthase. Moreover, detailed pathway analysis identified that MAG_14 and MAG_23 contain complete energy metabolism pathway for sulfur metabolism. Symbiotic bacteria in previous research show elevated expression of genes related to sulfur uptake and metabolism in symbiotic sample compared to the free-living state (Tsukada et al.,

2009). Our research underlined abundance of sulfur metabolism genes, what represents an important trait in symbiotic bacteria.

Three identified MAGs (5, 22, 23) showed many complete pathways for metabolism of cofactors and vitamins were recognized, including NAD biosynthesis, Coenzyme A biosynthesis, molybdenum cofactor biosynthesis, and siroheme biosynthesis. This finding suggests an importance of these mechanisms in survival of the microbial symbionts. Previous research suggests that vitamins and cofactors produced by diverse symbionts could be beneficial to the sponge host (Thomas et al., 2010). This finding correlates with other research and suggests that the sponges' nutrition is supplemented by symbiont-derived vitamins and cofactors (Bayer et al., 2018). Also, many modules were assigned as antimicrobial resistant, which might have notable potent therapeutic activities, and have the desirable properties required for drug development.

Figures:



C

Quality	Completeness	Contamination	No.Contigs	Mbp	N50.kbps
Bad :34	Min. : 0.00	Min. : 0.00	Min. : 43	Min. : 0.3132	Min. : 1.724
Good:27	1st Qu.: 33.81	1st Qu.: 0.80	1st Qu.: 342	1st Qu.: 2.7302	1st Qu.: 2.890
OK : 8	Median : 79.47	Median : 5.62	Median : 619	Median : 4.0046	Median : 6.999
NA	Mean : 62.62	Mean :11.95	Mean : 1510	Mean : 5.1649	Mean :18.056
NA	3rd Qu.: 93.10	3rd Qu.:15.17	3rd Qu.: 1489	3rd Qu.: 5.3828	3rd Qu.:30.695
NA	Max. :100.00	Max. :83.17	Max. :11000	Max. :25.7809	Max. :85.829

Figure 1: A) Genome Quality graph of 69 MAGs shows all good quality MAGs in green, OK quality MAGs in yellow, and bad quality MAGs in red. B) bar plots of completeness, contamination, genome size, and N50 of bad, good and OK quality MAGs. C) Statistical analysis table of all the MAGs and their qualitative characteristics, including quality, completeness, contamination, number of contigs, mbp, and N50.

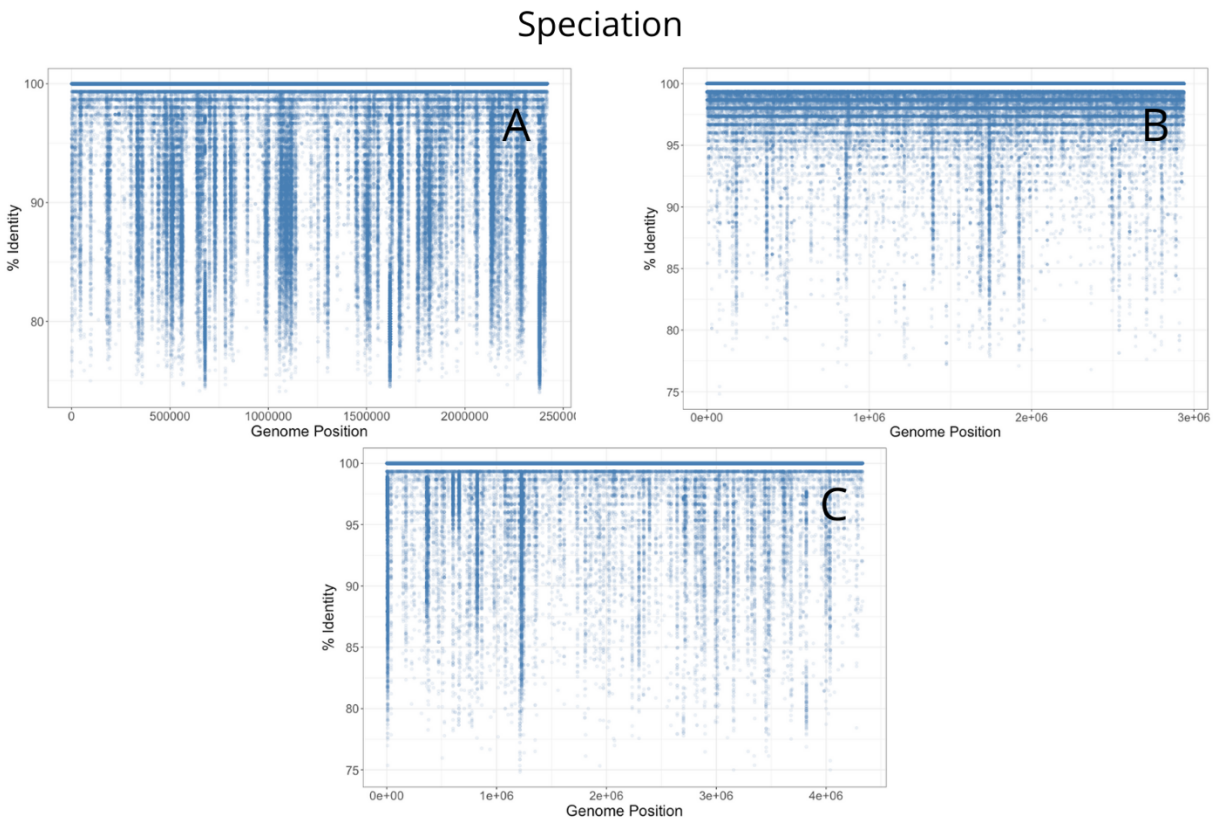


Figure 2. Recruitment plots of the most abundant three MAGs, based on the relative abundance, represent the reads, placed by genome position (X-axis) and identity percentage (Y-axis). All short reads from the sample metagenomes were searched against all contigs of the bin in a NCBI blast search. The recruitment plot displays the total number of short read-derived base-positions at given percent identities. Each dot represents a read. For the recruitment bit score >50 was used. A) MAG_2 B) MAG_1 C) MAG_3.

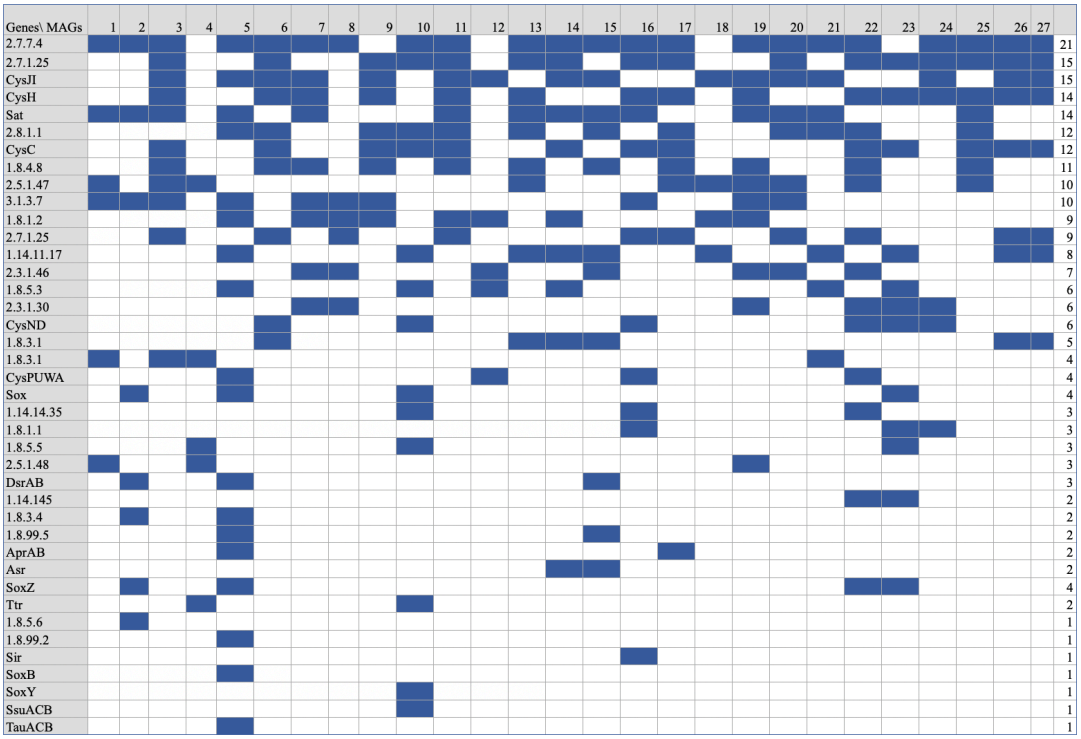


Figure 3. Sulfur metabolism functional potential of the 27 high quality metagenomically assembled genomes (MAGs) as deciphered by means of KEGG (Kyoto Encyclopedia of Genes and Genomes) modules. Colums represent MAGs; KEGG sulfur metabolism genes are arranged in rows. Blue color of the heatmap indicates the presence of identified genes. Last column shows the total number of represented gene in all MAGs.

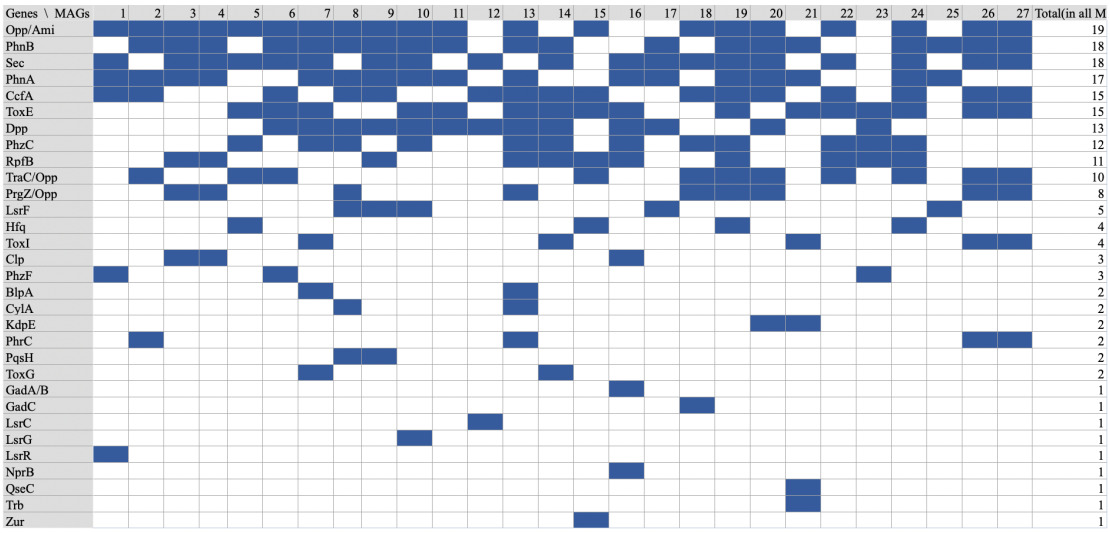


Fig 4. Quorum sensing functional potential of the 27 high quality metagenomically assembled genomes (MAGs) as deciphered by means of KEGG (Kyoto Encyclopedia of Genes and Genomes) modules. Colums represent MAGs; KEGG quorum sensing genes are arranged in rows. Blue color of the heatmap indicates the presence of identified genes. Last column shows the total number of represented gene in all MAGs.

Tables:

Table 1: Summary of found species and phylum of high-quality Metagenomically Assembled Genomes (MAGs) of *Cinachyrella* sponge sample. MAGs are sorted by Phylum.

BinID	SpeClosest species matchcies	PhyPhylumlum:
MAG27	*Opitutaceae bacterium TAV5*	*Verrucomicrobia*
MAG11	*Cenarchaeum symbiosum A*	*Thaumarchaeota*
MAG18	*Cenarchaeum symbiosum A*	*Thaumarchaeota*
MAG26	*Cenarchaeum symbiosum A*	*Thaumarchaeota*
MAG2	*"Thioalkalivibrio paradoxus"*	*Proteobacteria*
MAG3	*Nitrosomonas communis*	*Proteobacteria*
MAG9	*Desulfobacterium autotrophicum*	*Proteobacteria*
MAG10	*Rhodoplanes sp.*	*Proteobacteria*
MAG12	*Azospirillum brasilense*	*Proteobacteria*
MAG13	*Acidithiobacillus ferrivorans*	*Proteobacteria*
MAG14	*Haliangium ochraceum*	*Proteobacteria*
MAG16	*Thioalkalivibrio sulfidiphilus*	*Proteobacteria*
MAG20	*Pseudomonas citronellolis*	*Proteobacteria*
MAG23	*Rhizobiales bacterium*	*Proteobacteria*
MAG25	*Pseudomonas citronellolis*	*Proteobacteria*
MAG5	*Sulfurifustis variabilis*	*Proteobacteria*
MAG8	*Polymorphum gilvum*	*Proteobacteria*
MAG6	*Nitrospira defluvii*	*Nitrospirae*
MAG15	*Gemmatirosa kalamazoonesis*	*Gemmatimonadetes*
MAG22	*Gemmatirosa kalamazoonesis*	*Gemmatimonadetes*
MAG1	*Ardenticatena*	*Chloroflexi*
MAG21	*Caldilinea aerophila*	*Chloroflexi*
MAG19	*Simkania negevensis*	*Chlamydiae*

MAG4	*Ilumatobacter coccineus*	*Actinobacteria*
MAG17	*Mycolicibacterium chubuense*	*Actinobacteria*
MAG7	*Acidobacteria bacterium*	*Acidobacteria*
MAG24	*Chloracidobacterium thermophilum*	*Acidobacteria*

Table 2. The relative abundance of each MAG per Mb of genome. To estimate the relative abundance of each MAG, the total number of reads mapped to a MAG was divided by the total number of reads in the metagenome sample and later by the total genome size in Mb for the corresponding MAG.

Bin ID	Number reads mapped	Genome Size	Genome Size (mb)	Abundance per Mb of genome (%)
MAG_2	3803428	476295	0.476295	34.34%
MAG_1	12244319	3165170	3.16517	16.63%
MAG_3	3048256	2692900	2.6929	4.87%
MAG_20	657484	2277246	2.277246	1.24%
MAG_5	1445159	5203527	5.203527	1.19%
MAG_9	781205	2822265	2.822265	1.19%
MAG_14	647010	2441330	2.44133	1.14%
MAG_6	782632	3202857	3.202857	1.05%
MAG_10	1139182	4685360	4.68536	1.05%
MAG_7	589806	2456476	2.456476	1.03%
MAG_21	506869	2277246	2.277246	0.96%
MAG_15	621708	2898325	2.898325	0.92%
MAG_4	1086408	5149290	5.14929	0.91%
MAG_22	1369924	7376007	7.376007	0.80%
MAG_8	906089	4904957	4.904957	0.79%
MAG_16	810765	5382816	5.382816	0.65%
MAG_18	376686	4744898	4.744898	0.34%
MAG_24	427804	6326817	6.326817	0.29%
MAG_12	927703	17842486	17.842486	0.22%
MAG_19	219072	4744898	4.744898	0.20%
MAG_27	353051	9660623	9.660623	0.16%
MAG_26	332142	10309494	10.309494	0.14%

MAG_25	513534	16471589	16.471589	0.13%
MAG_17	539472	17601912	17.601912	0.13%
MAG_13	501869	17842486	17.842486	0.12%
MAG_11	571674	21917315	21.917315	0.11%
MAG_23	401038	25780865	25.780865	0.07%

Supplemental Tables:

Supplemental Table 1. Sequencing Statistics.

Supplemental Table 2. Functional potential of reconstructed MAGs with KEGG pathways. Functional assignments to KEGG categories revealed 282 functional categories.

Please refer to the Additional Files section in this thesis' record for supplemental data tables.

REFERENCES:

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K. L., Tyson, G. W. & Nielsen, P. H. (2013) Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat. Biotechnol. 31, 533–538

Al-Nakeeb K., Peterson, T.H., Sicheritz-Pontén T. (2017). Norgal: extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. BMC bioinfo. 510, 18. Doi: 10.1186/s12859-017-1927-y

Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995). Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiological reviews, 59(1), 143–169.

Anantharaman K., Hausmann B., Jungbluth S. P., Kantor R. S., Lavy A., Warren L. A., et al. (2018). Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. ISME J. 12 1715–1728. 10.1038/s41396-018-0078-0

Aziz, R.K., Bartels, D., Best, A.A. et al. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9, 75 (2008). https://doi.org/10.1186/1471-2164-9-75

Baquiran, J., Nada, M., Posadas, N., Manogan, D. P., Cabaitan, P. C., & Conaco, C. (2020). Population structure and microbial community diversity of two common tetillid sponges in a tropical reef lagoon. *PeerJ*, *8*, e9017. https://doi.org/10.7717/peerj.9017

Bayer, K., Jahn, M. T., Slaby, B. M., Moitinho-Silva, L., & Hentschel, U. (2018). Marine Sponges as Chloroflexi Hot Spots: Genomic Insights and High-Resolution Visualization of an Abundant and Diverse Symbiotic Clade. mSystems, 3(6), e00150-18. https://doi.org/10.1128/mSystems.00150-18

Baquiran, J., Nada, M., Posadas, N., Manogan, D. P., Cabaitan, P. C., & Conaco, C. (2020). Population structure and microbial community diversity of two common tetillid sponges in a tropical reef lagoon. PeerJ, 8, e9017. https://doi.org/10.7717/peerj.9017

Bowles M.W., Mogollon J.M., Kasten S., Zabel M., and Hinrichs K.U. (2014) Global rates of marine sulfate reduction and implications for sub-sea-floor metabolic activities. Science 344: 889–891.

Britstein, M., Saurav, K., Teta, R., Sala, G. D., Bar-Shalom, R., Stoppelli, N., Zoccarato, L., Costantino, V., & Steindler, L. (2018). Identification and chemical characterization of N-acyl-homoserine lactone quorum sensing signals across sponge species and time. FEMS microbiology ecology, 94(2), 10.1093/femsec/fix182. https://doi.org/10.1093/femsec/fix182

Brucker, R. M., & Bordenstein, S. R. (2012). Speciation by symbiosis. *Trends in ecology & evolution*, *27*(8), 443–451. https://doi.org/10.1016/j.tree.2012.03.011

Brucker, R. M., & Bordenstein, S. R. (2013). The capacious hologenome. *Zoology (Jena, Germany)*, *116*(5), 260–261. https://doi.org/10.1016/j.zool.2013.08.003

Burns, A., Padilla, C., Pratte, Z., Gilde, K., Regensburger, M., Hall, E., Dove, A. and Stewart, F. (2018). Broad Phylogenetic Diversity Associated with Nitrogen Loss through Sulfur Oxidation in a Large Public Marine Aquarium.

Cabello-Yeves, P. J., Zemskaya, T. I., Rosselli, R., Coutinho, F. H., Zakharenko, A. S., Blinov, V. V., & Rodriguez-Valera, F. (2017). Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal. Applied and environmental microbiology, 84(1), e02132-17. https://doi.org/10.1128/AEM.02132-17

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods. 7, 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. The ISME journal, 6(8), 1621-4.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U.S.A. 108(Suppl. 1), 4516–4522. doi: 10.1073/pnas.1000080107

Caro-Quintero A, Konstantinidis KT . (2012). Bacterial species may exist, metagenomics reveal. Environ Microbiol 14: 347–355.

Chaudhuri, R. R., Loman, N. J., Snyder, L. A., Bailey, C. M., Stekel, D. J., & Pallen, M. J. (2008). xBASE2: a comprehensive resource for comparative bacterial genomics. Nucleic acids research, 36(Database issue), D543–D546. https://doi.org/10.1093/nar/gkm928

Cleary, D., Polónia, A., Huang, Y. M., Putchakarn, S., Gomes, N., & de Voogd, N. J. (2019). A comparison of prokaryote communities inhabiting sponges, bacterial mats, sediment and seawater in Southeast Asian coral reefs. FEMS microbiology ecology, 95(12), fiz169. https://doi.org/10.1093/femsec/fiz169

Cleary, D. R., Polónia, A. R. M., Reijnen, B. T., Berumen, M. L., & de Voogd, N. J. (2020). Prokaryote Communities Inhabiting Endemic and Newly Discovered Sponges and Octocorals from the Red Sea. *Microbial Ecology*, 1-17.

Conway, C.A., Esiobu , N., Lopez, J.V. 2012 Co-cultures of Pseudomonas aeroginosa and Roseobacter denitrificans reveal shifts in gene expression levels compared to solo cultures . The Scientific World Journal. Vol 2012, Article ID 120108; http://www.tswj.com/2012/120108/

Cuvelier, M. L., Blake, E., Mulheron, R., McCarthy, P. J., Blackwelder, P., Thurber, R. L., & Lopez, J. V. (2014). Two distinct microbial communities revealed in the sponge Cinachyrella. Frontiers in microbiology, 5, 581. https://doi.org/10.3389/fmicb.2014.00581

DeBruyn JM, Nixon LT, Fawaz MN, Johnson AM, Radosevich M. (2011). Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. Appl Environ Microbiol 77:6295–6300. doi:10.1128/AEM.05005-11

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont T. (2015). Anvi'o: an advanced analysis and visualization platform for ‘omics data. PeerJ 3:e1319.doi:10.7717/peerj.1319.

Falkowski P.G., Fenchel T., DeLong E.F. (2008). The microbial engines that drive Earth's biogeochemical cycles. Science 320:1034–1039. doi:10.1126/science.1153213

Frishman, D., Mironov, A., Mewes, H. W., & Gelfand, M. (1998). Combining diverse evidence for gene recognition in completely sequenced bacterial genomes. *Nucleic acids research*, *26*(12), 2941–2947. https://doi.org/10.1093/nar/26.12.2941

Gilbert SF, Sapp J, Tauber AI. (2012). A symbiotic view of life: we have never been individuals. Q Rev Biol 87:325–341. doi:10.1086/668166.

González, J. E., & Keshavan, N. D. (2006). Messing with bacterial quorum sensing. Microbiology and molecular biology reviews : MMBR, 70(4), 859–875. https://doi.org/10.1128/MMBR.00002-06

Gutiérrez-Barranquero, J. A., Reen, F. J., Parages, M. L., McCarthy, R., Dobson, A., & O'Gara, F. (2019). Disruption of N-acyl-homoserine lactone-specific signalling and virulence in clinical pathogens by marine sponge bacteria. Microbial biotechnology, 12(5), 1049–1063. https://doi.org/10.1111/1751-7915.12867

Häussler S, Becker T (2008) The Pseudomonas Quinolone Signal (PQS) Balances Life and Death in Pseudomonas aeruginosa Populations. PLoS Pathog 4(9): e1000166. https://doi.org/10.1371/journal.ppat.1000166

Hentschel, U., Hopke, J., Horn, M., Friedrich, A. B., Wagner, M., Hacker, J.,et al. (2002). Molecular evidence for a uniform microbial community insponges from different oceans. Appl. Environ. Microbiol. 68, 4431–4440. doi:10.1128/AEM.68.9.4431-4440.2002

Hentschel U, Piel J, Degnan SM, Taylor MW. (2012). Genomic insights into the marine sponge microbiome. Nat Rev Microbiol 10:641–654. doi:10.1038/nrmicro2839.

Hirai, J., Nagai, S., & Hidaka, K. (2017). Evaluation of metagenetic community analysis of planktonic copepods using Illumina MiSeq: Comparisons with morphological classification and metagenetic analysis using Roche 454. PloS one, 12(7), e0181452. doi:10.1371/journal.pone.0181452

Human Microbiome Project Consortium (2012). Structure, function and diversity of the healthy human microbiome. Nature, 486(7402), 207–214. https://doi.org/10.1038/nature11234

Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome research*, *17*(3), 377–386. https://doi.org/10.1101/gr.5969107

Huson, D. H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H. J., & Tappu, R. (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS computational biology*, *12*(6), e1004957. https://doi.org/10.1371/journal.pcbi.1004957

Hyatt, D., Chen, G. L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC bioinformatics, 11, 119. https://doi.org/10.1186/1471-2105-11-119

Jensen, S., Fortunato, S. A., Hoffmann, F., Hoem, S., Rapp, H. T., Øvreås, L., & Torsvik, V. L. (2017). The Relative Abundance and Transcriptional Activity of Marine Sponge-Associated Microorganisms Emphasizing Groups Involved in Sulfur Cycle. Microbial ecology, 73(3), 668–676. https://doi.org/10.1007/s00248-016-0836-3

Jørgensen, B. B., Findlay, A. J., & Pellerin, A. (2019). The Biogeochemical Sulfur Cycle of Marine Sediments. Frontiers in microbiology, 10, 849. https://doi.org/10.3389/fmicb.2019.00849

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. Journal of molecular biology, 428(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Konstantinidis KT, DeLong EF . (2008). Genomic patterns of recombination, clonal divergence and environment in marine microbial populations. ISME J 2: 1052–1065.

McKane M, Milkman R. (1995). Transduction, re- striction andrecombinationpatterns in Escherichia coli. Genetics 139:35–43.

Kotak, M., Isanapong, J., Goodwin, L., Bruce, D., Chen, A., Han, C. S., Huntemann, M., Ivanova, N., Land, M. L., Nolan, M., Pati, A., Woyke, T., & Rodrigues, J. L. (2015). Complete Genome Sequence of the Opitutaceae Bacterium Strain TAV5, a Potential

Facultative Methylotroph of the Wood-Feeding Termite Reticulitermes flavipes. Genome announcements, 3(2), e00060-15. https://doi.org/10.1128/genomeA.00060-15

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., and Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl. Environ. Microbiol. 79, 5112–5120. doi: 10.1128/AEM.01043-13

Lloyd-Price, J., Mahurkar, A., Rahnavard, G., Crabtree, J., Orvis, J., Hall, A. B., Brady, A., Creasy, H. H., McCracken, C., Giglio, M. G., McDonald, D., Franzosa, E. A., Knight, R., White, O., & Huttenhower, C. (2017). Strains, functions and dynamics in the expanded Human Microbiome Project. Nature, 550(7674), 61–66. https://doi.org/10.1038/nature23889

Lutz, H. L., Jackson, E. W., Webala, P. W., Babyesiza, W. S., Kerbis Peterhans, J. C., Demos, T. C., Patterson, B. D., & Gilbert, J. A. (2019). Ecology and Host Identity Outweigh Evolutionary History in Shaping the Bat Microbiome. *mSystems*, *4*(6), e00511-19. https://doi.org/10.1128/mSystems.00511-19

Magoč, T., and Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963. doi: 10.1093/bioinformatics/btr507

McCutcheon, J. P., & von Dohlen, C. D. (2011). An interdependent metabolic patchwork in the nested symbiosis of mealybugs. Current biology : CB, 21(16), 1366-72.

McKnite, A. M., Perez-Munoz, M. E., Lu, L., Williams, E. G., Brewer, S., Andreux, P. A., Bastiaansen, J. W., Wang, X., Kachman, S. D., Auwerx, J., Williams, R. W., Benson, A. K., Peterson, D. A., & Ciobanu, D. C. (2012). Murine gut microbiota is defined by host genetics and modulates variation of metabolic traits. *PloS one*, *7*(6), e39191. https://doi.org/10.1371/journal.pone.0039191

McFall-Ngai, M., Heath-Heckman, E. A., Gillette, A. A., Peyer, S. M., & Harvie, E. A. (2011). The secret languages of coevolved symbioses: insights from the Euprymna scolopes-Vibrio fischeri symbiosis. Seminars in immunology, 24(1), 3-8.

Nurk, S., Bankevich, A., Antipov, D., Gurevich, A. A., Korobeynikov, A., Lapidus, A., et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J. Comput. Biol. 20, 714–737. doi: 10.1089/cmb.2013.0084

Obbard, D. J., Jiggins, F. M., Halligan, D. L., & Little, T. J. (2006). Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Current biology : CB*, *16*(6), 580–585. https://doi.org/10.1016/j.cub.2006.01.065

Pace, N.R., Stahl, D.A., Lane, D.J. and Olsen, G.J. (1985) Analyzing Natural Microbial Populations by rRNA Sequences. ASM News, 51, 4-12.

Padilla, C.C., Ganesh, S., Gnatt, S., Huhman, A., Parris, D.J., Sarode, N., Stewart, F.J. (2015). Standard filtration practices may significantly distort planktonic microbial diversity estimates. Front Microbio. 6,547. doi: 10.3389/fmicb.2015.00547

Papenfort, K., & Bassler, B. L. (2016). Quorum sensing signal-response systems in Gram-negative bacteria. Nature reviews. Microbiology, 14(9), 576–588. https://doi.org/10.1038/nrmicro.2016.89

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. doi: 10.1101/gr.186072.114

Riesenfeld, C. S., Schloss, P. D., and Handelsman, J. (2004). "Metagenomics: Genomic analysis of microbial communities," Annu. Rev. Genet. 38, 525-552. DOI: 10.1146/annurev.genet.38.072902.091216

Prescott, S. L. (2017). History of medicine: Origin of the term microbiome and why it matters. Human Microbiome Journal, 4, 24-25. doi:10.1016/j.humic.2017.05.004

Saurav, K., Borbone, N., Burgsdorf, I., Teta, R., Caso, A., Bar-Shalom, R., Esposito, G., Britstein, M., Steindler, L., & Costantino, V. (2020). Identification of Quorum Sensing Activators and Inhibitors in The Marine Sponge Sarcotragus spinosulus. Marine drugs, 18(2), 127. https://doi.org/10.3390/md18020127

Quail, M. A., Swerdlow, H., & Turner, D. J. (2009). Improved protocols for the illumina genome analyzer sequencing system. Current protocols in human genetics, Chapter 18, 10.1002/0471142905.hg1802s62. https://doi.org/10.1002/0471142905.hg1802s62

Rigaud, T., & Juchault, P. (1993). Conflict between feminizing sex ratio distorters and an autosomal masculinizing gene in the terrestrial isopod Armadillidium vulgare Latr. Genetics, 133(2), 247-52.

Richardson, E. J., & Watson, M. (2013). The automatic annotation of bacterial genomes. *Briefings in bioinformatics*, *14*(1), 1-12.

Shropshire, J. D., & Bordenstein, S. R. (2016). Speciation by Symbiosis: the Microbiome and Behavior. *mBio*, *7*(2), e01785. https://doi.org/10.1128/mBio.01785-15

Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., Deboy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., Madupu, R., … Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences of the United States of America, 102(39), 13950–13955. https://doi.org/10.1073/pnas.0506758102

Thompson, L. R., Williams, G. J., Haroon, M. F., Shibl, A., Larsen, P., Shorenstein, J., Knight, R., & Stingl, U. (2017). Metagenomic covariation along densely sampled environmental gradients in the Red Sea. The ISME journal, 11(1), 138–151. https://doi.org/10.1038/ismej.2016.99

Tsukada, S., Aono, T., Akiba, N., Lee, K. B., Liu, C. T., Toyazaki, H., & Oyaizu, H. (2009). Comparative genome-wide transcriptional profiling of Azorhizobium caulinodans ORS571 grown under free-living and symbiotic conditions. Applied and environmental microbiology, 75(15), 5037–5046. https://doi.org/10.1128/AEM.00398-09

Tung, J., Barreiro, L. B., Burns, M. B., Grenier, J. C., Lynch, J., Grieneisen, L. E., Altmann, J., Alberts, S. C., Blekhman, R., & Archie, E. A. (2015). Social networks predict gut microbiome composition in wild baboons. *eLife*, *4*, e05224. https://doi.org/10.7554/eLife.05224

Langille M., Zaneveld J., Caporaso J., McDonald D., Knights D., Reyes J., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31 814–821. 10.1038/nbt.2676

Lee, Y. K., & Mazmanian, S. K. (2010). Has the microbiota played a critical role in the evolution of the adaptive immune system?. Science (New York, N.Y.), 330(6012), 1768-73.

Lifemap-ncbi.univ-lyon1.fr. (2018). Lifemap ncbi. [online] Available at: http://lifemap-ncbi.univ-lyon1.fr [Accessed 18 Dec. 2018].

Pita, L., Rix, L., Slaby, B. M., Franke, A., & Hentschel, U. (2018). The sponge holobiont in a changing ocean: from microbes to ecosystems. Microbiome, 6(1), 46. https://doi.org/10.1186/s40168-018-0428-1

Reen, F. J., Gutiérrez-Barranquero, J. A., McCarthy, R. R., Woods, D. F., Scarciglia, S., Adams, C., Fog Nielsen, K., Gram, L., & O'Gara, F. (2019). Quorum Sensing Signaling Alters Virulence Potential and Population Dynamics in Complex Microbiome-Host Interactomes. Frontiers in microbiology, 10, 2131. https://doi.org/10.3389/fmicb.2019.02131

Reveillaud, J., Maignien, L., Murat Eren, A., Huber, J. A., Apprill, A., Sogin, M. L., & Vanreusel, A. (2014). Host-specificity among abundant and rare taxa in the sponge microbiome. The ISME journal, 8(6), 1198–1209. https://doi.org/10.1038/ismej.2013.227

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., Perez, T., Rodrigo, A., Schupp, P. J., Vacelet, J., Webster, N., Hentschel, U., & Taylor, M. W. (2012). Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. The ISME journal, 6(3), 564–576. https://doi.org/10.1038/ismej.2011.116

Schmidt, T. M., DeLong, E. F., & Pace, N. R. (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. Journal of bacteriology, 173(14), 4371–4378. https://doi.org/10.1128/jb.173.14.4371-4378.1991

Schuster, A., Lopez, J. V., Becking, L. E., Kelly, M., Pomponi, S. A., Wörheide, G., Erpenbeck, D., & Cárdenas, P. (2017). Evolution of group I introns in Porifera: new evidence for intron mobility and implications for DNA barcoding. BMC evolutionary biology, 17(1), 82. https://doi.org/10.1186/s12862-017-0928-9

Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. (2014);30:2068–9. doi: 10.1093/bioinformatics/btu153.

Simister, R. L., Deines, P., Botté, E. S., Webster, N. S., & Taylor, M. W. (2012). Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environmental Microbiology*, *14*(2), 517-524

Smith, E.C. (2013). De novo Transcriptome Analysis of the Marine Sponge Cinachyrella spp: A Potential Model Organism for Oil and Dispersant Ecotoxicology.

Shropshire, J. D., & Bordenstein, S. R. (2016). Speciation by Symbiosis: the Microbiome and Behavior. mBio, 7(2), e01785. doi:10.1128/mBio.01785-15

Slaby, B. M., Hackl, T., Horn, H., Bayer, K., & Hentschel, U. (2017). Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. The ISME journal, 11(11), 2465.

Steinert, G., Busch, K., Bayer, K., Kodami, S., Arbizu, P. M., Kelly, M., Mills, S., Erpenbeck, D., Dohrmann, M., Wörheide, G., Hentschel, U., & Schupp, P. J. (2020). Compositional and Quantitative Insights Into Bacterial and Archaeal Communities of South Pacific Deep-Sea Sponges (Demospongiae and Hexactinellida). Frontiers in microbiology, 11, 716. https://doi.org/10.3389/fmicb.2020.00716

Steinert G, Rohde S, Janussen D, Blaurock C, Schupp PJ. Host-specific assembly of sponge-associated prokaryotes at high taxonomic ranks. Sci Rep. 2017;7(1):2542. Published 2017 May 31. doi:10.1038/s41598-017-02656-6

Taylor, M. W., Tsai, P., Simister, R. L., Deines, P., Botte, E., Ericson, G., Schmitt, S., & Webster, N. S. (2013). 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. The ISME journal, 7(2), 438–443. https://doi.org/10.1038/ismej.2012.11

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J. R., Easson, C., Astudillo-García, C., Olson, J. B., Erwin, P. M., López-Legentil, S., Luter, H., Chaves-Fonnegra, A., Costa, R., Schupp, P. J., Steindler, L., Erpenbeck, D., Gilbert, J., Knight, R., Ackermann, G., Victor Lopez, J., Taylor, M. W., Thacker, R. W., Montoya, J. M., Hentschel, U., … Webster, N.

S. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. Nature communications, 7, 11870. doi:10.1038/ncomms11870

Thomas, T., Rusch, D., DeMaere, M. Z., Yung, P. Y., Lewis, M., Halpern, A., Heidelberg, K. B., Egan, S., Steinberg, P. D., & Kjelleberg, S. (2010). Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. The ISME journal, 4(12), 1557–1567. https://doi.org/10.1038/ismej.2010.7

Vavourakis, C. D., Mehrshad, M., Balkema, C., van Hall, R., Andrei, A. Ş., Ghai, R., Sorokin, D. Y., & Muyzer, G. (2019). Metagenomes and metatranscriptomes shed new light on the microbial-mediated sulfur cycle in a Siberian soda lake. BMC biology, 17(1), 69. https://doi.org/10.1186/s12915-019-0688-7

Verbeke, F., De Craemer, S., Debunne, N., Janssens, Y., Wynendaele, E., Van de Wiele, C., & De Spiegeleer, B. (2017). Peptides as Quorum Sensing Molecules: Measurement Techniques and Obtained Levels In vitro and In vivo. Frontiers in neuroscience, 11, 183. https://doi.org/10.3389/fnins.2017.00183

Wilkins, L., Leray, M., O'Dea, A., Yuen, B., Peixoto, R. S., Pereira, T. J., Bik, H. M., Coil, D. A., Duffy, J. E., Herre, E. A., Lessios, H. A., Lucey, N. M., Mejia, L. C., Rasher, D. B., Sharp, K. H., Sogin, E. M., Thacker, R. W., Vega Thurber, R., Wcislo, W. T., Wilbanks, E. G., … Eisen, J. A. (2019). Host-associated microbiomes drive structure and function of marine ecosystems. *PLoS biology*, *17*(11), e3000533. https://doi.org/10.1371/journal.pbio.3000533

Wooley JC, Godzik A, Friedberg I. (2010). A primer on metagenomics. PLoS Comput Biol. 2010;6(2):1000667. doi: 10.1371/journal.pcbi.1000667.

Wu, Y. W., Tang, Y. H., Tringe, S. G., Simmons, B. A., & Singer, S. W. (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome*, *2*, 26. https://doi.org/10.1186/2049-2618-2-26

Yada, S., Kamalesh, B., Sonwane, S., Guptha, I., & Swetha, R. K. (2015). Quorum sensing inhibition, relevance to periodontics. Journal of international oral health : JIOH, 7(1), 67–69.

Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R. N., Anokhin, A. P., Heath, A. C., Warner, B., Reeder, J., Kuczynski, J., Caporaso, J. G., Lozupone, C. A., Lauber, C., Clemente, J. C., Knights, D., Knight, R., … Gordon, J. I. (2012). Human gut microbiome viewed across age and geography. *Nature*, *486*(7402), 222–227. https://doi.org/10.1038/nature11053

Yoch D. C. (2002). Dimethylsulfoniopropionate: its sources, role in the marine food web, and biological degradation to dimethylsulfide. *Applied and environmental microbiology*, *68*(12), 5804–5815. https://doi.org/10.1128/aem.68.12.5804-5815.2002

Youngblut, N. D., Reischer, G. H., Walters, W., Schuster, N., Walzer, C., Stalder, G., Ley, R. E., & Farnleitner, A. H. (2019). Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nature communications*, *10*(1), 2200. https://doi.org/10.1038/s41467-019-10191-3